LIGAND-PROTEIN INTERACTIONS OF INHIBITORS WITH DENGUE VIRUS TYPE 2 SERINE PROTEASE: A STRUCTURAL STUDY

ROZANA OTHMAN

THESIS SUBMITTED IN FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

KUALA LUMPUR

2010

ABSTRACT

Dengue is a serious infectious disease that is endemic in over 100 countries. There has been an estimate of 50 million infection per year globally, with more than 2.5 billion people are at risk for epidemic transmission. Two principal illnesses associated with dengue are Dengue Fever (DF) and Dengue Haemorrhagic Fever (DHF). To date, there is no licensed vaccine or therapeutic drug available for these illnesses, although there have been reports of some vaccine candidates in clinical trials. Treatments have only been supportive thus far. This thesis describes part of our ongoing effort to search for a lead therapeutic agent for DF/DHF.

The early phase of this study involved the attempts to crystallize DEN-2 NS2B-NS3 protease complex. Optimization of the protein expression and purification procedures was carried out to yield protein purity suitable for crystallization trials. Protein crystallization techniques involved the hanging drop and sitting drop methods. The crystallization attempts were unsuccessful mainly due to the dynamic nature of the dengue protease which underwent auto-cleavage upon folding into its active conformation and produced degenerative products. The crystallized protease of DEN-2, however, was published in 2006 by D'Arcy *et al.*

Following the report of the crystallized protease, we performed validation of the *in silico* method that we established in predicting and building of the secondary structure profile of NS3 protease. The approach adopted was able to yield good prediction results and can be very useful, especially for cases where there are low homology sequence relationships between query and template proteins.

Subsequent experiments involved the computational docking of six non-competitive and one competitive inhibitor onto the DEN-2 protease complex. For the non-competitive inhibitors (ligands), blind docking was performed on the rigid ligand structures followed by flexible-ligand docking. Results obtained showed Lys 74 to be the most important residue in the binding site for interaction with the non-competitive inhibitors. Other residues were involved in the interactions with the inhibitors via H-bonds, van der Waals and hydrophobic interactions. Structure activity relationship study yielded the important features of the inhibitors that are responsible for activities to be the rigid structure of flavanone, the presence of the C5 hydroxyl and C7 methoxy groups on ring A, and the phenyl ring (B) in the molecules.

Computational docking of the competitive inhibitor, 4-hydroxypanduratin A, onto the protease active site, was done followed by QM/MM study using the ONIOM2 method. Results obtained proposed that the binding of the inhibitor to the active site to be mediated by H-bonding to Ser135. Structural investigation of the protein-inhibitor complex illustrated that the inhibitor was able to block the S1 pocket of the binding site, thus inhibiting the recognition of the pocket for two basic amino acids at the P1 position of a substrate. This blockade seemed to disable the substrate from entering the S1 pocket, rendering the protease inactive.

Information obtained from this study will be useful for the design of potential anti-dengue therapeutic agents.

(483 words)

ABSTRAK

Denggi adalah suatu penyakit berjangkit yang serius dan endemik di lebih dari 100 buah negara. Adalah dianggarkan 50 juta jangkitan setiap tahun secara global, dengan lebih dari 2.5 billion orang berisiko untuk transmisi epidemik. Dua penyakit utama dikaitkan dengan denggi adalah demam denggi (DF) dan demam denggi berdarah (DHF). Sehingga kini, tiada vaksin berlesen atau ubat terapeutik bagi penyakit-penyakit tersebut, walaupun terdapat laporan tentang beberapa calon vaksin dalam percubaan klinikal. Selama ini, rawatan adalah berbentuk rawatan sokongan sahaja. Tesis ini menerangkan sebahagian dari usaha berterusan kami untuk mencari satu agen terapeutik 'lead' bagi DF/DHF.

Bahagian awal kajian ini melibatkan usaha untuk menghablur kompleks proteas NS2B-NS3 DEN-2. Prosedur bagi ekspresi dan penulenan protein telah dioptimumkan bagi menghasilkan ketulenan protein yang sesuai untuk percubaan penghabluran. Teknik-teknk penghabluran protein melibatkan kaedah-kaedah 'hanging drop' dan 'sitting drop'. Percubaan-percubaan tersebut tidak berjaya menghasilkan hablur disebabkan oleh sifat semulajadi proteas denggi yang dinamik, yang mana protein tersebut mengalami pemotongan-auto sebaik sahaja ia melipat kepada konformasi aktif, lantas menghasilkan produk-produk degeneratif. Walaubagaimanapun, hablur protease bagi DEN-2 telah dilapurkan pada tahun 2006 oleh D'Arcy *et al.*

Susulan dari laporan tentang protease yang telah dihablur, pengesahan kaedah *in silico* yang telah dibina dijalankan untuk menjangkakan dan membina profil struktur sekunder bagi protease NS3. Pendekatan yang diambil telah dapat memberikan hasil jangkaan yang baik dan boleh menjadi sangat berguna, terutama bagi kes-kes di mana

terdapat hubungkait jujukan homologi yang rendah di antara protein yang dikaji dengan templat.

Kajian-kajian berikutnya melibatkan ‘docking’ berkomputer bagi enam perencat tak-bersaing dan satu perencat bersaing ke atas kompleks proteas DEN-2. Bagi perencat-perencat (ligan) tak-bersaing, ‘blind docking’ telah dijalankan ke atas struktur tegar bagi ligan, diikuti dengan ‘flexible-ligand docking’. Hasil-hasil yang diperolehi menunjukkan bahawa Lys74 adalah residu paling penting dalam tapak pengikatan untuk interaksi dengan perencat-perencat tak-bersaing. Residu-residu lain terlibat dalam interaksi dengan perencat-perencat melalui ikatan-H, interaksi van der Waals dan interaksi hidrofobik. Kajian hubungkait aktiviti-struktur menghasilkan ciri-ciri penting yang bertanggungjawab bagi aktiviti perencat iaitu: struktur tegar bagi flavanon, kehadiran kumpulan hidroksil bagi C5 dan kumpulan metoksi C7 pada gelang A, dan gelang (B) fenil.

‘Docking’ berkomputer bagi perencat bersaing, 4-hidroksipanduratin A, ke atas tapak aktif proteas, telah dijalankan diikuti dengan kajian QM/MM menggunakan kaedah ONIOM2. Hasil-hasil yang diperolehi mengesyorkan bahawa pengikatan perencat kepada tapak aktif adalah melalui ikatan-H kepada Ser135. Kajian struktur bagi kompleks protein-perencat menunjukkan bahawa perencat tersebut dapat menghalang poket S1 dalam tapak pengikatan, seterusnya menghindarkan pengenalpastian oleh poket tersebut terhadap dua asid amino berbes di kedudukan P1 pada suatu substrat. Halangan ini nampaknya menghindar substrat dari masuk ke dalam poket S1, menyebabkan proteas menjadi tidak aktif. Informasi yang diperolehi dari kajian ini adalah berguna bagi merekabentuk agen-agen terapeutik anti-denggi yang berpotensi.

Acknowledgements

I would like to express my special thanks and deepest gratitude to both my supervisors, Prof Dr Noorsaadah Abd Rahman and Prof Dr Rohana Yusof, for their kind supervision, guidance, patience, understanding and trust throughout the course of this study. Special thanks are also due to Assoc Prof Dr Habibah Abd Wahab from USM, for her kind help and expertise, and also for giving access to the facilities in her computer lab. I also would like to thank Prof David Rice of the Kreb's Institute, the University of Sheffield, for being a good supervisor and teacher during my six month attachment in his lab.

Special thanks to my lab mates especially Yean Kee, Hwee Ying, Teoh Kim Tat, Hadinur, Shatrah, Mudiana and Azahemy. Also to others in the Drug Design and Development Research Group (DDDRG), I thank you. I will cherish all the kind help that was given to me, and for making me laugh during hard times. To Dr. Tan Siew Kiat, thank you so much for your work has enabled me to continue with my study.

Not to forget, my appreciation is also for my lab friends from the Department of Molecular Biology and Biotechnology, the University of Sheffield, for their help and guidance during my attachment. In particular, to Berrisford for his help around the lab, and to Natalie and Ling, for sharing happy and sweet memories with me. My deepest appreciation is also due to my flat mates and others who had made my stay in Sheffield easier and full of memories. To Kak Min, thank you for everything; you have done it and now it is my turn.

My appreciation also goes to friends from USM, UPM, UKM and UiTM for the friendship, useful discussions and ideas. Special thanks to all staff from Universiti Malaya and the University of Sheffield for their assistance. I would also like to acknowledge the financial support provided by Universiti Malaya, Academy of Science Malaysia, and Ministry of Science, Technology and Innovation.

Finally, I owe very much to my family; my husband, Hermanto, and my children, Danial, Syauqee and Nabeel. You have been my main inspiration to persevere and complete my PhD project. To my parents, special thanks for all the guidance and teachings throughout my growing years.

And last, though not least, Alhamdulillah.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Docking of Noncompetitive Inhibitors into Dengue Virus Type 2 Protease: Understanding the Interactions with Allosteric Binding Sites

Rozana Othman,[†] Tan Siew Kiat,[§] Norzulaani Khalid,[||] Rohana Yusof,[‡] E. Irene Newhouse,[#] James S. Newhouse,[▽] Masqudul Alam,[○] and Noorsaadah Abdul Rahman*[,⊥]

Pharmacy Department and Department of Molecular Medicine, Faculty of Medicine and Institute of Biological Sciences and Chemistry Department, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia, Sunway University College, Bandar Sunway, 46150 Petaling Jaya, Malaysia, Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii at Manoa, 2565 McCarthy Mall, Keller 319, Honolulu, Hawaii 96822, Maui High Performance Computing Center, 550 Lipoa Parkway, Kihei, Hawaii 96753, and Department of Microbiology, University of Hawaii at Manoa, 2538 McCarthy Mall, Snyder 111, Honolulu, Hawaii 96822

Received October 26, 2007

A group of flavanones and their chalcones, isolated from *Boesenbergia rotunda* L., were previously reported to show varying degrees of noncompetitive inhibitory activities toward Dengue virus type 2 (Den2) protease. Results obtained from automated docking studies are in agreement with experimental data in which the ligands were shown to bind to sites other than the active site of the protease. The calculated K_i values are very small, indicating that the ligands bind quite well to the allosteric binding site. Greater inhibition by pinostrobin, compared to the other compounds, can be explained by H-bonding interaction with the backbone carbonyl of Lys74, which is bonded to Asp75 (one of the catalytic triad residues). In addition, structure−activity relationship analysis yields structural information that may be useful for designing more effective therapeutic drugs against dengue virus infections.

INTRODUCTION

Dengue virus belongs to the *Flaviviridae* family and is a widespread human pathogen that can cause diseases ranging from a harmless flulike illness to a severe hemorrhagic fever with high mortality rate, especially in children.[1] Dengue infections place some 2.5 billion people (or 40% of the world population) at risk and are a significant cause of mortality, especially in the tropical and subtropical regions,[2] and have been on the rise globally in recent years. In Malaysia, the Health Ministry's Parliamentary Secretary reported that dengue fever killed 44 people in the first four months of 2007, and a record number of suspected dengue cases occurred in 2007, with 900 cases in the first week of June alone. This is an increase of more than 100% compared to the same period of 2006.[3] Management of dengue fever is largely supportive, while its more severe hemorrhagic manifestation may require blood transfusions.

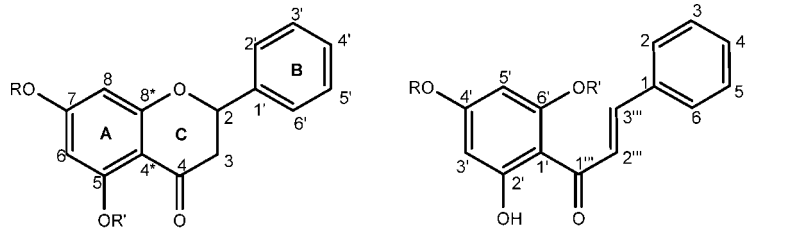
The most prevalent of the four dengue serotypes is Dengue virus type 2 (Den2), which contains a single-stranded RNA of positive polarity. The RNA genome codes for a single polyprotein precursor of 3,391 amino acids, arranged in the order C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5, comprising three structural and seven nonstructural proteins.[4] Flavivirus replication is dependent upon the correct cleavage of this polypeptide and requires both host cell proteases and the virus-encoded, two-component protease, NS2B-NS3.[5,6] NS3 contains a trypsin-like serine proteinase domain of 180 amino acid residues at its N-terminal, suggesting its role as the putative viral protease.[7,8] Analysis of virus sequence alignments indicates the catalytic triad of Den2 protease is His51, Asp75, and Ser135.[9] The N-termini of several nonstructural proteins are produced through cleavage by NS2B-NS3 at dibasic sites. Optimal catalytic activity of NS3 depends on the presence of NS2B. Biochemical studies and deletion analyses have mapped the required region on the NS2B to a central, hydrophilic 40 amino acid domain (Lys54-Glu93) in an otherwise relatively hydrophobic protein.[6,10,11] Hence, the NS2B-NS3 protease complex serves as a target in the development of antiviral drugs.

In spite of the efforts of many research groups,[12–17] no vaccines or antiviral drugs are currently available against dengue viral infections.[18] Thus, there is an immense ongoing interest in developing new antiviral therapeutic agents to fight diseases caused by dengue viruses.

Many research groups worldwide search for antiviral therapeutic agents from natural products. Nevertheless, this approach continues to furnish investigators with new and interesting findings. Several compounds from *Boesenbergia rotunda* (L.) Mansf. Kulturpfl. have shown noncompetitive inhibitory activity against the ability of Den2 protease to cleave fluorogenic peptide substrates.[12,19] *Boesenbergia rotunda* (L.) belongs to the ginger family (Zingiberaceae)

* Corresponding author phone: +603 7967 4254; fax: +603 7967 4193; e-mail: noorsaadah@um.edu.my.
[†] Pharmacy Department, Faculty of Medicine, Universiti Malaya.
[§] Sunway University College.
[||] Institute of Biological Sciences, Faculty of Science, Universiti Malaya.
[‡] Department of Molecular Medicine, Faculty of Medicine, Universiti Malaya.
[#] Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii at Manoa.
[▽] Maui High Performance Computing Center.
[○] Department of Microbiology, University of Hawaii at Manoa.
[⊥] Chemistry Department, Faculty of Science, Universiti Malaya.

10.1021/ci700388k CCC: $40.75
Published on Web 07/26/2008



R = OMe, R' = OH : Pinostrobin (**1**)
R = OH, R' = OH : Pinocembrin (**2**)
R = OH, R' = OMe : Alpinetin (**3**)

R = OMe, R' = OH : Pinostrobin chalcone (**4**)
R = OH, R' = OH : Pinocembrin chalcone (**5**)
R = OH, R' = OMe : Cardamonin (**6**)

Figure 1. Structures of the compounds extracted from *Boesenbergia rotunda* L.

and is a spice commonly used in Southeast Asia, especially in Malaysia and Indonesia. Some reports have indicated it has medicinal properties, and it has been used traditionally, mainly for problems and diseases of women, such as in postpartum protective medication, treatment for rheumatism, and as a tonic/lotion.[20] Scientific reports on the anti-inflammatory and anti-HIV activities of extracts from this plant have also been published.[21–23]

In this study, we report automated docking studies performed on the compounds which exhibited noncompetitive inhibitory activities toward NS2B-NS3 of Den2, using AutoDock 3.0.5 and Glide[24] software. The subjects of this study are the flavanones: pinostrobin, pinocembrin, and alpinetin and their chalcone derivatives: pinostrobin chalcone, pinocembrin chalcone, and cardamonin, respectively (Figure 1). The aim of this study is to understand the interactions involved in the binding of these compounds (referred to as ligands hereafter) to NS2B-NS3 of Den2 via computational docking methods and to gain insights into the experimental inhibition pattern.[19] It is hoped that information from this study will provide further understanding of the mechanism of inhibition of Den2 protease and enable the design of antiviral drugs which inhibit dengue virus replication.

EXPERIMENTAL METHODS

Purification and Screening of Ligands for Inhibition Activities. The purification techniques used to isolate the ligands involved in this study and the methods used in their biological screening for inhibition of proteolytic activity of Den2 NS2B-NS3 have been reported.[12,19] The fluorogenic substrate used in these studies was Boc-Gly-Arg-Arg-MCA. Den2 NS3 protease has been shown to cleave this substrate.

Building Protein and Ligand Structures for Computational Docking Experiments. The three-dimensional structure of Den2 NS2B-NS3 was retrieved from the Protein Data Bank (http://www.rcsb.org/pdb; accession code 2FOM). Chlorine atoms, water, and glycerol molecules were removed. Crystal coordinates of the flavanones were obtained from the Cambridge Crystallographic Data Centre and related publications.[25–27] Structures of the chalcones were built from the flavanones using Hyperchem Pro 6.0 software (Hypercube Inc.). Ligand structures were minimized with Hyperchem using the steepest descent and conjugate gradient methods (termination conditions set to a maximum of 500 cycles or 0.1 kcal/Å mol rms gradient).

Automated Rigid-Ligand Docking. Docking files were prepared using AutoDock Tools v.1.4 software.[28] For the protein molecule, polar hydrogen atoms were added and nonpolar hydrogen atoms were merged and Kollman charges and solvation parameters were assigned by default. For the ligands, Gasteiger charges were added, nonpolar hydrogen atoms were merged, and all bonds were made nonrotatable. All docking calculations were done with the AutoDock 3.0.5 software package using the Lamarckian genetic algorithm (LGA).[29] A population size of 150 and 250,000 energy evaluations were used for 100 search runs. The grid box, with grid spacing of 0.375 Å and dimension of 128 × 100 × 116 points along the x, y, and z axes, was centered on the macromolecule. After the docking searches were completed, clustering histogram analyses were performed, based on an rmsd (root-mean-square deviation) of not more than 1.0 Å. The conformation with the lowest docked energy was chosen from the most populated cluster and put through the next stage.

Automated Flexible-Ligand Docking. The methods involved were similar to the above, except that in ligand file preparation, torsional degrees of freedom were defined by allowing rotatable bonds to remain flexible. In addition, the grid box size was reduced to points, centered on the ligand. The number of energy evaluation was increased to 10 million. The conformations from the docking experiments were analyzed using Viewerlite 4.2 (http://www.accelrys.com), which also identified the H-bonding and van der Waals interactions between the protease and the ligands. LIGPLOT[30] was used to determine residues involved in hydrophobic interactions with the ligands. The HBPLUS[31] program in LIGPLOT calculates (by default) nonbonded/hydrophobic contacts, with cutoff in the range of 2.9−4.3 Å. The LIGPLOT algorithm[30] reads the three-dimensional (3D) structure of the ligands from the PDB files, together with the protein residues, and 'unrolls' each object about its rotatable bonds, flattening them out onto the 2D page.

Electrostatic Potential Map. Hyperchem Pro 6.0 was used to calculate single point energy of the ligands with the most favorable conformations using the semiempirical AM1 method. The electrostatic potentials of the ligands were then displayed as contour maps plotted in the form of 3-dimensional isosurfaces using Hyperchem. By default, Mulliken charges were used in the calculations to obtain the electrostatic potentials.

Glide Docking. The protein (PDB file as above) was prepared using the Protein Preparation Wizard of Schroedinger Suite 2007 (Schroedinger, Inc.). The suite's Site Map package was used to identify the potential binding sites; the noncompetitive site was selected for input into the Glide Receptor Grid Generation utility. The ligands were prepared either from sdf files found in the NCI database[32] or modified from similar compounds using the builder in the Maestro package of Schroedinger Suite 2007 and optimized using a

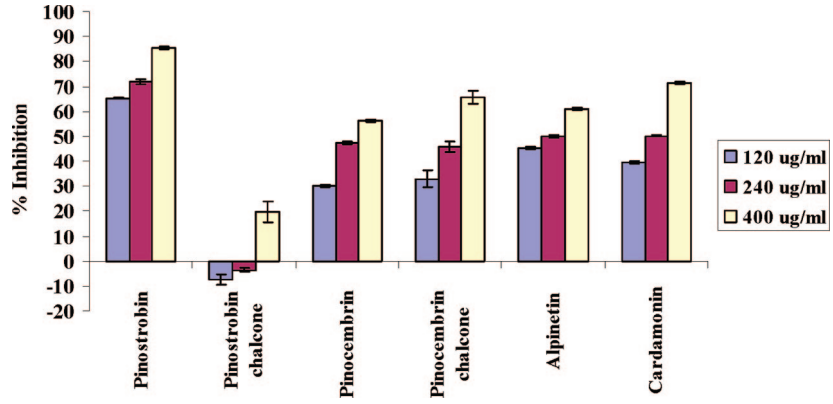


Figure 2. Percentage inhibition of DEN-2 NS2B/NS3 cleavage of substrate by ligands extracted from *Boesenbergia rotunda* L.[11,18] The standard 100 μL reaction mixtures comprised of 100 μM fluorogenic peptide subtrate Boc-Gly-Arg-Arg-MCA, 2 μM DEN-2 protease complex and with or without ligands of varying concentrations, buffered at pH 8.5 by 200 mM Tris-HCl. Each ligand was assayed at three different concentrations; 120 μg mL^{-1}, 240 μg mL^{-1}, and 400 μg mL^{-1}. Each test was done in quadruplicate. Four readings were taken—each at a time interval of 5 s per sample, and the three most consistent readings (% standard deviation <5%) were accepted.

Table 1. Automated Flexible-Ligand-Protein Docking Results Calculated Using AutoDock 3.0.5[a]

ligand	pinostrobin	cardamonin	alpinetin	pinocembrin chalcone	pinocembrin	pinostrobin chalcone
experimental IC_{50} (μg/mL)[18]	90.48	235.86	242.76	273.10	286.90	-
min. estimated free energy of binding, ΔG_{bind} (kcal/mol)	−8.30	−7.55	−8.84	−7.59	−9.25	−7.79
min. docked energy, E_{dock} (kcal/mol)	−8.82	−9.04	−9.45	−8.85	−9.57	−8.71
estimated inhibition constant, K_i (μM)	0.83	2.92	0.33	2.71	0.17	1.95
final intermolecular energy, ΔG_{inter} (kcal/mol)	−8.92	−8.80	−9.46	−8.53	−9.56	−9.03
final internal energy of ligand, ΔG_{intra} (kcal/mol)	+0.10	−0.25	+0.02	−0.32	0.00	+0.33
torsional free energy, ΔG_{tors} (kcal/mol)	+0.62	+1.25	+0.62	+0.93	+0.31	+1.25
rmsd (from ref structure) (Å)	0.61	3.15	0.60	2.05	0.88	4.60

[a] The IC_{50} values were obtained from experiments performed by Tan (2005).[18] The ligands are arranged in decreasing inhibition activity (increasing IC_{50} values) from left to right of the table. Estimated energy and K_i values shown were obtained from flexible docking of each ligand conformation which was initially the most favorable conformation obtained from rigid docking procedures (refer to the Experimental Methods section).

version of MacroModel also included. They underwent preparation using the Ligand Preparation utility of Schroedinger Suite 2007. The defaults, which were used, are set to generate states at pH 7 ± 2 and a maximum of 32 tautomers and stereoisomers, retaining specified chiralities, using the OPLS 2005 force field.[33] The outputs from Receptor Grid generation and Ligand Preparation are the inputs to Glide, which was run with standard precision (default settings), flexible docking, permitting 5- and 6-membered ring conformational interchange. Glide uses a grid approximation for the binding site, and an all-atom treatment of the ligands. Initial rough, rigid posing is followed by flexible energy minimization with an OPLS force field. The best poses are further refined using Monte Carlo methods. The output file is produced with best-scoring poses first and analyzed with Maestro's Pose Viewer utility.

RESULTS

Noncompetitive Inhibition of Den2 NS2B-NS3. Figure 2 shows the percentage of inhibition by the ligands, tested at three different concentrations. Results indicated pinostrobin to be the most active inhibitor, with up to 85.4% inhibition at 400 μg mL^{-1}, while its chalcone derivative, pinostrobin chalcone, did not exhibit any activity. However, in the other two cases, the chalcones were more active than their corresponding flavanones.

Tan et al. (2006) recently reported the noncompetitive inhibitory activities of some of these ligands.[12] The IC_{50} (ligand concentration at 50% inhibition) values are shown in Table 1, arranged in decreasing activity from left to right. The IC_{50} value of pinostrobin (90.48 μg mL^{-1}) is 2.6−3-fold smaller than the other ligands, indicating that it is the most active noncompetitive inhibitor.

Predicted Free Energy of Binding, Docking Energy, and K_i values. AutoDock was recently reported to be the most popular docking program.[34] Its high accuracy and versatility have expanded its use. In this study, rigid-ligand docking was initially performed using energetically optimized ligands at different potential binding sites of the whole protein (blind docking), since this reduces the time for the simulation to complete. Hetényi et al. (2002)[35] reported Autodock to be able to select the correct protein−ligand complexes, based on energy, without prior knowledge of the binding site (blind docking). Autodock has been proven to be efficient and robust in finding the binding pockets and binding orientations of the ligands, whether they are rigid (up to at least 30 heavy atoms) or flexible. As the ligands involved in this study are noncompetitive inhibitors, it was anticipated that they would dock to sites other than the active, requiring blind docking. SiteMap, part of the Schroedinger Suite 2007 (see Experimental Methods), identified largely the same noncompetitive binding site as Autodock. Glide[24] uses a ChemScore-derived algorithm to rank ligand binding; the more negative the score, the stronger the binding. The Glide scores are listed in Table 2.

Table 2. Automated Flexible, Extra-Precision-Ligand-Protein Docking Results Calculated Using Glide

ligand	pinostrobin	cardamonin	alpinetin	pinocembrin	pinocembrin chalcone	pinostrobin chalcone
Glide score	−7.14	−6.39	−6.86	−6.92	−5.55	−6.45

Table 3. Results of Rigid and Flexible Ligand Docking (AutoDock)[a]

		no. of free torsions		min. estimated ΔG_{bind} (kcal/mol)		final E_{dock} (kcal/mol)	
compounds	no. of atoms	rigid	flexible	rigid	flexible	rigid	flexible
pinostrobin	34	0	3	−8.23	−8.30	−8.23	−8.82
pinocembrin	31	0	3	−8.63	−9.25	−8.63	−9.57
alpinetin	34	0	3	−9.17	−8.84	−9.17	−9.45
pinostrobin chalcone	34	0	6	−7.74	−7.79	−7.74	−8.71
pinocembrin chalcone	31	0	6	−7.40	−7.59	−7.40	−8.85
cardamonin	34	0	6	−8.00	−7.55	−8.00	−9.04

[a] The energy values shown were obtained from flexible docking of each ligand conformation which initially was the most favorable conformation obtained from rigid docking procedures (refer to the Materials and Methods section).

Once the binding site for noncompetitive inhibition was identified, flexible-ligand docking was carried out at this site using Autodock. For this, a grid box of 60 × 60 × 60 number of points (grid spacing of 0.375 Å) was built around the binding region. Table 3 shows the calculated docking energy (E_{dock}) and free energy of binding (ΔG_{bind}) of both the rigid and flexible-ligand docking simulations. On the whole, energy values obtained from flexible docking were lower than those from rigid docking. Allowing the bonds in the ligands to rotate generated different ligand conformations, enabling more refined search for preferred binding sites.

The calculated energies and K_i (inhibitory constant) values do not follow the pattern of activity-ranking according to the observed IC_{50} values. Pinostrobin, the most active inhibitor, had the best Glide score. The remaining four inhibitors all had IC_{50} values between 200 and 300 μg mL^{-1}, so one might expect docking computations to be unable to discriminate between them. This highlights the challenge associated with scoring compound conformations and accurately predicting activity ranking using computational modeling techniques. Enthalpic and entropic effects drive ligand-binding processes, and either of these effects can dominate specific interactions.[36] However, if binding induces conformational changes in the protein, as often occurs in enzymatic interactions, using rigid binding sites limits the predictive ability to link experimental activity-ranking with calculated scoring functions.

In general, the estimated ΔG_{bind} for the flavanones obtained in this study were lower than those of their chalcone derivatives (Table 1). This is mainly because the chalcones have more rotatable bonds, increasing the torsional free energy (ΔG_{tor}) and lowering binding affinity (higher K_i values). On the other hand, the calculated K_i values for all the ligands are reasonably small (within the μM range), indicating the formation of stable enzyme−inhibitor complexes. Quantitative explanation of the experimental inhibitory activity of the ligands toward substrate-binding by the protease requires further structural insights.

Binding Site and Ligand Conformations. As expected, the automated docking experiments showed that all the ligands studied did not bind to the active site of the protease. Surprisingly, however, the binding site for these ligands was confined to a specific region of the protease (Figure 3a,b). The shape of the binding site can be divided into three parts (Figure 3b): the left region constitutes a hydrophobic hole which accommodates the aromatic rings of the ligands; the middle allows placement of flavanone ring C or the chains carrying the enone groups of the chalcones, enabling these ligands to interact with the surrounding residues; and the region on the right, bigger than the other two, allows more varied orientations of the ligands' pharmacophores. Presumably, the shape of the binding site complements the shapes of the ligands. Figure 3c,d shows the superimposition of the conformations of the flavanones and chalcones, respectively, at the binding sites computed with Autodock. Of the flavanones, pinocembrin and alpinetin take up similar poses. However, although pinostrobin lies on an axis similar to the other two ligands, it is oriented in the opposite direction, so that the phenyl ring projects into the region on the right side of the binding site.

Figure 3e shows the superposition of all six ligand conformations computed with Glide. Unlike Autodock, pinostrobin's orientation using Glide is similar to that of the other flavonoids. Like Autodock, pinostrobin chalcone binds differently from the five other ligands. The conformational difference between pinostrobin and the other flavanones obtained from Autodock is more clearly shown in Figure 4a−c which illustrates the three-dimensional (3D) isosurface plots of the electrostatic potential of the ligands. Similar plots were observed for alpinetin and pinocembrin, but the potential map for pinostrobin (colored green to indicate neutral to positive potential) is more elongated due to the presence of a methoxy group on ring A. Furthermore, the negative potentials (magenta) due to the carbonyl and hydroxyl groups on ring C and ring A, respectively, point toward the back of the plane of the model, while that due to the ether group on ring C points more toward the front of the plane of the model. The reverse was observed with alpinetin and pinocembrin. These observations infer the different chemical environment requirement of the binding site for docking of pinostrobin compared to alpinetin and pinocembrin. For the chalcones (Figure 4d−f), the shapes of the 3D isosurface plots of the electrostatic potential were different for each ligand. This could be attributable to the higher flexibility of the enone chains. This finding may also explain the requirement of a different (chemical) environment of the binding site by the ligands, hence the different orientations adopted by the ligands in the binding site.

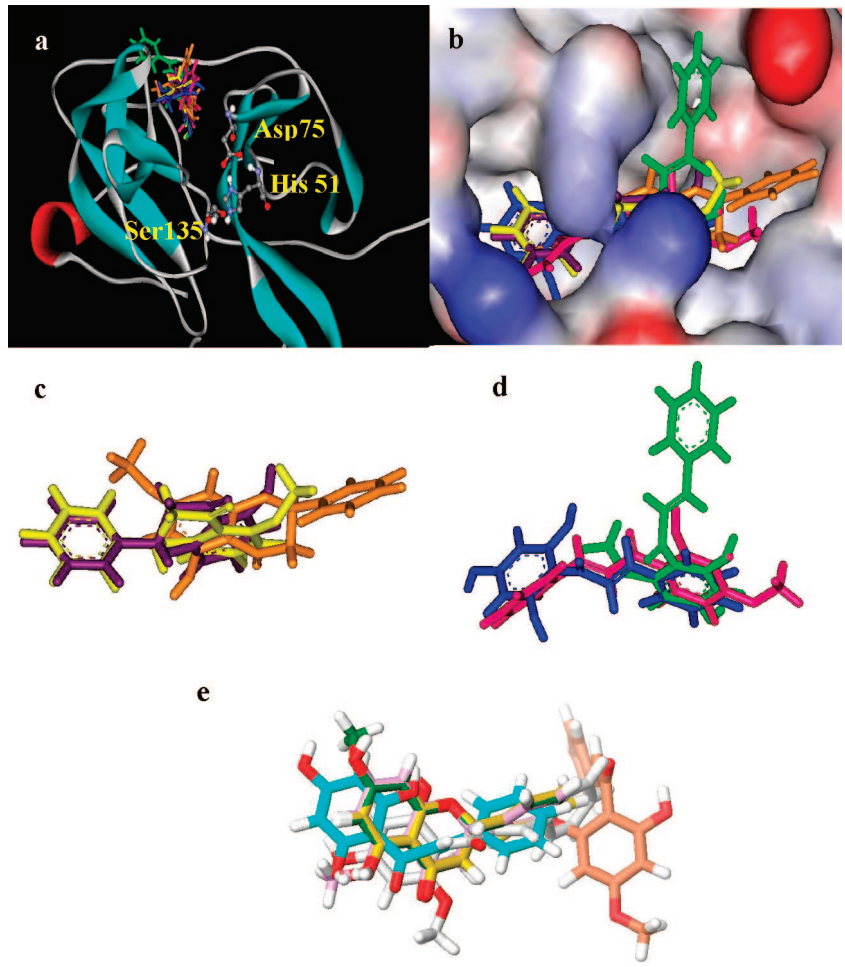


Figure 3. Computer generated models illustrating the superimposition of the flavanones, pinostrobin (orange), pinocembrin (purple), and alpinetin (yellow), and the chalcones, pinostrobin chalcone (pink), pinocembrin chalcone (blue), and cardamonin (green), at the binding site. The ligands are shown as sticks. (**a**) All the ligands are superimposed at the binding site. DEN-2 protease is represented as ribbons. The catalytic triads are labeled as His51, Asp75, and Ser135 and are shown as balls and sticks. (**b**) Connolly surface representation of the binding site colored according to electrostatic potential spectrum. (**c**) Superimposition of the flavanone poses as found in the binding site. (**d**) Superimposition of the chalcone poses as found in the binding site. (**e**) Glide poses of all ligands. Note that, as with Autodock, pinostrobin chalcone binds quite differently. The orange molecule is pinostrobin chalcone.

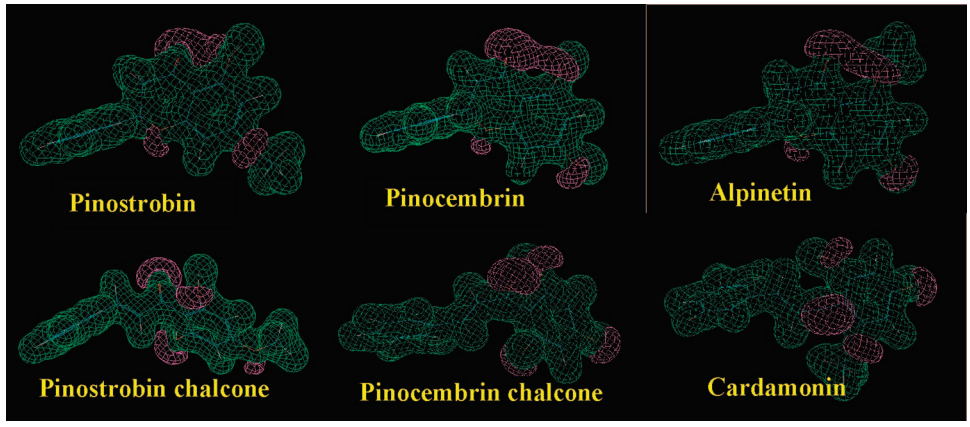


Figure 4. 3D isosurface plots of the electrostatic potentials of the ligands, using Hyperchem Pro 6.0 software. Green surfaces indicate neutral to positive potentials, while magenta surfaces indicate negative potentials.

DISCUSSION

Although both docking programs identified largely the same binding site, and docked pinostrobin chalcone differently from the other ligands, the details of the best poses from the two programs were different. Autodock's scoring appears to be more heavily tilted toward maximizing the number of hydrogen bonds than Glide. Glide was able to identify the best inhibitor but not the worst.

Binding Interactions. *Pinostrobin.* As shown in Table 1, pinostrobin showed the highest inhibition among all the ligands screened, with an IC_{50} value of 90.48 μg mL^{-1}, while pinostrobin chalcone was found to be inactive. Figure 5a,b

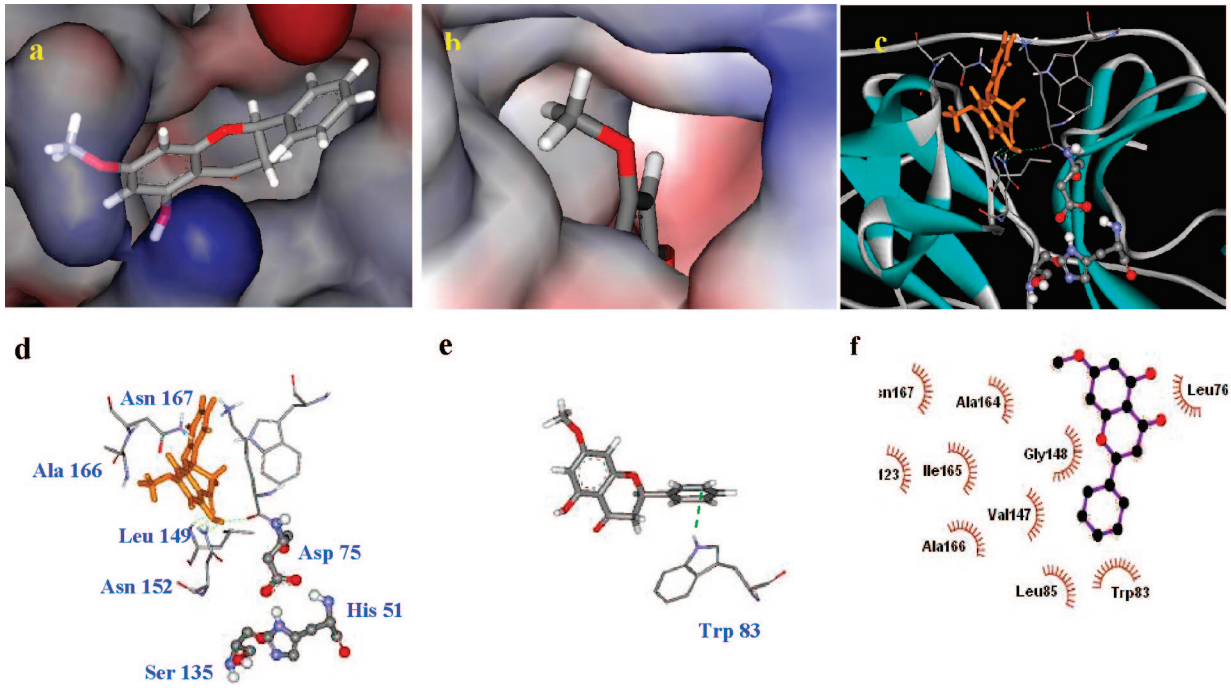


Figure 5. (**a**) Transparent Connolly surface representation of pinostrobin at the binding site. (**b**) View of pinostrobin in the binding site from a different angle. (**c**) View of pinostrobin (orange) at the binding site of Den2 protease (ribbons). Residues interacting with the ligand are shown as sticks. Catalytic triads are shown as balls and sticks. (**d**) Simplified view of pinostrobin interacting with surrounding residues. Residues labeled in green interact with the ligand via H-bonds, while those in red interact via van der Waals contacts. The catalytic triads are labeled in black. (**e**) Nonclassical H-bonding between Trp83 and phenyl ring B of pinostrobin. (**f**) 2D schematic diagram of residues in the binding site which exhibit hydrophobic interactions with pinostrobin, obtained using the Ligplot program. Keys for the plot are the following: (blue ball-purple stick-black ball) ligand bond; (His51 with curved railroad-like ties) nonligand residue involved in hydrophobic contact; and (black ball with railroad-like ties) corresponding atom involved in hydrophobic contact.

illustrated the binding site for pinostrobin. The shape of the binding pocket was observed to complement the shape (pose) of the ligand. The phenyl ring is protruded out toward the surface of the protein, while the rest of the molecule was embedded into the inner part of the protein. The surrounding residues involved in hydrogen bonding interactions with pinostrobin were Lys74, Leu149, and Asn152. The H-bonding interaction with Lys74 (Figure 5c,d) was not observed with the other flavanones and chalcones studied. The interaction between the hydroxyl H atom on ring A of pinostrobin with the backbone carbonyl O atom of Lys74 could account for the relatively high inhibition activity of pinostrobin. Since Lys74 is directly bonded to Asp75, the formation of H-bond between Lys74 and pinostrobin could have directly induced conformational change on Asp75, in particular, or the catalytic triad region, in general. This, presumably, could disrupt the electron transfer process required for substrate binding at the active site, hence affecting the activity of the protease.

Studies have shown that the formation of low-barrier hydrogen bond between Asp and His (of the catalytic triad) facilitates the nucleophilic attack by the β-OH group of Ser on the acyl carbonyl group of substrates.[37–39] Besides Lys74, Leu149 might play a role toward the activity of pinostrobin in two ways. First, Leu149 could be blocking the entry of the ligand into the active site due to its position in the protease, as observed in Figure 5c,d. Second, Leu149 was seen to protrude toward the adjacent β-barrel which carried the residues Asp75 and His51. Erbel et al. (2006)[40] reported that the NS3 protease domain adopts a chymotrypsin-like fold with two β-barrels and that the catalytic triad is located at the cleft between the two β-barrels. Thus, upon binding Leu149 with pinostrobin, a conformational change of the residue could occur in order to reduce steric clashes, affecting the spatial conformation of the surrounding residues, in particular the catalytic triad. In conjunction with this, the electron transfer process involved between Asp75 and His51 might be affected, reducing the capability of the active site to bind to the substrate. This process could further increase the inhibition capability of pinostrobin.

Figure 5f shows the residues involved in hydrophobic interactions with pinostrobin, obtained by the Ligplot program. Besides the hydrophobic interactions, Trp83 also exhibited nonclassical hydrogen bonding interaction between the H atom on $N_{\epsilon 1}$ of its indole ring with the phenyl ring (B) of pinostrobin (Figure 5e). In this case of nonclassical hydrogen bonding, the hydrogen bond acceptor is the aromatic ring.[41] Brocchieri and Karlin (1994) reported that the interplanar angle (dihedral angle between the extended planes of interacting planar groups, α) between the phenyl ring of Phe and the aromatic ring of Trp to be favorable at $30° < \alpha < 90°$.[42] They also reported that, generally, planar interactions of Trp involved mostly the five-atom ring which is capable of forming a hydrogen bond (involving its imino group), engaging π-cloud electrostatic interactions, and also undergoing hydrophobic interactions. Electrostatic charges associated with the phenyl ring include a weak negative charge about the center of the aromatic ring and a weak positive charge projected at the ring pheriphery.[43] In this study, interactions were also observed between Trp83 and pinostrobin, where the bond distance between the H atom on $N_{\epsilon 1}$ of Trp83 and the center of the ring of the ligand ($R_{centroid}$) was 2.67 Å, and α was approximately 61.0°.

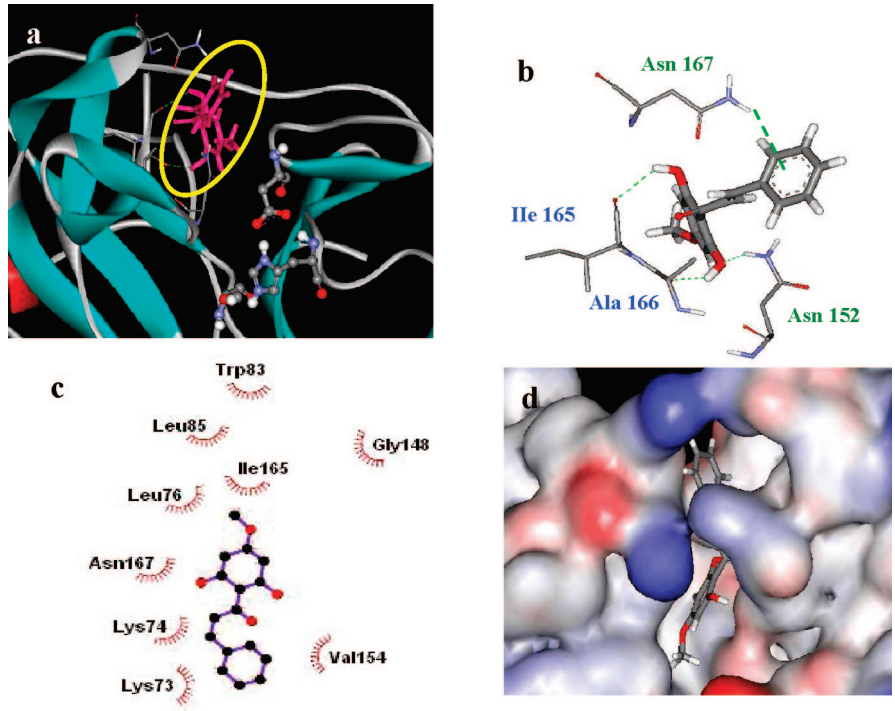


Figure 6. (**a**) View of pinostrobin chalcone (pink) at the binding site of Den2 protease (ribbons). Residues interacting with the ligand are shown as sticks. Catalytic triads are shown as balls and sticks. The yellow circle highlights the binding mode of the ligand which is confined to the C-terminal region of the protease, away from the catalytic triads. (**b**) Simplified view of pinostrobin chalcone interacting with surrounding residues. Residues labeled in green interact with the ligand via H-bonds, while those in blue exhibit both H-bond and van der Waals interactions. (**c**) 2D schematic diagram of residues in the binding site which exhibit hydrophobic interactions with pinostrobin chalcone obtained using the Ligplot program. Keys for the plot are the following: (blue ball-purple stick-black ball) ligand bond; (His51 with curved railroad-like ties) nonligand residue involved in hydrophobic contact; and (black ball with railroad-like ties) corresponding atom involved in hydrophobic contact. (**d**) Connolly surface representation of pinostrobin chalcone at the binding site nonligand residue involved in hydrophobic contact.

Pinostrobin Chalcone. Previous study had shown that pinostrobin chalcone did not exhibit inhibition activity (Table 1).[19] Investigation into the electrostatic interactions of pinostrobin chalcone with the binding site revealed that the surrounding amino acid residues of the protease involved in the interactions, via van der Waals and H-bond, were confined to the C-terminal region of the protease, with residues ranging from Asn152 to Asn167 (Figure 6a). This binding mode may not have any structural effect on the catalytic triad in promoting the disruption of electron transfer for the initiation of proteolytic processing. Presumably, binding activity of the ligand with its surrounding residues is confined to a region which may not impose any conformational change to the active site. Figure 6b illustrates the residues involved in H-bonding and van der Waals interactions with pinostrobin chalcone, in which a nonclassical H-bond is observed between Asn167 and the phenyl ring (B) of the ligand ($R_{centroid}$ = 3.64 Å, α = 40°). Figure 6c shows the residues involved in hydrophobic interactions with the ligand obtained from the Ligplot program. In addition to the interactions, the shape of the binding site was also complementary to the ligand's pose. From Figure 6d, observation of the hydrophobic pocket which accommodated the phenyl ring (B) of the ligand seemed to indicate that the ring was not placed directly in the center of the pocket, rather it was bent toward the left. This orientation may be attributed to the nonclassical H-bond interaction between the ring and Asn167.

The Other Ligands. Figure 7a−d shows the orientations of cardamonin, alpinetin, pinocembrin, and pinocembrin chalcone in the binding site of the protease, respectively. Except for pinocembrin chalcone, these ligands experienced H-bond and van der Waals interactions with almost the same surrounding residues. The common residues forming H-bond with the three ligands were Leu149 and Asn152. In addition, pinocembrin formed an additional H-bond with Val147 and Ile165. In terms of the van der Waals interactions, cardamonin interacted with Ile165, while alpinetin interacted with Val147 and Ile165, and pinocembrin with Asn167. Pinocembrin chalcone, however, did not demonstrate van der Waals interaction with any of the surrounding residues but interacted with Lys73 and Asn167 via H-bond.

The degree of inhibition offered by these four ligands was, in actual fact, similar to each other, even though both cardamonin and alpinetin exhibited slightly higher activities than pinocembrin chalcone and pinocembrin (Figure 2 and Table 1). The similarity in the degree of activities shown may be explained by the similar axis of orientation in the binding site (except for cardamonin; Figure 3) and similar mode of interactions as described above (except for pinocembrin chalcone; Figure 7). Unlike pinostrobin, these ligands did not form H-bond with Lys74, which may be the cause of the reduced activity observed when compared to pinostrobin. Nevertheless, Lys74 was involved in hydrophobic interaction with alpinetin, pinocembrin, and pinocembrin chalcone. Figure 8 illustrates the residues involved in hydrophobic interactions with the ligands using the Ligplot program. The observed activities of cardamonin, alpinetin, and pinocembrin could be due to the interactions of the ligands with Leu149, as shown in parts a, b, and d,

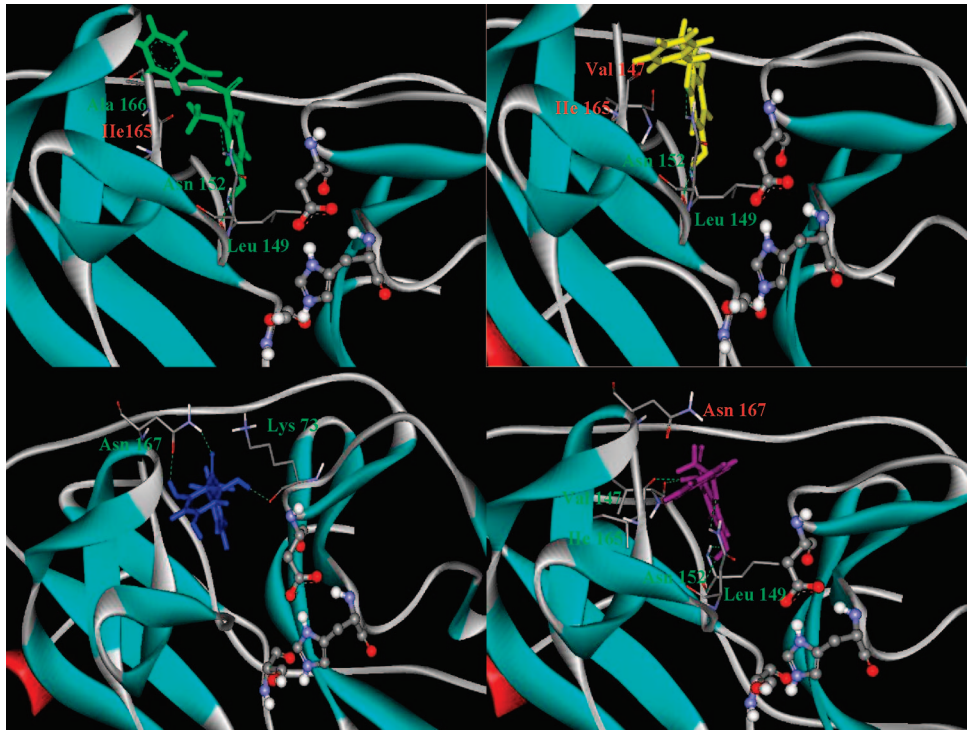


Figure 7. Models of ligands bound to the protease (ribbons) at the binding sites. The ligands shown are (**a**) cardamonin (green), (**b**) alpinetin (yellow), (**c**) pinocembrin chalcone (blue), and (**d**) pinocembrin (purple). Residues interacting with the ligands are shown as sticks. Catalytic triads are shown as balls and sticks. Residues labeled in green interact with the ligands via H-bonds, while those in red interact via van der Waals contacts.

respectively, of Figure 7. The effect of Leu149 on substrate-binding capability of the protease upon its binding with the ligands has been discussed earlier.

A different mode of interaction was observed with pinocembrin chalcone which did not bind to Leu149 to exhibit a similar inhibition effect. Rather, its H-bonding interaction with Lys73 could be the cause of the observed activity (Figure 7c). Lys73 is two residues downstream from Asp75, and interaction between the ligand with this residue may contribute directly to the conformational change of the active site, the extent of which, however, was not as great as that compared to the interaction between pinostrobin and Lys74. The fact that pinocembrin chalcone did not exhibit van der Waals interactions with the surrounding residues may be the cause for the lower activity seen in the pinocembrin chalcone than that of pinostrobin.

Structure−Activity Relationship (SAR) Analysis. The docking studies performed gave better structural insights and understanding on how the various ligands interacted with the protease in acting as noncompetitive inhibitors toward Den2 protease activity. Analysis of the structure−activity relationships of the ligands in this study may shed light on the important structure and conformation which could be applied in the design of new compounds.

This study affirmed that rigid conformation of the flavanones would ensure ligand activities. Opening up of the ring C, as found in the chalcones, would introduce higher flexibility to the ligands leading to a more extended conformation, thus requiring a different environment of the binding site (Figure 4). Higher flexibility of the chalcones also meant higher ΔG_{tor} compared to the corresponding flavanones and, hence, higher ΔG_{bind}. This would result in the chalcone binding being less favorable to the site when compared to the corresponding flavanones (Table 2). Pinocembrin showed the lowest values for the E_{dock} and ΔG_{bind}, followed by alpinetin and pinostrobin, indicating the parent structure of 7-hydroxyflavanone to have good binding capabilities with the protease.

The methoxy group at C7 on the ring A of pinostrobin may be important in creating the preferable electrostatic potential surface of the molecule in the 'search' for an optimal chemical environment within the binding site. In general, an electron-donating group at this position would be preferable for the design of new inhibitory compounds. The hydroxyl group at C5 on the ring A of pinostrobin could be considered to be an important pharmacophore. This hydroxyl group is very useful in forming H-bonds with the surrounding residues, particularly, between pinostrobin and Lys74 and between pinocembrin chalcone and Lys73 (the OH group is on C2′ of ring A for pinocembrin chalcone). For pinocembrin chalcone, the number of electrostatic interactions involved was the lowest among all the ligands reported. Thus, one might expect pinocembrin chalcone to exhibit a much lower inhibitory activity than the other ligands (except pinostrobin chalcone). However, from experiment, its activity was observed to be comparable to those of pinocembrin, cardamonin, and alpinetin. This could be attributed to its H-bonding interaction with Lys73 via the hydroxyl group, where Lys73 is positioned two residues downstream from Asp75. The conclusions derived from this study with regards to the importance of the methoxy and hydroxyl groups in determining the activity of the ligands are in accordance with those postulated by Tan (2005).[19]

The presence of the phenyl ring (B) in all the ligands is not of little importance and is essential in contributing to the hydrophobicity of the ligands. The entropic effects of

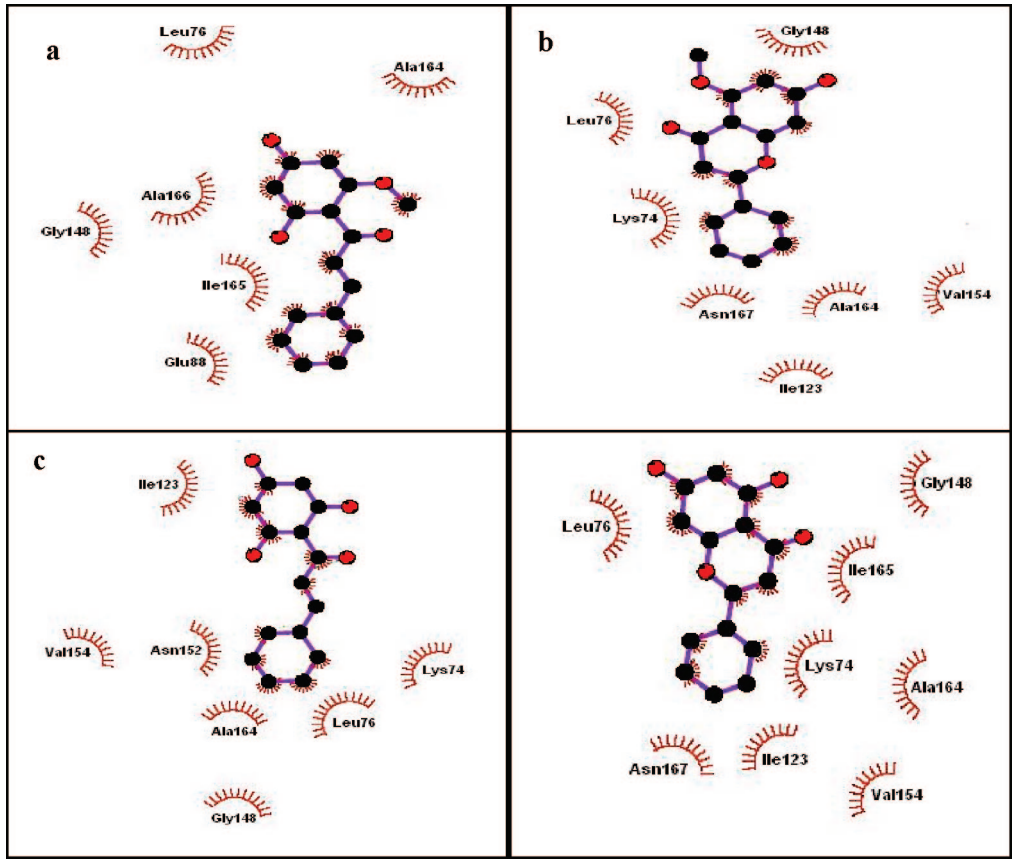


Figure 8. Schematic diagrams (2D) illustrating residues in the binding sites which are involved in hydrophobic interactions with (**a**) cardamonin, (**b**) alpinetin, (**c**) pinocembrin chalcone, and (**d**) pinocembrin. Keys for the plot are the following: (blue ball-purple stick-black ball) ligand bond; (His51 with curved railroad-like ties) nonligand residue involved in hydrophobic contact; and (black ball with railroad-like ties) corresponding atom involved in hydrophobic contact.

ligand binding vary with the degree of hydrophobic interactions in the system, contributing to the determination of the ΔG_{bind} value, and, as discussed earlier, the phenyl ring of pinostrobin and pinostrobin chalcone formed nonclassical H-bonding with Trp and Asn, respectively.

CONCLUSIONS

Three flavanones and three chalcones isolated from the plant *Boesenbergia rotunda* L. were docked onto the Den2 NS2B-NS3 protease using the AutoDock 3.0.5 software. Results obtained from this study are consistent with the experimental results illustrating the noncompetitive inhibitory activities for most of the ligands.[19] As expected, these ligands were found to bind to sites other than the active site of the Den2 serine protease. The calculated K_i values of these ligands were very small indicating that they bound considerably well to the allosteric binding site. The higher noncompetitive inhibitory activity shown by pinostrobin compared to the other compounds could be accounted for by H-bonding interaction with the backbone carbonyl of Lys74, which is bonded to Asp75 (one of the catalytic triad residues). This interaction was not observed with the other ligands. SAR analysis yielded some structural features which may be useful for the design of new compounds with potential inhibitory activities. These features are the rigid structure of flavanone, the C5 hydroxyl, and C7 methoxy groups on ring A and the phenyl ring (B).

Docking experiments were also performed with rigid ligands as well as with flexible ligands on the rigid protein structures (obtained from the PDB). To ensure that the observed interactions in the resulting structures were sustained, molecular dynamics calculations were often performed. Although molecular dynamics on the system was not carried out in the study, this additional technique could be carried out as a potential method of justification for the resulting structures. The results obtained did not illustrate any conformational changes that could occur after the ligand-protein binding process. To overcome such limitation, flexible-protein docking would be preferable. However, the development of computational strategies for this purpose is still in its infancy.[34] Several methods have been reported, and promising results have emerged from the application of combined methods such as the ensemble docking approach and an induced fit.[44]

Flexible-protein docking could be a way forward toward deeper insights into the present system which could aid in the design of new compounds for a therapeutic drug against dengue virus infections.

ACKNOWLEDGMENT

This work was supported in part by the Malaysian Ministry of Science, Technology and Innovation under the Top Down National Biotechnology Directory grant number 09-02-04-001BTK/TH/004 [UM 36-02-03-6008], the Academy of Science, Malaysia, under the Scientific Advancement Fund Allocation (SAGA 66-02-03-0049), and Universiti Malaya under the F-Vote grant number F0171/2005D.

REFERENCES AND NOTES

(1) Kautner, I.; Robinson, M. J.; Kuhnle, U. Dengue virus infection: epidemiology, pathogenesis, clinical presentation, diagnosis and prevention. *J. Pediatr.* **1997**, *131*, 516–524.

(2) Gubler, D. J. The global emergence/resurgence of arboviral diseases as public health problems. *Arch Med Res.* **2002**, *33*, 330–42. (WHO. 2007. http://www.who.int/csr/disease/dengue/impact/en/index.html (accessed May 20, 2008))

(3) Dengue kills 44 people in Malaysia, Sun Malaysia, 2007; May 20, 2007; http://story.malaysiasun.com/index.php/ct/9/cid/48cba686fe041718/id/250213/cs/1/ (accessed May 20, 2008).

(4) Irie, K.; Mohan, P. M.; Sasaguri, Y.; Putnak, R.; Padmanabhan, R. Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). *Gene* **1989**, *75*, 197–211.

(5) Falgout, B.; Pethel, M.; Zhang, Y. M.; Lai, C. J. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J. Virol.* **1991**, *65*, 2467–2475.

(6) Yusof, R.; Clum, S.; Wetzel, M.; Murthy, H. M.; Padmanabhan, R. Purified NS2B-NS3 serineprotease of dengue virus type 2 exhibits cofactor NS2B dependence for cleavage of substrates with dibasic amino acids in vitro. *J. Biol. Chem.* **2000**, *275*, 9963–9969.

(7) Gorbalenya, A. E.; Donchenko, A. P.; Koonin, E. V.; Blinov, V. M. N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. *Nucleic Acids Res.* **1989**, *17*, 3889–3897.

(8) Bazan, J. F.; Fletterick, R. J. Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. *Virology* **1989**, *171*, 637–639.

(9) Brinkworth, R. I.; Fairlie, D. P.; Leung, D.; Young, P. R. Homology model of the dengue 2 virus NS3 protease: putative interactions with both substrate and NS2B cofactor. *J. Gen. Virol.* **1999**, *80*, 1167–1177.

(10) Arias, C. F.; Preugschat, F.; Strauss, J. H. Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. *Virology* **1993**, *193*, 888–899.

(11) Clum, S.; Ebner, K. E.; Padmanabhan, R. Cotranslational membrane insertion of the serine proteinase precursor NS2B-NS3(Pro) of dengue virus type 2 is required for efficient in vitro processing and is mediated through the hydrophobic regions of NS2B. *J. Biol. Chem.* **1997**, *272*, 30715–30723.

(12) Tan, S. K.; Pippen, R.; Yusof, R.; Ibrahim, H.; Khalid, N.; Abd Rahman, N. Inhibitory activity of cycohexenyl chalcone derivatives and flavonoids of fingerroot, *Boesenbergia rotunda* (L.), towards dengue-2 virus NS3 protease. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3337–3340.

(13) Whitby, K.; Pierson, T. C.; Geiss, B.; Lane, K.; Engle, M.; Yi, Z.; Doms, R. W.; Diamond, M. S. Castanospermine, a potent inhibitor of Dengue virus infection in vitro and in vivo. *J. Virol.* **2005**, *79*, 8698–8706.

(14) Hrobowski, Y. M.; Garry, R. F.; Michael, S. F. Peptide inhibitors of dengue virus and West Nile virus infectivity. *Virol. J.* **2005**, *2*, 49.

(15) Putnak, J. R.; Coller, B.-A.; Voss, G.; Vaughn, D. W.; Clements, D.; Peters, I.; Bignami, G.; Hounga, H.-S.; Chena, R. C.-M.; Barvir, D. A.; Seriwatana, J.; Cayphas, S.; Garcon, N.; Gheysen, D.; Kanesa-thasan, N.; McDonell, M.; Humphreys, T.; Eckels, K. H.; Prieels, J.-P.; Innis, B. L. An evaluation of dengue type-2 inactivated, recombinant subunit, and live-attenuated vaccine candidates in the rhesus macaque model. *Vaccine* **2005**, *23*, 4442–4452.

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

(17) Diamond, M. S.; Zachariah, M.; Harris, E. Mycophenolic acid inhibits dengue virus infection by preventing replication of viral RNA. *Virology* **2002**, *304*, 211–221.

(18) Ray, D.; Shi, P.-Y. Recent Advances in Flavivirus Antiviral Drug Discovery and Vaccine Development. *Recent Pat. Anti-Infective Drug Discovery* **2006**, *1*, 45–55.

(19) Tan, S. K. *Flavonoids from Boesenbergia rotunda (L). Mansf.: Chemistry, bioactivity and accumulation*; Ph.D. Thesis, Universiti Malaya, Kuala Lumpur 2005.

(20) Ibrahim, H.; Rahman, A. A. *Several ginger plants (Zingiberaceae) of potential value. In Malaysian Traditional Medicine: Proceedings of the Seminar on Malaysian Traditional Medicine*; Soepadmo, E., Goh, S. H., Wong, W. H., Din, L., Chuah, C. H., Eds.; Institute of Advanced Studies, University of Malaya, and Malaysian Institute of Chemistry: Kuala Lumpur, 1989; pp 159−161.

(21) Tuchinda, P.; Reutrakul, V.; Claeson, P.; Pongprayoon, U.; Sematong, T.; Santisuk, T.; Taylor, W. Anti-inflammatory cyclohexenyl chalcone derivative in Boesenbergia pandurata. *Phytochemistry* **2002**, *59*, 169–173.

(22) Tewtrakul, S.; Subhadhirasakul, S.; Kummee, S. HIV-1 protease inhibitory effects of medicinal plants used as self medication by AIDS patients. *Songklanakarin J. Sci. Technol.* **2003**, *25*, 239–243.

(23) Tewtrakul, S.; Subhadhirasakul, S.; Puripattanavong, J.; Panphadung, T. HIV-1 protease inhibitory substances from the rhizomes of *Boesenbergia pandurata* Holtt. *Songklanakarin J. Sci. Technol.* **2003**, *25*, 503–508.

(24) Friesner, R. A.; Bank, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenking, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.

(25) Brown, M. J.; Henderson, D. E.; Hunt, C. Comparison of antioxidant properties of supercritical fluid extracts of herbs and the confirmation of pinocembrin as a principle antioxidant in Mexican oregano (Lippa gravolens). Electron. *J. Environ. Agric. Food Chem.* **2006**, *5*, 1265–1277.

(26) Jiang, R.-W.; He, Z.-D.; But, P. P.-H.; Chan, Y.-M.; Ma, S.-C.; Mak, T. C. W. A Novel 1:1 Complex of Potassium Mikanin-3-*O*-sulfate with Methanol. *Chem. Pharm. Bull.* **2001**, *49*, 1166–1169.

(27) Shoja, M. 5-Hydroxy-7-methoxyflavone. *Acta Crystallogr.* **1989**, *C45*, 828–829.

(28) Sanner, M. F. Python: A programming language for software integration and development. *J. Mol. Graphics Model.* **1999**, *17*, 57–61.

(29) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

(30) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **1995**, *8*, 127–134.

(31) McDonald, I. K.; Thornton, J. M. Satisfying hydrogen-bonding potential in proteins. *J. Mol. Biol.* **1994**, *238*, 777–793.

(32) National Cancer Institute, Developmental Therapeutics Program, Enhanced Database Browser. http://129.43.27.140/ncidb2/ (accessed May 20, 2008).

(33) Jorgensen, W. L.; Tirado-Rives, J. *Potential energy functions for atomic-level simulations of water and organic and biomolecular systems PNAS* **2005**, *102*, 6665–6670.

(34) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-Ligand Docking: Current Status and Future Challenges. *Proteins* **2006**, *65*, 15–26.

(35) Hetényi, C.; van der Spoel, D. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* **2002**, *11*, 1729–1737.

(36) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and application. *Nature Rev. Drug Discovery* **2004**, *3*, 935–949.

(37) Frey, P. A.; Whitt, S. A.; Tobin, J. B. A Low-Barrier Hydrogen Bond in the Catalytic Triad of Serine Proteases. *Science* **1994**, *264*, 1927–1930.

(38) Santis, L. D.; Carloni, P. Serine Proteases: An Ab Initio Molecular Dynamics Study. *Proteins* **1999**, *37*, 611–618.

(39) Hunkapiller, M. W.; Forgac, M. D.; Richards, J. H. Mechanism of Action of Serine Proteases: Tetrahedral Intermediate and Concerted Proton Transfer. *Biochemistry* **1976**, *15*, 5581.

(40) Erbel, P.; Schiering, N.; D'Arcy, A.; Renatus, M.; Kroemer, M.; Lim, S. P.; Yin, Z.; Keller, T. H.; Vasudevan, S. G.; Hommel, U. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat. Struct. Mol. Biol.* **2006**, *13*, 372–373.

(41) Gervasio, F. L.; Chelli, R.; Procacci, P.; Schettino, V. The Nature of Intermolecular Interactions between Aromatic Amino Acid Residues. *Proteins* **2002**, *48*, 117–125.

(42) Brocchieri, L.; Karlin, S. Geometry of interplanar residue contacts in protein structures. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 9297–9301.

(43) Burley, S. K.; Petsko, G. A. Weakly polar interactions in proteins. *Adv. Protein Chem.* **1988**, *39*, 125–189.

(44) Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. Representing Receptor Flexibility in Ligand Docking through Relevant Normal Modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632–9640.

CI700388K

In Silico Biology 7, 0022 (2007); ©2007, Bioinformation Systems e.V.

Analysis of secondary structure predictions of Dengue virus type 2 NS2B/NS3 against crystal structure to evaluate the predictive power of the *in silico* methods

Rozana Othman[1,5]*, Habibah Abdul Wahab[2,3], Rohana Yusof[4] and Noorsaadah Abd Rahman[5]

[1] Department of Pharmacy, Faculty of Medicine, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[2] School of Pharmaceutical Sciences, Universiti Sains Malaysia, 11800, USM, Penang, Malaysia
[3] Laboratory of Biocrystallography and Bioinformatics Structure, Universiti Sains Malaysia, 11800, USM, Penang, Malaysia
[4] Department of Molecular Medicine, Faculty of Medicine, Universiti Malaya, 50603 Kuala Lumpur, Malaysia
[5] Department of Chemistry, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

* Corresponding author

Email: rozanaothman@um.edu.my;
Phone: +60-3-79675796, +60-3-79674959; Fax: +60-3-79674964

Edited by E. Wingender; received September 18, 2006; revised February 08 and March 02, 2007; accepted March 03, 2007; published March 27, 2007

Abstract

Multiple sequence alignment was performed against eight proteases from the Flaviviridae family using ClustalW to illustrate conserved domains. Two sets of prediction approaches were applied and the results compared. Firstly, secondary structure prediction was performed using available structure prediction servers. The second approach made use of the information on the secondary structures extracted from structure prediction servers, threading techniques and DSSP database of some of the templates used in the threading techniques. Consensus on the one-dimensional secondary structure of Den2 protease was obtained from each approach and evaluated against data from the recently crystallised Den2 NS2B/NS3 obtained from the Protein Data Bank (PDB). Results indicated the second approach to show higher accuracy compared to the use of prediction servers only. Thus, it is plausible that this approach is applicable to the initial stage of structural studies of proteins with low amino acid sequence homology against other available proteins in the PDB.

Keywords: Dengue virus type 2, serine protease, secondary structure prediction, consensus, protein structure

Introduction

Dengue virus belongs to the Flaviviridae family and is a widespread human pathogen that can cause haemorrhagic fevers [Kautner *et al.*, 1997]. Dengue infections place some 2.5 billion people (or 40% of the world population) at risk and are a significant cause of mortality, especially in the tropical and subtropical regions [Yin *et al.*, 2006]. In the recent years, dengue fever has been on the rise globally. The most prevalent of the four dengue serotypes is the dengue virus type 2 (Den2). Den2 contains a single-stranded RNA of positive polarity. This RNA genome codes for a single polyprotein precursor arranged in the order C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5 [Irie *et al.*, 1989]. Flavivirus replication is dependent upon the correct cleavage of this polypeptide and requires both host cell proteases and a virus-encoded, two-component protease, NS2B/NS3 [Falgout *et al.*, 1991; Yusof *et al.*, 2000]. Hence, this protease complex serves to be a target for the development of antiviral drugs. Although there have been effort by many research groups, no vaccines or antiviral drugs are currently available against dengue virus infections. Thus, there is an immense interest in developing new antiviral therapeutic agents to fight diseases caused by dengue viruses.

Several studies were carried out by various research groups to gain structural insight into the protease complex of Dengue virus type 2 (Den2). Until very recently [D'Arcy *et al.*, 2006; Erbel *et al.*, 2006], no data on the crystal structure of NS2B/NS3 protease complex of the dengue virus was available. A homology model of the NS2B/NS3 of Den2 was built by Brinkworth *et al.*, 1999, based on the crystal structure coordinates of the hepatitis C virus NS3/NS4A as template (PDBid: 1JXP). The overall identity between the two sequences is only 14.8%. However, regions surrounding the putative catalytic residues, as defined by Bazan and Fletterick, 1989, indicate a high level of identity. The lack of structural details of the active protease from experiments did not offer substantial insights into its interaction with substrates. Thus, the design of inhibitors were mainly based on either kinetic studies, such as that reported by Kiat *et al.*, 2006, or theoretical understandings and explanations from *in silico* simulations [Brinkworth *et al.*, 1999; Lee *et al.* 2006].

Knowledge on the three-dimensional (3D) structure of a protein is crucial towards understanding its function. Since it is often difficult or impossible to determine a structure experimentally, computational techniques have become very popular in generating models of proteins. Comparative, or homology, modelling remains the only method that can reliably predict the 3D structure of a protein with accuracy comparable to that of a protein structure resolved at low-resolution via experimental means [Marti-Renom *et al.*, 2000]. This technique relies upon the alignment of a protein sequence of unknown structure (target) to a homologue of known structure (template). However, potential problems can occur in structural determination when the target protein and template have less than 25% sequence identity (based on an average domain length of 80 amino acids) [Sander and Schneider, 1991; Marti-Renom *et al.*, 2003]. Nevertheless, sufficiently long alignments can still infer structural similarity, even when the sequence similarity is below 25% [Lund *et al.*, 1997].

With no homologue of known structure from which to make a 3D model, a logical next step is to predict secondary structure, which aims to provide the location of α-helices and β-strands within a protein or protein family. Once the secondary structure of a protein has been determined, the protein fold recognition is carried out, followed by a prediction of the tertiary structure. There are many methods (web servers) available to perform secondary structure prediction. Kabsch and Sander, 1984, reported that low level of success of prediction without additional information was to be commonly attributed to the neglect of long-range interactions within a protein. However, Pan *et al.*, 1999, attributed the low levels in prediction accuracy to the limitation of the available protein database size or prediction algorithm. Nevertheless, the field of secondary

structure has achieved a break-through by combining algorithms from artificial intelligence with evolutionary information [Rost, 2003], boosting the current height of prediction accuracy to around 77%. Thus, secondary structure prediction can be accurate enough to be taken seriously as a tool to assist in the design of experiments to probe protein structure and function [Barton, 1995].

A secondary structure prediction study of 175 N-terminal amino acids of the NS3 protease using a combination of several prediction tools available over the website has been carried out in our lab, a couple of years before the publication of crystal structure of Den2 NS2B/NS3 protease complex (PDBid: 2FOM) [D'Arcy *et al.*, 2006; Erbel *et al.*, 2006]. The aim of this work is to map out the secondary structure of the different regions in the protease with the knowledge of structurally conserved regions obtained from multiple sequence alignment against NS3 proteases of other viruses from the Flaviviridae family. The structural data obtained from the recently crystallised Den2 NS2B/NS3 has enabled us to make an evaluation of the prediction and thus reported herein.

Methods

Multiple sequence alignment

Protein sequence alignments and comparisons were done with a BLAST (Basic Local Alignment Search Tool) program, blastp, against database specification of non-redundant protein which were available at the National Center for Biotechnology Information (NCBI) Web server, [Altschul *et al.*, 1997; http://www.ncbi.nlm.nih.gov/blast/]. Viruses for proteases used in this study were checked against the Universal Virus Database, ICTVdb (International Committee on Taxonomy of Viruses) [Büchen-Osmond, 2003]. Amino acid sequences were obtained from NCBI sequence viewer 2.0 available at http://www.ncbi.nlm.nih.gov. Multiple sequence alignments were done using ClustalW 1.82 [Thompson *et al.* 1994], available at the European Bioinformatic Institute (EBI) Web server. Consensus of amino acid sequence was obtained from Boxshade available at the European Molecular Biology Network (EMBnet) Web server (http://www.ch.embnet.org).

Secondary structure prediction

The 175 amino acid sequence of Den2 NS3 protease was submitted for automated prediction of secondary structures to the following programs via their web site interfaces: PSIPRED [Jones, 1999; McGuffin *et al.*, 2000; Bryson *et al.*, 2005; http://bioinf.cs.ucl.ac.uk/psipred/], PROF on PredictProtein Web server [Rost *et al.*, 2004; http://www.predictprotein.org], PHD on PredictProtein Web server [Rost, 1996], APSSP2 [Raghava, 2002; http://www.imtech.res.in/raghava/apssp2/] and Jnet on Jpred Web server [Cuff and Barton, 2000; http://www.compbio.dundee.ac.uk/~www-jpred/jnet/].

Threading programs used via their web site interfaces were: 123D+ [Alexandrov *et al.*, 1995; http://123d.ncifcrf.gov/123D+.html], 3D-PSSM [Kelley *et al.*, 2000; http://www.sbg.bio.ic.ac.uk/servers/3dpssm/] and LOOPP [Meller and Elber, 2001; Teodorescu *et al.* 2004; http://cbsuapps.tc.cornell.edu/loopp.aspx]. The secondary structure informations of some templates used in the threading programs were obtained from DSSP database available

online [Kabsch and Sander, 1983; http://swift.cmbi.ru.nl/gv/dssp/]. One-dimensional (1D) secondary structure predictions performed in this study employed two approaches:

1. By utilising available structure prediction servers only. Alignment of secondary structures obtained from prediction programs was performed to result in a consensus.
2. By gaining information on the secondary structures extracted from structure prediction servers, threading techniques and DSSP database of some of the templates used in the threading techniques. An alignment of all secondary structures obtained was performed to result in another consensus.

Comparison of the results obtained from the two tool sets (approach 1 and approach 2) were then carried out against the observed secondary structure of the recently crystallised NS2B/NS3 obtained from the Protein Data Bank [Berman *et al.*, 2000] (PDBid: 2FOM) and the percentage accuracy, A, was calculated as follows:

$$A = \frac{c}{N} \times 100\%$$

where c is the number of residues predicted correctly, and N is the total number of residues with predicted secondary structures aligned against structures from crystal data.

Percentage difference, D, in secondary structure between both approaches was calculated as follows:

$$D = \frac{d}{N_t} \times 100\%$$

where d is the number of residues with different secondary structure from both approaches, and N_t is the total number of residues.

The implementation of approach 2 was out of curiosity to see if more exhaustive techniques of secondary structure prediction would give rise to more reliable results.

Results and discussion

Fig. 1 illustrates the multiple sequence alignment of Den2 NS3 protease (EC 3.4.21.91) and other proteases belonging to the genus Flavivirus of the Flaviviridae family. Consensus of sequences obtained from the Boxshade program is shown and it can be seen that regions surrounding the catalytic triads are highly conserved [Bazan and Fletterick, 1989; Brinkworth, 1999]. Secondary structure profiles built following the two approaches (as stated in the methods section) are illustrated in Figs. 2 and 3. Consensus of the secondary structures is given as E (β-strand), H (α-helix) or C (coil). It can be seen that both approaches yield mainly β-strands, 39.43% from the whole structure using the first approach, while the latter results in 40% of β-strands. These results are in accordance with the fact that Den2 NS3 protease belongs to fold family of trypsin-like serine-proteases (S07.001; according to MEROPS) [Rawlings *et al.*, 2006], which falls into the all-β proteins class of protein structure (according to SCOP) [Bazan and Fletterick, 1989; Murzin, 1995].

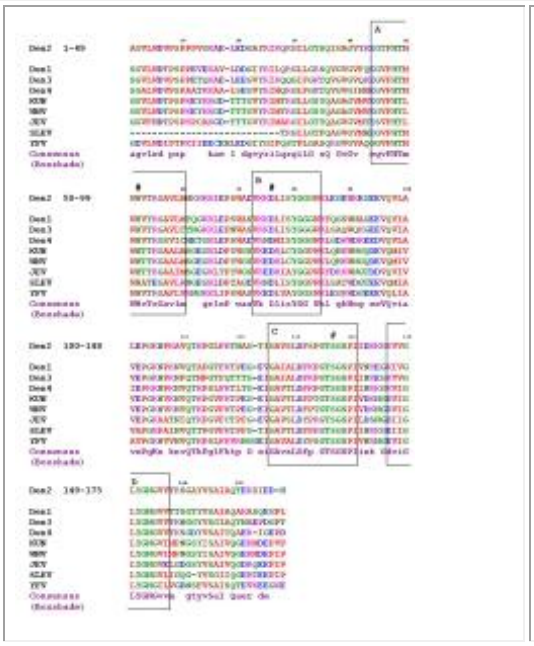


Figure 1: Multiple sequence alignment of Den2 NS3 protease and other proteases belonging to the genus Flavivirus. The catalytic triad residues: His51, Asp75 and Ser135, are labelled with the symbol # and are found in boxes labelled A, B and C. (The numbers in the diagram may not represent the exact positions of residues in the actual protein due to the insertion of gaps during alignment). The boxes labelled A, B, C and D identify regions of significant similarity surrounding the catalytic triad residues and residues that might form the substrate-binding pocket [Bazan and Fletterick, 1989]. Consensus of residues throughout the different viruses are indicated by Boxshade, with those having 100% similarity written in the upper case font. Sequences used in the analyses: Den1 - Dengue virus type 1, Den3 - Dengue virus type 3, Den4 - Dengue virus type 4, KUN - Kunjin virus, WNV - West Nile virus, JEV - Japanese encephalitis virus, SLEV - St. Louis encephalitis virus and YFV - Yellow fever virus.

Figure 2: Profile of 1D secondary structure prediction of Den2 NS3 protease based on Approach 1 (see Method section): combination of automated prediction online programs. Consensus of secondary structure are given as E = β-strand, H = α-helix or C = coil. The catalytic triad residues: His51, Asp75 and Ser135, are labelled with the symbol #.

Figure 3: Profile of 1D secondary structure prediction of Den2 NS3 protease based on Approach 2 (see Method section): combination of automated prediction tools, threading programs and observed secondary structure information (taken from structural templates) used in threading methods (obtained from DSSP database). The templates were chosen based on E-values (E-value < 0.05, highly confident; E-value ≤ 1.00, worthy of attention) or Z-scores (Z-score > 3, high confidence). Template sequences used in the analyses: 1dy9 - Hepatitis C virus NS3 protease/helicase, 1svp - Sindbid virus capsid protein, 1vcp - Semiliki Forest virus capsid protein, 1jxp - Human Hepatitis C virus NS3 protease, 1ep5 - Venezuelan Equine Encephalitis virus capsid protein, and 1cqq - Type 2 Rhinovirus 3C protease. Codes for secondary structures are given by: E = extended β-strand, B = residue in isolated β-bridge, H = α-helix, G = 3/10 helix, T = hydrogen-bonded turn, S = bend, L = loop and C = coil [Kabsch and Sander, 1983]. Regions with no codings are noted as coils. Consensus of secondary structure is given as E, H or C. Underlined secondary structures in the consensus line refers to regions which differ from 2FOM. (2FOM: secondary structure of crystallised Den2 NS3 protease available on PDB).

Comparison of the prediction results with secondary structure of crystallised NS3 protease of Den2 (2FOM), shows that the second approach yields higher accuracy (A = 76%) compared to the use of prediction servers only (A = 72.67%). These results show that the application of programs used in the second approach gives a more reliable 1D secondary structure profile. Difference in secondary structure between both approaches was calculated to be 7.43%. The alignment of various predictions provides much more structural information than a pairwise alignment. As opposed to the suggestions made by Russell, 2002, the techniques used in this study do not involve the usage of human insight to reach the consensus. With only a 'palmful' of knowledge on the Den2 NS3 protein structure during the early part of this study, we wondered about the closeness of the prediction results to the true secondary profile of the system. However, the recently published crystal model of Den2 NS2B/NS3 protease [Erbel *et al.*, 2006], which resolved to 1.5 Å, assured the quality of the prediction results at 76% accuracy for the results from the second approach against the secondary profile of the crystal structure. This accuracy value is reasonably high. The application of the second approach towards the building up of secondary structure profile of proteins can also be suggested to other groups of proteins from the same family of virus, i. e. Flaviviridae (which can further be classified into 3 genus). Nevertheless, the percentage of accuracy of the prediction results compared with the true structure will also depend on the prediction tools chosen. The selection of the appropriate prediction tools can be determined by referring to CASP (the Critical Assessment of Structure Prediction) [Bourne, 2003] experiments.

The crystal structure of Den2 NS2B/NS3 revealed that the protease adopt a β-barrel fold [Erbel *et al.*, 2006]. It was also reported that residues 51-57 of the cofactor NS2B contributed one β-strand to the N-terminal β-barrel. Furthermore, the crystal structure of West Nile virus NS2B/NS3, complexed with inhibitor, revealed direct interactions of NS2B with active site of NS3, thus underlining the work by Yusof *et al.*, 2000, which indicated the dependence of the protease on the cofactor for cleavage of substrates with dibasic amino acids. However, this study only concentrated on the secondary structure of the protease domain (NS3) and it was assumed, at this stage, that the involvement (or not) of NS2B cofactor in the prediction will not affect the building up of 1D secondary structure profile of the NS3 protease, since alignment of the protein was made with other proteins of the same domain (when using prediction tools).

Analysis of conservation in the protein families is effective in secondary structure predictions performed before knowledge of the protein structure [Cuff and Barton, 1999]. A strong correlation can be made between structurally conserved regions (Fig. 3) and regions with highly conserved sequences across the different proteases (Fig. 1), especially regions surrounding the catalytic triad. However, two regions were missed by the prediction methods, where 2FOM defined residues Trp50-Arg54 and Ser131-Ser135 (each carrying residues comprising the catalytic triad, i.e. His51 and Ser135, respectively) to be α-helices. On the other hand, it is quite unexpected for α-helix to span the region Ser131-Ser135 which also included Pro132, and proline is well-known to be a 'helix-breaker' due to its rigid ring conformation [Garrett and Grisham, 1997]. Looking at the prediction results in Fig. 3 for region Ser131-Ser135, there is 100% consensus for this region to be coiled. Perhaps the α-helices will be recognised after the secondary structure profile is put through fold recognition procedures.

Conclusions

This study illustrates the application of automated prediction servers and information of secondary structures from threading programs to build a consensus of secondary structure (1D)

of Den2 protease. In an early stage, the secondary structure prediction was carried out before the PDB deposition of the crystal structure (2FOM) for NS2B/NS3. Soon after the publication of the 3D-structure in the year 2006, we conducted a validation of the predictive power of the tools. The conclusions were drawn by comparing a combination of tools (approaches 1 and 2) against the observed secondary structure of 2FOM. The consensus obtained in the comparison studies showed higher similarity in approach 2 than approach 1 to 2FOM (Fig. 4). In the present case and presumably other cases of low homology sequence relationships, a significantly better prediction can be attributed to approach 2 because it comprises heuristic (similarity of amino acid properties), probabilistic (from structural data collections) and more sophisticated overlaying techniques, i.e. threading the query into the template backbone to detect hidden phylogenetic resemblance.

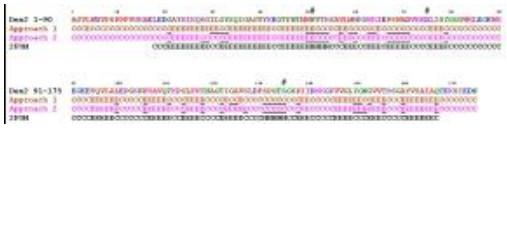


Figure 4: Alignment of secondary structure prediction consensus obtained from approaches 1 and 2, against the secondary structure of 2FOM. Underlined secondary structures in the consensus lines refer to regions which differ from 2FOM. The catalytic triad residues: His51, Asp75 and Ser135, are labelled with the symbol #.

Acknowledgements

The authors thank Mohd Firdaus Raih from the National Institute for Genomics and Molecular Biology (Interim Laboratory), and School of Biosciences and Biotechnology, Universiti Kebangsaan Malaysia, Malaysia, for his kind intellectual contribution towards this study. This work was supported in part by the Malaysian Ministry of Science, Technology and Innovation under the Top Down National Biotechnology Directory grant number 09-02-04-001BTK/TH/004 [UM 36-02-03-6008], and the Academy of Science, Malaysia, under the Scientific Advancement Fund Allocation (SAGA 66-02-03-0049).

References

- Alexandrov, N. N., Nussinov, R. and Zimmer, R. M. (1995). Fast protein fold recognition via sequence to structure alignment and contact capacity potentials. In: Pacific Symposium on Biocomputing '96, Hunter, L. & Klein, T. E. (eds), World Scientific Publishing Co., Singapore, pp.53-72.

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**, 3389-3402.

- Barton, G. J. (1995). Protein secondary structure prediction. Curr. Opin. Struct. Biol. **5**, 372-376.

- Bazan, J. F. and Fletterick, R. J. (1989). Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. Virology **171**, 637-639.

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). The Protein Data Bank, Nucleic Acids Res. **28**, 235-242.

- Bourne, P. E. (2003). CASP and CAFASP experiments and their findings. Methods Biochem. Anal. **44**, 501-507.

- Brinkworth, R. I., Fairlie, D. P., Leung, D. and Young, P. R. (1999). Homology model of the dengue 2 virus NS3 protease: putative interactions with both substrate and NS2B cofactor. J. Gen. Virol. **80**, 1167-1177.

- Bryson, K., McGuffin, L. J., Marsden, R. L., Ward, J. J., Sodhi, J. S. and Jones, D. T. (2005). Protein structure prediction servers at University College London. Nucleic Acids Res. **33**, W36-W38.

- Büchen-Osmond, C. (2003). The Universal Virus Database ICTVdB. Comput. Science Engineer. **5(3)**, 16-25.

- Cuff, J. A. and Barton, G. J. (1999). Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. Proteins **34**, 508-519.

- Cuff, J. A. and Barton, G. J. (2000). Application of enhanced multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins **40**, 502-511.

- D'Arcy, A., Chaillet, M., Schiering, N., Villard, F., Lim, S. P., Lefeuvre, P. and Erbel, P. (2006). Purification and crystallization of dengue and West Nile virus NS2B-NS3 complexes. Acta Cryst. Sec. F **62**, 157-162.

- Erbel, P., Schiering, N., D'Arcy, A., Renatus, M., Kroemer, M., Lim, S. P., Yin, Z., Keller, T. H., Vasudevan, S. G. and 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.

- Falgout, B., Pethel, M., Zhang, Y.-M. and Lai, C.-J. (1991). Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J. Virol. **65**, 2467-2475.

- Garrett, R. H. and Grisham, C. M. (1997). Proteins: Secondary, tertiary, and quarternary structure. *In*: Principles of biochemistry, with a human focus, Vondeling, J. and Kiselica, S. (eds.), Thomson Learning, Inc. Thomson Learning, USA, pp. 113-152.

- Irie, K., Mohan, P. M., Sasaguri, Y., Putnak, R. and Padmanabhan, R. (1989). Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). Gene **75**, 191-211.

- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. **292**, 195-202.

- Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers **22**, 2577-2637.

- Kabsch, W. and Sander, C. (1984). On the use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations. Proc. Natl. Acad. Sci. USA **81**, 1075-1078.

- Kautner, I., Robinson, M. J. and Kuhnle, U. (1997). Dengue virus infection: epidemiology, pathogenesis, clinical presentation, diagnosis and prevention. J. Pediatr. **131**, 516-524.

- Kelley, L. A., MacCallum, R. M. and Sternberg, M. J. E. (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. **299**, 499-520.

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

- Lee, Y. K., Othman, R., Abdul Wahab, H., Yusof, R. and Rahman, N. A. (2006). A revisit into the Den2 NS2B/NS3 virus protease homology model: Structural verification and comparison with crystal structure of HCV NS3/4A and Den2 NS3. Malaysian J. Sci. **25**, 15-22.

- Lund, O., Frimand, K., Gorodkin, J., Bohr, H., Bohr, J., Hansen, J. and Brunak, S. (1997). Protein distance constraints predicted by neural networks and probability density functions. Protein Eng. **10**, 1241-1248.

- Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F. and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. **29**, 291-325.

- Marti-Renom, M. A., Fiser, A., Madhusudhan, M. S., John, B., Stuart, A., Eswar, N., Pieper, U., Shen, M.-Y. and Sali, A. (2003). Modeling protein structure from sequence. Curr. Prot. Bioinfo., John Wiley & Sons, Inc. 5(3), 5.1.1-5.1.42.

- McGuffin, L. J., Bryson, K. and Jones, D. T. (2000). The PSIPRED protein structure prediction server. Bioinformatics **16**, 404-405.

- Meller, J. and Elber, R. (2001). Linear optimisation and a double statistical filter for protein threading protocols. Proteins **45**, 241-261.

- Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. **247**, 536-540.

- Pan, X.-M., Niu, W.-D. and Wang, Z.-X. (1999). What is the minimum number of residues to determine the secondary structural state? J. Protein Chem. **18**, 579-584.

- Raghava, G. P. S. (2002). APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. CASP5, A-132.

- Rawlings, N. D., Morton, F. R. and Barrett, A. J. (2006) MEROPS: the peptidase database. Nucleic Acids Res. **34**, D270-D272.

- Rost, B. (1996). PHD: predicting 1D protein structure byprofile based neural networks. Methods Enzymol. **266**, 525-539.

- Rost, B. (2003). Rising accuracy of protein secondary structure prediction. *In*: Protein structure determination, analysis, and modeling for drug discovery, Chasman, D. (ed.), Dekker, New York, pp. 207-249.

- Rost, B., Yachdav, G. and Liu, J. (2004). The PredictProtein Server. Nucleic Acids Res. **32**, W321-W326.

- Russell, R. (2002). A guide to structure prediction, v2.1. http://speedy.embl-heidelberg.de/gtsp/.

- Sander, C. and Schneider, R. (1991). Database of homology derived protein structures and the structural meaning of sequence alignment. Proteins **9**, 56-69.

- Teodorescu, O., Galor, T., Pillardy, J. and Elber, R. (2004). Enriching the sequence substitution matrix by structural information. Proteins **54**, 41-48.

- Thompson, J. D., Higgins, D. G. and Gibson, T. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**, 4673-4680.

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

- Yusof, R., Clum, S., Wetzel, M., Murthy, H. M. K. and Padmanabhan, R. (2000). Purified NS2B/NS3 serine protease of Dengue virus type 2 exhibits cofactor NS2B dependence for the cleavage of substrates with dibasic amino acids in vitro. J. Biol. Chem. **275**, 9963-9969.

Malaysian Journal of Science 25(1): 15-22 (2006)

A Revisit into the DEN2 NS2B/NS3 Virus Protease Homology Model: Structural Verification and Comparison with Crystal Structure of HCV NS3/4A and DEN2 NS3

Lee Yean Kee,[1] Rozana Othman,[1] Habibah Abdul Wahab,[2] Rohana Yusof[3] and Noorsaadah Abd. Rahman.[1,*]

1. Dept. of Chemistry, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur
2. School of Pharmacy, Universiti Sains Malaysia, Pulau Pinang
3. Dept. of Mol. Medicine, Faculty of Medicine, Universiti Malaya, 50603 Kuala Lumpur.

ABSTRACT

Although the crystal structure of DEN2 NS3 serine protease has been reported, the proteolytic mechanism of this enzyme in the presence of NS2B as co-factor which greatly enhance its activity is not well-understood. Using the homology model of DEN2 NS3 co-complexed with NS2B co-factor based on HCV NS3/4A as the proposed template, the model of the DEN2 NS2B/3 complex was reproduced and its structure evaluated through PROCHECK, VERIFY3D and ERRAT. Comparison of the homology model with the crystal structure of DEN2 NS3 revealed that this homology model served as a better choice for the NS2B/3 protein model.

Keywords: Dengue, NS3, serine protease, structural verification, HCV.

INTRODUCTION

Dengue fever and dengue haemorrhagic fever is a serious disease with almost 40% of the world population currently at risk. This disease is easily spread by the *Aedes aegypti* and *Aedes albopictus* mosquitoes which are highly populated in tropical country. Studies revealed that about 50 million cases of dengue infection were reported every year. There is currently no available drug or vaccine to combat this disease. However, various efforts are currently being put into finding effective vaccines or inhibitors for this disease (Kautner et al., 1997; Monath, 1994).

The dengue fever and dengue haemorrhagic fever is caused by the dengue virus that grouped under the flavivirus family. Currently, 4 serotypes of dengue virus have been discovered with the DEN2 serotype being the most prevalent amongst the four. The single-

stranded, positive-sense RNA of dengue virus has approximately 10,723 nucleotides (for New Guinea-C strand). The genomic RNA has a single open reading frame that encodes a polyprotein of 3,391 amino acids. These amino acids are processed into 3 structural (C, prM, and E) proteins that are further assembled into the virion and seven non-structural proteins, NS1 to NS5, and are expressed in infected cells (Irie et al., 1989). Studies has revealed that the second largest protein encoded by virus, NS3 to contain a serine proteinase catalytic triad within terminal region of 180 amino acid residues which require 40 amino acid residues of NS2B for protease activity (Falgout et al., 1993).

The polyprotein precursor processing occurs co-translationally as well as post-translationally and is performed by either the host signalase in association with the membranes of the endoplasmic reticulum or the viral protease. The protease cleaves the viral polyprotein at four junctions, NS2A-NS2B (Arg-Ser), NS2B-NS3 (Arg-Ala), NS3-NS4A (Lys-Ser), and NS4B-NS5 (Arg-Gly) where a pair of dibasic amino acids at the P2 and P1 positions followed by a small, non-branched amino acid at P1' was found as the consensus of substrate cleavage motif (Yusof et al., 2000 and references therein).

Many approaches have been employed to understand the mechanism, structure and molecular interaction between the serine protease of NS2B/3 complex and its substrates. The minimum domain size required for protease activity of the 69-kDa NS3 protein has been mapped to 167 residues at the N terminus (Li et al., 1999). The virus sequence alignments analyzed revealed that structural motifs as well as the characteristic catalytic triad (His-Asp-Ser) of mammalian serine proteases are conserved in all flaviviruses (Bazan and Fletterick, 1989; Gorbalenya et al. 1989). Sequence comparison among the serine proteases and mutational analysis verified that a catalytic triad of NS2B/3 comprised of the residues His51, Asp75, and Ser135, and that replacement of the catalytic Ser135 residue by alanine resulted in an enzymatically inactive NS3 protease (Yan et al., 1998). The presence of a peptide co-factor is also found to be essential for optimal catalytic activity of the flaviviral proteases with natural polyprotein substrates (Bartenschlager et al., 1995; Chambers, 1991).

Although the dengue virus NS3 protease exhibits NS2B-independent activity with model substrates for serine proteases such as *N*-α-benzoyl-L-arginine-*p*-nitroanilide, the enzymatic cleavage of dibasic peptides is markedly enhanced in NS2B/3 complex. In addition, the presence of the NS2B co-factor has been shown to be an absolute requirement for *trans*-cleavage of a cloned polyprotein substrate (Yusof et al, 2000). A genetically engineered NS2B(H)-NS3pro protease containing a non-cleavable nonamer glycine linker between the NS2B activation sequence and the protease moiety exhibited higher specific activity with *para*-nitroanilide peptide substrates than the NS2B(H)-NS3pro molecule (Leung et al., 2001). The NS2B-NS3pro protease incorporating a full-length NS2B cofactor sequence could catalyze the cleavage of 12-mer peptide substrates representing native polyprotein junctions (Khumthong et al., 2002; Khumthong et al., 2003). However, this protein appeared to be completely resistant to proteolytic self-cleavage.

A model of the NS2B/NS3 dengue virus protease was first constructed through homology modeling using the crystal structure of HCV NS3/4A complex as template by Brinkworth and co-workers, with the suggestion that the 40 amino acid residues of dengue virus NS2B co-factor could be reduced to 12 hydrophobic residues (Brinkworth et al., 1999). Experimental data on hepatitis C virus protease showed some structural and mechanistic explanations for the protease activation by its co-factor, where the NS4A co-factor was found to affect the folding of the NS3 protease. This resulted in the conformational rearrangements of the N-terminal 28 residues of the protease and a strand displacement that lead to the formation of a well-ordered array of three β-sheets with the co-factor as an integral part of the protease fold (Kim et al., 1996; Yan et al., 1998). These conformational changes reorient the residues of the catalytic triad making it more favorable for proton shuttling during proteolytic process.

In the absence of the crystal structure for NS3 pro complexed with its co-factor NS2B, it is rather difficult to understand the structure and conformation of the dengue virus serine protease as well as its mechanism of action. Comparative modeling could provide an alternative method to understand the structure and conformation of protease and further aid in

predicting the binding interactions of the substrate with competitive substrate-based or non-substrate-based inhibitors.

Although the model of the NS2B/3 complex of the dengue virus protease has been proposed based on a homology modeling on HCV (Brinkworth et al., 1999), the verifications of this model were not extensively carried out. In addition, there was no detailed discussion on the comparisons between the homology structure and the dengue NS3 crystal structure to enable better insights into the role of NS2B co-factor in the influencing the function of NS3. In this work, the dengue virus protease model was reproduced through homology modeling and is subjected to structural verification and evaluation using server-based protein structure verification program, namely PROCHECK, VERIFY 3D and ERRAT. The results of these verifications are discussed in greater details in this paper. The similarities and differences between the crystal structures and computer model of NS3 generated in this work as well as that of Brinkworth's (Brinkworth et al., 1999) are also discussed.

MATERIALS AND METHODS

Homology model of DEN2 NS2B/3 Serine Protease

Homology model of NS2B/3 of dengue virus type 2 was built using the HCV serine protease NS3/4A (pdb ID: 1jxp) as the template. The Modeller (mod6v2) software package was used to perform model building. The sequence alignment was done based on the published results of Brinkworth *et. al*., 1999. The quality of the backbone of rough model generated from Modeller was then evaluated using PROCHECK (Laskowski *et. al*., 1993), VERIFY3D (Bowie *et. al*. 1991) and ERRAT (Colovos and Yeates, 1993) on the UCLA bioinformatics server. Energy minimization (100 steps of steepest decent plus 50 steps of conjugate gradient) was performed onto the model, using Hyperchem software package (Hypercube, Inc.) to reduce the bumps and bad contacts while keeping the backbone of the protein restrained. The model evaluation was then repeated.

Comparison of the homology model with crystal structures of and DEN2 NS3 and HCV NS3/4A

The similarities and differences of the structure and conformation around the catalytic triad in of the constructed homology model of DEN2 NS2B/3 serine protease were evaluated using the crystal structures of DEN2 NS3 (pdb ID: 1bef) and the HCV NS3/4A (pdb id: 1jxp).

RESULTS

Homology model building and model evaluation

The Ramachandran plot obtained from PROCHECK (Figure 1) showed an overall 100 % non-glycine residue to be in the allowed region, which implies a good protein backbone structure and folding, where the distribution of the φ/ϕ angle of the model were within the allowed region. In addition, VERIFY 3D showed 90.4% of the residues having a 3D-1D score greater than 0.2, suggesting a reasonable conformation of the residues in the model. However, region with a 3D-1D score lower than 0.2 was found in the range of Glu91-Gln110, indicating a lower confidence in its conformations and folding, implying a lower homology between DEN2 serine protease and HCV serine protease in this particular region. ERRAT was used to check the non-bonded structures of the queried structure and to compare with a database of reliable high-resolution structures. The DEN2 N2SB/3 homology model gave 77.1% overall quality factor of the sequence to be below 95% rejection limit for each chain in the input structure (see figure 1, 2 and 3 for more details) after several iterations of geometry optimizations. This indicated a better three-dimensional profile of the protein from pre-generated homology structure (data not shown). All these verification procedures performed on the NS2B/3 protease model indicated that the model has reached a satisfactory fold quality. Hence, no further loop modeling was carried out on the model.

DISCUSSIONS

Comparison of the homology model with crystal structures of and DEN2 NS3 and HCV NS3/4A

Overall, the homology model showed almost the same folding of protein backbone as observed in the DEN2 NS3 crystal structure where in the first domain (NH_2-terminal domain) of NS3, one alpha-helix and 6 beta sheets are found, both in the homology model as well as

in the DEN2 NS3 crystal structure (Figure 4). The differences between two models, however, were observed in the second NS3 domain where more loop regions are found in the crystal structure of NS3 compared to those observed in the homology model. In addition, the crystal structure of NS3 indicated the presence of only one alpha helix and 7 beta sheets in C terminal region, whilst one extra beta sheets in the same domain was observed in the homology model.

It has been reported that in the crystal structure of NS3 with NS4A co-factor of HCV, the incorporation of NS4A co-factor into the N-terminal domain β-sheet led to a more rigid and precise framework for "prime-side" substrate binding channel residues, thus providing a better catalytic cavity that makes the NS3 enzyme active in proteolytic process (Kim et al., 1996). Superimposition (Figure 2d) of the NS3 structure from the crystallographic data with that of the homology model revealed a difference in the folding between Gly114-Val126 of crystals of NS3 compared to the homology model. These may explain the importance of the NS2B co-factor's role in re-packing the NS3 protein into a more rigid and stabilized structure, especially at the C- terminal domain, where more secondary structure was observed as compared to the NS3 crystal in the absence of the NS2B co-factor.

The catalytic triad residues for HCV NS3/4A and DEN2 NS2B/3 serine proteases were found to be in the structurally conserved region and there is no significant conformational differences observed between them. The RMSD value found between the catalytic triad residues of the HCV NS3/4A (His57, Asp81 and Ser139) crystal with the homology model of DEN2 NS2B/3 (His51, Asp75 and Ser135) is 0.6 while that of the homology model of DEN2 NS2B/3 and the DEN2 NS3 crystal structures is 1.1. The hydrogen bonding between the hydroxyl group of Ser135 and cycloimine of His51 sidechain was observed in the catalytic triad of the reported Den2 NS3 crystal structure (Figure 5a). The side chain carboxyl oxygen of Asp75, however, is observed to be oriented away from His51 (Figure 5a). This caused it to be unable to form a hydrogen bond between carboxyl group of Asp75 and cycloamine of His51 and hence, the proton transfer from Asp75 to Ser 135 which is required to activate the proteolytic process could not transpire.

In the homology model, the carboxyl oxygen of Asp75 and His51, as well as that of His75 and the hydroxyl of Ser135, was found to be at 1.6 Å which is within the hydrogen bonding distance (Figure 5b), suggesting a better arrangement of catalytic residues.

As demonstrated in Table 1, all the structures gave a reasonable reading of the Ramachandran plot, where more than 90% of non-glycine residues were located in the allowed region and no residues were located in disallowed region was observed in all the protein structures. This indicated that the backbone of the serine protease of HCV and DEN2 have reasonably high degree of homology, in spite of their low sequence identity (Brinkworth et al., 1999). Interestingly, the homology model for DEN2 NS2B/3 displayed a better reading in VERIFY3D and ERRAT as compared to crystal structure of NS3, suggesting a better side chain packing in the computer model. The absence of the NS2B co-factor fragment in the crystal structure of NS3 is attributed to a lower quality 3D structure of the crystals. This information has provided some insights into the role of the protease co-complexed with NS2B co-factor, which seems to re-orientate the active pocket of the DEN2 NS3, exhibiting a better side chain packing for a more efficient proteolytic cleavage (Yusof et al, 2000). However, it is still not viable to use the NS3 crystal structure as a template for generating the model structure of NS2B/NS3 dengue virus protease since the structural verification studies of various methods performed showed low confidence in the structural information. On the other hand, structural verifications performed on the crystal structure of HCV NS3/4A showed a remarkably high level of confidence. Hence, the homology model generated using HCV crystal structure as a template should provide a better and more accurate picture of the DEN2 serine protease structure.

CONCLUSIONS

In summary, the homology model of DEN2 NS2B/3 complex serine protease has provided information on the similarities and differences between the structures of HCV NS3/4A crystal and uncomplexed DEN2 NS3 crystal. The crystal structure of HCV NS3/4A appears to be a better choice to be used as a template in order to generate a model for DEN2 serine protease since results indicated better structure verification values for HCV NS3/4A than that of DEN2 NS3 crystal.

ACKNOWLEDGEMENT

The authors acknowledge financial support for this project provided by the Malaysian Ministry of Science, Technology and Innovation under the *Top Down National Biotechnology Directory* grant no 09-02-04-001BTK/TH/004 (UM 36-02-03-6008).

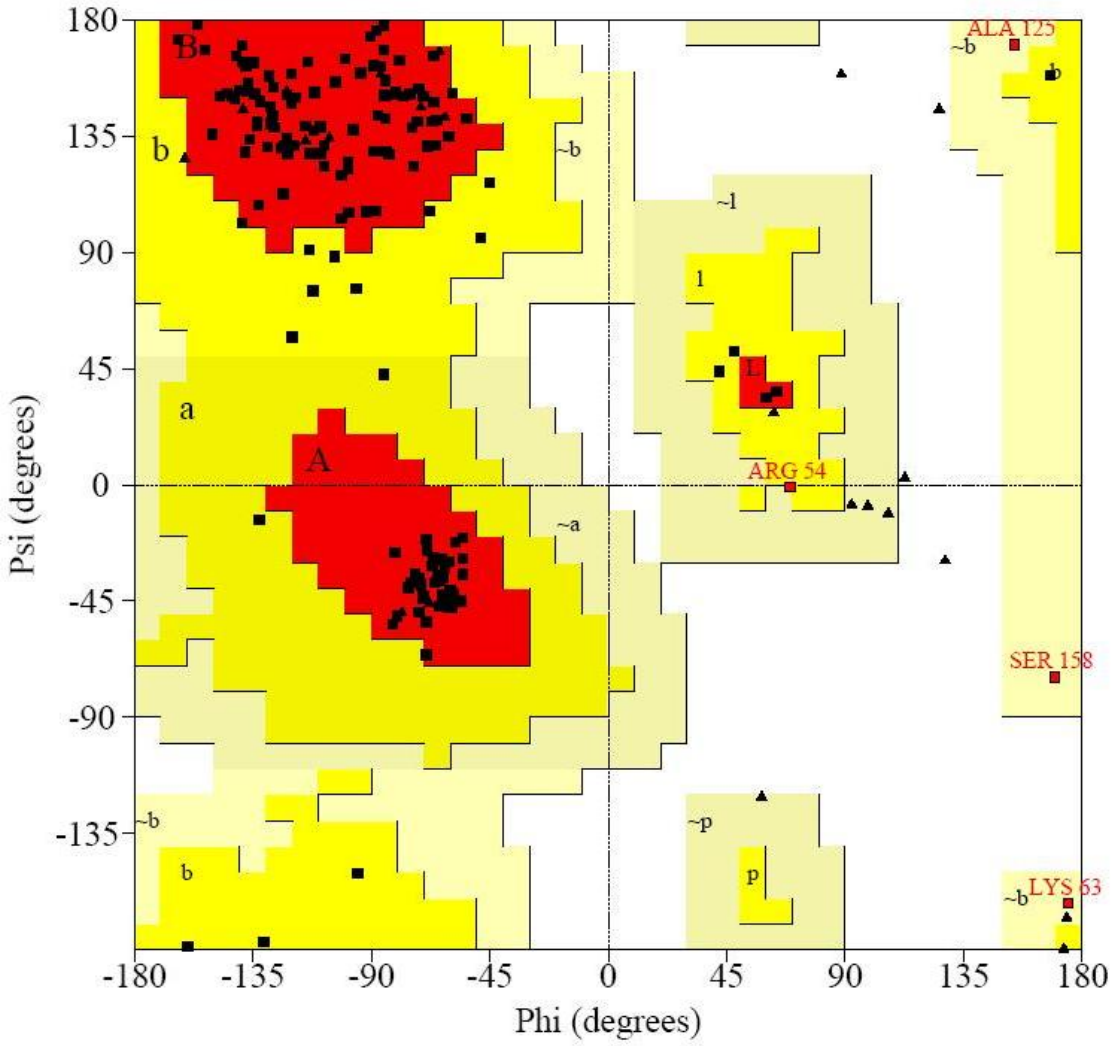


Figure 1: Ramachandran plot of built homology model of DEN2 NS2B/3 complex.

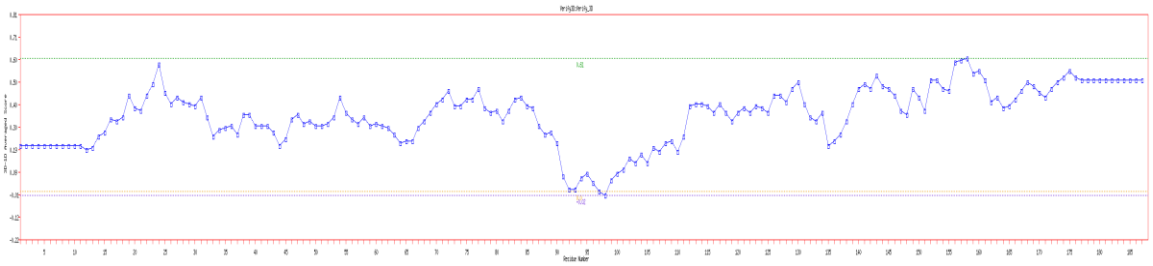


Figure 2: Verify 3D plot of NS2B/3 homology model

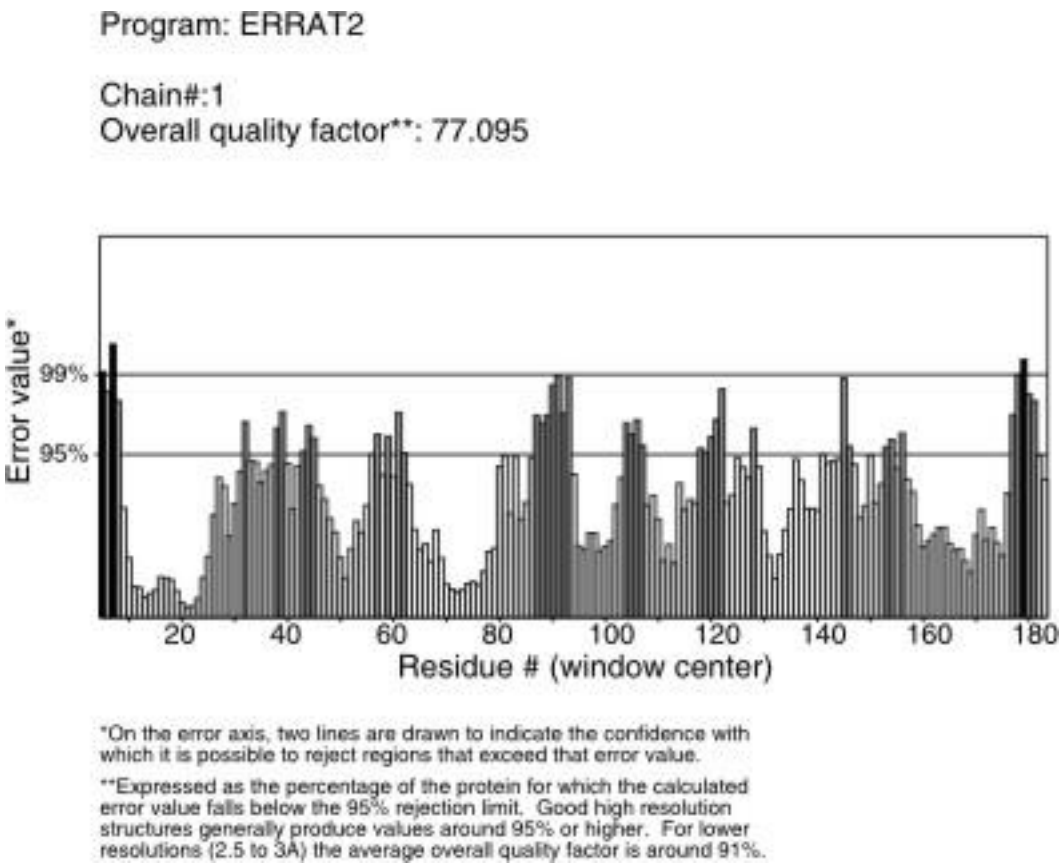


Figure 3: ERRAT analysis of NS2B/3 homology model

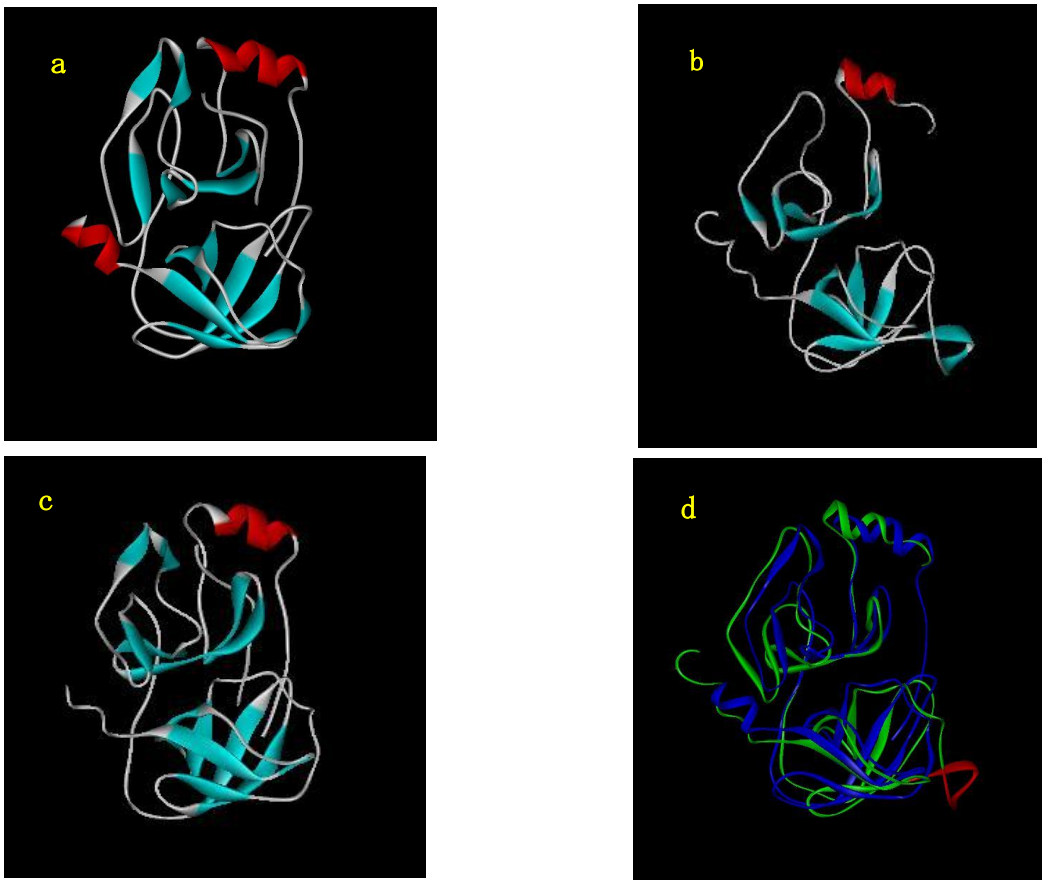


Figure 4: Structures of flavivirus serine proteases; a: DEN2 NS2B/3 complex homology model, b: DEN2 NS3 crystal structure, c: HCV NS3/4A complex crystal structure, d: superimposition of DEN2 NS3 crystal (green) and homology model (blue). In figure 2d, fragment that exhibiting the difference in the protein folding is shown in red.

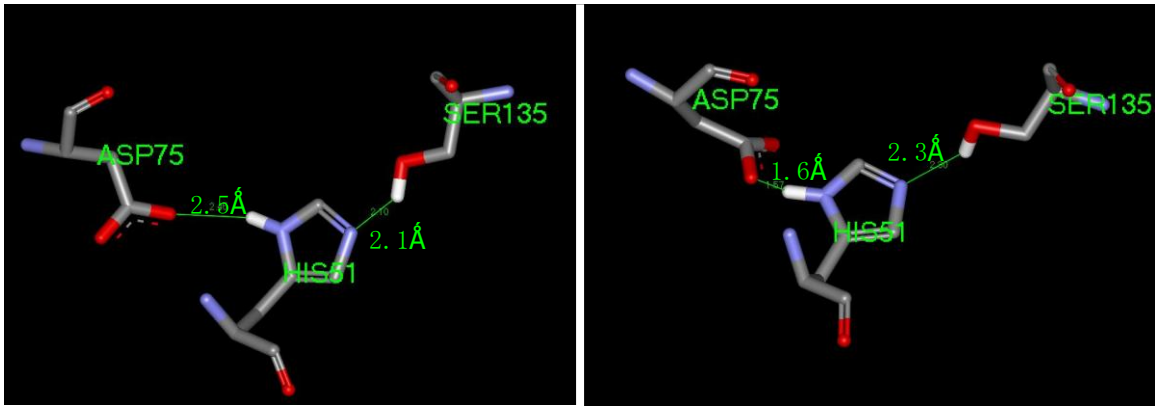


Figure 5: Spatial arrangement of catalytic triad in a: DEN2 NS3 crystal structure (pdb id: 1BEF) and b: DEN2 NS2B/3 complex homology model. Distance between the carboxyl oxygen of Asp75 and His51, as well as that of His75 and the hydroxyl of Ser135 are indicated.

Table 1: Structural verification (PROCHECK, VERIFY3D, ERRAT) comparison between structure of HCV NS3/4A crystal, homology model of DEN2 NS2B/3 and DEN2 NS3 crystal.

Structural Verification	HCV crystal	NS2B/3 Homology model	NS3 crystal
Ramachandran Plot			
Core	80.7	86.3	82.7
allowed	18.0	11.0	15.1
Generously allowed	1.3	2.7	2.2
Disallowed	0	0	0
VERIFY3D	97.4	90.4	52.2
ERRAT	92.0	77.1	49.4

REFERENCES

Bartenschlager, R., V. Lohman, T. Wilkinson, and J. O. Koch. (1995) Complex formation between the NS3 serine-type proteinase of the hepatitis C virus and NS4A and its importance for polyprotein maturation. *J. Virol.* 69:7519–7528.

Bazan, J. F., and Fletterick, R. J. (1989) Detection of a trypsin-like serine protease domain in flaviviruses and pestiviruses. *Virology* 171, 637–639

Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253(5016):164-170

Brinkworth, R. I., D. P. Fairlie, D. Leung, and P. R. Young. (1999) Homology model of the dengue 2 virus NS3 protease: putative interactions with both substrate and NS2B cofactor. *J. Gen. Virol.* 80:1167–1177.

Chambers, T. J., A. Grakoui, and C. M. Rice. (1991) Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. *J. Virol.* 65 (11):6042–6050.

Colovos C, Yeates TO. (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.*, 2(9):1511-1519.

Falgout, B., R. H. Miller, and C.H. Lai (1993) Deletion Analysis of Dengue Virus Type 4 Nonstructural Protein NS2B: Identification of a Domain Required for NS2B-NS3 Protease Activity. *J. Virol.* 67(4): 2034-2042.

Gorbalenya, A. E., Donchenko, A. P., Koonin, E. V., and Blinov, V. M. (1989) N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases. *Nucleic Acids Res.* 17(10): 3889–3897.

Irie, K., Mohan, P. M., Sasaguri, Y., Putnak, R., and Padmanabhan, R. (1989) Sequence analysis of cloned dengue virus type 2 genome (New Guinea-C strain). *Gene,* 75(2): 197–211.

Kautner, I., Robinson, M. J., and Kuhnle, U. (1997) Dengue virus infection: epidemiology, pathogenesis, clinical presentation, diagnosis, and prevention. *J. Pediatr.* 131(4): 516–524.

Khumthong, R., C. Angsuthanasombat, S. Panyim, and G. Katzenmeier (2002) In vitro

determination of dengue virus type 2 NS2B-NS3 protease activity with fluorescent peptide substrates. *J. Biochem. Mol. Biol.* 35(2): 206–212.

Khumthong, R., P. Niyomrattanakit, S. Chanprapaph, C. Angsuthanasombat, S. Panyim, and G. Katzenmeier (2003) Steady-state cleavage kinetics for dengue virus type 2 NS2B-NS3(pro) serine protease with synthetic peptides. *Protein Peptide Lett.* 10(1): 19–26.

Kim, J. L., K. A. Morgenstern, C. Lin, T. Fox, M. D. Dwyer, J. A. Landro, S. P. Chambers, W. Markland, C. A. Lepre, E. T. O'Malley, S. L. Harbeson, C. M. Rice, M. A. Murcko, P. R. Caron, and J. A. Thomson. (1996) Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide. *Cell,* 87(2): 343–355.

Laskowski R A, MacArthur M W, Moss D S & Thornton J M (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26, 283-291.

Leung, D., K. Schroeder, H. White, N. X. Fang, M. J. Stoermer, G. Abbenante, J. L. Martin, P. R. Young, and D. P. Fairlie. (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(49):45762–45771.

Li, H., S. Clum, S. You, K. E. Ebner, and R. Padmanabhan. (1999). The serine protease and RNA-stimulated nucleoside triphosphatase and RNA helicase functional domains of dengue virus type 2 NS3 converge within a region of 20 amino acids. *J. Virol*. 73(4):3108–3116.

Monath, T. P. (1994) Dengue: the risk to developed and developing countries. *Proc. Natl. Acad. Sci. U. S. A.* 91(7): 2395–2400

Valle, R. P. C., and B. Falgout. (1998) Mutagenesis of the NS3 protease of dengue virus type 2. *J. Virol.* 72(1):624–632.

Yan, Y., Y. Li, S. Munshi, V. Sardana, J. L. Cole, M. Sardana, C. Steinkuehler, L. Tomei, R. De Francesco, L. C. Kuo, and Z. Chen. (1998) Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 A resolution structure in a hexagonal crystal form. *Protein Sci.* 7(4): 837–847.

Yusof, R., S. Clum, M. Wetzel, H. M. Murthy, and R. Padmanabhan. (2000) Purified

NS2B/NS3 serine protease of dengue virus type 2 exhibits cofactor NS2B dependence for cleavage of substrates with dibasic amino acids in vitro. *J. Biol. Chem.* 275:9963–9969.