เอกสารและสิ่งอ้างอิง

1. Assavalapsakul, W., Tirasophon, W., Panyim, S. Antiserum to the gp116 glycoprotein of yellow head virus neutralizes infectivity in primary lymphoid organ cells of Penaeus monodon (2005) Diseases of Aquatic Organisms, 63 (1), 85-88.
2. Beate S., Nadia N. and Regina S. Protein Detection with Aptamer Biosensors (2008) Sensors, 8, 4296-4307
3. Bellecave, P., Cazenave, C., Rumi, J., Staedel, C., Cosnefroy, O., Andreola, M.-L., Ventura, M., Tarrago-Litvak, L., Astier-Gin, T. Inhibition of hepatitis C virus (HCV) RNA polymerase by DNA aptamers: Mechanism of inhibition of in vitro RNA synthesis and effect on HCV-infected cells (2008) Antimicrobial Agents and Chemotherapy, 52 (6), 2097-2110.
4. Butz, K., Denk, C., Fitscher, B., Crnkovic-Mertens, I., Ullmann, A., Schröder, C.H., Hoppe-Seyler, F. Peptide aptamers targeting the hepatitis B virus core protein: A new class of molecules with antiviral activity (2001) Oncogene, 20 (45), 6579-6586.
5. Cheng, C., Dong, J., Yao, L., Chen, A., Jia, R., Huan, L., Guo, J., Shu, Y., Zhang, Z. Potent inhibition of human influenza H5N1 virus by oligonucleotides derived by SELEX (2008) Biochemical and Biophysical Research Communications, 366 (3), 670-674.
6. De Soultrait, V.R., Lozach, P.-Y., Altmeyer, R., Tarrago-Litvak, L., Litvak, S., Andréola, M.L. DNA aptamers derived from HIV-1 RNase H inhibitors are strong anti-integrase agents (2002) Journal of Molecular Biology, 324 (2), 195-203.
7. Dey, A. K., C. Griffiths, S. M. Lea, and W. James. Structural characterization of an anti-gp120 RNA aptamer that neutralizes R5 strains of HIV-1. (2005) RNA, 11, 873-884
8. Ellington AD and Szostak JW. In vitro selection of RNA molecules that bind specific ligands. (1990) *Nature,* 346, 818–822

9. 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 23:1623-1641

10. James, W. Aptamers in the virologists' toolkit (2007) Journal of General Virology, 88 (2), 351-364.

11. Jayasena SD. Aptamers: an emerging class of molecules that rival antibodies in diagnostics. (1999) *Clin Chem.*, 45, 1628–1650

12. Jitrapakdee S., Unajak S., Sittidilokratna N., Richard AJ Hodgson, Jeff A. Cowley, Peter J. Walker, Panyim S. and Boonseang V. Identification and analysis of gp116 and gp64 structural glycoproteins of yellow head nidovirus of *Penaeus monodon* shrimp. (2003) J. Gen. Virol, 84(4), 863-873.

13. Jochmans, D. Novel HIV-1 reverse transcriptase inhibitors (2008) Virus Research, 134 (1-2), 171-185.

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

15. Joshi, P.J., North, T.W., Prasad, V.R. Aptamers directed to HIV-1 reverse transcriptase display greater efficacy over small hairpin RNAs targeted to viral RNA in blocking HIV-1 replication (2005) Molecular Therapy, 11 (5), 677-686.

16. Kim CS, Kosuke Z, Nam YK, Kim SK, Kim KH Protection of shrimp (Penaeus chinensis) against white spot syndrome virus (WSSV) challenge by double-stranded RNA. (2007) Fish Shellfish Immunol, 23, 242–246

17. Khati, M., M. Schuman, J. Ibrahim, Q. Sattentau, S. Gordon, and W. James. Neutralization of infectivity of diverse R5 clinical isolates of human immunodeficiency virus type 1 by gp120-binding 2'F-RNA aptamers. (2003), J. Virol, 77, 12692-12698

18. Koizumi M and Breaker RR. Molecular recognition of cAMP by an RNA aptamer. *Biochemistry* (2000), 39, 8983–8992.

19. Lee, J.F., Stovall, G.M., Ellington, A.D. Aptamer therapeutics advance (2006) Current Opinion in Chemical Biology, 10 (3), 282-289.

20. Lee, S., Kim, Y.S., Jo, M., Jin, M., Lee, D.-k., Kim, S. Chip-based detection of hepatitis C virus using RNA aptamers that specifically bind to HCV core antigen (2007) Biochemical and Biophysical Research Communications, 358 (1), 47-52.

21. Li H and Rothberg L. Colorimetric detection of DNA sequences based on electrostatic interactions with unmodified gold nanoparticles (2004) Proc. Natl. Acad. Sci. USA, 101, 14036–14039

22. Liu YingChun, ZHANG Yan, Ye GuoZhu, Yang ZhenJun, Zhang LiangRen, Zhang LiHe. In vitro selection of G-rich RNA aptamers that target HIV-1 integrase (2008) Science in China Series B: Chemistry, 401-413

23. Lindorff-Larsen K, Piana S, Palmo K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950-1958

24. Lochrie MA, Waugh S, Pratt Jr DG, Clever J,Parslow TG, Polisky B. In vitro selection of RNAs that bind to the human immunodeficiency virus type-1 gag polyprotein 1997 Nuc.. Acids Res. 25 2902-2910Mee Young Kim and Sunjoo Jeong. RNA Aptamers that Bind the Nucleocapsid Protein Contain Pseudoknots. (2003) Mol. Cells, Vol. 16(3), 413-417

25. Métifiot, M., Leon, O., Tarrago-Litvak, L., Litvak, S., Andréola, M.-L. Targeting HIV-1 integrase with aptamers selected against the purified RNase H domain of HIV-1 RT (2005) Biochimie, 87 (9-10), 911-919.

26. Nicholas OF, Theodore MT, and Jeffrey BT. Protein Detection via Direct Enzymatic Amplification of Short DNA Aptamers Anal Biochem (2008) 373, 121–128

27. Nickens, D.G., Patterson, J.T., Burke, D.H. Inhibition of HIV-1 reverse transcriptase by RNA aptamers in *Escherichia coli* (2003) RNA, 9 (9), 1029-1033

28. Osborne, S.E., Matsumura, I., Ellington, A.D. Aptamers as therapeutic and diagnostic reagents: Problems and prospects (1997) Current Opinion in Chemical Biology, 1 (1), 5-9.

29. Porschewski P, Grattinger MA, Klenzke K and Erpenbach A, Blind MR, Schafer F. Using aptamers as capture reagents in bead-based assay systems for diagnostics and hit identification. (2006) *J Biomol Screen,* 11, 773–781.

30. Shekhar MS. and Lu Y. Application of Nucleic-acid-based Therapeutics for Viral Infections in Shrimp Aquaculture (2009) Mar Biotechnol, 11, 1–9

31. Robalino J, Bartlett T, Shepard E, Prior S, Jaramillo G, Scura E. Double-stranded RNA induces sequence-specific antiviral silencing in addition to nonspecific immunity in a marine shrimp: Convergence of RNA interference and innate immunity in the invertebrate antiviral response? (2005) J Virol, 79, 13561–13571

32. Sittidilokratna, N., Dangtip, S., Cowley, J.A., Walker, P.J. RNA transcription analysis and completion of the genome sequence of yellow head nidovirus (2008) Virus Research, 136 (1-2), 157-165.

33. Stoltenburg, R., Reinemann, C., Strehlitz, B. SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands (2007) Biomolecular Engineering, 24 (4), 381-403.

34. Tirasophon, W., Roshorm, Y., Panyim, S. Silencing of yellow head virus replication in penaeid shrimp cells by dsRNA (2005) Biochemical and Biophysical Research Communications, 334 (1), 102-107.

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

36. Tok JB, Cho J and Rando RR. RNA aptamers that specifically bind to a 16S ribosomal RNA decoding region construct. (2000) *Nucleic Acids Res.* 28, 2902–2910

37. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. (1990) *Science*, 249, 505–510

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

39. Westenberg M, Heinhuis B, Zuidema D, Vlak JM., siRNA injection induces sequence independent protection in Penaeus monodon against white spot syndrome virus (WSSV). (2005) Virus Res, 114, 133–139

40. Wijegoonawardane PKM, Sittidilokratna N, Petchampai N, Cowley JA, Gudkovs N, Walker PJ. Homologous genetic recombination in the yellow head complex of nidoviruses

infecting Penaeus monodon shrimp. (2009) Virology, 390, 79–88.

41. Wu Y, Lu L, Yang L, Weng S, Chan S, He J. Inhibition of white spot syndrome virus in Litopenaeus vannamei shrimp by sequence-specific siRNA. (2007) Aquaculture, 271, 21–30

42. Xu H, Mao X, Zeng Q, Wang S, Kawde A and Liu G. Aptamer-Functionalized Gold Nanoparticles as Probes in a Dry-Reagent Strip Biosensor for Protein Analysis (2009) *Anal. Chem.* , *81*, 669–675

43. Xu J, Han F, Zhang X. Silencing shrimp white spot syndrome virus (WSSV) genes by siRNA. (2007) Antiviral Res, 73, 126–131

44. Yu Young Jeong, Seon Hee Kim, Soo In Jang & Ji Chang You. Examination of specific binding activity of aptamer RNAs to the HIV-NC by using a cell-based in vivo assay for protein-RNA interaction. (2008) BMB reports, 511-515

45. Zhang J, Wang L, Pan D, Song S, Boey FYC, Zhang H and Fan C. Visual Cocaine Detection with Gold Nanoparticles and Rationally Engineered Aptamer Structures (2008) Small 2008, 4(8), 1196–1200

46. Zhou, J., Swiderski, P., Li, H., Zhang, J., Neff, C.P., Akkina, R., Rossi, J.J. Selection, characterization and application of new RNA HIV gp120 aptamers for facile delivery of Dicer substrate siRNAs into HIV infected cells (2009) Nucleic Acids Research, 37 (9), 3094-3109.

47. Ziebuhr J., Bayer S., Cowley JA., Gorbalenya AE., The 3C-Like Proteinase of an Invertebrate Nidovirus Links Coronavirus and Potyvirus Homologs (2003) J of Virol, 77(2), 1415-1426

กิจกรรมอื่นๆที่เกี่ยวข้อง ได้แก่

(1) การไปเสนอผลงาน

1. การประชุม JSPS-NRCT International Symposium 2010 "Productivity techniques and effective utilization of aquatic animal resources to the new century-VI" 26-27 February, 2010 Shinagawa Campus, Tokyo University of Marine Science and Technology, Japan

2. การประชุม JSPS-NRCT Asian Core Program Symposium 2011 "Development of new bio-technology for aquaculture and risk management of aquaculture products-I", Friday 17th June 2011, Room 303, Boon Intarumphun Building, Faculty of Fisheries, Kasetsart University, Bangkok, Thailand

(2) การเชื่อมโยงทางวิชาการกับนักวิชาการอื่นๆทั้งในและต่างประเทศ

1. มีความร่วมมือกับ Prof Takashi Aoki , Prof Ikuo Hirono , และ Assoc Prof Hedeihiro Kondo Laboratory of Fish and Shrimp Genomes, Tokyo University of Marine Science and Technology, Tokyo, Japan ในด้านถ่ายทอดเทคโนโลยีการผลิต aptamer

2. มีความร่วมมือกับ ผศ.ดร. เกียรติทวี ชูวงศ์โกมล ภาควิชาชีวเคมี มหาวิทยาลัยเกษตรศาสตร์ ด้านการจำลองโครงสร้างสามมิติและการวิเคราะห์หาสารยับยั้งการทำงานของเอนไซม์ protease จากไวรัสหัวเหลือง

Journal of Molecular Modeling

ology modeling and Virtual Screening for Antagonists of Protease from Yellow Head Virus

--Manuscript Draft--

ipt Number:	JMMO-D-13-00424R1
:	Homology modeling and Virtual Screening for Antagonists of Protease from Yellow Head Virus
ype:	Original paper
ls:	homology modeling; yellow head virus protease; virtual screening; molecular docking; molecular dynamics simulations; in vitro inhibiting assay
onding Author:	Kiattawee Choowongkomon, Ph.D. Kasetsart University Bangkok, THAILAND
onding Author Secondary ion:	
onding Author's Institution:	Kasetsart University
onding Author's Secondary n:	
hor:	Sasimanas Unajak, Ph.D.
hor Secondary Information:	
Authors:	Sasimanas Unajak, Ph.D.
	Orathai Sawatdichaikul, Ph.D.
	Napat Songtawee, Ph.D.
	Siriluk Rattanabunyong, M.S.
	Anchalee Tassnakajon, Ph.D.
	Nontawith Areechon, Ph.D.
	Ikuo Hirono, Ph.D.
	Hidehiro Kondo, Ph.D.
	Pongsak Khunrae, Ph.D.
	Triwit Rattanarojpong, Ph.D.
	Kiattawee Choowongkomon, Ph.D.
Authors Secondary Information:	
	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 has not much effectively protected the mass mortality in shrimp. New approaches for the antiviral drug development for such a virus have been focused in inhibition of several potent viral enzymes, and thus the YHV protease is one of the target for develop antiviral drug 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 the target for virtual screening - molecular docking against approximately 1,364 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 inhibiting among them while others showed more than 50% inhibiting in our assay condition. These compounds could be potential inhibitors for curing YHV in the shrimp farming.

ıe to Reviewers:	Both molecular dynamics simulation and in vitro inhibiting assay were added in the manuscript.

Homology modeling and Virtual Screening for Antagonists of Protease from Yellow Head Virus.

Sasimanas Unajak[1], Orathai Sawatdichaikul[2], Napat Songtawee[3], Siriluk Rattanabunyong[1], Anchalee Tassnakajon[4], Nontawith Areechon[5], Ikuo Hirono[6], Hidehiro Kondo[6], Pongsak Khunrae[7], Triwit Rattanarojpong[7] and Kiattawee Choowongkomon[1, 8]*

[1]Department of Biochemistry, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan Rd, Chatuchak, Bangkok 10900, Thailand

[2]Institute of Food Research and Product Development, Kasetsart University, 50 Ngam Wong Wan Rd, Chatuchak, Bangkok 10900, Thailand

[3]Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok, Thailand

[4]Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10300, Thailand

[5]Department of Aquaculture, Faculty of Fisheries, Kasetsart University, 50 Ngam Wong Wan Rd, Chatuchak, Bangkok 10900, Thailand

[6]Laboratory of Genome Science, Graduate School of Tokyo University of Marine Science and Technology, Konan 4-5-7, Minato-ku, Tokyo 108-8477, Japan

[7]Microbiology Department, King Mongkut's University of Technology Thonburi, Bangmod, Toongkru, Bangkok 10140, Thailand

[8]Center for Advanced Studies in Tropical Natural Resources, National Research University-Kasetsart University, Thailand (CASTNAR, NRU-KU, Thailand)

* To whom correspondence should be addressed. Phone: +66-86-562-5555 ext 2052; Fax: +66-2-5614627

Author e-mail address: fsciktc@ku.ac.th (K. 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 has not much effectively protected the mass mortality in shrimp. New approaches for the antiviral drug development for such a virus have been focused in inhibition of several potent viral enzymes, and thus the YHV protease is one of the target for develop antiviral drug 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 the target for virtual screening – molecular docking against approximately 1,364 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 inhibiting among them while others showed more than 50% inhibiting in our assay condition. These compounds could be potential inhibitors for curing YHV in the shrimp farming.

Keywords: homology modeling, yellow head virus protease, virtual screening, molecular docking, molecular dynamics simulations, *in vitro* inhibiting assay

Introduction

Yellow head virus has been emerging for past two decades in Thailand. Virulence of the virus is known causing devastation in shrimp farming 2-3 days after onset of the disease. The infected shrimp exhibits yellow coloration on cephalothorax area and pale color throughout the body. Because of mass mortality of shrimp farming by Yellow head disease (YHD) during 1991-1995 led the enormous economically lost to Thailand, the promising prevention method of the disease have to be considered. The retardation or inhibition of YHV propagation is one of the prevention method for develop antiviral agent nowadays. Yellow head virus (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 proteins component of the virus composed of 3 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 Gill-associated virus (GAV) that is less virulent (or provides less virulence). A recent study revealed that they 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 are all involved in the replication complexes, and it also contains the metal-ion binding domain (MIB) [4].Chymotrypsin-like proteinase (3CLP) was firstly 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 with 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 successive in inhibition of YHV propagation was only accomplished by YHV-protease dsRNA [9,10]. Therefore, this viral protease should be a potential drug target for future therapeutic approach, likewise the development of viral protease inhibitors (PI) for HIV-1, HCV, and SARS, illustrated 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 determination the three-dimensional structure thus allows for an 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 to the successful therapeutic approach for HIV-1, HCV and SARS, the YHV-3CLP protease should be the best target for potent inhibition of 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 to against the YHV protease model was performed. The program Autodock4 is used as the docking tool in this study with several reasons. First, AutoDock is the 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 of the 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 DOPE (discrete optimized protein energy) 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 Å convergence). The stereochemical qualities of the energy-minimized model were assessed by the PROCHECK v3.5 [15] via submitting the coordinates to the JCSG structure validation server (www.jcsg.org).

Ligand preparation

The 1,364 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]. According to the ligands in this diversity set were well collected from the almost 140,000 compounds vailable on plates with the criteria of DTP, such as hydrogen bond acceptor, hydrogen bond donor, positive harge, aromatic, hydrophobic, acid, base and pharmacophores, resulted in the selection of 1,364 compounds. 'hereby, 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 ond order and the chiral center check (stereo-chemistry) using LigPrep 2.2 with default setting [17]. The output

1,523 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, 19, 20].

Docking procedure

The minimized structure of YHV-protease model was obtained from 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 present in Fig2c). 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 was performed 50 trial runs using Lamarckian genetic algorithm search. The number of 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*

n the same time, all of these results were rescored using FRED program [21]. The anchor binding mode of igands 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 olecular dynamics (MD) simulations. The AMBER99SB-ILDN force field [23] was used to simulate the rotein structures and the ionization state of amino acid residues was set according to the standard protocol. To imic catalytic conditions of cysteine proteases, the side chain of His30 was protonated at both δ- and ε- trogen atoms (HIP form), while the side chain of Cys152 was deprotonated (minus cysteine; CYM form). The - and C-terminal ends of amino acid chains were capped with acetyl (ACE) and methyl amino (NME) groups, spectively. The GROMACS topology files for each ligand were generated using the ACPYPE script [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/nm on any atom or until additional steps result in a potential energy change of less than 1 kJ/mol to reduce undesirable atomic contacts. At first, positioned restraints with a force constant of 1000 kJ/mol/nm 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. Afterwards, 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/nm 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-fs 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 $\leq 30°$. 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 75nM 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 to % activity.

Results and discussions

Structural model of YHV protease

We have 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 between target and template sequences is one of the key step in homology modeling, and thus it should be cautiously done in case that the target contains 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 allowing the development of our own method used for the target-template sequence alignment as described in *Materials and Methods* section. The model templates used 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 (Figure 1 and Table 1). The selected templates are the 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 are in the most favored and additional allowed regions, indicating that a good quality model could be obtained (Supplementary Data Fig. S1). This final model adopts the two-domain antiparallel β-barrel fold-like structure resembling the templates (Figure 2a and 2b) with less than 3 Å of the C_α-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) are highly conserved among the four viral C-type cysteine proteases; yellow head, tobacco etch, rhino- and hepatitis A viruses (Figure 2c). These also indicated high accuracy of the modeled YHV protease and it is reasonable to use this model for subsequent docking experiments.

From the structural analysis of 3c-proteases in several species reveal that there are three conserved residues which might role as the catalytic triad, His, Asp or Glu, and Cys. The neighbor pocket of this site represent the region from S_2 to S'_2 positions for serve with P_2 to P'_2 positions of peptide substrate. This neighbor pocket is named "Pocket A" which composes of Lys148, Asp149, Gly170, Ala172 and Ser173. The residues of S1 pocket are generally Thr147 and His167.

Molecular docking with NCI database

In order to identify the small molecules specifically bind to YHV protease, a docking study analysis was performed against small molecules from NCI diversity dataset. All of 1,364 compounds from NCI diversity I 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 the 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 (Figure 3). As shown in the SiMMap moieties, there are four spheres which represent two types of interactions. *H moiety* refers to the hydrogen bonding, while *V moiety* refers to van der Waals force. *Anchor H1* illustrates 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 as 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 role to interact with ether, nitro, 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 followed the interaction modes.

If classified all 43 docked compounds from the interaction modes, they are divided into three groups. Considering at group one, there are 32 compounds which are able to form hydrogen bond to residues in the catalytic cleft and/or the pocket A (Figure 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 () which its structure mimic peptide bond. In group two composed of only two compounds which presenting only cation-π interaction. Compounds in group three consist of various types of core structure. 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 as present in the known; leupeptins, Saquinavir, Nelfinavir, Ritonavir, Indianavir, Amprenavir, Lopinavir, Chymostatins, Antipain, lastinal, and β-MAPI. These features were also found in our eleven candidate compounds from 32 compounds in group one which contains aromatic moiety, carboximide, hydroxy, carbonyl, ether, and ester groups. All even 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 presentative 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 there are

three common types of binding between these two compounds as presented in Figure 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 negatively dipole and act 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 *VI anchor*, and the fused rings of both compound have been monitored. These observations suggested that the screened compounds might posses the protease inhibitor properties as the known one, Nelfinavir. These set of ligands not only possessed strong hydrogen bonds with the surrounded residues in the binding pocket but also be reported as the potent inhibitory agents targeting several proteins including the malarial parasite plastid (Table 2).

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 Figure 4). The backbone RMSD values of each snapshot with respect to the initial structures obtained before nergy minimizations were plotted over the course of 15 ns simulation time (Supplementary Data Fig. S2). lthough the RMSD values for the apo simulation tended to be more deviated from the initial structure, those or the complex simulations had reached equilibrium well after t = 5 ns and remained fluctuated not more than ≈ .1 nm until the end of the simulations, indicating stability of the complex systems. Potential energies for nteractions between the hit compounds and the enzyme were calculated as the sum of the electrostatic and van er Waals (vdW) energies over the last 5 ns of the simulations (Figure 6). It would be noted that vdW teractions were found to be stronger than electrostatic interactions in all cases. The compounds NSC345647 nd NSC319990 contribute more interaction energies than the other three complexes. These results are nsistent with the docking observations in that both compounds are the first two hits among the five. Numbers hydrogen bonds between the five hit compounds and the enzyme were then investigated throughout the 15 ns all complex simulations (Supplementary Data Fig. S3). List of the interaction pairs and percentage of cupancy were shown in Table 3. The analyses showed that the compounds NSC345647 and NSC319990 hibit the largest hydrogen bond contributions among the five complexes as expected (Table 3). In particular, compound NSC319990 makes the strong hydrogen bonds with several residues including the catalytic triad

of the protease (Figure 7). Among these, the interaction between the N3 and N4 atoms of the compound and the side chain of Asp70 exhibits stable with the highest bond occupancy (95.8% and 69.5%, respectively). The Asp149 side chain (in Pocket A) also makes the 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 figure 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 docked-score compounds, NSC122819 also showed the strongest inhibition. The further experiments will be necessary for 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 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.

Acknowledgements

This work was supported by Thailand Research Fund (MRG4980039), Kasetsart University Research and Development Institute, and NRCT-JSPS Asian Core Program in Fisheries Science for SU, Faculty of Science, Kasetsart University and the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission for KC. Furthermore, we would like to thank Dr. Pongsak Khunrae for YHV protein, National Nanotechnology Center (NANOTEC), the National Science

and Technology Development Agency (NSTDA), Thailand for providing the Discovery Studio 2.5 programs. Finally, we would like to acknowledge the OpenEye Scientific Software for the academic license.

Figure legends

Figure 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.

Figure 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, respectively. (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 C_α atom of the YHV protease, 1LVM, 1CQQ and 1HAV is shown in green, orange, pink and yellow, respectively.

Figure 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).

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

Figure 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.

Figure 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.

Figure 7 Representation for the interactions between the 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.

Figure 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 75nM.

References

1. Jitrapakdee S, Unajak S, Sittidilokratna N, Hodgson RAJ, 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 (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 (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 (6):1473-1484

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

7. Anderson J, Schiffer C, Lee S-K, Swanstrom R (2009) Viral Protease Inhibitors.

Antiviral Strategies. In: Kräusslich H-G, Bartenschlager R (eds), vol 189. Handbook of Experimental Pharmacology. Springer Berlin Heidelberg, pp 85-110.

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

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 Bioph Res Co 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. Antiviral Research 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, Jr. NHB, Deerfield DWI (1997) GeneDoc: Analysis and Visualization of Genetic Variation. EMBNET.news, p 4

13. Šali 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, Dunbrack RL, 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 (2):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. Discovery 1:178-183

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

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 19:1639-1662

21. Fred. FRED program OpenEye Scientific Software, Inc., Santa Fe, NM, USA., www.eyesopen.com

22. Chen YF, Hsu KC, Lin SR, Wang WC, Huang YC, Yang JM (2010) SiMMap: a web server for nferring site-moiety map to recognize interaction preferences between protein pockets and compound moieties. Nucleic Acids Res 38 Suppl:W424-430

23. Lindorff-Larsen K, Piana S, Palmo K, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 78:1950-1958

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 25:1157-1174

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 23:1623-1641

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

28. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701-1718

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 4:435-447

30. Pronk S, Pall S, Schulz R, et al. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845-854

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

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

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

34. Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. Mol Phys 50:1055-1076

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

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 98:10089-10092

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

38. van der Spoel D, Lindahl E, Hess B, et al. (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 Protocols 5 (4):725-738

42. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protocols 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 B-H, Das R, Grishin NV, Baker D (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. Proteins: Structure, Function, and Bioinformatics 77 (S9):89-99.

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

46. Buchan DWA, Ward SM, Lobley AE, Nugent TCO, Bryson K, Jones DT (2010) Protein annotation and modelling servers at University College London. Nucleic Acids Res 38 (suppl 2):W563-W568.

47. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32 (suppl 2):W526-W531.

48. Phan J, Zdanov A, Evdokimov AG, Tropea JE, Peters HK, Kapust RB, Li M, Wlodawer A, Waugh S (2002) Structural Basis for the Substrate Specificity of Tobacco Etch Virus Protease. J Biol Chem 277 (2):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. P Natl Acad Sci USA 96 (20):11000-11007.

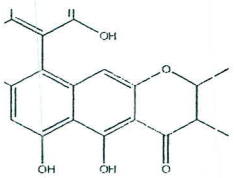
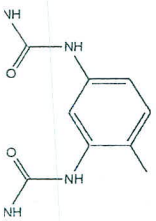
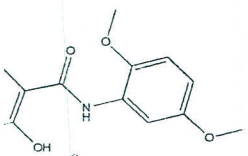
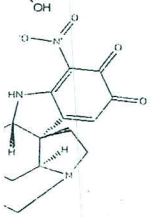
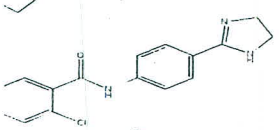
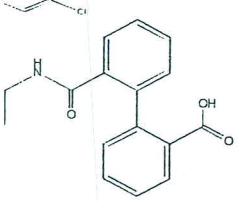
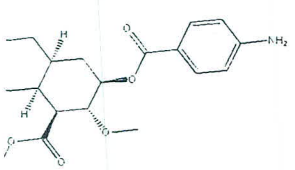
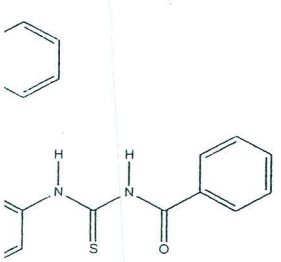
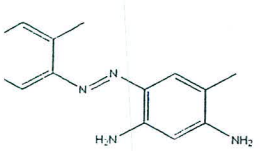
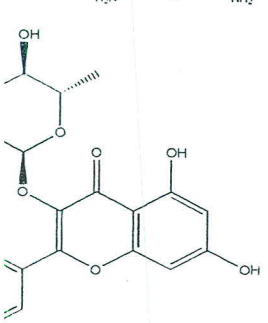
50. Bergmann EM, Mosimann SC, Chernaia MM, Malcolm BA, James MN (1997) The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. J Virol 71 (3):2436-2448

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: Structure, Function, and Bioinformatics 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 5 (6):813-824

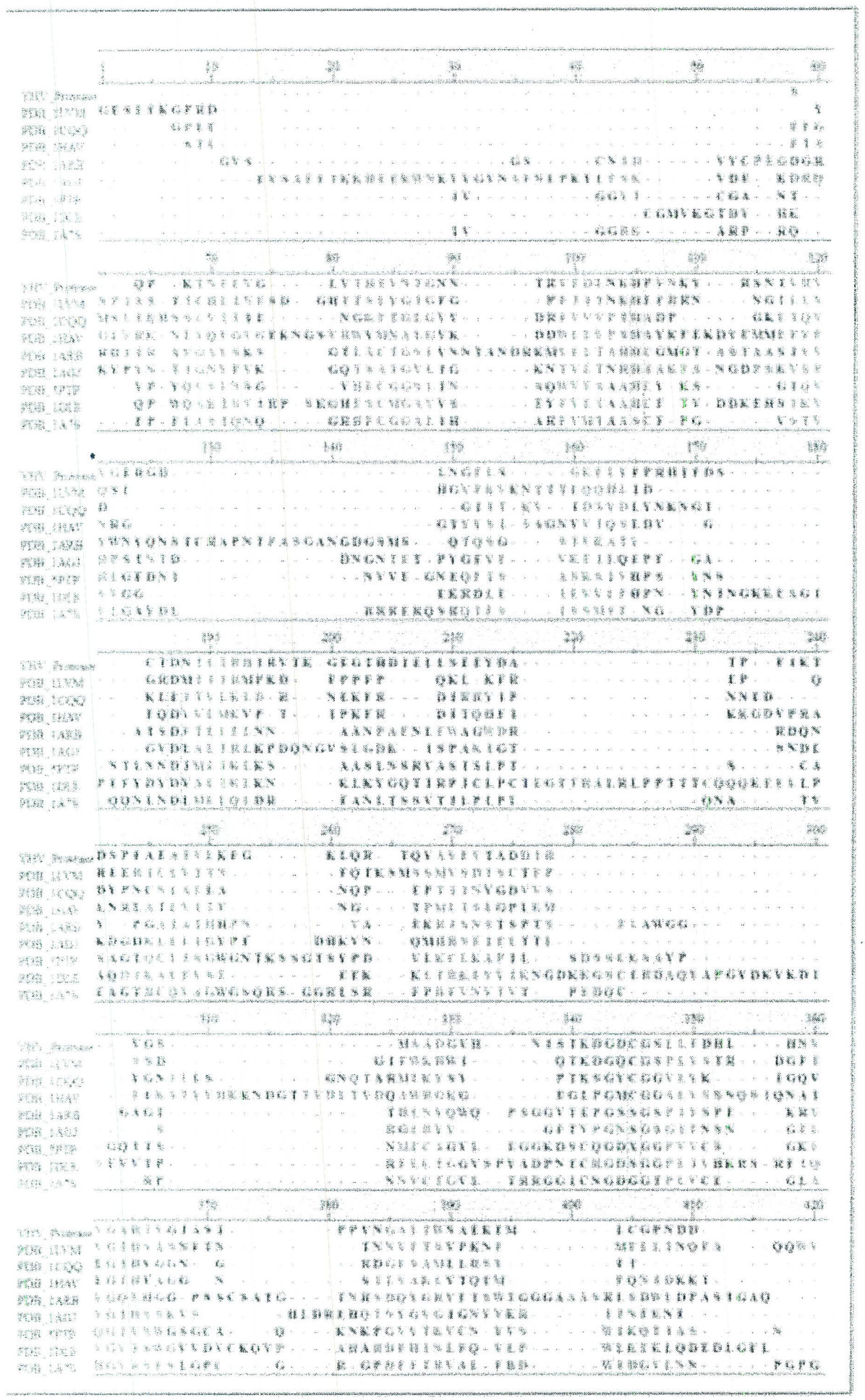
54. Tsunasawa S, Masaki T, Hirose M, Soejima M, Sakiyama F (1989) The primary structure and structural characteristics of Achromobacter lyticus protease I, a lysine-specific serine protease. J Biol Chem 264 (7): 3832-3839

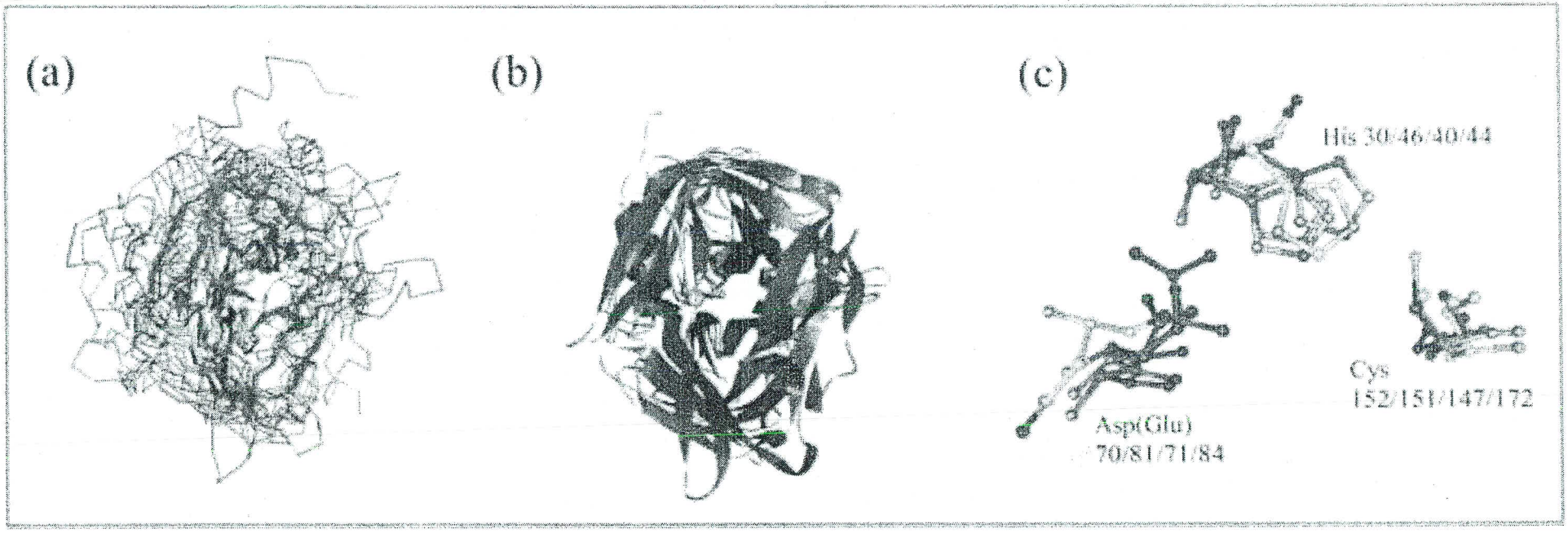
	53277	162783	• Ser/Thr protein phosphatase 2A (• HIV1-integrase
	330740	1578626	• Histone-lysine N-methyltransfera • DNA polymerase beta • Luciferin 4-monooxygenase
	67239	156747	• Lysine-specific demethylase 4D-l • Aldehyde dehydrogenase 1A1 • Histone-lysine N-methyltransfera • Sphingomyelin phosphodiesteras
	221125		
	249307	1578626	• Histone-lysine N-methyltransfera • DNA polymerase beta • Luciferin 4-monooxygenase
	280492	1172912	• MAP kinase ERK2 • Dengue virus type 2 NS3 protein • etc.
	232432	1326032	• Beta-lactamase • DNA prolymerase iota • etc.
	3361436	1426572	• Aldehyde dehydrogenase 1A1 • Putative fructose-1,6-bisphosphat • Peptidyl-prolyl cis-trans isomerise interacting 1 • etc.
	24674	1602553	• Pyruvate kinase • DNA polymerase kappa & iota • etc.
	5281673	454576	

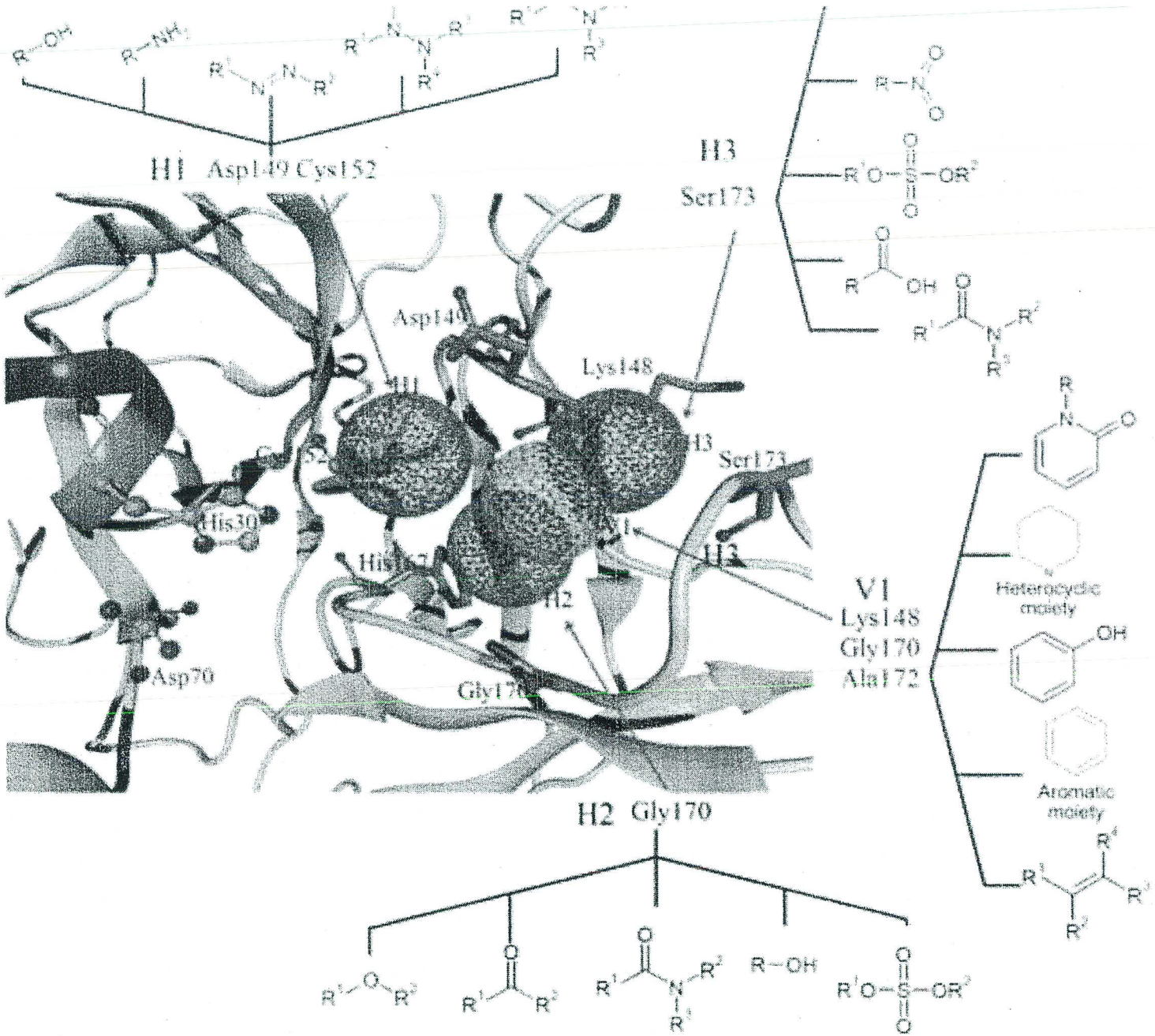
nload table: Table 3.xls

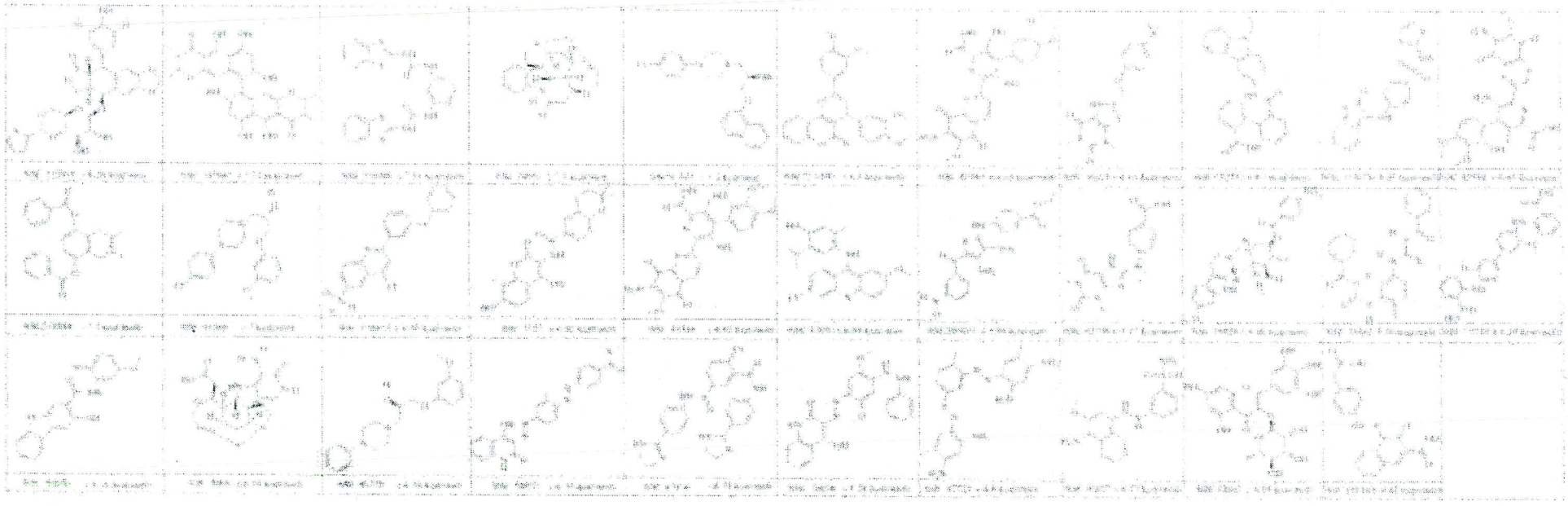
3 List of hydrogen bonds of the YHV protease - hit compound interactions ted during the last 5 ns from each complex simulation.

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

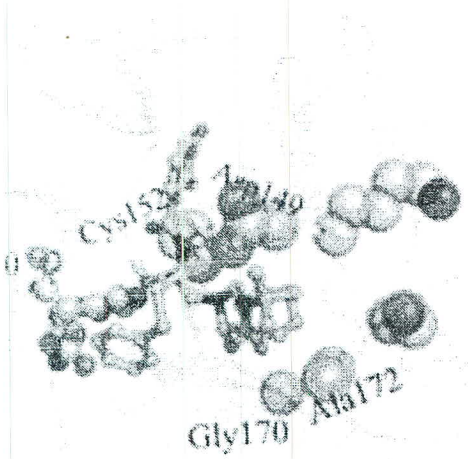


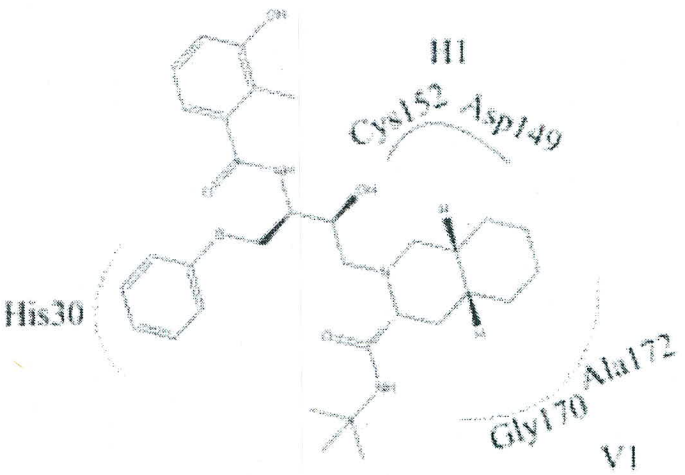

(a)
(b)
(c)
His 30/46/40/44
Cys
152/151/147/172
Asp(Glu)
70/81/71/84

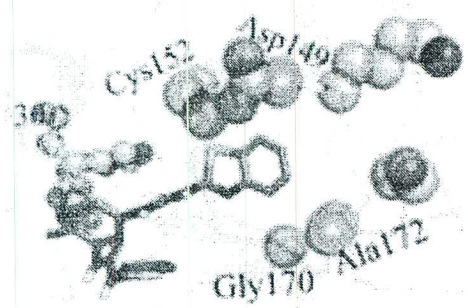

H1 Asp149 Cys152
H3 Ser173
Asp140
Lys148
His30
Asp70
Gly170
H2 Gly170
V1 Lys148 Gly170 Ala172
Heterocyclic moiety
Aromatic moiety
OH

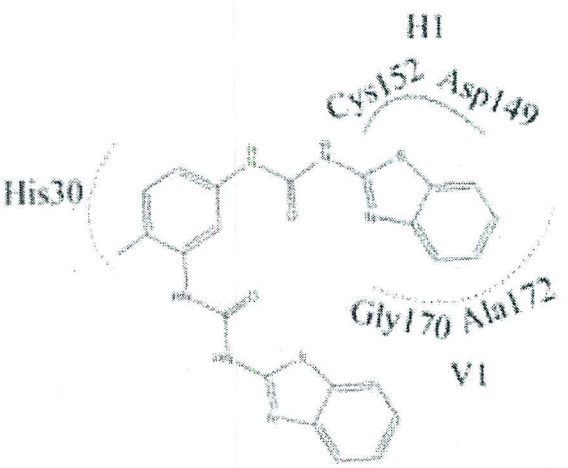


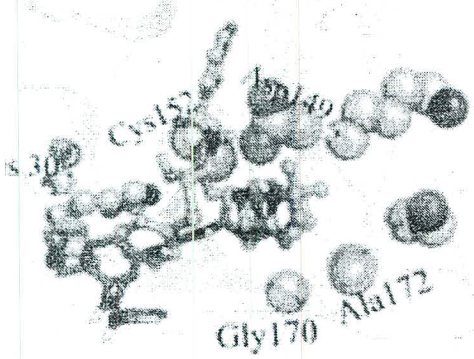
oad high resolution image

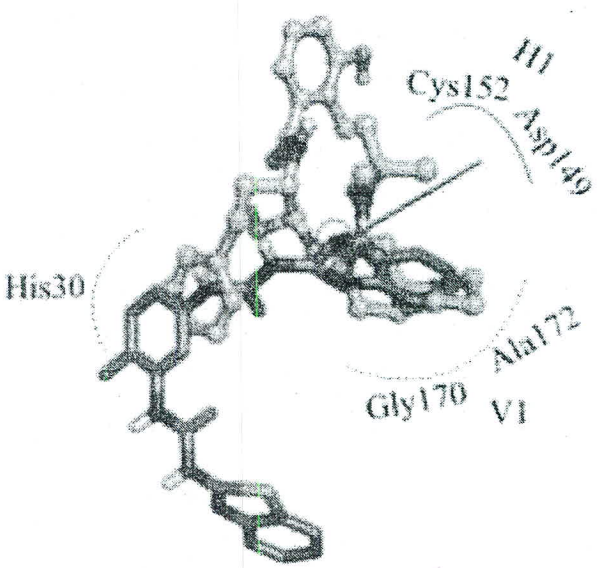


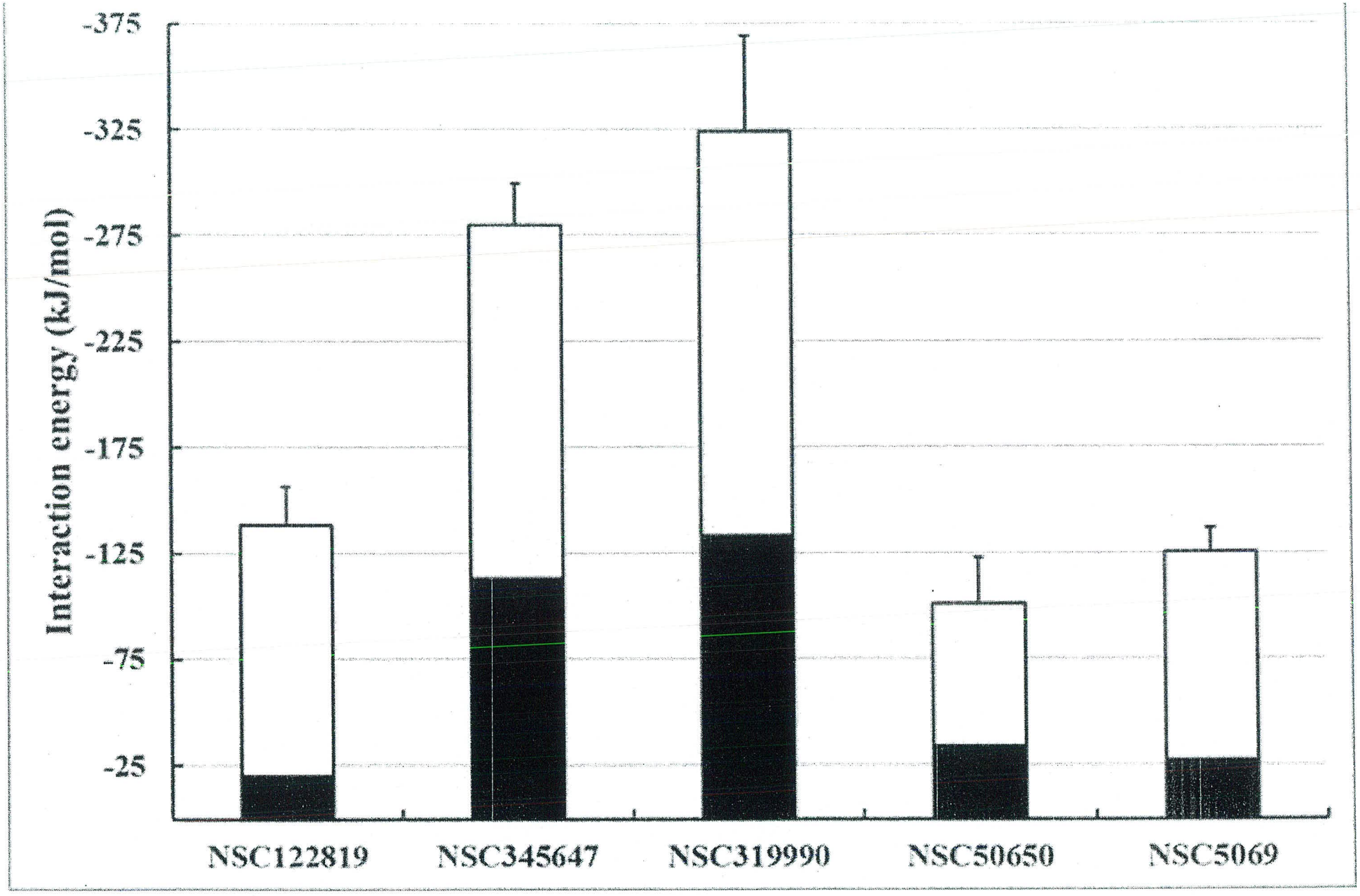


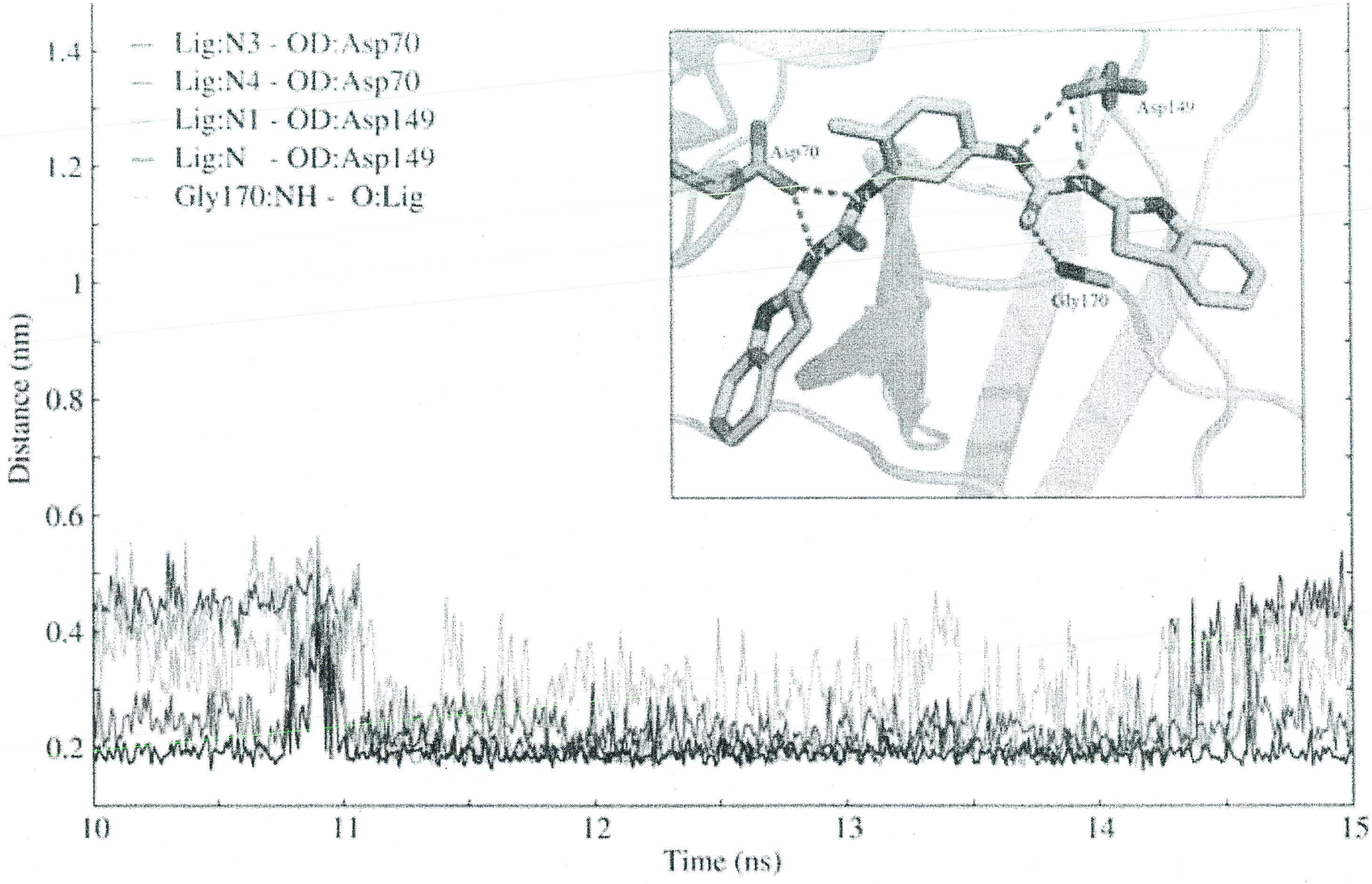


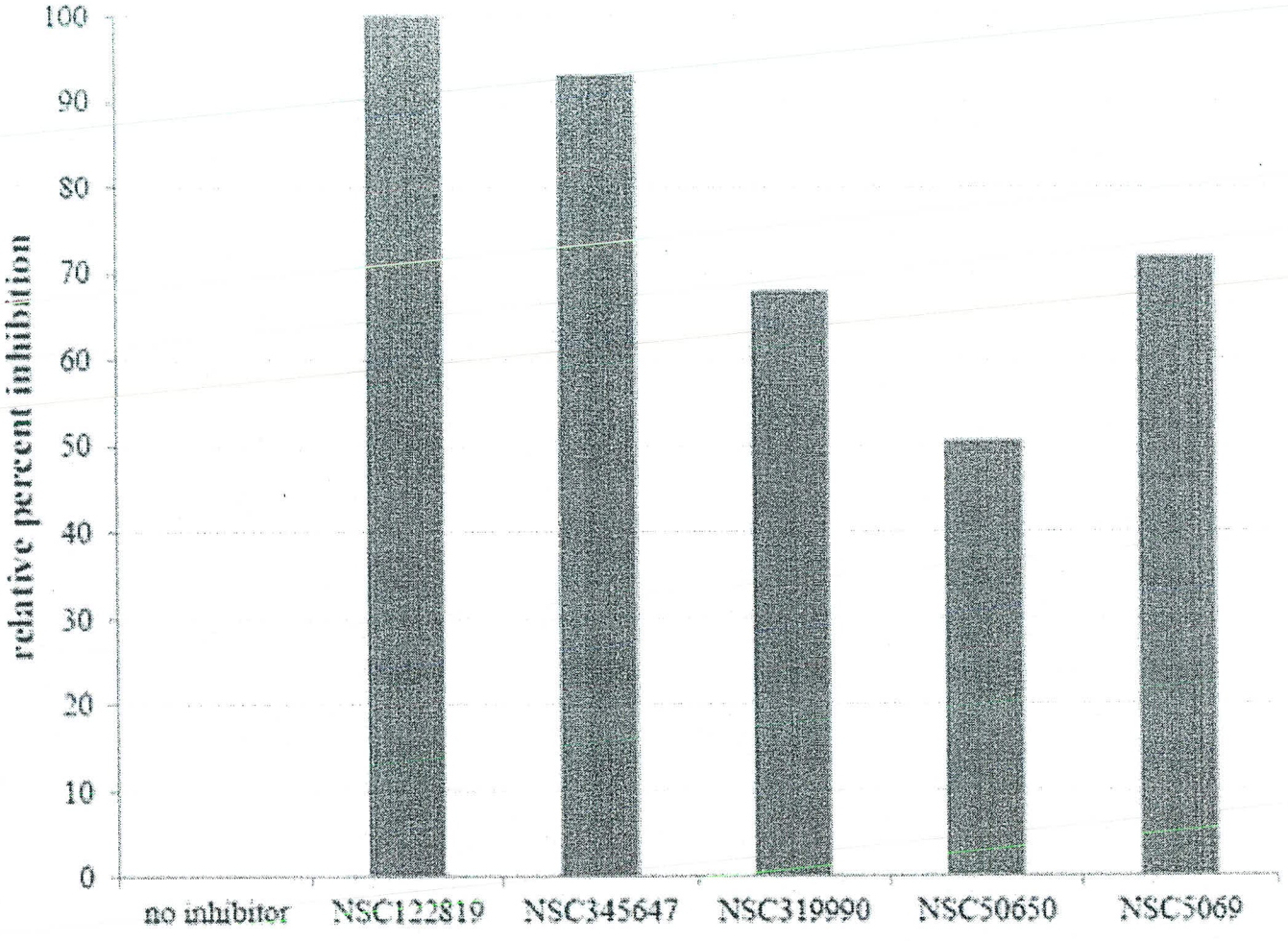







-375
-325
-275
-225
-175
-125
-75
-25
Interaction energy (kJ/mol)
NSC122819
NSC345647
NSC319990
NSC50650
NSC5069


Lig:N3 - OD:Asp70
Lig:N4 - OD:Asp70
Lig:N1 - OD:Asp149
Lig:N - OD:Asp149
Gly170:NH - O:Lig
Asp70
Asp149
Gly170
Distance (nm)
Time (ns)
1.4
1.2
1
0.8
0.6
0.4
0.2
10
11
12
13
14
15


relative percent inhibition
100
90
80
70
60
50
40
30
20
10
0
no inhibitor
NSC122819
NSC345647
NSC319990
NSC50650
NSC5069

upplementary Data

lomology modeling and Virtual Screening for ntagonists of Protease from Yellow Head Virus

simanas Unajak[1], Orathai Sawatdichaikul[2], Napat Songtawee[3], Siriluk Rattanabunyong[1], Anchalee ssnakajon[4], Nontawith Areechon[5], Ikuo Hirono[6], Hidehiro Kondo[6], Pongsak Khunrae[7], Triwit attanarojpong[7] and Kiattawee Choowongkomon[1, 8]*

Department of Biochemistry, Faculty of Science, Kasetsart University, 50 Ngam Wong Wan Rd, Chatuchak, angkok 10900, Thailand

nstitute of Food Research and Product Development, Kasetsart University, 50 Ngam Wong Wan Rd, hatuchak, Bangkok 10900, Thailand

Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, angkok, Thailand

Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10300, Thailand

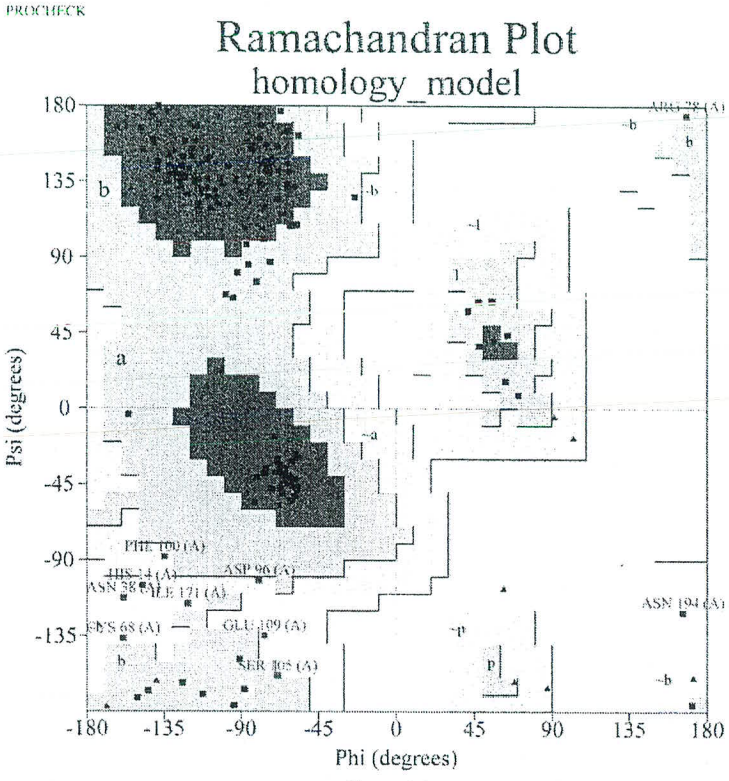
Department of Aquaculture, Faculty of Fisheries, Kasetsart University, 50 Ngam Wong Wan Rd, Chatuchak, angkok 10900, Thailand

Laboratory of Genome Science, Graduate School of Tokyo University of Marine Science and Technology, onan 4-5-7, Minato-ku, Tokyo 108-8477, Japan

Microbiology Department, King Mongkut's University of Technology Thonburi, Bangmod, Toongkru, angkok 10140, Thailand

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

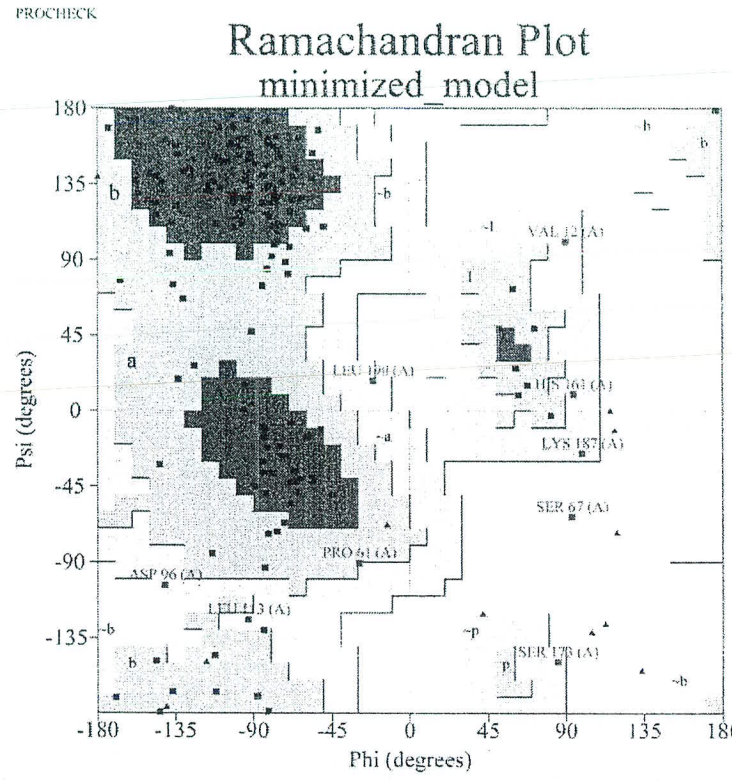
Figure 31 Ramachandran plots for the homology and optimized models of the HIV protease.



Plot statistics

Residues in most favoured regions [A,B,L]	130	76.9%
Residues in additional allowed regions [a,b,l,p]	29	17.2%
Residues in generously allowed regions [~a,~b,~l,~p]	8	4.7%
Residues in disallowed regions	2	1.2%
Number of non-glycine and non-proline residues	169	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	17	
Number of proline residues	8	
Total number of residues	196	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions

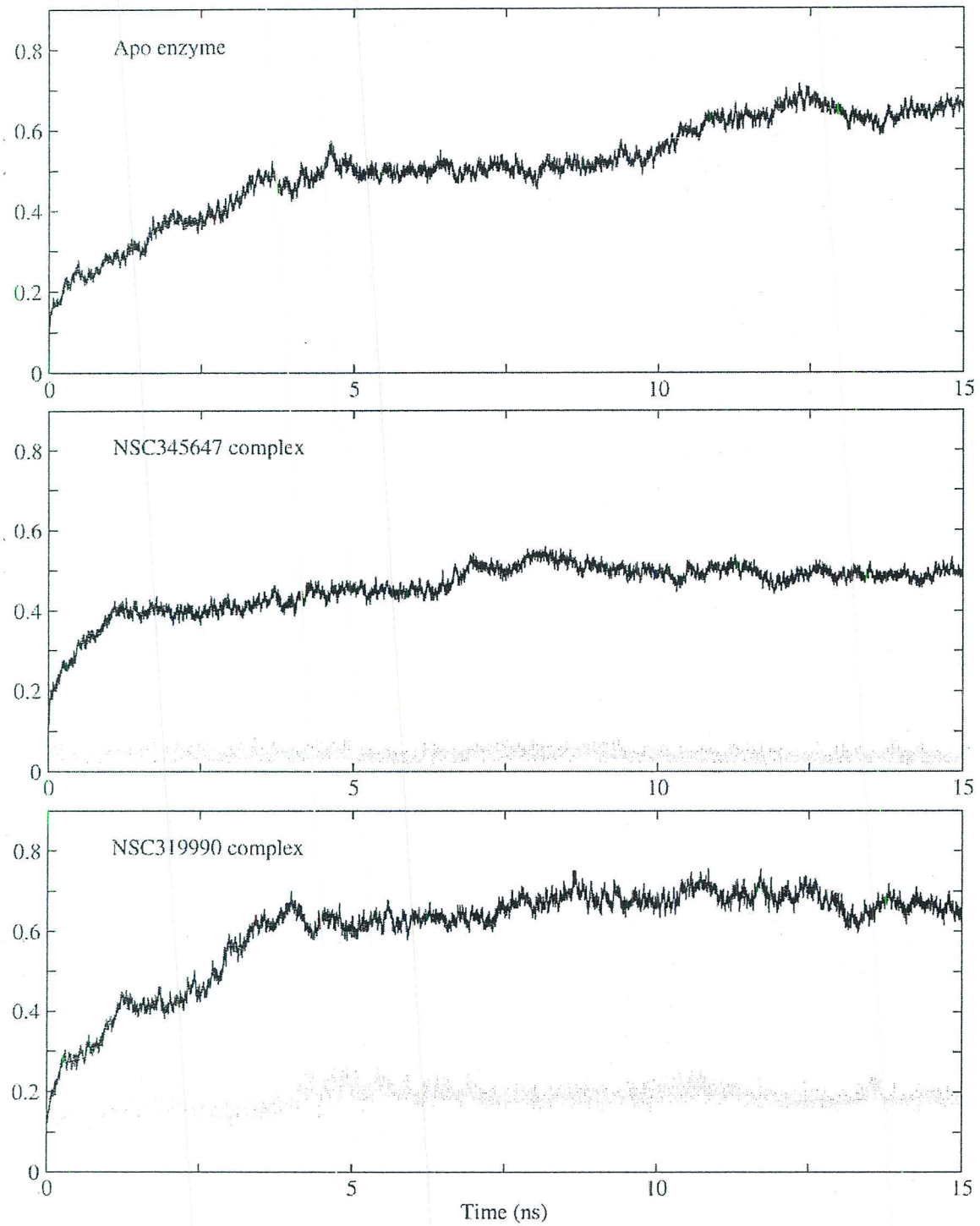


Plot statistics

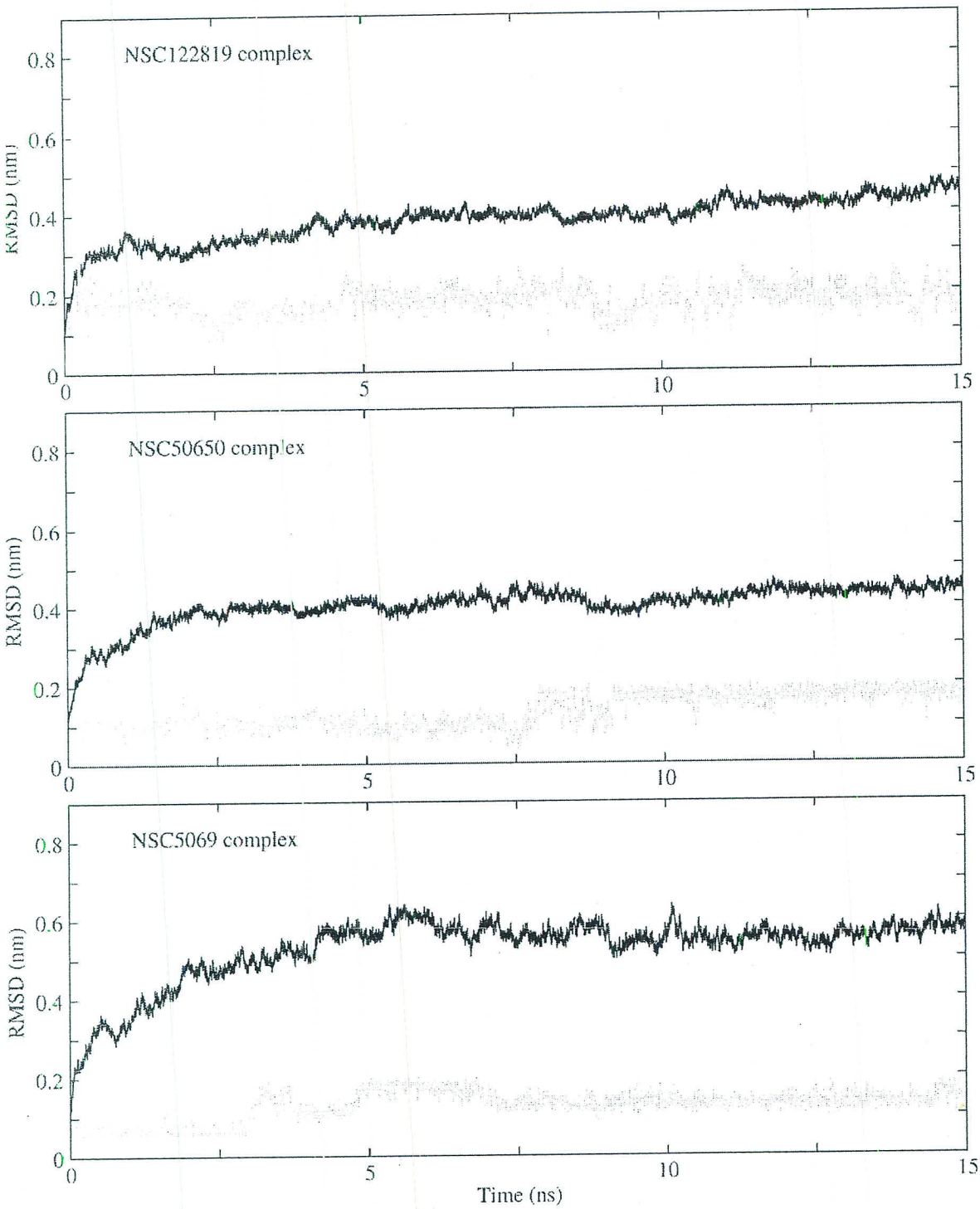
Residues in most favoured regions [A,B,L]	122	72.2%
Residues in additional allowed regions [a,b,l,p]	39	23.1%
Residues in generously allowed regions [~a,~b,~l,~p]	7	4.1%
Residues in disallowed regions	1	0.6%
Number of non-glycine and non-proline residues	169	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	17	
Number of proline residues	8	
Total number of residues	196	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions

gure S2 Time evolution of the RMSD calculated on backbone atoms of the proteins (black .es) and heavy atoms of the compounds (gray lines) with respect to the initial structure efore energy minimization) for the apo and complex simulations.

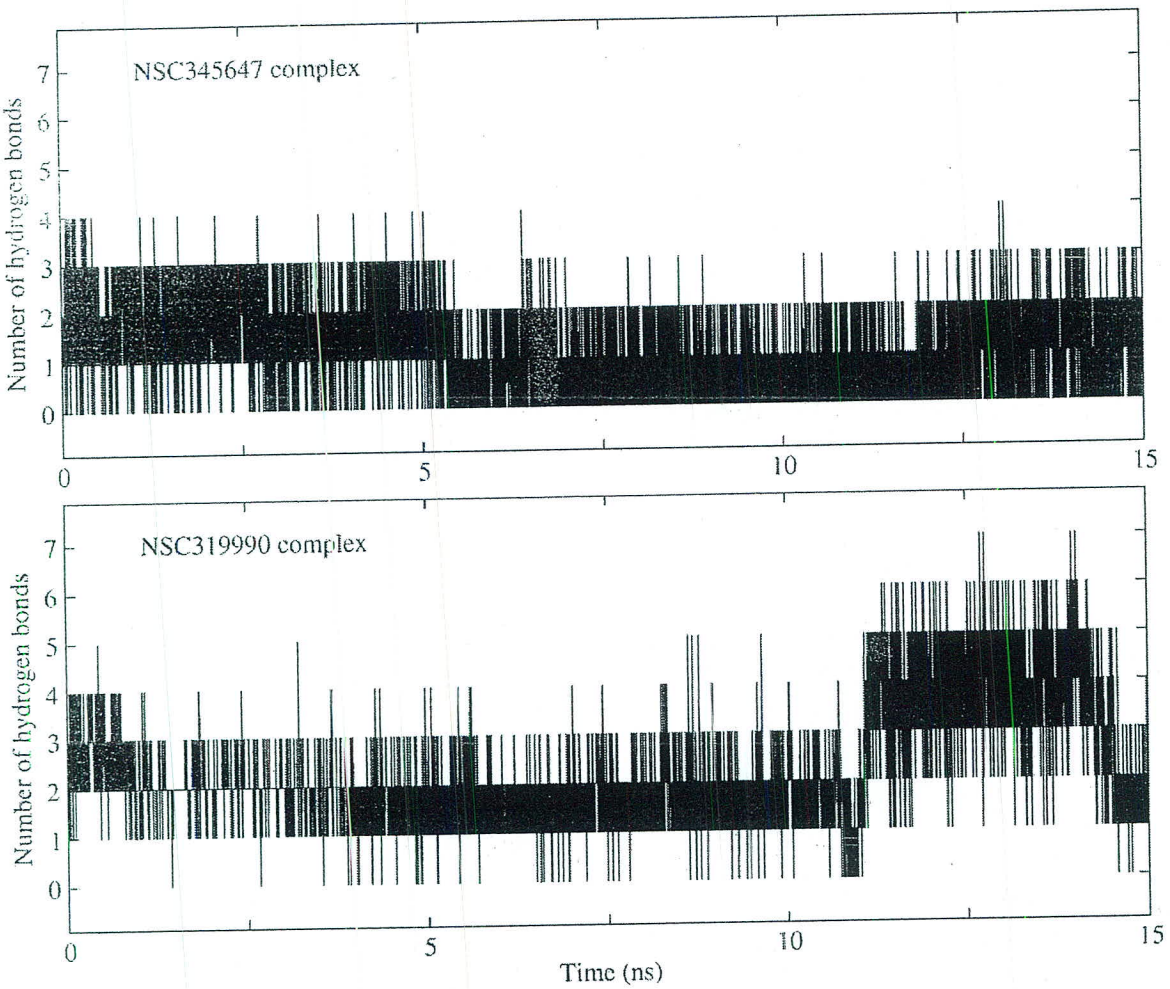


igure S2 (continue).



igure S3 Time evolution of the number of hydrogen bonds between the YHV protease and

t compounds extracted over the course of the simulations.



igure S3 (continue).