Homology modeling and virtual screening for antagonists of protease from yellow head virus

**Sasimanas Unajak · Orathai Sawatdichaikul · Napat Songtawee ·
Siriluk Rattanabunyong · Anchalee Tassnakajon · Nontawith Areechon ·
Ikuo Hirono · Hidehiro Kondo · Pongsak Khunrae · Triwit Rattanarojpong ·
Kiattawee Choowongkomon**

Abstract Yellow head virus (YHV) is one of the causative agents of shrimp viral disease. The prevention of YHV infection in shrimp has been developed by various methods, but it is still insufficient to protect the mass mortality in shrimp. New approaches for the antiviral drug development for viral infection have been focused on the inhibition of several potent viral enzymes, and thus the YHV protease is one of the interesting targets for developing antiviral drugs according to the pivotal roles of the enzyme in an early stage of viral propagation. In this study, a theoretical modeling of the YHV protease was constructed based on the folds of several known crystal structures of other viral proteases, and was subsequently used as a target for virtual screening—molecular docking against approximately 1364 NCI structurally diversity compounds. A complex between the protease and the hit compounds was investigated for intermolecular interactions by molecular dynamics simulations. Five best predicted compounds (NSC122819, NSC345647, NSC319990, NSC50650, and NSC5069) were tested against bacterial expressed YHV. The NSC122819 showed the best inhibitory characteristic among the candidates, while others showed more than 50 % of inhibition in the assay condition. These compounds could potentially be inhibitors for curing YHV infection.

Keywords Homology modeling · In vitro inhibiting assay · Molecular docking · Molecular dynamics simulations · Virtual screening · Yellow head virus protease

S. Unajak · S. Rattanabunyong · K. Choowongkomon (✉)
Department of Biochemistry, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan Rd, Chatuchak, Bangkok 10900, Thailand
e-mail: fsciktc@ku.ac.th

O. Sawatdichaikul
Institute of Food Research and Product Development, Kasetsart University, 50 Ngam Wong Wan Rd, Chatuchak, Bangkok 10900, Thailand

N. Songtawee
Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand

A. Tassnakajon
Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10300, Thailand

N. Areechon
Department of Aquaculture, Faculty of Fisheries, Kasetsart University, 50 Ngam Wong Wan Rd, Chatuchak, Bangkok 10900, Thailand

I. Hirono · H. Kondo
Laboratory of Genome Science, Graduate School of Tokyo University of Marine Science and Technology, Konan 4-5-7, Minato-ku, Tokyo 108-8477, Japan

P. Khunrae · T. Rattanarojpong
Microbiology Department, King Mongkut's University of Technology Thonburi, Bangmod, Toongkru, Bangkok 10140, Thailand

K. Choowongkomon
Center for Advanced Studies in Tropical Natural Resources, National Research University-Kasetsart University, (CASTNAR, NRU-KU, Thailand), Bangkok, Thailand

Introduction

Yellow head virus (YHV) has been emerging for the past two decades in Thailand. Virulence of the virus is known to cause devastation in shrimp farming 2–3 days after the onset of disease. The infected shrimp exhibits yellow coloration on cephalothorax area and pale color throughout the body. The enormous economical loss of shrimp production in Thailand by YHV infection started from 1991–1995. In the latter years of yellow head disease outbreak, the diagnostic of YHV contamination in post-larvae must be determined prior to culturing. Therefore, the promising preventive methods of the disease have to be considered. The attenuation or inhibition of YHV propagation is one of the preventive methods for developing antiviral agents nowadays. YHV is classified in the order of *Nidovirales*, genus *Okavirus*, and family *Roniviridae*. The rod-shaped virus contains positive single-stranded RNA as a genomic material with approximately 26 kbp. Structural protein components of the virus are composed of three major types including gp116, gp64 spike glycoprotein, and p20 nucleocapsid protein [1, 2]. Among viral genetic variation, the genotypes of YHV are closely related to the Gill-associated virus (GAV) that is less virulent (or provides less virulence). A recent study revealed that YHV are composed of five open reading frames (ORFs) [3]. Among those, the ORF1ab encodes for chymotrypsin-like proteinase (3CLP), nidovirus accessory protease, and RNA-dependent RNA polymerase (RdRp), which have been known to be involved in the replication complexes, and also contained the metal-ion binding domain (MIB) [4]. Chymotrypsin-like proteinase (3CLP) was first and well characterized in GAV. The enzyme employs a "Cys-His catalytic dyad" and possesses *trans*-cleavage activity [4, 5]. The sequence alignment showed that 3CLP of GAV successfully located between I^{2866} and G^{3006} in YHV and interestingly, the putative 3CLP of YHV shared remarkably high identities (up to 87.9 %) in their amino acid sequences. Moreover, the two proteolytic cleavage sites for 3CLP; ^{2838}LVTHE↓VNTGN^{2847} and ^{6455}KVNHE↓LYHVA^{6464} (the cleavage site is indicated by ↓) are conserved between the GAV and YHV sequences. However, the specific role of 3CLP in YHV is yet unclear. It has been postulated that the 3CLP in YHV is involved in the regulation of viral replication complex activity [4], like 3CLP in other Coronaviruses that process the downstream replicase domains including RdRp and helicase enzyme [6]. The other encoding protease that is found in both YHV and CLP is Nidovirus accessory protease that is widely known as papain-like protease (PLP). This enzyme contains the canonical "Cys-His catalytic dyad" and a distinctive α+β fold that is typically found in proteins that function as transcription factor involved in subgenomic messenger RNA (sgmRNA) synthesis [3]. However, the sequence alignment showed that, unlike the 3CLP, the PLP of YHV and GAV share only 73.7 % identities in their amino acid sequences [3].

The prevention of viral propagation has focused on this key enzymatic reaction as well [7, 8]. The inhibition of YHV propagation was only accomplished by YHV-protease dsRNA [9, 10]. Therefore, this viral protease could possibly be a potential target in therapeutic approach, likewise the development of viral protease inhibitors (PI) for HIV-1, HCV, and SARS, illustrate a high achievement in drug pharmaceutical. The 3CLP-specific inhibitors were rationally designed based on the three-dimensional structure of the viral proteases with computer-assisted molecular modeling. The X-ray crystallographic determining the three-dimensional structure thus provide a better in-depth understanding of the association of proteins and their other target peptides. Unfortunately, crystal structure of YHV protease is not yet available; a computer-assisted homology modeling of 3CLP of YHV is currently in progress in order to obtain the molecular basis of protease cleavage site.

Therefore, regarding the successful therapeutic approach targeting viral components for HIV-1, HCV and SARS, the YHV-3CLP protease should be the best target inhibiting viral propagation in YHV. In this study, the homology modeling approach was used to construct the 3D structure of the YHV protease. Molecular docking based virtual screening of synthetic small molecules from the National Cancer Institute (NCI) diversity set against the YHV protease model was performed. The program Autodock4 is used as the docking tool in this study for several reasons. First, AutoDock is widely used among docking programs in the market. In 2006, AutoDock was the most cited docking software. Second, AutoDock is free of charge to access. Last, AutoDock has also been shown to be useful in blind docking in which the location of the binding site is not known. It is very fast, provides high quality predictions of ligand conformations, and good correlations between predicted inhibition constants and experimental ones. The complexes of the protease – hit ligand were subsequently investigated for intermolecular interactions by molecular dynamics simulations.

Materials and methods

Multiple sequence alignment and homology modeling

Sequence analyses, homology modeling and molecular mechanics calculation were performed on the Discovery Studio 2.5 package (Accelrys Inc., CA, USA). The amino acid sequence of the yellow head virus (YHV) protease was retrieved from the NCBI database (www.ncbi.nlm.nih.gov), covering 196 amino acid residues (Ser^{2828} to Asp^{3023}) of the YHV replicase polyprotein 1a (Accession number: ACA21300). All of the template coordinates used for homology modeling were retrieved from the protein data bank (www.pdb.org), which include the eight protease enzymes from both

microorganisms (tobacco etch virus, entry 1LVM; rhinovirus, entry 1CQQ; hepatitis A virus, entry 1HAV; *Staphylococcus aureus*, entry 1AGJ; *Achromobacter lyticus*, entry 1ARB) and higher organisms (bovine β-trypsin, entry 5PTP; human factor B, entry 1DLE; human heparin binding protein, entry 1A7S). All eight template sequences were multiple-aligned according to their structural superimposition of the Cα atoms on the "*align and superimpose proteins*" protocol. The YHV protease sequence was then automatically aligned to a set of aligned template sequences on the "*align multiple sequences*" protocol of which the pairwise alignment algorithm is modified from the ClustalW 1.8 [11]. This alignment protocol allows keeping the existing aligned sequences of the templates as a rigid element and thus the pre-aligned target sequence could be mobile. Sequence identities and similarities between the target and template sequences were calculated on the GeneDoc v2.7 [12]. The structural model of the YHV protease was constructed based on the unconventional sequence alignment between the target and template profile using the Modeller 9v4 [13] implemented in the "*build homology models*" protocol. Insertion and deletions in the target sequence with respect to the templates were defined as loops and further refined within this protocol. The best quality model with highest discrete optimized protein energy (DOPE) score was subsequently subjected to the "*minimization*" protocol based on the charmm22 force field [14] to remove unreasonable atomic contacts (steepest descent and conjugate gradient methods until the model reaching $0.1 \ kcal \ mol^{-1} \ Å^{-1}$ convergence). The stereochemical qualities of the energy-minimized model were assessed by PROCHECK v3.5 [15] via submitting the coordinates to the JCSG structure validation server (www.jcsg.org).

Ligand preparation

The 1364 files (in sdf MDL MOL format) of the NCI diversity dataset were obtained from the Office of the Associated Director of the Developmental Therapeutics Program (DTP), Division of Cancer Treatment and Diagnosis, National Cancer Institute, more information is available at NCI/DTP Open Chemical Repository [16]. As a description from the official NCI/DTP page, these small molecules from NCI dataset have been collected from almost 140,000 compounds on plates. They were screened following the criteria of DTP, which are hydrogen bond acceptor, hydrogen bond donor, positive charge, aromatic, hydrophobic, acid, base, and pharmacophores. Therefore, the selections of 1364 compounds have passed the formal mentioned criteria. Thereby, this set of ligand is suitable to use as a model for finding the antagonists of protease from YHV.

The 3D structure of ligands were prepared by adding hydrogen atoms and rectifying an appropriate bond order and the chiral center check (stereo-chemistry) using LigPrep 2.2 with the default setting [17]. The output 1523 ligand conformations were stored in SYBYL mol2 format. The rotational bonds of ligands were treated as flexible using python script, prepare_ligand4.py. This script was implemented in the AutoDock program [18–20].

Docking procedure

The minimized structure of YHV-protease model was obtained from the previous section (homology modeling and minimization processes). AutoDockTools version 1.5.2 was used to assign the Kollman united atom charges and the solvent parameter to this protein, as well as to perform the grid set-up process. The grid boxes were generated around the catalytic residues (C152, H30, D70; as presented in Fig. 2c). The H30 protonated state was set to be HIP (add H atoms both δ and ε positions) by the N position point to D40. The grid points spacing was 0.375 Å in the box-sizes 60, 60, and 60 Å in each dimension. AutoGrid 4.0 was used to calculate the grid affinity maps for following atom types: A (aromatic carbon), C, HD, N, NA (hydrogen-bond-accepting N), OA (hydrogen-bond-accepting O), S, SA (hydrogen-bond-accepting S), Cl, F, Br, I, P, and e (electrostatic).

The protein-ligand docking performed 50 trial runs using Lamarckian genetic algorithm search. The population size was set to 150. The docking results were sorted by the lowest *estimated free energy of binding* (AutoDock4 score) of each ligand. This score is calculated from the following equation:

$$Estimated \ Free \ Energy \ of \ Binding = Final \ Intermolecular \ Energy + Final \ Total \ Internal \ Energy + Torsional \ Free \ Energy - Unbound \ System's \ Energy$$

At the same time, all of these results were rescored using FRED program [21]. The anchor binding mode of ligands and protein were analyzed using SiMMap server (http://simmap.life.nctu.edu.tw/) [22].

Molecular dynamics simulations

The ligand—YHV protease complexes obtained from molecular docking were subsequently applied to molecular

dynamics (MD) simulations. The AMBER99SB-ILDN force field [23] was used to simulate the protein structures and the ionization state of amino acid residues was set according to the standard protocol. To mimic catalytic conditions of cysteine proteases, the side chain of His30 was protonated at both δ- and ε-nitrogen atoms (HIP form), while the side chain of Cys152 was deprotonated (minus cysteine; CYM form). The N- and C-terminal ends of amino acid chains were capped with acetyl (ACE) and methyl amino (NME) groups, respectively. The GROMACS topology files for each ligand were generated using the ACPYPE script Sousa 2012 [24], in which the general AMBER force field (GAFF) parameters [25] were applied and atomic partial charges were calculated using the AM1-BCC method [26], implemented in the Antechamber module. Each complex was solvated in a rectangular box of TIP3P water [27] keeping a distance of 1.0 nm between the solutes and the sides of the solvent box. Sodium ions were added to neutralize the charge of the system.

The MD simulations were carried out on the explicit-solvent periodic boundary conditions using the GROMACS v4.5.5 [28–30]. Each solvated system was energy-minimized by two steps using the steepest descent method either until the maximum force is smaller than 1000 kJ mol^{-1} nm^{-1} on any atom or until additional steps result in a potential energy change of less than 1 kJ mol^{-1} to reduce undesirable atomic contacts. At first, positioned restraints with a force constant of 1000 kJ mol^{-1} nm^{-1} were applied to all heavy atoms of the proteins and ligands (if presented), allowing water molecules and counterions to relax their position. Second, the restrains on proteins and ligands were released, then allowing all atoms in a system to freely move in turn. Afterward, the energy-minimized systems were equilibrated in three phases with the positioned restraints described earlier. The first step is heating up each system from 50 to 300 K over 50 ps under the NVT condition using the Berendsen thermostat [31, 32]. The following step is conducted under the NPT condition at 1 bar pressure over 100 ps using the Parrinello-Rahman barostat [33, 34]. Once each system was sufficiently equilibrated around the target temperature and pressure, the positioned restraints were then gradually reduced to zero kJ mol^{-1} nm^{-1} with four rounds of 50 ps—NPT simulations. After the equilibrations, an unrestrained dynamics production was subsequently performed under the NPT condition with snapshots collected every 1 ps. For all dynamics runs, the LINCS algorithm [35] was applied to fix all hydrogen related bond lengths, facilitating the use of a 2 f. time step. A short-range nonbonded interaction cut-off distance of 1.0 nm was used. The particle mesh Ewald method was used to account for long-range electrostatics [36, 37].

All dynamics analyses were performed using tools available within the GROMACS suite [38]. Hydrogen bonds observed in the ligand—YHV protease complexes and their percentage occupancy were also analyzed from MD

trajectories using the tool available in the GROMACS suite. A hydrogen bond is defined by the default geometrical criteria: donor-acceptor distance ≤ 0.35 nm and proton-donor-acceptor angle ≤ 30°. The Grace v5.1.2 [39] was used to plot the 2D data and the 3D images were created and rendered with the PyMOL v1.3.1 [40].

YHV inhibition assay

Fluorogenic peptide YHV substrate was dissolved in dimethyl sulfoxide (DMSO) at 25 mM and diluted with YHV reaction buffer (20 mM Tris–HCl pH 7.5, 200 mM NaCl, 1 mM EDTA and 1 mM DTT) to obtain 100 μM YHV substrate. YHV protease inhibitors (gifted from NCI) were resuspended in dimethyl sulfoxide (DMSO) to obtain 6.7 μM and diluted with YHV reaction buffer to obtain 75 nM YHV protease inhibitors. Purified YHV protease enzyme (gifted from Dr. Pongsak Khunrae) was dissolved in YHV reaction buffer (20 mM Tris–HCl pH 7.5, 200 mM NaCl, 1 mM EDTA and 1 mM DTT). The determination of YHV inhibition assay was performed by incubating 30 μl YHV protease with equal volume of 75 nM inhibitor at room temperature for 3 min. Then, fluorogenic peptide YHV substrate diluted in YHV reaction buffer was added. The YHV protease in reaction mixture without inhibitor was used as negative control while the reaction mixture excluded YHV protease was used as positive control. The reactions were measured by using fluorescence microplate reader with excitation/emission at 340/485 nm for 5 min. Rate of reaction was calculated as % activity.

Results and discussion

Structural model of YHV protease

We performed a comparative modeling approach using multiple templates followed by an energy minimization to construct a theoretical model for the YHV protease. Because a multiple alignment is the key step in homology modeling, our model was cautiously done since there is only limited sequence identity to such templates. Ab initio modeling and protein threading are now very effective and invaluable methods to use for structural prediction of non- or low-template identity proteins, and their online version are currently accessible [41–47]. In our case, the search for template coordinates available from the protein data bank performed on several BLAST and threading web-servers did not provide us with the consensus template structures or a satisfied model (data not shown), and none of them exhibits acceptable percentage of identities/similarities to the YHV protease. A tertiary structure of the YHV protease should have significant conserved fold like other proteases, and thus
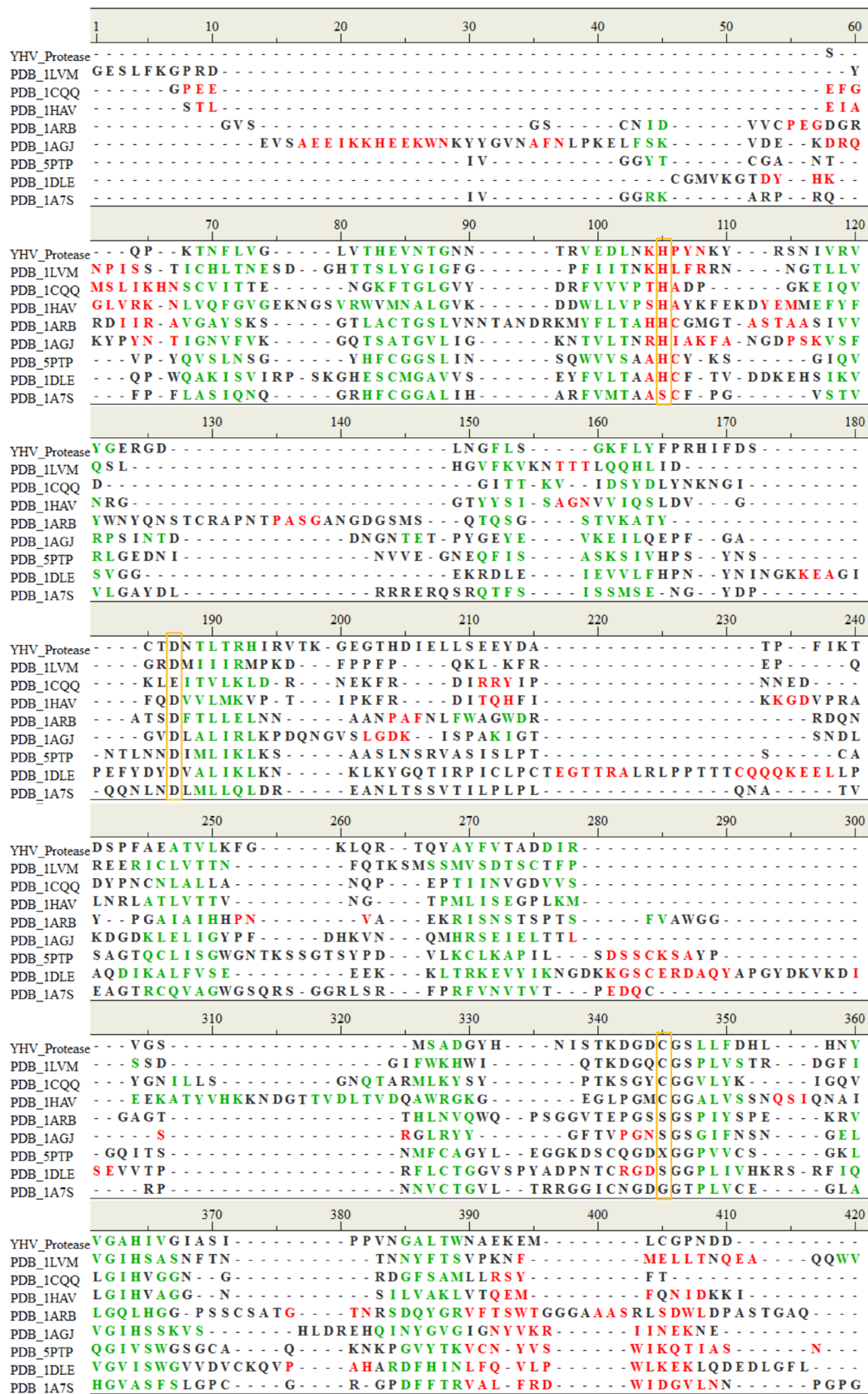
Fig. 1 Structural comparison of YHV protease to eight template proteases. Secondary structure assignment is based on the sequence alignment as described in Methods. The amino acid residues responsible for β-strand and α-helical structures are shown in *green* and *red* colors, respectively. The catalytic triad residues are indicated in *yellow rectangles*

Table 1 Sequence identity (similarity) and Cα-atom RMSD between the YHV protease with templates

PDB template	% Identity (similarity)	Cα-RMSD (Å)
1LVM	16 (34)	2.88
1CQQ	20 (30)	2.56
1HAV	14 (30)	2.8
1AGJ	14 (30)	2.4
1ARB	14 (37)	2.75
5PTP	21 (26)	1.94
1DLE	17 (37)	2.49
1A7S	16 (27)	1.84

allow the development of our own method used for the target-template sequence alignment as described in the Materials and methods section. The model templates used in this study were selected according to an extremely high structural similarity in the two-domain antiparallel β-barrel fold, which is the hallmark of trypsin-like proteases even though these structures exhibit a limited sequence identity to the YHV protease, ranging from 14 to 20 % of the aligned regions (Fig. 1 and Table 1). The selected templates were three cysteine proteases from tobacco etch virus (PDB entry 1LVM) [48], rhinovirus (PDB entry 1CQQ) [49], hepatitis A virus (PDB entry 1HAV) [50], and the other five proteases of non-viral proteins, bovine β-trypsin (PDB entry 5PTP) [51], protease domain of human factor B (PDB entry 1DLE) [52], epidermolytic toxin B from *Staphylococcus aureus* (PDB entry 1AGJ), lysine-specific protease I from *Achromobacter lyticus* (PDB entry 1ARB) [53], and trypsin-like folded human heparin binding protein (PDB entry 1A7S) [54].

After energy minimization of the homology model, results of Ramachandran plots from PROCHECK showed that more than 95 % of the residues were in the most favored and additional allowed regions, indicating a good quality model could be obtained (Supplementary data Fig. S1). This final model adopted the two-domain antiparallel β-barrel fold-like structure resembling the templates (Fig. 2a and b) with less than 3 Å of the Cα-atom root mean square deviation (RMSD) between the model and such templates (Table 1). Although the overall fold of the YHV protease seemed closer to those of the non-viral serine proteases than the three viral cysteine proteases, the atomic coordinates of the catalytic triad residues, His30, Asp(Glu)70, and Cys152 (numbering of YHV protease) were highly conserved among the four viral C-type cysteine proteases; yellow head, tobacco etch, rhino- and hepatitis A viruses (Fig. 2c). These also indicated a high accuracy of the modeled YHV protease and it was reasonable to use this model for subsequent docking experiments.

Structural analysis of 3c-proteases in several species revealed that there were three conserved residues, His, Asp or Glu, and Cys, known as the catalytic triad. They located on the center between two lobes of anti-parallel b-strands. According to the nomenclature of Schechter and Berger [55], these catalytic residues positioned around the S_2 to S'_2 sub-sites of protease enzyme to serve P_2 to P'_2 positions of peptide substrate. Furthermore, most of the 3c-proteases preferred glutamine as the P1 site for substrate, and in some cases the minor preference seems to be glutamic acid [4]. Generally, residues of S_1 pocket are Thr147 and His167 (YHV protease numbering). In addition, the surrounding residues around the S_1 sub-site were also conserved among 3c-protease. This neighbor pocket is named "pocket A" which is composed of Lys148, Asp149, Gly170, Ala172, and Ser173.

Molecular docking with NCI database

In order to identify the small molecules that specifically bind to YHV protease, a docking study analysis was performed
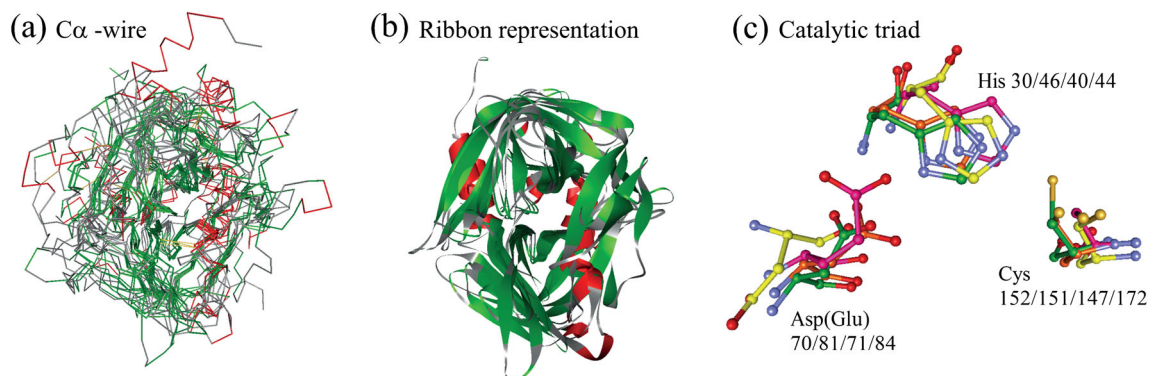
Fig. 2 a Cα-wire representation for the structural alignment of the homology model to the eight templates. The regions responsible for β-stranded and α helical structures are shown in *green* and *red* colors, respect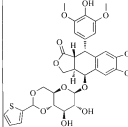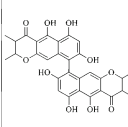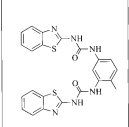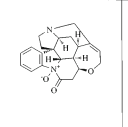ively. **b** Ribbon diagram for such an alignment of the YHV protease on the other three viral cysteine proteases (entry 1LVM, 1CQQ and 1HAV) and **c** only the catalytic triad His/Asp(Glu)/Cys represented in a ball & stick model. The Cα atom of the YHV protease, 1LVM, 1CQQ and 1HAV is shown in *green*, *orange*, *pink*, and *yellow*, respectively
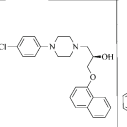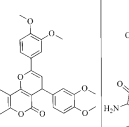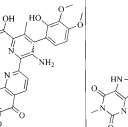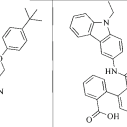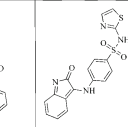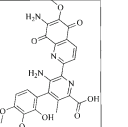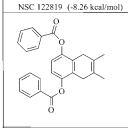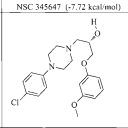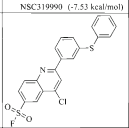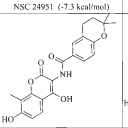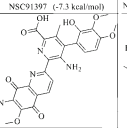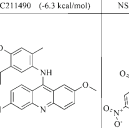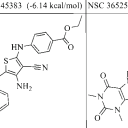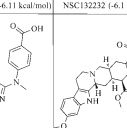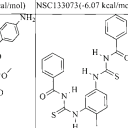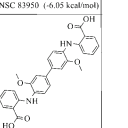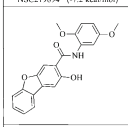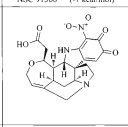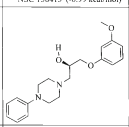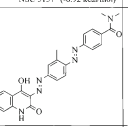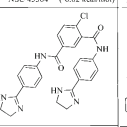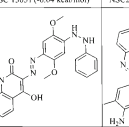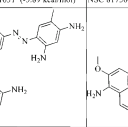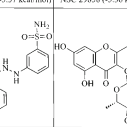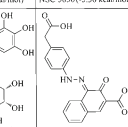
Fig. 3 The catalytic triad residues (*pink ball and stick*), the peptide binding site (*green stick*), moiety anchors of ligand - YHV protease interactions (*green and gray ball meshes* represent hydrogen bond anchor and van der Waals anchor, respectively)
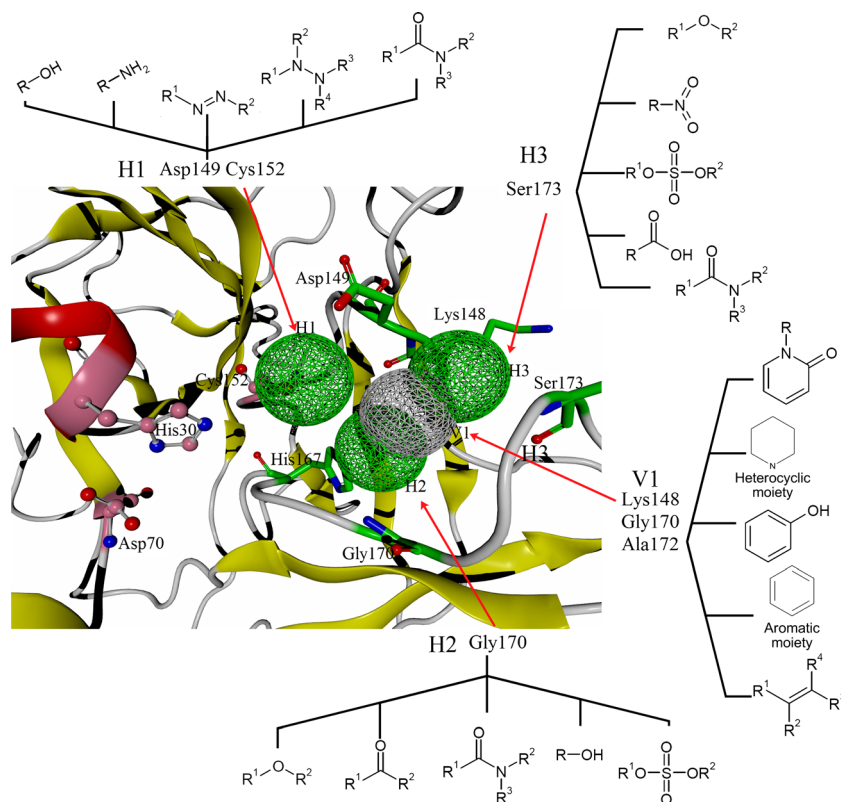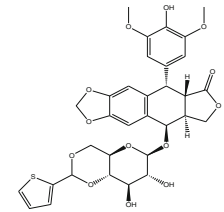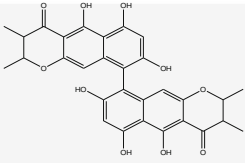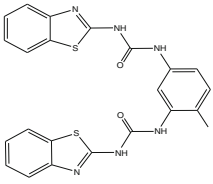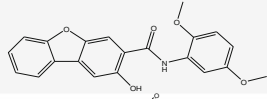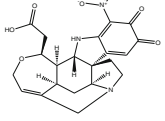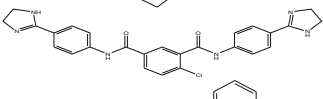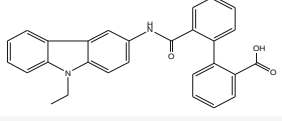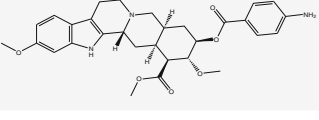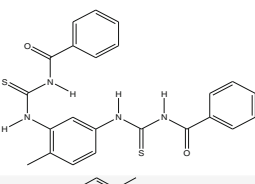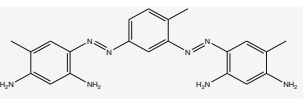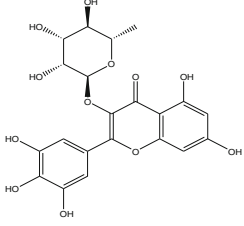

against small molecules from NCI diversity dataset. All 1364 compounds from NCI diversity II were docked into the modeled structure of YHV-protease, 43 compounds were selected from this screening step based on docking scores. The expected results showed several sets of the ligand-protein binding mode. Therefore, the FRED program combinations with SiMMap server (http://simmap.life.nctu.edu.tw/) were used to justify the reliable possibility of the binding mode. The SiMMap server was used to identify the site-moiety map between the functional group(s) of these ligands to the residues around the binding site (Fig. 3). As shown in the SiMMap

moieties, there were four spheres which represent two types of interactions. *H moiety* refered to the hydrogen bonding, while *V moiety* refered to van der Waals force. *Anchor H1* illustrated the binding residues; Asp149 and Cys152 prefer to form hydrogen bonding interactions with the following functional groups; hydroxyl, 1° amine, azo (also named diimide or diazene compound), hydrazine and carboxamide moieties. Similar to anchor *H1*, *H2* specified the consistent functional groups with the position Gly170 which are ether, carbonyl, amide, alcohol, and sulfonyl groups, respectively. For anchor *H3*, Ser173 plays the role that interacts with ether, nitro,

Fig. 4 2D-structure of each compound from group1 and their AutoDock4 scores

Table 2 Eleven compounds which present strong hydrogen bond(s) with surrounding residues around binding pocket

NSC no.	Two dimensional structure	PubChem CID	ChEMBL ID	Protein target (s)
122819		34698	1237078	• DNA topoisomerase I and II (α & β) • Monoamine oxidase A and B
345647		53277	162783	• Ser/Thr protein phosphatase 2A (PP2B) • HIV1-integrase
319990		330740	1578626	• Histone-lysine N-methyltransferase MLL • DNA polymerase beta • Luciferin 4-monooxygenase
50650		67239	156747	• Lysine-specific demethylase 4D-like • Aldehyde dehydrogenase 1A1 • Histone-lysine N-methyltransferase MLL • Sphingomyelin phosphodiesterase
5069		221125		
64736		249307	1578626	• Histone-lysine N-methyltransferase MLL • DNA polymerase beta • Luciferin 4-monooxygenase
132232		280492	1172912	• MAP kinase ERK2 • Dengue virus type 2 NS3 protein • etc.
29858		232432	1326032	• Beta-lactamase • DNA prolymerase iota • etc.
5856		3361436	1426572	• Aldehyde dehydrogenase 1A1 • Putative fructose-1,6-bisphosphate aldolase • Peptidyl-prolyl cis-trans isomerise NIMA-interacting 1 • etc.
47722		24674	1602553	• Pyruvate kinase • DNA polymerase kappa & iota • etc.
19803		5281673	454576	

sulphonyl, carboxy, and carboxamide groups. In additional, the van der Waals interactions between three residues around binding pocket; Lys148, Gly170, and Ala172, and the hydrophobic moieties; phenol, benzene, alkene etc. were represented by *V1 anchor*. These analyses assisted us to understand the mode of binding of ligands and to classify ligand-group which followed the interaction modes.

The 43 screened small molecules can be categorized into three groups. Considering at group one, there are 32 compounds which are able to form hydrogen bonds with residues in the catalytic cleft and/or pocket A (Fig. 4). Moreover, these compounds are also representing cation-π and π-π interactions with those mentioned residues.

Compounds in this group share common aromatic nucleus and also carboxamide group, the peptide-bond-like structure ( ) where its structure mimics peptide bond. Group two is composed of only two compounds which present only cation-π interactions. Compounds in group three consist of various types of core structures. These seven compounds interact with residues around the catalytic cleft with weak van der Waals interactions.

After all viral protease inhibitors with crystal structures were deeply analyzed, the general features of the protease inhibitors always consist of the carboximide group and aromatic moieties and at present are: Leupeptins, Saquinavir, Nelfinavir, Ritonavir, Indianavir, Amprenavir, Lopinavir,

Fig. 5 The comparison between **a** the known protease inhibitor, Nelfinavir and **b** NSC319990 oriented on the binding site of YHV protease, both 2D and 3D views, as well as **c** the superimpose structures and the three common interactions
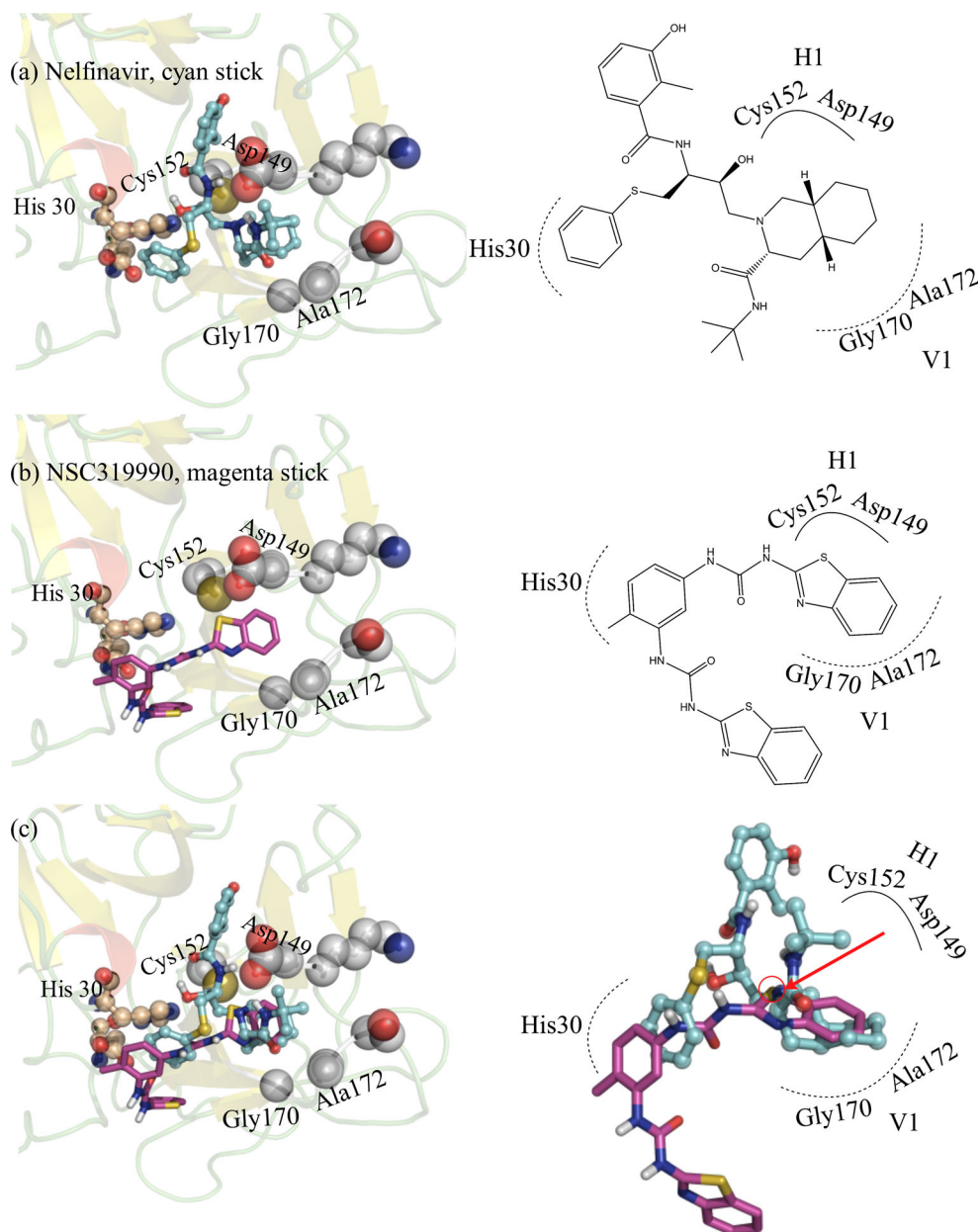
Chymostatins, Antipain, Elastinal, and β-MAPI. These features were also found in our 11 candidate compounds from 32 compounds in group one which contains aromatic moiety, carboximide, hydroxy, carbonyl, ether, and ester groups. All 11 two-dimensional structures as well as properties were presented in Table 2. In order to better understand the mode of actions of these small molecules via strong hydrogen bonds, the orientations of Nelfinavir, a representative of known protease inhibitors, and compound NSC319990 (the best docking score with carboximide functional group, Rank3) have been compared. The post-docking analyses revealed that one-half of NSC319990 is oriented at the same area as Nelfinavir. These two molecules lay along the sub-sites S_2 (H30, S67-D70, I168 and V169) to S_1 (T147, K148, D149 and H167) of YHV protease. There are three common types of binding between these two compounds as presented in Fig. 5. First, π-π interaction between 6-membered aromatic ring and residue His30 (one of the catalytic triad) have been observed from both compounds. Second, the dipole-dipole electrostatic interaction between ligands and binding residues in pocket A, Asp149, and Cys152 were formed. N-atom of decahydroisoquinoline of Nelfinavir, and S-atom of benzo(d)thiazole of NSC319990 held the negative dipole and acted as nucleophile to form interaction with either thiol group of Cys152 or carboxylic group of Asp149. Third, the van der Waals interactions between the residues Gly170 and Ala172, which were defined as *V1 anchor*, and the fused rings of both compounds have been monitored. Considering at the compound NSC319990, as mentioned above, one-half of this compound is perfectly occupied on the binding pocket along the sub-site S_2 to S_1 of YHV protease. Moreover, another-half of NSC319990 lies on the S3 sub-pocket (M137, I168-I171, and A180-T182). These observations suggested that the

screened compounds might possess the protease inhibitor properties as the known one, Nelfinavir. These sets of ligands not only possessed strong hydrogen bonds with the surrounded residues in the binding pocket but also are reported as the potent inhibitory agents targeting several proteins including the malarial parasite plastid (Table 2).

Table 3 List of hydrogen bonds of the YHV protease—hit compound interactions extracted during the last 5 ns from each complex simulation

| Complex	Residue atom		Ligand atom	% occupancy
NSC122819				
	Ala172:NH	–	O11:Lig	31.5
	Ala172:NH	–	O12:Lig	6.7
NSC345647				
	Lys29:Nζ	–	O1:Lig	< 1
	Asn33:Oδ	–	O6:Lig	< 1
	Lys34:Nζ	–	O4:Lig	2.4
	Asp149:Oδ	–	O4:Lig	43.3
	Ile168:CO	–	O5:Lig	3.0
	Gly170:NH	–	O8:Lig	42.3
	Gly170:CO	–	O8:Lig	< 1
NSC319990				
	His30:Nδ	–	N3:Lig	< 1
	Ser67:Oγ	–	O1:Lig	3.7
	Asp70:Oδ	–	N3:Lig	95.8
	Asp70:Oδ	–	N4:Lig	69.5
	Asp149:NH	–	N2:Lig	3.2
	Asp149:Oδ	–	N1:Lig	63.4
	Asp149:Oδ	–	N:Lig	61.6
	Gly170:NH	–	O:Lig	19.4
	Gly170:NH	–	N1:Lig	1.8
	Gly170:NH	–	N2:Lig	< 1
NSC50650				
	Arg133:Nε	–	O1:Lig	21.3
	Arg133:Nε	–	N:Lig	< 1
	Arg133:Nη	–	O1:Lig	< 1
	Gly135:NH	–	O2:Lig	5.6
	Ser136:Oγ	–	O:Lig	< 1
NSC5069				
	Lys148:Nζ	–	O3:Lig	< 1
	Asp149:NH	–	O2:Lig	< 1
	Asp149:NH	–	O3:Lig	76.7
	Asp149:NH	–	O4:Lig	5.5
	Gly170:NH	–	O2:Lig	3.4
	Ser173:NH	–	O5:Lig	2.1
	Ser173:NH	–	N2:Lig	< 1
	Ser173:Oγ	–	O:Lig	< 1
	Ser173:Oγ	–	O1:Lig	< 1
	Ser173:Oγ	–	O5:Lig	< 1

Fig. 6 Potential energies for interactions between the YHV protease and each five hit compound. The electrostatic and vdW contributions are shown in *black* and *white*, respectively. *Error bars* represent the standard deviation in the sum of the interactions. The data were extracted from the last 5 ns of the simulations

Fig. 7 Representation for the interactions between compound NSC319990 and YHV protease. The compound is shown in stick (C atoms in cyan) and three strong binding residues (Asp70, Asp149, and Gly170) are indicated. Hydrogen bonds are shown in *magenta dotted lines*. Distances of the important hydrogen bonds were plotted over the course of 15 ns simulation time
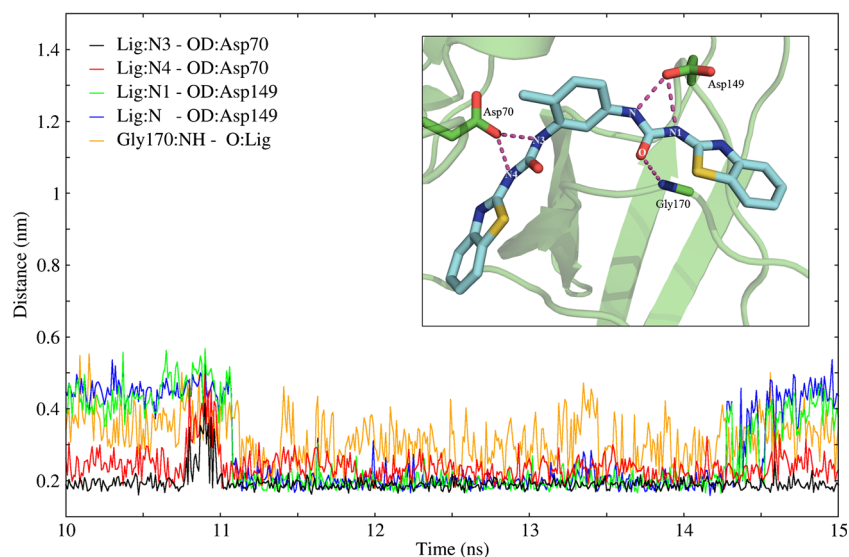

MD simulations of protease—ligand complexes

Stability of the complexes between the first five hit compounds ((NSC122819, NSC345647, NSC319990, NSC50650, and NSC5069) and YHV protease was evaluated by performing MD simulations (Fig. 4). The backbone RMSD values of each snapshot with respect to the initial structures obtained before energy minimizations were plotted over the course of 15 ns simulation time (Supplementary data Fig. S2). Although the RMSD values for the apo simulation tended to be more deviated from the initial structure, those for the complex simulations had reached equilibrium well after $t=$ 5 ns and remained fluctuated not more than ≈0.1 nm until the end of the simulations, indicating stability of the complex systems. Potential energies for interactions between the hit compounds and the enzyme were calculated as the sum of the electrostatic and van der Waals (vdW) energies over the last 5 ns of the simulations (Fig. 6). It is noted that vdW interactions were found to be stronger than electrostatic interactions in all cases. The compounds NSC345647 and NSC319990 contribute more interaction energies than the other three complexes. These results are consistent with the docking observations in that both compounds are the first two hits among the five. Numbers of hydrogen bonds between the five hit compounds and the enzyme were then investigated throughout the 15 ns in all complex simulations (Supplementary data Fig. S3). List of the interaction pairs and percentage of occupancy were shown in Table 3. The analyses showed that the compounds NSC345647 and NSC319990 exhibit the largest hydrogen bond contributions among the five complexes as expected (Table 3). In particular, the compound NSC319990 makes strong hydrogen bonds with several residues including the catalytic triad of the protease (Fig. 7). Among these, the interaction between the N3 and N4 atoms of the compound and the side chain of Asp70 exhibits stability with the highest

bond occupancy (95.8 % and 69.5 %, respectively). The Asp149 side chain (in pocket A) also makes strong hydrogen bonds with the N atoms of the compound (63.4 % and 61.6 %). Furthermore, the π-π interaction between the 6-membered aromatic ring of the compound and His30 was observed (data not shown). These strong hydrogen bonds observed in the NSC319990 complex are in agreement with the observations from molecular docking.

YHV inhibitory assay

The same first five hit compounds were tested for YHV inhibitory assay using bacterial expressed YHV. The relative inhibition results were shown in Fig. 8. Both NSC122819 and NSC345647 showed a strong inhibitory against YHV at our assay condition while NSC319990, NSC50650, and NSC5069 showed more than 50 % inhibitory. The best
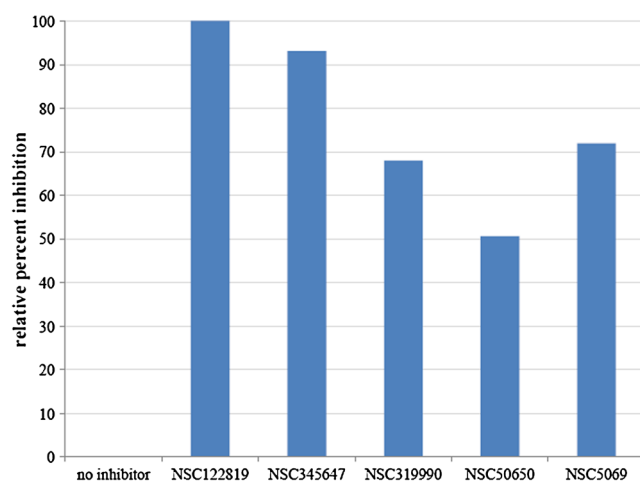
Fig. 8 Percentage of YHV inhibition of the first five hit NCI compounds against bacterial expressed YHV protease using fluoroscopic assay. The final concentration of compounds were 75 nM

docked-score compounds, NSC122819 also showed the strongest inhibition. Further experiments will be necessary to understand the inhibiting mechanism of these compounds.

Conclusions

In this study, the homology model of YHV protease was created from multiple template technique. The modeled structure was used in virtual screening studies against a set of compounds in the NCI database. This investigation proposed a set of promising small molecules to inhibit or decrease the activity of protein YHV protease from the in silico studies model. The inhibiting activity of the best five docked-score compounds were tested against YHV protease in vitro. All of them showed a good degree of inhibition. The best docked-score, NSC122819, also showed the best inhibition. This confirmed our successful homology and docking protocols to predict the potential inhibitors against YHV. These compounds can be further tested against yellow head virus in cells and farm levels in the future.

References

1. Jitrapakdee S, Unajak S, Sittidilokratna N, Hodgson RA, Cowley JA, Walker PJ, Panyim S, Boonsaeng V (2003) Identification and analysis of gp116 and gp64 structural glycoproteins of yellow head nidovirus of Penaeus monodon shrimp. J Gen Virol 84(Pt 4):863–873

2. Hertzig T, Scandella E, Schelle B, Ziebuhr J, Siddell SG, Ludewig B, Thiel V (2004) Rapid identification of coronavirus replicase inhibitors using a selectable replicon RNA. J Gen Virol 85(Pt 6):1717–1725

3. Sittidilokratna N, Dangtip S, Cowley JA, Walker PJ (2008) RNA transcription analysis and completion of the genome sequence of yellow head nidovirus. Virus Res 136(1–2):157–165

4. Ziebuhr J, Bayer S, Cowley JA, Gorbalenya AE (2003) The 3C-like proteinase of an invertebrate nidovirus links coronavirus and potyvirus homologs. J Virol 77(2):1415–1426

5. Cowley JA, Dimmock CM, Spann KM, Walker PJ (2000) Gill-associated virus of Penaeus monodon prawns: an invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. J Gen Virol 81(Pt 6):1473–1484

6. Ziebuhr J, Snijder EJ, Gorbalenya AE (2000) Virus-encoded proteinases and proteolytic processing in the Nidovirales. J Gen Virol 81(Pt 4):853–879

7. Anderson J, Schiffer C, Lee SK, Swanstrom R (2009) Viral protease inhibitors. Handb Exp Pharmacol 189:85–110

8. Natarajan S (2010) NS3 protease from flavivirus as a target for designing antiviral inhibitors against dengue virus. Genet Mol Biol 33(2):214–219

9. Yodmuang S, Tirasophon W, Roshorm Y, Chinnirunvong W, Panyim S (2006) YHV-protease dsRNA inhibits YHV replication in Penaeus monodon and prevents mortality. Biochem Biophys Res Commun 341(2):351–356

10. Tirasophon W, Yodmuang S, Chinnirunvong W, Plongthongkum N, Panyim S (2007) Therapeutic inhibition of yellow head virus multiplication in infected shrimps by YHV-protease dsRNA. Antivir Res 74(2):150–155

11. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22(22):4673–4680

12. Nicholas KB, Nicholas HBJ (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author. http://wwwpscedu/biomed/genedoc

13. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779–815

14. MacKerell AD, Bashford D, Bellott DRL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins†. J Phys Chem B 102(18):3586–3616

15. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) {PROCHECK}: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291

16. NCI/DTP. http://dtp.nci.nih.gov/branches/dscb/repo_open.html

17. Ligprep, v2. 2. Schrödinger, LLC, New York, NY (2008)

18. Huey R, Goodsell DS, Morris GM, Olson AJ (2004) Grid-based hydrogen bond potentials with improved directionality. Lett Drug Des Discov 1(2):178–183

19. Huey R, Morris GM, Olson AJ, Goodsell DS (2007) A semiempirical free energy force field with charge-based desolvation. J Comput Chem 28(6):1145–1152

20. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19(14):1639–1662

21. FRED program OpenEye Scientific Software, Inc., Santa Fe, NM, USA

22. Chen YF, Hsu KC, Lin SR, Wang WC, Huang YC, Yang JM (2010) SiMMap: a web server for inferring site-moiety map to recognize interaction preferences between protein pockets and compound moieties. Nucleic Acids Res 38(Web Server issue):W424–W430

23. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78(8):1950–1958

24. Sousa da Silva AW, Vranken WF (2012) ACPYPE—AnteChamber PYthon Parser interfacE. BMC Res Notes 5:367

25. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25(9):1157–1174

26. Jakalian A, Jack DB, Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. J Comput Chem 23(16):1623–1641

27. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935

28. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718

29. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4(3):435–447

30. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29(7):845–854

31. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81(8):3684–3690

32. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126(1), 014101

33. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 52(12): 7182–7190

34. Nose S, Klein M (1983) Constant pressure molecular dynamics for molecular systems. Mol Phys 50(5):1055–1076

35. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18(12):1463–1472

36. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J Chem Phys 98(12):10089–10092

37. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. J Chem Phys 103(19): 8577–8593

38. van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Feenstra KA, van Drunen R, Berendsen HJC (2010) Gromacs user manual version 4.5.4. www.gromacs.org

39. Turner PJ Grace. http://plasma-gate.weizmann.ac.il/Grace/

40. DeLano WL The PyMOL molecular graphics system, Version 1.3r1. http://pymol.org/educational/

41. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5(4):725–738

42. Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4(3):363–371

43. Jones DT (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 287(4): 797–815

44. Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei J, Kim D, Kellogg E, DiMaio F, Lange O, Kinch L, Sheffler W, Kim BH, Das R, Grishin NV, Baker D (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins 77(Suppl 9):89–99

45. Kurowski MA, Bujnicki JM (2003) GeneSilico protein structure prediction meta-server. Nucleic Acids Res 31(13):3305–3307

46. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones DT (2010) Protein annotation and modelling servers at University College London. Nucleic Acids Res 38(Web Server issue):W563–W568

47. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32(Web Server issue):W526–W531

48. Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK 3rd, Kapust RB, Li M, Wlodawer A, Waugh DS (2002) Structural basis for the substrate specificity of tobacco etch virus protease. J Biol Chem 277(52):50564–50572

49. Matthews DA, Dragovich PS, Webber SE, Fuhrman SA, Patick AK, Zalman LS, Hendrickson TF, Love RA, Prins TJ, Marakovits JT, Zhou R, Tikhe J, Ford CE, Meador JW, Ferre RA, Brown EL, Binford SL, Brothers MA, DeLisle DM, Worland ST (1999) Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. Proc Natl Acad Sci U S A 96(20): 11000–11007

50. Bergmann EM, Mosimann SC, Chernaia MM, Malcolm BA, James MN (1997) The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. J Virol 71(3):2436–2448

51. Finer-Moore JS, Kossiakoff AA, Hurley JH, Earnest T, Stroud RM (1992) Solvent structure in crystals of trypsin determined by X-ray and neutron diffraction. Proteins 12(3):203–222

52. Jing H, Xu Y, Carson M, Moore D, Macon KJ, Volanakis JE, Narayana SVL (2000) New structural motifs on the chymotrypsin fold and their potential roles in complement factor B. EMBO J 19(2): 164–173

53. Cavarelli J, Prévost G, Bourguet W, Moulinier L, Chevrier B, Delagoutte B, Bilwes A, Mourey L, Rifai S, Piémont Y, Moras D (1997) The structure of Staphylococcus aureus epidermolytic toxin A, an atypic serine protease, at 1.7 Å resolution. Structure (London, England: 1993) 5(6):813–824

54. Tsunasawa S, Masaki T, Hirose M, Soejima M, Sakiyama F (1989) The primary structure and structural characteristics of Achromobacter lyticus protease I, a lysine-specific serine protease. J Biol Chem 264(7):3832–3839

55. Schechter I, Berger A (1967) On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun 27(2): 157–162