2015

In silico drug discovery targeting Chikungunya virus

Phuong Thuy Viet Nguyen
University of Wollongong

UNIVERSITY OF WOLLONGONG
AUSTRALIA

School of Chemistry

IN SILICO DRUG DISCOVERY TARGETING CHIKUNGUNYA VIRUS

Phuong Thuy Viet Nguyen

MSc. in Pharmacy

This thesis is presented as part of the requirements for the award of the Degree of Doctor of Philosophy of the University of Wollongong

March 2015

For Mom and Dad

CERTIFICATION

I, Phuong Thuy Viet Nguyen, declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Chemistry, Faculty of Science, Medicine and Health, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. The document has not been submited for qualifications at any other academic institution.

Phuong Thuy Viet Nguyen
March 2015

PUBLICATIONS

1. Nguyen, P. T. V.; Yu, H.; Keller, P. A. Discovery of *in silico* hits targetting the nsP3 macrodomain of chikungunya virus. *J. Mol. Model*. **2014**, 20, 1-12.

2. Nguyen, P. T. V.; Yu, H.; Keller, P. A. Identification of chikungunya virus nsP2 protease inhibitors using structure-base approaches. *J. Mol. Graph. Model*. **2015**, 57, 1-8.

ACKNOWLEDGEMENTS

I would like to take this opportunity to acknowledge the people who have helped me to complete my PhD study. I feel fortunate to have completed the PhD thesis in such a friendly environment.

First and foremost, I would like to sincerely thank my supervisors Prof. Paul Keller and Dr Haibo Yu, for their guidance, advice, and support during my PhD journey. It has been a great honour to work under your supervision during the past three years. To Paul, I am always thankful to you for giving me an opportunity to join your group and all of the opportunities you have given me. Thanks so much for being encouraging, friendly, and very patient in every respect. To Haibo, thank you so much for your insights into my project. I have always considered you as one of my main supervisors, and I appreciate all of the comments and suggestions you have given me which led to enlightening discussions which helped me to figure out the problems. I have learnt a lot from you, and am glad to have had the chance to work with you.

I would not have been able to conduct my study without financial contribution from the Vietnamese Government (VIED-MOET), the University of Wollongong (IPTA) scholarships, and then the Matching Postgraduate Research Scholarship. I would also like to acknowledge High Performance Computing (HPC) of UOW for computational and research support. Many thanks to Dr Rui Yang for supporting and caring for the HPC cluster and for fixing things that needed to be fixed.

To the Keller group, I cannot record all of the memories I have of people in the group so must summarize. It is an international group with lovely people from different countries. It was really enjoyable time for me to get to know international friends. Many thanks Paul for the group meetings with research updates and weekly problems. It is very helpful and I have learnt and reviewed the chemistry knowledge from all of you. Thanks to all current and former members of the group for sharing many memorable moments, experiences, and valuable academic discussions; along with the many other aspects of life, as well as for the help and friendship. Special thanks to Mohammed for

my first days with the group; and your support and enthusiasm to help me become familiar with molecular modeling. Thanks to Adel for suggestions and support in the project. Thanks Alex for such a very lovely friendship. Thanks so much to Ari for the creative and promising plans in the future. Many thanks to Matthew for always listening and discussing problems with helped me to find the solutions. Thanks Rudi for lovely chatting in the office and the caring. Huge thanks to Jamie and Nick for helping me to double-check my thesis and giving me encouragement. Thanks also to Andrew, Steve Bailey, Steve Wales, Guy and Josh for the conversations. You all are so lovely and I enjoy the time to talk with you very much.

To the Yu group, I would like to thank Haibo for the group meetings with thanks to all people in the group as well. I learned so much computational knowledge from all of you. I will not forget Tom for always being there to help and the friendship we have shared. I appreciate it very much. Thanks to Tiatian for the help; and Andrew, Bhavani, and Kela for friendly talks.

I greatly appreciate opportunities given to me by Dr Glennys O'Brien and Dr Simon Bedford in teaching undergraduate chemistry students. It gave me a wonderful opportunity to have teaching experience in English which was something I had never done before. I learned a lot from it.

Thanks to all the people in the School of Chemistry for creating such an inspiring and welcoming environment for international students. I am grateful to Ellen Manning for her great support and enthusiasm. Thanks also to Chris Hyland for the concern and lovely chats in the school.

For some of my special friends in the school, I would like to say thanks to Zorik for many discussions about computational chemsitry. I could see your passion for science and I learn from it. Thanks to Alex Martyn for a lovely friendship.

To my Vietnamese friends in Wollongong, thanks so much for your friendship over the years. Special thanks to Lien for becoming a good friend during my study here. It was

always relaxing to have chats with you about many things in study and life. Thanks, too to Trang for the friendship.

Outside the school, I am thankful to all Illawara Committee for International Students (ICIS) members for your friendship, your help, and your support. It has been a privilege and a pleasure for me to work with so many ICIS members who are very friendly and enthusiastic which has made my life a happy one here in Wollongong. I always feel relax and refresh from the PhD research after having the social chats with you. Especially, I would like to express my gratitude to Kel Magrath and Bob Colvin who always help, support, and encourage me. My heartfelt thanks to Kel for your love and caring. You are like my dad in Wollongong, thanks so much for that. Many thanks to Bob for helping me with proof-reading and English grammar. My English writing would not be good without your help.

I must not forget to thank my teachers and colleagues in the Faculty of Pharmacy, University of Medicine and Pharmacy in Hochiminh city, Vietnam (my home Univerity). Special thanks to my close colleague, Ms Huong T. Nguyen who is like my older sister, always giving me support. I must also offer thanks to her family for their love and caring. To my close friend, Quyen Duong for the friendship and caring. Also, thanks to Huyen Pham, Loi Huynh and Ha T. Nguyen for their friendships over the years. Thanks for all the sharing emails which encouraged me while I was studying in Wollongong. A big thank to Khac-Minh Thai for his encouragement from proof-reading of the papers work and friendship. It meant so much to me and gave me more motivation to pursue my research.

Finally, I will dedicate my Doctoral thesis to my family for their unconditional love, caring and encouragement. To my parents, thank you for always giving me the freedom to choose my own path, and giving me support to succeed. To my brother and my sister-in-law, thank you for being there whenever I need you. Without your support and belief in me, a higher education in a foreign country would always have remained my dream.

Phuong Thuy Viet Nguyen
March 2015

ABSTRACT

In recent years, there has been an emergence or re-emergence of Chikungunya virus (CHIKV), a member of the alphavirus. The virus is one of the arboviruses, and it is classified as a neglected tropical disease in more than 55 different countries in the world, including many African and Asian countries, Europe, Americas, and Australia. In 2008, it was listed in the US National Institute of Allergy and Infectious Disease (NIAID) category C priority pathogen due to its morbidity and mortality rates. In addition to damaging global health, the virus also imposes a huge economic burden on affected countries. However, there is currently no licensed vaccine or effective drug to combat the disease. Up to now, there have been few studies focusing on finding potential inhibitors of CHIKV. Taking advantage of all available data about CHIKV and a combination of different computational methods, this study aimed to discover and develop an approach leading to identifying inhibitors against this virus. The study targeted the non-structural proteins, nsP3 macrodomain and nsP2 protease, which play crucial roles in the viral replication and transcription (Chapter 2 and Chapter 3), and the envelope glycoprotein complexes responsible for virus entry and attachment (Chapter 4). Initially, this study searched for potential binding pockets of the CHIKV protein structures. A combination of computational tools including molecule docking, virtual screening, molecule dynamics simulations, and binding free energy calculations were used in this approach. A number of lead compounds to fight CHIKV disease were identified. The insights into the interactions between CHIKV inhibitors and their targets were elucidated. Our findings open a way which would be helpful for the further research on antiviral rational drug design, especially design of inhibitors for CHIKV and also contribute to the guidelines for the drug discovery and development.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

1D	First Dimension
2D	Two Dimension
3D	Three Dimension
AMBER	Assisted Model Building with Energy Refinement
CADD	Computer-Aided Drug Design
CC_{50}	50% Cytotoxic Concentration
cDNA	Complementary DNA
CHARMM	Chemistry At Harvard Molecular Mechanics
CHIKV	Chikungunya Virus
cryo-EM	Cryo-Electron Microscopy
DNA	Deoxyribonucleic Acid
dsRNA	Double-stranded RNA
EC_{50}	50% Effective Concentration
ECSA	East Central South African
EEEV	Eastern Equine Encephalitis Virus
elec	Electrostatic
ELISA	Enzyme-Linked Immunosorbent Assay
FEP	Free Energy Perturbation
GAFF	Generalized AMBER force field
GARD	Generally Applicable Replacement for RMSD
GROMACS	GROningen MAchine for Chemical Simulations
GTP	Guanosine 5'-Triphosphate
H-A	Hydrogen bond acceptor
HBs	Hydrogen Bonds
H-D	Hydrogen bond donor
HSP	Heat shock protein
IC_{50}	50% Inhibitory Concentration
ID	Identification
IgG	Immunoglobulin G
IgM	Immunoglobulin M

IMP dehydrogenase	Inosine 5'-Monophosphate Dehydrogenase
IRES	Internal Ribosome Entry Sequence
IU	International Unit
K_d	Dissociation Constant
K_i	Inhibitory Constant
L	Ligand
LIE	Linear Interaction Energy
logP	A calculated octanol-water partition coefficient
MC	Monte Carlo
MD	Molecular dynamics
MM-GBSA	Molecular Mechanics-Generalized Born Surface Area
MM-PBSA	Molecular Mechanics-Poisson Boltzmann Surface Area
MW	Molecular weight
NAMD	Not (just) Another MD Program
NCBI	National Center for Integrative Biomedical Informatics
NCI	National Cancer Institute
NIAID	National Institute of Allergy and Infectious Diseases
NMR	Nuclear Magnetic Resonance
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
NTP	N: number of particles, T: temperature, P: pressure
OAS3	Oligoadenylate Synthetase 3
OPLS	Optimized Potentials for Liquid Simulations
ORFs	Opening Reading Frames
P	Protein
PDB id	Protein Data Bank Identification Code
pKa	The logarithm of the acid dissociation constant
PL	Protein-Ligand
PME	Particle Mesh Ewald
RC	Replication Complex
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic acid

ROC	Receiver Operating Characteristic
ROC AUCs	Area under ROC curves
RT-LAMP	Real-Time Loop-Mediated Isothermal Amplification
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SASA	Solvent accessible surface area
SFV	Semliki Forest virus
SINV	Sindbis virus
siRNA	Small interfering RNA
SO	Swam Optimization
TI	Thermodynamic Integration
TIP3P	Transferable intermolecular potential 3 points
TM	Transmembrane
UB	Urey-Bradley
US	United States
vdW	van der Waals
VEEV	Venezeulan Equine Encephalitis Virus
VMD	Visual Molecular Dynamics
VST	Virtual Screening
WNDR	Weighted Desolvation Non-Polar Ratio
ZINC	Zinc Is Not Commercial

CHAPTER 1. INTRODUCTION

1.1 OVERVIEW OF CHIKUNGUNYA VIRUS

1.1.1 GLOBAL EXPANSION OF CHIKUNGUNYA DISEASE

Chikungunya disease is caused by the chikungunya virus (CHIKV), one of the arthropod-borne viruses (arboviruses, or virus spread by mosquitoes) that has emerged[1-2] or re-emerged[3-4] in recent years. This virus is considered a neglected tropical disease; in 2008, it was listed as a category C priority pathogen by the US National Institute of Allergy and Infectious Diseases (NIAID) due to its morbidity and mortality rates.[5] The CHIKV produces a dengue-like illness, which may lead to misdiagnosis. The disease is non-fatal,[6] however its debilitating symptoms such as fever, rash, headache, myalgia (muscle pain), and arthralgia cause enormous health problems,[7] affecting millions of people in nearly 55 different countries around the world.[8] Importantly, the polyarthralgia can exist in some infected patients for months.[7, 9-10] The history of the virus shows the first recorded case was in Tanganyika, Africa, in 1952.[11] An explanation for the name CHIKV is that it was derived from a local dialect, meaning "that which bends up"[11] in order to describe the stooped posture of patients who suffer from joint pains for weeks to years.[6, 11] The virus has been largely neglected due to its sporadic re-emergence,[4, 6] however it has recently attracted interest given a rise in epidemics occurring since 2006 in different countries; from Africa to Asia, Europe, Americas, and Australia, spread by infected travellers.[6, 10, 12] The expansion of the virus has drawn global attention with the prevalence listed in Table 1.1 and Figure 1.1. Some extensive and expanding epidemics occurred in some large cities, affecting potentially millions of people. However, there is currently no cure for the CHIKV.

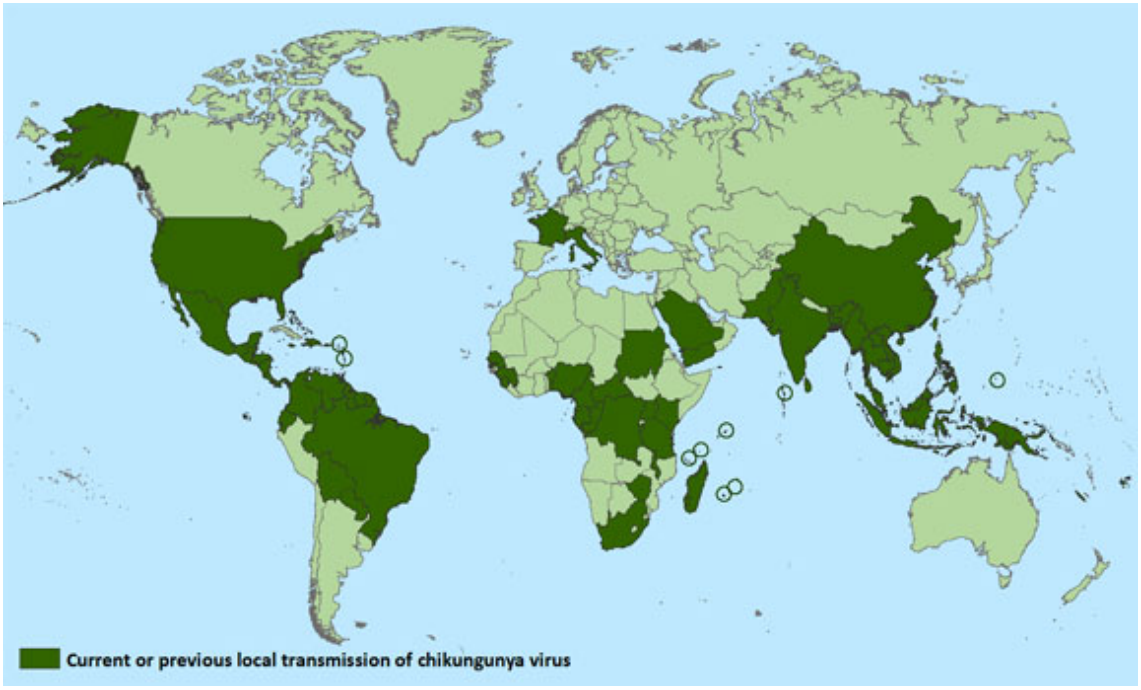


Figure 1.1. Global expansion of CHIKV (taken from the website of Centers for Disease Control and Prevention (http://www.cdc.gov/chikungunya/geo/index.html) with countries and territories where chikungunya cases have been reported with endemic or epidemic, updated 24th of February, 2015).

Table 1.1. Prevalence of CHIKV in terms of year and infected countries.

Year	Infected countries or regions	Outbreak
1950s	Tanganyika,[11, 13] Uganda,[14] Thailand,[15] Phillipines[16]	
1960s	India, Sri Lanka, Myanmar and Thailand,[17] Cambodia,[18] Vietnam, Laos,[19] Pakistan, Malaysia,[20] Taiwan,[6] and Phillipines[15]	≥ 100,000 cases and caused 200 dealths
1970s	India and Southeast Asia[15]	
1980s	Thailand, Philippines, and Indonesia[21]	
1990s	Malaysia[22-24] (514,271 cases), Congo[25] (50,000 cases)	
2001-2003	Indonesia, Cambodia, Vietnam, Myanmar, Pakistan, Thailand,[15] the Indian Ocean islands of the Mauritius, Mayotte, Seychelles, Madagascar, and La Reunion[22]	
2004	Comoros[5] (5,000 cases)	
2005-2006	The Indian Ocean islands of the Mauritius (15,760 cases), Mayotte (6,346 cases), Seychelles, and La Reunion (244,000 cases/total population of 770,000),[26] Indonesia,[19] Guinea,[27] India,[28-29] US[30]	Indian Ocean (300,000 cases with 237 deaths)

Year	Infected countries or regions	Outbreak
2006-2007	Europe, US, Australia,[31] Maldives (11,879 cases), Gabon (48,000 cases)[32], Mayotte (6,346 cases, Mauritius (15,760 cases), Italy (248 cases), Sri Lanka (40,000 cases), Indonesia (15,000 cases),[33] Cameroon,[34] India[29, 35]	India (1.4-6.5 million)
2008-2009	India (95,000 cases in 2008; 68,000 cases in 2009), Singapore (1,033 cases),[36] Thailand,[37-38] Malaysia[39] (7,000 cases), Taiwan,[40] US,[41-42] Indonesia,[33] Bangladesh[43]	
2010	France[44], China,[45] Thailand,[17] Canada, Myanmar[33]	
2011	India, Cambodia[18]	
2012	Australia[46], Bhutan,[47] Canada, Cambodia,[18] Papua New Guinea[48]	
2013	Caribbean, Canada, Thailand[49]	
2014	France,[50] Europe, US, Caribbean[51]	
2015	US[52]	

1.1.2 THE SPREAD OF CHIKV

CHIKV is transmitted from human-mosquito-human (urban cycle) or animal-mosquito-human (sylvatic cycle) by the bite of infected mosquitoes.[22] The virus is classified into three genotypes: the Asian, West African, and East Central South African (ECSA).[6] Genetic analyses revealed that the CHIKV originated in Africa and expanded to Asia. The spread of CHIKV from Asian and African countries to other areas is due to an association of various factors; namely worldwide distribution of the transmission vectors, climatic conditions, and inadequate mosquito control. The global expansion of this disease is solely primarily due to an increase in international travel, with disease transportation from infected travellers.[1-3, 22] The mosquito species, *Aedes aegypti* was a principal vector in many outbreaks, while *Aedes albopictus*, the Asian tiger mosquito, has been considered a primary re-emergence factor since 2005.[5-6, 22] The vector switch was found to be a result of reduced populations of *Aedes aegypti*, meaning that viral transmission was primarily caused by the larger *Aedes albopictus* population.[3] Furthermore, a number of adaptive mutations have allowed for exploitation of the new epidemic vector.[53] In particular, a mutation in the E1 envelope protein, Ala226Val, was responsible for a dramatic increase in CHIKV infectivity for *Aedes albopictus* since

2005 in Africa and Asia.[54-56] In addition, a substitution of lysine by glutamic acid at the position 211 of the E1 resulted in adaption of the virus to the *Aedes albopictus*.[5, 22, 53] Recently, the E2-Ile211Thr substitution was shown to help the virus adapt to *Aedes albopictus*, as this mutation could set up the foundation for the E1-Ala226Val mutation that provides the enhanced infectivity of CHIKV.[33] Also, the E2-Gly60Asp was a determinant factor in CHIKV infectivity for both these species.[57]

1.1.3 VIRUS LIFECYCLE

As with other alphaviruses, the CHIKV life cycle begins with attachment to a host cell via receptor mediated endocytosis in clathrin coated vesicles,[8, 22, 58] a process described in Figure 1.2. Under the acidic pH of the endosome, there are conformational changes in the structures of the envelope glycoprotein complexes. The complex E1 and E2 heterodimers dissociate to form the E1 trimers. The trimers use the hydrophobic fusion loop to insert into the host membrane, and refold to form a hairpin-like structure. The nucleocapsid and viral genome are then released into the host cell cytoplasm. During translation, the nsP123 (a polyprotein precursor) from the viral genome binds to the nsP4 to form the replication complex (RC). The RC then produces the full length minus strand (negative strand RNA) required for replication. When the nsP123 concentration increases, the nP123 is cleaved to non-structural proteins nsP1, nsP2, nsP3, and nsP4. As these non-structural proteins and the host cell proteins serve as the plus strand (positive strand) in replication, they produce the 26S subgenomic RNAs and genomic (49S) RNAs. The 26S subgenomic positive stranded RNA encodes the polyprotein precursor for structural proteins. The cleavage process takes place in the Golgi complex, and then the products (non-structural proteins and structural proteins) are transported to the plasma membrane. The viral RNA is packaged into the nucleocapsid, and the mature virions bud out of the plasma membrane.

.

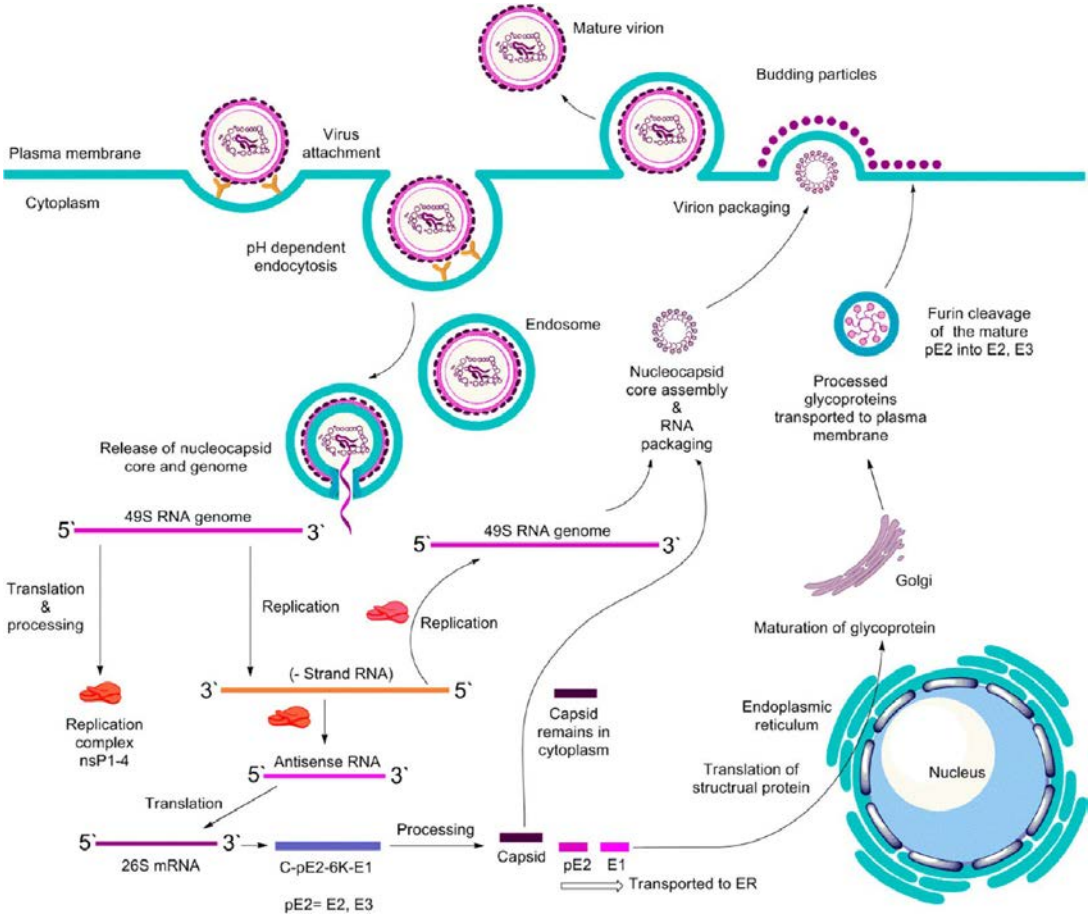


Figure 1.2. Replication cycles of CHIKV. The picture is reproduced from Rashad A. *et al*[8] with permission).

1.1.4 CLINICAL SYMPTOMS

Following the transmission, CHIKV replicates and expands to liver and joints.[5] There are similar symptoms with other viruses, such as dengue fever, yellow fever, or Ross River fever. After an incubation period of 2-4 days, acute clinical symptoms start with the onset of high fever, rigors, headache, fatigue, nausea, vomiting, myalgia, and rash.[5] In particular, intense joint pain (polyarthralgia) is the most characteristic which incapacitate patients.[5] This acute stage lasts from 1-10 days. The chronic stage depends on the acute stage, and is characterized by polyarthralgia. It can last from weeks to years,[8] which affects individual patients, and results in social health impacts[10] due to debilitating infection in large working population. Neurological disorders[59] and eye infection[60] are also reported in infected patients, while meningoencephalitis and

haemorrhagic manifestations such as haematemesis and malaena leading to death have been reported.[6]

1.1.5 DIAGNOSIS

Since no vaccines or effective therapeutics are available, early detection and proper diagnosis are becoming increasingly important in treating the CHIKV disease. Viral culture is considered a gold standard for CHIKV diagnosis. It is based on innoculation of mosquito cell cultures, mammalian cell cultures, or mice.[61] However, for a rapid diagnosis, techniques such as a detection of reverse transcription polymerase chain reaction (RT-PCR), real-time RT-PCR, real-time loop-mediated isothermal amplification (RT-LAMP) are recommended.[1, 62] More frequently, serodiagnostic methods;[63-64] such as enzyme-linked immunosorbent assay (ELISA), indirect immunofluorescent method, hemagglutination inhibition, or neutralization techniques were used effectively as reliable techniques in the identification and characterization of CHIKV. The detection of immunoglobulin M (IgM) and immunoglobulin G (IgG) antibodies using capture enzyme-linked immunosorbent assay (MAC-ELISA) is a rapid and reliable technique in serology.

1.1.6 CHIKV GENOME AND STRUCTURE

A full understanding of the structure and genome of CHIKV is crucial for the development of drugs to combat the virus. CHIKV is a positive-sense and single-stranded RNA virus in the alphavirus genus, Togaviridae family.[65] In 1984, Simizu *et al* used African and Asian strains of CHIKV to analyse the structural proteins of CHIKV.[66] In 2002, Khan *et al* identified the full genomic sequence of CHIKV (S27, African prototype),[67] opening up further investigations into the elucidation of the structure and genome of this virus. CHIKV's genome includes two opening reading frames (ORFs) that consist of 11,805 nucleotides in total without including the cap at the 5' end, a l-poly (A) tract, and a poly (A) tail at 3' end.[67] The first ORF has 2474 amino acids encoding non-structural proteins (nsP1, nsP2, nsP3, and nsP4) at the 5' region, while the second consists of 1244 amino acids encoding structural proteins (the capsid C, envelope glycoproteins E1, E2, E3, and 6K). Between these two ORFs, there is a

junction region. In addition, there is a 7-methylguanosine group capped at the 5' end and the polyadenylated group at the 3' end. The arrangement of the genome can be as follows: 5'-cap-nsP1-nsP2-nsP3-nsP4-(junction region)-C-E3-E2-6K-E1-poly(A)3' (Figure 1.3).

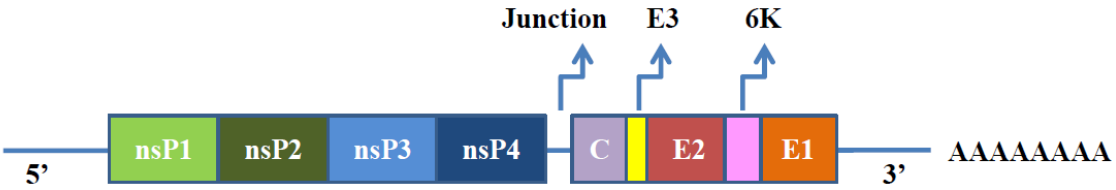


Figure 1.3. Genome organisation of CHIKV (adapted from Singh S. K. and Unni S. K.[22]).

The role and function of non-structural and structural proteins of CHIKV have been examined based on information of other alphaviruses, such as Sindbis virus (SINV) and Semliki Forest virus (SFV). According to these studies, the non-structural proteins of CHIKV play an important role in the formation of the transcription/replication complex of the virus and the negative strand synthesis.[58, 68-70]

Studies of SFV and SINV nsP1 indicate that nsP1 is a multifunctional protein.[71] It is involved in the formation of the virus cap,[70, 72-73] and directing RNA replication complex to membranes or liposomes,[74] as well as associating with endosomes and lysosomes at the cytoplasmic surface of membranes.[72, 75] It has also been implicated with both guanine-7-methyltransferase and guanylyltransferase activities.[74]

Investigations of nsP2 from SINV and SFV show that the nsP2 has multiple enzymatic activities.[76-77] Primarily, it belongs to the papain superfamily of cysteine protease.[78] It has functions of proteolytic enzyme at C-terminal domain whereas at the N-terminal domain, it posses the activities of ATPase and GTPase,[76-77] RNA helicase,[79] and RNA triphosphatase.[80] Moreover, it is involved in the regulation of synthesis of the 26S subgenome[81-82] and activates the switch from early to late stages,[82] as well as playing a role in the translocation of 50% of the translated nsP2 into the nucleus.[83] The nsP2 plays a vital role in the replication of virus; which combined with the exhibition of some degree of sequence specificity,[76] has resulted in it being an attractive target for drug

design. More recently, the crystal structure of nsP2 protease of CHIKV was solved (PDB id: 3TRK).

The specific functions, roles, and activities of the nsP3 protein remain relatively elusive; however, genetic and functional analysis of the nsP3 from SINV revealed that it is a phosphoprotein participating in the process of synthesis of the minus strand and the subgenome of the virus.[69, 73, 84] It has been reported that deletion of phosphorylation sites in the SFV nsP3 decreases the level of RNA synthesis.[85] The nsP3 protein consists of two domains; the N-domain which is highly conserved, and the C-domain which is not.[73] There is an "X domain", or macrodomain (a first 160 amino acid domain with unknown function), present at the N-terminal region of the nsP3 with a determined crystal structure (PDB id: 3GPG).[73] The structure on its conserved adenosine binding site was also obtained (PDB id: 3GPO).[73] In addition, it was found that the mutation of amino acids at the position Asn10 and Ala24 in ADP-ribose binding of the nsP3 macrodomain in SINV affects the replication and viral RNA synthesis, though it has no effect on the binding region.[86-87]

Several attempts have been made on SINV to propose the roles of the nsP4 in the RNA-dependent RNA polymerase activity[88] and in replication, and transcription of the virus.[88-91] The results showed that the N-terminal Tyr residue of the nsP4 of SINV may be substituted with Phe, Trp, or His without changing the wild-type phenotype in cultured cells. However, other substitutions, except for Met, were lethal or quasilethal.[91] The nsP4 is also stable and remains active during the infection cycle.[89] Previous research on the mutants reported that the mutations in the nsP4 were Glu191 subtituted for Leu and Glu315 to Gly, Val, or Lys, together with one mutation in the nsP1 (Thr349 to Lys); which suppress the minus strand RNA synthesis.[91] Arg183 of SINV nsP4 polymerase was found to have an important role in alphavirus minus strand RNA synthesis.[92]

Most studies of the structural proteins of CHIKV are based on the biology and pathogenesis of the virus, as infection is mediated by these glycoproteins. The E2 is responsible for receptor binding, while membrane fusion is supported by the E1.[93] In 2008, Santhosh *et al*[56] found the mutation in E1 Ala226Val, leading to the epidemic outbreaks of CHIKV in India. Ongoing insights into the structure of CHIKV have been

revealed by two crystal structures of surface glycoprotein complexes; namely precursor p62-E1 heterodimer (PDB id: 3N40), and the mature E3-E2-E1 (PDB id: 3N42) determined in 2010.[93] The E3 protein plays an important role in the proper folding of p62 and the formation of the p62-E1 heterodimer,[94-95] but the E3 is not in the component of mature CHIKV.[66] The 6K associates with the complex p62-E1 and is transported to the plasma membrane before assembly.[65] The assembly process takes places due to the interactions of the genomic RNA and the nuclear capsid protein. The 6K protein facilitates particle morphogenesis, but it is not stoichiometrically incorporated into virions.[65] The structural change of envelope glycoproteins in membrane fusion was investigated by Li *et al.*[96] The resulting roles of the E2 in receptor interactions, and the existence of epitopes, are essential for the design of a vaccine against this virus.

Recently, structural analysis of CHIKV at pseudo-atomic level resolution was reported,[65] by combining electron cryo-microscopy (cryo-EM) techniques for the whole virus and X-ray crystallography for the component of structural proteins, together with the published crystal structure of the CHKV E1-E2 glycoprotein heterodimer.[93] A 5.3 Å resolution cryo-EM map of CHIKV-like particles was interpreted, and the mechanisms of neutralization of antibodies were proposed. The study revealed that like other alphaviruses, CHIKV has an icosahedral spherical structure with T=4 quasi-icosahedral symmetry (diameter of about 60-70 mm, Figure 1.4 A). This structure consists of 80 spikes: including 20 icosahedral “i3” spikes (located on the icosahedral 3-fold axes), and 60 quasi-3-fold “q3” spikes (located in general positions) with a quasi-3-fold axis.[65] The spikes i3 and q3 are significantly different, possibly indicating different stages of generation of fusogenic E1 trimers. The complete q3 spike combines with one-third of an i3 spike to form a single T=4 icosahedral asymmetric unit (Figure 1.4 A). These spikes are made by the envelope glycoproteins E1, E2, and together with virus membrane, transmembrane (TM) helix, and a capsid covering genome RNA to form a nucleocapsid (Figure 1.4 B). The E1 glycoprotein is composed of 439 amino acids including 404 N-terminal residues, 30 residues which comprise the TM helix, and 5 amino acids which form the cytoplasmic domain, as well as an N-linked glycosylation site at Asn141. The E1 is divided into three domains, namely domain I, II, and III, with domain I located in between domain II and III. The E2 consists of 364 residues, a 26

residue TM helix, and 33 residues cytoplasmic domain. The E2 also has three domains, domain A, B, and C, which are known as distinct immunoglobulin-fold domains. Domain A takes part in receptor-binding process, lying in between domain B and domain C. Domain B covers the fusion loop in domain II of E1 in the mature structure. Under the pH acidic environment, the virus becomes fusogenic by combining three E1 to make a trimer. The fusion loop is then exposed to insert into the host cell membrane. Therefore, hiding the "fusion loop" may interfere with the virus entry and infection of human tissue.

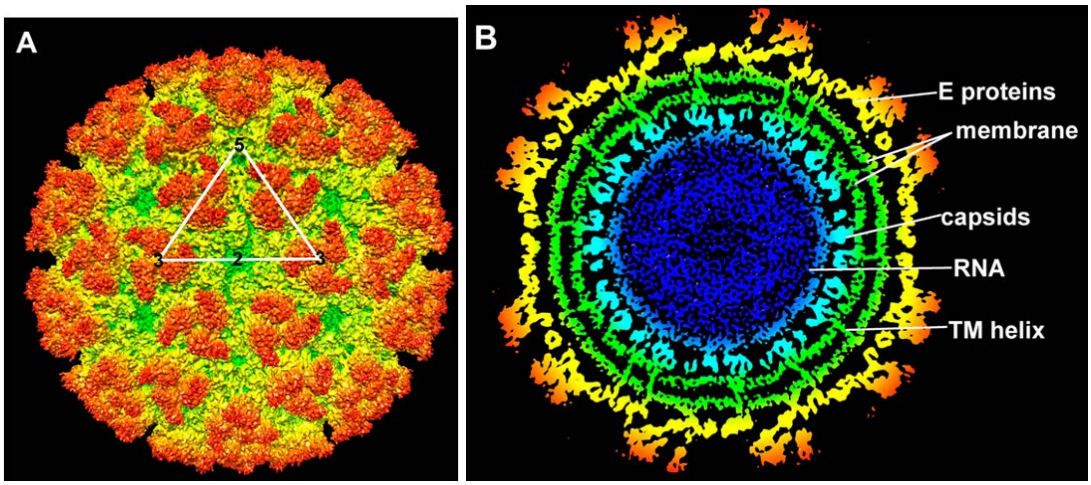


Figure 1.4. Structure of the CHIKV, showing the structure of virus particles in a T=4 icosahedral symmetry **(A)** and the components of a single unit with a nucleocapsid consisting of spikes (made by envelope glycoproteins), virus membrane, and transmembrane (TM) helix; and a capsid covering genome RNA inside **(B),** (adapted from Sun S. e*t al*.[65])

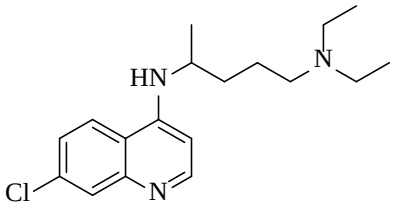
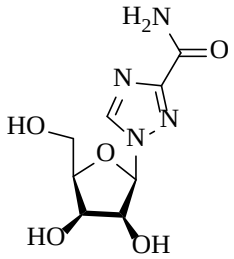
1.2 CURRENT TREATMENT FOR CHIKV INFECTION

There is no effective vaccine or antiviral agent currently available for CHIKV infection.[10] Treatment is mostly based on alleviating symptoms by using analgesics (non-salicylate analgesics), antipyretics,[1, 6] anti-inflammatory agents (corticosteroids and non-steroidal anti-inflammatory (NSAIDs));[6, 9, 21] along with taking bed rest, and undertaking an extra fluid diet.[1] Some agents for the treatment of acute CHIKV include Methotrexate; NSAIDs: Rofecoxib, Celecoxib, Parecoxib; corticosteroids: Prednisolon; antirheumatic: Sulfasalazine and Methotrexate; non-salicylate analgesics: Paracetamol, Morphine; traditional herbal medicine: *Fernelia* spp species.

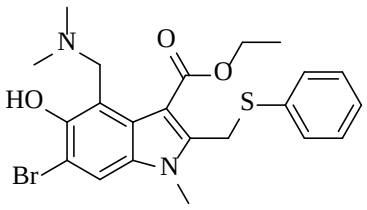
1.2.1 SPECIFIC INHIBITORS OF CHIKV

While some inhibitors have been shown to be effective against CHIKV *in vitro*, currently, there is no antiviral drug available for CHIKV treatment. Further biological testing (*in vivo*) is required. These inhibitors are listed in Table 1.2, along with the current update of inhibitory effects and the clinical trials.

Table 1.2. Some potential inhibitors for CHIKV.

Compound	Inhibitory effects and current update
Chloroquine  **(1)** IC_{50} = 1.1 µg/ml	Chloroquine **(1)** was reported as *in vitro* antiviral compound more than 35 years ago.[97] However, recently, a mouse model showed that this compound enhances virus replication, and aggravates the disease.[9, 98] Chloroquine phosphate has also been used for chronic chikungunya arthritis with anti-inflammatory, rather than antiviral effects.[99-101] Some studies revealed that chloroquine was considered as an entry inhibitor as it could interact with the endosome-mediated internalization process in the infection cycle.[98, 102] Chloroquine was tested in Phase III clinical trials in France in 2006; however, this compound was terminated in 2007 with no anti-CHIKV effect.[8]
Ribavirin  **(2)** EC_{50} = 83.3 µg/ml	Ribavirin **(2)** is well-known as an antiviral inhibitor *in vitro*.[103] The mechanism is varied between different viruses, such as interaction with intracellular viral RNA production, inhibition of inosine 5'-monophosphate dehydrogenase (IMPDH), leading to depletion of cellular GTP pools, and action as a potent mutagen for some RNA viruses (error catastrophe mechanism), even though the precise mechanism is still unclear.[9] Ribavirin can be used either alone or in combination with α-interferon to get a synergistic effect *in vitro* against CHIKV (concentration of α-interferon 3.9 IU/ml, and ribavirin 18.75 µg/mL).[104-105] However, up to now, no evidence on clinical efficacy of ribavirin on CHIKV or ribavirin in combination with α-interferon for anti-CHIKV is reported.[9]
Arbidol	Arbidol **(3)** was developed 20 years ago in Russia for treatment of acute respiratory infections.[8] This compound elicits a broad effect on RNA, DNA, enveloped, and non-enveloped viruses.[106] The mechanism may be interference with the mediated

Arbidol

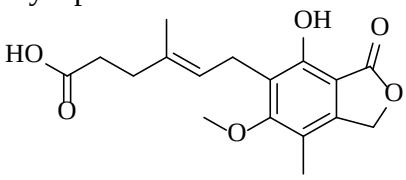


(3)

IC_{50} = 12.2 μM

fusion,[107] blocking the viral entry into the target cells through inhibition of glycoprotein conformational changes.[108-110] In 2011, arbidol and its derivatives were used for *in vitro* testing for CHIKV. The results reported that arbidol could inhibit CHIKV but very weak effects even though the IC_{50} value is much lower than the toxic concentration (IC_{50} = 12.2 μM, $CC_{50} \geq 200$ μg/mL).[111] Evidence from *in vivo* studies is required to validate activity of arbidol on CHIKV.[8]

Mycophenolic acid

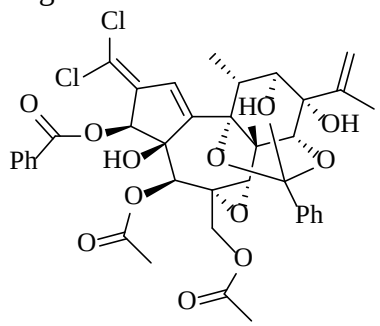


(4)

IC_{50} = 0.2 μM

Mycophenolic acid **(4)** was discovered approximately 100 years ago.[112] It is an inhibitor for the enzyme IMPDH involved in *de novo* biosynthesis of guanine nucleotide.[113] It was known as having good activity with antiproliferation, anticancer, as an antiviral agent, and an immunosuppressant.[8] Recently, this compound showed inhibition of the CHIKV replication using virus-induced cell death.[114] However, the compound was reported as suffering from a metabolic drawback associated with rapid conjugation of the C-7 phenolic hydroxyl group with glucuronic acid,[113] thus *in vivo* studies are required.[8]

Trigocherrin A

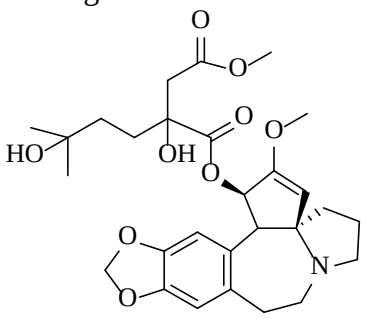


(5)

EC_{50} = 1.5 μM

Trigocherrin A **(5)** is a natural compound, isolated from the bark of *Trigonostemon cherrieri Veillon* (*Euphorbiaceae*),[115] a tree in New Caledonia, or the species found in tropical Asia, India, and Sri Lanka to New Guinea.[8] Recently, in testing CHIKV inhibitory effect, this compound showed inhibition of viral replication function on virus-induced cell death, using a virus-cell-based assay.[116]

Harringtonine[117]



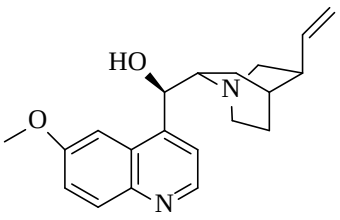
(6)

EC_{50} = 0.24 μM

Harringtonine **(6)** is an alkaloid compound, isolated from *Cephalotaxus harringtonia* tree in Japan.[8] It displayed CHIKV inhibition, by affecting the early stages of infection after cellular endocytosis.[117] Also, it was found to affect the CHIKV RNA production inside the infected cell and viral protein expression through the nsP3 and the E2 proteins.[117] This compound is still undergoing *in vivo* testing.[8]

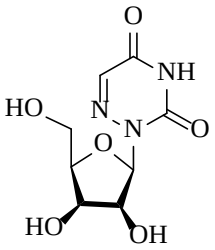
Some other compounds are less common, thus requiring further testing to be considered promising lead compounds (Figure 1.5). However, they were not approved for human use. For example, quinine **(7)** inhibited the *in cellulo* growth of CHIKV.[9] Increasing concentrations of quinine affected the nsP1, as mutations were observed suggesting an impairment of the function of the viral guanylyltransferase activity. Another compound, named 6-azauridine **(8)**, showed its inhibition on both DNA and RNA virus replication, and orotidine monophosphate decarboxylase, an enzyme involved in the *de novo* biosynthesis of pyrimidine, cytidine, and thymidine.[118] Trigowiin A **(9)** and prostratin **(10)** were isolated from the bark of *Trigonostemon howii* of the *Euphorbiaceae* species in central Vietnam during the biological testing for CHIKV.[119] This compound showed a weak anti-CHIKV activity, however, it possessed its selectivity for CHIKV, against the SINV and SFV. Prostratin[119] showed better inhibitory effects on CHIKV than **(9)** in a CHIKV inhibition assay. 12-o-tetradecanoylphorbol 13-acetate **(11)** was reported to inhibit CHIKV through the activation of the signal transduction enzyme protein kinase C, also selective for CHIKV inhibition.[119] Lupenone **(12)** and β-Amyrone **(13)** are isolated from the leaves of *Anacolosa pervilleana* (Madagascan plant). **(12)** showed moderate anti-CHIKV activity in a virus-cell-based assay,[120] while (**13)** had a moderate anti-CHIKV activity in the assay. Polycytidylic acid or poly(I:C) **(14)** is a synthetic double-stranded RNA (dsRNA) analogue. It displayed an immunostimulant action as an inducer for the interferon via interaction with the toll-like receptor 3 in CHIKV infection assay.[121]

Quinine



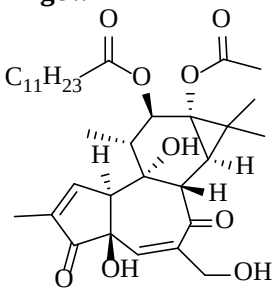
(7)

6-Azauridine



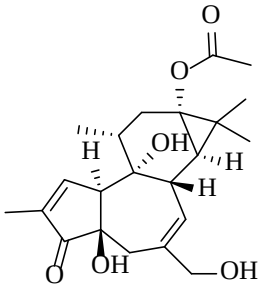
(8)

Trigowiin A



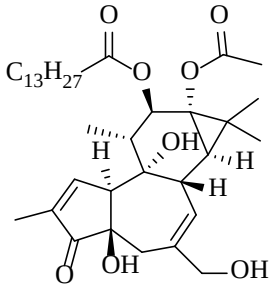
(9)

Prostratin



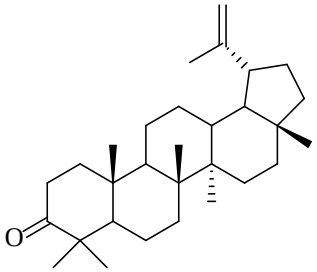
(10)

12-O-Tetradecanoylphorbol 13-acetate



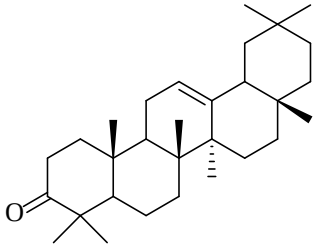
(11)

Lupenone



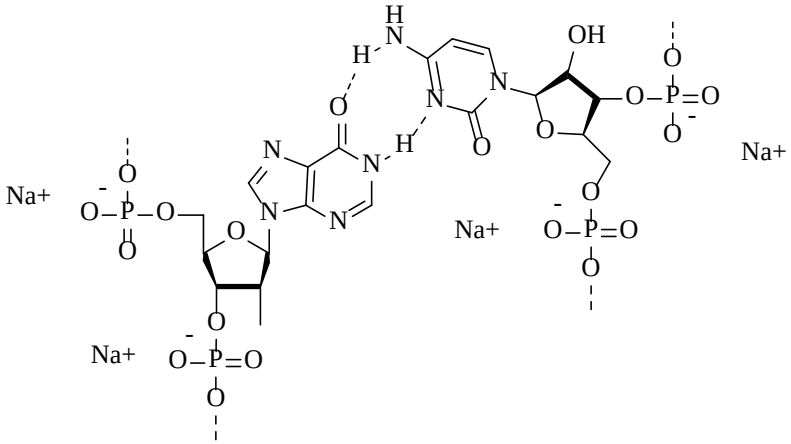
(12)

β-Amyrone



(13)

Polycytidylic acid



(14)

Figure 1.5. Less common structures of some potential inhibitors for CHIKV.

1.2.2 CURRENT APPROACH FOR CHIKV VACCINE

With the widespread distribution of CHIKV, there is a need for a safe and effective vaccine. Currently, there is no effective vaccine for CHIKV although there have been some efforts on this approach. An understanding of antibody-mediated and cell-mediated immune responses is important for vaccine development.[5] Unfortunately, there is very little information pertaining to the interaction of CHIKV infection and adaptive immune system in re-infection.[5, 8] Some potential vaccines which need to be developed are listed in Table 1.3.[122]

Table 1.3. Potential vaccines for CHIKV.

Type of vaccine	Vaccine	Production	Developed year	Status
Inactivated vaccines	Formalin-inactivated vaccine[123]	A whole virus grown in monkey cell cultures	1970s	Phase II: discontinue
	Tween 80-ether extraction[124]	A whole virus grown in monkey cell cultures	1970s	Preclinical
	Vero adapted formalin inactivated vaccine[125]	A virus on Vero cells	2006	Preclinical
Live-attenuated vaccines	181/clone 25 vaccine strain[126]	Serial passage CHIKV strain in culture cells	1986	Phase II
	TSI-GSD-218 CHIKV vaccine[127]	An attenuated strain CHIKV	2000	Underway Phase III
Genetically engineered vaccines	Chimeric virus vaccine[128]	VEEV (attenuated vaccine strain TC-83), an attenuated strain of EEEV or SINV and the structural protein genes of CHIKV	2008	Preclinical
	Recombinant adenovirus vaccine[129]	A non-replicating complex adenovirus vector encoding the structural polyprotein cassette of the CHIKV	2010	Preclinical
	CHIKV-IRES vaccine[130-131]	A cDNA clone generated from the wild-type La Reunion strain	2010	Preclinical
	DNA vaccines[132-133]	Single or three individual plasmids	2008	Preclinical
	Subunit protein vaccines[134-137]	Recombinant CHIKV envelope proteins	2012	Preclinical
	Virus-like particle vaccine[138-139]	Non-infectious virus-like particles coated with the same protein	2010	Completed Phase I

1.2.3 ALTERNATE APPROACHES

Some other approaches are targeting virus entry and maturation, viral nucleic acids, and cellular receptors. For example, agents used on each interference can be listed as follows: using furin inhibitors, or decanoyl-RVKR-cholormethyl ketone to impair the maturation of the E2 glycoprotein;[9, 140] small hairpin RNA molecules to interfere RNAs;[141] 2',5'-oligoadenylate synthetase (OAS3),[142-143] cellular IMPDH enzyme,[114] and visperin[144] in controlling CHIKV replication or human antibodies. Recently, silencing of HSP-90 (a chaperone protein related to heat shock) using small interfering RNA (siRNA) has been shown to disturb CHIKV replication in cultured cells.[145]

1.3 CURRENT RESEARCH IN CHIKV DRUG DISCOVERY

1.3.1 CELL-BASED HIGH THROUGHPUT SCREENING APPROACHES

Recently, high throughput screening methods have been developed to identify potential CHIKV inhibitors. For example, a CHIKV replicon and a concomitant screen with SFV surrogate infection model were used to screen 356 natural compounds, and clinically approved drugs.[146] A cell-based high throughput screening assay using resazurin against a kinase inhibitor library of 4,000 compounds, combined with the image-based high content assay approach was applied.[147] A phenotypic assay was also used to identify one natural compound that partially blocks nsP2 activity and inhibits CHIKV replication *in vitro*.[148]

1.3.2 *IN SILICO* APPROACHES

In recent years, *in silico* approaches are promising to save time and the cost of the drug discovery process. With the availability of the crystal structures of several proteins of the CHIKV genome and other related alphaviruses, structure-based approaches using molecular docking and virtual library screening can be applied to identify potential inhibitors (hits) from a large database of compounds. These hit compounds may be experimentally tested, or used to investigate a structure-activity relationship to optimize

the compound's activity. Generally, drug targets are key proteins which is identified through cellular and protein biochemical processes associated with the disease. These biomolecules are known as being involved in signalling or metabolic pathways that are specific to the disease process[149-150] For CHIKV, non-structural proteins and envelope glycoproteins were considered potential targets.[8] Non-structural proteins of the virus play an integral role in viral replication and transcription, so they are attractive targets for designing potent inhibitors of CHIKV.[151] Recently, two crystal structures, namely the nsP2 protease (PDB id: 3TRK) and the nsP3 macrodomain (PDB id: 3GPG)[73] have been available for use as a starting point in antiviral research. Homology models for structure nsP4 was also proposed to provide structures for drug design.[152] The envelope glycoprotein complexes were determined by X-ray crystallography, the immature form (PDB id: 3N40) and the mature form (PDB id: 3N42).[153] These complexes are also of interest as targeting the envelope glycoproteins can affect the virus entry.

1.4 STRUCTURE-BASED DRUG DISCOVERY FOR CHIKV

1.4.1 OVERVIEW OF COMPUTER-AIDED DRUG DISCOVERY

In fighting disease, drug discovery and development are very expensive and time-consuming processes.[154] The drug discovery process involves different stages such as target identification (target ID), hit identification (hit ID), lead discovery and optimization, biological testing (preclinical trials), and clinical trials.[149, 155] Therefore, utilizing computational techniques not only increases the efficiency and effectiveness of research, but also saves time and reduces costs; in particular during lead discovery and lead optimization.

The field of computer-aided drug design (CADD) was started in the 1960s and in the late 1980s up to now, this has been growing and has become an integral part of drug discovery with the development of computer hardware and software, and the increasing availability of protein structures of biochemical targets of pharmaceutical interest (Figure 1.6).[156] There have been a number of studies reported on describing specific computational methods, clarifying their role and importance, and highlighting recent

advances, as well as their successful applications and challenges.[156-169] These contain many applications of this approach in antiviral drug design, in particular in analysis of viral protein target. The prediction and simulation of conformational, steric, and physicochemical properties helps to elucidate the characteristics of the target. The binding pockets and interactions may be identified using computational approaches. Rationalization of drug action and virtual screening to identify potential inhibitors, together with the structure-activity relationship of these ligands, were applied.[163]

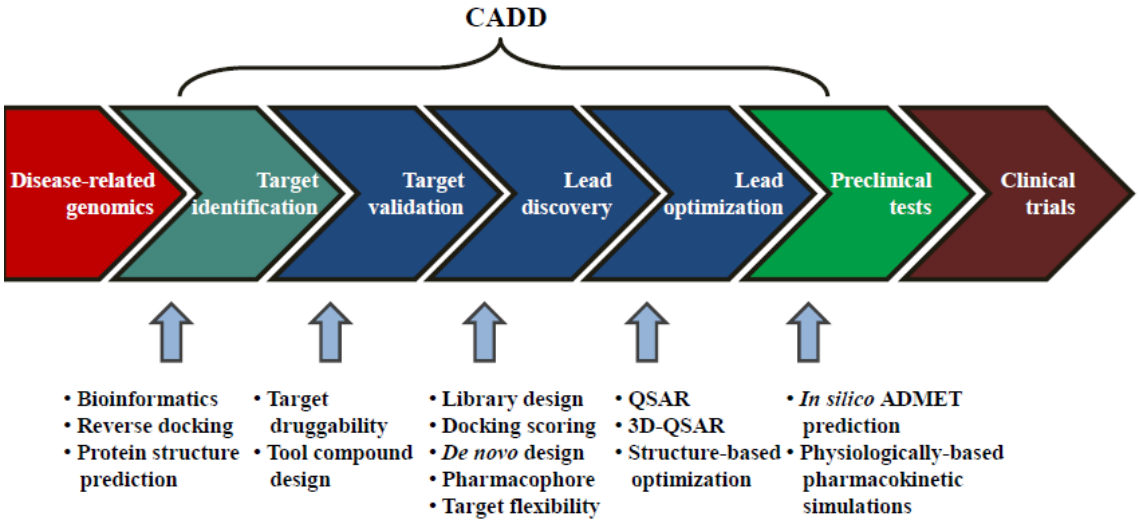


Figure 1.6. Applications of CADD to the various stages of drug development (adapted from Tang Y. *et al.*[170]).

Computational approaches in CADD can be classified into two main groups: the ligand-based approaches and the structure-based approaches.

Ligand-based approaches: This is an indirect approach which is useful in the case of a lack of an experimental receptor structure together with the difficulty of designing a reliable model.[171] Ligand-based methods are particularly valuable in the early stage drug discovery.[163] These approaches rely on known active compounds and utilize their similar descriptors taken from their molecular characteristics, properties, and biological activity data to elucidate the structural and physicochemical properties of the ligands.[163] Molecular characteristics may be physicochemical descriptors in one dimension, 1D (e.g., molecular weight, atom counts, logP, and pKa); two dimensions, 2D (e.g., topological descriptors); and three dimensions, 3D (physicochemical properties such as location, constraints, and shape descriptors). Ligand-based approaches introduce a definition of a pharmacophore as the ensemble of steric and electronic features,

necessary to account for the common molecular interactions with the protein target, and to trigger (or to block) its biological response.[154, 163] A pharmacophore model describes the three-dimensional chemical features, using pharmacophoric descriptors such as hydrogen bond donors, hydrogen bond acceptors, hydrophobic, aromatic, positive ionizable groups, and negative ionizable groups. There are many different ways to build a pharmacophore model. It can be based on chemical structures of known active compounds from different chemical scaffolds or diverse chemical structures for compounds (with IC_{50} or K_i values ranging over more than three orders of magnitude). Similarities of molecular properties such as pharmacophore features, shaped-based models or a quantitative structure-activity relationship are also utilized. The underlying assumption here is that ligands having similar physiochemical properties are likely to show comparable activity spectra.

Structure-based approaches: Unlike the ligand-based methods, structural data of a protein target is a prerequisite in a structure-based approach.[166] The full three-dimensional (3D) structure of the protein may be obtained from X-ray crystallography or nuclear magnetic resonance (NMR) techniques. In the case where the structure of protein is unknown, homology modeling can be used based on the similarity of the genomic sequences with other viruses.[150] Protein-ligand docking and structure-based virtual screening are examples of these approaches. They utilize information of the protein structure to identify and optimize drug candidates by examining molecular interactions between ligands and target macromolecules, as detailed below.[154]

1.4.2 PROTEIN-LIGAND DOCKING ASSISTS IN DRUG DISCOVERY

Pioneered in the early 1980s,[172] protein-ligand docking has developed as an integral part of drug discovery.[173] It has become an invaluable tool that assists efficiency in drug discovery to understand interactions of protein-ligand complexes.[174-176] In the 2000s, a series of reviews summarized the methodology, and highlighted the successful applications, recent advances, and challenges of this approach.[162, 166-167, 173, 177-187]

Given a small molecule and a protein target of a virus, docking attempts to insert the small molecule into a binding site of the protein target. In other words, the aim of

docking is to get the "best match" of a protein-ligand complex,[178] or to accurately predict the orientation and conformation of a small molecule (ligand), i.e. the lowest binding energy (known as binding mode or binding pose), and then estimate binding affinity of the ligand into a known structure (Figure 1.7).

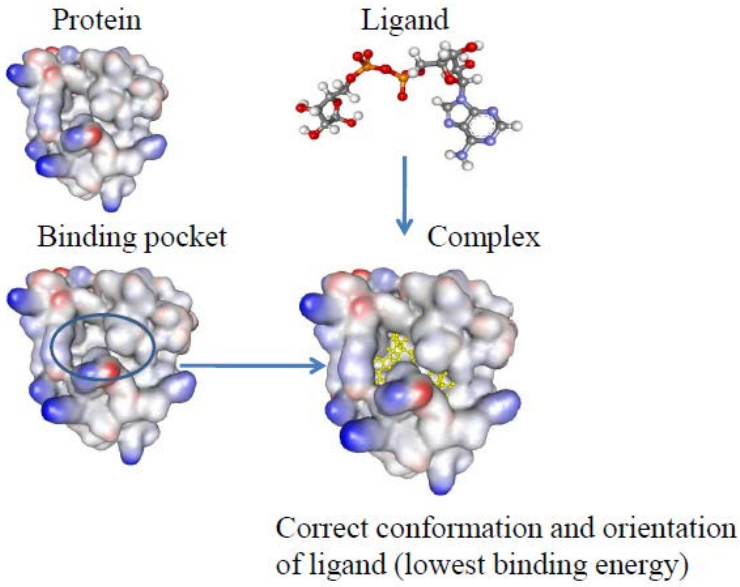


Figure 1.7. Representation of the strategy used for protein-ligand docking.

Therefore, protein-ligand docking includes two processes: docking (geometric sampling of potential ligand/protein binding mode); and scoring (using an equation and specific parameters to estimate a ligand binding affinity). Consequently, a protein-ligand docking program consists of these two essential components: sampling (ligand sampling and protein flexibility) and scoring. There are three main kinds of sampling setups: first, the protein and ligand are kept rigid, only translational and rotational manipulations of the ligand are investigated; second, the protein is rigid and ligand is assumed to be flexible (all degrees of freedom of the ligand are explored); and third, the protein is fully or partly flexible as well as the ligand. Some popular docking programs are DOCK[188], AutoDock,[189] FlexX,[190] GOLD,[191] and GLIDE;[192] most of which utilise the flexible ligand and a rigid receptor approach.

To make protein-ligand docking more practical, improvements from docking with rigid structure to partly or full flexible receptor have been undertaken during the last years.[173, 176-179, 183, 186] However, protein flexibility is still the major hurdle where ligand sampling

is concerned. In addition, there has been no universally applicable scoring function available. Therefore, finding the most appropriate algorithm for the prediction of protein-ligand binding of each specific target is a major focus of research. Many search algorithms and scoring functions have been developed over the years, alongside numerous scoring functions, taking into consideration speed and accuracy.[173]

1.4.2.1 Search algorithms

As previously mentioned, most docking programs account for ligand flexibility. In order to search for precise conformation and configuration, an optimal search algorithm should explore all of degrees of freedom to sample sufficiently so as to include the true binding modes. These algorithms may be classified into three main groups: systematic searches (incremental construction, conformational search, exhaustive or databases), random or stochastic methods (Monte Carlo, genetic algorithms or evolutionary algorithm, Tabu search), and deterministic or simulation searches (molecular dynamics and energy minimization).[173, 176-177, 179, 186] Some algorithms implement a hybrid approach combining two or all three of the methods.

Systematic search: In this search algorithm, all of the degrees of freedom of ligand are investigated. It can systematically rotate all rotatable bonds of a ligand through 360° using a fixed increment, to obtain all possible combinations for evaluation. This approach is, however, succeptable to a combinatorial explosion in computational cost. Another approach is that the ligand is divided into different fragments or a core fragment and flexible parts (sidechains), and then these components are docked into the active site. The results will be covalently linked together (other incremental search algorithms) or rebuilt the ligand from fragments (Hammerhead algorithm). Libraries of pre-generated conformations may be utilized to tackle the combinatorial explosion problem.

Random (or stochastic) search: The search makes random changes for a single ligand or a population of ligands. A pre-defined probability function is used to evaluate the ligand. The common groups are Monte Carlo (MC) and genetic algorithms. In the MC method, following random changes and energy minimization for the generated conformations, the Boltzmann probability function is used as the criteria of evaluation.

If a change in temperature is also combined to increase the probability, this is called Simulated annealing. In the genetic algorithms or evolutionary algorithms, the theory of the evolution in biological system is applied to search for the correct ligand binding mode.[162] The Tabu search is another method to explore areas of conformational space, and the value of root mean square deviation (RMSD) between the calculated molecular coordinates, and every molecule's previously recorded conformation should be less than a cutoff value to be accepted. Swam optimization tries to search for a search space for the ligand using Swam Intelligent method by taking the information of the best positions of its neighbors.[162]

Deterministic (simulation) method: These algorithms use molecular dynamics (MD) simulation or energy minimization to explore the conformations. The ensembles of populations are generated and are docked rather than a single conformer. MD is the most popular approach, however has weaknesses in its inability to cross high-energy barriers within simulation timescale, leading to arrest of ligands in a local minima of the free energy surface.[177] To overcome this, simulations can be carried out for the different parts of a protein-ligand system at different temperatures, or ligand starting positions. Energy minimization is often used in combination with other approaches.

The search algorithms for protein flexibility in docking are still limited, though there are some reviews and studies on it.[193-197] MD and MC, are usually applied for a search of protein conformations. Another strategy is the use of rotamer libraries to model protein conformational space based on experimental data, and favourable sidechain conformations. A protein ensemble grid is different approach in which the algorithm uses an ensemble of conformations of protein for docking rather than a single one, and then maps them on a grid representation.

1.4.2.2 Scoring function

For a docking process to be successful, the adopted scoring function is the deterministic factor for obtaining an accurate prediction of conformation of a protein-ligand complex, and a correct ranking of final structures. In other words, scoring is used to predict the binding affinities, and differentiate between correct and incorrect orientations, and

conformations (poses). The ideal docking program should satisfy both computational efficiency and reliability. Therefore, speed and accuracy are two key components of a scoring function. Free-energy simulation can be based on atomistic MD simulation, applied for quantitative modeling of protein-ligand interactions, and binding affinity predictions; however, this approach is expensive. Most docking programs still do not include entropic effects explicitly. There are different scoring functions which are classified in terms of shape, chemical complementation, force fields, empirical results or system knowledge. The three main categories are force-field-based, empirical, and knowledge-based scoring functions.[162, 166, 173, 176-177, 179, 181-182, 186, 198-199] Beside these scoring functions, some other approaches to improve scoring such as consensus scoring (combining the information from multiple scoring functions), or clustering, and entropy-based scoring methods.[162]

Force-field-based scoring: This type of scoring function tries to model many types of interactions involved in protein-ligand binding by utilizing physics-based functional forms, and parameters derived from experiments or quantum mechanical simulations.[166] There are different types of force field scoring functions. They have similar functional forms but their force field parameters are different, for example G-Score with Tripos force field,[200] and AutoDock with AMBER force field.[176] Most of them only consider a single protein conformation to make it possible to omit calculation of the internal protein energy. In molecular mechanics force fields, the binding free energy is calculated as a sum of two energies, receptor-ligand interaction energy (van der Waals (vdW) and electrostatic interactions) and internal ligand energy (steric strain induced by binding). In AMBER force fields, the vdW term is referred to a Lennard-Jones potential function (such as 12-6 Lennard-Jones).[166] The general AMBER force field (GAFF) including some parameters is suitable for simulating small molecules.[201] The CHARMM force field is similar to AMBER force fields, but has some additional terms. The CHARMM22 force field is is usually used for modelling protein. The recent CGenFF force field can be applied as a general force field for small molecules.[202] The disadvantage of standard force-field-based scoring function is that it does not include solvation, and entropic term explicitly.[176]

Empirical scoring: This type of scoring uses a sum of various empirical energies to estimate the binding energy. A set of weighted empirical energy terms may be composed of vdW, electrostatic, hydrogen bonding energy, desolvation term, entropy, and hydrophobicity term.[162] This scoring is far simpler than force-field-based scoring, however, it depends on the molecular data sets used to perform regression and fitting, leading to reproducibility of the experimental binding affinity data.

Knowledge-based scoring: The binding affinity can be calculated based on the information of experimental protein-ligand complexes data. This is estimated as a sum of free energies of protein-ligand atom-pair interactions (the potential mean force). The frequencies or probability distributions of interatomic distances between two atoms are converted into distance-dependent interaction free energies of protein-ligand atom pairs using the inverse Boltzmann method.[162]

Some popular docking programs together with their searching algorithms and their scoring functions are summarized in Table 1.4.

In conclusion, there are many docking programs with different scoring functions. However, no single one is the best. The ideal scoring function is still not available. The problem is that all of the scoring functions mentioned above, are based on different assumptions and simplifications, and do not fully take into account of entropic and solvation effects. Compromises between the conformational searching algorithms and scoring functions could improve docking algorithms, however, it may not improve binding affinity prediction. Thus, it is difficult to obtain a chemical accuracy, and the results depend on the specific system.

Table 1.4. Docking programs with corresponding search algorithm and scoring function.

Docking program	Fees	Ligand searching algorithm	Scoring function
DOCK[188]	No (academics); Yes (profit)	Incremental build	Force field or contact score
AutoDock[189]	No (academics); Yes (profit)	Lamarckian algorithm	Force field
AutoDock Vina[203]	No (academics); Yes (profit)	Lamarckian algorithm	Combining knowledge-based potentials and empirical scoring functions
FlexX[190]	Yes	Fragmentation and Incremental construction	Empirical score
GOLD[191]	Yes	Genetic algorithm	Empirical score
Glide[192]	Yes	Exhaustive search (Monte Carlo)	Empirical score
FRED[204]	Yes	Conformational ensembles (Rigid body docking)	Gaussian score or empirical scores
ICM[205]	Yes	Pseudo-Brownian sampling and local minimization (Metropolis Monte Carlo)	Mixed force field and empirical score

1.4.2.3 Setting up a docking protocol

To perform docking, preparations of protein and ligand are indispensable. For the target protein, the structure is usually obtained by X-ray crystallography or by NMR structure determination. The structure may be apo, holo (complexed with another compound) or if the structure is not available, it may be predicted by threading or homology modeling. If the function of protein is unknown, it is crucial to search for possible binding sites in the structure (discussed below). Some crucial factors should be considered to check the structure carefully. Initially, structural integrity is needed to check. Polar hydrogen atoms are also added. The next step is to assign proper protonation and tautomeric states of ionizable residues including aspartate (Asp), glutamic acid (Glu), arginine (Arg), lysine (Lys), and histidine (His). The orientation of asparagine (Asn) and glutamine (Gln) residues also require checking. The active site or binding site needs to be defined before docking, and the treatment of water molecules in docking is considered. Water molecules may affect the formation of the complex as they can form hydrogen bonds with the ligand and the protein. Geometry refinement of the protein/receptor-ligand is required to correct small artefacts in the protein or ligand such as strained bond

lengths/angles or intermolecular steric clashes.[187] Energy minimization of the structure should be used. Preperation of the ligand, like the protein, requires great care. Ligands can be taken from various sources, such as a database or electronic vendor catalogues. Filtering using the drug-like properties is recommended to eliminate molecules, with unfavourable properties: such as poor solubility, pharmacokinetics characteristics relating to adsorption, distribution, metabolism and excretion; oral bioavailability and toxicity. The most popular filter used in drug design is Lipinski "rule of five"[206] that is, molecular weight (MW) lower than 500 daltons, logP less than 5, number of hydrogen bond donors less than 5 and number of hydrogen bond acceptors less than 10.[207] Usual, compounds are represented in 2D format and then converted to a 3D representation. Several methods are available for generation of a 3D structure. Special care must take into account of tautomeric and protonation states, stereochemistry of chiral centres. The next step is docking with sampling to generate different binding poses; scoring and ranking. Most docking programs as mentioned previously, focus on docking with rigid docking and flexible ligands. Protein flexibility may be treated for some specific sidechains of protein residues.

1.4.2.4 Prediction of binding sites

In structure-based drug design with molecular docking and virtual screening approaches, the question of where in the structure of the protein the ligand binds is of interest. Therefore, it is a requirement to understand the structure and function of a protein target, in particular, knowledge of locations that small molecules (ligands) could bind in the structure is of key importance to help rationally design of ligands, to fit with high binding affinities and specificity that can modulate the functions of this target.[208-210] A binding process is the sum of many contributions such as environment (pH, ionic strength), and presence of water molecules, in which the shape complementarity of protein and ligand together with their physicochemical properties are balanced.[209] Proteins can change conformations upon ligand binding, which may influence the steric accessibility of a binding cleft and can interfere with the ability of an algorithm to identify a potential binding site.[208] Small molecules are known to usually bind in the largest pockets on the surface of protein, and their binding free energies is a result of enthalpy-entropy compensation.[209]

There have been many different approaches developed to identify binding sites of a protein, depending on whether the co-crystallised structure of a protein-ligand complex is known or not.[209] Thus, these approaches can use a co-crystallisation of a protein with a ligand if available, use structural or sequence similarity with a known binding site, or use *in silico* prediction methods. For computational methods, some algorithms utilize protein surface analysis to identify ligand binding pockets in the protein. The others use probe clustering and energy contour analysis which is based on analysis of binding energies of probes placed on a grid around the protein.[210] Other types of characteristics such as surface accessibility, the net charge on the protein residues in a protein as a function of pH, and sequence conservation can be used. MD simulations are also used to generate dynamic ensembles of protein conformations for binding site detection. In general, computational approaches for prediction of binding sites on a protein can be classified into three main groups: geometric-based (using geometrical pocket description or free accessible volume calculation to identify the sterically favoured cavities among all clefts on the protein's surface); energetic-based (finding energetically favoured positions, e.g. by calculation of interactions with different probes and the protein's surface); and knowledge-based (using structure and sequence comparison to generate templates of similar sequence or functionality to identify evolutionarily conserved regions to rank cavities generated with geometric approaches). The geometric-based methods are fast and easy to use while the others are more time-consuming and require user expertise.[209]

1.4.2.5 Structure-based virtual screening

The most notable application of protein-ligand docking is structure-based virtual screening.[211] It has become increasingly important to improve the speed and efficiency of drug discovery and development.[211-215] The main purpose is to reduce the large number of compounds from databases and select the most promising compounds for biological testing.[176] It is known as a fast tool for drug design which is valuable to discover lead compounds complementary to experimental methods, for example high-throughput screening. A large volume of studies have shown its importance and successful application in this field of drug design and development.[211-224] Some popular programs available for this purpose are DOCK,[188] FlexX,[190] GOLD,[191, 225] AutoDock[189,

[226] and AutoDock Vina.[203, 227] However, based on docking, virtual screening has similar challenges in accuracy of scoring and ranking. If many ligands are docked, the time for scoring and ranking in screening a large sample is of utmost important. More reliable scoring functions should include calculations for nonbonded interactions such as cation-π interaction, CH-π interaction, and π-π stacking interactions.[224] The choice of ligand libraries for screening is also of concern. Several databases are available, for instance, the National Cancer Institute (NCI) Diversity Set (dtp.nci.nih.gov/branches/dscb/diversity_explanation.html), NCBI PubChem (pub-chem.ncbi.nlm.nih.gov), eMolecules (www.emolecules.com) and ZINC[228] composed of a variety of compounds or libraries of natural products,[229] metabolome,[230-231] and nutraceuticals.[232] If in the case of a new target, where no information about binding sites for ligands is known, blind docking for the entire protein is applied to identify sites that ligands bind tightly. Other computational approaches can be used to predict the binding sites as discussed previously. Validation of the approaches is the most critical due to the inaccuracies of the scoring functions, which will affect the results of ranking. The main factors should be taken into account are the quality of the obtained docking poses, and the ability of the methods to discriminate known active and inactive compounds after docking in the same target. False-positive hits or decoy molecules have similar physical structures but are chemically distinct from 'true' hits, and can be used as competitive binders to a protein.[233-235]

1.4.2.6 Validation of docking method

Docking is usually validated by the ability to reproduce the experimental data in predicting the binding pose, and binding affinity to distinguish the active and inactive compounds. The commonly accepted criterion for docking success is pose selection by a comparison of the RMSD between the docked structure (top scoring pose) and the experimental structure (the co-crystal structure). With the protein-ligand co-crystal structure, a docking protocol can be evaluated and improved. However, the errors exiting in the bound structures due to the poorly refined ligand geometries could lead to misleading interpretation of key binding interactions. They can be errors in the ligand structure (for instance missing atoms, incorrect bond orders or other connectivity issues); or incorrect bond distances, angles or dihedral angles; or conformational errors

(for example, cis- or twisted amides, distorted rings, non planar aromatic groups, or groups of planar structures of not being planar); incorrect orientations or bad steric clashes between protein and ligand. Therefore, it has become commonly to apply protein preparation tools to structure.[236] If the RMSD value is no more than 2 Å, it is considered a docking success.[167, 179, 184, 237-238] Moreover, other criteria to validate docking include Generally applicable replacement for RMSD (GARD),[238] enrichment factors, receiver operating characteristic (ROC) factors,[227] area under ROC curves (ROC AUCs),[239] and binding affinity calculations.[240] Using a decoy set of inactive compounds to dock and after ranking by scoring, enrichment is calculated by dividing the ratio of active molecules in the entire dataset by the ratio of active molecules in the top 5 or 10% of the top ranked molecules. The enrichment plot or ROC curves are plotted. The sensitivity of a given docking/scoring combination and specificity are shown in ROC plots.

1.4.3 MOLECULAR DYNAMICS SIMULATIONS IN DRUG DISCOVERY

Given the structure of a macromolecule and its complexes, molecular dynamics (MD) simulation is one of the computational approaches to investigate the motions of a system of particles.[241] MD simulation of biological macromolecules (proteins) of interest was introduced in 1977.[242] Since then, this technique has become a powerful computational method, quickly showing wide application to many fields, and constant improvement quantitatively and qualitatively in structure-based drug discovery to understand drug-receptor interactions.[243-246] Especially with the recent advances in algorithms and computer hardware and software, long-timescale from microsecond to millisecond for thousands of atoms has been achieved.[247] The interaction for the particles is calculated and included sequentially to the time of simulating. The result of the process (MD trajectory) provides information on the atomic level of positions, velocities and energies. Statistical mechanics related to distribution and motion of atoms are required to connect microscopic simulations with macroscopic observables such as changes of conformations, binding free energy, and mechanism of reaction. MD simulation has been applied to a wide range of biological problems, including protein folding,[248-249] protein-ligand interactions,[250] electron transfer state in photosynthesis,[251] enzyme reactions,[252-253] determination of protein structures from NMR,[254] refinement of

protein X-ray crystal structures,[255] and calculation of free energy changes from mutations in proteins.[245-246, 252, 254, 256]

Classical all-atom MD simulations use a Newtonian equation of motion ($F = ma$ (Equation 1.1), with F is the force on the particle, m is its mass, and a is its acceleration) to simulate the movement of each atom in the system. The MD method starts with the initial set of coordinates obtained from X-ray crystal structures, NMR structures or theoretical models (homology models), or a combination of these. Given the atomic model, the interactions for all atoms must be defined. The Newtonian equation is subsequently applied to calculate the force after the systems are minimized to eliminate high energy interactions, such as steric clash, prior to simulation. The integration of the equations of motion after equilibration generates an ensemble of equilibriated states, including coordinates and velocities of the atoms as a function of time. The simulations require three components: initial coordinates (obtained from experimental structures or from models, or some combination of both), a potential (obtained from a force field and the coordinates), and algorithms for propagation.[244] The process of how an MD simulations proceeds is set out in Figure 1.8.

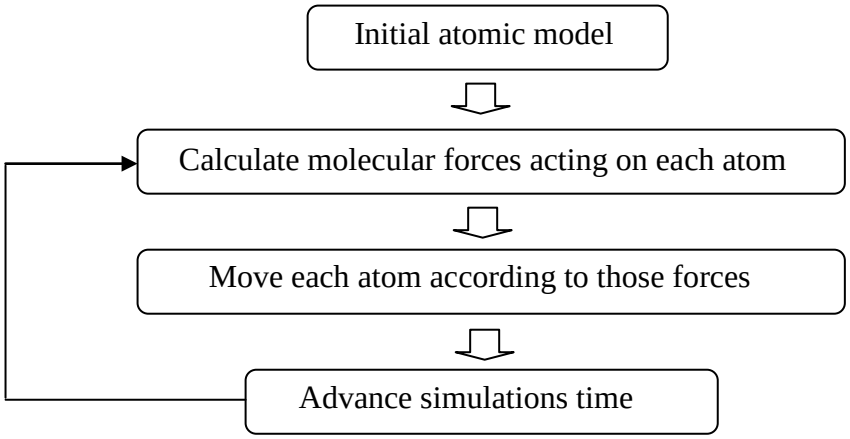


Figure 1.8. How MD simulations proceed (adapted from Durrant D. *et al*[246]).

1.4.3.1 Force fields

The mathematical functions describing the potential energy of a system and their related parameters are called a “force field”, which is set to describe the interactions between atoms and molecules. A typical molecular CHARMM force field for a molecular system

is estimated as follows in Equation 1.2. In brief, a molecular force field usually include six terms; namely the bond, angles, dihedral, improper dihedral angles, nonbonded, and Urey-Bradley (UB) terms.

$$U(\{\vec{R}\}) = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\emptyset[1 + \cos(n\emptyset - \delta)] + \sum_{\text{impropers}} K_\varphi(\varphi - \varphi_0)^2 + \sum_{\text{UB}} K_{UB}(r_{1,3} - r_{1,3;0})^2 + \sum_{\text{nonbonded}} \varepsilon_{ij}\left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min,ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi D r_{i,j}}$$

Equation 1.2. Equation used to calculate the atomic forces in MD simulations, from MacKerell A. D. *et al*,[202, 257] where K_b, K_θ, K_χ, K_φ and K_{UB} are the bond, angles, dihedrals angles, improper dihedral angles and Urey-Bradley force constants, respectively. b, θ, ∅, φ, $r_{1,3}$ are the bond length, bond angle, dihedral angle, improper torsion angle, and the Urey-Bradley 1,3-distance, respectively; with the subscript zero means the equilibrium values. The Urey-Bradley is a harmonic term, used for bond-stretching or angle bending in distance between atoms 1 and 3 (the two terminal atoms) in an angle. It is important to more accurately model in vibration spectra.[258] n and δ are the values of dihedral multiplicity and phase. The nonbonded includes the van der Waalsand electrostatic interactions. The van der Waals interaction is using Lennard-Jones 6-12 potential with $R_{min,ij}$ is the radius in the Lennard-Jones term; q_i and q_j are the partial atomic charge of atom i and j, respectively.ε_{ij} is the effective dielectricconstant, and r_{ij} is the distance between atoms i and j.

The different force fields have a critical influence on the results of MD simulations.[245] The most common currently used force fields are AMBER,[259] CHARMM,[257] GROMOS[260] and OPLS.[261] The choice of force field usually depends on the preference of the molecular simulation suite.[244] Among of the most commonly-used simulation packages are the AMBER,[262] CHARMM,[263] GROMACS[264] and NAMD.[265] These programs share common basic features, but differences lie in their capacities, and underlying philosophies.[244]

1.4.3.2 Setting up and running MD simulations

To set up and run a MD simulation, some important ingredients should be considered carefully: such as the initial atomic coordinates of the system, the choice of force field, simulation program (integration method), time steps, type of ensemble, boundary conditions, salvation, and time length of the simulation.[266-267] The length of simulation

time depends on the investigated system and the aim of the study. The integration method will decide parameter of time steps. Any change to these factors will affect the outcomes of the simulation, as well as the requirements of computational time.[268] In addition, the type of ensemble must be selected. Traditional MD for biomolecular systems often use NTP (N, number of particles, T, temperature, and P, pressure) or isothermal-isobaric ensemble, which include the constant NTP. Furthermore, in MD simulations, treatment of electrostatic interactions (long range coulombic forces) is very important.[267] Nowadays, one can use the Ewald summation method, such as the Particle-Mesh Ewald (PME) for electrostatics.[269-270]

Moreover, boundary conditions and solvent models are required. Due to the influence of interactions at the boundaries of the system on energy calculations, the boundaries must be taken into account. The most common way in simulation of biomolecular systems is using periodic boundary conditions. Periodic boundaries cover the system in a cell (typically a cubic box or a sphere or other geometric shape), and surround it with mirror replica cells of the system. The size of the box must be large enough so that the molecule does not "see" itself.[267] The interaction energies can be calculated across the cell boundaries, so the boundary effect is minimized. Two types of solvent models are commonly used, namely implicit (or continuum) and explicit models. The implicit solvation means the solvent is represented with a continuum medium. Two common algorithms used to calculate the solvent electrostatic effects are the Poisson Boltzmann surface area (PBSA) and the Generalized Born surface area (GBSA) model.[244] These models are less expensive, but the major problem is that they do not take into account entropic effects, and their difficulty in dealing with heterogeneous environment, which affect many biological processes. On the other hand, the explicit solvation is the model in which the solvent molecules and counterions are treated as explicitly surrounding the biomolecule.[244] This method is the most accurate, however it is also time-consuming. Among different water models developed, TIP3P water model is adopted for the CHARMM force field.[271-272]

In general, a MD simulation consists of minimization, equilibration, and production steps. The initial system is minimized to relax potential steric clashes in the structure. The minimized structure is then equilibrated. When equilibration is reached, the values

of average temperature, pressure, and energies are stabilized. The production phase will be the next step and used to calculate the desired properties.

1.4.3.3 Analysis of a MD simulation

The results of MD simulations are analyzed using trajectories to gain insights into the structure; such as protein stability and flexibility. The frequently used analyses are global measures by calculating a RMSD[273] and a root mean square fluctuation (RMSF) or B-factors;[274] secondary structure analysis, hydrogen bonds (HBs), and hydrophobic contacts; clustering analysis, quasi-harmonic, and principal component analyses/correlation function, and binding free energy calculation.[244]

RMSD is usually used to investigate the global stability of the system,[275] while RMSF is used to obtain the local structural flexibility and stability. They are calculated as following equations:

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i^m - x_i^l\right)^2 + \left(y_i^m - y_i^l\right)^2 + \left(z_i^m - z_i^l\right)^2}$$

Equation 1.3. The RMSD between atoms of the trajectory frames and the corresponding atoms of the initial structure, where N is the number of atoms, x^m, y^m, z^m are the Cartesian coordinates of the initial structure and x^l, y^l, z^l are the Cartesian coordinates of trajectory at frame t.[275]

$$\mathrm{RMSF} = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(x_i - \bar{x})^2}$$

Equation 1.4. The RMSF of an atom, where T is the number of trajectory frames, and $\bar{x}$ is the time-averaged position.[275]

The resulting trajectories are analyzed by the CHARMM22[263] and VMD (version 1.9.1).[276] A simple geometry criterion was used to define a hydrogen bond: the distance between proton donor (D) and acceptor (A) atoms less than 3.5 Å, and the angle $D - H \cdots A$ greater than 120°.[70] If the percentage of HBs occupation is higher than 50%, they are considered as the medium, whereas the strong HBs are determined by HBs occupations of greater than 75%.[70] The hydrophobic contacts between the carbon atoms of non-polar parts of residues of proteins were also monitored with a cutoff distance of

4.0 Å.[277-278] Clustering analysis was conducted using the Clustering Plugin in VMD for all of the snapshots from the trajectories.

Clustering analysis involves the grouping of similar samples of data, joining the ensembles of data into the group to identify the most populated conformations sampled. Structural clustering is a useful method to reduce the sample size for conformational analysis to understand the molecular motion within conformational space.[279] Principal-component analysis uses a constructed matrix of atomic fluctuations to find the lowest modes, which represent the most of fluctuations. Quasi-harmonic analysis gives normal modes in the harmonic system, while correlations functions are used to measure the correlation of two fluctuating quantities over the time.[244]

1.4.3.4 Combining molecular docking and MD simulations

Docking and MD simulations have their own strengths and weaknesses, listed in Table 1.5. Some reviews highlighted the improvement of computational protocols using a combination of MD simulations in docking procedures.[280-281] MD simulations before docking may explore the conformational space of the protein receptor, while using MD simulations after docking helps to optimize the final structures, analyze protein flexibility, and stability of different complexes, account for solvent effects, and obtain accurate energetic properties.[167]

Table 1.5. Strengths and weaknesses for docking and MD simulations[280-281]

	Docking	**MD simulations**
Strengths	Fast and inexpensive: docking explore conformation space of ligands in a short time, allowing the scrutiny of large libraries of drug-like compounds at a reasonable cost.	Accurate because it treats both ligand and protein in a flexible way, allowing for an induced fit of the receptor-ligand. The effect of explicit water molecules can be studied directly.
Weaknesses	Lack of or poor flexibility of the protein. Absence of a unique and widely applicable scoring function, necessary to generate a reliable ranking of the final complexes.	Costly and time-consuming. The system can get trapped in local minima.

1.4.4 BINDING FREE ENERGY CALCULATIONS

Given a protein (P) and a ligand (L), the ligand can bind to a protein to form a complex PL. The binding affinity of PL in an equilibrium concentration can be computed simply by using the following equation:

$$K = \frac{[L][P]}{[PL]} \qquad \text{(Equation 1.5)}$$

This binding affinity can then be related to the free energy of binding (free energy change to changes in enthalpy and entropy) using:

$$\Delta G_{bind} = -RT\ln K_d = \Delta H - T\Delta S \quad \text{(Equation 1.6)}$$

Where ΔG_{bind} is the change in free energy of a binding process, ΔH and ΔS are the corresponding changes in enthalpy and entropy, respectively. R is a gas constant with $R=8.314\ JK^{-1}mol^{-1}$, and T is the temperature of the system in Kelvin degree, and K_d is dissociation constant.

The binding free energies (or binding affinities) are used as a criterion for differentiation of inhibitors from other small molecules (binders), and also selection of strong ligands based on their protein binding strengths. It has become a major interest in structure-based drug design. The computational approaches can be used to estimate the binding free energies together with the experimental assays. However, obtaining accurate values of binding free energies remains a challenge.[282,283] Upon protein binding, the protein and ligand may be affected by conformational changes influenced by water and ions. Recently, there has been a very large number of approaches developed to solve this problem using different atomistic models.[283-298] The approaches can have the simplicity of a scoring function (docking) or the complexity and sophistication of free energy methods.[292] Most of them are still under active study, and have different trade-offs between accuracy and computational efficiency.

As mentioned above, docking and scoring use a single bound conformation containing a simplified energy model, such as an empirical force field, with a simple solvent model. Thus, it can provide approximate binding affinities (scores) of ligand and protein. However, the docking results are system-dependent.[292] The starting configuration of the protein–ligand complex is also considered.[299] Some attempts have been made to

improve docking such as taking account redistribution of ligand charges (potential energy models), solvent models, protein flexibility, and considering changes of configuration entropy.[292] In contrast to docking, free energy methods can give more accurate binding energy, but need more computational cost. These approaches generate thermodynamic averages (converged results) using a conformational sampling. They can be classified into two methods: end point and pathway methods. The end point methods require simulation of the bound and free states of the ligands to generate conformations of both states, and compute the binding free energy based on the difference between them. The pathway methods calculate binding free energy using the simulation of many immediate states to sum up all of small changes along a multistep pathway.[292] A choice of calculating absolute or relative free energies using implicit or explicit solvent, and the length of the simulation will hugely impact computational time, the accuracy, and efficiency of the calculation. Popular methods include free energy perturbation (FEP), thermodynamic integration (TI), linear interaction energy (LIE), molecular mechanics-PBSA (MM-PBSA), and molecular mechanics-GBSA (MM-GBSA), discussed below.

1.4.4.1 Free energy perturbation (FEP)

The FEP and TI approach are the theoretically rigorous methods, and offer accurate binding free energy; including the absolute binding free energy of a ligand, and relative binding energies between two ligands, X and Y, bound to the same protein (P).[300] This approach depends on a thermodynamic cycle (Figure 1.9). The perturbation theory is that different binding free energy between the first state (before binding) and final state (after binding) can be calculated by the formula $\Delta G = -RT\ln\langle e^{-\Delta U/RT}\rangle$ (Equation 1.7) in which ΔU is change of the energy function between two states, and the angle brackets are a Boltzmann average (ensemble average) obtained from MC or MD simulations.[301]

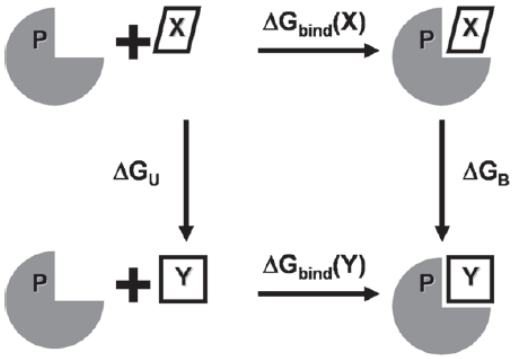


Figure 1.9. Thermodynamic cycle for calculating relative binding free energies between two ligands bound to the same protein (from Michel J. *et al*[300]).

Based on this theory, the difference of binding free energies of two ligands X and Y bound in the same protein in Figure 1.9 is calculated as follows:

$$\Delta\Delta G_{bind} = \Delta G_{bind}(Y) - \Delta G_{bind}(X) = \Delta G_B - \Delta G_U \text{ (Equation 1.8)}$$

Where $\Delta G_{bind}(X)$ and $\Delta G_{bind}(Y)$ are binding free energies for ligand X, Y, respectively; and ΔG_U and ΔG_B are unphysical transmutation free energy from ligand X to ligand Y in the unbound and bound state, respectively.

In the case where ligand X and Y are similar, the values of ΔG_U and ΔG_B are easier to obtain than $\Delta G_{bind}(X)$ and $\Delta G_{bind}(Y)$ because the mutation from ligand X to ligand Y is assumed to cause only localized changes. If ligand X and Y are too different, large changes between them may cause sampling problems. So, in the FEP method, the path of transformation is divided gradually into many small steps (intermediate) to allow smooth conversion of ligand X to Y.[300] The binding free energy difference between two states, X and Y, is the sum of the contributions from all steps. The formula to calculate the value for each step is:

$$\Delta G = -RT \sum_{i=1}^{N-1} \ln \langle \exp\left(-\frac{H(\lambda_{i+1}) - H(\lambda_i)}{RT}\right) \rangle_{\lambda_i} \text{ (Equation 1.9)}$$

Where ΔG is the free energy difference between two states, X and Y. λ_i varies from 0 (state X) to 1 (state Y); $H(\lambda_i)$ and $H(\lambda_{i+1})$ represents Hamiltonian of the system at λ_i, λ_{i+1}; and $\langle \quad \rangle_{\lambda_i}$ indicates an ensemble average. Absolute binding free energies can be obtained from FEP method by setting the interaction potential of the ligand to zero in one of the states. That means transforming the ligand into dummy atoms that do not interact with their surroundings.[250]

The FEP is currently considered the most powerful and promising approach.[293-294] It takes into account entropic contributions to binding affinities arising from solvent effects and protein/ligand flexibility. However, the FEP results could have high precision, but low accuracy.[302-303] Most of the computational time is spent on "perturbations", meaning uninteresting configurations corresponding from unphysical paths (X to Y), which makes the method difficult for application.[295]

1.4.4.2 Thermodynamic integration (TI)

Another pathway approach, similar to FEP is TI.[300] In the TI method, the difference in free energy between two states, A and B, is calculated based on using multiple intermediate states, defined by a coupling parameter λ.[304-305] The average of the derivatives of the Hamiltonian at each λ, H(λ) is calculated and then TI uses numerical integration over λ to calculate the free energy difference between two states, where λ has the same meaning as in FEP:

$$\Delta G = \int_0^1 \langle \frac{\partial H(\lambda)}{\partial \lambda} \rangle \, d\lambda \quad \text{(Equation 1.10)}$$

Where ΔG is the Gibbs energy difference between two states, ⟨ ⟩ is an ensemble average obtained at λ. The pathway of intermediates between the states of interest can be parameterized between λ=0 and λ=1.[284]

1.4.4.3 Molecular mechanics-Poisson Boltzmann surface area (MM-PBSA) or Molecular mechanics-Generalised Born surface area (MM-GBSA)

Compared to the pathway methods, the end point methods such as MM-PBSA or MM-GBSA are more computationally efficient, and widely applied for the estimation of the accurate relative binding free energies of related compounds.[250, 306-309] The methods combine the Poisson Boltzmann (PB) or Generalised Born (GB) electrostatics with molecular mechanics (MM), and solvent accessibility (SA) models, or continuum solvent approaches, to estimate binding energies.[299, 307-309] An initial MD simulation using continuum solvent approach provides a thermally average ensemble of structures. Several snapshots are then processed, removing all water and counterion molecules, and used to calculate the total binding free energy of the system with the equations:

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad \text{(Equation 1.11)}$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{elec} + \Delta E_{vdw} \quad \text{(Equation 1.12)}$$

$$\Delta G_{solv} = \Delta G_{PB/GB} + \Delta G_{SA} \quad \text{(Equation 1.13)}$$

Where ΔE_{MM} is the change of the MM energy in the gas phase, which include $\Delta E_{internal}$ (corresponding to the bond, angle, and dihedral energies), ΔE_{elec} (electrostatic energy) and ΔE_{vdw} (van der Waals energy); ΔG_{sol} is the solvation free energy which is sum of electrostatic solvation energy, $\Delta G_{PB/GB}$ (polar contribution, calculated by solvent accessible surface area (SASA)) and the non-electrostatic solvation ΔG_{SA} (nonpolar contribtion). The conformational entropy change, $-T\Delta S$, is the most difficult term to evaluate, estimated using quasi-harmonic analysis or normal mode analysis of the trajectory. The entropy change can be assumed to be cancelled if only the relative binding energies of a series of structurally similar compounds is required; however if the absolute energy is important, or if the compounds are notably different, then the contribution to the final free energy cannot be ignored.[285]

Although the MM-PBSA and MM-GBSA have shown their successful application in biochemical systems,[310-314] especially as post-docking methods in virtual screening,[308] their performances rely on the system in question. The modification in simulation protocols can affect the approach; such as the sampling strategy of generating snapshots, methods to calculate entropy, and other parameters (charges models, force fields, the solute dielectric constant, and radius parameters in continuum solvent models). In general, the MM-PBSA is more sensitive to the parameters of the systems than MM-GBSA.[309]

1.4.4.4 Linear interaction energy (LIE)

A new semi-empirical method for calculating binding free energies for ligands from MD simulations has recently been introduced.[315] This method is based on a linear approximation of polar and non-polar free energy contributions from the MD averages. The idea originated from the problem that with a diversity of compounds with “small perturbations,” it would be very difficult to calculate their binding free energies using FEP. Therefore, the absolute binding free energy of a ligand is calculated as the change

in free energy when the ligand is transferred from aqueous solution (unbound or free state) to its solvated receptor binding site (bound state).[316] In other words, two simulations are required: one with the ligand free in solution, and one with it bound to solvated receptor. The equation for binding free energy is used as follows:

$$\Delta G_{bind} = \alpha\, \Delta\langle V_{l-s}^{vdw}\rangle + \beta\, \Delta\langle V_{l-s}^{elec}\rangle + \gamma \quad \text{(Equation 1.14)}$$

Where $\Delta\ \langle\ \ \rangle$ are differences between the averages of the nonbonded van der Waals (vdW) and electrostatic (elec) interactions in the bound or unbound states, collected from MD simulation averages between the ligand and its surrounding environment (l-s). The parameters are the weight coefficients α and β for the non-polar and polar binding energy contributions, respectively; and an additional constant, γ.

Since its initial use, there have been a large number of studies showing LIE as a promising method for computation of binding free energy for protein-ligand.[285, 295, 315-330] LIE approach is considered to be a good alternative compared to other approaches such as FEP and TI as it estimates the absolute binding free energies; slower than scoring of single conformations, but faster than rigorous FEP approach.[320] When examining the general validity of the electrostatic linear response approximation, the differences in electrostatic response properties between protein and water solvent were investigated by introducing different electrostatic scaling coefficients; β_{wat} and β_{prot}, and α_{wat} and α_{prot}. The constant term γ in solvation energy was also discussed and it depends linearly on surface area.[319, 331-334] So the following LIE equation was:

$$\Delta G_{bind} = \alpha_{prot}\langle V_{l-s}^{vdw}\rangle_{bound} - \alpha_{wat}\langle V_{l-s}^{vdw}\rangle_{unbound} - \beta_{prot}\langle V_{l-s}^{elec}\rangle_{bound} - \beta_{wat}\langle V_{l-s}^{elec}\rangle_{unbound} + \gamma \quad \text{(Equation 1.15)}$$

The general equation with the values $\alpha_{prot} = \alpha_{wat}$, $\beta_{prot} = \beta_{wat}$ and $\gamma = 0$ gave a significant improvement compared to the orginial one. Aqvist and Hansson reported the relationships between electrostatic free energies and the solvation energetic for several model compounds in different solvents.[316] Some deviations from linear respone were found, in particular, for neutral dipolar solutes and for uncharged ligands having certain dipolar groups in the case of water solvent through effect of hydrogen bonding network. The optimal value of $\beta = 0.5$ was suggested to be reconsidered, for instance $\beta < 0.5$ to gain more accurate predictions. The α is an empirical constant which can be fitted to experimental binding free energies. In addition, the coefficients α and β converge to the

same values in both bound and unbound states. These findings supported the use of the basic LIE equation 1.14.

A refined LIE model, the FEP-derived model showed that the value of β varies depending on the number of hydroxyl groups. The more hydroxyl groups the compound has, the lower the value of β (Table 1.6).[316]

Table 1.6. Values of the β parameter as a function of the chemical nature of the ligand, according to Hansson *et al*.[320]

Chemical nature	**β**
Charged compounds	0.5
Neutral compounds	0.43
Neutral compounds bearing a single hydroxyl group	0.37
Neutral compounds bearing two or more hydroxyl groups	0.33

So, the final binding energy is calculated as:

$$\Delta G_{bind} = \alpha \langle V_{bound}^{vdW} - V_{unbound}^{vdW} \rangle + \beta \langle V_{bound}^{elec} - V_{unbound}^{elec} \rangle + \gamma \quad \text{(Equation 1.16)}$$

Where $\langle V_{bound}^{elec} - V_{unbound}^{elec} \rangle$ represents the averages change in electrostatic energy between the bound and unbound (free or unbound or just solvent) states, and $\langle V_{bound}^{vdW} - V_{unbound}^{vdW} \rangle$ the average change in vdW from an aqueous solution to a protein environment. α, β and γ are empirically determined constants. The α, β is for the non-polar, polar contributions, respectively; and are the same values in the bound and unbound state. [316] Applying a value of α = 0.18 has shown to successfully reproduce the experimental binding free energies in a wide variety of ligand-protein systems.[328]

β_{FEP} values can be assigned to each chemical group present in the ligand, as shown in Equation 1.17 and the values are provided in Table 1.7. The weighting factors depend on salvation energies of each chemical group. The value of w_i is 1.0 for all neutral groups or 11.0 for the anions and cations.[326] The advantage of this approach is that the β coefficient is flexible and provides higher accuracy, since deviations from the linear response due to chemical groups such as amides, amines, or carboxylic acids is explicitly taken into account.

$$\beta = \beta_0 + \frac{\sum_i w_i \Delta\beta_i}{\sum_i w_i} \quad \text{(Equation 1.17)}$$

Table 1.7. Values for the β parameter in Equation 1.17, according to Almlöf M. *et al.*[326]

Parameter	Value	Chemical nature
$\Delta\beta_i$	-0.06	Alcohols
$\Delta\beta_i$	-0.04	1°, 2° -Amines
$\Delta\beta_i$	-0.02	1° Amides
$\Delta\beta_i$	-0.03	Carboxylic acid
$\Delta\beta_i$	+0.02	Anions
$\Delta\beta_i$	+0.09	Cations

In addition, a correlation of β and the hydrophobicity of the binding site using a weighted desolvation non-polar ratio (WNDR) has been investigated.[321] The result suggested that the β is predictable by calculating the WNDR; in particular for systems in which different ligands bind to different binding sites of the same protein. The parameter γ is influenced by the hydrophobicity of the binding site.[324]

1.5 PROJECT BACKGROUND AND AIMS

In recent years, there has been an emergence or re-emergence of some alphaviruses in various countries; in particular the CHIKV. This presents a worldwide threat to human health, and creates an economic burden for the affected countries. However, there are currently no vaccines or effective drugs available for the treatment of CHIKV virus. In addition, there has been little research to find anti-CHIKV compounds. Therefore, a significant need for research into medicines to combat the virus exists. During the past decade, computational approaches have become an increasingly powerful tool in drug discovery and development. It has not only helped scientists succeed in developing many therapeutic compounds for specific diseases, but has also aided the development of time-saving and cost-effective procedures.

With this in mind, we aim to discover and develop an approach leading to the identification of a number of lead compounds to combat CHIKV disease. This study will primarily utilise computational techniques in all stages of the process. We will use all available information of CHIKV, together with a combination of computational tools, to maximize the efficiency of this study. The study has three principal objectives:

1. To identify potential inhibitors for CHIKV using a structure-based approach with molecular docking and virtual screening.
2. To investigate the stability and flexibility of protein-hit compounds complexes with molecular dynamics simulations.
3. To obtain accurate binding free energies from molecular dynamics simulations and provide guidance in rational optimization of hit compounds.

CHAPTER 2. DISCOVERY OF INHIBITORS TARGETING CHIKV NON-STRUCTURAL PROTEIN 3

2.1 INTRODUCTION

2.1.1 FUNCTION AND ROLE OF THE NSP3 OF CHIKV

The nsP3 protein is considered an attractive target for CHIKV drug discovery[72] due to its participation in the early stages of the transcription processes of viral replication, though the specific functions of the nsP3 protein remain elusive.[8, 69, 84, 87] The nsP3 is the third non-structural protein in the CHIKV genome. It consists of two domains, the N-domain and the C-domain;[70, 73] the N-domain is highly conserved, but the C-domain is not.[73] The C-domain is phosphorylated on serine and threonine residues on up to 16 positions.[73, 85, 335-337] The role of this phosphorylation is still unclear, but deletion of the residues involved in the phosphorylated process has been shown to decrease the level of RNA synthesis.[8, 73, 336] The N-domain contains the X-domain or a macrodomain, the region comprising the first 160 residues; is commonly present in eukaryotic organisms, bacteria, archea; and also many positive-strand RNA viruses such as hepatitis E, rubella, coronavirus, and alphaviruses.[70] The alphavirus macrodomain has a highly positively charged patch on the surface, at the crevice of ADP-ribose 1"-phosphate active site and its periphery.[73] The other side of the protein, far from the active site, possesses a negative charge. Thus, the nsP3 macrodomain is considered to complex with ADP-ribose derivatives and RNA. It is believed to control the metabolism of ADP-ribose 1"-phosphate and/or other ADP-ribose derivatives with regulatory functions in the cell.[8, 338]

In addition, studies based on the SINV reported that the nsP3 phosphoprotein is an essential component of the viral replication and transcription process.[87] Functional analysis of the effects of mutations of nsP3 on RNA synthesis demonstrated that alterations may cause a loss of capacity for minus strand synthesis, or a failure to increase plus strand synthesis. A change of Ala68 to Gly leading to a modification of the His-Ala-Val peptide was predicted to form part of the active site of the conserved

nsP3 macrodomain.[87] However, no effects on the ADP-ribose binding was found.[73] In addition, the mutation of amino acids at the position Asn10 and Ala24 in ADP-ribose binding of nsP3 macrodomain in SINV affected the replication and viral RNA synthesis, without affecting the binding region.[86-87]

Recent findings revealed that the CHIKV nsP3 was described to have a novel function as a regulator of the cellular stress response.[339] Studies of SINV-infected cells indicated the importance of nsP3 in the interactions with alphavirus-host.[340] Functional analysis of SFV at the C-terminal region of nsP3 showed that the mutations, where 10 residues at the C-terminal are lacking, suppresses the establishment of infection; while lacking the 30 C-terminal residues led to reduced synthesis of subgenomic RNA.[341] The nsP3 macrodomain has been shown to be responsible for SINV and SFV replication in neurons and neurovirulence in mice.[85-86]

2.1.2 EARLY ATTEMPTS TO DISCOVER CHIKV NSP3 INHIBITORS

The crystal structure of the nsP3 macrodomain of CHIKV was determined in 2010.[73] This structure includes four subunits; and the asymmetric unit consists of six-stranded β-sheets and four α-helices (Figure 2.1). The core β-sheet and positions of the α-helices have proven to be highly conserved. Also present in the nsP3 structure is the ligand ADP-ribose.[70, 73] The active site is in the crevice, between the top of the β-strands 2, 4 and 5 and is surrounded by two loops between β2-α1 and β5-α3.

Currently, there is only one publication regarding molecular modelling for the CHIKV nsP3 macrodomain based on the crystal structures (PDB id: 3GPG) and its complex with ADP-ribose (PDB id: 3GPO).[70] The study focused on an understanding of the specific binding of the ADP-ribose to the nsP3 macrodomain of CHIKV, while also comparing with VEEV. The results of MD simulations of the structure with ADP-ribose identified the binding modes and the key residues for interactions between ADP-ribose and the nsP3.[70] The negatively charged $PO4^{2-}$ component of ADP-ribose showed the strongest interaction with the protein, and the binding free energies estimated from MD simulations were in good agreement with previous experimental data.

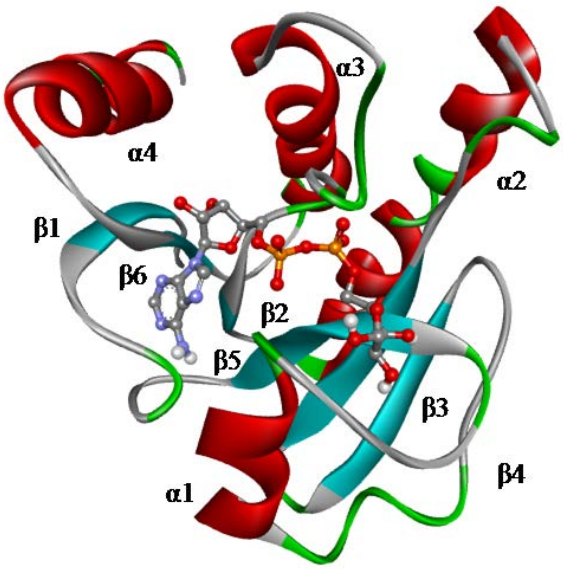


Figure 2.1. X-ray crystal structure of the macrodomain of CHIKV in complex with the ADP-ribose.[73]

2.2 OVERVIEW OF THIS STUDY

There is very little information available on the nsP3 macrodomain and its inhibitor. Therefore, this study focused on using a combination of computational approaches, including molecular docking, virtual screening, MD simulations, and binding free energy calculations to discover potential lead compounds that inhibit the nsP3 in CHIKV and propose compounds for biological testing (Figure 2.2).

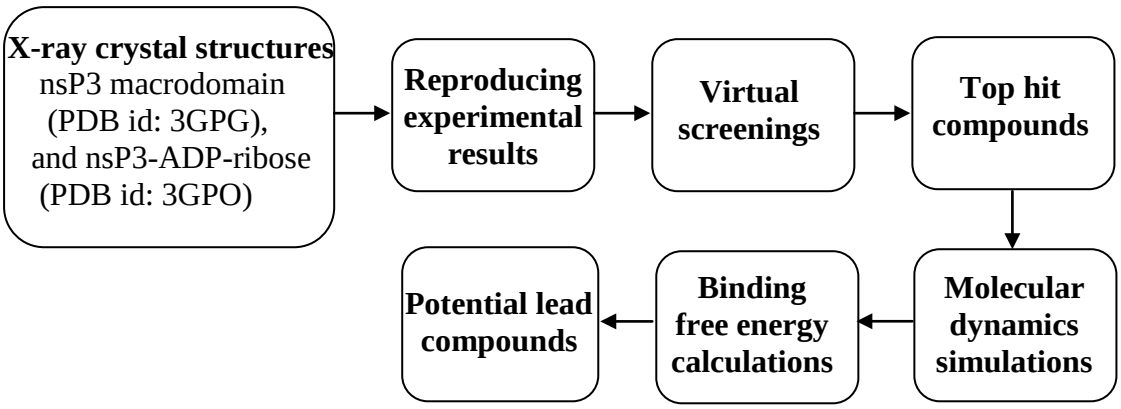


Figure 2.2. Schematic diagram of *in silico* approaches.

2.2.1 MOLECULAR DOCKING AND VIRTUAL SCREENING

A docking protocol was established (details described in the Experimental procedures and methods section), which included the following stages:

The nsP3 macrodomain protein was downloaded from the Protein Data Bank (PDB id: 3GPG) and this was used as the receptor for docking. The protein structure was submitted to WhatIF website to correct manually chirality errors, and the conformation of sidechains of His, Asn, and Gln.

The next step was to carry out energy minimization to relax the structure and remove steric overlaps using the CHARMM22 force field in the Accelrys Discovery Studio (DS) 2.0 software package.[342] The steepest descent algorithm with 3,000 steps was applied. With a C_α RMSD of 0.59 Å between the minimized structure and the X-ray structure, the minimized structure was utilized for the subsequent docking process. Polar hydrogen atoms were added with AutoDock Tools (version 1.5.4). The ligand ADP-ribose was extracted from the complex crystal structure (PDB id: 3GPO) and prepared by AutoDock Tools for docking in Vina. The other ligands employed for virtual screening were taken from National Cancer Institute (NCI) Diversity Set II.

Parameters of the grid box size and centre of the box were defined, and other parameters including a search space or exhaustiveness (E) and number of binding modes to generate (num_modes), were selected using AutoDock Tools. For docking, the grid box needs to be large enough to accommodate the ligands. Initially, in this case, the grid box size and its location of the binding site were defined based on the place where the ligand ADP-ribose was bound in the X-ray structure. During virtual screening, the location and the size of the grid box were carefully investigated via blind docking (in which the box is sufficiently large to cover the whole protein), and focused docking (in which a smaller box was centred on potential binding sites of interest). Blind docking can reveal potential binding sites in the nsP3. The parameters for docking were chosen such as E and the maximum num_modes set to the default values with E = 8 and num_modes = 9.

The docking protocol was validated by re-docking ADP-ribose to compare with the crystal structure of the nsP3-ADP-ribose complex. The re-docking results were evaluated in terms of RMSD value and the binding affinity. After evaluation, the established docking protocol was used for different virtual screenings. In an effort to identify potential inhibitors (hit compounds), 1541 compounds from the NCI Diversity Set II were screened by docking against the nsP3 macrodomain of CHIKV. Based on the starting point of docking with ADP-ribose at ADP-ribose binding site, virtual screening (VST) was carried out with three different setups. The first was a focused docking centred on the ADP-ribose binding site (Pocket 1: VST1 and VST2). The second setup was a blind docking centred either at the middle of the Pocket 1 (VST3) or the protein (VST4) with the box large enough to cover the whole protein. The third setup was a focused docking centred at the predicted binding sites by MetaPocket,[343] (Pocket 2: VST5 and Pocket 3: VST6). The focused dockings at Pocket 2 (VST5) and Pocket 3 (VST6) were also carried out. Interactions of ligands and protein were analyzed from docking results using Accelrys DS 3.5. The Lipinski's rule was also used to give a general drug-likeness information for hits.[206]

2.2.2 MD SIMULATIONS

The program NAMD[265] was used for MD simulations to investigate the stability and flexibility of the hit-target complexes, and study their interactions. The apo protein and the complexes of protein-hit compounds were prepared to run simulations (details of the procedure are given in the Experimental procedures and methods Chapter). The protein atoms were represented with the CHARMM22 force field,[257] and the corresponding parameters for the ligands were generated with AmberTools.[262] The systems were solvated under periodic boundary condition with explicit solvent model TIP3P and 0.15 M NaCl. The Langevin algorithm was used to maintain the temperature at 298.15 K and pressure at 1 atm. The PME algorithm was used to compute long range electrostatic interactions.[269] The cutoff distance for vdW interactions were set at 12 Å and the pair-list distance was 13.5 Å. The minimization process was applied first and followed by equilibrium simulations with weak harmonic restraints on the heavy atoms for 3 ns. The production runs were continued for 50 ns.

The trajectories for analysis were saved every 10 ps. To determine the system stability, the RMSDs of the heavy atoms over 50 ns was calculated with respect to the starting structure versus the simulation time. The RMSF of Cα atoms during the simulations was measured to obtain information on local flexibility of the system. The resulting trajectories such as HBs and hydrophobic contact interactions were analyzed by the CHARMM22[263] and VMD (version 1.9.1),[276] (details in Experimental procedures and methods Chapter, section 5.2). The Clustering Plugin Tool in VMD is used for clustering analyses.

2.2.3 BINDING FREE ENERGY CALCULATIONS

Having obtained the simulation results of protein and its complexes, the simulations of ligands in solvent (water) were run to apply LIE approach to estimate the absolute binding free energies for ligands in complexes with the protein targets. Preparation of simulations for ligands are described in Chapter Experimental procedures and methods. NAMDEnergy plugin in VMD was utilized to compute the energy components over the frames obtained from the MD simulations.[265, 276] The α, β, and γ in the LIE equation need to be defined based on the properties such as hydrophobicity of the ligand and binding site to estimate absolute free energies of binding.

2.3 RESULTS AND DISCUSSION

2.3.1 DOCKING RESULTS WITH ADP-RIBOSE

The docking outcomes of ADP-ribose into the nsP3 protein were evaluated and compared with the available co-crystal structure (PDB id: 3GPO). The different ligand conformations were ranked based on their predicted binding affinities with the default scoring function in Vina (Table 2.1). The RMSD value was calculated between the docked structure and the initial structure (Table 2.1). The best docked pose had a binding affinity of -10.2 kcal/mol.

Table 2.1. Poses in the docking of ADP-ribose into the nsP3. RMSD refers to the heavy-atom RMSD from the co-crystal structure for ADP-ribose with the nsP3.

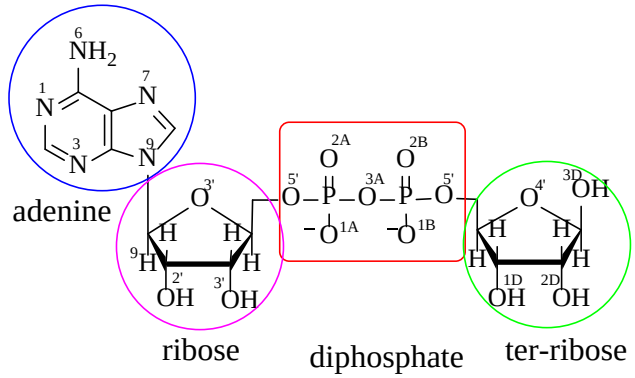
Poses	Binding affinity (kcal/mol)	RMSD (Å)
1	-10.2	0.6
2	-8.9	8.6
3	-8.9	1.8
4	-8.9	5.5
5	-8.6	8.4
6	-8.5	4.4
7	-8.4	5.7
8	-8.3	10.0
9	-7.8	4.5

Details of analysis of interactions between the complex of the best docked of ADP-ribose and the nsP3 are shown in Table 2.2 and Figure 2.3. The residues in the active site of nsP3, namely Ile11, Ala23, Asn24, Asp31, Val33, Leu108, Gly112, Val113, Tyr114, Tyr142, and Arg144, formed HBs with ADP-ribose. Most of the hydrogen bond donors arose from the protein residues, with corresponding acceptors contained in the ADP-ribose. The only exception is that the ribose component of ADP-ribose can be the donor in interactions with Tyr142. In addition, the diphosphate component of ADP-ribose showed the strongest interaction compared to the rest of this ligand with the greatest number of HBs.

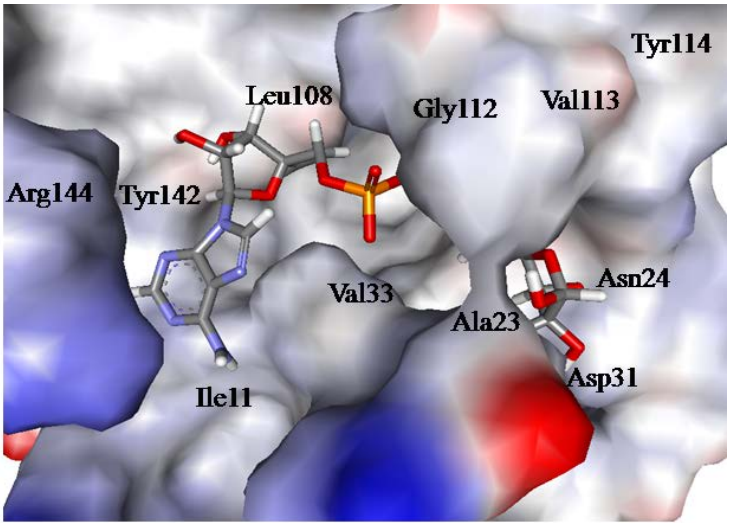
Table 2.2. Analysis of interactions between the best docked of ADP-ribose and the nsP3 macrodomain.

Part of ADP-ribose	Number of HBs	Interactions (Å)
Adenine	2	Arg144(HH21)-N1=2.5 Ile11(HN)-N1=2.1
Ribose	2	H9-Tyr142(O)=2.4 Leu108(HN)-O3'=2.4
Diphosphate	5	Val33(HN)-O1A=2.4 Val113(HN)-O2A=2.4 Val113(HN)-O2B=2.4 Gly112(HN)-O2B=2.1 Tyr114(HN)-O2B=2.1
Ter-ribose	3	Asp31(HN)-O1D=1.9 H9-Ala23(O)=2.4 Asn24(HD21)-O3D=2.1

A



B



C

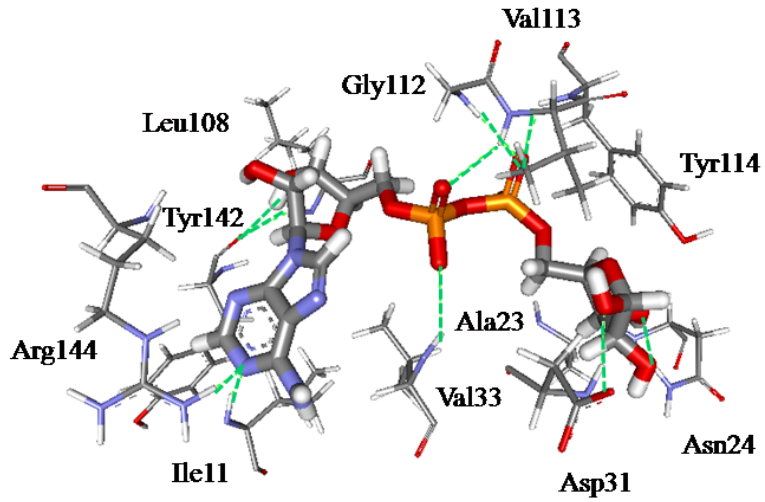


Figure 2.3. Re-docking ADP-ribose **(A)** into the active site of the nsP3: **(B)** The best docking pose of ligand ADP-ribose is represented as a stick model (coloured by atom type) while the protein nsP3 is shown in the solvent surface (coloured by interpolated charge with a probe radius of 1.4 Å). **(C)** The interactions of this pose and the nsP3 residues show hydrogen bonding interactions at the binding site of the nsP3.

Moreover, the accuracy and reliability of docking were evaluated by superimposing both the docked structure and the X-ray structure. The heavy atom RMSD was 0.6 Å, smaller than the 2.0 Å (often used as a criterion for the correct bound structure prediction)[344] (Table 2.1 and Figure 2.4), indicating that molecular docking reproduced the binding mode in the co-crystal structure. A comparison of interactions of docking results with published data (Table 2.3) confirms that there was a good agreement with the key interactions, and showed the current docking protocol was able to reproduce the correct pose. The differences in important residues from forming HBs and hydrophobic contacts were acceptable due to different methods used.

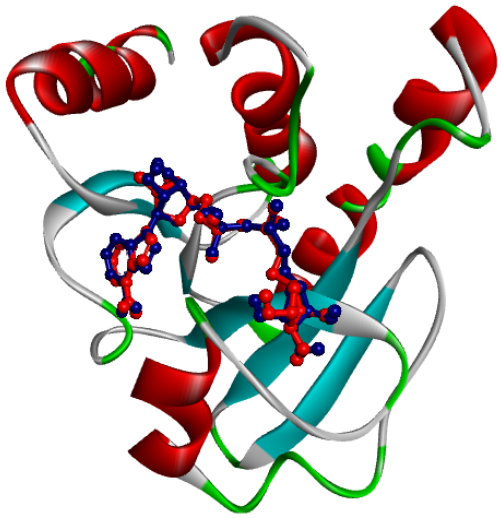


Figure 2.4. Superimposition of the ADP-ribose after docking (in red, the top pose) and its structure in the co-crystal structure (in blue) at the active site of nsP3. The heavy-atom RMSD between the two structures is 0.6 Å.

Table 2.3. Comparison of the identified hydrogen bonding interactions in the nsP3-ADP-ribose docked complex with the previously published data. In Ref [73], key residues including bonding residues (in bold), were identified by experimental work with the crystal structure of complex nsP3-ADP-ribose (3GPO) while residues in Ref [70] were determined by MD simulations of ADP-ribose in the nsP3 based on the above crystal structure.

	Current work	**Ref[73]**	**Ref[70]**
HBs	11	11	11
Interacting residues	**Ile11**, Ala23, **Asn24**, **Asp31**, Val33, Leu108, **Gly112**, **Val113**, **Tyr114**, Tyr142, **Arg144**	Asp10, **Ile11**, **Asn24**, **Asp31**, Thr111, **Gly112**, **Val113**, **Tyr114**, Tyr142, **Arg144**	Asp10, **Ile11**, **Asn24**, **Asp31**, Val33, Ser110, Thr111 **Gly112**, **Val113**, **Tyr114**, **Arg144**

2.3.2 IDENTIFICATION OF INHIBITORS FOR THE NSP3 MACRODOMAIN

Results of virtual screenings based on bind docking and focused docking are listed in the Table 2.4. The top ten compounds for each VST and their binding affinities were selected (structures appended in Appendix 1).

In addition to the proposed ADP-ribose binding site (the active site, Pocket 1), two additional binding sites (Pocket 2 and 3) were identified based on blind docking. The residues making up each pocket are listed in Table 2.5.

Pocket 1 and Pocket 3 share a number of interacting residues including Asn24, Asp31, Val33, Gly112, Val113, and Tyr114. Pocket 2 was found on the opposite side and behind Pocket 1. The locations of pockets in the nsP3 and the locations of top hit ligands in the three pockets are illustrated in the Figure 2.5.

Table 2.4. Results of the top ten compounds of different virtual screens for the nsP3. The binding affinities are shown in kcal/mol.

VST1[a]	VST2[b]	VST3[c]	VST4[d]	VST5[e]	VST6[f]
1. NCI_25457 (-10.8)	1. NCI_34567_a (-10.9)	1. NCI_61610 (-11.1)	1. NCI_61610 (-11.1)	1. NCI_127133 (-8.3)	1. NCI_670283 (-10.6)
2. NCI_116702 (-10.7)	2. NCI_37553 (-10.9)	2. NCI_293778 (-11.0)	2. NCI_37553 (-11.0)	2. NCI_293778 (-8.2)	2. NCI_319990 (-10.2)
3. NCI_309892 (-10.3)	3. NCI_25457 (-10.8)	3. NCI_345647_a (-10.9)	3. NCI_345647_a (-10.9)	3. NCI_338042 (-7.6)	3. NCI_80731 (-10.1)
4. NCI_109451 (-10.2)	4. NCI_116702 (-10.7)	4. NCI_25457 (-10.8)	4. NCI_25457 (-10.8)	4. NCI_132232 (-7.5)	4. NCI_84100_b (-10.1)
5. NCI_127133 (-10.2)	5. NCI_58052 (-10.6)	5. NCI_58052 (-10.6)	5. NCI_293778 (-10.8)	5. NCI_310326 (-7.4)	5. NCI_372287_a (-10.0)
6. NCI_328101 (-10.2)	6. NCI_127133 (-10.5)	6. NCI_127133 (-10.5)	6. NCI_127133 (-10.7)	6. NCI_328101 (-7.4)	6. NCI_84100_a (-9.9)
7. NCI_372275_a (-10.2)	7. NCI_293778 (-10.5)	7. NCI_372499_b (-10.3)	7. NCI_116702 (-10.6)	7. NCI_69359_a (-7.3)	7. NCI_97920 (-9.6)
8. NCI_45545 (-10.2)	8. NCI_670283 (-10.5)	8. NCI_37553 (-10.3)	8. NCI_58052 (-10.6)	8. NCI_90737 (-7.3)	8. NCI_58502 (-9.5)
9. NCI_84100_b (-10.2)	9. NCI_328101 (-10.4)	9. NCI_309892 (-10.2)	9. NCI_670283 (-10.5)	9. NCI_122819_a (-7.2)	9. NCI_227186_a (-9.4)
10. NCI_37168 (-10.1)	10. NCI_372499_b (-10.3)	10. NCI_37168 (-10.2)	10. NCI_324623 (-10.3)	10. NCI_400976 (-7.2)	10. NCI_293778 (-9.4)

(a) In VST1, the grid box is fixed at the centre of Pocket 1 (9.7 Å, 43.0 Å, -13.2 Å) with a dimension of 16 Å × 16 Å × 16 Å. **(b)** In VST2, the grid box is fixed at the centre of Pocket 1 (9.7 Å, 43 Å, -13.2 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(c)** In VST3, the grid box is fixed at the centre of Pocket 1 (9.7 Å, 43 Å, -13.2 Å) with a dimension of 50 Å × 50 Å × 56 Å. **(d)** In VST4, the grid box is fixed at the centre of the protein (7.7 Å, 45.3 Å, -5.3 Å) with a dimension of 50 Å × 50 Å × 56 Å. **(e)** In VST5, the grid box is fixed at the centre of Pocket 2 (7.7 Å, 45.4 Å, 11.5 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(f)** In VST6, the grid box is fixed at the centre of Pocket 3 (2.3 Å, 44.6 Å, -18.3 Å) with a dimension of 20 Å × 20 Å × 20 Å.

Table 2.5. Pocket residues in the nsP3 macrodomain.

Pocket	Pocket residues
Pocket 1	Asp10, Ile11, Ala22, Ala23, Asn24, Gly30, Asp31, Gly32, Val33, Cys34, Gly70, Pro107, Leu108, Leu109, Ser110, Thr111, Gly112, Val113, Tyr114, Tyr142, Cys143, Arg144, Asp145, Trp148
Pocket 2	His-1, His0, Ala1, Pro2, Ser3, Tyr4, Phe129, Met132, Asp133, Ser134, Thr135, Asp136, Ala137, Asp138, Val139, Ile156, Gln157, Arg159, Thr160
Pocket 3	Ala22, Ala23, Asn24, Pro25, Arg26, Leu28, Pro29, Gly30, Asp31, Gly32, Val33, Cys34, Pro51, Val52, Gly70, Pro71, Asn72, Tyr76, Leu108, Ser110, Thr111, Gly112, Val113, Tyr114

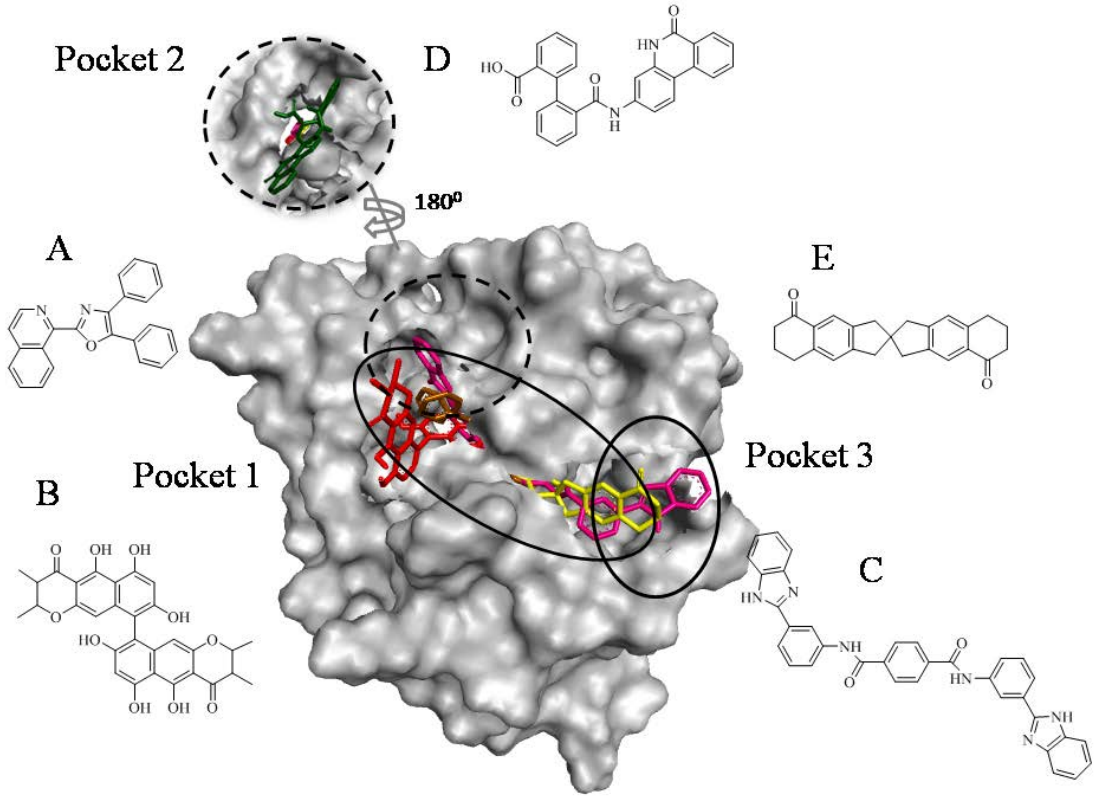


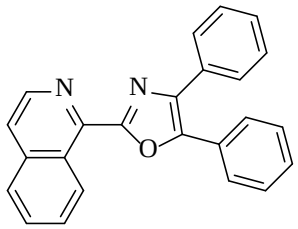
Figure 2.5. Representation of three binding pockets identified in the nsP3 with top hit compounds binding in the pockets. Pocket 1 is the ADP-ribose binding site with ligand NCI_25457 (A, in burgundy), NCI_345647_a (B, in red), and NCI_61610 (C, in pink). Pocket 3 shares some residues with Pocket 1 with ligand NCI_670283 (E, in yellow). Pocket 2 is in the other side of Pocket 1 with ligand NCI_127133 (D, in dark green).

In the focused dockings targeting Pocket 1 (the ADP-ribose binding site, VST1 and VST2), the top hits were NCI_25457 (-10.8 kcal/mol) and NCI_345647_a (-10.9 kcal/mol) (Figure 2.6). Among the top ten hits, four are shared between VST1 and VST2, which differ in the size of the grid box used. It is worth noting that the change in

size of the grid box affected the searching process in Vina, with some new hits being identified. However, the binding affinity values between them are not significantly different from each other. To find more inhibitors, and other potential binding sites in the structure of nsP3, blind docking was used to dock into the entire protein with the grid box centred either at the middle of the ADP-ribose binding site (VST3) or the protein (VST4). For the blind docking (VST3 and VST4), six of the top ten hits are common to VST3 and VST4, and their binding affinities were reproduced within 1.0 kcal/mol. This indicated that the blind dockings are likely to have converged. The results show that most of the top ten ligands fitted well in Pocket 1, and that this pocket can accommodate ligands of different size. However, ligands with bulky structures, such as NCI_293778, NCI_58052 and NCI_61610 (in both VST3 and VST4, Figure 2.6), protruded from the binding site. Therefore, the other pockets surrounding Pocket 1 may serve as alternative binding sites for potential inhibitors.

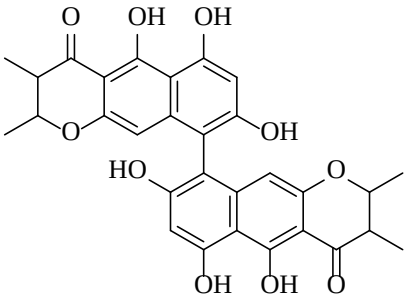
Screenings in VST5 and VST6 produced hits already identified from previous screens, along with some new hits (Table 2.4). For virtual screening, changes to the size of the grid box and its location affected the searching process in Vina. An increase in the dimension of the box is likely to be suitable for larger molecules. For instance, in the blind docking (VST3 and VST4), NCI_61610 (-11.1 kcal/mol) was identified as a top hit, but it did not belong to the top 10 hits in VST1 and VST2. It was also important to note that the majority of hit compounds have a tighter binding in Pocket 1 compared to those in Pocket 2. Most compounds effectively occupied Pocket 2 and Pocket 3 with the significant interactions. Ligands NCI_127133 (-8.3 kcal/mol) and NCI_670283 (-10.6 kcal/mol) (Figure 2.6) bind in Pockets 1, 2, and 3, though in different conformations. Interestingly, the ligand NCI_293778 appeared able to bind in all three pockets, and it may infer that Pocket 1 was more favourable for binding, given the binding affinity of -10.5 kcal/mol compared to -9.4 kcal/mol (Pocket 3) and -8.3 kcal/mol (Pocket 2).

NCI_25457



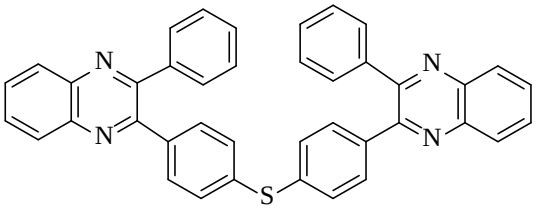
(15)

NCI_345647_a



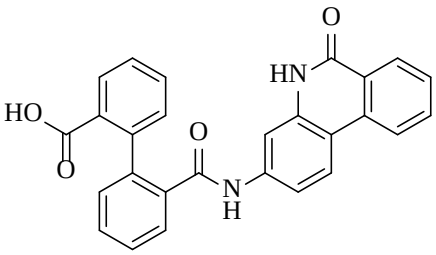
(16)

NCI_293778



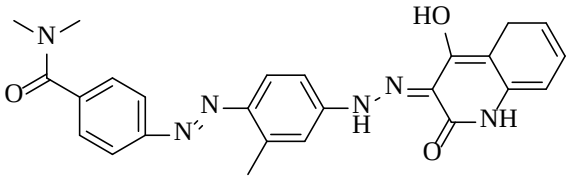
(17)

NCI_127133



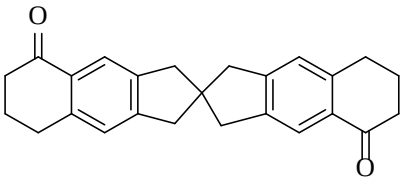
(20)

NCI_58052



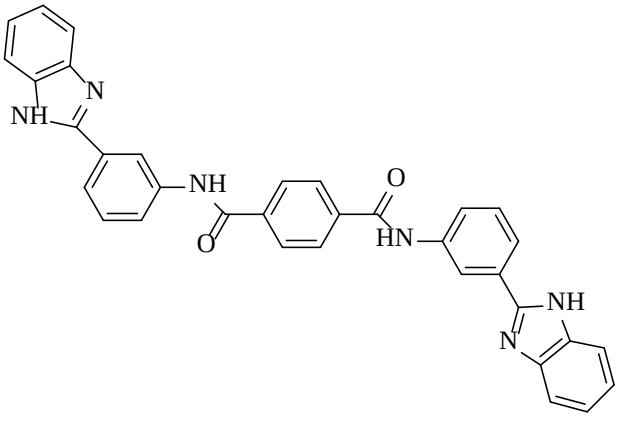
(18)

NCI_670283



(21)

NCI_61610

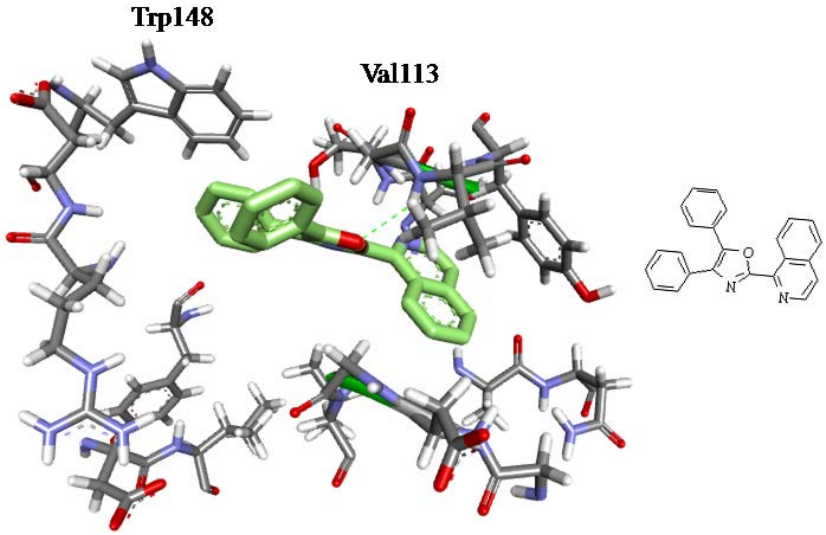


(19)

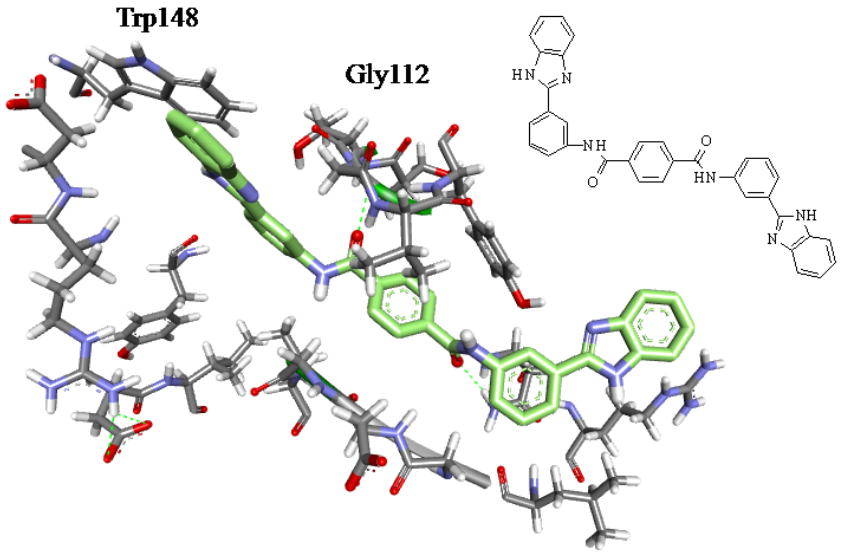
Figure 2.6. Structures of the top hit compounds, obtained from screenings for the nsP3.

Detailed analyses of the interactions between the ligands and protein target were carried out in regard to HBs and hydrophobic contacts (Appendixes 2-7).[224] The results showed that the hydrogen bonding interactions played a more important role in the binding to Pocket 1 and Pocket 3, whereas hydrophobic contacts were responsible for most interactions in the binding to Pocket 2. The contribution of aromatic rings by π-stacking or π-network was emphasized in the enzyme. Most of the ligands can fit very well in the Pocket 1 by forming hydrogen bonds with backbone nitrogen of Val113 or Thr111, and/or interacting through π-stacking or π-network with aromatic ring of Tyr114 or Trp148. For example, top-hit ligands NCI_25457 and NCI_61610 could interact with protein through HBs with residues Val113 and Gly112, respectively; and also π-stacking with Trp148 (Figure 2.7). In addition, the residues located in the region 110-114 play a crucial role in ligand binding to Pocket 1. Mostly, ligands served as hydrogen bonding acceptors while residues of protein were donors. Only interactions between ligands and Thr111, in some cases, this residue can change its role to be an acceptor. In agreement with previous reports,[70] residues Ser110, Thr111, Gly112, and Tyr114 define Pocket 1 and were key residues in forming interactions with ligands. In addition, we found these residues concurrently define Pocket 3. Among them, Tyr114 formed HBs with ligands, or interacted through the π-network on the aromatic ring with most of ligands in both Pockets 1 and 3. Residues Asp31 and Asn75 contributed in forming hydrogen bonds for ligands in Pocket 3. For ligands bound in the Pocket 2, we found residues Tyr4 or Met132, Asp133, and Thr135 are key residues in forming hydrogen bonding interactions between ligands and protein. Additionally, ligand interactions were observed by π-stacking from the aromactic ring of Tyr4.

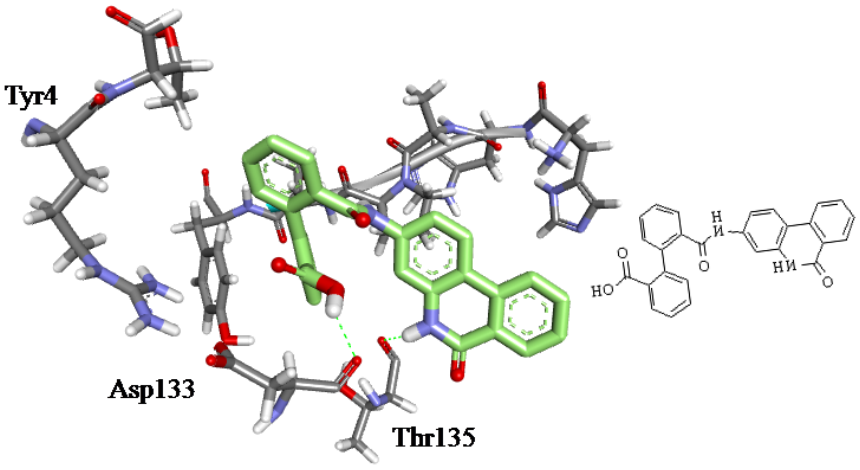
A



B



C



D

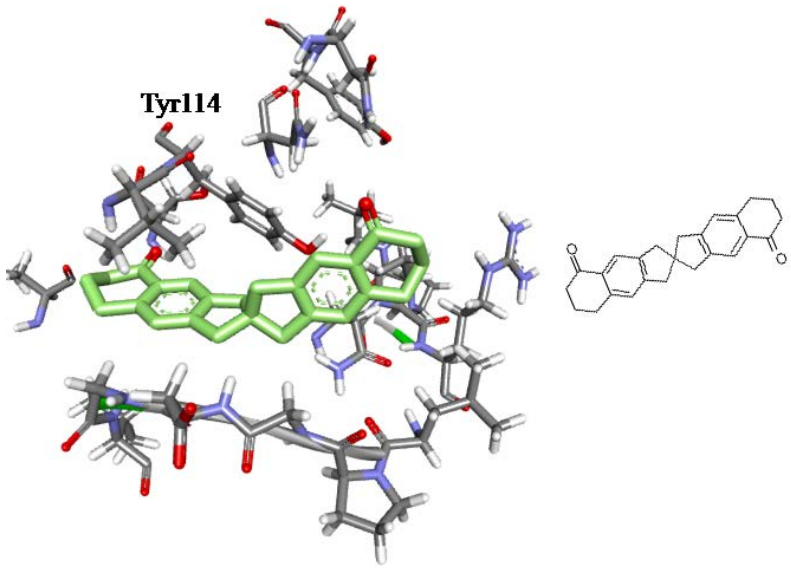
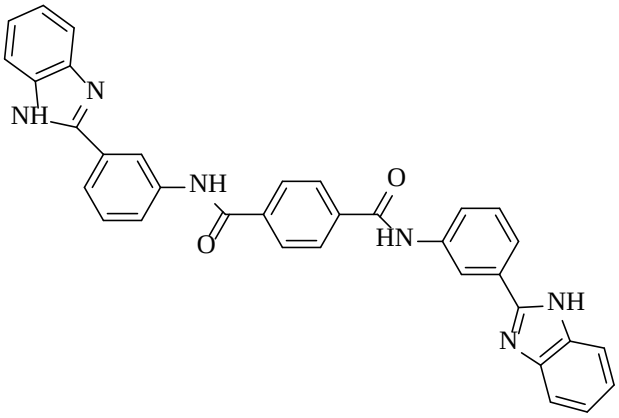
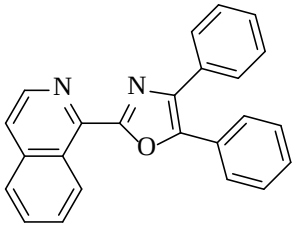
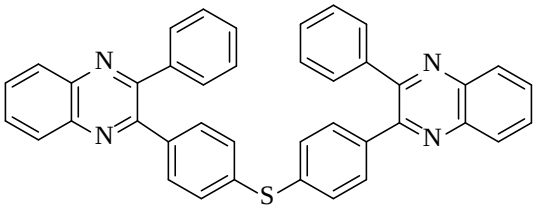
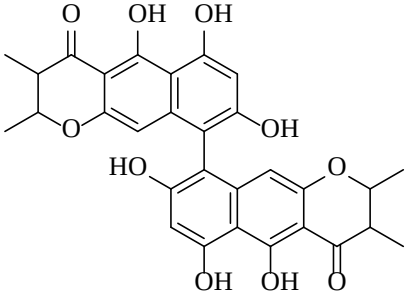
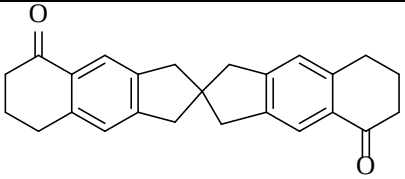


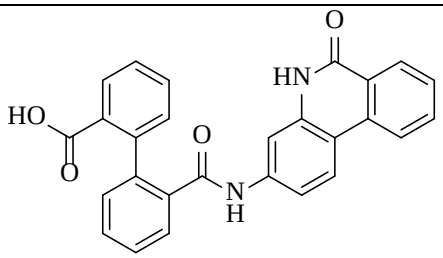
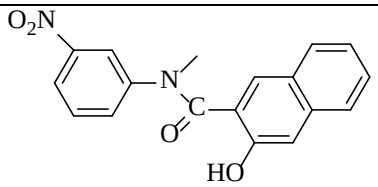
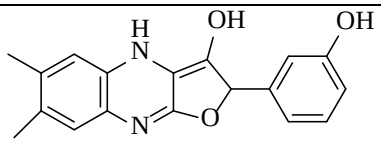
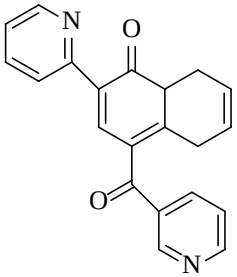
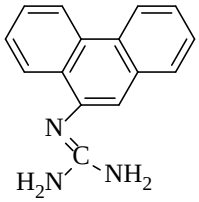
Figure 2.7. Binding pose and interactions of hit compounds in the nsP3 macrodomain: **(A)** NCI_25457 in Pocket 1: HBs with Val113 and π-π interaction with Trp148; **(B)** NCI_61610 in Pocket 1: HBs with Gly112 and π-π interaction with Trp148; **(C)** NCI_127133 in Pocket 2: HBs with Asp133; **(D)** NCI_670283 in Pocket 3: Hydrophobic contacts only. The ligands (in cyan) and the residues surrounding the ligands (in grey) were displayed in sticks and coloured by atoms (carbon in cyan in ligand or in grey in residues, nitrogen in blue, oxygen in red).

2.3.3 MD SIMULATIONS

MD simulations were undertaken to investigate the stability of the protein and its complexes as well as gain insights into the accurate binding modes of the protein and its inhibitors. The top-hit compounds NCI_61610, NCI_25457, NCI_345647_a, NCI_670283, and NCI_127133; and the tenth-hit compounds NCI_37168, NCI_372499_b, NCI_37168, NCI_324623, NCI_400976, and NCI_293778 from each screening were also subsequently submitted to MD simulations (Table 2.6). MD simulations were carried out with the NAMD package with the CHARMM force field for 50 ns following 3 ns equilibrium simulations.

Table 2.6. Chemical structures of five top hit compounds for the nsP3 macrodomain and their properties.

Compound's name	Structural formula	Binding affinity (kcal/mol)	Lipinski's values
NCI_61610		-11.1	LogP[a]: 5.31 H-D[b]: 4 H-A[c]: 4 MW[d]: 548.60
NCI_25457		-10.8	LogP: 5.37 H-D: 0 H-A: 3 MW: 348.39
NCI_293778		-9.4	LogP: 10.87 H-D: 0 H-A: 4 MW: 594.73
NCI_345647_a		-10.9	LogP: 2.17 H-D: 6 H-A: 10 MW: 546.52
NCI_670283		-10.6	LogP: 4.81 H-D: 0 H-A: 2 MW: 356.45

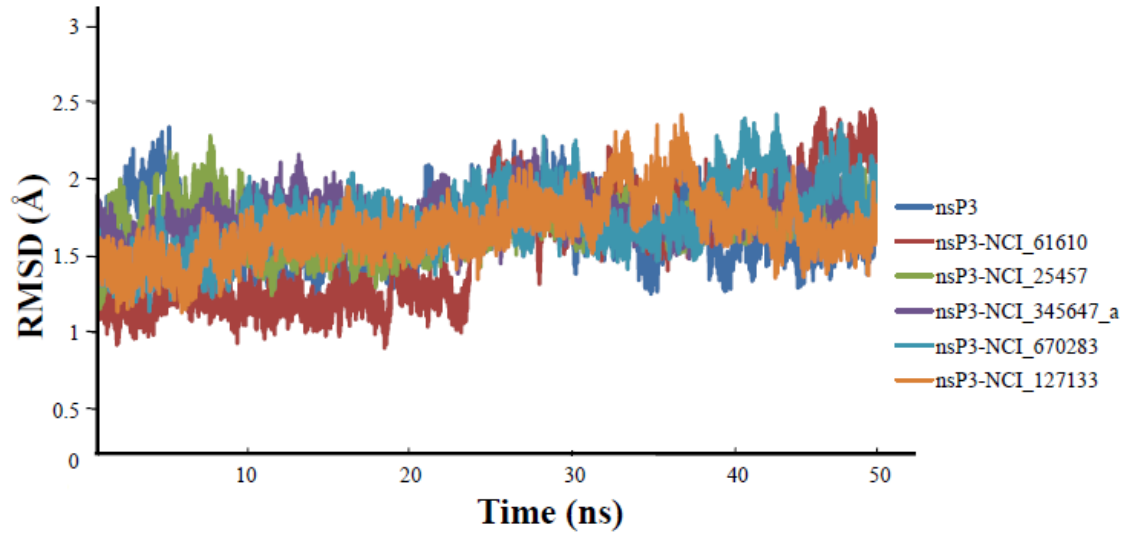
NCI_127133		-8.3	LogP: 4.52 H-D: 3 H-A:4 MW: 434.4
NCI_37168		-10.1	LogP: 2.98 H-D: 1 H-A: 5 MW: 322.32
NCI_372499_b		-10.3	LogP: 1.91 H-D: 0 H-A: 3 MW: 308.34
NCI_324623		-10.3	LogP: 1.37 H-D: 0 H-A: 4 MW: 328.37
NCI_400976		-7.2	LogP: 2.50 H-D: 3 H-A:4 MW: 237.31

(a) A calculated octanol-water partition coefficient; **(b)** H-D: Hydrogen bond donor; **(c)** H-A: Hydrogen bond acceptor; **(d)** MW: Molecular weight.

2.3.3.1 Overall stability of the nsP3 and its complexes

In order to assess overall stability, the values of positional RMSD for backbone from the starting structure are used as a major criterion. Backbone RMSD curves for the nsP3 and its complexes with different ligands with respect to the starting structure after the systems reached equilibrium within 3 ns, are shown in Figure 2.8. The plots showed that most of the systems were relatively stable during the 50 ns simulations within 1-2.5 Å.

A



B

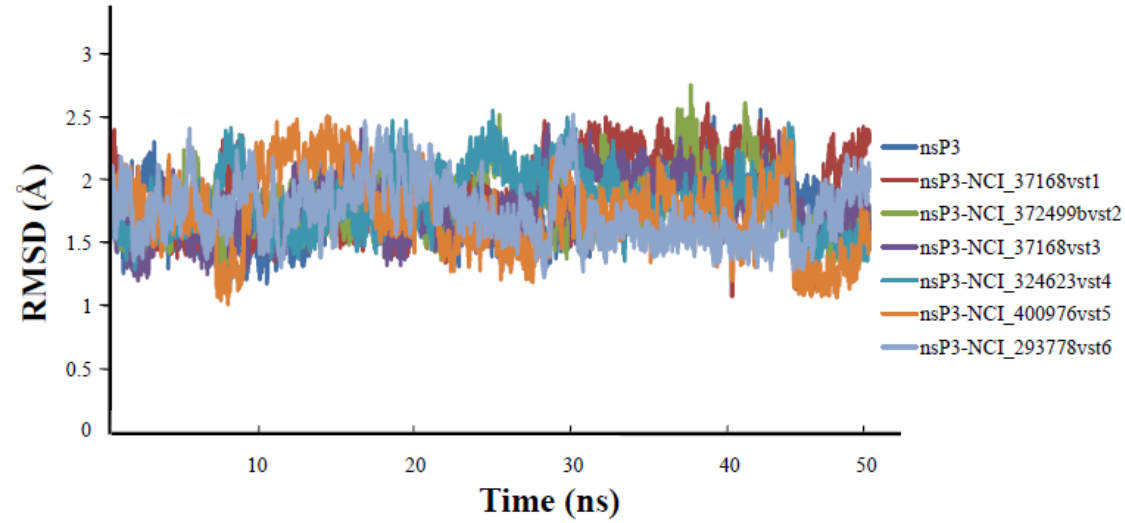
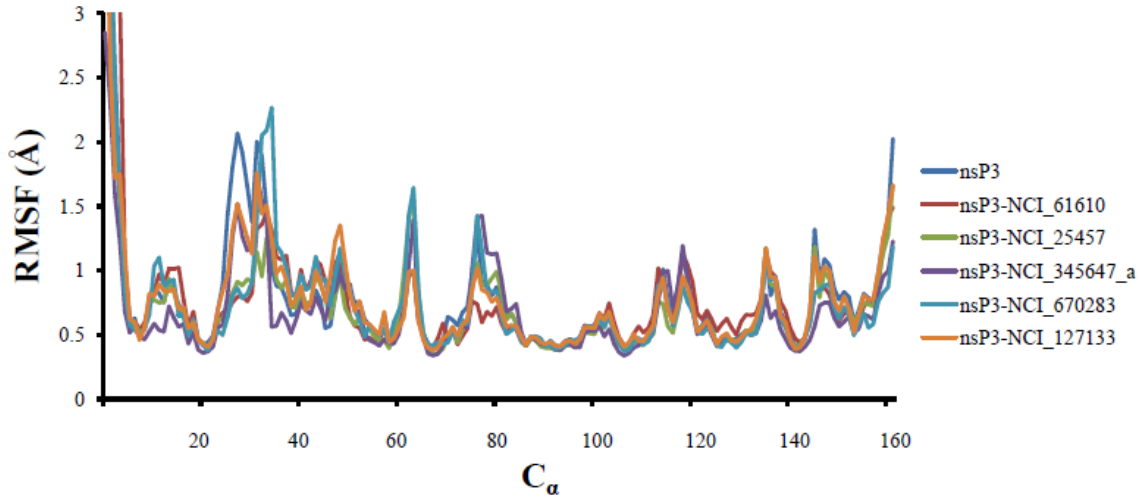


Figure 2.8. The backbone RMSD profiles for the apo protein nsP3 and its different complexes during MD simulations: **(A)** Complexes of the nsP3 and top-hit compounds; **(B)** Complexes of the nsP3 and tenth-hit compounds.

2.3.3.2 Investigating the flexibility of the nsP3 and its complexes

To understand the flexibility of the complex, the RMSF of the Cα atoms of each residue was calculated from the trajectory data for 50 ns for the protein nsP3 and its complexes. The RMSF profiles presented in Figure 2.9 show they are comparatively similar between the apo protein and the complexes.

A



B

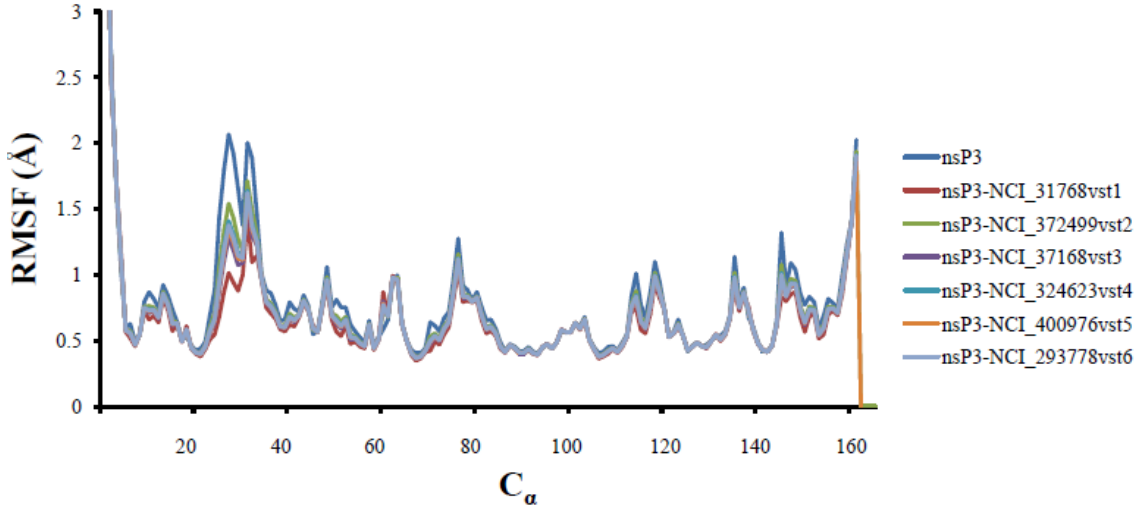


Figure 2.9. RMSF values of C_α atoms of the apo protein nsP3 and its different complexes during MD simulations: **(A)** Complexes of the nsP3 and top-hit compounds; **(B)** Complexes of the nsP3 and tenth-hit compounds.

With regards to hit compounds in Figure 2.9 (A) and (B); the residues making up the binding pockets were quite stable during the simulations (the fluctuation within 1.0 Å). Subtle differences were observed for a few regions, including the loop at residue 31-34. It is worth noting that for three ligands NCI_61610, NCI_25457, and NCI_345647_a bound to the protein at Pocket 1, the RMSFs for the binding loop region 31-34 were decreased compared to those in the apo protein. However, this RMSF for this loop was not significantly perturbed for the ligands NCI_670283 and NCI_127133 when bound to Pocket 3 and Pocket 2, respectively.

2.3.3.3 Atomic interaction between the protein nsP3 and ligands

Detailed analysis of the interactions between the ligands and nsP3 were carried out on the HBs interactions and hydrophobic contacts. The outcomes are listed in Table 2.7 and Table 2.8.

Table 2.7. Hydrogen bonding analyses on the trajectories sampled in MD simulations of hit compounds for the nsP3.

	Number	Details of HBs	% occupancy
NCI_61610	5	Asn24 (HD22)-O1	98
		Tyr114 (HN)-O	92
		Gly112 (HN)-O	88
		Thr111 (OG1)-H1	13
		Cys34 (HG1)-O1	10
NCI_25457	3	Val113 (HN)-N	29
		Val33 (HN)-N	21
		Val33 (HN)-O	20
NCI_345647_a	7	Ile11 (HN)-O4	39
		Ile11 (HN)-O6	12
		Gly112 (HN)-O1	18
		Val33 (HN)-O	17
		Gly32 (O)-O5	14
		Thr111 (HN)-O1	13
		Arg144 (HE)-O3	10
NCI_670283	4	Thr111 (HN)-O	77
		Gly112 (HN)-O	65
		Ser110 (HN)-O	25
		Thr111 (HG1)-O	10
NCI_127133	1	Arg159 (HH12)-O2	22
NCI_37168vst1	7	Asn24 (HD22)-N1	18
		Asn24 (HD22)-O2	20
		Asn24 (HD22)-O3	16
		Asp32 (HN)-N1	26
		Asp31 (HN)-O2	18
		Asp31 (HN)-O3	16
		Cys34 (HG1)-O2	10
NCI_372499vst2	4	Asp145 (OD1)-H	13
		Asp145 (OD1)-H1	19
		Asp45 (OD2)-H	19
		Asp145 (OD2)-H1	26
NCI_37168vst3	13	Asn24 (HD22)-N1	25
		Asn24 (HD22)-O2	24
		Asn24 (HD22)-O3	24
		Gly30 (HN)-O3	10
		Cys34 (HN)-N1	12
		Cys34 (HN)-O2	12
		Cys34 (HN)-O3	11
		Leu108 (O)-H1	92

NCI_37168vst3		Thr111 (HN)-O1	92
		Gly112 (HN)-O	90
		Gly112 (HN)-O1	69
		Val113 (HN)-O	49
		Tyr114 (HN)-O	90
NCI_324623vst4	5	Asn24 (HD22)-O	56
		Val33 (HN)-O1	49
		Cys34 (HG1)-N1	11
		Thr111 (HN)-O	24
		Gly112 (HN)-O	34
NCI_400976vst5	1	Tyr4 (HN)-N1	11
NCI_293778vst6	4	Asn24 (HD22)-N	81
		Arg26 (HE)-N2	32
		Arg36 (HH12)-N2	12
		Asp31 (HN)-N1	13

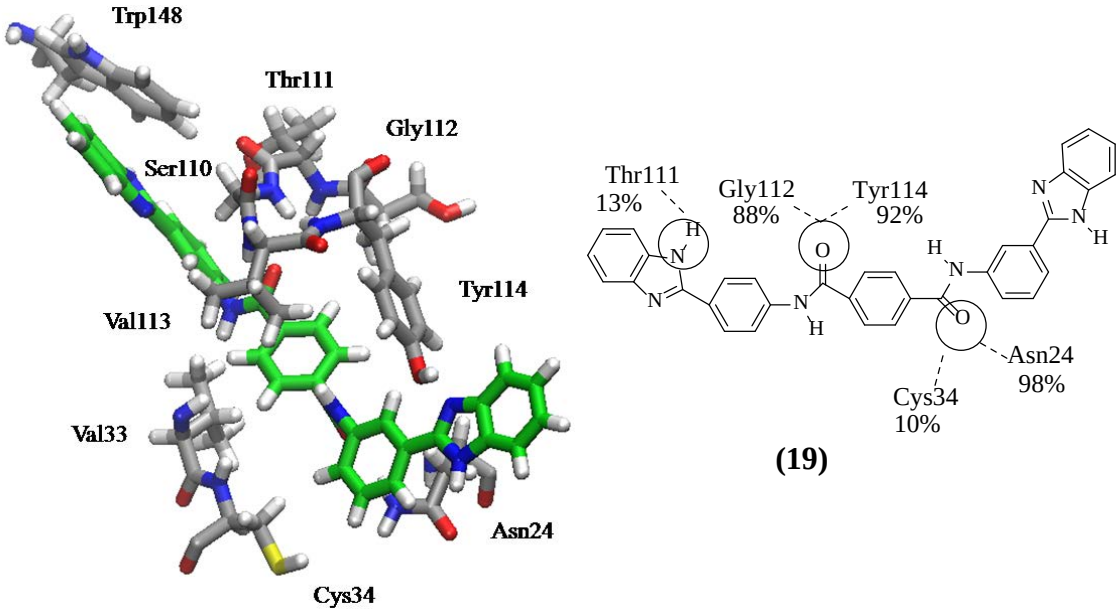
Table 2.8. Hydrophobic contact analyses on the trajectories sampled in the MD simulations of hit compounds for the nsP3.

Ligand	**Non-polar part of residues**
NCI_61610	Ala22, Pro25, Leu28, Val33, Pro107, Val113, Tyr114, Trp148
NCI_25457	Ala22, Val33, Pro107, Val113, Tyr114, Trp148
NCI_345647_a	Ile11, Val33, Ala36, Val113, Tyr114, Trp148
NCI_670283	Ala22, Leu28, Val33, Pro107, Val113, Tyr114
NCI_127133	Ala1, Pro2, Tyr4
NCI_37168vst1	Ala22, Ala23, Val33, Pro107, Thr111, Val113, Tyr114, Arg144
NCI_372499vst2	Val33, Arg144, Trp148
NCI_37168vst3	Ala22, Ala23, Val33, Pro107, Val113, Tyr114, Trp148
NCI_324623vst4	Ala22, Val33, Phe45, Pro107, Val113, Tyr114
NCI_400976vst5	Ala1, Pro2, Tyr4, Phe129, Arg159
NCI_293778vst6	Ala22, Ala23, Pro25, Arg26, Leu28, Val113, Tyr114

For hit compounds, hydrogen bonding and hydrophobic contact analyses indicated that all investigated ligands are stabilized the protein by a number of HBs (Tables 2.7 and 2.8). Complementary to the docking where the protein was kept rigid, MD simulations revealed that when the ligands bind to the nsP3, the ligand and/or the residues in the binding pockets fluctuate and adapt their structure in order to better accommodate the ligands by optimizing HBs and/or hydrophobic contacts. Most ligands at Pocket 1 and Pocket 3 bound strongly to the nsP3; always displaying strong HBs, particularly with Asn24, Val33, Cys34, Thr111, Gly112, Val113, and Arg114, while Tyr4 is important in Pocket 2.

Moreover, based on the occupancy of HBs for each ligand, the results showed that Pocket 1 could have the highest potential for ligand binding, as Pocket 1 had more HBs with higher occupancy than those in Pocket 3 > Pocket 2. As mentioned above, some residues, such as Val33, Val113, Tyr114, Arg144, and Trp148 (for compounds binding to Pocket 1), displayed noticeable movement upon binding. For the ligands NCI_127133 (Pocket 2), and NCI_670283 (Pocket 3), the results showed the fluctuation of residues Ala1, Pro2, and Tyr4 (at Pocket 2), and residues Val113, and Tyr114 (Pocket 3), were required for a correct fit into the nsP3. In particular, some ligands (NCI_61610 and NCI_37168) established the strong HBs interactions with the nsP3 at Pocket 1, emphasized on the residues at region of 110-114. For instance, occupancy of HBs of NCI_61610 and residues at Pocket 1, namely Asn24, Tyr114, Gly112 was 98%, 92%, and 88%, respectively. Ligand NCI_37168 hydrogen-bonded with residues Thr111 (92%), Gly112 (90%), Tyr114 (90%); and two medium hydrogen bonds between NCI_670283 and Thr111 (77%); and Gly112 (65%) at Pocket 3 were observed. For ligand NCI_61610 at Pocket 1, half of the ligand was quite stable, while the other half was flexible enough to fit well through interacting with Trp148 by π-π interaction and forming HBs with residues in the region 110-114 (illustrated in Figure 2.10). For ligand NCI_670283 at Pocket 3, the part of ligand which interacted with the region of residues from 110-114 was optimized to fit well in the pocket, even though the frequency of HBs interactions were medium with Thr111 (77%), Gly112 (65%), and weak with Ser110 (25%), respectively (Figure 2.10).

A



B

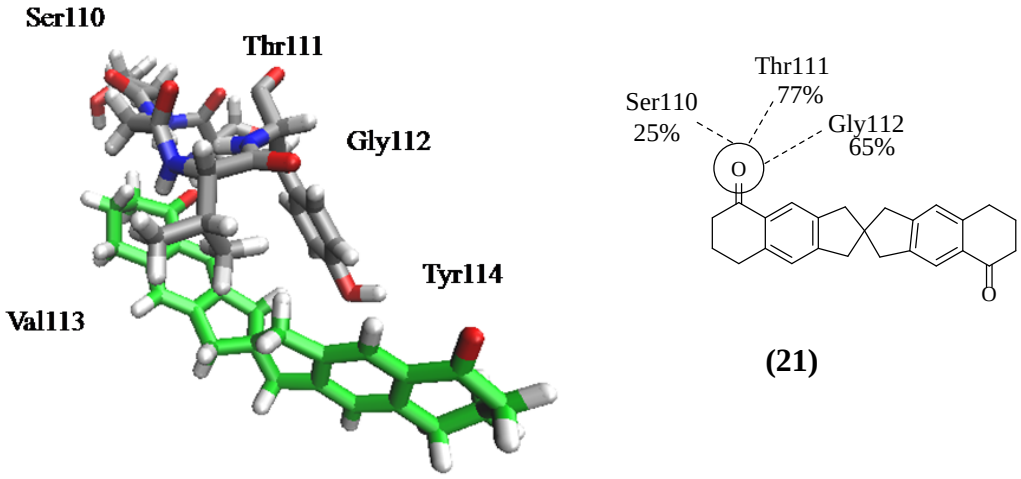


Figure 2.10. Hydrogen bonding interactions between the nsP3 and ligands: **(A)** Ligand NCI_61610 at Pocket 1 and **(B)** Ligand NCI_670283 at Pocket 3, with representation of ligands and key residues for interactions surrounding the ligands (in stick).

Furthermore, the key residues interacting in the region 110-114 of the protein at Pockets 1 and 3, served as hydrogen bond donors, except in the complex nsP3-NCI_61610, where residue Thr111 served as an acceptor. This observation is in close agreement with docking results as well as with earlier simulations and experimental data.[70] In addition, it emerged that the structure of NCI_61610 and NCI_345647_a are more polar, thus more hydrogens bonds were found in their complexes than others.

The solvent accessible surface areas (SASA) were also calculated to monitor the possible solvation environment change upon ligand binding (Appendix 8). It was expected that the SASA for hydrophobic interacting residues in the complex protein-ligand would be decreased compared to those in apo protein. At Pocket 1, when the ligands bind to protein, it can be seen that the SASA of residue Tyr114 showed a decrease from 70.2 $Å^2$ in the apo nsP3 to 63.4 $Å^2$ for the nsP3-NCI_61610 complex, and 63.8 $Å^2$ for nsP3-NCI_345647_a complex. Also, the SASA of Val33 displayed a reduction from 68.8 $Å^2$ in the apo state to 57.4 $Å^2$ and 52.9 $Å^2$ for the bound state in nsP3-NCI_25457, and NCI_345647_a, respectively. However, changes observed in SASA for Val33, Val113, and Trp148 are not consistent for different ligands. It can be rationalised that these residues were not only able to form hydrophobic contacts, but also form polar hydrogen bonding interactions. Thus, the change in SASA will be compromised by the polar interactions and both of protein and ligands will modulate the SASA values.

2.3.3.4 Clustering analysis

Throughout the simulations, the complex structure of the protein and ligands could vary under the effects of environment, so structural clustering was used to identify the most popular conformation during the simulation, and more important to compare the structures from MD simulation and docking. Clustering analysis was carried out on all of the snapshots from the trajectories, and the clusters were visualized and superimposed with the initial structure. The value of RMSD was used to evaluate the difference between clustering structures and initial structure. The different conformations of protein and its complexes at 0 ns, 10 ns, 20 ns, 30 ns, 40 ns, and 50 ns were superimposed. Slightly fluctuations were found in all complexes, though had very little significance. For example, for the ligand NCI_61610, in Figure 2.11, superimposition of the most popular conformations of protein obtained from simulations and RMSD value for the ligand was 1.2 Å (at 10 ns), 1.1 Å (at 20 ns), 1.8 Å (at 30 ns), 2.2 Å (at 40 ns), and 2.0 Å (at 50 ns), with respect to the initial structure.

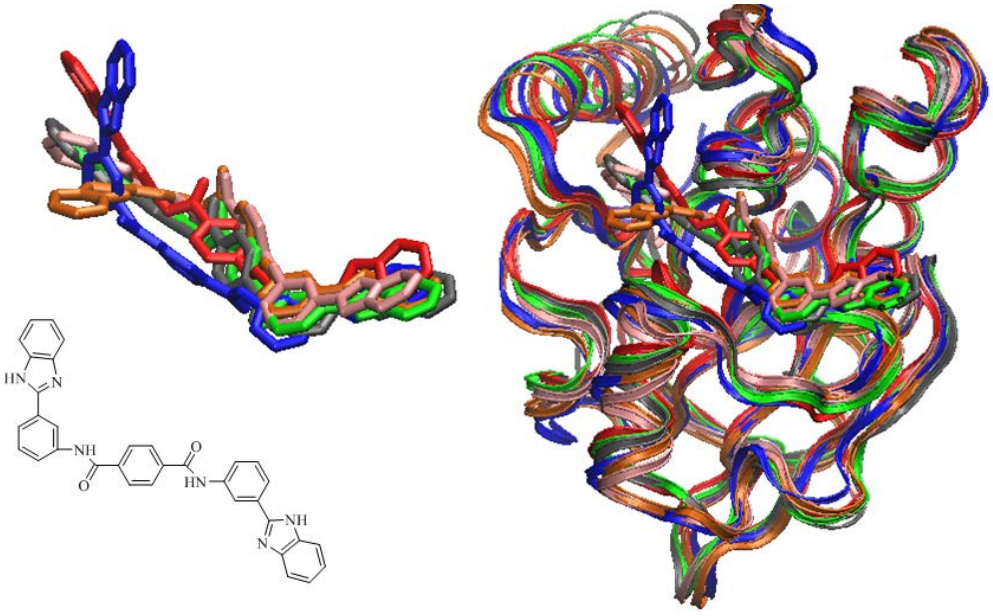


Figure 2.11. Superimposition of the different conformations of ligand and complexed ligand NCI_61610-nsP3 during simulation with the initial structure (red: at 0 ns, grey: at 10 ns, green: at 20 ns, pink: at 30 ns, orange: at 40 ns, and blue: at 50 ns).

2.3.3.5 Combination of MD simulations and docking for the nsP3

In order to probe the effects of the static protein structure used in the docking, multiple docking simulations and virtual screenings were carried out based on the sampled conformations of protein nsP3-NCI_61610 at the different timepoints (5 ns, 10 ns, 15 ns, and 20 ns). This complex was selected as NCI_61610 showed the highest potential for interacting with the protein after analyzing the outcomes of docking and simulations. These results are listed in Table 2.9 and Table 2.10. The binding affinities of these docking runs in Table 2.10 were not significantly different from those obtained from previous docking based on the X-ray structure. Additionally, the binding modes are similar (details of interactions analysis in Appendixes 9-12). For virtual screening, most of the top hits were the same compounds as previous screening indicated although there were some new hits. That indicates that in the case of nsP3, the docking results were not very sensitive for the static structure used.

Table 2.9. Re-docking results for complex nsP3-NCI_61610 with different conformations of the nsP3 protein taken from the different timepoints in simulations at Pocket 1.

nsP3 conformation	Binding affinity (kcal/mol)	Interaction between the inhibitor and residues of protein (with distance in Å)
At 0 ns	-11.1	Tyr114(HH)-N=2.0 Val33(HN)-O=2.3 Asn24(HD21)-O=2.4
At 5 ns	-10.3	Tyr114 (OH)-H1=1.9
At 10 ns	-11.3	Asn24 (HD22)-O=2.1 Tyr114 (OH)-H1=2.3
At 15 ns	-10.6	Asn24 (HD22)-O1=2.5 Ser110 (HN)-O=1.9
At 20 ns	-11.4	Tyr114 (OH)-H1=2.3

Table 2.10. Virtual screening results for blind docking into Pocket 1 with different conformations of the nsP3 taken from the different timepoints in simulations. The binding affinities are shown in kcal/mol.

VST-5ns[a]	VST-10ns[b]	VST-15ns[c]	VST-20ns[d]
1. NCI_293778 (-11.6)	1. NCI_37553 (-11.1)	1. NCI_293778 (-10.5)	1. NCI_37553 (-12.1)
2. NCI_84100_b (-10.7)	2. NCI_293778 (-11.1)	2. NCI_308835 (-10.3)	2. NCI_61610 (-11.9)
3. NCI_84100_a (-10.6)	3. NCI_60785_a (-11.0)	3. NCI_37553 (-10.3)	3. NCI_670283 (-11.7)
4. NCI_80997_b (-10.5)	4. NCI_ 59620_a (-10.8)	4. NCI_97920 (-10.0)	4. NCI_293778 (-11.3)
5. NCI_37553 (-10.4)	5. NCI_27592_a (-10.7)	5. NCI_84100_b (-9.9)	5. NCI_60785_a (-11.1)
6. NCI_61610 (-10.3)	6. NCI_670283 (-10.7)	6. NCI_59620_a (-9.8)	6. NCI_63680 (-11.1)
7. NCI_670283 (-10.2)	7. NCI_82802_a (-10.7)	7. NCI_84100_a (-9.8)	7. NCI_82802_a (-11.0)
8. NCI_59620_a (-10.2)	8. NCI_328101 (-10.5)	8. NCI_37627 (-9.8)	8. NCI_219894 (-10.8)
9. NCI_308835 (-10.2)	9. NCI_59620_a (-10.5)	9. NCI_60785_b (-9.7)	9. NCI_80997_b (-10.7)
10. NCI_60785_a (-10.2)	10. NCI_308835 (-10.5)	10. NCI_25457 (-9.7)	10. NCI_328101 (-10.7)

(a) VST-5ns, the grid box is fixed at the centre of Pocket 1 (9.7 Å, 43.0 Å, -13.2 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(b)** In VST-10ns, the grid box is fixed at the centre of Pocket 1 (9.7 Å, 43 Å, -13.2 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(c)** In VST-15ns, the grid box is fixed at the centre of Pocket 1 (9.7 Å, 43 Å, -13.2 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(d)** In VST-20ns, the grid box is fixed at the centre of the protein at the centre of Pocket 1 (9.7 Å, 43 Å, -13.2 Å) with a dimension of 20 Å × 20 Å × 20 Å.

2.3.4 BINDING FREE ENERGY CALCULATIONS FOR THE LIGANDS BINDING TO THE NSP3 PROTEIN

The converged trajectories of complexes obtained from simulations were used to calculate the binding free energy for the hit compounds. The average values of vdW and elec interactions for each ligand in the two states, bound state (V_{bound}) and unbound state ($V_{unbound}$) were calculated (Table 2.11).

In applying the LIE equation, the vdW values of the ligands complexed with the enzyme were more negative than when only in solution, showing that the ligand has more favourable vdW interactions when complexed. The electrostatic interactions of ligands in all the complexes were significantly less favourable than those of the ligand in water. When using the values of empirical parameters α, β and γ, for example β = 0.43 for neutral compounds, α = 0.18 and γ = 0.0,[328] the value of binding free energies of all of complexes did not agree with the docking results (ΔG^1 in Table 2.11). However, as it has been shown that α and γ need to be recalibrated depending on the different systems; the α value has been suggested to be dependent on the system and the force field used in the LIE calculations while the γ relies on the nature of the binding site. In this case, due to the lack of experimental data of complexes, the chemical nature of the ligands and the binding sites were taken into the consideration. Most binding sites are composed of both polar and non-polar residues, so the magnitude of hydrophobicity in the selection of γ is not easy to define in practice, the exact value of γ does not affect the relative ranking. A larger value of α = 1.043 was adopted as it had been shown to provide a better estimate in the study of cytochrome P450-camphor analogue complexes.[324] The results of ΔG^2, with α set to 1.043, are shown in Table 2.11, and revealed better agreement with the binding affinity obtained from docking. This gave a good explanation when comparing the binding free energies for different ligands at the same pocket. It could be explained that van der Waals and electrostatic results were compromised by hydrogen bonding interactions and hydrophobic contacts between the ligands and the nsP3 that contributes to the binding free energy.

Table 2.11. Average binding free energies (kcal/mol) of top-hit compounds and tenth-hit compounds for the nsP3 calculated by LIE method using data trajectories from the MD simulations: ΔG^1 (in kcal/mol using $\alpha = 0.18$, $\beta = 0.43$, and $\gamma = 0$) or ΔG^2 (in kcal/mol using $\alpha = 1.043$, $\beta = 0.43$, and $\gamma = 0$).[328] ΔG is the predicted binding affinity by Vina (in kcal/mol).

Compound	$\mathbf{V^{vdw}_{bound}}$	$\mathbf{V^{vdw}_{unbound}}$	$\mathbf{V^{elec}_{bound}}$	$\mathbf{V^{elec}_{unbound}}$	**ΔvdW**	**ΔElec**	$\mathbf{\Delta G^1}$	$\mathbf{\Delta G^2}$	**ΔG**
NCI_61610	-75.5	-54.3	-76.9	-93.1	-21.2	16.2	3.6	-15.2	-11.1
NCI_25457	-47.7	-36.2	-28.3	-32.9	-11.6	4.7	-0.1	-10.0	-10.8
NCI_345647_a	-64.3	-49.5	-36.9	-44.7	-14.8	7.8	0.7	-12.1	-10.9
NCI_670283	-51.7	-37.7	-28.9	-32.4	-14.0	3.4	-1.1	-13.2	-10.6
NCI_127133	-45.5	-36.6	-66.3	-65.8	-8.9	-0.6	-1.8	-9.6	-8.3
NCI_37168vst1	-43.6	-29.7	-41.2	-47.0	-13.9	5.7	-0.04	-12.0	-10.1
NCI_372499vst2	-36.3	-27.8	-49.9	-53.4	-8.5	3.5	-0.03	-7.3	-10.3
NCI_37168vst3	-45.0	-30.1	-50.0	-44.4	-14.9	-5.6	-5.1	-17.9	-10.2
NCI_324623vst4	-41.5	-29.8	-41.7	-49.2	-11.6	7.50	1.12	-8.9	-10.3
NCI_400976vst5	-31.9	-23.2	-33.7	-70.8	-8.7	37.1	14.4	6.9	-7.2
NCI_293778vst6	-72.2	-55.9	-38.1	-48.1	-16.3	10.1	1.4	-12.7	-9.4

The binding free energies in Table 2.11 indicated that most of hits compounds showed good interactions with Pocket 1 and Pocket 3, as the values of ΔG^2 were higher than those obtained from docking, except the ligand NCI_400976vst5 at Pocket 2. A comparison of binding free energies results of top-hit compounds complexes and tenth-hit compounds complexes revealed that the binding free energies of most of top-hit compounds were better than those of tenth-hit compounds even though they were similar in docking results, which supported a close agreement between docking results and MD simulations results (Table 2.11). Two ligands NCI_37168vst3 and NCI_293778vst6 were tenth-hits which had the binding free energies lower then those in top-hits. However, these two ligands showed better binding affinities than some top-hits. These result emphasized important roles of MD simulations for investigating the binding affinity for ligand. Further work and results from experimental data are required to clarify the issue.

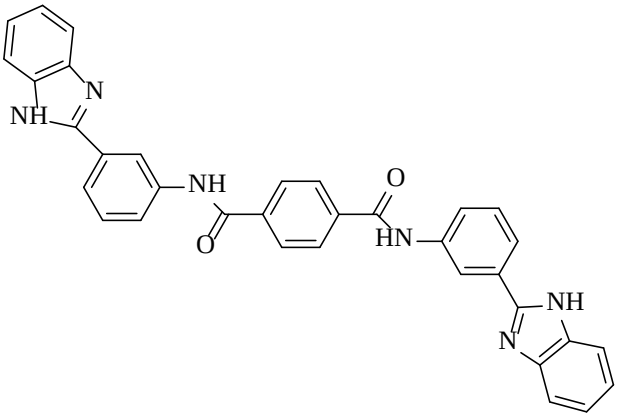
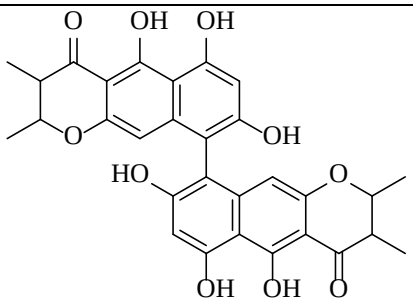
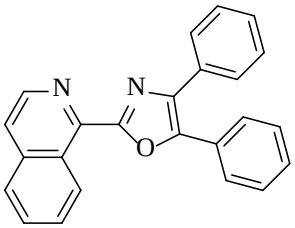
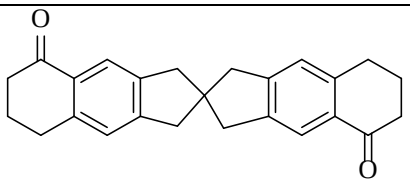
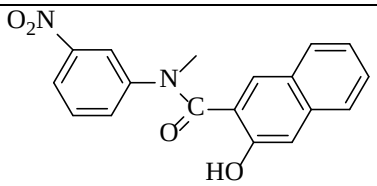
2.3.5 ANALYSIS AND SELECTING LEADS FOR BIOLOGICAL TESTING

In docking and virtual screening, based on the binding affinities and interaction analysis, the highest priority potential compounds would be top-hit compounds, namely NCI_61610 (-11.1 kcal/mol), NCI_345647_a (-10.9 kcal/mol), NCI_25457 (-10.8 kcal/mol), NCI_670283 (-10.6 kcal/mol), NCI_37168 (around -10.0 kcal/mol), NCI_372499_b (-10.3 kcal/mol), and NCI_324623 (-10.3 kcal/mol).

Throughout analyzing MD simulations and binding free energy calculations using simulation results, insights into the interactions of ligands and the protein nsP3 were analyzed at an atomic level. Therefore, the compounds considered would be NCI_61610 (-15.2 kcal/mol), NCI_670283 (-13.2 kcal/mol), NCI_293778 (-12.7 kcal/mol), NCI_345647_a (-12.1 kcal/mol), NCI_37168 (-12.0 kcal/mol), and NCI_25457 (-10.0 kcal/mol).

Taking the results from docking, simulations and binding free energy calculations together, a list of compounds was proposed for biological testing would be NCI_61610, NCI_345647_a, NCI_25457, NCI_670283, and NCI_37168 (Table 2.12).

Table 2.12. Potential lead compounds for the nsP3 proposed for biological testing.

No	Compounds	Chemical structure
1	NCI_61610	
2	NCI_345647	
3	NCI_25457	
4	NCI_670283	
5	NCI_37168	

2.4 CONCLUSIONS

In conclusion, with a combination of molecular docking, virtual screening, and MD simulations, we identified potential inhibitors for CHIKV against the nsP3 macrodomain. So far, very little information is known in the literature about the nsP3 macrodomain and its inhibitors. Therefore, this study is of great importance in the identification of potential novel inhibitors for the chikungunya virus. It is also the first in combining computational tools to identify potential inhibitors targeting CHIKV through taking advantages of the recently determined high-resolution crystal structure of nsP3 in complex with ADP-ribose. Considering the presence of a positive charge at the ADP-ribose binding site, this binding pocket may have the tendency to accommodate negatively charged ligands, which may not have acceptable pharmacokinetic properties. Therefore, we decided to identify all possible binding pockets based on available methods. Starting with re-docking ADP-ribose, the bound structure and the key interacting residues in the ADP-ribose binding pocket were successfully reproduced. It was evident that the binding affinity of ADP-ribose was less than -10 kcal/mol showing this negatively-charged ligand binds very tightly into the nsP3. This study also supported that the current docking protocol utilising AutoDock Vina is robust in reproducing experimentally determined binding modes.

Subsequently, virtual screening with the NCI Diversity Set II was undertaken to identify inhibitors targeting CHIKV. In addition to the well-characterised adenine binding pocket (Pocket 1), two additional binding pockets (Pocket 2 and Pocket 3) were identified through blind docking and MetaPocket. Pocket 3 was overlapped partly by Pocket 1, while Pocket 2 was found on the opposite side of the protein. A comparison of the binding affinities at the three different binding pockets revealed that all the hits at Pocket 1 bind to the protein well, with binding affinities of less than -10 kcal/mol. This suggests that the Pocket 1 would be the most favourable for ligand binding. In contrast, Pocket 2 might not be a good place for binding compared to Pocket 1 and Pocket 3 as those had higher binding affinities. It is also important to note that some ligands were repeated many times in the list of virtual screening results. With the different conformations, these ligands can bind to different pockets e.g. the ligand NCI_127133

binds to Pocket 1 and 2; ligand NCI_670283 binds to Pocket 1 and 3; and ligand NCI_293778 binds to Pocket 1, 2, and 3. In such circumstances, these ligands present better prospects as inhibitors.

By analysing the hydrogen bonding and hydrophobic interactions between the ligands and the nsP3, key residues were identified. The region of residues 110-114 was predicted as the most important area in interacting with ligands. This is consistent with previous work by Rungrotmongkol *et al*.[70]

Furthermore, equilibrium MD simulations were carried out to validate the molecular docking results. Analysis of the simulations confirmed that the docked complex of most of the ligand-nsP3 were stable over the simulation of 50 ns. However, subtle structural rearrangements in the nsP3 were observed to better accommodate the ligands. MD simulations also confirmed that Pocket 1 is the more favourable pocket.

Binding free energy calculations were conducted using MD simulations results. Most of the top-hit compounds were found to have higher potential inhibitory than the tenth-hit compounds in terms of the binding free energy calculations values.

In future work, the detailed binding modes for the identified inhibitors will be better characterized, with more extensive computational free energy calculations, and the inhibitory effects will be verified by experimental studies. Additionally, our current work also provides important input for constructing a pharmacophore for future ligand-based drug design and generating the chemical libraries for inhibitors targeting CHIKV.

CHAPTER 3. DISCOVERY OF INHIBITORS TARGETING CHIKV NON-STRUCTURAL PROTEIN 2

3.1 INTRODUCTION

3.1.1 FUNCTIONAL IMPORTANCE OF THE NSP2 IN DRUG DESIGN

The second non-structural protein in the CHIKV genome, nsP2, is a multifunctional enzyme,[8, 345] and an attractive target for anti-CHIKV drug discovery and development. In its free form, the protein induces cytotoxicity, and may be responsible for transcriptional shut-off.[68] The CHIKV nsP2 also inhibits the cellular response to type I[346-347] and type II[347] interferons (IFN) to block the host antiviral response. The complete nucleotide sequence of the CHIKV genome reveals that the nsP2 protein is the largest non-structural protein, being 798 amino acids long, and possessing a large net positive charge.[67] Nucleoside triphosphatase, helicase, and RNA-dependent 5'-triphosphatase activities reside in the N-terminal domain while activities of cysteine protease, also known as thiol protease, are present in the the C-terminal domain.[345, 348-349] The nsP2 protease plays an essential role in viral replication cycle through a cleavage of the non-structural polyproteins into non-structural proteins.[345] The mechanism of protease catalysis of VEEV (a virus in the same family of CHIKV) is related to the deprotonation of a thiol group of cysteine residue at the active site by assistance of an adjacent histidine residue (Figure 3.1).[8] This could be explained by the reaction of imidazole ring of histidine and the thiol group of cysteine to form a thiolate/imidazolium ion pair which is highly nucleophilic.[350] The ion pair attacks the scissile amide bond and the intermediated products are stabilized by the oxyanion. Through catalysis, the intermediate converts to the thioester and releases the C-terminal substrate fragment. Hydrolysis regenerates the active site and the N-terminal substrate fragment. In addition, based on a study on the nsP2 protease of the SINV (another alphavirus), the substitution of amino acids at the catalytic dyad, either cysteine or histidine at the active site, can completely abolish this enzyme function.[346]

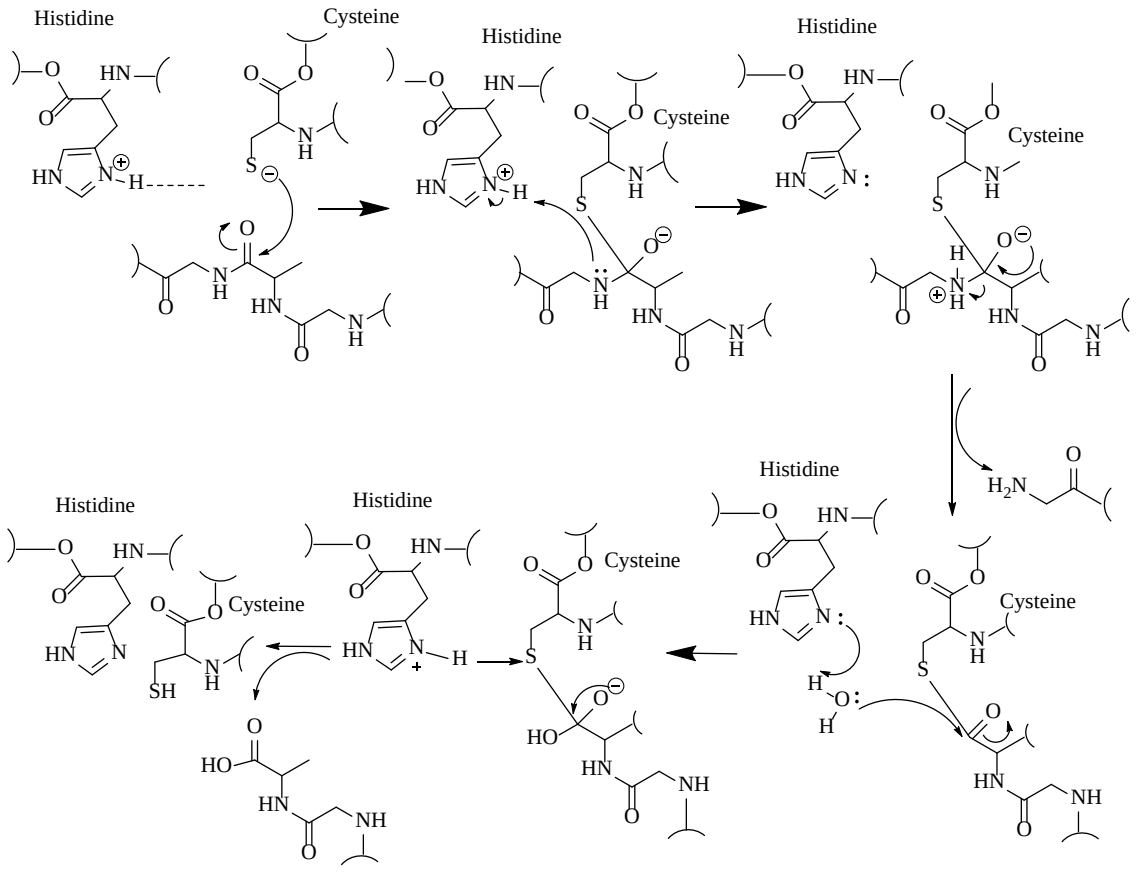


Figure 3.1. Proposed mechanism of protease catalytic of the nsP2 (adapted from Andrew T. R. *et al*).[350]

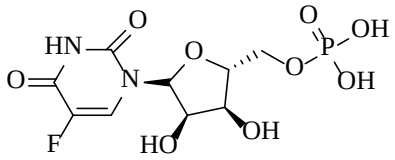
3.1.2 EARLY ATTEMPTS TO DISCOVER CHIKV NSP2 INHIBITORS

Although the function and role of the nsP2 protease is well-known, few studies have focused on targeting inhibitors for this CHIKV enzyme. Several studies constructed homology models of the nsP2 protease based on the counterparts of the VEEV to assist in drug design for CHIKV,[351-353] before a crystal structure of the nsP2 protease became available in late 2011. With the homology models, inhibitors of the nsP2 enzyme were screened, allowing compound designs based on the pharmacophore models.[351-353] The active site was predicted to be in the major surface groove of the C-terminal-domain region, and likely binds to the substrate polyprotein sequences in the cleavage process.[8] The key residues responsible for interactions between the nsP2 protease and ligands were established and pharmacophore features of inhibitors were suggested.[351] Based on docking results, the four best compounds, Compound **22**-**25** (Figure 3.2) were proposed as potential inhibitors of CHIKV nsP2. In another study, the structure-activity

relationships of the hit compounds were evaluated.[353] Compound **28** and compound **29** were reported as promising lead compounds with predictions of CHIKV replication inhibition at low μM concentrations, (compound **28** with EC_{50}=5.0 μM; compound **29** with EC_{50}=3.2 μM). The structures of mentioned compounds were in Figure 3.2

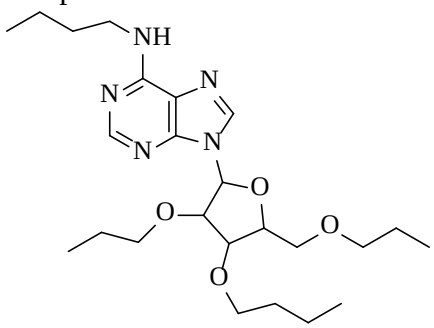
Overall, these studies have provided useful information for understanding ligand binding and key interacting residues for the nsP2 protease. However the possible limitation is that they were conducted using the homology models of the nsP2 protease. The recently determined crystal structure of the nsP2 protease opened a new pathway for drug design targeting this enzyme.

Compound
(ID: 27943-Binding Database)[351]



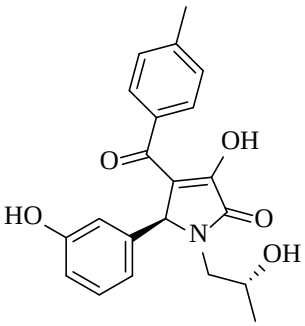
(22)

Compound[352]



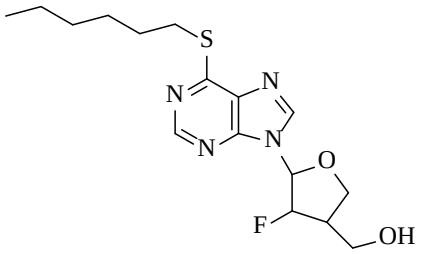
(26)

Compound
(ID: 21362-TosLab)[351]



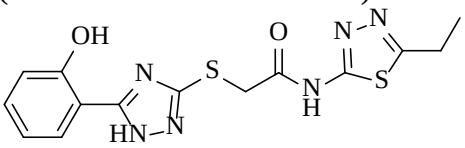
(23)

Compound[352]



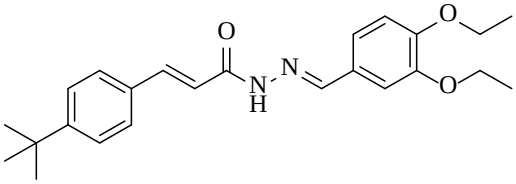
(27)

Compound
(ID: ASN 01541696-Asinex)[351]



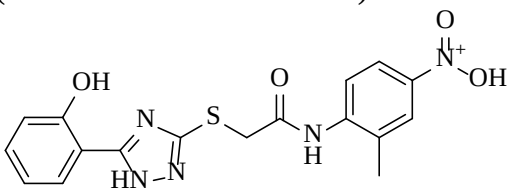
(24)

Compound[353]



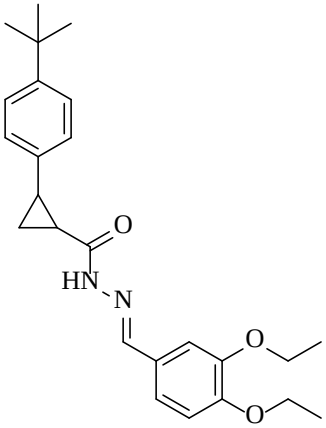
(28)

Compound
(ID: ASN 01107557-Asinex)[351]



(25)

Compound[353]



(29)

Figure 3.2. Structures of some potential inhibitors for the nsP2.

3.2 OVERVIEW OF THIS STUDY

Based on the recently reported crystal structure of CHIKV nsP2 protease (PDB id: 3TRK), this current study aimed to discover potential new inhibitors of the CHIKV nsP2 protease through structure-based approaches, by combining molecular docking and MD simulations. The same process with discovering inhibitors targeting the nsP3 (described in Chapter 2) was applied for the nsP2 protease (more details in Experimental procedures and methods section).

There was modification to the protocol in that the best four compounds **22-25** taken from Ref[351] were docked in the binding site of the nsP2 protease to confirm its location. This binding site was assigned binding site 1 (Pocket 1). In addition, identification of the active site of the enzyme was discussed (see Results section), and the proposed active site was named Pocket 4 in this study.

In silico virtual screening based on docking was then performed to explore inhibitors and potential binding sites of the nsP2 protease. Blind dockings and focused dockings centred at each of the potential pockets identified were carried out. Additionally, the potential binding pockets were predicted with the MetaPocket algorithm.[343] The compounds from library NCI Diversity Set II were docked to the entire protein (the nsP2 protease) with a grid box placed at either the centre at Pocket 1 (VST1) or at the centre of the protein (VST2). The binding modes of the compounds were ranked by their predicted binding affinities. In total, there were five potential binding pockets (labeled as Pockets 1 to 5, shown in Figure 3.3) identified for the nsP2 protease (discussed in the Results section). The VST3/VST4, VST5, VST6, VST7, and VST8 screenings were focused at Pocket 1, Pocket 2, Pocket 3, Pocket 4, and Pocket 5, respectively. The blind docking data initially revealed four potential binding sites in the nsP2 protease (Pockets 1, 3, 4, 5). Encouragingly, the four binding pockets identified in the blind docking were reproduced by the MetaPocket method, which also detected Pocket 2 (near Pocket 1) despite the methods being based on very different algorithms.

The top hit compounds were identified based on their binding affinities using the default scoring function in Vina. The modes of interaction between the ligands and the protein were analyzed in Accelrys DS 2.0. The drug-like properties of these compounds were subsequently evaluated using the Lipinski Guidelines.[206]

To understand not only the rigid structures but also the dynamic behaviour, MD simulations were subsequently carried out to provide details of the motion, and flexibility of the nsP2 protease and its complexes upon binding of small molecules. Analysis of simulations was carried out from the trajectories, saved every 10 ps. The value of RMSD of the heavy atoms over 50 ns and the RMSF of Cα atoms during the MD simulations was measured with respect to the starting structure versus the simulation time, to obtain information on global stability, and local flexibility of the system. The interactions in terms of hydrogen bonding and hydrophobic contact were analyzed by the CHARMM22 and VMD (version 1.9.1).[276] Clustering analysis was also conducted using the Clustering Plugin in VMD. Detail was described in Chapter 2, section 2.2.2.

The binding free energies for docked structures binding to the enzyme were estimated based on MD simulations using the LIE method (see Experimental procedures and methods chapter). A combination of all these methods helped to propose the list of potential hits for biological testing.

3.3 RESULTS AND DISCUSSION

3.3.1 MOLECULAR DOCKING AND VIRTUAL SCREENING

3.3.1.1 Docking to confirm the binding site

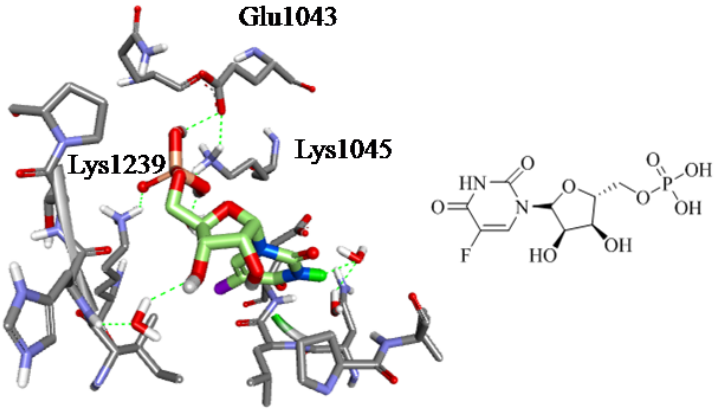
The results of docking the four best compounds **22-25** from Ref[351] into the Pocket 1, are listed in Table 3.1. Pocket 1 includes the key residues for interactions Lys1045, Gly1176, His1222, and Lys1239, which was taken from the previous data by Singh *et al.*[351]

Table 3.1. Docking results of the best compounds for the nsP2 taken from previous study [from Singh K. D. *et al*[351]] with the binding affinities (kcal/mol).

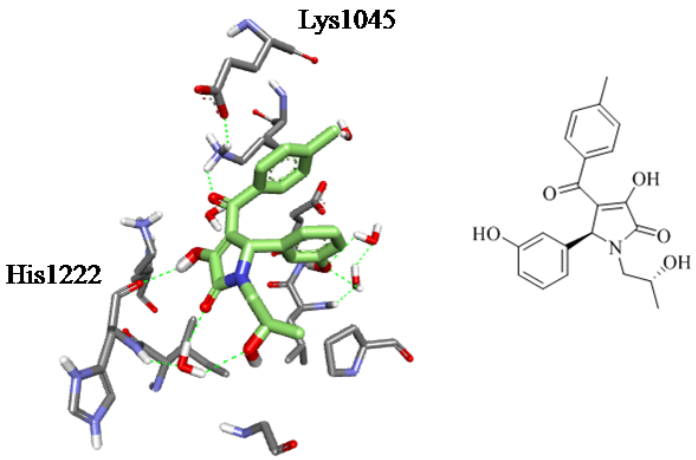
Binding affinity (kcal/mol)	**Compound 22**	**Compound 23**	**Compound 24**	**Compound 25**
	-6.6	-6.3	-6.3	-7.1

The binding affinities (Table 3.1) along with analyses of the interactions of compounds and protein were in agreement with those published. The key residues for interactions at the active site (Figure 3.3) were re-confirmed, although the binding affinities of these compounds were just medium level with the values of around -6.5 kcal/mol.

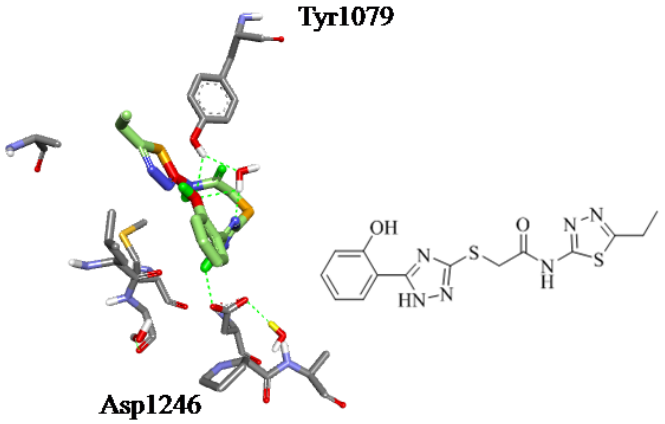
A



B



C



D

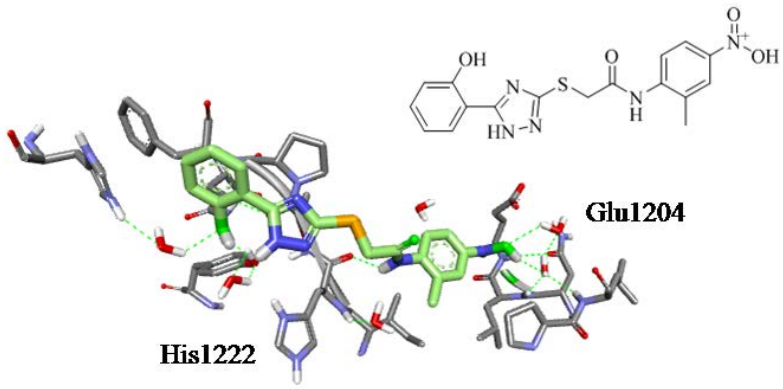


Figure 3.3. Binding pose and interactions of compounds 22-25 in the nsP2 protease: **(A)** Compound 22 in Pocket 1: HBs with Glu1043, Lys1045 and Lys1239; **(B)** Compound 23 in Pocket 1: HBs with Lys1045 and His1222; **(C)** Compound 24 in Pocket 1: HBs with Tyr1079 and Asp1246; **(D)** Compound 25 in Pocket 1: HBs with Glu1204 and His1222. Compound 22, 23 and 25 were in good position in the pocket, except compound 24 was in the rear and nearly out of the pocket. The ligands (in cyan) and the residues surrounding the ligands (in grey) were displayed in sticks and coloured by atoms (carbon in cyan in ligand or in grey in residues, nitrogen in blue, oxygen in red, sulphur in organge).

3.3.1.2 Blind and focused dockings to identify potential binding sites and hit compounds for the nsP2 protease

The outcomes of two blind dockings as mentioned above, one with the centre at Pocket 1 and one at the centre of protein, are displayed in Table 3.2. The results of top ten

compounds and their binding affinities in focused dockings at each pocket are presented in Table 3.3.

Table 3.2. Results of the top ten hit compounds from the blind dockings for the nsP2. The binding affinities ΔG are in kcal/mol.

VST1[a]			**VST2**[b]		
Hits	ΔG	Location	Hits	ΔG	Location
1. NCI_61610	-10.6	Pocket 1	1. NCI_293778	-10.3	Pocket 4
2. NCI_293778	-10.3	Pocket 4	2. NCI_61610	-10.1	Pocket 1
3. NCI_116702	-9.2	Pocket 3	3. NCI_670283	-9.7	Pocket 4
4. NCI_37553	-9.2	Pocket 4	4. NCI_116702	-9.2	Pocket 3
5. NCI_84100_a	-9.2	Pocket 3	5. NCI_217697	-9.2	Pocket 1
6. NCI_84100_b	-9.1	Pocket 3	6. NCI_84100_a	-9.2	Pocket 1
7. NCI_25457	-9.0	Pocket 3	7. NCI_84100_b	-9.2	Pocket 1
8. NCI_670283	-9.0	Pocket 1	8. NCI_298892_b	-9.1	Pocket 1
9. NCI_97920	-9.0	Pocket 3	9. NCI_37553	-9.1	Pocket 5
10. NCI_58052	-8.8	Pocket 3	10. NCI_25457	-9.0	Pocket 3

a) In VST1, the grid box is centred at Pocket 1 of the nsP2 protease (12.4 Å, 34.3 Å, 28.6 Å) with dimensions of 66 Å × 86 Å × 60 Å. b) In VST2, the grid box is centred at the centre of the nsP2 protease (14.3 Å, 25.5 Å, 22.3 Å) with dimensions of 60 Å × 70 Å × 60 Å. The origin and axes of the coordinate systems were set the same as in the PDB structure (PDB id: 3TRK).

Table 3.3. Results of the top ten compounds of focused dockings for the nsP2. The binding affinities are in kcal/mol.

VST3[a]	VST4[b]	VST5[c]	VST6[d]	VST7[e]	VST8[f]
1. NCI_217697 (-9.3)	1. NCI_293778 (-9.8)	1. NCI_293778 (-9.8)	1. NCI_37553 (-9.6)	1. NCI_293778 (-10.4)	1. NCI_293778 (-10.2)
2. NCI_670283 (-9.3)	2. NCI_37553 (-9.7)	2. NCI_37553 (-9.4)	2. NCI_319990 (-9.4)	2. NCI_670283 (-9.7)	2. NCI_80734 (-9.2)
3. NCI_298892_b (-9.2)	3. NCI_328101 (-9.3)	3. NCI_269148_a (-8.5)	3. NCI_116702 (-9.2)	3. NCI_121868_a (-9.3)	3. NCI_61610 (-9.1)
4. NCI_328101 (-9.2)	4. NCI_670283 (-9.3)	4. NCI_146771 (-8.3)	4. NCI_61610 (-9.2)	4. NCI_37553 (-9.2)	4. NCI_37553 (-9.0)
5. NCI_25457 (-9.1)	5. NCI_25457 (-9.2)	5. NCI_227186_b (-8.3)	5. NCI_25457 (-9.0)	5. NCI_61610 (-9.2)	5. NCI_80735 (-8.7)
6. NCI_116702 (-9.0)	6. NCI_298892_b (-9.2)	6. NCI_308835 (-8.3)	6. NCI_293778 (-9.0)	6. NCI_319990 (-8.9)	6. NCI_80997_a (-8.4)
7. NCI_80997_a (-9.0)	7. NCI_61610 (-9.2)	7. NCI_7524_b (-8.3)	7. NCI_97920 (-9.0)	7. NCI_328101 (-8.9)	7. NCI_345845 (-8.1)
8. NCI_109451 (-8.9)	8. NCI_84100_a (-9.2)	8. NCI_227186_a (-8.2)	8. NCI_67436 (-8.9)	8. NCI_67436 (-8.9)	8. NCI_5157 (-8.1)
9. NCI_93427 (-8.9)	9. NCI_84100_b (-9.2)	9. NCI_328101 (-8.2)	9. NCI_80997_b (-8.9)	9. NCI_227186_a (-8.8)	9. NCI_58052 (-8.1)
10. NCI_127133 (-8.8)	10. NCI_217697 (-9.1)	10. NCI_362639 (-8.2)	10. NCI_84100_b (-8.9)	10. NCI_122819_b (-8.7)	10. NCI_67436 (-8.1)

(a) In VST3, the grid box is centred at Pocket 1 (12.4 Å, 34.3 Å, 28.6 Å) with a dimension of 16 Å × 16 Å × 16 Å. **(b)** In VST4, the grid box is centred at Pocket 1 (12.4 Å, 34.3 Å, 28.6 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(c)** In VST5, the grid box is centred at Pocket 2 (28.1 Å, 36.9 Å, 30.6 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(d)** In VST6, the grid box is centred at Pocket 3 (30.4 Å, 42.1 Å, 14.8 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(e)** In VST7, the grid box is centred at Pocket 4 (3.2 Å, 31.9 Å, 18.1 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(f)** In VST8, the grid box is centred at Pocket 5 (15.2 Å, 13.9 Å, 11.5 Å) with a dimension of 20 Å × 20 Å × 20 Å.

The blind docking data initially revealed four potential binding sites in the nsP2 protease (Pockets 1, 3, 4, 5) where the hit compounds could potentially bind. Encouragingly, as mentioned above, a further pocket (Pocket 2) near Pocket 1 was also found by the the MetaPocket method together with four binding pockets identified in the blind docking. The top docked structures of each virtual screening are represented in Figure 3.3, showing the locations of five pockets.

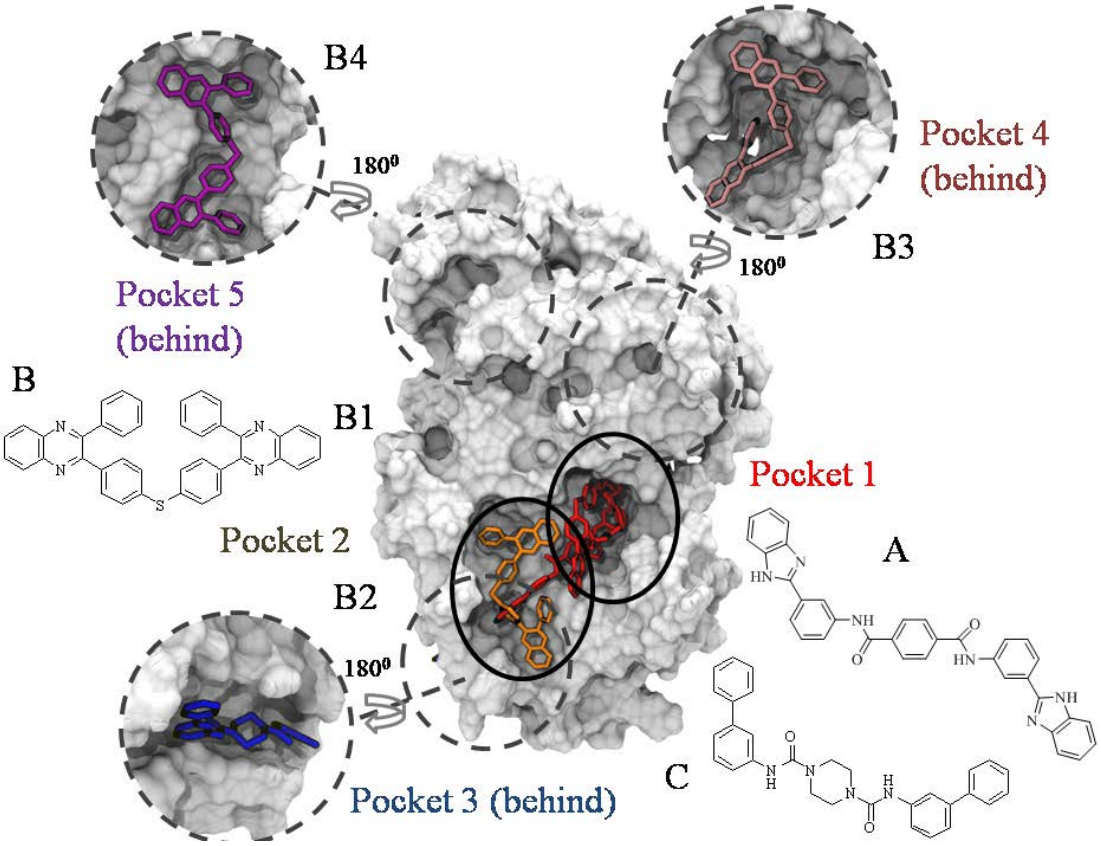


Figure 3.4. Representation of docked structures of top hit compounds in different virtual screenings at five different binding sites of the nsP2 protease with Pocket 4 being the active site of the nsP2 protease. Ligand NCI_61610 (A) and NCI_293778 (B1) in Pocket 1; ligand NCI_293778 (B2) in Pocket 2; ligand NCI_37553 (C) in Pocket 3; ligand NCI_293778 (B3) in Pocket 4; and NCI_293778 (B4) in Pocket 5. Ligand NCI_293778 (B) with different conformations, B1-B4 could bind to different pockets Pocket 1, 2, 4, and 5, respectively.

Pocket 1 and Pocket 4 (Figure 3.3) were groove-like while Pocket 2 (adjacent to Pocket 1, sharing some residues with Pocket 1 such as Tyr1177 and His1222), Pocket 3 (behind) and Pocket 5 (behind) were shallow. Pocket 3 was below Pocket 1, Pocket 4 was on the opposite site behind Pocket 1, and Pocket 5 was in the rear side behind

Pocket 1. Pocket 4, containing the catalytic dyad Cys1013 and His1083, is the active site. The residues making up each pocket are presented in Table 3.4.

Table 3.4. The important residues in each pocket of the nsP2 protease with the key residues in bold involved in forming HBs and hydrophobic contacts between the protein and ligands.

Binding pocket	Residues making up the pocket
Pocket 1	Gln1039, Ala1040, Glu1043, **Lys1045**, Ala1046, Tyr1047, Gly1176, Tyr1177, Pro1191, Leu1192, Gly1193, Val1194, Asn1202, Leu1203, Glu1204, Ile1221, **His1222**, Thr1223, Asp1235, His1236, Met1238, **Lys1239**, Met1242 and Cys1233
Pocket 2	His1151, Pro1153, Val1154, **Lys1155**, Gly1156, Glu1157, Arg1158, Met1159, **Glu1160**, **Tyr1177**, Asn1178, Leu1179, **Ala1180, His1222**, Thr1223, Pro1224, **Phe1225**, **Gln1232**, Arg1260, Tyr1262, Ser1293, Thr1292 and **Thr1295**
Pocket 3	Glu1157, Arg1158, Met1159, **Trp1161**, Leu1162, Lys1165, Ile1166, **Asn1167**, Gly1254, **Ser1256**, Arg1281, Ser1282, Arg1284, Leu1286, Lys1287, Pro1288, **Pro1289, Cys1290**, Leu1300, Ser1302 and Asn1303
Pocket 4	Cys1013, Ala1046, Tyr1047, **Ser1048,** Pro1049, Glu1050, Val1051, Tyr1078, **Tyr1079**, Asn1082, His1083, **Trp1084**, Gly1090, **Lys1091**, Phe1093, Tyr1201, Asn1202, Glu1204, **Leu1205**, Gly1206, Pro1208, Ala1209, Met1238, Gln1241, Met1242, Gly1245, Asp1246 and Arg1249
Pocket 5	Asn1040, Glu1050, Leu1053, Asp1064, Leu1065, **Asp1066**, **Ser1067**, Gly1068, Leu1069, **Phe1070**, Ser1071, Lys1091, Phe1093, Asn1096, Glu1098, **Ala1099**, Ile1102, Leu1103, Lys1106, Tyr1107, **Asn1140**, Arg1141, **Arg1142**, Leu1143, Pro1144, Arg1267, **Glu1270**, **Arg1271**, Cys1274, Val1275, Arg1278, Thr1313, His1314 and Asn1317

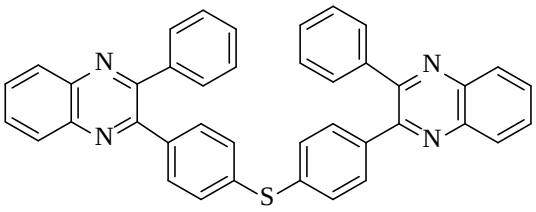
In the docking VST1, six out of ten ligands preferred binding at Pocket 3 along with two ligands at Pocket 1 and two ligands at Pocket 4. In VST2, most of the ligands (five) occupied Pocket 1, two ligands occupied Pocket 3, two ligands occupied Pocket 4, and one ligand occupied Pocket 5. In VST4 with an increased box size, larger compounds were identified as top hits despite not being present in VST3; for example NCI_293778 (-9.8 kcal/mol), NCI_84100_a (-9.2 kcal/mol), and NCI_84100_b (-9.2 kcal/mol) (their chemical structures in Figure 3.4). It is likely that the larger grid box accommodates the binding of these molecules better. Some hit compounds were present in different virtual screenings, that means they could bind into different pockets with varying conformations such as NCI_293778 (Pocket 1, 2, 3, 4, and 5); NCI_37553 (Pocket 1, 2, 3, 4, and 5); and NCI_61610 (Pocket 1, 4, and 5); NCI_670283 (Pocket 1 and 4); NCI_84100_b (Pocket 1 and 3); and NCI_37553 (Pocket 3 and 5), the chemical structures presented in Figure 3.4. The binding affinities of all hit compounds were

approximately -9 kcal/mol, which indicate favourable interactions between the compounds and the nsP2 protease.

Binding modes obtained from docking at potential binding sites were analysed to show details of interactions and to identify key residues through quantification of HBs or hydrophobic contacts. The results of the important residues involved in forming HBs and hydrophobic contact for each pocket are detailed in Appendixes 13-20. The residues surrounding ligands within 5 Å which make up each pocket are listed with the key residues involved in forming HBs and hydrophobic contacts between the protein and ligands in bold (Table 3.4). Most ligands were hydrogen bond donors, and the residues of protein were acceptors in Pocket 1 and Pocket 3 while they can exchange their roles in the other pockets. As mentioned earlier, the active site was Pocket 4 with Cys1013 and His1083 in good positions to be a catalytic dyad. The residue Trp1084 is close to His1083 in Pocket 4 which is essential for interactions, as previously reported in the study of the nsP2 protease of SINV.[346] Among the five pockets (with respect to ligands containing aromatic rings), π-stacking interactions and π-network interactions were often present with His1222 at Pocket 1; Tyr1177 at Pocket 2; and Tyr1079 and Trp1084 at Pocket 4.

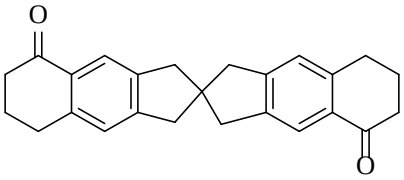
It was interesting to analyse interactions of protein and ligands to look at the different conformations and how they bind to the different pockets. Intriguingly, for the top-hit NCI_293778, a high binding affinity was observed, though analysis revealed no HBs were formed at Pocket 5 (Figure 3.5). It could not explain why at Pocket 5, this ligand gained the higher binding affinity at this pocket than those at other pockets. This might be due to the strong compromise of degree of complementarity of protein and ligands as the binding affinity does not hold all flexible complexes.

NCI_293778



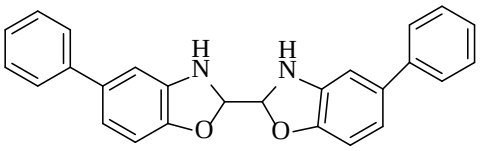
(17)

NCI_670283



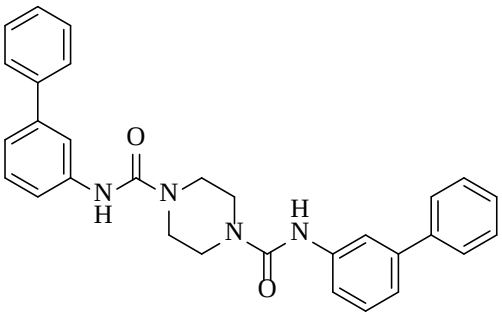
(21)

NCI_84100



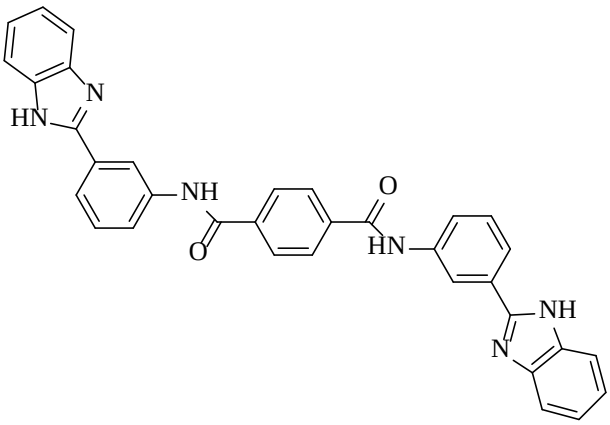
(30)

NCI_37553



(31)

NCI_61610



(19)

Figure 3.5. Structures of some top hit compounds for the nsP2.

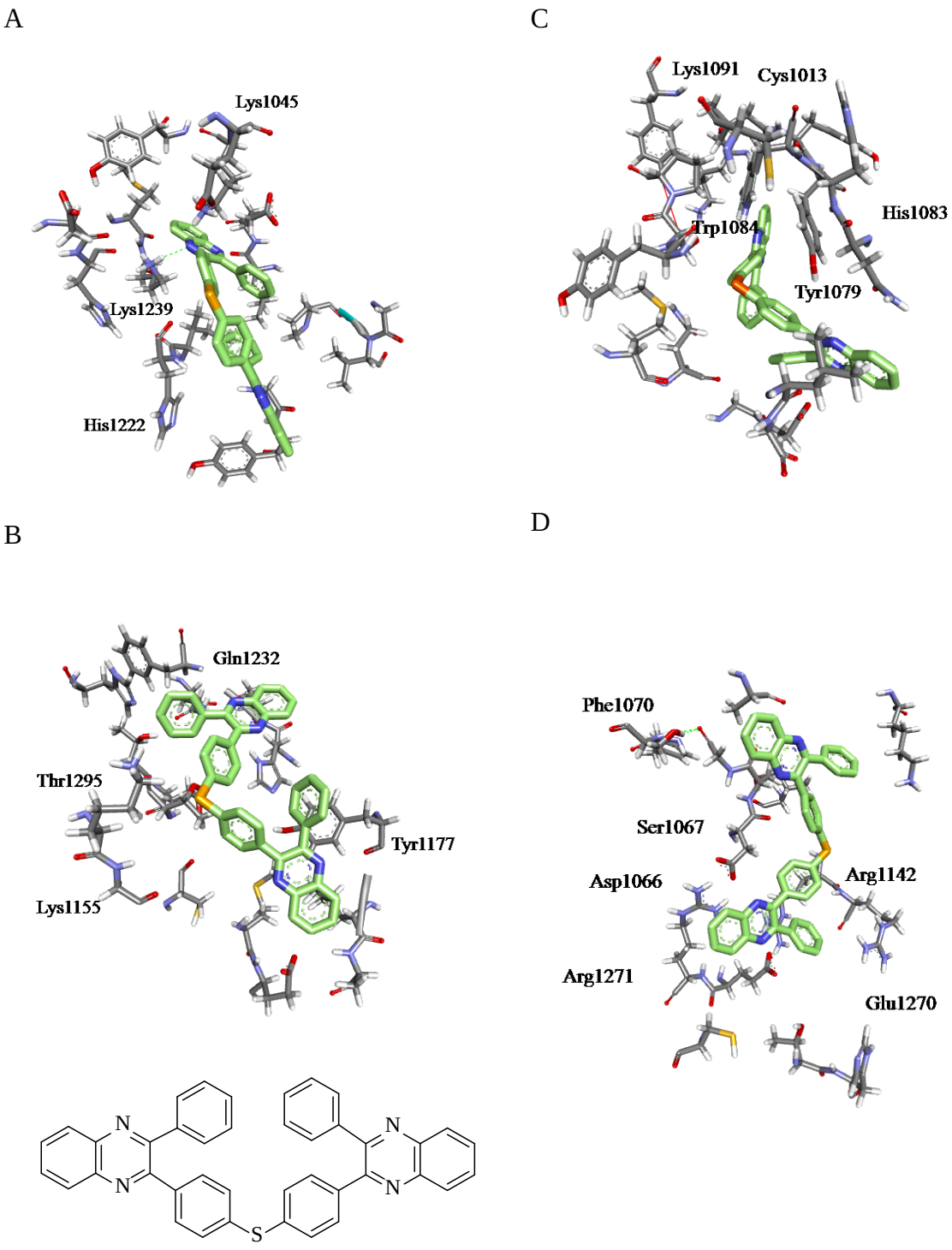


Figure 3.6. Binding poses and interactions of hit compound NCI_293778 at different binding pockets and key residues for interactions at each pocket: **(A)** At Pocket 1: HBs with Lys1239; **(B)** At Pocket 2: π-π interactions with Tyr1177; **(C)** At Pocket 4: π-π interactions with Trp1084; **(D)** At Pocket 5: Hydrophobic contact only. The ligands (in cyan) and the residues surrounding the ligands (in grey) were displayed in sticks and coloured by atoms (carbon in cyan in ligand or in grey in residues, nitrogen in blue, oxygen in red.

3.3.1.3 Identification of the active site of the nsP2 protease

It is known that the nsP2 protease carries out its function through a conserved catalytic dyad with a cysteine and a histidine; however, these residues have not been conclusively identified in the CHIKV nsP2 protease. For instance, Bassetto *et al.* reported that Cys1013 and His1083 (the numbering is according to the CHIKV nsP2 sequence) in the homology model of the nsP2 protease were the catalytic residues,[353] compared to Cys477 and His546 in the VEEV nsP2 protease through a sequence alignment between the CHIKV nsP2 protease and the VEEV nsP2 protease. In contrast, with a similarly based the sequence alignment, Singh *et al*. predicted that the active site residues interacting with the peptide substrate include Lys1045, Gly1176, His1222, and Lys1239, without explicitly identifying the catalytic residues.[351] Structural alignment of the nsP2 proteases from three alphaviruses, CHIKV (PDB id: 3TRK), VEEV (PDB id: 2HWK), and SINV (PDB id: 4GUA) were carried out with MUSTANG[354] (Appendix 21). The sequence identity between the CHIKV nsP2 protease sequence and the counterparts in VEEV and SINV sequences are 40% and 44%, respectively. This structural alignment revealed that Cys1013 and His1083 are the two catalytic residues in the nsP2 protease of CHIKV (Figure 3.6). Their catalytic roles are supported by their close proximity in the crystal structure. Cys1233 and His1222 in the C-terminal domain, identified in the homology model by Singh *et al*., are separated by 15 Å in the solved crystal structure (between the S atom in Cys and the N atom in His); thus, they are not likely to be the catalytic residues unless there is a structural transition upon substrate binding. Conversely, Cys1013 and His1083 in the N-terminal domain are sufficiently close enough to carry out proton transfer (with the distance of 4.8 Å between the S atom in Cys1013 and the N atom in His1083); thus, they are properly positioned to be a catalytic dyad. In addition, the presence of Trp1084 close to His1083 was proposed to be necessary for a functional protease.[355]

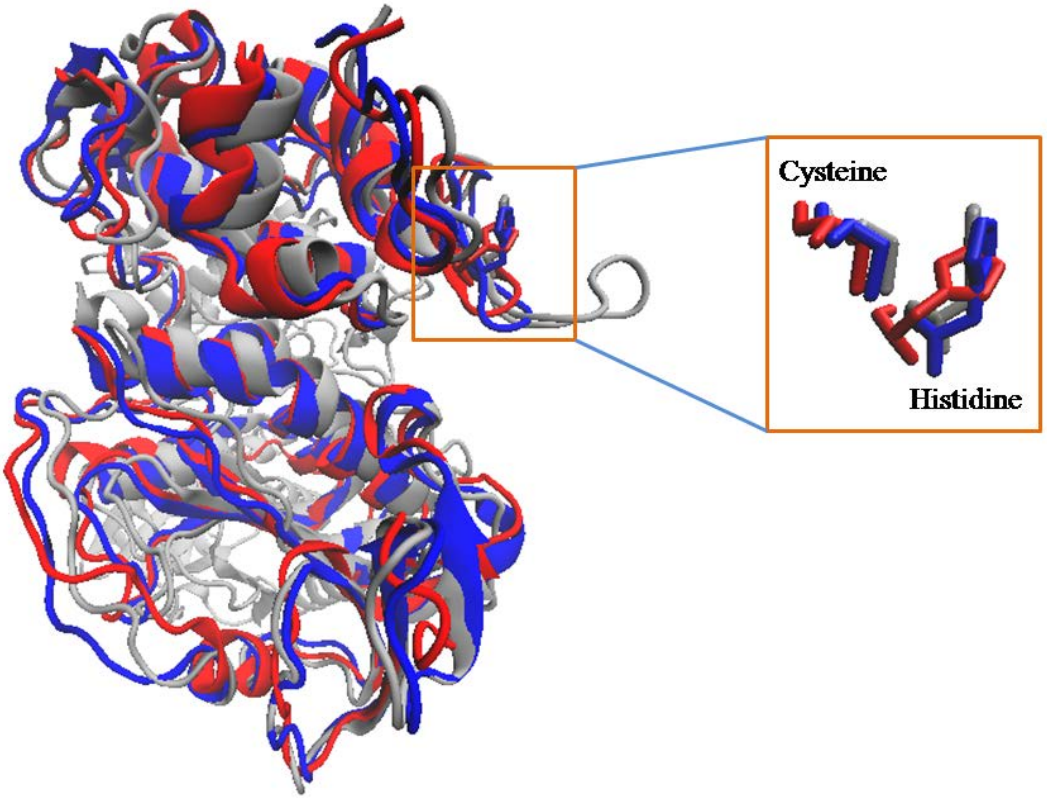
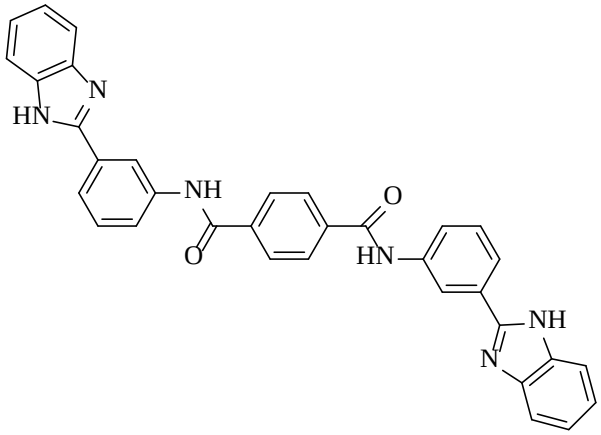
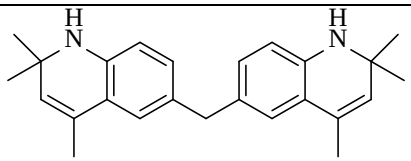
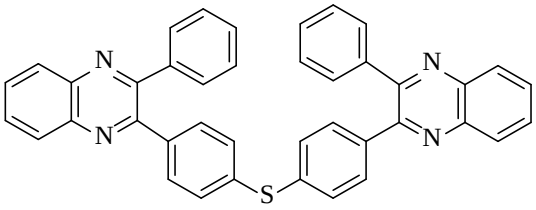
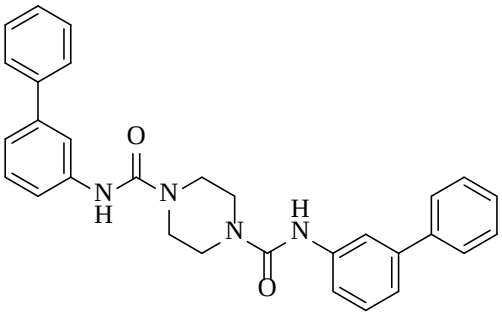
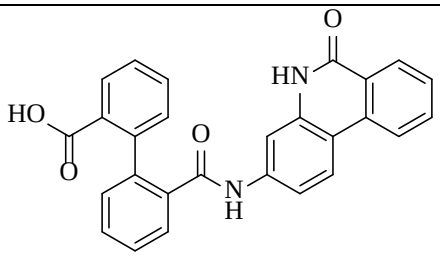


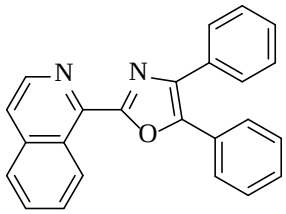
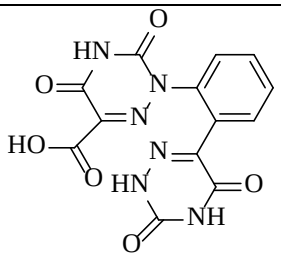
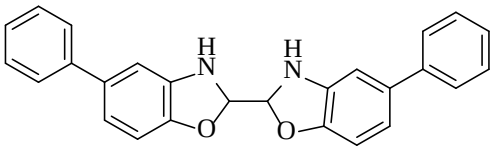
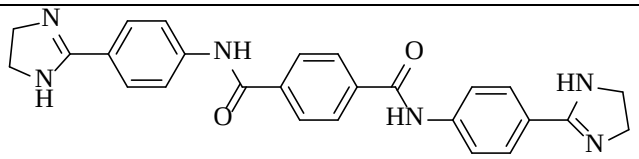
Figure 3.7. Superimposition of three crystal structures, namely the CHIKV nsP2 protease (PDB id: 3TRK, in blue), the VEEV nsP2 protease (PDB id: 2HWK, in red), and the structure of SINV (PDB id: 4GUA, in grey). The conserved catalytic residues, cysteine and histidine (in licorice), Cys1013 and His1083 in the CHIKV nsP2 protease (in blue), Cys477 and His546 in the VEEV nsP2 protease (in red); and Cys1021 and His1098 in the SINV structure (in gray) are also shown.

3.3.2 MD SIMULATIONS

The top-hit compounds (NCI_217697, NCI_61610, NCI_37553, and NCI_293778) and tenth-hit compounds (NCI_217697w, NCI_25457w, NCI_362639, NCI_84100b, and NCI_67436) obtained from virtual screenings were subjected to MD simulations using the NAMD package with the CHARMM force field (Table 3.5). In particular, ligand NCI_293778, which had different conformations at the different binding sites (NCI_293778vst4 at Pocket 1, NCI_293778vst5 at Pocket 2, NCI_293778vst2 and NCI_293778vst7 at Pocket 4, and NCI_293778vst8 at Pocket 5), was investigated. MD simulations were performed to study the stability and flexibility of the nsP2 protease and its complexes as well as to obtain the accurate binding modes. MD simulations were run for 50 ns following 3 ns equilibrium simulations.

Table 3.5. Chemical structures of hit compounds for nsP2 and their properties.

Compound's name	Structural formula	ΔG[a]	Physical properies
NCI_61610		-10.6	LogP[b]: 5.31 H-D[c]: 4 H-A[d]: 4 MW[e]: 548.60
NCI_217697		-9.3	LogP: 4.97 H-D: 2 H-A: 0 MW: 358.50
NCI_293778		-9.8[f], -10.4[g] and -10.2[h]	LogP: 10.87 H-D: 0 H-A: 4 MW: 594.73
NCI_37553		-9.6	LogP: 5.03 H-D: 2 H-A: 2 MW: 476.57
NCI_127133		-8.8	LogP: 4.52 H-D: 3 H-A:4 MW: 434.40

NCI-25457		-9.0	LogP: 5.37 H-D: 0 H-A: 3 MW: 348.39
NCI_36239		-8.2	LogP: -0.68 H-D: 4 H-A: 8 MW: 344.24
NCI_84100_b **(**$C_{26}H_{20}N_2O_2$**)**		-8.9	LogP: 6.30 H-D: 2 H-A: 2 MW: 392.15
NCI_67436		8.1	LogP: 2.96 H-D: 4 H-A: 4 MW: 452.52

(a) ΔG refers to the binding affinities by Vina in kcal/mol. **(b)** The logarithm of the partition coefficient between n-octanol and water calculated using ChemBioDraw Ultra 14.0. **(c)** Number of hydrogen bond donor atoms. **(d)** Number of hydrogen bond acceptor atoms. **(e)** Molecular weight. **(f)** Located in Pocket 1 and Pocket 2. **(g)** Located in Pocket 4. **(h)** Located in Pocket 5.

3.3.2.1 Overall stability of the nsP2 protease and its complexes

The overall stability of the nsP2 protease and its complexes was evaluated by the values of backbone atomic positional RMSD. The results of backbone RMSD profile for the nsP2 enzyme and its complexes with different ligands with respect to the starting structure are showed in Figure 3.7.

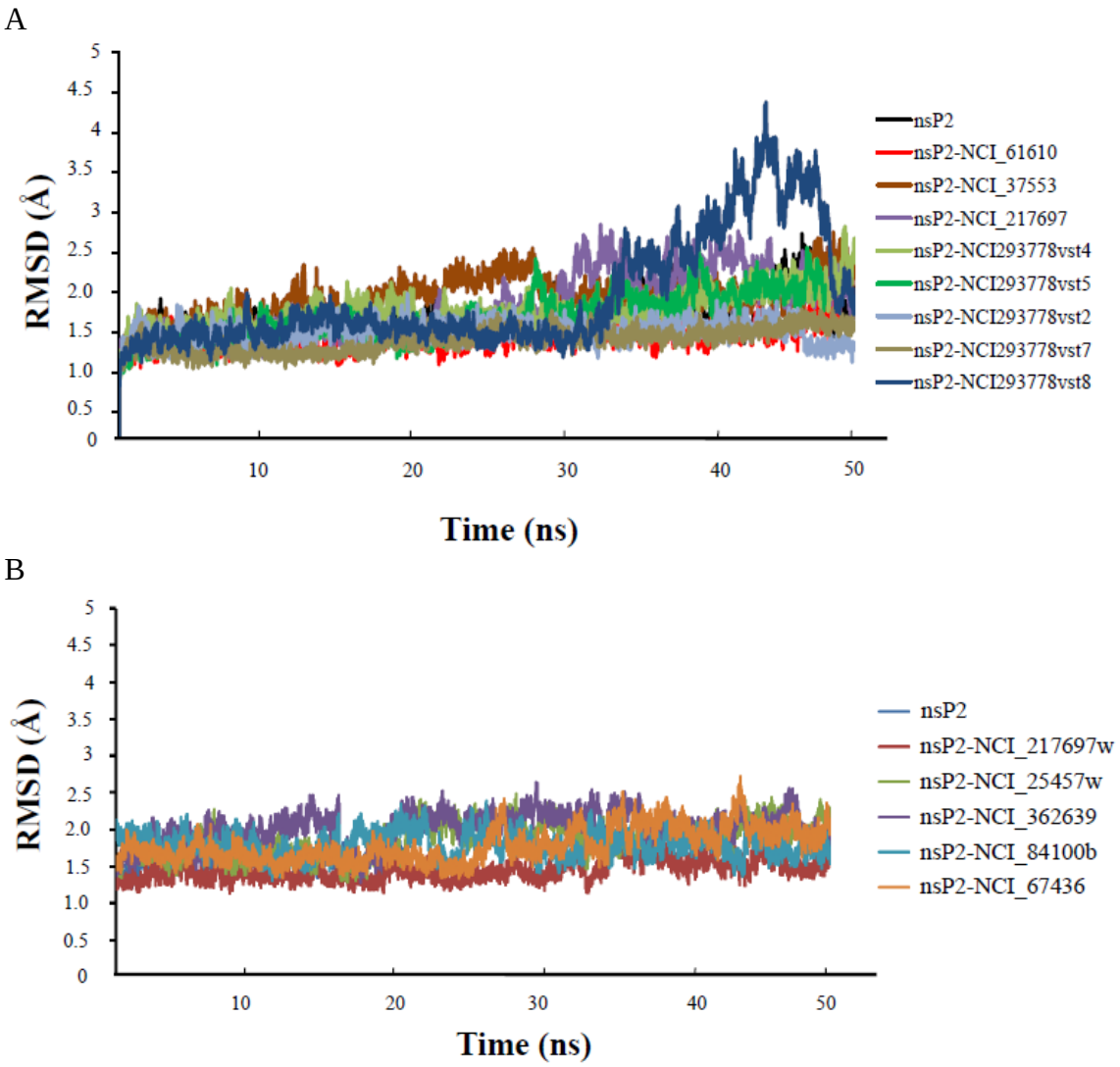


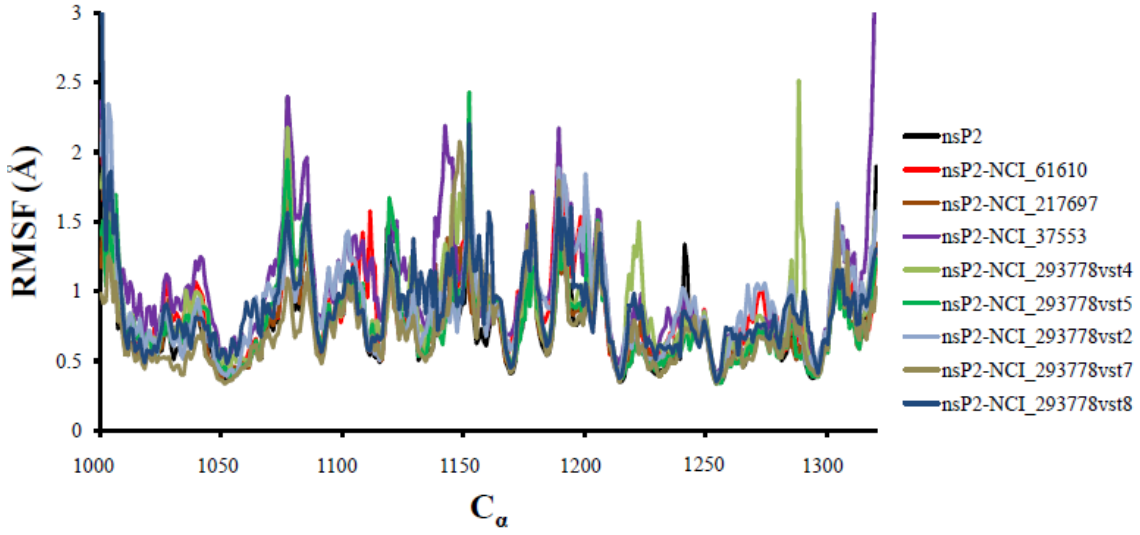
Figure 3.8. The backbone RMSD profiles for the apo protein nsP2 and its different complexes during MD simulations from 4 ns to 53 ns: **(A)** Complexes of the nsP2 and top-hit compounds; **(B)** Complexes of the nsP2 and tenth-hit compounds.

The plots showed that most of the systems were relatively stable during the 50 ns simulations within 1-2.5 Å. In simulation of the complex of ligand NCI_293778vst8 at Pocket 5, a large RMSD was observed around 47 ns (Figure 3.7 A). Examination of the structures revealed that it was due to the relative movement between two domains and the fluctuation of loops. This ligand was also found to gradually move from a position whereby part of the ligand was positioned out of the pocket, to the whole ligand being out of the pocket but still stuck at the rear after 40 ns, and finally dissociated from the pocket after 47 ns.

3.3.2.2 Investigation of flexibility of the nsP2 protease and its complexes

For the flexibility of the nsP2 protease and its complexes, the RMSF curves of the Cα atoms in the nsP2 protease and its complexes were calculated, displayed in Figure 3.8.

A



B

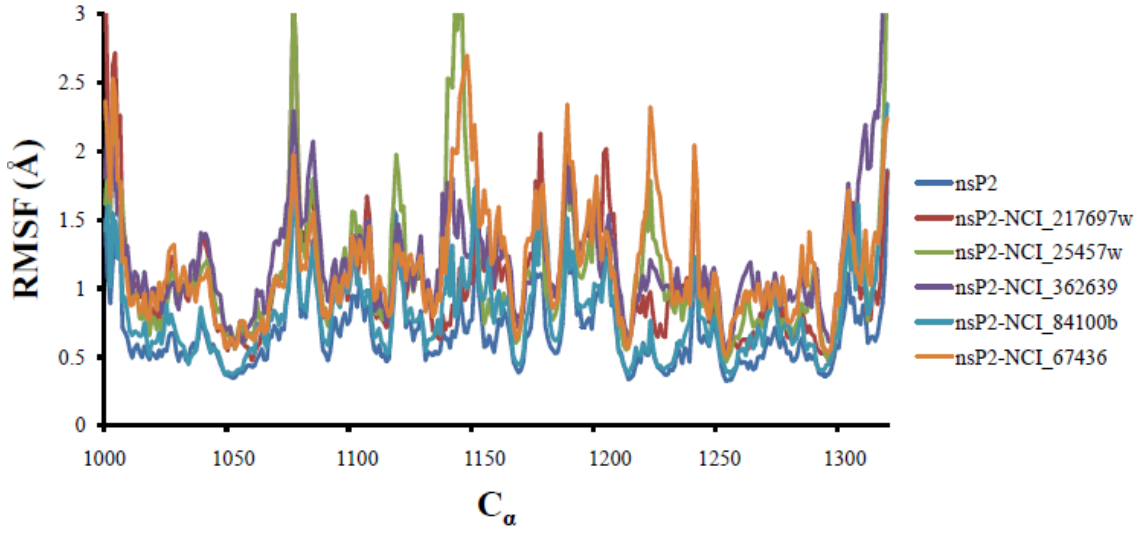


Figure 3.9. RMSFs values of C_α atoms of the apo protein nsP2 and its different complexes during MD simulations: **(A)** Complexes of the nsP2 and top-hit compounds; **(B)** Complexes of the nsP2 and tenth-hit compounds.

For the flexibility, comparative analysis of the RMSF values of the apo protein and the complexes was focused on the binding sites of ligands. Most of the residues making up the binding sites were quite stable during the simulation with fluctuations within 2 Å.

Furthermore, no significantly reduced fluctuations were found in the binding pockets of the protein between the nsP2 protease (apo protein) and its complexes with the nsP2 protease-ligand.

3.3.2.3 Atomic interaction between the nsP2 protease and hit compounds

Details of atomic interactions were obtained by HBs and hydrophobic contacts. The residues involved in forming hydrophobic contacts based on the trajectories of the MD simulations are listed in Table 3.6 and Table 3.7.

Table 3.6. Hydrogen bonding analyses on the trajectories sampled in the MD simulations for hit compounds in complexed with the nsP2 protease. The HBs with occupancy more than 10% are in highlighted in bold.

Ligands	Details of HBs	% occupancy
NCI_61610	Lys1239(HZ1)-N1	7.5
	Lys1239(HZ2)-N1	6.0
	Lys1239(HZ3)-N1	5.6
	His1222(HN)-O1	4.5
	Glu1204(HN)-O	3.1
	Lys1239(HZ3)-O	2.9
	Lys1239(HZ2)-O	2.7
	Lys1239(HZ1)-O	2.6
	His1222(HN)-O	1.8
	Lys1045(HZ3)-N1	0.3
	Lys1045(HZ1)-N1	0.3
	Lys1045(HZ2)-N1	0.3
	Lys1045(HZ2)-O	0.1
	Gly1176(HN)-O1	0.1
NCI_217697	Gln1039(OE1)-H	0.2
	Glu1043(OE2)-H	0.1
	Leu1193(O)-H1	1.7
	Asn1202(OD1)-H1	0.3
	Glu1204(OE1)-H	0.5
	Glu1204(O21)-H	0.4
NCI_37553	**Arg1284(HH12)-O1**	**80.5**
	Arg1284(HH12)-N1	**26.3**
	Ser1302(HG1)-O1	**10.4**
	Lys1165(HZ2)-O	7.2
	Lys1165(HZ1)-O	2.9
	Lys1165(O)-H1	1.7
NCI_293778vst4	**His1222(HD1)-N3**	**10.7**
	Asn1202(HD21)-N3	0.5
	Lys1239(HZ1)-N	0.1
	Lys1239(HZ2)-N	0.1

	Lys1045(HZ2)-N3	0.1
	Lys1045(HZ3)-N3	0.1
	Leu1192(HN)-N2	0.1
NCI_293778vst5	Gly1156(HN)-N	1.4
	Lys1155(HZ2)-N	0.8
	Lys1155(HZ1)-N	0.5
	Lys1155(HZ3)-N	0.5
	Met1159(HN)-N	0.1
NCI_293778vst2	Tyr1079(HH)-N3	8.6
	Asn1082(HD22)-N2	5.8
	Trp1084(HE1)-N3	3.4
	Asn1082(HD21)-N3	3.6
	Asn1082(HD21)-N2	0.4
	Tyr1079(HN)-N1	0.2
	Leu1207(HN)-N2	0.1
NCI_293778vst7	Ser1048(HG1)-N2	0.6
	Tyr1079(HH)-N	0.1
	Asn1082(HD21)-N1	0.1
NCI_293778vst8	Ser1067(HG1)-N2	0.2
	Arg1105(HE)-N3	0.2
	Lys1106(HZ2)-N2	0.1
	Lys1106(HZ3)-N2	0.1
	Arg1141(HH11)-N3	0.1
NCI_217697w	Glu1043(OE1)-H	1.9
	Glu1043(OE2)-H	2.7
	Leu1192(O)-H1	2.9
	Asn1202(OD1)-H1	6.6
	Glu1204(OE1)-H	1.6
NCI_25457w	Lys1165(HZ1)-N	1.8
	Lys1165(HZ1)-N	1.4
NCI_362639	His1152(HD1)-N1	0.5
	His1152(NE2)-H2	0.5
	Pro1153(O)-H1	1.3
	Val1154(HN)-O4	1.2
	Phe1224(HH)-O2	1.3
NCI_84100b	**Leu1162(O)-H1**	**28.0**
	Trp1161(O)-H1	3.3
	Lys1165(O)-H1	4.9
	Arg1284(HH12)-O1	7.3
NCI_67436	**Asn1140(HD22)-N1**	**65.3**
	Arg1142(HN)-O	**63.4**
	Glu1270(OE1)-H2	**49.2**
	Glu1270(OE2)-H2	**49.5**
	Cys1274(HG1)-N5	**28.2**
	His1314(HD1)-O1	**22.2**
	Asn1317(HD21)-O1	**26.1**

Table 3.7. Hydrophobic contact analyses on the trajectories sampled in the MD simulations for hit compounds in complexed with the nsP2 protease.

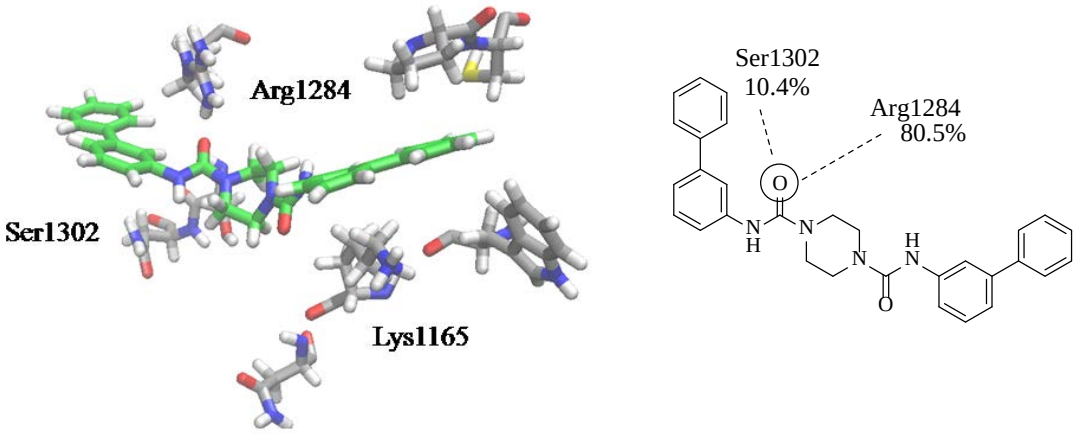
Ligand	Interacting non-polar parts of the residues
NCI_61610	Ala1040, Tyr1177, Pro1191, Leu1203, Ile1221, Leu1243
NCI_217697	Pro1191, Leu1243
NCI_37553	Trp1161, Leu1162, Leu1286, Pro1288, Pro1289, Leu1300
NCI_293778vst4	Pro1191, Leu1192, Leu1203, Ile1221, Leu1243
NCI_293778vst5	Tyr1177, Pro1224
NCI_293778vst2	Ala1046, Tyr1079, Trp1084, Leu1205
NCI_293778vst7	Ala1046, Val1077, Tyr1079, Trp1084, Leu1205
NCI_293778vst8	Ile1102, Leu1103
NCI_217697w	Ala1040, Lys1045, Tyr1047, Pro1191, Leu1192, Gly1193, Val1194, Glu1204, Ile1221, Hsd1222,, Met1238, Lys1239, Leu1243
NCI_25457w	Glu1157, Trp1161, Leu1162, Lys1165, Leu1285, Pro1289
NCI_362639	Thr1292, Ser1293
NCI_84100b	Glu1157, Trp1161, Leu1162, Lys1165, Ile1166, Leu1286, Pro1288, Pro1289, Leu1300
NCI_67436	Asp1066, Ser1067, Asn1140, Arg1141, Arg1142, Pro1144, Hsd1145, Glu1270, Arg1271, Cys1274, Val1275

For most of the investigated hit compounds, the ligand or/and the residues in the binding sites fluctuate and adapt their structure in order to better accommodate the ligands by optimizing HBs, and/or hydrophobic contacts, compared to those in docking where protein was kept rigid. Most of their interactions with the nsP2 protease were maintained by hydrophobic contacts rather than HBs, as most of the HBs occupancy was found to be low (less than 10%, Table 3.6).

Considering the structure of the ligands, the presence of both donor and acceptor atoms in the ligand, NCI_61610 and NCI_37553 could form more HBs than ligands NCI_217697 and NCI_293778. As expected, HBs were mostly formed with NCI_61610 and NCI_37553. Ligand NCI_61610 showed HBs interactions with residues Lys1239, Glu1204, Leu1203, Tyr1177, Gly1176, and Lys1045, although these interactions were weak (occupancies less than 10%). The complex nsP2 protease-ligand NCI_37553 at Pocket 3 was maintained through strong HBs between its oxygen and Arg1284 (80.5%), Ser1302 (10.4%); or between its nitrogen and Arg1284 (26.3%), illustrated in Figure

3.9. Ligand NCI_67436 fitted in Pocket 5 by moderate HBs interactions during the simulations with residues of the enzyme such as Asn1140 (65.3%), Arg1142 (63.4%), Glu1270 (49.2%), and some weak HBs with Cys1274 (28.2%), His1314 (22.2%), and Asn1317 (26.1%), (Figure 3.9). In particular, this ligand plays the roles of donor and acceptor in forming the HBs. The interactions of ligand NCI_84100b and the nsP2 enzyme were maintained through weak HBs with Leu1162 (28%).

A



B

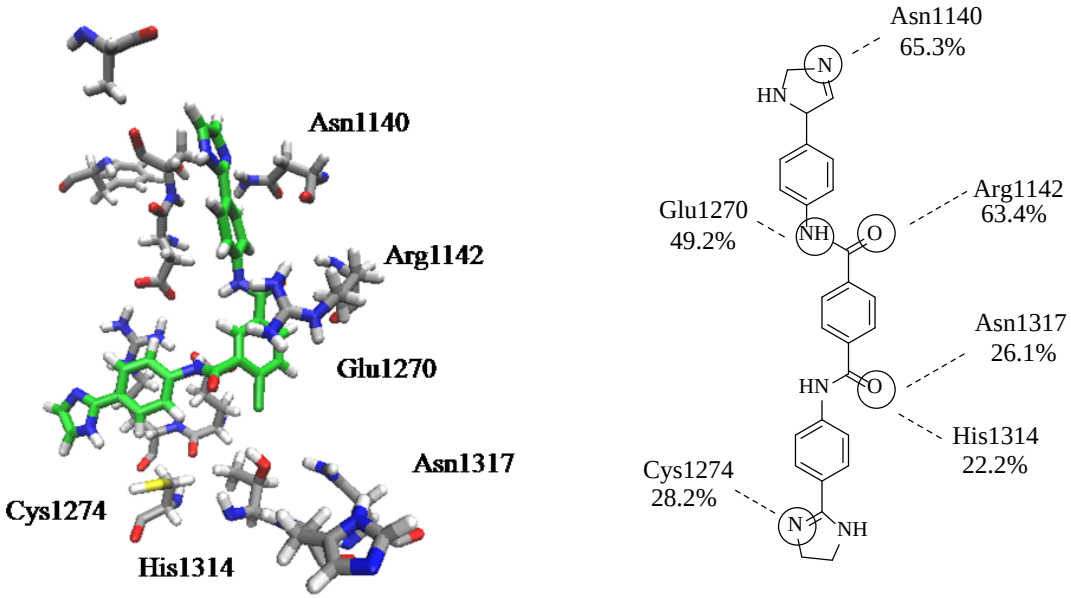


Figure 3.10. Hydrogen bonding interactions between the nsP2 enzyme and ligands: **(A)** Ligand NCI_37553 at Pocket 3, **(B)** Ligand NCI_67436 at Pocket 5; with representation of ligands and key residues for interactions surrounding the ligands (in licorice), showing the interactions maintained between the ligands and residues of protein through strong HBs interactions.

Interestingly, a comparison of NCI_293778 bound at different binding sites revealed that the nitrogen of this ligand could form one hydrogen bond (occupancy of 10.7%) with the residue His1222 at Pocket 1, accompanied by surrounding hydrophobic interactions (Figure 3.10). In contrast, NCI_293778 often lacked strong HBs at other pockets (with low occupancies, see Table 3.6). This indicates that NCI_293778 was more likely to be accommodated by Pocket 1 than other pockets, even though that the binding affinity obtained for this ligand was -9.8 kcal.mol; less than for Pocket 4 (-10.4 kcal/mol), and Pocket 5 (-10.2 kcal/mol). However, the aromatic ring of this ligand could form π-stacking interactions or a π-network with Tyr1177 (Pocket 2); Tyr1079, and Trp1084 (Pocket 4) to maintain the interactions between the protein nsP2 protease and the ligand.

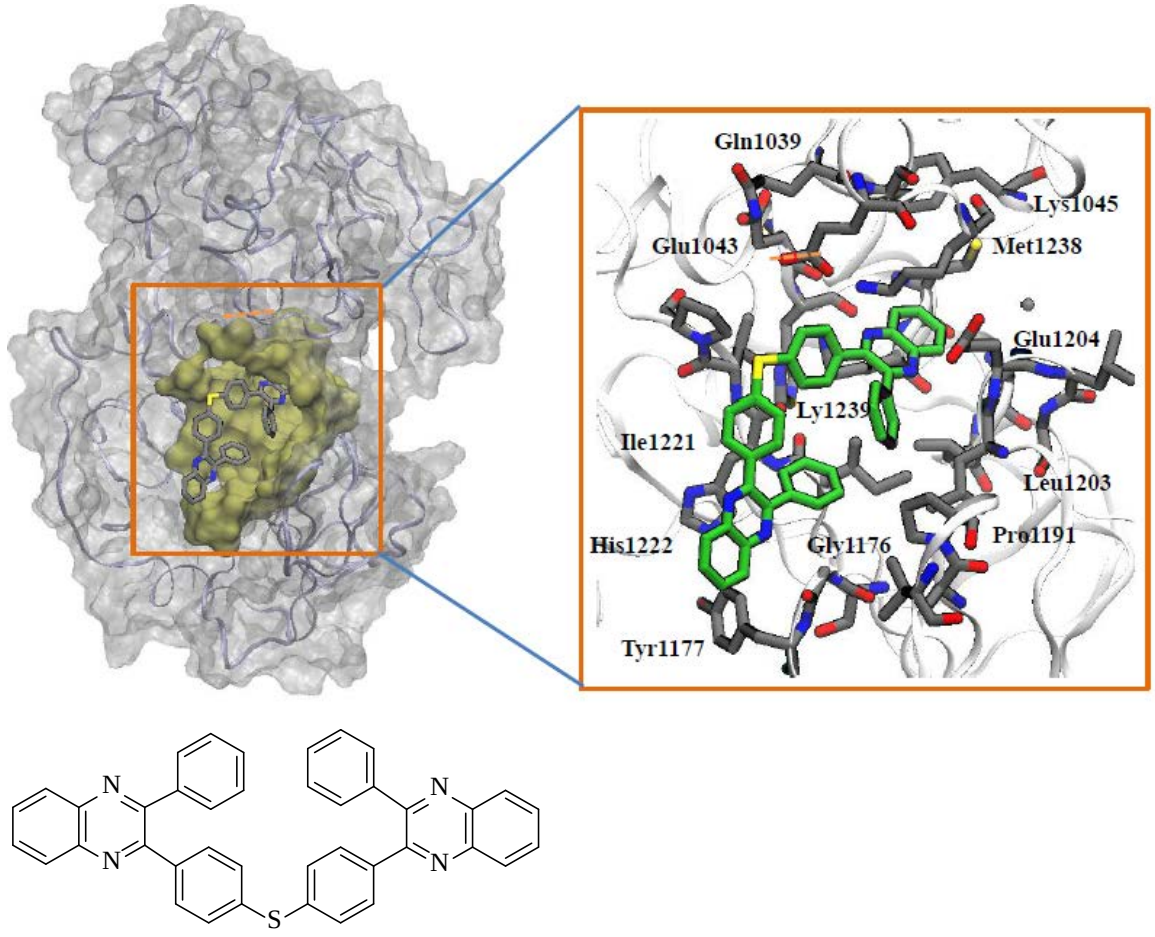


Figure 3.11. The docked structure of ligand NCI_293778vst4 (in green) in Pocket 1 of the nsP2 protease, showing the residues forming Pocket 1 and the ligand-protein interactions (in grey).

Compound NCI_67436 played both donor and acceptor roles in the HBs interactions. For the other ligands, there was no significant contribution from HBs in the interactions, even though those ligands have donor and acceptor groups available for hydrogen bonding formations.

The SASA used to monitor the possible solvation environment change upon ligand binding was expected to get a reduction of hydrophobic interacting residues in the complex protein-ligand compared to those in apo protein (Appendix 22). As mentioned in Chapter 2, section 2.3.3.3, the SASA is heavily influenced by interactions between protein and ligands. However, most of the involved residues were inconsistent for different ligands in the nsP2 protein. This could be explained by the formation of not only hydrophobic contacts but also polar hydrogen bonding interactions. Thus, the change in SASA will be compromised by the polar interactions, and both of protein and ligands will modulate the SASAs.

3.3.2.4 Clustering analysis, comparison of simulations and docking results for the complexed nsP2 protease-ligands

Throughout the simulations, structural clustering was used to identify the most popular conformation, and the interaction between these clusters and the nsP2 at the different binding sites was examined to compare with docking. Clustering analysis was undertaken on all of the snapshots from the trajectories, and the clusters were visualized and superimposed with the initial structure. The superimposition and the value of RMSD were used to evaluate between the different clustering structures and initial structure at 0 ns, 10 ns, 20 ns, 30 ns, 40 ns, and 50 ns were calculated. For example, the potential hit compound NCI_37553, superimposition of the most popular conformations of complex; and the RMSD value for the ligand was 1.7 Å (at 10 ns), 2.0 Å (at 20 ns), 2.0 Å (at 30 ns), 1.9 Å (at 40 ns), and 2.2 Å (at 50 ns) with respect to the initial structure (Figure 3.11). The docking results were in close agreement with MD simulations, even though there was flexibility of residues in the protein and ligands in forming HBs or creating hydrophobic contacts. There was no major change in the important residues at the binding sites of the nsP2. Moreover, understanding the effects of ligand binding to other binding pockets on the active site was also of interest. Experimental validation is

required to confirm whether there is an allosteric site or a synergistic effect upon binding at different sites within the active site of the nsP2.

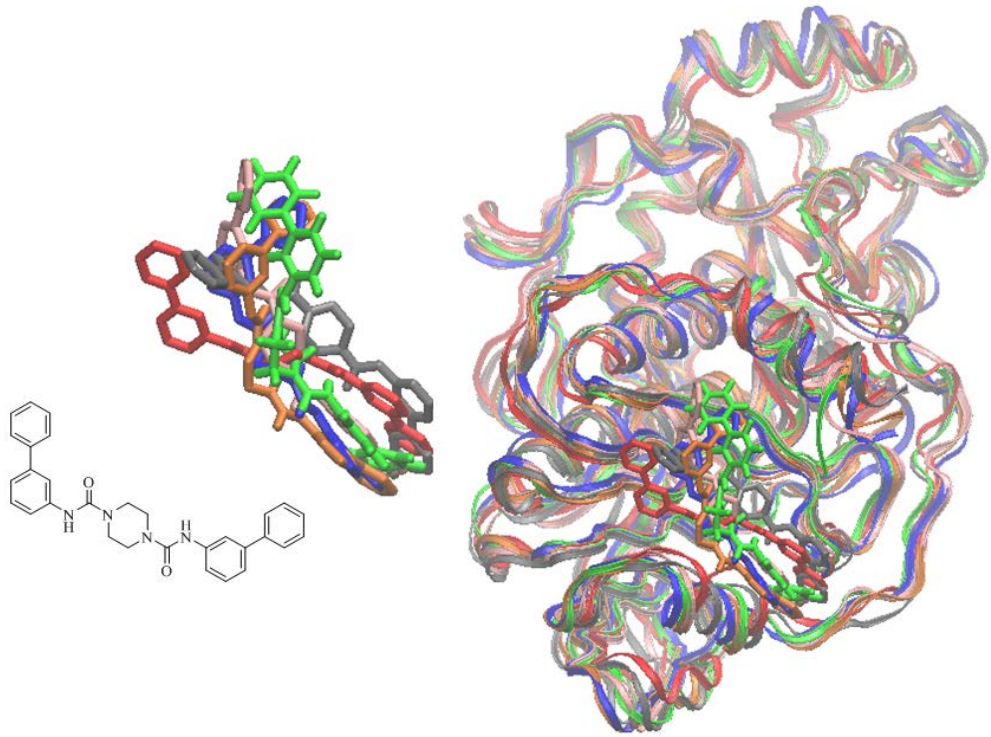


Figure 3.12. Superimposition of the different conformations of ligand NCI_37553 and complexed with the nsP2 at Pocket 3 of the nsP2 during simulations with respect to the initial structure (red: at 0 ns, grey: at 10 ns, green: at 20 ns, pink: at 30 ns, orange: at 40 ns, and blue: at 50 ns).

3.3.2.5 Combination of simulations and docking for the nsP2

Based on MD simulations results, the sampled conformations of protein nsP2-NCI_67436 at the different timepoints (10 ns, 20 ns, 30 ns, 40 ns, and 50 ns) were used for docking and virtual screening, so as to probe the effects of the static protein structure used in the docking. This complex was selected as NCI_67436 showed higher potential after analyzing the outcomes of docking and simulations, listed in Table 3.8 and Table 3.9. The binding affinities of the docking runs in Table 3.8 were not significantly different from those obtained by docking based on the X-ray structure. A comparison of complexes of different protein conformations and the initial docked structure showed there were changes in the binding mode, but always maintaining HBs of the ligand with the protein to maintain the interactions of ligand and the residues in the binding pocket during simulations (details of interactions analyses in Appendixes 23-27). For virtual

screening, the docking results were sensitive in the exact static structure used as most of the hits introduced were new in comparison of top hits lists in the screening results based on the X-ray structure.

Table 3.8. Re-docking results for complex nsP2-NCI_67436 with different conformations of the nsP2 taken from the different timepoints of simulations at Pocket 5. The binding affinity is in kcal/mol.

nsP2 conformation	Binding affinity (kcal/mol)	HBs interaction (Å)
At 0ns	-8.1	No HBs (hydrophobic contacts only)
At 10 ns	-7.7	No HBs (hydrophobic contacts only)
At 20 ns	-8.3	Arg1142(HH11)-O=2.3
At 30 ns	-7.0	H-Arg1267(O)=2.2 Arg1267(HE)-N=2.4 Leu1276(HN)-O=2.1 H-Thr1268(O)=2.4
At 40 ns	-6.8	Arg1278(HH11)-N=2.2
At 50 ns	-7.1	Arg1267(HH21)-N=2.4

Table 3.9. Virtual screening results for blind docking into Pocket 1 with different conformations of the nsP2 taken from the different timepoints of simulations. The binding affinities ΔG are in kcal/mol.

VST-10ns[a]	**VST-20ns[b]**	**VST-30ns[c]**	**VST-40ns[d]**	**VST-50ns[e]**
1. NCI_293778 (-8.6)	1. NCI_345647_a (-9.0)	1. NCI_293778 (-8.3)	1. NCI_293778 (-8.5)	1. NCI_61610 (-8.2)
2. NCI_37553 (-8.5)	2. NCI_338042 (-8.8)	2. NCI_345647_a (-8.0)	2. NCI_61610 (-7.9)	2. NCI_84100_a (-7.2)
3. NCI_59620 (-8.4)	3. NCI_293778 (-8.7)	3. NCI_97920 (-8.0)	3. NCI_122819_b (-7.7)	3. NCI_146771 (-7.2)
4. NCI_7524_b (-8.3)	4. NCI_67436 (-8.3)	4. NCI_ 327702 (-7.9)	4. NCI_671136 (-7.6)	4. NCI_84100_b (-7.1)
5. NCI_308835 (-8.2)	5. NCI_80997_a (-8.3)	5. NCI_156516 (-7.9)	5. NCI_19990_a (-7.6)	5. NCI_25457 (-7.0)
6. NCI_227186_a (-8.0)	6. NCI_338963 (-8.2)	6. NCI_80313 (-7.8)	6. NCI_37641_b (-7.5)	6. NCI_116702 (-7.0)
7. NCI_328101 (-8.0)	7. NCI_319990 (-8.2)	7. NCI_116709_b (-7.8)	7. NCI_269148_b (-7.4)	7. NCI_37553 (-7.0)
8. NCI_339161 (-7.8)	8. NCI_25457 (-8.0)	8. NCI_37553 (-7.8)	8. NCI_319990 (-7.4)	8. NCI_80731 (-7.0)
9. NCI_270738_b (-7.8)	9. NCI_670283 (-7.9)	9. NCI_326757 (-7.6)	9. NCI_156219_b (-7.4)	9. NCI_55691 (-7.0)
10. NCI_58052 (-7.8)	10. NCI_326182 (-7.9)	10. NCI_154585 (-7.6)	10. NCI_59620_b (-7.3)	10. NCI_332670 (-7.0)

(a) In VST-10ns[a], the grid box is fixed at the centre of Pocket 3 (15.2 Å, 13.9 Å, 11.5 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(b)** In VST-20ns[b], the grid box is fixed at the centre of Pocket 3 (15.2 Å, 13.9 Å, 11.5 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(c)** In VST-30ns[c], the grid box is fixed at the centre of Pocket 3 (15.2 Å, 13.9 Å, 11.5 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(d)** In VST-40ns[d], the grid box is fixed at the centre of the protein at the centre of Pocket 3 (15.2 Å, 13.9 Å, 11.5 Å) with a dimension of 20 Å × 20 Å × 20 Å. **(e)** In VST-50ns[e], the grid box is fixed at the centre of the protein at the centre of Pocket 3 (15.2 Å, 13.9 Å, 11.5 Å) with a dimension of 20 Å × 20 Å × 20 Å.

3.3.3 BINDING FREE ENERGY CALCULATIONS FOR THE LIGANDS BINDING TO THE NSP2 ENZYME

The prediction of binding free energy is usually a challenge for docking and scoring in computational drug design. In this study, the converged trajectories of complexes obtained from simulations were used to calculate the binding free energy for the hit compounds. The average values of vdW and elec interactions for each ligand in the two states, bound state (V_{bound}) and unbound state ($V_{unbound}$) were calculated (Table 3.10). The results in Table 3.10 indicated that the van der Waals value of the ligands in the complex with the enzyme were much more negative than when only in solution, showing that the ligand has more favourable van der Waals interactions in the complex. However, it is striking that the electrostatic interactions of ligands in all the complexes were significantly less favourable than those of the ligand in water.

Parameters selection proceeded with the same approach as for the nsP3 (see Chapter 2, section 2.3.4 for explanation of selecting parameters). The results of ΔG^2 with α set to 1.043 are shown in Table 3.10 and revealed better agreement with the binding affinity obtained from docking. This provided good agreement when comparing the binding free energies for different ligands at the same pocket; such as in the case of NCI_61610, NCI_217697, and NCI_293778vst4. For instance, ligand NCI_293778vst4, with the presence of stronger HBs (indicated by the higher value of HBs occupancy) had higher binding affinities compared to other ligands.

A comparison of binding free energies of top-hit compound complexes and tenth-hit compound complexes was also carried out. The results revealed that most of the top-hit compounds had better binding free energies than the tenth-hit compounds, which supported a close agreement between docking results and MD simulations results (Table 3.10). Ligand NCI_67436 showed the highest binding affinities compared to the value obtained from docking and also compared to the other compounds; while the ligand NCI_362639 had the lowest binding affinity. This is in accordance with the simulations results analysed above, as the ligand NCI_67436 fit well in the pocket by forming strong HBs with the residues of the pocket.

Table 3.10. Average binding free energies (kcal/mol) of top-hit compounds and tenth-hit compounds for the nsP2 calculated by LIE method using data trajectories from the MD simulations: ΔG^1 (in kcal/mol using $\alpha = 0.18$, $\beta = 0.43$, and $\gamma = 0$) or ΔG^2 (in kcal/mol using $\alpha = 1.043$, $\beta = 0.43$, and $\gamma = 0$).[328] ΔG is the predicted binding affinity by Vina (in kcal/mol).

Compound	V^{vdw}_{bound}	$V^{vdw}_{unbound}$	V^{elec}_{bound}	$V^{elec}_{unbound}$	ΔvdW	ΔElec	ΔG^1	ΔG^2	ΔG
NCI_61610	-65.0	-54.4	-79.2	-91.0	-10.6	11.8	3.2	-6.0	-10.6
NCI_217697	-50.0	-39.1	-30.4	-36.2	-10.9	5.8	0.5	-8.9	-9.3
NCI_37553	-60.1	-48.6	-62.5	-73.1	-11.5	10.6	2.5	-7.4	-9.6
NCI_293778vst4	-69.6	-54.6	-35.6	-48.0	-15.0	12.4	2.6	-10.3	-9.8
NCI_293778vst5	-68.5	-54.4	-42.3	-47.8	-14.1	5.5	-0.2	-12.3	-9.8
NCI_293778vst2	-69.2	-54.7	-38.9	-47.7	-14.5	8.8	1.1	-11.3	-10.3
NCI_293778vst7	-72.7	-55.6	-40.9	-48.2	-17.1	7.3	0.1	-14.6	-10.4
NCI_293778vst8	-65.6	-54.4	-43.2	-47.4	-11.2	4.2	-0.2	-9.9	-10.2
NCI_217697w	-50.6	-39.1	-29.7	-36.4	-11.4	6.7	0.8	-9.0	-9.1
NCI_25457w	-46.5	-35.4	-20.5	-32.0	-11.1	11.5	2.9	-6.6	-9.0
NCI_362639	-28.7	-26.8	-76.9	-78.5	-1.9	1.6	0.4	-1.3	-8.2
NCI_84100b	-51.4	-40.8	-30.5	-42.8	-10.6	12.3	3.4	-5.7	-8.9
NCI_67436	-62.7	-44.4	-81.3	-89.7	-18.3	8.4	0.3	-15.5	-8.1

However, between the binding affinities obtained from docking and the binding free energies estimated from simulations, three ligands, NCI_61610, NCI_217697, and NCI_37553 showed the lower binding free energies in simulations: -6.0 kcal/mol, -8.9 kcal/mol, -7.4 kcal/mol, respectively compared these in docking: -10.6 kcal/mol, -9.3 kcal/mol, -9.6 kcal/mol, respectively. Interestingly, in the case of NCI_293778, with different conformations binding at the different binding sites, the binding affinities in docking and from simulations indicated a consistent trend that binding at Pocket 4 (-14.6 kcal/mol) better than Pocket 2 (-12.3 kcal/mol) > Pocket 1 (-10.3 kcal/mol) > Pocket 5 (-9.9 kcal/mol); whereas in analysis simulations above, with a presence of HBs, Pocket 1 may be as the better place for this ligand compared to others. In addition, HBs and hydrophobic contacts in some ligands (such as NCI_293778vst4 and NCI_37553) were maintained. These ligands were expected to have lower binding free energies, however this was not observed with LIE predictions. It may be deduced that there was a compromise between van der Waals and electrostatic results from HBs interactions and hydrophobic contacts between the ligands and the enzyme nsP2. More work and results from experimental data are required to clarify the issue.

3.3.4 ANALYSIS AND SELECTING LEADS FOR BIOLOGICAL TESTING

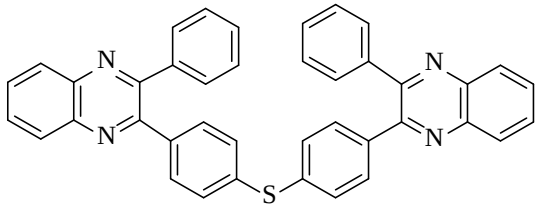
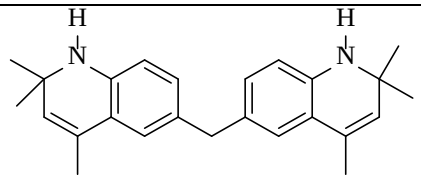
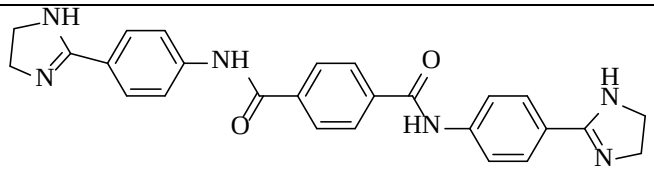
The outcomes of dockings and virtual screening showed the potential compounds to be considered are NCI_61610 (-10.6 kcal/mol), NCI_293778 (from -9.8 to around -10.2 kcal/mol), NCI_37553 (-9.6 kcal/mol), NCI_217697 (-9.3 kcal/mol), and NCI_25457 (-9.0 kcal/mol).

After analyzing MD simulations and binding free energy calculations, the compounds to be selected as leads are NCI_293778 (from -9.9 to around -13 kcal/mol), NCI_67436 (-15.5 kcal/mol), NCI_217697 (-9.0 kcal/mol) as they have higher binding affinities than the others and the interactions were maintained during MD simulations.

From a combination of these two outcomes (from simulations and docking), a list of compounds which have good binding affinities and the remained interactions

between them and the nsP2 protein during MD simulations, was proposed for biological testing; namely NCI_293778, NCI_217697, and NCI_67436 (Table 3.11).

Table 3.11. Potential lead compounds for the nsP2 proposed for biological testing.

No	Compounds	Chemical structure
1	NCI_293778	
2	NCI_217697	
3	NCI_67436	

3.4 CONCLUSIONS

In this work, the X-ray crystal structure of nsP2 protease was utilized to conduct a combination of molecular docking, virtual screening, and MD simulations to search for potential inhibitors. Starting with blind dockings and published information, we identified top hit compounds, together with the five potential binding pockets of the nsP2 protease. Subsequently, the focused dockings into these different binding sites were investigated to discover more hit compounds, and observe further binding modes at these pockets. The top hit compounds were then subjected to MD simulations for 50 ns after equilibration. The simulation results demonstrated the different binding affinities of different ligands through the number of HBs and hydrophobic contacts. Previous studies largely focused on Pocket 1 as the active site. However, in this study, Pocket 4, in the N-terminal domain of the nsP2 protease, was recognised as the active site by the presence of catalytic residues, Cys1013 and His1083. Importantly, the current work offers more opportunities to identify potential inhibitors. The effect upon the active site of ligands binding into different pockets requires further experimental testing. Finally, the trajectories data of MD

simulations was utilized for the linear interaction energy to obtain accurate binding free energies. More experimental data is required to achieve a compromise between vdW and electrostatic interactions. A good agreement between docking results and MD simulations were confirmed by better binding free energies of top-hit compounds than tenth-hit compounds. Finally, it is worth emphasizing that this is the first instance where molecular modelling with more accurate data from X-ray structure has been studied. Our findings open up a promising approach in combining docking and MD simulations to assist in rational drug design, especially providing useful information for the design of inhibitors for CHIKV to help combat the disease.

CHAPTER 4. DISCOVERY OF INHIBITORS TARGETING CHIKV ENVELOPE GLYCOPROTEINS

4.1 INTRODUCTION

4.1.1 ENVELOPE GLYCOPROTEINS AS POTENTIAL TARGETS FOR CHIKV DRUG DISCOVERY

The CHIKV envelope glycoproteins (E1, E2, and E3) are promising targets for drug discovery as they play crucial roles in the virus attachment and entry.[8, 68] Like other alphavirus in Togaviridae family, CHIKV has an icosahedral structure comprising a nucleocapsid enclosed within a phospholipid envelope. Structural analysis of CHIKV particles revealed that there are 20 icosahedral "i3" spikes (located on the icosahedral 3-fold axes), and 60 quasi-3-fold "q3" spikes (located in general positions) that consist of a quasi-3-fold axis to form T=4 symmetry structure for the virus.[65] Additionally, the virus particles contain 80 spikes that make glycoprotein shells with 240 copies of each of two glycoproteins, E1 and E2 heterodimers.[68] The CHIKV genome, is divided into two open reading frames (ORF), encoding four non-structural proteins and five structural proteins (the capsid C; envelope glycoproteins E1 and E2; and two small peptides, E3 and 6K).[67] The non-structural proteins are essential for virus replication, protein modification, and immune antagonism, while the structural proteins are products obtained from a cleavage of a polyprotein by an auto-protease and signalase.[65]

The viral entry process is a receptor-mediated endocytosis in clathrin coated vesicles, which is controlled by two viral envelope glycoproteins, E1 and E2. In the acidic pH of the endosome environment, a conformational rearrangement of the surface glycoprotein shell causes a dissociation of the heterodimers p62-E1 (as known as pE2-E1) and formation of E1 homotrimers.[356] This process induces a fusion of the virus and endosomal membranes, resulting in release of the viral nucleocapsid into the cytoplasm. The E1 has a hydrophobic fusion loop which invades the cell membrane during the membrane fusion. The E2 contributes to receptor binding and protects the loop at a

neutral pH. During replication, the capsid is released, and the precursors of structural proteins are processed in the Golgi complex, then transported to the plasma membrane.

Structural and functional characterization of the envelope glycoproteins as well as the mechanism of neutralization with an interpretation of the cryo-electron microscopy structure of the CHIKV-like particles has been reported.[68, 356-357] The mutations of four highly conserved residues of E1, namely Gly91, Val178, Ala226, and His230 were found to reveal their important roles in the cell fusion process.[8, 358] Two highly conserved residues, Gly91 and His230, are important for membrane fusion functionality. A substitution of Gly91 with Glu91 caused a loss of E1 fusogenicity whereas any replacement of His230 into Ala230 resulted in the disappearance of this activity.[358] On the other hand, Val178 and Ala226 are less conserved residues whose mutations do not lose E1 fusion capacity but these changes depend on pH and cholesterol.[358] In particular, a change of Ala to Val in the position of 226 of E1 protein resulted in the reduction of cholesterol dependence to infect mosquito hosts.[6] Moreover, the study on CHIKV E2 mutants identified the acid-sensitive region in the E2 (amino acid at 229, 231, 232, 233, and 234 positions).[359] The results suggested that the E2 amino acids 229 to 234 region was responsible for neutralizing antibodies, and that these amino acids could prevent the conformational change through interacting with antibodies, leading to initiate the viral fusion and entry.[359] Also, the region of 229 to 234 was found to be a crucial role for viral replication partly due to its participation of inducing pH-dependent conformational changes.[359] In addition, the E2 glycoprotein interacts with the the host receptor protein at the 216 residue which is involved in initiating infection.[360]

4.1.2 EARLY RESEARCH TARGETING ENVELOPE GLYCOPROTEINS

Numerous studies in recent years have focused on the structural characterization of the envelope proteins of alphaviruses.[7] However, little is known about changes of their conformation or structural intermediates during fusogenic activity or cell attachment.[65] With regards to CHIKV drug discovery, there have been few studies to identify potential inhibitors. Mostly, targets for CHIKV drug discovery are non-structural proteins.[70, 351, 353, 361] Envelope glycoproteins are another class of attractive targets for

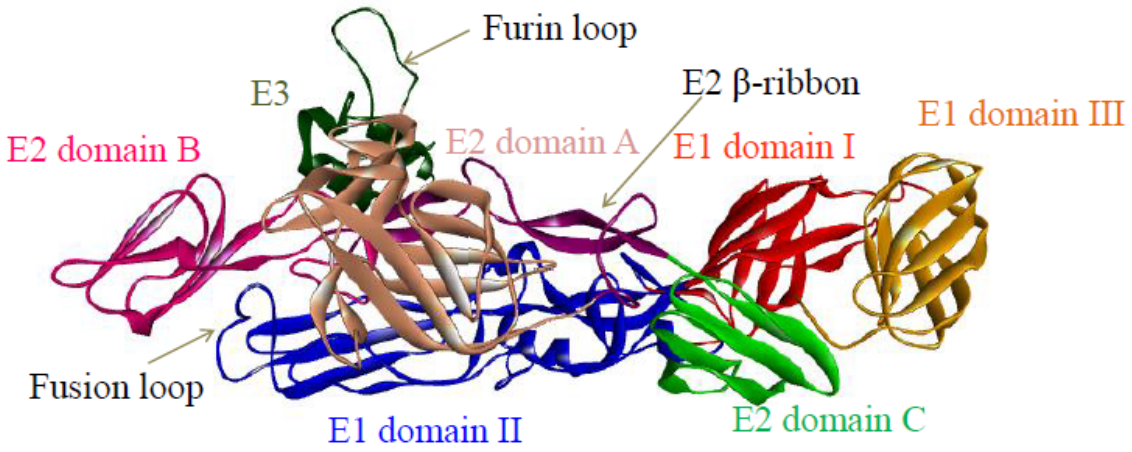
interfering with virus entry or virus attachment through inhibition of the viral envelope glycoproteins, to block the *in vitro* CHIKV infection.

Preliminary, computational investigation of the envelope glycoprotein complexes (the immature and mature forms)[153] has previously been explored within our group,[362] however a more rigorous exploration based on blind docking was deemed essential for full characterization of potential target sites. Virtual screenings based on blind docking and focused docking are a powerful approach in discovering the potential binding sites and hit compounds for CHIKV. Interaction analysis of hit compounds and these glycoprotein complexes provide useful information for CHIKV drug design.

4.1.3 COMPARISON OF TWO ENVELOPE GLYCOPROTEIN COMPLEXES

In order to identify inhibitors targeting the CHIKV envelope glycoproteins, characterization of the two complexes, the immature (PDB id: 3N40) and mature form (PDB id: 3N42) was necessary. As shown in Figure 4.1, the structures of precursor p62-E1 heterodimer (the immature form) and the complex E3-E2-E1 heterodimer (the mature form) are similar with an association of E1 domains (E1 domain I, domain II, and domain III), E2 domains (E2 domain A, domain B, and domain C), and E3. Additionally, in both structures, E2 makes contact with E3, while E1 does not create any interactions with E3. The only major difference between the two complexes is the presence of a furin loop, located in the E1 of the immature 3N40, but not present in the mature 3N42.[153] This loop becomes disordered in the cleavage process. In addition, furin maturation of p62 into E3 and E2 will prime the spikes for fusogenic activiation during cell entry. Moreover, the dissociation of E2-E1 heterodimer occurs under an acidic environment, which rearranges E1 into fusogenic homotrimers that induce fusion of viral and endosomal membranes. Therefore, residues in the region of the furin loop and the fusion loop were of key interest.

A



B

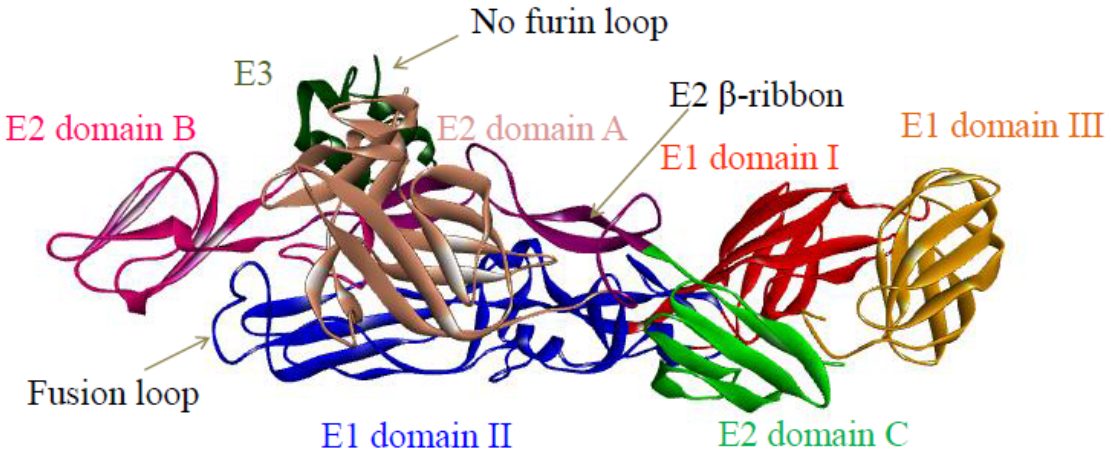


Figure 4.1. Structure of the envelope glycoprotein complexes: **(A)** The immature structure (PDB id: 3N40); **(B)** The mature structure (PDB id: 3N42). These structures are similar and the only difference is in the furin loop.

4.2 OVERVIEW OF THIS STUDY

In an attempt to discover hit compounds for the CHIKV targeting the envelope glycoproteins, this study used an *in silico* approach. The same process of docking and virtual screenings as used for the nsP3 and the nsP2 were undertaken. A docking protocol was established (described details in the Experimental procedures and methods chapter). Two X-ray crystal structures of envelope glycoprotein, an immature form (PDB id: 3N40) and a mature form (PDB id: 3N42), were downloaded from the Protein Data Bank. Initially, blind docking was carried out to identify the potential binding

sites, which were identified based on the locations of binding ligand. The dimensions of the box for covering 3N40 and 3N42 were 108 Å × 126 Å × 62 Å and 84 Å × 126 Å × 74 Å, respectively. The value of E in docking was changed from the default value E=8 to E=16, 32, 128, 256, and 1024 in order to investigate the sampling convergence in docking with AutoDock Vina. Potential binding sites were revealed by the place where ligand binds into the target (discussed in the following section). The next step was a focused docking, in which a smaller box (20 Å × 20 Å × 20 Å), was centred at the potential binding site of interest to confirm their likehood. Another aim of this process is to identify more hit compounds focusing on these binding sites, in order to achieve a better sampling. The sets were carried out for each glycoprotein complexes, 3N40 and 3N42. Moreover, identification of potential binding sites was also investigated and compared with the previous published results by using a receptor cavities tool in Accelerys DS 3.5 and the MetaPocket program.[343]

4.3 RESULTS AND DISCUSSION

4.3.1 MOLECULAR DOCKING AND VIRTUAL SCREENINGS

The outcomes of screenings using blind docking and focused docking for two glycoprotein complexes are shown in Tables 4.1, 4.2, 4.3, and 4.4.

In Table 4.1, blind dockings showed that there were two potential binding pockets identified in the structure 3N42, and one in 3N40. However, analysis of docking results with the change to the E value in blind docking (Table 4.2 and Table 4.3) introduced some new binding pockets (discussed below). In total, two potential binding pockets (3N40_Pocket 1 and 3N40_Pocket 2) were identified in the immature 3N40; and four binding pockets (3N42_Pocket 1, 3N42_Pocket 2, 3N42_Pocket 3, and 3N42_Pocket 4) were in the mature structure 3N42, respectively. 3N40_Pocket 1 and 3N40_Pocket 2 are located in the similar location with 3N42_Pocket 1 and 3N42_Pocket 2.

Table 4.1. Results of the top ten compounds of blind dockings with locations of pockets in the immature (PDB id: 3N40) and mature forms (PDB id: 3N42). The binding affinities ΔG are in kcal/mol.

VST1[a]			**VST2[b]**		
Hits	ΔG	Pocket	*Hits*	ΔG	Pocket
1. NCI_293778	-13.2	1	1. NCI_293778	-13.7	1
2. NCI_61610	-12.2	1	2. NCI_37553	-11.4	1
3. NCI_37553	-11.7	1	3. NCI_61610	-11.2	1
4. NCI_156219_b	-11.5	1	4. NCI_156219_b	-11.1	1
5. NCI_84100_b	-10.9	1	5. NCI_84100_a	-11.0	1
6. NCI_84100_a	-10.9	1	6. NCI_227186_b	-10.9	1
7. NCI_116702	-10.8	1	7. NCI_84100_b	-10.9	2
8. NCI_19990_a	-10.8	1	8. NCI_7524_a	-10.9	1
9. NCI_227186_a	-10.7	1	9. NCI_116702	-10.8	1
10. NCI_227186_b	-10.6	1	10. NCI_308835	-10.8	1

(a) In VST1, the grid box is fixed at the centre of the structure (-23.7 Å, 1.8 Å, 23.4 Å) with dimensions of 108 Å × 126 Å × 62 Å (PDB id: 3N40). **(b)** In VST2, the grid box is centred at the centre of the structure (-26.4 Å, 7.1 Å, -24.5 Å) with dimensions of 84 Å × 126 Å × 74 Å (PDB id: 3N42). The origin and axes of the coordinate systems were set as in the PDB structure.

Pocket 1 and Pocket 2 were found to be similar in both the immature and mature form structures, even though some residues from the E2 domain forming the pocket were different. This may be due to a difference in these two structures in the virus attachment. For the both structures, immature and mature structures, Pocket 1 was in between E2 domain A, E2 domain C, and E1 domain II which was a favourable place for most of the ligands. Pocket 2 located at the E2 β-ribbon. Pocket 1 was corresponding to the combined Site 1 and Site 2 identified previously.[362] In the mature structure, Pocket 3 was in the region of between E2 domain C, E1 domain I, and E1 domain III while Pocket 4 was behind the fusion loop and between E2 domain A, E2 domain B, and E1 domain II. Pocket 2 was a novel binding pocket identified in the 3N40. Pocket 2 and Pocket 3 were two new binding pockets in the 3N42, while Pocket 1 and Pocket 4 have been reported.[362] Most of the pockets included some E1 and some E2 residues. The locations of the pockets along with the top hit compounds, are illustrated in Figure 4.2, and the residues making up each pocket are listed in Table 4.5. Using an increase of the value of a search space parameter (E) from the default value E=8 to E=16, 32, 128, 256, and 1024 to investigate the sampling convergence in Vina, the results were presented in Table 4.2 and 4.3 for two glycoprotein complexes.

Table 4.2. Results of the top ten compounds of blind dockings for the immature structure. The binding affinities are in kcal/mol.

VST3[a]	VST4[b]	VST5[c]	VST6[d]	VST7[e]	VST8[f]
1. NCI_293778 (-13.1)	1. NCI_293778 (-13.1)	1. NCI_293778 (-13.1)	1. NCI_293778 (-13.1)	1. NCI_293778 (-13.1)	1. NCI_293778 (-13.0)
2. NCI_61610 (-11.9)	2. NCI_61610 (-12.3)	2. NCI_61610 (-12.3)	2. NCI_61610 (-12.3)	2. NCI_61610 (-12.3)	2. NCI_37553 (-11.7)
3. NCI_37553 (-11.7)	3. NCI_156219_b (-11.5)	3. NCI_37553 (-11.7)	3. NCI_37553 (-11.7)	3. NCI_37553 (-11.7)	3. NCI_156219_b (-11.5)
4. NCI_156219_b (-11.5)	4. NCI_308835 (-11.0)	4. NCI_156219_b (-11.5)	4. NCI_156129_b (-11.5)	4. NCI_308835 (-11.5)	4. NCI_260594 (-11.4)
5. NCI_84100_b (-10.9)	5. NCI_84100_a (-10.9)	5. NCI_260594 (-11.4)	5. NCI_260594 (-11.4)	5. NCI_156219_b (-11.5)	5. NCI_308835 (-11.4)
6. NCI_84100_a (-10.9)	6. NCI_84100_b (-10.9)	6. NCI_308835 (-11.1)	6. NCI_84100_b (-11.4)	6. NCI_260594 (-11.4)	6. NCI_227186_a (-10.9)
7. NCI_19990_a (-10.8)	7.NCI_37553 (-10.8)	7.NCI_84100_a (-11.0)	7.NCI_7524_a (-11.0)	7. NCI_84100_a (-11.0)	7. NCI_19990_a (-10.8)
8. NCI_116702 (-10.8)	8.NCI_116702 (-10.8)	8.NCI_84100_b (-10.9)	8.NCI_227186_b (-10.9)	8. NCI_60785_b (-11.0)	8. NCI_202386 (-10.8)
9. NCI_227186_a (-10.7)	9.NCI_67436 (-10.6)	9.NCI_227186_a (-10.9)	9.NCI_116702 (-10.9)	9. NCI_227186_a (-10.9)	9. NCI_116702 (-10.8)
10. NCI_7524_a (-10.7)	10. NCI_227186_b (-10.6)	10. NCI_19990_a (-10.8)	10. NCI_308835 (-10.8)	10. NCI_84100_b (-10.9)	10. NCI_37641_b (-10.7)

The grid box is fixed at the centre of the structure (-23.7 Å, 1.8 Å, 23.4 Å) with dimensions of 108 Å × 126 Å × 62 Å (PDB id: 3N40): **(a)** VST3: E=8; **(b)** VST4: E = 16; **(c)** VST5: E = 32; **(d)** VST6: E = 128; **(e)** VST7: E = 256; **(f)** VST8: E = 1024

Table 4.3. Results of the top ten compounds of blind dockings for the mature structure. The binding affinities are in kcal/mol.

VST3[a]	VST4[b]	VST5[c]	VST6[d]	VST7[e]	VST8[f]
1. NCI_293778 (-13.7)	1. NCI_293778 (-13.7)	1. NCI_293778 (-13.7)	1. NCI_293778 (-13.7)	1. NCI_293778 (-13.7)	1. NCI_293778 (-13.7)
2. NCI_37553 (-11.5)	2. NCI_37553 (-11.5)	2. NCI_61610 (-11.5)	2. NCI_61610 (-11.7)	2. NCI_61610 (-11.6)	2. NCI_37553 (-11.5)
3. NCI_61610 (-11.3)	3. NCI_61610 (-11.4)	3. NCI_37553 (-11.4)	3. NCI_37553 (-11.5)	3. NCI_37553 (-11.5)	3. NCI_156219_b (-11.1)
4. NCI_156219 (-11.1)	4. NCI_156219_b (-11.1)	4. NCI_156219_b (-11.1)	4. NCI_156219_b (-11.1)	4. NCI_156219_b (-11.1)	4. NCI_25457 (-10.8)
5. NCI_84100_a (-11.0)	5. NCI_227186_b (-10.9)	5. NCI_84100_a (-11.0)	5. NCI_84100_a (-11.0)	5. NCI_84100_a (-11.0)	5. NCI_121868_a (-10.8)
6. NCI_7524_a (-10.9)	6. NCI_84100_a (-10.9)	6. NCI_227186_b (-10.9)	6. NCI_7524_a (-10.9)	6. NCI_84100_b (-10.9)	6. NCI_308835 (-10.8)
7. NCI_84100_b (-10.9)	7. NCI_7524_a (-10.9)	7. NCI_7524_a (-10.9)	7. NCI_227186_b (-10.9)	7. NCI_7524_a (-10.9)	7. NCI_116702 (-10.8)
8. NCI_25457 (-10.8)	8. NCI_84100_b (-10.9)	8. NCI_84100_b (-10.9)	8. NCI_84100_b (-10.9)	8. NCI_227186_b (-10.9)	8. NCI_122819_b (-10.7)
9. NCI_116702 (-10.8)	9. NCI_308835 (-10.8)	9. NCI_308835 (-10.8)	9. NCI_156219_b (-10.8)	9. NCI_116702 (-10.8)	9. NCI_156219_a (-10.7)
10. NCI_122819 (-10.7)	10. NCI_116702 (-10.8)	10. NCI_25457 (-10.8)	10. NCI_25457 (-10.8)	10. NCI_308835 (-10.8)	10. NCI_227186_b (-10.7)

The grid box is fixed at the centre of the structure (-26.4 Å, 7.1 Å, -24.5 Å) with dimensions of 84 Å × 126 Å × 74 Å (PDB id: 3N42): **(a)** VST3: E = 8; **(b)** VST4: E = 16; **(c)** VST5: E = 32; **(d)** VST6: E = 128; **(e)** VST7: E = 256; **(f)** VST8: E = 1024

Table 4.4. Results of the top ten compounds of focused dockings for the immature and mature structures. The binding affinities are in kcal/mol.

VST^{1a}	**VST^{2a}**	**VST^{1b}**	**VST^{2b}**	**VST^{3b}**	**VST^{4b}**
1. NCI_293778 (-13.1)	1. NCI_293778 (-11.4)	1. NCI_293778 (-13.7)	1. NCI_61610 (-11.6)	1. NCI_61610 (-11.5)	1. NCI_121868_a (-7.6)
2. NCI_37553 (-11.4)	2. NCI_37641_b (-10.7)	2. NCI_37553 (-11.4)	2. NCI_37553 (-11.1)	2. NCI_293778 (-11.0)	2. NCI_24951_b (-10.6)
3. NCI_156219_b (-11.3)	3. NCI_37553 (-10.0)	3. NCI_61610 (-11.3)	3. NCI_308835 (-11.1)	3. NCI_37553 (-9.9)	3. NCI_24951_a (-10.2)
4. NCI_19990_a (-10.9)	4. NCI_328101 (-10.0)	4. NCI_25457 (-10.8)	4. NCI_84100_b (-10.9)	4. NCI_670283 (-9.4)	4. NCI_332186 (-9.7)
5. NCI_84100_b (-10.8)	5. NCI_116702 (-9.9)	5. NCI_116702 (-10.8)	5. NCI_84100_a (-10.8)	5. NCI_332670 (-9.4)	5. NCI_332670 (-9.6)
6. NCI_84100_a (-10.8)	6. NCI_37641_a (-9.9)	6. NCI_308835 (-10.7)	6. NCI_82802_b (-0.7)	6. NCI_328101 (-9.3)	6. NCI_3391_a (-9.6)
7. NCI_116702 (-10.7)	7. NCI_121868_a (-9.8)	7. NCI_97920 (-10.6)	7. NCI_670283 (-10.5)	7. NCI_60785_a (-9.2)	7. NCI_80313 (-9.5)
8. NCI_60785_b (-10.5)	8. NCI_308835 (-9.7)	8. NCI_7524_b (-10.4)	8. NCI_7524_b (-10.4)	8. NCI_50648 (-9.2)	8. NCI_310326 (-9.3)
9. NCI_308835 (-10.5)	9. NCI_24951_a (-9.5)	9. NCI_671136 (-10.3)	9. NCI_293778 (-10.4)	9. NCI_60785_b (-9.2)	8. NCI_645330 (-9.3)
10. NCI_227186_b (-10.4)	10. NCI_127133 (-9.5)	10. NCI_84100_a (-10.2)	10. NCI_328101 (-10.3)	10. NCI_308835 (-9.1)	9. NCI_97920 (-9.2)

(a) In VST^{1a}, VST^{2a} the grid box is fixed at the centre of Pocket 1 and Pocket 2 of the immature structure (-22.3 Å, 18.3 Å, 26.3 Å; -15.9 Å, 1.5 Å, 19.7 Å; respectively) with dimensions of 20 Å × 20 Å × 20 Å (PDB id: 3N40). **(b)** In VST^{1b}, VST^{2b}, VST^{3b}, and VST^{4b}, the grid box is fixed at the centre of Pockets 1, 2, 3, and 4 of the mature structure (-21.3 Å, -11.7 Å, -24.9 Å; -22.4 Å, -4.2 Å, -11.9 Å; -5.7 Å, 31.2 Å, -8.9 Å, -44.7 Å, -39.8 Å, -38.2 Å respectively) with dimensions of 20 Å × 20 Å × 20 Å (PDB id: 3N42). The origin and axes of the coordinate systems were set the same as in the PDB structure.

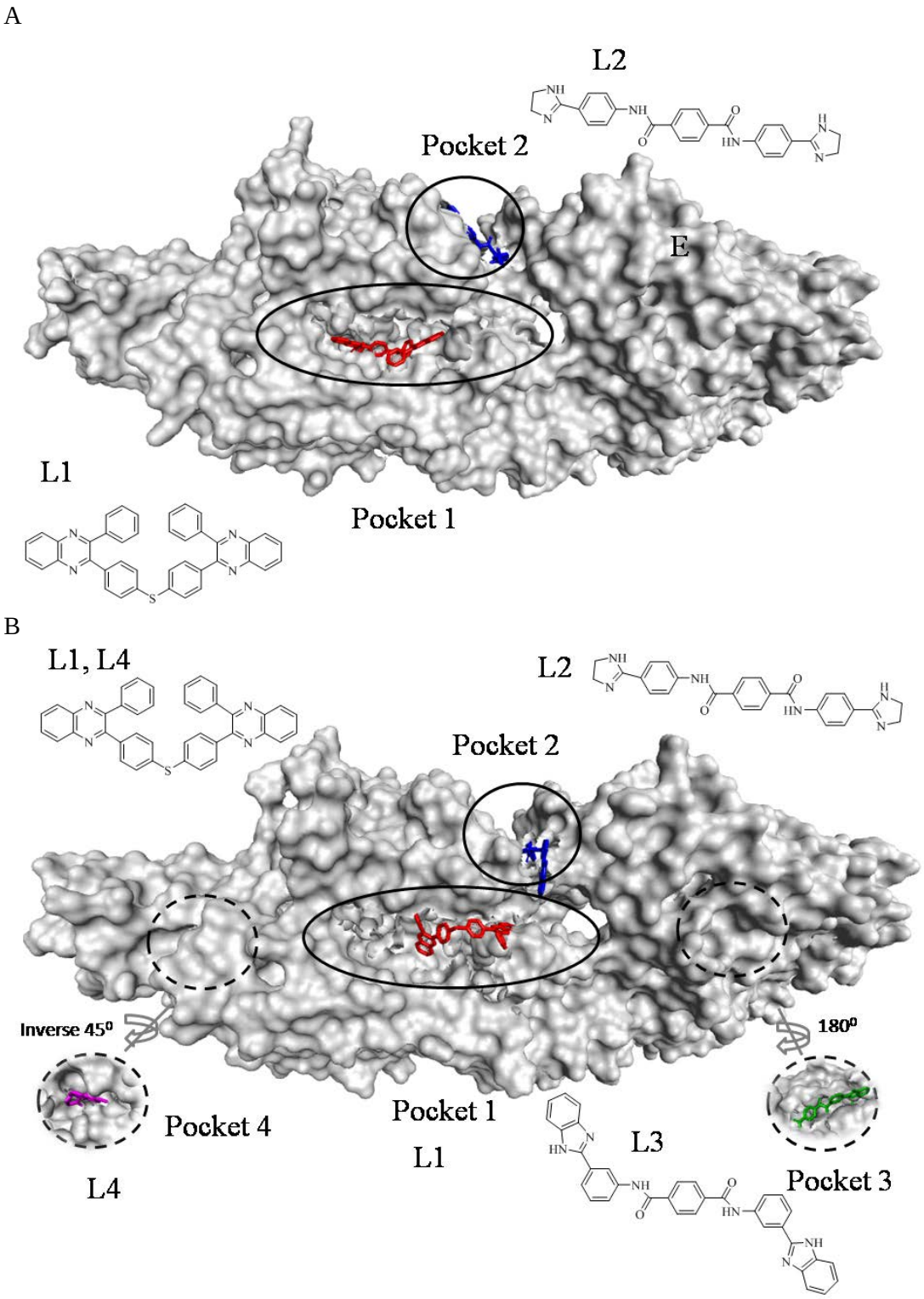


Figure 4.2. Representation of docked structures of top hit compounds in different virtual screenings showing the location of the pockets: **(A)** Pocket 1 with ligand NCI_293778 (L1) and Pocket 2 with ligand NCI_67436 (L2) in the immature structure (PDB id: 3N40); **(B)** Pockets in the mature structure (PDB id: 3N42): ligand NCI_293778 (conformation L1) in Pocket 1, ligand NCI_67436 (L2) in Pocket 2, ligand NCI_61610 in Pocket 3 (L3), and ligand NCI_293778 (conformation L4) in Pocket 4.

Table 4.5. Residues making up the pockets for two structures, the immature (PDB id: 3N40) and mature structure (PDB id: 3N42).

	Location	The immature structure	The mature structure
Pocket 1	Between E2 domain A, E1 domain II, and E2 domain C	**E1 residues:** Glu50, Tyr51, Lys52, Thr53, Val54, Ile55, Tyr180, Lys181, Asp183, Tyr185, Gln222, Tyr242, His230, Val231, Pro232, Tyr233, Gln235, Ser238, Lys241, Tyr242, Leu244, Lys245, Arg247, Gly248, Ala249, Ser250, His253 **E2 residues:** His82, Glu99, Arg100, Ile101, Arg102, Asn103, Thr106, His191, Pro192, Phe193, His194, Asp196, Pro198, Val199, Ile200, Gly201, Leu215, Pro216, Glu232, Tyr301, Asn302, Pro304, Leu305	**E1 residues:** Glu50, Tyr51, Lys52, Thr53, Val54, Ile55, Tyr180, Lys181, Asp183, Tyr185, Gln222, Tyr242, His230, Val231, Pro232, Tyr233, Gln235, Ser238, Lys241, Tyr242, Leu244, Lys245, Arg247, Gly248, Ala249, Ser250, His253 **E2 residues:** Glu35, Arg36, Ile37, Gln39, Glu40, Ala41, Thr42, His127, Pro128, Phe129, His130, His131, Glu168, Tyr237, Asn238, Pro240, Leu241
Pocket 2	E2 β-ribbon	**E2 residues:** Pro198, Val199, Phe205, His206, Arg208, Arg331	**E2 residues:** Asp43, Arg104, Pro133, Pro134, Val135, Ile136, Lys140, Phe141, Arg144, Thr155, Val264, Cys266, Arg267
Pocket 3	Between E2 domain C, E1 domain I, and E1 domain III		**E1 residues:** Gly12, Val13, Pro14, Ser310, Asp311, Phe312, Phe321, Ala336, Ala359, Asn389, Tyr390, Pro391, Ala392, Ser393, Ala336, Thr338, Ser357, Thr358 **E2 residues:** Pro335, Tyr336, Lys337
Pocket 4	Behind the E1 fusion loop and between E2 domain B, E2 domain A, and E1 domain II		**E1 residues:** Lys61, Cys62, Tyr93, Phe95 **E2 residues:** Asp174, Arg198, Tyr199, Lys200, Glu208, Gly209, His226, Ala228, Glu247

4.3.2 IDENTIFICATION OF POTENTIAL BINDING POCKETS USING DIFFERENT METHODS

The receptor cavities tool in Accelrys DS software and MetaPocket, were utilized to identify possible binding sites in the structures. Using Accelrys DS, four pockets mentioned above were found in both structures. Some additional pockets are also found in two structures (Appendixes 28-29). No potential binding site was detected using MetaPocket. Table 4.6 shows a comparison of identified binding pockets using different methods.

Table 4.6. Comparison of locations of identified binding pockets in both structures using different methods, blind dockings and the receptor cavities tool in Accelrys Discovery Studio program, and compared with previous study.

Accelerys Discovery Studio	**Blind dockings**		**Previous study**[358]
	3N40	3N42	
Between E1 domain II and E2 domain A, and E2 domain C	Pocket 1	Pocket 1	Site 1
Between E1 domain II and near the E2 β-ribbon	Pocket 1	Pocket 1	Site 2
At the E2 β-ribbon	Pocket 2	Pocket 2	No
Behind the fusion loop, between E1 domain II, E2 domain B, E2 domain A	No	Pocket 4	Site 4
Between E1 domain I, E1 domain III, and E2 domain C	No	Pocket 3	No

Among the pockets, Pocket 1 was the largest (with its volume 352 $Å^3$ in the 3N40 and 621 $Å^3$ in the 3N42), accommodating most ligands, while the other pockets were narrower and shallower (Pocket 2 in 3N40, 3N42 were 123 $Å^3$ and 156 $Å^3$, respectively; Pocket 3 in 3N40, 3N42 were 42 $Å^3$ and 129 $Å^3$, respectively; and Pocket 4 in 3N40, 3N42 were in turn 26 $Å^3$ and 523 $Å^3$, respectively), illustrated by the volumes in Appendixes 28-29 for the 3N40 and 3N42, respectively. Additionally, a comparison of the volumes of these four pockets between the immature and mature forms revealed that Pocket 4 near the fusion loop was much bigger in the mature form (523 $Å^3$) than the immature form (26 $Å^3$). Therefore, small molecules can bind more easily to this pocket in the mature structure with significant effects; including blocking the relative movement of E2 domains A and B, and also freezing the fusion loop by stabilizing interactions; this could lead to prevention of the exposure of the

fusion loop. These findings were in accordance with the previous results.[362] However, because the Accelrys program also considered small cavities as potential binding pockets, the cavities (Appendixes 28-29) were located by Accelrys which were not found by docking.

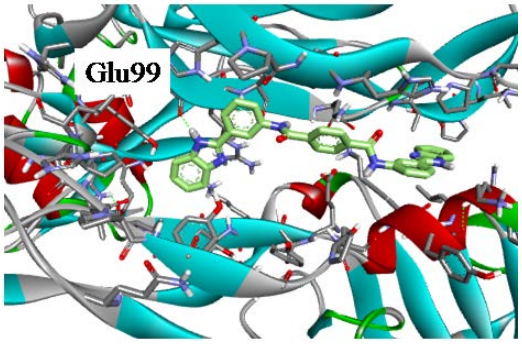
4.3.3 ANALYSIS OF INTERACTIONS BETWEEN HIT COMPOUNDS AND THEIR BINDING POCKETS

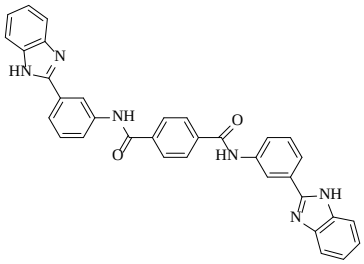
The binding affinities of all of hit compounds were approximately -10 kcal/mol, which indicated favourable interactions with the two structures in all of the identified binding pockets. It should be noted that focused dockings introduce new hits compounds. Docking analysis revealed some key residues for interactions between both the 3N40 and 3N42 structures and the ligands, listed in Table 4.7. The interaction of these key residues are illustrated in Figures 4.3 and 4.4 (details in Appendixes 30-49). For Pocket 1 and Pocket 4, the key residues in Table 4.7 were in close agreement with the published findings.[362] Interestingly, there was a change of key residues in E2 of Pocket 1 from the immature to mature structure; for example residues involved in forming hydrogen bonds in E2 were Arg100, Tyr301 in the immature form corresponding to Arg36, Tyr237 in the mature form. The ligands bound to Pocket 1 were expected to have higher affinities than others due to their interactions with E1 or E2 through formation of hydrogen bonds, which may stabilize the E1-E2 heterodimer and prevent dissociation.[362] The residues in Pocket 4 could interact with the E1 fusion loop including Gly91 and His230, which emphasize their importance in impairment of the fusion process. In particular, His230 was reported to be important in stabilization of the fusion loop. In most of the hydrogen bonding interactions in both structures, the ligands acted as donors; for example, ligands NCI_61610, NCI_84100_a, NCI_116702, NCI_156219_b, and NCI_227186_a in the immature form; or NCI_7524_a, NCI_61610, NCI_84100_b, NCI_116702, NCI_156219_b, and NCI_227186_b in the mature form. In contrast, some ligands, such as NCI_227186_b in the 3N40 or NCI_61610 in the 3N42, were hydrogen bonding acceptors.

Table 4.7. Key residues for interactions in both envelope glycoprotein structures at each of binding site.

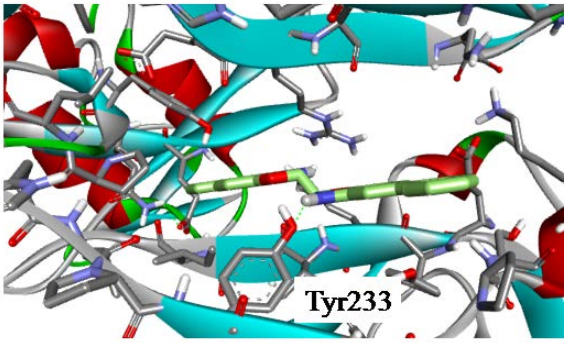
Binding pocket	Immature structure (3N40)	Mature structure (3N42)
Pocket 1	**E1 residues**: Tyr51, Thr53, Tyr233, Ser238 **E2 residues**: Arg100, Ile101, Asn103, Pro192, Tyr301	**E1 residues**: Glu50, Tyr51, Lys52, Thr53, Ile55, Tyr242, Ser238 **E2 residues**: Arg36, Pro128, Tyr237
Pocket 2	**E2 residues**: Arg168, Pro170, Pro198, Val199, Ile200, Phe205, His206	**E2 residues**: Asp43, Ile136, Arg144, Arg267
Pocket 3		**E1 residues**: Phe312, Tyr390 **E2 residues**: Tyr336
Pocket 4		**E1 residues**: Lys61, Cys62 and Tyr93 **E2 residues**: Ile37

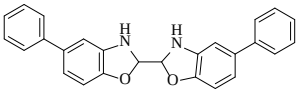
A



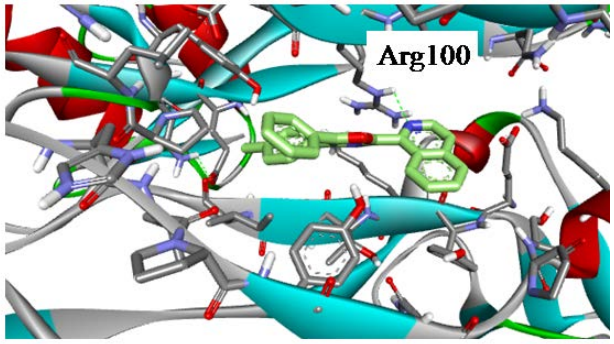


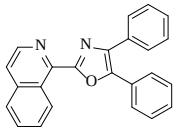
B



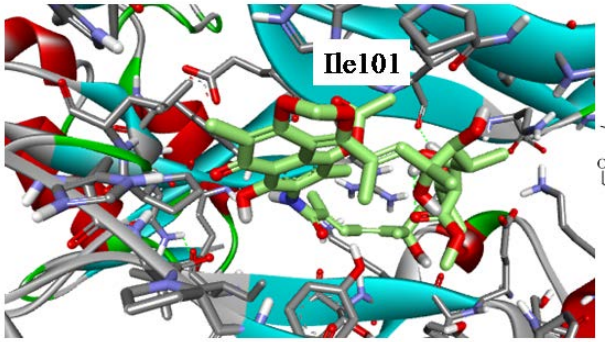


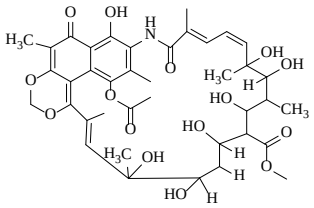
C



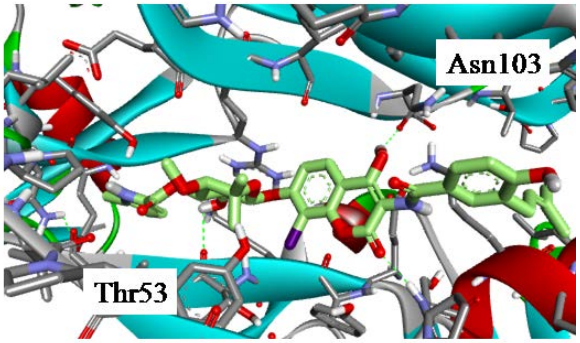


D





E



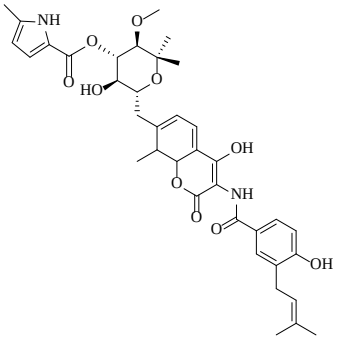
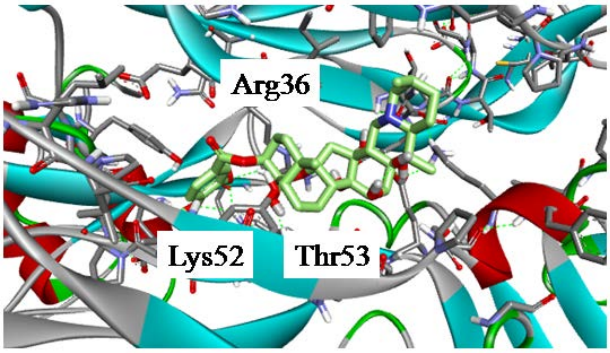
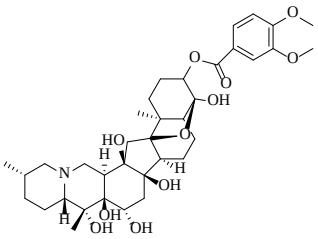


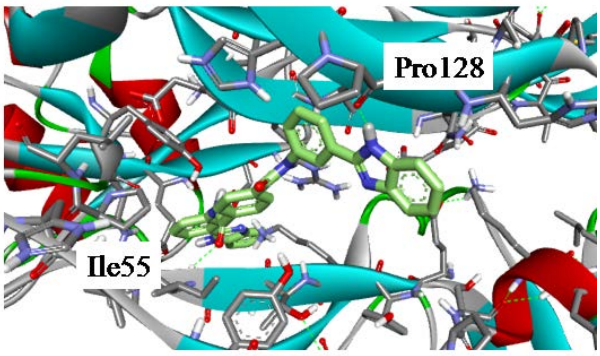
Figure 4.3. Hydrogen bonding analysis of compounds with the immature structure in docking: **(a)** NCI_61610, **(b)** NCI_84100_a, **(c)** NCI_116702, **(d)** NCI_156219_b, **(e)** NCI_227186_a. The key residues involved in the interactions between glycoproteins and ligands are shown. The ligands (in cyan) and the residues surrounding the ligands (in grey) were displayed in sticks and coloured by atoms (carbon in cyan in ligand or in grey in residues, nitrogen in blue, oxygen in red, sulphur in organge).

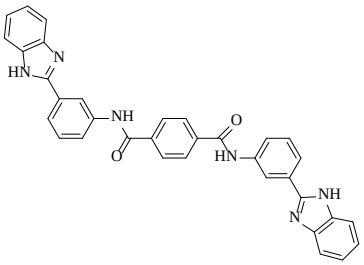
A



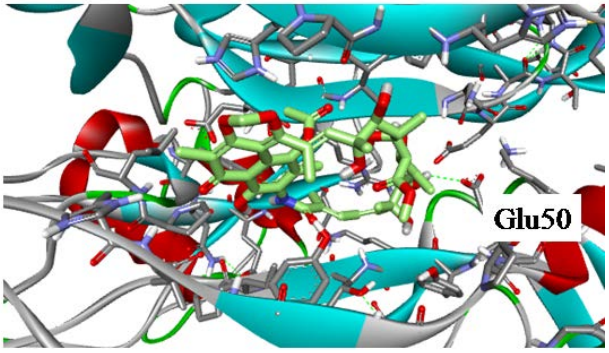


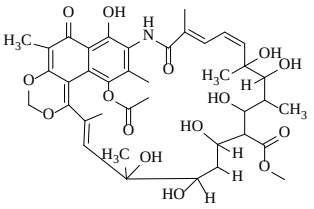
B



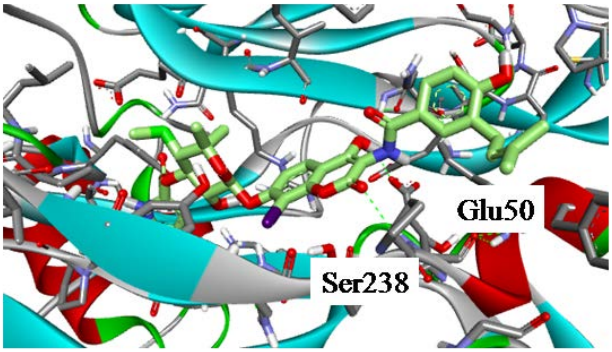


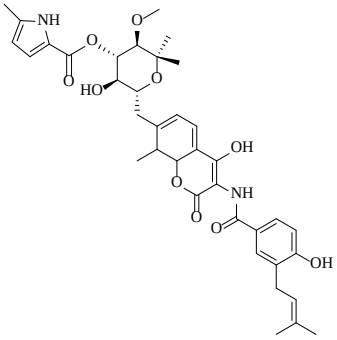
C





D





E

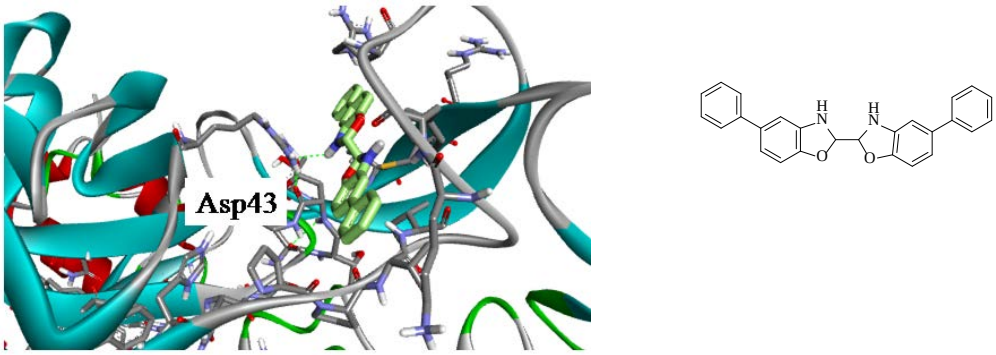


Figure 4.4. Hydrogen bonding analysis of compound with the mature structure: **(a)** NCI_7524_a, **(b)** NCI_61610, **(c)** NCI_156219_b, **(d)** NCI_227186_b, **(e)** NCI_84100_b. The key residues involved in the interactions are also shown. The ligands (in cyan) and the residues surrounding the ligands (in grey) were displayed in sticks and coloured by atoms (carbon in cyan in ligand or in grey in residues, nitrogen in blue, oxygen in red, sulphur in organge).

4.3.4 SAMPLING CONVERGENCE IN AUTODOCK VINA

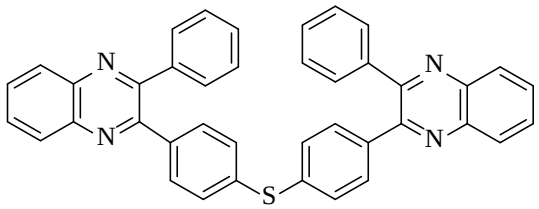
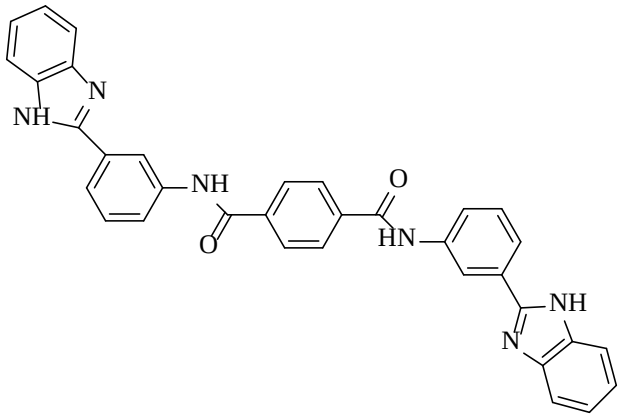
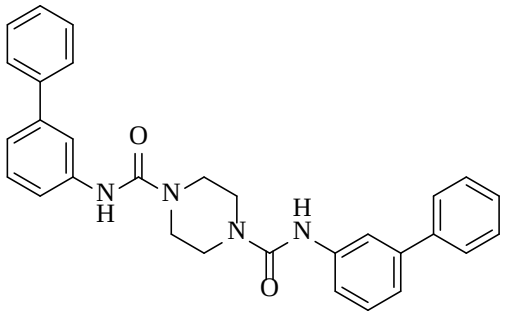
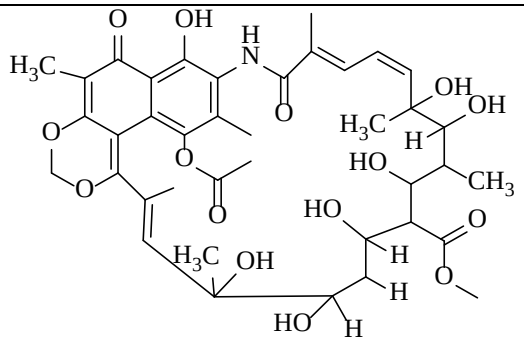
As previously mentioned, the value of exhaustiveness (E) or a search space parameter in AutoDock Vina was increased from the default value E=8 to E=16, 32, 128, 256, and 1024 to investigate the searching convergence in docking. The results of top hit compounds are listed above in Table 4.2 and Table 4.3. These tables demonstrate that the exhaustiveness parameter greatly affected the searching process for ligands conformations in docking. Analysis of docking results revealed that the blind dockings with an increase in the exhaustiveness allowed new binding pockets. For instance, in the immature structure, Pocket 2 appeared in docking for E = 16-1024, though was not present for E=8. In the mature structure, Pocket 2 and Pocket 3 were introduced in docking with E = 16, 32, 128 instead of using E=8. It is encouraging to see the frequent presence of top hit compounds in most blind docking procedures. Interestingly, the hit compounds maintained their conformations at the same binding site despite changes to the E value, which brings further confidence in sampling convergence.

4.3.5 ANALYSIS AND SELECTION OF HIT COMPOUNDS FOR BIOLOGICAL TESTING

The results obtained from all dockings showed that some compounds could bind in many docking procedures. They seem to possess higher potential than other compounds. The binding affinities of most of them were from -10 to -13 kcal/mol indicating good binding affinities between the ligands and the envelope glycoprotein complexes. As mentioned previously, the compounds which bound to Pocket 1 and Pocket 4 were expected to be inhibitors targeting the glycoprotein complexes, as they could interact with the residues involved in the process of virus attachment and virus entry. Therefore, based on the binding affinities of hit compounds obtained from blind and focused dockings, and analysis of the interactions between the compounds and the envelope glycoproteins complexes, the proposed potential lead compounds for biological testing are NCI_293778 (from -11.4 to -13.7 kcal/mol), NCI_61610 (from -11.6 to -12.3 kcal/mol), NCI_37553 (-11.5 kcal/mol), and NCI_156219 (-11.1 kcal/mol). (Table 4.8).

In docking, the protein structure was kept rigid, so the selection of hit compounds from molecular docking is a good starting point for anti-CHIKV drug discovery. More accurate binding free energies are expected to gain from molecular dynamics simulations, however, given the larger sizes of these glycoproteins complexes, this procedure could not be conducted in this study due to lack of time on researches.

Table 4.8. Potential lead compounds for the envelope glycoproteins proposed for biological testing.

No	Compounds	Chemical structure
1	NCI_293778	
2	NCI_61610	
3	NCI_37553	
4	NCI_156219	

4.4 CONCLUSIONS

Taking advantage of the envelope glycoprotein complexes, virtual screening based on blind docking and focused dockings explored the potential binding pockets and inhibitors for both the immature and mature structures of envelope proteins. Promising hit compounds were identified for two complexes of the envelope glycoproteins. Pocket 2 was a novel binding site in the immature structure. Pocket 2 and Pocket 3 were novel binding sites for the mature complex of the glycoproteins. The key residues involved in stabilizing the complex or participating in the fusion process were confirmed. This study also supported the current docking protocol utilising AutoDock Vina as robust and with a good accuracy, and could be used to indentify inhibitors. However, due to larger size of two complexes of glycoproteins, the immature and mature structures, molecular dynamics simulations could not be carried out during the project. Therefore, the results of hit compounds and their binding modes obtained from docking are a good starting point for further studies. Further experiments are required to test the inhibitory effects for the anti-CHIKV envelope glycoproteins compounds.

CHAPTER 5. DISCUSSION OF FINDINGS

As there is little information available investigating CHIKV drug discovery, this study undertook to establish a computational framework to identify novel inhibitors for CHIKV. This combined the strength of molecular docking, virtual screening, molecular dynamics simulations and free energy calculations. The X-ray crystal structures used were of high resolution and the results provide valuable information about inhibitors of CHIKV, not previously available from previous studies involving homology modeling. The information could also be used for further steps to optimize the hit compounds into lead compounds. This chapter extends the discussion from each chapter in an attempt to gain a comprehensive of the various strengths and limitations.

5.1 SELECTION OF LEAD COMPOUNDS FOR BIOLOGICAL TESTING FOR CHIKV

The potential lead compounds for CHIKV were identified based on the results of molecular docking, molecular dynamics simulations and binding free energy calculations for each protein targets of CHIKV, including nsP3 macrodomain, nsP2 protease and envelope glycoproteins.

Virtual screening based on docking has shown successfully the ability to predict binding affinities of ligands when they interact with the protein of interest. The top ten hits (or the best binders), with high binding affinities, were selected as potential inhibitors. They all showed a good fit in the binding sites through interacting with protein residues such as H-bonds or hydrophobic contacts. For each docking run, the real difference between binding affinities for the top ranked compounds was in the diversity of chemical structures as well as the shape and size of binding site. There appears to be little correlation between structures and the binding affinities. However, the smaller molecules may not utilise the full binding potential of the pocket. Most pockets are hydrophobic and of the resulting hits therefore contained aromatic rings such as benzimidazole, quinoline, quinoxaline, oxazole, and thiophene. These rings could interact with the aromatic rings of protein residues Tyr,

Trp, Phe and His in the pockets by π-π stacking or π-network to stabilize and maintain the interaction of protein and ligands. The phenyl rings made hydrophobic contacts with protein residues. The ligand backbone containing nitrogen and oxygen atoms in combination with functional groups such as OH-, NH-, C=O in the side chains contributed in forming H-bonds with protein residues as the hydrogen bonding acceptors mostly and occasionally donors. These key interactions between protein and ligands are defined based on the chemical features of ligands. This study demonstrated that depending on each protein target, hydrophobic interactions or hydrogen bonding interactions would play an important role in stabilizing the binding. For example, in the case of nsP3 protein, H-bonds showed more contribution in Pocket 1 and Pocket 3 while hydrophobic interactions are more favourable in Pocket 2. For the nsP2 protease, in the active site, interactions were stabilized by forming H-bonds with the catalytic dyad. Suggestions for regions of modifying the structure to aid the design of new compounds with higher affinities were considered, *e.g*, the adding of heteroatom into the aromatic rings or functional groups to balance the hydrophobicity of ligands. Further, a pharmacophore model for CHIKV inhibitors, could be constructed in the next step. For instance, the case of targeting CHIKV nsP2 protease, it is likely that a model would include an aromatic ring, a hydrophobic and two hydrogen bond donor sites.

In addition, it is also apparent that the size of compounds with diverse spatial orientations of aromatic rings or the conformational complexity of compounds, as well as the shape and the size of binding sites would affect their binding affinities. It was found that some hits having bulky structures molecules, for example NCI_61610, NCI_293778, could fit well in the large binding sites, for example, in the case of Pocket 1 of nsP3, so they would show good binding scores. Conversely, in the case of ligands of bigger size or the binding size not particular large, the molecules protruded and did not bind correctly, such as in the Pocket 2 of nsP3.

The application of Lipinski's Rule of Five to the top ten hits does not provide good immediate outcomes. This procedure would help to evaluate the drug-likeness of these specific derivatives for oral drugs. The results revealed that some compounds did not satisfy the conditions, however they were still selected for biological testing.

The major reason for that as there is not a current drug for CHIKV treatment, the need for a treatment means that the lack of oral administration requires alternative administration, *e.g.* intravenous route delivery. More importantly, the revealed hit compounds were never actually considered to be possible drugs themselves, however, as hits, medicinal chemistry studies investigate modifications to allow for the synthesis of more drug-like compounds. For instance, NCI_293778 has a logP of 10.87 and although it showed good interactions and a high binding affinity, modification would be required for them to be considered in the future as orally available drugs.

Furthermore, when combing the results of all three targets, some virtual hits overlap multiple times with different pockets of a single protein, and also appear in different protein targets. It could also be the nature of the molecules themselves that makes them in the top of different screens. A prime example of this is the NCI_61610 molecule, which was the top hit for both protein nsP3 and nsP2 and one of the highest scoring inhibitors for envelope glycoproteins. Thus, this could be a promising lead compound for multi-pathway inhibitions, however, it may require studies of structure-activity relationship to avoid non-specific activity. For example, NCI_61610 has also shown anti-inflammatory effect in mice model.[363]

Molecular dynamics simulations and linear interaction energy calculations were subsequently carried out which helped to look at insights into the interactions between proteins and ligands and to obtain more accurate binding affinities. Importantly, the outcomes also confirmed the stability of the complexes and validated the docking outcomes. The results revealed that the systems were stabilized during 50 ns simulations. Although there are some fluctuations, as discussed in each case in the previous chapters, the interactions of lead compounds and proteins were also maintained by H-bonds or hydrophobic contacts. The residues of protein and ligands adapted to enhance these interactions. For example, some H-bonds in docking were weak and not sufficiently stable, but were enhanced after the molecular dynamics simulations to new, stronger H-bonds, improving the binding in the case of nsP2 protease. Interestingly, with the ligands that could bind to the different binding pockets, the simulation results provided details of intermolecular interactions through

occupancy of H-bonds and hydrophobic contacts to compare the binding affinities for ligand. Therefore, the more accurate binding energies were gained from the simulations results. However, there was a place for discussion of a challenge in the binding free energy calculations in the cases.

Finally, analysis of the results showed that one of the next vital steps would be the building of scaffolds for CHIKV inhibitors based on the obtained lead compounds. More importantly, identified leads should be further validated in the laboratory for their ability to actually bind to the protein target and modulate its activity, and whether there are some allosteric sites in the protein targets or not.

5.2 IDENTIFICATION OF POTENTIAL BINDING SITES IN A PROTEIN TARGET

There are different methods to explore binding pockets in the protein (outlined in the Introduction chapter). Because the ligand is not co-crystalised explicitly with the protein nsP2 or envelope glycoproteins, the location of binding site is not defined. Blind docking into an entire protein is a common approach to see which residues the ligand associates with. By analysing in a 5Å radius from where the ligand binds, these residues may be detected, and thus inferred in making up the binding pocket. The most important aspect is its ability to differentiate the real binding site from non-specific and energetically unfavorable ones based on the inbuilt scoring function. The ligand binding assay is required to verify the data in this case as there is no literature published about on this area.

Using the other programs such as MetaPocket or detection of cavities tool in Accelrys Discovery Studio can search for potential binding sites. There was an agreement between them in terms of location of binding sites. In particular, the MetaPocket program combining eight programs showed better correlations with blind docking. However, limitations discussed previously in the studies involving the envelope glycoproteins were that some detected sites were not large enough for ligand binding.

5.3 EVALUATION OF COMPUTATIONAL APPROACHES AND THEIR COMBINATION

Molecular docking is a computationally efficient method which helps to screen a large library of compounds with efficient time and low cost. Notwithstanding, it also shows some methodological deficits, for example, using one conformation of protein with the protein kept rigid, ignoring protein desolvation and most entropic terms, neglecting internal energies of docking molecules and using simplified models of solvation energies for the ligand. Therefore, a critical assessment of docking and virtual screening was undertaken. First was the evaluation of the ability to replicate a protein-bound conformation. In this study, a docking protocol was evaluated by reproducing results from experimental data. Docking provided multiple docking poses for each ligand and a RMSD for all poses with respect to the top hit. The results of the RMSD for ligand heavy atoms was within 2 Å and also showed high binding affinities with the similar residues for interactions in the nsP3-ADP-ribose complex. The scoring function, with a semi-empirical algorithm, was successful in identifying the pose similar to the co-crystal structure. Different binding conformations for each individual ligand are ranked based on affinities and the highest for each ligand compared to determine the best lead. In addition, the accuracy of docking was examined carefully by investigation of the searching convergence of docking. The increasing of a searching parameter indicated some new hits (Section 4.3.4). However, decoy molecules, which are competitive binding with ligands were not used in the screenings. Therefore, experimental studies such as ligand binding assays and cell-based assays are needed to obtain the IC_{50} value, to evaluate the accuracy of scoring *i.e.* to eliminate false positives, and to determine the validity of the screens which answer the question of whether docking provides a true ligand among top scoring compounds.

Molecular dynamics simulations were applied to consider protein flexibility and the solvent effect which are missing in docking. The simulation required significantly more computing-time and power than the docking approach, and convergence of simulation ensured. Furthermore, the similar chemical structures were clustered to search for the most popular conformation during simulations. The resulting

simulations were used to calculate binding free energy more accurately than the results from docking. However, the challenge in the LIE method lies in the parameterization of the required co-efficients, which rely on the availability of experimental data. The final limitation was the time and computational cost for simulations, such as with large protein, for example envelope glycoproteins, which could be rectified given more time.

In summary, a combination of docking and MD simulations showed a potential approach that balances the computational cost and accuracy. More importantly, the procedure can be applied in discovering therapeutic compounds for other diseases. It can also help to generate the library for CHIKV inhibitors.

CHAPTER 6. EXPERIMENTAL PROCEDURES AND METHODS

6.1 GENERAL PROCEDURE FOR MOLECULAR DOCKING AND VIRTUAL SCREENING

The process of docking and virtual screening was carried out using AutoDock Vina (version 1.1.2).[203] AutoDock Vina is an open-source program which can be downloaded from the website http://vina.scripps.edu. This program is designed in a similar manner to AutoDock, based on pairing an empirical-weighted scoring function with a global optimization algorithm.[227] Vina significantly improves the accuracy of binding mode predictions and ranking for larger molecules in terms of molecular weight and rotatable bonds.[203, 227] Vina also shortens running time in screening compared to AutoDock by using multiple CPU or CPU cores on the system. Vina can be run on 32-bit binaries or modern 64-bit machines with Windows, Linux, and Mac operating systems. In general, the protein was kept rigid while the ligands were fully flexible, however, some sidechains of residues in the protein can be selected as flexible residues in docking. The Lamarckian genetic algorithm was used to search space for docking and the binding affinity of a complex was calculated on the basis of a set of weighted energies in empirical-scoring function (a united-atom scoring function involving the heavy atoms). Calculating grid maps and assigning atom charges are not required in Vina.

A protocol was established to carry out docking and *in silico* screenings (described below). The key docking parameters were defined by the location of the docking site (centre x, y, z) and the size of a grid box of search space. The size of the box was no less than 15 Å in each dimension and larger than the size of the ligand, and no less than 22.5 Å total.[344] For virtual screening, the location and the size of the grid box were carefully investigated via blind docking (in which the box is sufficiently large to cover the whole protein) and focused docking (in which a smaller box was centred on the potential binding site of interest).

After docking, the results were analyzed and compared with available experimental data. The top ten compounds ("hits") were identified according to their binding affinities using the default scoring function in Vina. The modes of interaction between the ligands and the protein were analyzed in Accelrys DS 3.5.

6.1.1 DOCKING PROTOCOL

The docking protocol using AutoDock Vina (Figure 5.1), includes three main steps:

Step 1: Prepare a protein and a ligand (file *.pdbqt).

Step 2: Define parameters for the grid box size and centre of the box, and other parameters such as exhaustiveness and number of binding modes to generate (num_modes).

Step 3: Run molecular docking.

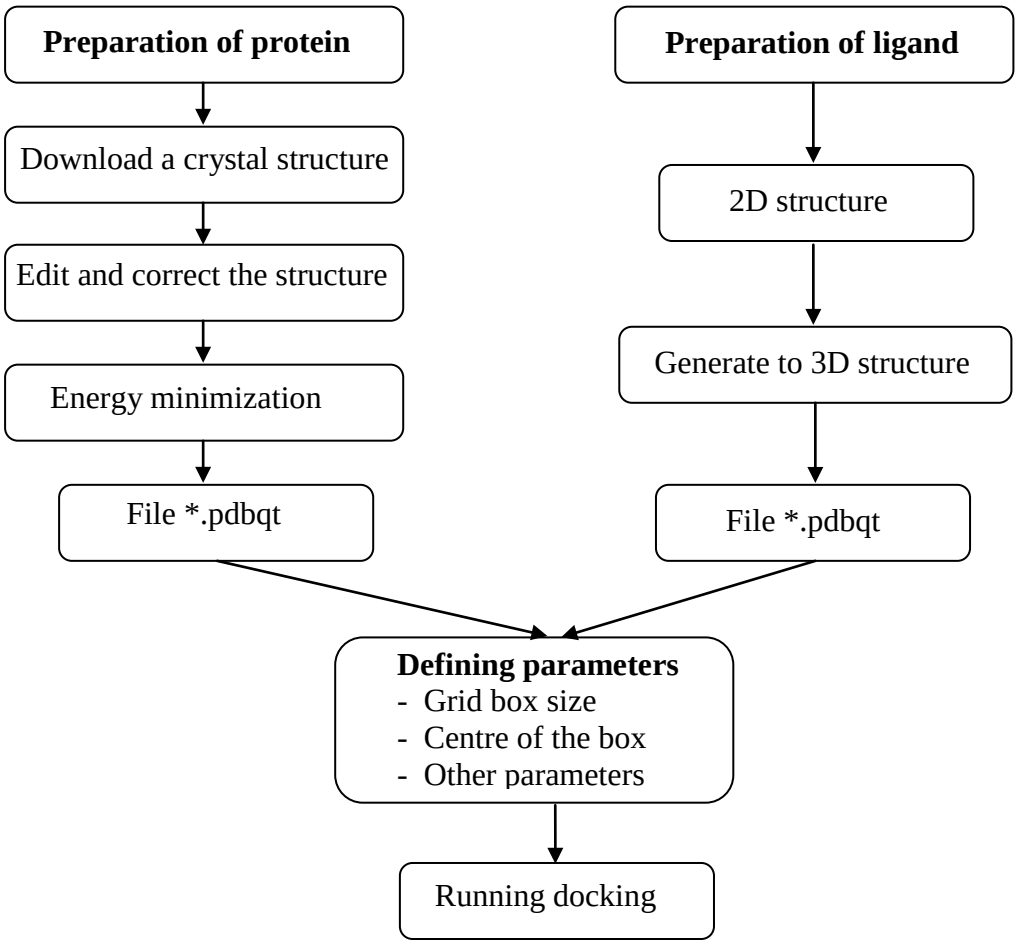


Figure 6.1. A diagram of the docking procedure in AutoDock Vina.

6.1.1.1 Preparation of a protein target

The 3D crystal structure of the protein (file *.pdb) was obtained from the Protein Data Bank website (http://www.rcsb.org), and served as the protein target in molecular docking. Details of preparation procedure were presented as follows:

- Edit and correct this structure by using Procheck (WhatIF) website (http://swift.cmbi.ru.nl/servers/html/) to upload the file to detect potential errors.
- Correct the errors from chirality (atoms CG2, CG1), side chain flip of amino acids residue His (atoms CD2, ND1, NE2 and CE1), Asn (atoms ND2 and OD1), Gln (atoms OE1 and NE2). Missing residues for a protein were added.
- Minimize the structure energy in Accelrys DS visualizer 2.0 software[342]: Apply CHARMM22 force field and set up parameters for minimization such as steepest-descent algorithm (3000 steps), and default for other parameters.
- Prepare file *.pdbqt for protein and ligand: Using AutoDock Tools to generate file *.pdbqt automatically by uploading the file pdb of protein and ligand, and then add polar hydrogens.

6.1.1.2 Preparation of a ligand

The ligand structure was generated by OpenBabel 2.3.1 based on the ChemDraw formula. In a case there is a co-crystal structure, the ligand structure could be extracted from the PDB file. The chemical libraries of the NCI Diversity Set II, which contains diverse drug-like compounds, were selected for screening. The potential hits obtained from docking can be ordered from the Open Chemical Repository Collection for future activity testing.

6.1.1.3 Docking parameters

The grid box size was chosen to be large enough to cover the ligand but as small as possible to ensure sufficient conformational sampling within available searching space. The value of search space should be no bigger than 30 Å × 30 Å × 30 Å, otherwise the

search algorithm may suffer from insufficient sampling. An increase to the value of the exhaustiveness is thus applied in the case of blind docking.

The preparation of the grid parameter file was done by using AutoDock Tools 1.5.4: Adjust the grid box manually to get the suitable parameters which cover the entire protein or the binding site, and taking note the x, y, z coordinates of the centre and sizes in the x, y, z dimension (Å) of the grid box.

Other parameters are the number of CPUs to use (--cpu), explicit random seed (--seed), exhaustiveness of the global search (--exhaustiveness, E), maximum number of binding modes to generate (--num_modes), maximum energy difference between the best binding mode and the worst one (--energy_range, unit: kcal/mol), and weight of docking algorithm (--weight). The docking parameters can be set with default values (listed in Table 5.1) or changed depending on the aims of the study. E value can be increased to expand the search space for searching conformations of complex.

Table 6.1. Default values of docking parameters.

Docking parameter	**Default value**
Number of CPUs	1
Exhaustiveness	8
Num_modess	9
Energy_range	4
Weight	1

6.1.1.4 Preparation of a configuration file (conf.txt)

File configuration includes protein and ligand file inputs (*.pdbqt), the values of centre and size of of a grid box (centre_x, centre_y, centre_z; and size_x, size_y, size_z), and docking parameters.

Details for molecular docking with different CHIKV protein targets in this study were described as follows.

6.1.2 PREDICTION OF BINDING SITES ON PROTEIN

In docking with AutoDock Vina, the grid box used for docking will be defined easily if the location of ligand binding is known. However, in some cases where the binding sites are unknown or one would like to explore alternative binding sites, computational methods such as blind docking, MetaPocket, Accelrys DS can be applied to predict potential binding sites.

6.1.2.1 MetaPocket 2.0[343]

The procedure includes three stages, namely calling based methods, meta-pocket site generation, and mapping binding residues. Given protein structure, there will be 8 predictors, namely LIGSITEcs, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity, and POCASA to identify pocket sites on the protein surface. All the predictors are run in parallel and a ranking-scores comparable, z-score is calculated separately for each pocket site in different predictors. The pockets will be clustered in terms of their spatial similarity and total z-score values. The final pocket sites are the potential ligand binding sites on protein surface. The potential binding pockets were obtained by the webserver http://projects.biotec.tu-dresden.de/metapocket/index.php, taking 5-10 minutes to get the potential pockets depending on the protein size.

6.1.2.2 Receptor cavities tool in Accelery Discovery Studio 3.5

Accelerys DS 3.5 package[342] was also applied to explore potential binding sites using a receptor cavities tool. Based on a grid search and “eraser” algorithm, the program defines location of binding sites from receptor cavities in the structure of protein. The binding sites were displayed as a set of points (points count), and their volumes were calculated as the product of number of site points and the cube of the grid spacing. These pockets were compared to those identified in blind dockings in terms of location for both of the envelop glycoprotein structures.

6.1.3 APPLYING A DOCKING PROCEDURE

6.1.3.1 The nsP3 macrodomain

The nsP3 macrodomain protein was downloaded from the Protein Data Bank (PDB id: 3GPG) and this was used as the receptor for docking. The docking protocol was established by re-docking the ADP-ribose into the nsP3 as observed in the co-crystal structure (PDB id: 3GPO). The ligand ADP-ribose was extracted from the complex crystal structure. The protein structure was edited and uploaded into the Procheck (WhatIF) website, and manually corrected some errors present in sidechains of His, Asn, and Glu flips. The next step is running energy minimization to relax the structure and remove steric overlaps. With a Cα RMSD of 0.59 Å between the minimized structure and the X-ray structure, the minimized structure was utilized for the docking process. Polar hydrogen atoms were added and save into file *.pdbqt format in AutoDock Tools.

Using AutoDock Tools, the binding site was defined as the place where ADP-ribose binds. The values of different grid box dimensions and its centre were used for docking and screenings, listed in Table 5.2, (See more discussion in Chapter 2).

Table 6.2. Parameters of a grid box in different docking and virtual screenings for the nsP3 protein.

Docking and virtual screening	x, y, z coordinates	Dimension of a grid box
Pocket 1	9.7 Å, 43.0 Å, -13.2 Å	16 Å × 16 Å × 16 Å
Pocket 2	9.7 Å, 43 Å, -13.2 Å	20 Å × 20 Å × 20 Å
Blind docking (at the centre of Pocket 1)	9.7 Å, 43 Å, -13.2 Å	50 Å × 50 Å × 56 Å
Blind docking (at the centre of the protein)	7.7 Å, 45.3 Å, -5.3 Å	50 Å × 50 Å × 56 Å
Pocket 2	7.7 Å, 45.4 Å, 11.5 Å	20 Å × 20 Å × 20 Å
Pocket 3	2.3 Å, 44.6 Å, -18.3 Å	20 Å × 20 Å × 20 Å

6.1.3.2 The nsP2 enzyme

The 3D crystal structure of the nsP2 protease of CHIKV was obtained from the Protein Data Bank (PDB id: 3TRK) and served as a protein target in docking. Using Procheck (WhatIF), no errors were found. Water molecules and the other co-factors were deleted.

The structures of the four best compounds taken from the previous study[351] were drawn in ChemDraw (ChemBioDraw Ultra 13.0) and then converted to 3D structure using OpenBabel. These compounds were docked into the nsP2 protease based on the key residues of Pocket 1 identified in the previous study.[351] The preparation of docking was the same as previously, and Pocket 1 was confirmed. The blind dockings, with a grid box size of 60 Å × 70 Å × 60 Å at the centre of protein, and with a grid box size of 66 Å × 86 Å × 70 Å at the centre of the binding site, were carried out. Focused docking with a smaller box (20 Å × 20 Å × 20 Å), centred on the potential binding sites of interest was then undertaken. Additionally, the MetaPocket was used to identify potentially pocket sites of the nsP2 other than the active site.[343]

The values of different grid box dimensions and its centres were used for docking and different screenings, listed in Table 5.3, (See more discussion in Chapter 3).

Table 6.3. Parameters of a grid box in different docking and virtual screenings for the nsP2 protease.

Docking and virtual screening	x, y, z coordinates	Dimension of a grid box
Blind docking (at the centre of Pocket 1)	12.4 Å, 34.3 Å, 28.6 Å	66 Å × 86 Å × 60 Å
Blind docking (at the centre of Pocket 2)	14.3 Å, 25.5 Å, 22.3	60 Å × 70 Å × 60 Å
Pocket 1	12.4 Å, 34.3 Å, 28.6 Å	16 Å × 16 Å × 16 Å
Pocket 1	12.4 Å, 34.3 Å, 28.6 Å	20 Å × 20 Å × 20 Å
Pocket 2	28.1 Å, 36.9 Å, 30.6 Å	20 Å × 20 Å × 20 Å
Pocket 3	30.4 Å, 42.1 Å, 14.8 Å	20 Å × 20 Å × 20 Å
Pocket 4	3.2 Å, 31.9 Å, 18.1 Å	20 Å × 20 Å × 20 Å
Pocket 5	15.2 Å, 13.9 Å, 11.5 Å	20 Å × 20 Å × 20 Å

6.1.3.3 The envelope glycoproteins

Two X-ray crystal structures of envelope glycoprotein, an immature form (PDB id: 3N40) and a mature form (PDB id: 3N42), were downloaded from the Protein Data Bank. The dimensions of the box covering 3N40 and 3N42 were set to 108 Å × 126 Å × 62 Å and 84 Å × 126 Å × 74 Å, respectively. Potential binding sites were revealed by the location of ligands binding into the target. The next step was a focused docking, in which a smaller box (20 Å × 20 Å × 20 Å), was centred at the potential binding site of interest. The receptor cavities tool in Accelrys DS and MetaPocket program were also used to predict other potential binding sites of the envelope proteins.[343]

The values of different grid box dimensions, and their centres, were used for docking and different screenings, listed in Table 5.4, (See more discussion in Chapter 4).

Table 6.4. Parameters of a grid box in different docking and virtual screenings for envelope glycoprotein complexes.

Docking and virtual screening	x, y, z coordinates	Dimension of a grid box
Blind docking (at the centre of the 3N40)	23.7 Å, 1.8 Å, 23.4 Å	108 Å × 126 Å × 62 Å
Blind docking (at the centre of the 3N42)	26.4 Å, 7.1 Å, -24.5 Å	84 Å × 126 Å × 74 Å
Pocket 1 of the 3N40	-22.3 Å, 18.3 Å, 26.3 Å	20 Å × 20 Å × 20 Å
Pocket 2 of the 3N40	-15.9 Å, 1.5 Å, 19.7 Å	20 Å × 20 Å × 20 Å
Pocket 1of the 3N42	-21.3 Å, -11.7 Å, -24.9 Å	20 Å × 20 Å × 20 Å
Pocket 2 of the 3N42	-22.4 Å, -4.2 Å, -11.9 Å	20 Å × 20 Å × 20 Å
Pocket 3 of the 3N42	-5.7 Å, 31.2 Å, -8.9 Å	20 Å × 20 Å × 20 Å
Pocket 4 of the 3N42	-44.7 Å, -39.8 Å, -38.2 Å	20 Å × 20 Å × 20 Å

6.2 MOLECULAR DYNAMICS SIMULATIONS

The program NAMD[265] was used for MD simulations to investigate the stability and flexibility of the hit-target complexes and study their interactions. In this study, a general strategy for MD simulations was performed as followings (Figure 5.2):

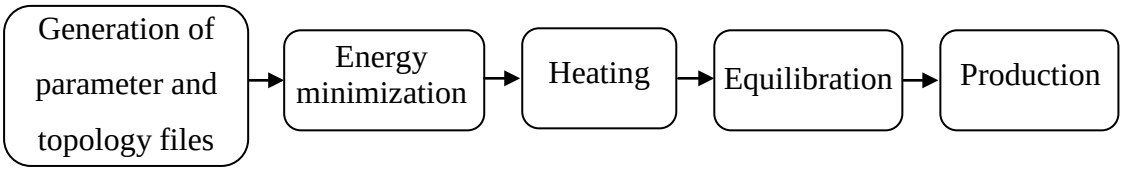


Figure 6.2. A general scheme for MD simulations.

It includes the following steps:

Step 1: Generation of the force field parameters (including the parameter and topology files) for simulation for protein and ligand. The protein atoms were treated with the CHARMM22 force field[257] and the corresponding parameters for the ligands were generated with AmberTools.[262]

Step 2: The systems were solvated in a cubic box of TIP3P water molecules that extends 10 Å from the edge. Sodium counterions were added to neutralize the system by randomly replacing the water molecules to mimic the physiological concentration of 0.15 M NaCl. All simulations were run under the periodic boundary conditions.

Step 3: Energy minimization for the system was conducted using a conjugate gradient method to remove the bad contacts.

Step 4: The system was subjected to a heating procedure. The temperature was set to 298.15 K, and the pressure to 1 atm. The Langevin algorithm was used to maintain the temperature and pressure coupling. The PME algorithm was used to compute long range electrostatic interactions.[269] The cutoff distance for vdW interactions was set at 12 Å, and the pair-list distance was 13.5 Å.

Step 5: Running equilibrium and production simulations. The equilibrium phase with weak harmonic restraints on heavy atoms for 3 ns was carried out. This brings the system to equilibration state when the calculated average temperature, pressure, and energies were converged or stabilized. The production phase of the simulation was run subsequently and results of MD simulations were analysed based on the production phase.

All of the simulations were run on the high performance computation (HPC) clusters. The HPC cluster is comprised of large compute nodes (Model Dell PowerEdge C6145) with 16 core-processors, and 256 GB memory. System software provided includes Linux or Mac OS client and Windows system operating system. The clusters are connected with Interconnect technology from the Linux OS or MS Windows or DOS.

Time for running depends on system size. For the nsP3 macrodomain and its complexes: 12 simulations were run, including one simulation of apo protein and 11 for the complexes of nsP3 and ligands. For the nsP2 protease and its complexes: the number of systems was 14 including one simulation of apo protein and 13 for the complexes of nsP2 enzyme and ligands. For envelope glycoprotein complexes: due to the large sizes of these glycoproteins complexes, the simulations for these complexes could not be done in this study.

Analysis of simulations: The trajectories for analysis were saved every 10 ps. To determine the system stability, the value of RMSD of the heavy atoms over 50 ns was calculated with respect to the starting structure versus the simulation time. The RMSF of Cα atoms during the MD simulations was measured to obtain information on the local flexibility of the system. The resulting trajectories such as hydrogen bonding and hydrophobic contact interactions were analyzed by the CHARMM22[263] and VMD (version 1.9.1).[276] A hydrogen bond was defined as the distance between proton donor (D) and acceptor (A) atoms was $\leq$ 3.5 Å, and the angle $D-H\cdots A$ was $\geq 120^{\circ}$.[70] If occupancy of HBs was more than 75%, they were considered as strong HBs whereas HBs occupations of less than 50%, were designated as medium.[70] The hydrophobic contacts were also monitored based on a cutoff distance between the carbon atoms of non-polar parts of residues of proteins and of ligand of 4.0 Å.[277-278] The Clustering Plugin in VMD is applied for clustering analysis for the snapshots from the trajectories.

6.2.1 SIMULATIONS FOR PROTEIN

MD simulations were conducted for the apo protein and the hit-target complexes obtained from virtual screening.

6.2.1.1 The nsP3 macrodomain

MD simulations were carried out for the apo state (PDB id: 3GPG), and 11 ligand-nsP3 complexes. The total number of atoms (including protein, water, and counterions) in the CHIKV system was about 18,000; and the size of the initial water box was about 58 Å × 58 Å × 58 Å.

6.2.1.2 The nsP2 enzyme

MD simulations were conducted for the apo state (PDB id: 3TRK), the 13 complexes of the nsP2 protease, and hit compounds obtained from virtual screening. The total number of atoms of protein, water, and counterions in the CHIKV system was about 70,000; and the dimension of the initial water box was about 90 Å × 90 Å × 90 Å.

6.2.2 SIMULATIONS FOR LIGANDS

Additional simulations of ligands in water were conducted to calculate the binding free energies with the LIE method, described in 5.3 section. The other parameters of setting up MD simulations were the same as previously used for protein.

6.3 BINDING FREE ENERGY CALCULATIONS

There have been different approaches developed over the years to calculate the ligand binding free energies such as free energy perturbation, thermodynamic integration, and LIE.[316, 328] In this study, an LIE approach was applied to calculate the absolute binding free energies for ligands in complexes with the protein targets, the nsP3 macrodomain and the nsP2 enzyme. The main reason for this approach is that this method compromises between speed and accuracy, and so is more practically feasible for large ligands. In addition, we can utilize the results of MD simulations of proteins. Derived from the linear response theory, the binding free energies is approximated as

$$\Delta G_{bind} = \alpha \langle V_{bound}^{vdw} - V_{unbound}^{vdw} \rangle + \beta \langle V_{bound}^{elec} - V_{unbound}^{elec} \rangle + \gamma \qquad \text{(Equation 1.16)}$$

The various van der Waals and electrostatic components can be obtained from two simulations: one with the ligand in aqueous solution, and the other with the protein (nsP3 or nsP2)-ligand complex in aqueous solution. The NAMDEnergy plugin in VMD was utilized to compute the energy components over the frames obtained from the MD simulations.[265, 276]

In Equation 1.16, α, β, and γ are empirical parameters. α is often set to 0.18 for a wide variety of ligand-protein systems. The β represents a function of the chemical nature of the ligand, so in principle the value of β can be parameterized from explicit solvent free energy calculations of different chemical entities. The γ parameter can be set $\gamma = 0$ or $\gamma \neq 0$ depending on the hydrophobicity of the binding site to estimate absolute free energies of binding.

CHAPTER 7. OUTLOOK AND FUTURE DIRECTIONS

During the past decade, computational approaches have shown their success and power in assisting drug discovery and development. With the emergence or re-emergence of some alphaviruses, such as CHIKV, a neglected tropical disease in more than 55 countries in the word, searching for potential inhibitors to combat the virus in efficient time, and in a cost-effective manner is significantly important. Taking advantage of all available information on CHIKV, this study examined major potential targets of CHIKV. The non-structural proteins nsP3 macrodomain and nsP2 protease were used to target the viral replication and transcription (Chapter 2 and Chapter 3). The envelope glycoprotein complexes were another potential target as they direct virus entry and attachment (Chapter 4). A combination of computational tools; including molecular docking, virtual screening, molecular dynamics simulations, and binding free energy calculations, were used to identify a number of lead compounds for potential CHIKV disease treatment. The insights into the CHIKV inhibitors and their complexes, as well as their interaction energies, were elucidated. Our findings provide a foundation which would be helpful for further research on rational antiviral drug design, especially the design of CHIKV inhibitors.

7.1 NSP3 MACRODOMAIN PROTEIN

Chapter 2 described the attempts at drug discovery targeting the nsP3 macrodomain. The nsP3 is considered a potential target for CHIKV drug discovery because of its crucial role in the early stages of transcription for viral replication. With the goal of identifying potential novel inhibitors for the CHIKV, this study utilised the available crystal structures, nsP3 macrodomain (PDB id: 3GPG) and its complex with ADP-ribose (PDB id: 3GPO) as a starting point. Initially, reproducing the experimental data with respect to the co-crystal structure of nsP3-ADP-ribose showed that ligand ADP-ribose binds tightly into the nsP3 (binding affinity of less than -10 kcal/mol). This indicated a successful and robust docking protocol with AutoDock Vina. Virtual screening based on this protocol was conducted with the library database compounds (NCI Diversity Set II). The top hit compounds and three binding pockets were

identified. In virtual screening, blind dockings covering the entire protein, followed by focused dockings to a specific pocket of protein, were a robust strategy to detect potential binding pockets. An additional strategy used the MetaPocket program to confirm these pockets, with consistent results. Pocket 1 is the well-characterised adenine binding pocket, and Pocket 2/Pocket 3 are two additional binding pockets. Moreover, in terms of binding affinity, Pocket 1 appears to be the most favourable for ligand binding (binding affinities of less than -10 kcal/mol), while Pocket 3 appears to be the least favourable one. Some ligands bind to different pockets with the different conformations. Based on the binding affinities and analysis of interactions, some ligands showed greater potential than others. They are NCI_61610 (-11.1 kcal/mol), NCI_345647_a (-10.9 kcal/mol), NCI_25457 (-10.8 kcal/mol), NCI_670283 (-10.6 kcal/mol), NCI_37168 (around -10.0 kcal/mol), NCI_372499_b (-10.3 kcal/mol), and NCI_324623 (-10.3 kcal/mol). However, the question of ligand selectivity to the pocket requires more experimental work. Analysis in terms of HBs and hydrophobic interactions revealed key residues for interactions with the ligands in the nsP3.

Furthermore, equilibrium MD simulations for most of the top docked ligand-nsP3 complexes were carried out to validate the molecular docking results. The stability of these complexes over 50 ns was reported, and there were subtle structural rearrangements in the nsP3 observed to better accommodate the ligands. Throughout analysis of MD simulations and binding free energy calculations, insights into the interactions between ligands and the protein nsP3 were observed at an atomic level. Consequently, the compounds to be considered for biological testing, are NCI_61610 (-15.2 kcal/mol), NCI_670283 (-13.2 kcal/mol), NCI_293778 (-12.7 kcal/mol), NCI_345647_a (-12.1 kcal/mol), NCI_37168 (-12.0 kcal/mol), and NCI_25457 (-10.0 kcal/mol).

Utilizing the simulation results of the nsP3 and its complexes, the LIE approach was applied to calculate binding free energies for the top hits obtained from docking. The LIE method is one of the most powerful approaches in the estimation of binding free energy for a ligand. Given the approximations used in the LIE approach, careful selection of parameters was provided. The comparison between the top-hit compounds and the tenth-hit compounds in terms of their binding free energy showed a close

agreement of docking and simulations in predicting binding affinities. The combination of the results from docking, MD simulations and binding free energy calculations provides a list of compounds was proposed for biological testing including NCI_61610, NCI_345647_a, NCI_25457, NCI_670283, and NCI_37168.

For further work, the identified potential inhibitors should be experimentally tested their ability to inhibit CHIKV verified by *in vivo* studies. The key residues for interaction will be further investigated on their contributions, e.g. through HBs interactions, or the vdW to probe their influence on the binding process. The results are also useful for constructing a pharmacophore for future ligand-based drug design to generate the chemical libraries for novel inhibitors targeting CHIKV. Strategies for the synthesis and subsequent lead optimization are also a major next consideration.

7.2 NSP2 PROTEASE ENZYME

Drug discovery targeting the nsP2 protease was presented in Chapter 3, using the same process as described in Chapter 2. The nsP2 protease is a potential target with an important role in the cleavage of the non-structural polyproteins into non-structural proteins in the viral replication cycle, and transcription shut-off. This study is the first instance where molecular modelling using the X-ray structure has been undertaken. Due to lack of the structural information about the binding between ligands and the nsP2 protease, no information about its potential binding pockets was forthcoming; blind docking was crucial to discover the possible binding pockets in the enzyme, and to detect hit compounds which may bind to these sites. Blind dockings were also based on previously published information, defining the binding site based on a homology model of the nsP2 protease and its potential inhibitors. In our study, the top hit compounds and five potential binding pockets of the nsP2 protease were identified. Focused dockings into these different binding sites offered more hit compounds and provided further binding modes at these pockets. Pocket 4 in the N-terminal domain of the nsP2 protease was identified based on structural alignment as the active site (through the presence of catalytic residues, Cys1013 and His1083) while most of previous studies reported Pocket 1 as the active site. The results from docking and virtual screenings indicate that

the potential compounds are NCI_61610 (-10.6 kcal/mol), NCI_293778 (from -9.8 to around -10.2 kcal/mol), NCI_37553 (-9.6 kcal/mol), NCI_217697 (-9.3 kcal/mol), and NCI_25457 (-9.0 kcal/mol).

The results of top hit compounds after equilibrium MD simulations for 50 ns indicated most of the systems were maintained by interactions from hydrophobic contacts rather than hydrogen bonds. The LIE approach was used to obtain the accurate binding affinities of different ligands based on their trajectories from MD simulations. The different binding energies were explained by the contribution of varying numbers of HBs and hydrophobic contacts. The lead compounds selected were NCI_293778 (from -9.9 to around -13 kcal/mol), NCI_67436 (-15.5 kcal/mol), and NCI_217697 (-9.0 kcal/mol). The relative ranking of binding free energies of top-hit compounds gave a good agreement between docking results and MD simulations results.

After combining the results obtained from docking, MD simulations and binding free energy calculations, proposed compounds for biological testing including NCI_293778, NCI_217697, and NCI_67436.

Moreover, as there are some potential binding pockets other than the active site in the structure of nsP2 enzyme, so further experimental work is required to investigate the effect upon binding to these alternate pockets. More experimental data is also needed to elucidate the contribution and influence of vdW, and electrostatic energy on ligand binding. Calculations of interaction energy components from key residues should be considered. The study opens up a promising approach to rational drug design (pharmacophore constructing with the generated scaffold) for the design of inhibitors for CHIKV. Synthesis of the lead compounds followed by lead optimization, is also an important next step.

7.3 ENVELOPE GLYCOPROTEINS

Envelope glycoproteins are another class attractive targets for inhibition of virus entry or virus attachment as inhibition of the viral envelope glycoproteins can block the *in*

vitro CHIKV infection. Chapter 4 focused on the discovery of inhibitors for CHIKV targeting the envelope glycoproteins complexes. Computational investigation of the envelope glycoproteins complexes (the immature and mature forms) has previously been explored within our group, however a more rigorous exploration based on blind docking was deemed essential to the full characterization of potential target sites. Virtual screenings based on blind docking and focused docking explored the potential binding pockets and inhibitors for both the immature and mature structure of envelope proteins. Promising hit compounds were identified for two complexes of the envelope glycoproteins. Pocket 2 was a novel binding pocket in the immature structure. Pocket 2 and Pocket 3 were detected as novel binding sites for the mature complex of the glycoproteins. The key residues involved in stabilizing the complex or participating in the fusion process were confirmed. This study also investigated the convergence of conformational sampling in AutoDock Vina. The results demonstrated that this parameter affected a searching process for conformations of ligand in docking however in the current study, the different screening provided a semi-quantitatively consistent picture. The binding affinities of most of top hits compounds were about -10 to -13 kcal/mol indicating good binding between the ligands and the envelope glycoprotein complexes. The compounds bound into Pocket 1 and Pocket 4 were expected to be promising targeting the glycoprotein complexes, as they could interact with the residues involved in the process of virus attachment and virus entry. Therefore, the potential compounds proposed for biological testing include NCI_293778 (from -11.4 to around -13.7 kcal/mol), NCI_61610 (from -11.6 to -12.3 kcal/mol), NCI_37553 (-11.5 kcal/mol), and NCI_156219 (-11.1 kcal/mol).

This work supported the current docking protocol, utilising AutoDock Vina as a robust strategy, which could be used to identify inhibitors and understand the binding modes. However, in docking, the protein structure was kept rigid and this might affect both the docking and scoring process. More accurate binding free energies are expected to be obtained from explicit solvent based MD simulations, however, due to the larger sizes of these envelope glycoprotein complexes, this could not be conducted within the time frame of this work.

In future work, experimental studies are required to determine the inhibitory effects of hit compounds on these membrane proteins. It would be useful to investigate the stability and flexibility of the systems. The mechanism of virus entry and virus attachment through the envelope glycoproteins would require further MD simulations experimental data. The binding free energy calculations would be also needed to guide rational antiviral drug design.

REFERENCES

1. Cavrini, F.; Gaibani, P.; Pierro, A. M.; Rossini, G.; Landini, M. P.; Sambri, V., Chikungunya: an emerging and spreading arthropod-borne viral disease. *J. Infect. Dev. Ctries.* **2009**, 3, 744-752.
2. Thiboutot, M. M.; Kannan, S.; Kawalekar, O. U.; Shedlock, D. J.; Khan, A. S.; Sarangan, G.; Srikanth, P.; Weiner, D. B.; Muthumani, K., Chikungunya: a potentially emerging epidemic? *Plos Negl. Trop. Dis.* **2010**, 4, 1-8.
3. Staples, J. E.; Breiman, R. F.; Powers, A. M., Chikungunya fever: an epidemiological review of a re-emerging infectious disease. *Clin. Infect. Dis.* **2009**, 49, 942-948.
4. Townson, H.; Nathan, M. B., Resurgence of chikungunya. *Trans. R. Soc. Trop. Med. Hyg.* **2008**, 102, 308-309.
5. Schwartz, O.; Albert, M. L., Biology and pathogenesis of chikungunya virus. *Nat. Rev. Microbiol.* **2010**, 8, 491-500.
6. Sudeep, A. B.; Parashar, D., Chikungunya: an overview. *J. BioSci.* **2008**, 33, 443-449.
7. Thiberville, S.-D.; Moyen, N.; Dupuis-Maguiraga, L.; Nougairede, A.; Gould, E. A.; Roques, P.; de Lamballerie, X., Chikungunya fever: epidemiology, clinical syndrome, pathogenesis and therapy. *Antivir. Res.* **2013**, 99, 345-370.
8. Rashad, A. A.; Mahalingam, S.; Keller, P. A., Chikungunya Virus: Emerging Targets and New Opportunities for Medicinal Chemistry. *J. Med. Chem.* **2014**, 57, 1147-1166.
9. Xavier de, L.; Laetitia, N.; Remi, C., Antiviral Treatment of Chikungunya Virus Infection. *Infect. Disord. Drug Targets* **2009**, 9, 101-104.
10. Parashar, D.; Cherian, S., Antiviral Perspectives for Chikungunya Virus. *BioMed. Res. Int.* **2014**, 2014, 1-11.
11. Robinson, M. C., An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952-53. I. Clinical features. *Trans. R. Soc. Trop. Med. Hyg.* **1955**, 49, 28-32.
12. Scott, C. W., Arrival of Chikungunya Virus in the New World: Prospects for Spread and Impact on Public Health: e2921. *Plos Neglect. Trop. Dis.* **2014**, 8, e2921: 1-4.
13. Lumsden, W. H. R., An epidemic of virus disease in Southern province, Tanganyika territory, in 1952-53. II. General description and epidemiology. *Trans. R. Soc. Trop. Med. Hyg.* **1955**, 49, 33-57.
14. Weinbren, M. P.; Haddow, A. J.; Williams, M. C., The occurrence of chikungunya virus in Uganda I. Isolation from mosquitoes. *Trans. R. Soc. Trop. Med. Hyg.* **1958**, 52, 253-262.
15. Pulmanausahakul, R.; Roytrakul, S.; Auewarakul, P.; Smith, D. R., Chikungunya in Southeast Asia: understanding the emergence and finding solutions. *Int. J. Infect. Dis.* **2011**, 15, e671-e676.
16. Burt, F. J.; Rolph, M. S.; Rulli, N. E.; Mahalingam, S.; Heise, M. T., Chikungunya: a re-emerging virus. *Lancet* **2012**, 379, 662-671.
17. Sasayama, M.; Kurosu, T.; Chaichana, P.; Maneekan, P.; Ikuta, K.; Ramasoota, P.; Okabayashi, T.; Singhasivanon, P.; Luplertlop, N.;

Benjathummarak, S.; Kawashita, N.; Rukmanee, P.; Sangmukdanun, S.; Masrinoul, P.; Pitaksajjakul, P.; Puiprom, O.; Wuthisen, P., Chikungunya virus was isolated in Thailand, 2010. *Virus Gen.* **2014**, 49, 485-489.
18. Duong, V.; Laurent, D.; Hok, B.; Roces, M. C.; Ong, S.; Char, M. C.; Deubel, V.; Tarantola, A.; Buchy, P.; Andries, A.-C.; Ngan, C.; Sok, T.; Richner, B.; Asgari-Jirhandeh, N.; Bjorge, S.; Huy, R.; Ly, S., Reemergence of Chikungunya virus in Cambodia. *Emerg. Infect. Dis.* **2012**, 18, 2066-2069.
19. Laras, K.; Hartati, S.; Beckett, C.; Sedyaningsih, E. R.; Beecham, r. H. J.; Corwin, A. L.; Sukri, N. C.; Larasati, R. P.; Bangs, M. J.; Kosim, R.; Djauzi; Wandra, T.; Master, J.; Kosasih, H., Tracking the re-emergence of epidemic chikungunya virus in Indonesia. *Trans. R. Soc. Trop. Med. Hyg.* **2005**, 99, 128-141.
20. Powers, A. M.; Logue, C. H., Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *J. Gen. Virol.* **2007**, 88, 2363-2377.
21. Her, Z.; Kam, Y.-W.; Lin, R. T. P.; Ng, L. F. P., Chikungunya: a bending reality. *Micro. Infect.* **2009**, 11, 1165-1176.
22. Singh, S. K.; Unni, S. K., Chikungunya virus: host pathogen interaction. *Rev. Med. Virol.* **2011**, 21, 78-88.
23. Lam, S. K.; Chua, K. B.; Hooi, P. S.; Rahimah, M. A.; Kumari, S.; Tharmaratnam, M.; Chuah, S. K.; Smith, D. W.; Sampson, I. A., Chikungunya infection- An emerging disease in Malaysia. *Southeast Asian J. Trop. Med. Public Health* **2001**, 32, 447-451.
24. AbuBakar, S.; Sam, I. C.; Wong, P.-F.; MatRahim, N.; Hooi, P.-S.; Roslan, N.; NorAziyah, M., Reemergence of endemic Chikungunya, Malaysia. *Emerg. Infect. Dis.* **2007**, 13, 147-149.
25. Pastorino, B.; Muyembe-Tamfum, J. J.; Bessaud, M.; Tock, F.; Tolou, H.; Durand, J. P.; Peyrefitte, C. N., Epidemic resurgence of Chikungunya virus in democratic Republic of the Congo: identification of a new central African strain. *J. Med. Virol.* **2004**, 74, 277-282.
26. Hochedez, P.; Jaureguiberry, S.; Debruyne, M.; Bossi, P.; Hausfater, P.; Brucker, G.; Bricaire, F.; Caumes, E., Chikungunya infection in travelers. *Emerg. Infect. Dis.* **2006**, 12, 1565-1567.
27. Jentes, E. S.; Coulibaly, M.; Bausch, D. G.; Bryan, J.; Robinson, J.; Johnson, B. W.; Conde, I.; Sakouvougui, Y.; Iverson, J.; Beecher, S.; Bah, M. A.; Diakite, F., Acute arboviral infections in Guinea, West Africa, 2006. *Am. J. Trop. Med. Hyg.* **2010**, 83, 388-394.
28. Krishnamoorthy, K.; Harichandrakumar, K. T.; Krishna Kumari, A.; Das, L. K., Burden of chikungunya in India: estimates of disability adjusted life years (DALY) lost in 2006 epidemic. *J. Vector Borne Dis.* **2009**, 46, 26-35.
29. Mavalankar, D.; Shastri, P.; Raman, P., Chikungunya epidemic in India: a major public-health disaster. *Lancet Infect. Dis.* **2007**, 7, 306-307.
30. Centers for Disease, C.; Prevention, Chikungunya fever diagnosed among international travelers-United States, 2005-2006. *Morb. Mort. Week. Rep.* 2006, pp 1040-1042.
31. Johnson, D. F.; Druce, J. D.; Chapman, S.; Swaminathan, A.; Wolf, J.; Richards, J. S.; Korman, T.; Birch, C.; Richards, M. J., Chikungunya virus infection in travellers to Australia. *Med. J. Aust.* **2008**, 188, 41-43.
32. Peyrefitte, C. N.; Daries, W.; Ollivier, L.; Pages, F.; Martin, R.; Boniface, F.; Tolou, H. J.; Grandadam, M.; Bessaud, M.; Pastorino, B. A. M.; Gravier, P.;

Plumet, S.; Merle, O. L.; Moltini, I.; Coppin, E.; Tock, F., Circulation of Chikungunya virus in Gabon, 2006–2007. *J. Med. Virol.* **2008**, 80, 430-433.
33. Ng, L. C.; Hapuarachchi, H. C., Tracing the path of Chikungunya virus - Evolution and adaption. *Infect. Genet. Evol.* **2010**, 10, 876-885.
34. Peyrefitte, C. N.; Paupy, C.; Vessiere, A.; Imbert, P.; Tchendjou, P.; Durand, J.-P.; Tolou, H. J.; Grandadam, M.; Rousset, D.; Pastorino, B. A. M.; Pouillot, R.; Bessaud, M.; Tock, F.; Mansaray, H.; Merle, O. L.; Pascual, A. M., Chikungunya virus, Cameroon, 2006. *Emerg. Infect. Dis.* **2007**, 13, 768-771.
35. Kumar, N. C. V. M.; Gopal, S. D. V. R., Reemergence of Chikungunya virus in Indian Subcontinent. *Indian J. Virol.* **2010**, 21, 8-17.
36. Ng, L.-C.; Lin, R. T. P.; Leo, Y.-S.; Tan, B.-H.; Han, H.-K.; Ooi, P.-L. S.; James, L.; Khoo, S.-P.; Tan, L.-K.; Tan, C.-H.; Tan, S. S. Y.; Hapuarachchi, H. C.; Pok, K.-Y.; Lai, Y.-L.; Lam-Phua, S.-G.; Bucht, G., Entomologic and virologic investigation of chikungunya, Singapore. *Emerg. Infect. Dis.* **2009**, 15, 1243-1249.
37. Tiawsirisup, S., Chikungunya virus in Thailand; An update. *Thai J. Vet. Med.* **2011**, 41, 133-134.
38. Rianthavorn, P.; Prianantathavorn, K.; Wuttirattanakowit, N.; Theamboonlers, A.; Poovorawan, Y., An outbreak of chikungunya in southern Thailand from 2008 to 2009 caused by African strains with A226V mutation. *Int. J. Infect. Dis.* **2010**, 145, e161-e165.
39. Sam, I. C.; Chan, S. Y.; Chan, Y. F.; Loong, S. K.; Chin, H. K.; Hooi, P. S.; Ganeswrie, R.; AbuBakar, S., Chikungunya virus of Asian and Central/East African genotypes in Malaysia. *J. Clin. Virol.* **2009**, 46, 180-183.
40. Huang, J.-H.; Yang, C.-F.; Su, C.-L.; Chang, S.-F.; Cheng, C.-H.; Yu, S.-K.; Lin, C.-C.; Shu, P.-Y., Imported chikungunya virus strains, Taiwan, 2006-2009. *Emerg. Infect. Dis.* **2009**, 15, 1854-1856.
41. Lanciotti, R. S.; Kosoy, O. L.; Laven, J. J.; Panella, A. J.; Velez, J. O.; Lambert, A. J.; Campbell, G. L., Chikungunya virus in US travelers returning from India, 2006. *Emerg. Infect. Dis.* **2007**, 13, 764-767.
42. Gibney, K. B.; Fischer, M.; Prince, H. E.; Kramer, L. D.; St. George, K.; Kosoy, O. L.; Laven, J. J.; Staples, J. E., Chikungunya Fever in the United States: A Fifteen Year Review of Cases. *Clin. Infect. Dis.* **2011**, 52, e121-e126.
43. Hassan, R.; Rahman, M. M.; Moniruzzaman, M.; Rahim, A.; Barua, S.; Biswas, R.; Biswas, P.; Mowla, S. G. M.; Chowdhury, M. J., Chikungunya - an emerging infection in Bangladesh: a case series. *J. Med. Case Rep.* **2014**, 8, 1-3.
44. Grandadam, M.; Leparc-Goffart, I.; Desprès, P.; Caro, V.; Plumet, S.; Thiberge, J. M.; Souarès, Y.; Failloux, A.-B.; Tolou, H. J.; Budelot, M.; Cosserat, D., Chikungunya virus, southeastern France. *Emerg. Infect. Dis.* **2011**, 17, 910-913.
45. Wu, D.; Zhou, H.; He, J.; Li, L.; Yang, X.; Wu, J.; Zhang, Q.; Zhong, H.; Ke, C.; Deng, X.; Guan, D.; Li, H.; Zhang, Y., Chikungunya outbreak in Guangdong Province, China, 2010. *Emerg. Infect. Dis.* **2012**, 18, 493-495.
46. Viennet, E.; Knope, K.; Faddy, H. M.; Williams, C. R.; Harley, D., Assessing the threat of chikungunya virus emergence in Australia. *Commun. Dis. Intell. Q. Rep.* **2013**, 37, e136-e143.

47. Wangchuk, S.; Chinnawirotpisan, P.; Dorji, T.; Tobgay, T.; Dorji, T.; Yoon, I.-K.; Fernandez, S., Chikungunya fever outbreak, Bhutan, 2012. *Emerg. Infect. Dis.* **2013**, 19, 1681-1684.
48. Horwood, P. F.; Siba, P. M.; Pavlin, B. I.; Reimer, L. J.; Dagina, R.; Susapu, M.; Bande, G.; Katusele, M.; Koimbu, G.; Jimmy, S.; Ropa, B., Outbreak of chikungunya virus infection, Vanimo, Papua New Guinea. *Emerg. Infect. Dis.* **2013**, 19, 1535-1538.
49. Wanlapakorn, N.; Thongmee, T.; Linsuwanon, P.; Chattakul, P.; Vongpunsawad, S.; Payungporn, S.; Poovorawan, Y., Chikungunya outbreak in Bueng Kan province, Thailand, 2013. *Emerg. Infect. Dis.* **2014**, 20, 1404-1406.
50. Nhan, T.-X.; Claverie, A.; Roche, C.; Teissier, A.; Colleuil, M.; Baudet, J.-M.; Cao-Lormeau, V.-M.; Musso, D., Chikungunya virus imported into French Polynesia, 2014. *Emerg. Infect. Dis.* **2014**, 20, 1773-1774.
51. Van Bortel, W.; Cesaire, R.; Najioullah, F.; Ardillon, V.; Balleydier, E.; Carvalho, L.; Lemaître, A.; Noel, H.; Servas, V.; Six, C.; Zurbaran, M.; Dorleans, F.; Leon, L.; Guinard, A.; van den Kerkhof, J.; Henry, M.; Fanoy, E.; Braks, M.; Reimerink, J.; Swaan, C.; Georges, R.; Brooks, L.; Rosine, J.; Freedman, J.; Sudre, B.; Zeller, H.; Blateau, A.; Rousset, D.; Matheus, S.; Leparc-Goffart, I.; Flusin, O.; Prat, C., Chikungunya outbreak in the Caribbean region, December 2013 to March 2014, and the significance for Europe. *Euro. Commun. Dis. Bullet.* **2014**, 19, 1-11.
52. Chikungunya virus in the United States, 2015. http://www.cdc.gov/chikungunya/geo/united-states.html.
53. Tsetsarkin, K. A.; Chen, R.; Sherman, M. B.; Weaver, S. C., Chikungunya virus: evolution and genetic determinants of emergence. *Curr. Opin. Virol.* **2011**, 1, 310-317.
54. Tsetsarkin, K. A.; Vanlandingham, D. L.; McGee, C. E.; Higgs, S., A single mutation in chikungunya virus affects vector specificity and epidemic potential. *Plos Pathog.* **2007**, 3, 1895-1906.
55. Kumar, N. P.; Joseph, R.; Kamaraj, T.; Jambulingam, P., A226V mutation in virus during the 2007 chikungunya outbreak in Kerala, India. *J. Gen. Virol.* **2008**, 89, 1945-1948.
56. Santhosh, S. R.; Dash, P. K.; Parida, M. M.; Khan, M.; Tiwari, M.; Lakshmana Rao, P. V., Comparative full genome analysis revealed E1: A226V shift in 2007 Indian Chikungunya virus isolates. *Virus Res.* **2008**, 135, 36-41.
57. Tsetsarkin, K. A.; McGee, C. E.; Volk, S. M.; Vanlandingham, D. L.; Weaver, S. C.; Higgs, S., Epistatic roles of E2 glycoprotein mutations in adaption of Chikungunya virus to Aedes albopictus and Ae. Aegypti mosquitoes. *Plos One* **2009**, 4, e6835: 1-12.
58. Jose, J.; Snyder, J. E.; Kuhn, R. J., A structural and functional perspective of alphavirus replication and assembly. *Future Microbiol.* **2009**, 4, 837-856.
59. Chandak, N. H.; Kashyap, R. S.; Kabra, D.; Karandikar, P.; Saha, S. S.; Morey, S. H.; Purohit, H. J.; Taori, G. M.; Daginawala, H. F., Neurological complications of Chikungunya virus infection. *Neurol. India* **2009**, 57, 177-180.
60. Mahendradas, P.; Avadhani, K.; Shetty, R., Chikungunya and the eye: a review. *J. Ophthalmic Inflamm. Infect.* **2013**, 3, 1-9.

61. Mohan, A.; Kiran, D. H. N.; Manohar, I. C.; Kumar, D. P., Epidemiology, clinical manifestations, and diagnosis of Chikungunya fever: lessons learned from the re-emerging epidemic. *Indian J. Derma.* **2010**, 55, 54-63.
62. Edwards, C. J.; Welch, S. R.; Chamberlain, J.; Hewson, R.; Tolley, H.; Cane, P. A.; Lloyd, G., Molecular diagnosis and analysis of Chikungunya virus. *J. Clin. Virol.* **2007**, 39, 271-275.
63. Litzba, N.; Schuffenecker, I.; Zeller, H.; Drosten, C.; Emmerich, P.; Charrel, R.; Kreher, P.; Niedrig, M., Evaluation of the first commercial chikungunya virus indirect immunofluorescence test. *J. Virol. Methods* **2008**, 149, 175-179.
64. Cho, B.; Jeon, B.-Y.; Kim, J.; Noh, J.; Kim, J.; Park, M.; Park, S., Expression and Evaluation of Chikungunya Virus E1 and E2 Envelope Proteins for Serodiagnosis of Chikungunya Virus Infection. *Yonsei Med. J.* **2008**, 49, 828-835.
65. Sun, S.; Xiang, Y.; Akahata, W.; Holdaway, H.; Pal, P.; Zhang, X.; Diamond, M. S.; Nabel, G. J.; Rossmann, M. G., Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *eLife* **2013**, 2, 1-27.
66. Simizu, B.; Yamamoto, K.; Hashimoto, K.; Ogata, T., Structural proteins of Chikungunya virus. *J. Virol.* **1984**, 51, 254-258.
67. Khan, A. H.; Morita, K.; Parquet, M. d. C.; Hasebe, F.; Mathenge, E. G. M.; Igarashi, A., Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. *J. Gen. Virol.* **2002**, 83, 3075-3084.
68. Gould, E. A.; Delogu, I.; Forrester, N.; Khasnatinov, M.; Gritsun, T.; de Lamballerie, X.; Canard, B.; Coutard, B.; Malet, H.; Morin, B.; Jamal, S.; Weaver, S.; Gorbalenya, A.; Moureau, G.; Baronti, C., Understanding the alphaviruses: Recent research on important emerging pathogens and progress towards their control. *Antivir. Res.* **2010**, 87, 111-124.
69. Wang, Y. F.; Sawicki, S. G.; Sawicki, D. L., Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. *J. Virol.* **1994**, 68, 6466-6475.
70. Rungrotmongkol, T.; Nunthaboot, N.; Malaisree, M.; Kaiyawet, N.; Yotmanee, P.; Meeprasert, A.; Hannongbua, S., Molecular insight into the specific binding of ADP-ribose to the nsP3 macro domains of chikungunya and venezuelan equine encephalitis viruses: Molecular dynamics simulations and free energy calculations. *J. Mol. Graph. Model.* **2010**, 29, 347-353.
71. Laakkonen, P.; Hyvönen, M.; Peränen, J.; Kääriäinen, L., Expression of Semliki Forest virus nsP1-specific methyltransferase in insect cells and in Escherichia coli. *J. Virol.* **1994**, 68, 7418-7425.
72. Gould, E. A.; Delogu, I.; Forrester, N.; Khasnatinov, M.; Gritsun, T.; de Lamballerie, X.; Canard, B.; Coutard, B.; Malet, H.; Morin, B.; Jamal, S.; Weaver, S.; Gorbalenya, A.; Moureau, G.; Baronti, C., Understanding the alphaviruses: Recent research on important emerging pathogens and progress towards their control. *Antivir. Res.* **2010**, 87, 111-124.
73. Malet, H.; Lafitte, D.; Ferron, F.; Lescar, J.; Gorbalenya, A. E.; de Lamballerie, X.; Canard, B.; Coutard, B.; Jamal, S.; Dutartre, H.; Papageorgiou, N.; Neuvonen, M.; Ahola, T.; Forrester, N.; Gould, E. A., The crystal structures of Chikungunya and Venezuelan equine encephalitis virus

nsP3 macro domains define a conserved adenosine binding pocket. *J. Virol.* **2009**, 83, 6534-6545.
74. Ahola, T.; Lampio, A.; Auvinen, P.; Kääriäinen, L., Semliki Forest virus mRNA capping enzyme requires association with anionic membrane phospholipids for activity. *The EMBO J.* **1999**, 18, 3164-3172.
75. Peränen, J.; Laakkonen, P.; Hyvönen, M.; Kääriäinen, L., The Alphavirus Replicase Protein nsP1 Is Membrane-Associated and Has Affinity to Endocytic Organelles. *Virology* **1995**, 208, 610-620.
76. Zhang, D.; Tözsér, J.; Waugh, D. S., Molecular cloning, overproduction, purification and biochemical characterization of the p39 nsp2 protease domains encoded by three alphaviruses. *Protein Expres. Purif.* **2009**, 64, 89-97.
77. Rikkonen, M.; Peränen, J.; Kääriäinen, L., ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2. *J. Virol.* **1994**, 68, 5804-5810.
78. Aleksei, L.; Valeria, L.; Kairit, T.; Tero, A.; Andres, M., Molecular Determinants of Substrate Specificity for Semliki Forest Virus Nonstructural Protease. *J. Virol.* **2006**, 80, 5413-5422.
79. Gomez de Cedrón, M.; Ehsani, N.; Mikkola, M. L.; García, J. A.; Kääriäinen, L., RNA helicase activity of Semliki Forest virus replicase protein nsP2. *FEBS Lett.* **1999**, 448, 19-22.
80. Karpe, Y. A.; Aher, P. P.; Lole, K. S., NTPase and 5'-RNA triphosphatase activities of Chikungunya virus nsP2 protein. *Plos One* **2011**, 6, e22336:1-8.
81. Suopanki, J.; Sawicki, D. L.; Sawicki, S. G.; Kaariainen, L., Regulation of alphavirus 26S mRNA transcription by replicase component nsP2. *J. Gen. Virol.* **1998**, 79, 309-319.
82. Dorothea, L. S.; Silvia, P.; John, M. P.; Stanley, G. S., Role for nsP2 Proteins in the Cessation of Alphavirus Minus-Strand Synthesis by Host Cells. *J. Virol.* **2006**, 80, 360-371.
83. Peränen, J.; Rikkonen, M.; Liljeström, P.; Kääriäinen, L., Nuclear localization of Semliki Forest virus-specific nonstructural protein nsP2. *J. Virol.* **1990**, 64, 1888-1896.
84. LaStarza, M. W.; Lemm, J. A.; Rice, C. M., Genetic analysis of the nsP3 region of Sindbis virus: evidence for roles in minus-strand and subgenomic RNA synthesis. *J. Virol.* **1994**, 68, 5781-5791.
85. Helena, V.; Tero, A.; Minna, T.; Andres, M.; Leevi, K. Ã. r. i., Elimination of Phosphorylation Sites of Semliki Forest Virus Replicase Protein nsP3. *J. Biol. Chem.* **2001**, 276, 5745-5752.
86. Park, E.; Griffin, D. E., The nsP3 macro domain is important for Sindbis virus replication in neurons and neurovirulence in mice. *Virology* **2009**, 388, 305-314.
87. Dé, I.; Fata-Hartley, C.; Sawicki, S. G.; Sawicki, D. L., Functional analysis of nsP3 phosphoprotein mutants of Sindbis virus. *J. Virol.* **2003**, 77, 13106-13116.
88. Rubach, J. K.; Wasik, B. R.; Rupp, J. C.; Kuhn, R. J.; Hardy, R. W.; Smith, J. L., Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. *Virology* **2009**, 384, 201-208.
89. Shirako, Y.; Strauss, J. H., Regulation of Sindbis virus RNA replication: uncleaved P123 and nsP4 function in minus-strand RNA synthesis, whereas

cleaved products from P123 are required for efficient plus-strand RNA synthesis. *J. Virol.* **1994**, 68, 1874-1885.
90. Jonathan, C. R.; Natasha, J.; Richard, W. H., Requirement for the Amino-Terminal Domain of Sindbis Virus nsP4 during Virus Infection. *J. Virol.* **2011**, 85, 3449-3460.
91. Shirako, Y.; Strauss, J. H.; Strauss, E. G., Suppressor Mutations That Allow Sindbis Virus RNA Polymerase to Function with Nonaromatic Amino Acids at the N-Terminus: Evidence for Interaction between nsP1 and nsP4 in Minus-Strand RNA Synthesis. *Virology* **2000**, 276, 148-160.
92. Fata, C. L.; Sawicki, S. G.; Sawicki, D. L., Alphavirus minus-strand RNA synthesis: Identification of a role for Arg183 of the nsP4 polymerase. *J. Virol.* **2002**, 76, 8632-8640.
93. Voss, J. E.; Vaney, M.-C.; Duquerroy, S.; Vonrhein, C.; Girard-Blanc, C.; Crublet, E.; Thompson, A.; Bricogne, G.; Rey, F. A., Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* **2010**, 468, 709-712.
94. Mulvey, M.; Brown, D. T., Involvement of the molecular chaperone BiP in maturation of Sindbis virus envelope glycoproteins. *J. Virol.* **1995**, 69, 1621-1627.
95. Carleton, M.; Lee, H.; Mulvey, M.; Brown, D. T., Role of glycoprotein PE2 in formation and maturation of the Sindbis virus spike. *J. Virol.* **1997**, 71, 1558-1566.
96. Li, L.; Jose, J.; Xiang, Y.; Kuhn, R. J., Structural changes of envelope proteins during alphavirus fusion. *Nature* **2010**, 468, 705-708.
97. Inglot, A. D., Comparison of the Antiviral Activity in vitro of some Non-steroidal Anti-inflammatory Drugs. *J. Gen. Virol.* **1969**, 4, 203-214.
98. Maheshwari, R. K.; Srikantan, V.; Bhartiya, D., Chloroquine enhances replication of Semliki Forest virus and encephalomyocarditis virus in mice. *J. Virol.* **1991**, 65, 992-995.
99. Delogu, I.; de Lamballerie, X., Chikungunya disease and chloroquine treatment. *J. Med. Virol.* **2011**, 83, 1058-1059.
100. Brighton, S. W., Chloroquine phosphate treatment of chronic Chikungunya arthritis. An open pilot study. *South Afr. Med. J.* **1984**, 66, 217-218.
101. De Lamballerie, X.; Boisson, V.; Reynier, J.-C.; Enault, S.; Charrel, R. N.; Flahault, A.; Roques, P.; Le Grand, R., On chikungunya acute infection and chloroquine treatment. *Vector Borne Zoonotic Dis.* **2008**, 8, 837-840.
102. Coombs, K.; Mann, E.; Edwards, J.; Brown, D. T., Effects of chloroquine and cytochalasin B on the infection of cells by Sindbis virus and vesicular stomatitis virus. *J. Virol.* **1981**, 37, 1060-1065.
103. Pieter, L.; Erik De, C.; Johan, N., The Anti-Yellow Fever Virus Activity of Ribavirin Is Independent of Error-Prone Replication. *Mol. Pharmacol.* **2006**, 69, 1461-1467.
104. Ravichandran, R.; Manian, M., Ribavirin therapy for Chikungunya arthritis. *J. Infect. Dev. Ctries.* **2008**, 2, 140-142.
105. Briolant, S.; Garin, D.; Scaramozzino, N.; Jouan, A.; Crance, J. M., In vitro inhibition of Chikungunya and Semliki Forest viruses replication by antiviral compounds: synergistic effect of interferon-alpha and ribavirin combination. *Antivir. Res.* **2004**, 61, 111-117.

106. Boriskin, Y.; Leneva, I.; Pecheur, E. I.; Polyak, S., Arbidol: A Broad-Spectrum Antiviral Compound that Blocks Viral Fusion. *Curr. Med. Chem.* **2008**, 15, 997-1005.
107. Villalaín, J., Membranotropic effects of arbidol, a broad anti-viral molecule, on phospholipid model membranes. *J. Phys. Chem. B* **2010**, 114, 8544-8554.
108. Pécheur, E.-I.; Lavillette, D.; Alcaras, F.; Molle, J.; Boriskin, Y. S.; Roberts, M.; Cosset, F.-L.; Polyak, S. J., Biochemical mechanism of hepatitis C virus inhibition by the broad-spectrum antiviral arbidol. *Biochem.* **2007**, 46, 6050-6059.
109. Leneva, I. A.; Russell, R. J.; Boriskin, Y. S.; Hay, A. J., Characteristics of arbidol-resistant mutants of influenza virus: Implications for the mechanism of anti-influenza action of arbidol. *Antivir. Res.* **2009**, 81, 132-140.
110. Elodie, T.; François, P.; Eve-Isabelle, P.; Giorgia, Z.; Antoine, L.; Dimitri, L.; Jean-Pierre, L.; Roland, M.; François-Loïc, C.; Anja, B.; Beat, H. M., Mechanism of Inhibition of Enveloped Virus Membrane Fusion by the Antiviral Drug Arbidol: e15874. *Plos One* **2011**, 6, 1-13.
111. Delogu, I.; Pastorino, B.; Baronti, C.; Nougairède, A.; Bonnet, E.; de Lamballerie, X., In vitro antiviral activity of arbidol against Chikungunya virus and characteristics of a selected resistant mutant. *Antivir. Res.* **2011**, 90, 99-107.
112. Bentley, R., Mycophenolic Acid: A One Hundred Year Odyssey from Antibiotic to Immunosuppressant. *Chem. Rev.* **2000**, 100, 3801-3826.
113. Sweeney, M. J.; Hoffman, D. H.; Esterman, M. A., Metabolism and biochemistry of mycophenolic acid. *Cancer Res.* **1972**, 32, 1803-1809.
114. Khan, M.; Dhanwani, R.; Patro, I. K.; Rao, P. V. L.; Parida, M. M., Cellular IMPDH enzyme activity is a potential target for the inhibition of Chikungunya virus replication and virus induced apoptosis in cultured mammalian cells. *Antivir. Res.* **2011**, 89, 1-8.
115. Allard, P.-M.; Martin, M.-T.; Dau, M.-E. T. H.; Leyssen, P.; Guéritte, F.; Litaudon, M., Trigocherrin A, the first natural chlorinated daphnane diterpene orthoester from Trigonostemon cherrieri. *Org. Lett.* **2012**, 14, 342-345.
116. Bourjot, M.; Leyssen, P.; Neyts, J.; Dumontet, V.; Litaudon, M., Trigocherrierin A, a potent inhibitor of chikungunya virus replication. *Molecules* **2014**, 19, 3617-3627.
117. Parveen, K.; Meerra, T.; Regina Ching Hua, L.; Huixin, C.; Karen Caiyun, C.; Mah Lee, N.; Justin Jang Hann, C., Inhibition of Chikungunya Virus Replication by Harringtonine, a Novel Antiviral That Suppresses Viral Protein Expression. *Antimicro. Agents Chemother.* **2013**, 57, 155-167.
118. Rada, B.; Dragúň, M., Antiviral action and selectivity of 6-Azauridine. *Ann. N. Y. Acad. Sci.* **1977**, 284, 410-417.
119. Bourjot, M.; Delang, L.; Nguyen, V. H.; Neyts, J.; Guéritte, F.; Leyssen, P.; Litaudon, M., Prostratin and 12-O-tetradecanoylphorbol 13-acetate are potent and selective inhibitors of Chikungunya virus replication. *J. Nat. Prod.* **2012**, 75, 2183-2187.
120. Bourjot, M.; Leyssen, P.; Eydoux, C.; Guillemot, J.-C.; Canard, B.; Rasoanaivo, P.; Guéritte, F.; Litaudon, M., Chemical constituents of Anacolosa pervilleana and their antiviral activities. *Fitoterapia* **2012**, 83, 1076-1080.
121. Li, Y.-G.; Takeda, N.; Anantapreecha, S.; Siripanyaphinyo, U.; Tumkosit, U.; Noranate, N.; A-Nuegoonpipat, A.; Pan, Y.; Kameoka, M.; Kurosu, T.; Ikuta,

K., Poly (I:C), an agonist of toll-like receptor-3, inhibits replication of the Chikungunya virus in BEAS-2B cells. *J. Virol.* **2012**, 9, 114-114.
122. Singh, P.; Chhabra, M.; Mittal, V.; Sharma, P.; Rizvi, M. A.; Chauhan, L. S.; Rai, A., Current research and clinical trials for a vaccine against Chikungunya virus. *Vaccine: Dev. Ther.* **2013**, 3, 35-46.
123. Harrison, V. R.; Eckels, K. H.; Bartelloni, P. J.; Hampton, C., Production and Evaluation of a Formalin-Killed Chikungunya Vaccine. *J. Immunol.* **1971**, 107, 643-647.
124. Kenneth, H. E.; Venton, R. H.; Frank, M. H., Chikungunya Virus Vaccine Prepared by Tween-Ether Extraction. *Appl. Microbiol.* **1970**, 19, 321-325.
125. Tiwari, M.; Parida, M.; Santhosh, S. R.; Khan, M.; Dash, P. K.; Rao, P. V. L., Assessment of immunogenic potential of Vero adapted formalin inactivated vaccine derived from novel ECSA genotype of Chikungunya virus. *Vaccine* **2009**, 27, 2513-2522.
126. Levitt, N. H.; Ramsburg, H. H.; Hasty, S. E.; Repik, P. M.; Cole, F. E.; Lupton, H. W., Development of an attenuated strain of chikungunya virus for use in vaccine production. *Vaccine* **1986**, 4, 157-162.
127. Edelman, R.; Tacket, C. O.; Wasserman, S. S.; Bodison, S. A.; Perry, J. G.; Mangiafico, J. A., Phase II safety and immunogenicity study of live chikungunya virus vaccine TSI-GSD-218. *Am. J. Trop. Med. Hyg.* **2000**, 62, 681-685.
128. Wang, E.; Volkova, E.; Adams, A. P.; Forrester, N.; Xiao, S.-Y.; Frolov, I.; Weaver, S. C., Chimeric alphavirus vaccine candidates for chikungunya. *Vaccine* **2008**, 26, 5030-5039.
129. Wang, D.; Einfeld, D.; Dong, J. Y.; Suhrbier, A.; Penn-Nicholson, A.; Woraratanadharm, J.; Gardner, J.; Luo, M.; Le, T. T.; Anraku, I.; Sakalian, M., A complex adenovirus vaccine against chikungunya virus provides complete protection against viraemia and arthritis. *Vaccine* **2011**, 29, 2803-2809.
130. Plante, K.; Stinchcomb, D. T.; Osorio, J. E.; Frolov, I.; Weaver, S. C.; Wang, E.; Partidos, C. D.; Weger, J.; Gorchakov, R.; Tsetsarkin, K.; Borland, E. M.; Powers, A. M.; Seymour, R., Novel chikungunya vaccine candidate with an IRES-based attenuation and host range alteration mechanism. *Plos Pathog.* **2011**, 7, e1002142: 1-11.
131. Partidos, C. D.; Paykel, J.; Weger, J.; Borland, E. M.; Powers, A. M.; Seymour, R.; Weaver, S. C.; Stinchcomb, D. T.; Osorio, J. E., Cross-protective immunity against o'nyong-nyong virus afforded by a novel recombinant chikungunya vaccine. *Vaccine* **2012**, 30, 4638-4643.
132. Muthumani, K.; Kim, J. J.; Vijayachari, P.; Weiner, D. B.; Lankaraman, K. M.; Laddy, D. J.; Sundaram, S. G.; Chung, C. W.; Sako, E.; Wu, L.; Khan, A.; Sardesai, N., Immunogenicity of novel consensus-based DNA vaccines against Chikungunya virus. *Vaccine* **2008**, 26, 5128-5134.
133. Mallilankaraman, K.; Sundaram, S. G.; Muruganandam, N.; Sarangan, G.; Srikanth, P.; Khan, A. S.; Lewis, M. G.; Kim, J. J.; Sardesai, N. Y.; Muthumani, K.; Weiner, D. B.; Shedlock, D. J.; Bao, H.; Kawalekar, O. U.; Fagone, P.; Ramanathan, A. A.; Ferraro, B.; Stabenow, J.; Vijayachari, P., A DNA vaccine against chikungunya virus is protective in mice and induces neutralizing antibodies in mice and nonhuman primates. *Plos Neglect. Trop. Dis.* **2011**, 5, e928:1-13.

134. Metz, S. W.; Pijlman, G. P., Arbovirus vaccines; opportunities for the baculovirus-insect cell expression system. *J. Invert. Pathol.* **2011**, 107, S16-S30.
135. Anade, P.; Martina, B. E.; Geertsema, C.; Oers, v. M. M.; Vlak, J. M.; Pijlman, G. P.; Heldens, J.; Metz, S. W. H.; Goldbach, R. W., Functional processing and secretion of Chikungunya virus E1 and E2 glycoproteins in insect cells. *J. Virol.* **2011**, 8, 353-353.
136. Kumar, M.; Sudeep, A. B.; Arankalle, V. A., Evaluation of recombinant E2 protein-based and whole-virus inactivated candidate vaccines against chikungunya virus. *Vaccine* **2012**, 30, 6142-6149.
137. Khan, M.; Dhanwani, R.; Rao, P. V. L.; Parida, M., Subunit vaccine formulations based on recombinant envelope proteins of Chikungunya virus elicit balanced Th1/Th2 response and virus-neutralizing antibodies in mice. *Virus Res.* **2012**, 167, 236-246.
138. Akahata, W.; Rao, S.; Nabel, G. J.; Yang, Z.-Y.; Andersen, H.; Sun, S.; Holdaway, H. A.; Kong, W.-P.; Lewis, M. G.; Higgs, S.; Rossmann, M. G., A virus-like particle vaccine for epidemic Chikungunya virus protects nonhuman primates against infection. *Nat. Med.* **2010**, 16, 334-338.
139. Metz, S. W.; Gardner, J.; Geertsema, C.; Le, T. T.; Goh, L.; Vlak, J. M.; Suhrbier, A.; Pijlman, G. P., Effective chikungunya virus-like particle vaccine produced in insect cells. *Plos Negl. Trop. Dis.* **2013**, 7, e2124: 1-11.
140. Ozden, S.; Mouly, V.; Desprès, P.; Gessain, A.; Butler-Browne, G.; Chrétien, M.; Tangy, F.; Vidalain, P.-O.; Seidah, N. G.; Lucas-Hourani, M.; Ceccaldi, P.-E.; Basak, A.; Valentine, M.; Benjannet, S.; Hamelin, J.; Jacob, Y.; Mamchaoui, K., Inhibition of Chikungunya virus infection in cultured human muscle cells by furin inhibitors: impairment of the maturation of the E2 surface glycoprotein. *J. Biol. Chem.* **2008**, 283, 21899-21908.
141. Dash, P. K.; Tiwari, M.; Santhosh, S. R.; Parida, M.; Lakshmana Rao, P. V., RNA interference mediated inhibition of Chikungunya virus replication in mammalian cells. *Biochem. Biophys. Res. Commun.* **2008**, 376, 718-722.
142. Bréhin, A.-C.; Casadémont, I.; Frenkiel, M.-P.; Julier, C.; Sakuntabhai, A.; Desprès, P., The large form of human 2′,5′-Oligoadenylate Synthetase (OAS3) exerts antiviral effect against Chikungunya virus. *Virology* **2009**, 384, 216-222.
143. Henrik Gad, H.; Paulous, S.; Belarbi, E.; Diancourt, L.; Drosten, C.; Kümmerer, B. M.; Plate, A. E.; Caro, V.; Desprès, P., The E2-E166K substitution restores Chikungunya virus growth in OAS3 expressing cells by acting on viral entry. *Virology* **2012**, 434, 27-37.
144. Teng, T.-S.; Leo, Y.-S.; Merits, A.; Chin, K.-C.; Ng, L. F. P.; Foo, S.-S.; Simamarta, D.; Lum, F.-M.; Teo, T.-H.; Lulla, A.; Yeo, N. K. W.; Koh, E. G. L.; Chow, A., Viperin restricts chikungunya virus replication and pathology. *J. Clinic. Invest.* **2012**, 122, 4447-4460.
145. Rathore, A. P. S.; Haystead, T.; Das, P. K.; Merits, A.; Ng, M.-L.; Vasudevan, S. G., Chikungunya virus nsP3 and nsP4 interacts with Hsp-90 to promote virus replication: Hsp-90 inhibitors reduce CHIKV infectin and inflammation in vivo. *Antivir. Res.* **2014**, 103, 7-16.
146. Pohjala, L.; Utt, A.; Varjak, M.; Lulla, A.; Merits, A.; Ahola, T.; Tammela, P., Inhibitors of alphavirus entry and replication identified with a stable Chikungunya replicon cell line and virus-based assays. *Plos One* **2011**, 6, e28923:1-13.

147. Cruz, D. J. M.; Freitas-Junior, L. H.; Bonotto, R. M.; Gomes, R. G. B.; da Silva, C. T.; Taniguchi, J. B.; No, J. H.; Lombardot, B.; Schwartz, O.; Hansen, M. A. E., Identification of novel compounds inhibiting chikungunya virus-induced cell death by high throughput screening of a kinase inhibitor library. *Plos Negl. Trop. Dis.* **2013**, 7, e2471:1-15.
148. Lucas-Hourani, M.; Munier-Lehmann, H.; Lupan, A.; Desprès, P.; Thoret, S.; Pamlard, O.; Dubois, J.; Guillou, C.; Tangy, F.; Vidalain, P.-O., A phenotypic assay to identify Chikungunya virus inhibitors targeting the nonstructural protein nsP2. *J. Biomol. Screen.* **2013**, 18, 172-179.
149. Hubbard, R. E., *Structure-based drug discovery: An overview*. The Royal Society of Chemistry: 2006.
150. Mandal, S.; Mandal, S. K.; Moudgil, M. n., Rational drug design. *Eur. J. Pharmacol.* **2009**, 625, 90-100.
151. Bhakat, S.; Karubiu, W.; Jayaprakash, V.; Soliman, M. E. S., A perspective on targeting non-structural proteins to combat neglected tropical diseases: Dengue, West Nle and Chikungunya viruses. *Eur. J. Med. Chem.* **2014**, 87, 677-702.
152. Kumar, S.; Kapopara, G.; Patni, I.; Pandya, A.; Jasrai, T.; Patel, K., Exploring the polymerase activity of chikungunya viral non structural protein 4 (nsP4) using molecular modeling, e-pharmacophore and docking studies. *Int. J. Pharm. Life. Sci.* **2012**, 3, 1752-1765.
153. Voss, J.; Vaney, M. C.; Duquerroy, S.; Vonrhein, C.; Ginard-Blanc, C.; Crublet, E.; Thompson, A.; Bricogne, G.; Rey, F. A., Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* **2010**, 468, 709-712.
154. Kapetanovic, I. M., Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chem. Biol. Interact.* **2008**, 171, 165-176.
155. Giovanna, S., Structural Biology and Drug Discovery. *Curr. Pharm. Des.* **2006**, 12, 2087-2097.
156. Clark, D. E., What has computer-aided molecular design ever done for drug discovery? *Expert Opin. Drug Discov.* **2006**, 1, 103-110.
157. Kore, P. P.; Mutha, M. M.; Antre, R. V.; Oswal, R. J.; Kshirsagar, S. S., Computer-aided drug design: An innovative tool for modeling. *J. Med. Chem.* **2012**, 2, 139-148.
158. Klebe, G., Recent developments in structure-based drug design. *J. Mol. Med.* **2000**, 78, 269-281.
159. Kubinyl, H., Success stories of Computer-Aided Design. In *Comput. Appl. Pharm. Res. Dev.*, John Wiley & Sons: 2006.
160. Claudio, C.; Andrew, W. O., Ligand Docking and Structure-based Virtual Screening in Drug Discovery. *Curr. Top. Med. Chem.* **2007**, 7, 1006-1014.
161. Tanaji, T.; Santosh, K.; Alan, R., Successful Applications of Computer Aided Drug Discovery: Moving Drugs from Concept to the Clinic. *Curr. Top. Med. Chem.* **2010**, 10, 127-141.
162. Huang, S.-Y.; Zou, X., Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* **2010**, 11, 3016-3034.
163. Kirchmair, J.; Distinto, S.; Liedl, K. R.; Markt, P.; Rollinger, J. M.; Schuster, D.; Spitzer, G. M.; Wolber, G., Development of anti-viral agents using molecular modeling and virtual screening techniques. *Infect. Disord. Drug Targets* **2011**, 11, 64-93.

164. Aparoy, P.; Reddy, K. K.; Reddanna, P., Structure and ligand based drug design strategies in the development of novel 5- LOX inhibitors. *Curr. Med. Chem.* **2012**, 19, 3763-3778.
165. Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R. P.; Foloppe, N., Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *J. Comp. Struct. Biotech.* **2013**, 5, e201302011:1-14.
166. Grinter, S. Z.; Zou, X., Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules* **2014**, 19, 10150-10176.
167. Yuriev, E.; Agostino, M.; Ramsland, P. A., Challenges and advances in computational docking: 2009 in review. *J. Mol. Recognit.* **2011**, 24, 149-164.
168. Van Drie, J. H., Computer-aided drug design: the next 20 years. *J. Comput.-Aided Mol. Des.* **2007**, 21, 591-601.
169. Rafael, V. C. G.; Glaucius, O.; Adriano, D. A., Structure- and ligand-based drug design approaches for neglected tropical diseases. *Pure Appl. Chem.* **2012**, 84, 1857-1866.
170. Tang, Y.; Zhu, W.; Chen, K.; Jiang, H., New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. *Drug Discov. Today: Tech.* **2006**, 3, 307-313.
171. Veselovsky, A. V.; Ivanov, A. S., Strategy of computer-aided drug design. *Curr. Drug Targets Infect. Dis.* **2003**, 3, 33-40.
172. Kuntz, I. D.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T., A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, 161, 269-288.
173. Sousa, S. F.; Fernandes, P. A.; Ramos, M. J., Protein-ligand docking: current status and future challenges. *Proteins* **2006**, 65, 15-26.
174. Gschwend, D. A.; Good, A. C.; Kuntz, I. D., Molecular docking towards drug discovery. *J. Mol. Recognit.* **1996**, 9, 175-186.
175. Shoichet, B. K.; McGovern, S. L.; Wei, B.; Irwin, J. J., Lead discovery using molecular docking. *Curr. Opin. Chem. Biol.* **2002**, 6, 439-446.
176. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004**, 3, 935-949.
177. Taylor, R. D.; Jewsbury, P. J.; Essex, J. W., A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, 16, 151-166.
178. Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R., Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Struct., Funct., and Genet.* **2002**, 47, 409-443.
179. Brooijmans, N.; Kuntz, I. D., Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, 32, 335-373.
180. Rester, U., Dock around the Clock – Current Status of Small Molecule Docking and Scoring. *QSAR Comb. Sci.* **2006**, 25, 605-615.
181. Venkatraman, M.; Alan, G.; Maxwell, C.; Edward, J.; Renee, D., Docking: Successes and Challenges. *Curr. Pharm. Des.* **2005**, 11, 323-333.
182. Coupez, B.; Lewis, R. A., Docking and scoring-Theoretically easy, practically impossible? *Curr. Med. Chem.* **2006**, 13, 2995-3003.
183. Dias, R.; Walter, F. d. A. J., Molecular docking algorithms. *Curr. Drug Targets* **2008**, 9, 1040-1047.

184. Novikov, F. N.; Chilov, G. G., Molecular docking: theoretical background, practical applications and perspective. *Mendeleev. Commun.* **2009**, 19, 237-242.
185. Yuriev, E.; Ramsland, P. A., Latest developments in molecular docking: 2010-2011 in review. *J. Mol. Recognit.* **2013**, 26, 215-239.
186. Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E., Receptor–ligand molecular docking. *Biophys. Rev.* **2014**, 6, 75-87.
187. Waszkowycz, B.; Clark, D. E.; Gancia, E., Outstanding challenges in protein-ligand docking and structure-based virtual screening. *Adv. Rev.* **2011**, 1, 229-259.
188. Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D., DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, 15, 411-428.
189. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, 19, 1639-1662.
190. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G., A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J. Mol. Biol.* **1996**, 261, 470-489.
191. Jones, G.; Willett, P.; Glen, R. C., Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, 245, 43-53.
192. Friesner, R. A.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, 47, 1739-1749.
193. B-Rao, C.; Subramanian, J.; Sharma, S. D., Managing protein flexibility in docking and its applications. *Drug Discov. Today* **2009**, 14, 394-400.
194. Kokh, D. B.; Wade, R. C.; Wenzel, W., Receptor flexibility in small-molecule docking calculations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, 1, 298-314.
195. Lill, M. A., Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochem.* **2011**, 50, 6157-6169.
196. Flick, J.; Tristram, F.; Wenzel, W., Modeling loop backbone flexibility in receptor-ligand docking simulations. *J. Comput. Chem.* **2012**, 33, 2504-2515.
197. Asses, Y.; Venkatraman, V.; Leroux, V.; Ritchie, D. W.; Maigret, B., Exploring c-Met kinase flexibility by sampling and clustering its conformational space. *Proteins: Struct. Funct. Bioinf.* **2012**, 80, 1227-1238.
198. Huang, S.-Y.; Grinter, S. Z.; Zou, X., Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* **2010**, 12, 12899-12908.
199. Jain, A. N., Scoring functions for protein-ligand docking. *Curr. Protein Pept. Sci.* **2006**, 7, 407-420.
200. Kramer, B.; Rarey, M.; Lengauer, T., Evaluation of the FlexX incremental construction algorithm for protein- ligand docking. *Proteins: Struct. Funct. Gen.* **1999**, 37, 228-241.

201. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, 25, 1157-1174.
202. Vanommeslaeghe, K.; Vorobyov, I.; Mackerell Jr, A. D.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, 31, 671-690.
203. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, 31, 455-461.
204. McGann, M., FRED pose prediction and virtual screening accuracy. *J. Chem. Inform. Model.* **2011**, 51, 578-596.
205. Abagyan, R.; Totrov, M.; Kuznetsov, D., ICM - A new method for protein modeling and design -Applications to docking and structure prediction from the distorted native conformations. *J. Comput. Chem.* **1994**, 15, 488-506.
206. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, 46, 3-26.
207. Walters, W. P.; Murcko, M. A., Prediction of "drug-likeness". *Adv. Drug Deliv. Rev.* **2002**, 54, 255-271.
208. Simon, L.; Sebastian, S.; Martin, Z., In Silico Prediction of Binding Sites on Proteins. *Curr. Med. Chem.* **2010**, 17, 1550-1562.
209. Henrich, S.; Salo-Ahen, O. M. H.; Huang, B.; Rippmann, F. F.; Cruciani, G.; Wade, R. C., Computational approaches to identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.* **2009**, 1-11.
210. Laurie, A. T. R.; Jackson, R. M., Methods for the Prediction of Protein-Ligand Binding Sites for Structure-Based Drug Design and Virtual Ligand Screening. *Curr. Protein Pept. Sci.* **2006**, 7, 395-406.
211. Tuccinardi, T., Docking-based virtual screening: recent developments. *Comb. Chem. High T. Scr.* **2009**, 12, 303-314.
212. Barril, X.; Hubbard, R. E.; Morley, S. D., Virtual screening in structure-based drug discovery. *Mini Rev. Med. Chem.* **2004**, 4, 779-791.
213. Mclnnes, C., Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **2007**, 11, 494-502.
214. Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N., Virtual screening in drug discovery-A computational perspective. *Curr. Protein Pept. Sci.* **2007**, 8, 329-351.
215. Sousa, S. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J., Virtual screening in drug design and development. *Comb. Chem. High T. Scr.* **2010**, 13, 442-453.
216. Vyas, V.; Jain, A.; Jain, A.; Gupta, A., Virtual screening: A fast tool for drug design. *Sci. Pharm.* **2008**, 76, 333-360.
217. Tingjun, H.; Xiaojie, X., Recent Development and Application of Virtual Screening in Drug Discovery: An Overview. *Curr. Pharm. Des.* **2004**, 10, 1011-1033.
218. Muegge, I.; Oloff, S., Advances in virtual screening. *Drug Discov. Today Tech.* **2006**, 3, 405-411.
219. Klebe, G., Virtual ligands screening: strategies, perspective and limitations. *Drug Discov. Today* **2006**, 11, 580-594.

220. Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T. M.; Murray, C. W.; Taylor, R. D.; Watson, P., Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, 44, 793-806.
221. Cavasotto, C. N.; Orry, A. J. W., Ligand docking and structure-based virtual screening in drug discovery. *Curr. Top. Med. Chem.* **2007**, 7, 1006-1014.
222. Breda, A.; Basso, L. A.; Santos, D. S.; F., d. A. J. W., Virtual screening of drugs: score functions, docking, and drug design. *Curr. Comput.-Aided Drug Des.* **2008**, 4, 265-272.
223. Guido, R. V. C.; Oliva, G.; Andricopulo, A. D., Virtual screening and its integration with modern drug design technologies. *Curr. Med. Chem.* **2008**, 15, 37-46.
224. Badrinarayan, P.; Sastry, G. N., Virtual high throughput screening in new lead identification. *Comb. Chem. High T. Scr.* **2011**, 14, 840-860.
225. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R., Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, 267, 727-727.
226. Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J., Virtual screening with AutoDock: theory and practice. *Expert Opin. Drug Discov.* **2010**, 5, 597-607.
227. Chang, M. W.; Ayeni, C.; Breuer, S.; Torbett, B. E., Virtual screening for HIV protease inhibitors: A comparison of AutoDock 4 and Vina. *Plos One* **2010**, 5, 1-9.
228. Irwin, J. J.; Shoichet, B. K., ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inform. Model.* **2005**, 45, 177-182.
229. Dunkel, M.; Fullbeck, M.; Neumann, S.; Preissner, R., SuperNatural: a searchable database of available natural compounds. *Nucleic Acids Res.* **2006**, 34, D678-D683.
230. Wishart, D. S.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.-A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Tzur, D.; Amegbey, G.; Block, D.; Hau, D. D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G. E.; Knox, C.; MacInnis, G. D.; Weljie, A. M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B. D.; Vogel, H. J.; Eisner, R.; Querengesser, L.; Guo, A. C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D., HMDB: The human metabolome database. *Nucleic Acids Res.* **2007**, 35, D521-D526.
231. Wishart, D. S.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Knox, C.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Guo, A. C.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E., HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, 37, D603-D610.
232. Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M., DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008**, 36, D901-D906.
233. Graves, A. P.; Brenk, R.; Shoichet, B. K., Decoys for docking. *J. Med. Chem.* **2005**, 48, 3714-3728.

234. Wallach, I.; Lilien, R., Virtual decoy sets for molecular docking benchmarks. *J. Chem. Inform. Modell.* **2011**, 51, 196-202.
235. Huang, N.; Shoichet, B. K.; Irwin, J. J., Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, 49, 6789-6801.
236. Reynolds, C. H., Protein-ligand cocrystal structures: we can do better. *ACS Med. Chem. Lett.* **2014**, 5, 727-729.
237. Kolb, P.; Irwin, J. J., Docking screens: right for the right reasons? *Curr. Top. Med. Chem.* **2009**, 9, 755-770.
238. Baber, J. C. B.; Thompson, D. C.; Cross, J. B.; Humblet, C., GARD: A generally applicable replacement for RMSD. *J. Chem. Inform. Model.* **2009**, 49, 1889-1900.
239. Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C., Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J. Chem. Inform. Model.* **2009**, 49, 1455-1474.
240. Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R., Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inform. Model.* **2009**, 49, 1079-1093.
241. Karplus, M. P. G. A., Molecular dynamics simulations in biology. *Nature* **1990**, 347, 631-639.
242. McCammon, J. A.; Gelin, B. R.; Karplus, M., Dynamics of folded proteins. *Nature* **1977**, 267, 585-590.
243. Rognan, D., Molecular dynamics simulations: A tool for drug design. *Perspect. Drug Discov. Des.* **1998**, 9-11, 181-209.
244. Salsbury, F. R., Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Curr. Opin. Pharmacol.* **2010**, 10, 738-744.
245. Ode, H.; Nakashima, M.; Kitamura, S.; Sugiura, W.; Sato, H., Molecular dynamics simulation in virus research. *Front. Microbiol.* **2012**, 3, 1-9.
246. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, 9, 71-71.
247. Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E., Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Opin. Struct. Biol.* **2009**, 19, 120-127.
248. Tiana, G.; Simona, F.; De Mori, G. M. S.; Broglia, R. A.; Colombo, G., Understanding the determinants of stability and folding of small globular proteins from their energetics. *Protein Sci.* **2004**, 13, 113-124.
249. Daggett, V., Molecular dynamics simulations of the protein unfolding/folding reaction. *Acc. Chem. Res.* **2002**, 35, 422-429.
250. Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A., Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, 30, 211-243.
251. Allen, T. W.; Andersen, O. S.; Roux, B., Energetics of Ion Conduction through the Gramicidin Channel. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 117-122.
252. Karplus, M.; Kuriyan, J., Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102, 6679-6685.

253. Warshel, A., COMPUTER SIMULATIONS OF ENZYME CATALYSIS: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, 32, 425-443.
254. Karplus, M.; McCammon, J. A., Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, 9, 646-652.
255. Brunger, A. T. A. P. D., Molecular dynamics applied to x-ray structure refinement. *Acc. Chem. Res.* **2002**, 35, 404-412.
256. Dodson, G. G.; Lane, D. P.; Verma, C. S., Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Rep.* **2008**, 9, 144-150.
257. MacKerell, A. D.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Bashford, D.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M.; Bellott; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, 102, 3586-3616.
258. Adcock, S. A.; McCammon, J. A., Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* **2006**, 106, 1589-1615.
259. Ponder, J.; Case, D., Force fields for protein simulations. *Adv. Protein Chem.* **2003**, 66, 27-85.
260. Oostenbrink, C.; Villa, A.; Mark, A.; Van Gunsteren, W. F., A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, 25, 1656-1676.
261. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, 118, 11225-11236.
262. Case, D. A.; Woods, R. J.; Cheatham, r. T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, J. K. M.; Onufriev, A.; Simmerling, C.; Wang, B., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, 26, 1668-1688.
263. Brooks, B. R.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Brooks, r. C. L.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Mackerell, J. A. D.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C., CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, 30, 1545-1614.
264. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4, Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, 4, 435-447.
265. Phillips, J. C.; Schulten, K.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, 26, 1781-1802.
266. Wan, S.; Coveney, P. V.; Flower, D. R., Molecular dynamics simulations: Bring biomolecular structures alive on a computer. *Methods Mol. Biol.* **2007**, 409, 321-339.
267. Hansson, T.; Oostenbrink, C.; van Gunsteren, W. F., Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2002**, 12, 190-196.

268. Van Gunsteren, W. F.; Mark, A. E., Validation of molecular dynamics simulation. *J. Chem. Phys.* **1998**, 108, 6109-6116.
269. Ulrich, E.; Lalith, P.; Max, L. B.; Tom, D.; Hsing, L.; Lee, G. P., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, 103, 8577-8593.
270. Harvey, M. J.; De Fabritiis, G., An implementation of the smooth Particle Mesh Ewald method on GPU hardware. *J. Chem. Theory Comput.* **2009**, 5, 2371-2377.
271. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, 79, 926-935.
272. Neria, E.; Fischer, S.; Karplus, M., Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, 105, 1902-1921.
273. Zheng, J.; Ma, B.; Tsai, C.-J.; Nussinov, R., Structural Stability and Dynamics of an Amyloid-Forming Peptide GNNQQNY from the Yeast Prion Sup-35. *J. Biophys.* **2006**, 91, 824-833.
274. Kuzmanic, A.; Zagrovic, B., Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors. *J. Biophys.* **2010**, 98, 861-871.
275. Tatsis, V. A.; Tsoulos, I. G.; Stavrakoudis, A., Molecular Dynamics Simulations of the TSSPSAD Peptide Antigen in Free and Bound with CAMPATH-1H Fab Antibody States: The Importance of the [beta]-Turn Conformation. *Int. J. Pept. Res. Ther.* **2009**, 15, 1-9.
276. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graph. Model.* **1996**, 14, 33-38.
277. Bruno, A.; Guadix, A. E.; Costantino, G., Molecular dynamics simulation of the heterodimeric mGluR2/5HT(2A) complex. An atomistic resolution study of a potential new target in psychiatric conditions. *J. Chem. Inform. Model.* **2009**, 49, 1602-1616.
278. Rajgaria, R.; McAllister, S. R.; Floudas, C. A., Towards accurate residue-residue hydrophobic contact prediction for alpha helical proteins via integer linear optimization. *Proteins* **2009**, 74, 929-947.
279. Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham Iii, T. E., Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* **2007**, 3, 2312-2334.
280. Alonso, H.; Bliznyuk, A. A.; Gready, J. E., Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, 26, 531-568.
281. Okimoto, N.; Futatsugi, N.; Fuji, H.; Suenaga, A.; Morimoto, G.; Yanai, R.; Ohno, Y.; Narumi, T.; Taiji, M., High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. *Plos Comput. Biol.* **2009**, 5, e1000528:1-13.
282. Riccardo, B., *Computational Drug Discovery and Design, Methods in Molecular Biology*. Springer Science & Business Media, LLC 2012: 2012; Vol. 819.
283. Parenti, M. D.; Rastelli, G., Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnol. Adv.* **2012**, 30, 244-250.
284. Merz, K. M.; Reynolds, C. H.; Ringe, D., *Drug design: structure- and ligand-based approaches*. Cambridge University Press: Cambridge, 2010.

285. Brandsdal, B. O.; Osterberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Aqvist, J., Free energy calculations and ligand binding. *Adv. Protein Chem.* **2003**, 66, 123-158.
286. Michel, J.; Essex, J. W., Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput.-Aided Mol. Des.* **2010**, 24, 639-658.
287. Hansen, N.; Van Gunsteren, W. F., Practical aspects of free-energy calculations: A review. *J. Chem. Theory Comput.* **2014**, 10, 2632-2647.
288. de Ruiter, A.; Oostenbrink, C., Free energy calculations of protein–ligand interactions. *Curr. Opin. Chem. Biol.* **2011**, 15, 547-552.
289. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M., Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, 107, 9535-9551.
290. Wang, J.; Deng, Y.; Roux, B., Absolute binding free energies calculations using molecular dynamics simulations with restraining potentials. *J. Biophys.* **2006**, 91, 2798-2814.
291. Rodinger, T.; Pomes, R., Enhancing the accuracy, the efficiency and the scope of free energy simulations. *Curr. Opin. Struct. Biol.* **2005**, 15, 164-170.
292. Gilson, M. K.; Zhou, H.-X., Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, 36, 21-42.
293. Beveridge, D. L.; DiCapua, F. M., Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, 18, 431-492.
294. Straatsma, T. P.; McCammon, J. A., Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, 43, 407-435.
295. Aqvist, J.; Luzhkov, V. B.; Brandsdal, B. O., Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002**, 35, 358-365.
296. Guo, Z.; Brooks, C. L.; Kong, X., Efficient and Flexible Algorithm for Free Energy Calculations Using the λ-Dynamics Approach. *J. Phys. Chem. B* **1998**, 102, 2032-2036.
297. Simonson, T.; Archontis, G.; Karplus, M., Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* **2002**, 35, 430-437.
298. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, 11, 425-445.
299. Foloppe, N.; Hubbard, R., Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* **2006**, 13, 3583-3608.
300. Michel, J.; Foloppe, N.; Essex, J. W., Rigorous free energy calculations in structure-based drug design. *Mol. Inf.* **2010**, 29, 570-578.
301. Zwanzig, R. W., High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, 22, 1420-1426.
302. Lu, N.; Kofke, D. A., Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling. *J. Chem. Phys.* **2001**, 114, 7303-7311.
303. Lu, N.; Kofke, D. A., Accuracy of free-energy perturbation calculations in molecular simulation. II. Heuristics. *J. Chem. Phys.* **2001**, 115, 6866-6875.
304. van Gunsteren, W. F.; Berendsen, H. J. C., Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry. *J. Comput.-Aided. Mol. Des.* **1987**, 1, 171-176.

305. Lawrenz, M.; Wereszczynski, J.; Ortiz-Sánchez, J. M.; Nichols, S. E.; McCammon, J. A., Thermodynamic integration to predict host-guest binding affinities. *J. Comput.-Aided Mol. Des.* **2012**, 26, 569-576.
306. Kollman, P. A.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham Iii, T. E.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y., Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, 33, 889-897.
307. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inform. Model.* **2011**, 51, 69-82.
308. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem.* **2011**, 32, 866-877.
309. Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* **2014**, 16, 16719-16729.
310. Wang, W.; Kollman, P. A., Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 14937-14942.
311. Wang, W.; Kollman, P. A., Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model1. *J. Mol. Biol.* **2000**, 303, 567-582.
312. Kuhn, B.; Kollman, P. A., Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* **2000**, 43, 3786-3791.
313. Huo, S.; Wang, J.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D., Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: Insight into structure-based ligand design. *J. Med. Chem.* **2002**, 45, 1412-1419.
314. Gohlke, H.; Kiel, C.; Case, D. A., Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J. Mol. Biol.* **2003**, 330, 891-913.
315. Aqvist, J.; Medina, C.; Samuelsson, J. E., A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, 7, 385-391.
316. Aqvist, J.; Marelius, J., The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High T. Scr.* **2001**, 4, 613-626.
317. Aqvist, J., Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **1996**, 17, 1587-1597.
318. Hansson, T.; Aqvist, J., Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein Eng.* **1995**, 8, 1137-1144.

319. Jones-Hertzog, D. K.; Jorgensen, W. L., Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *J. Med. Chem.* **1997**, 40, 1539-1549.
320. Hansson, T.; Marelius, J.; Aqvist, J., Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, 12, 27-35.
321. Wang, W.; Wang, J.; Kollman, P. A., What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins* **1999**, 34, 395-402.
322. Pierce, A. C.; Jorgensen, W. L., Estimation of binding affinities for selective thrombin inhibitors via Monte Carlo simulations. *J. Med. Chem.* **2001**, 44, 1043-1050.
323. Ljungberg, K. B.; Marelius, J.; Musil, D.; Svensson, P.; Nordem, B.; Aqvist, J., Computational modelling of inhibitor binding to human thrombin. *Eur. J. Pharm. Sci.* **2001**, 12, 441-446.
324. Almlöf, M.; Brandsdal, B. O.; Åqvist, J., Binding affinity prediction with different force fields: Examination of the linear interaction energy method. *J. Comput. Chem.* **2004**, 25, 1242-1254.
325. Tounge, B. A.; Rajamani, R.; Baxter, E. W.; Reitz, A. B.; Reynolds, C. H., Linear interaction energy models for beta-secretase (BACE) inhibitors: Role of van der Waals, electrostatic, and continuum-solvation terms. *J. Mol. Graph. Model.* **2006**, 24, 475-484.
326. Almlöf, M.; Carlsson, J.; Åqvist, J., Improving the Accuracy of the Linear Interaction Energy Method for Solvation Free Energies. *J. Chem. Theory Comput.* **2007**, 3, 2162-2175.
327. de Amorim, H. L. N.; Caceres, R. A.; Netz, P. A., Linear interaction energy (LIE) method in lead discovery and optimization. *Curr. Drug Targets* **2008**, 9, 1100-1105.
328. Guitierrez-de-Teran, H.; Aqvist, J., Linear Interaction Energy: Method and Applications in Drug Design. In *Comput. Drug Dis. Des.*, Springer New York: 2012; Vol. 819, pp 305-323.
329. Nicolotti, O.; Convertino, M.; Leonetti, F.; Catto, M.; Cellamare, S.; A., C., Estimation of the binding free energy by linear interaction energy models. *Mini Rev. Med. Chem.* **2012**, 12, 551-561.
330. Khan, Y. S.; Gutierrez-de-Teran, H.; Boukharta, L.; Aqvist, J., Toward an optimal docking and free energy calculations scheme in ligand design with application to COX-1 inhibitors. *J. Chem. Inform. Model.* **2014**, 54, 1488-1499.
331. Ben-Naim, A.; Marcus, Y., Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **1984**, 81, 2016-2027.
332. Ben-Naim, A.; Mazo, R. M., Size dependence of the solvation free energies of large solutes. *J. Phys. Chem.* **1993**, 97, 10829-10834.
333. Verkhivker, G.; Appelt, K.; Freer, S. T.; Villafranca, J. E., Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng.* **1995**, 8, 677-691.

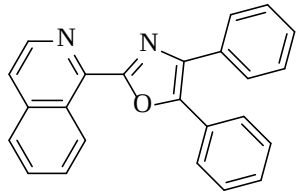
334. Wallqvist, A.; Jernigan, R. L.; Covell, D. G., A preference-based free-energy parameterization of enzyme-inhibitor binding. Applications to HIV-1-protease inhibitor design. *Protein Sci.* **1995**, 4, 1881-1903.
335. Li, G.; La Starza, M. W.; Reef Hardy, W.; Strauss, J. H.; Rice, C. M., Phosphorylation of sindbis virus nsP3 in vivo and in vitro. *Virology* **1990**, 179, 416-427.
336. Lastarza, M. W.; Grakoui, A.; Rice, C. M., Deletion and Duplication Mutations in the C-Terminal Nonconserved Region of Sindbis Virus nsP3: Effects on Phosphorylation and on Virus Replication in Vertebrate and Invertebrate Cells. *Virology* **1994**, 202, 224-232.
337. Helena, V.; Juhani, S., Phosphorylation Site Analysis of Semliki Forest Virus Nonstructural Protein 3. *J. Biol. Chem.* **2000**, 275, 27775-27783.
338. Neuvonen, M.; Ahola, T., Differential Activities of Cellular and Viral Macro Domain Proteins in Binding of ADP-Ribose Metabolites. *J. Mol. Biol.* **2009**, 385, 212-225.
339. Jelke, J. F.; Natalia, E. D.; Jim, B.; Corinne, G.; Jacky, F.; Just, M. V.; Gorben, P. P., Chikungunya Virus nsP3 Blocks Stress Granule Assembly by Recruitment of G3BP into Cytoplasmic Foci. *J. Virol.* **2012**, 86, 10873-10879.
340. Rodion, G.; Natalia, G.; Elena, F.; Ilya, F., Different Types of nsP3-Containing Protein Complexes in Sindbis Virus-Infected Cells. *J. Virol.* **2008**, 82, 10088-10101.
341. Margus, V.; Eva, Ž.; Andres, M., Novel Functions of the Alphavirus Nonstructural Protein nsP3 C-Terminal Region. *J. Virol.* **2010**, 84, 2352-2364.
342. Accelrys Software Inc *Discovery Studio Modeling Environment*, 4.0; Accelrys Software Inc.: San Diego, 2013.
343. Huang, B., MetaPocket: a meta approach to improve protein ligand binding site prediction. *OMICS J. Integrat. Biol.* **2009**, 13, 325-330.
344. Trott, O.; Olson, A. J., Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, 31, 455-461.
345. Pastorino, B. A. M.; Peyrefitte, C. N.; Almeras, L.; Grandadam, M.; Rolland, D.; Tolou, H. J.; Bessaud, M., Expression and biochemical characterization of nsP2 cysteine protease of Chikungunya virus. *Virus Res.* **2008**, 131, 293-298.
346. Strauss, E. G.; De Groot, R. J.; Levinson, R.; Strauss, J. H., Identification of the active site residues in the nsP2 proteinase of Sindbis virus. *Virology* **1992**, 191, 932-940.
347. Jelke, J. F.; Alexander, A. K.; Gorben, P. P.; Wen Jun, L.; Natalie, A. P.; Corinne, G.; Maarten, L.; Dana, L. V.; Esther, S.; Just, M. V.; Andreas, S., Chikungunya Virus Nonstructural Protein 2 Inhibits Type I/II Interferon-Stimulated JAK-STAT Signaling. *J. Virol.* **2010**, 84, 10877-10887.
348. Lidia, V.; Andres, M.; Andrey, G.; Valeria, S.; Leevi, K. Ã. r. i.; Tero, A., Regulation of the Sequential Processing of Semliki Forest Virus Replicase Polyprotein. *J. Biol. Chem.* **2003**, 278, 41636-41645.
349. Kim, K. H.; Rümenapf, T.; Strauss, E. G.; Strauss, J. H., Regulation of Semliki Forest virus RNA replication: a model for the control of alphavirus pathogenesis in invertebrate hosts. *Virology* **2004**, 323, 153-163.

350. Russo, A. T.; Malmstrom, R. D.; White, M. A.; Watowich, S. J., Structural basis for substrate specificity of alphavirus nsP2 proteases. *J. Mol. Graph. Model.* **2010**, 29, 46-53.
351. Singh, K. D.; Kirubakaran, P.; Nagarajan, S.; Sakkiah, S.; Muthusamy, K.; Velmurgan, D.; Jeyakanthan, J., Homology modeling, molecular dynamics, e-pharmacophore mapping and docking study of Chikungunya virus nsP2 protease. *J. Mol. Model.* **2011**, 18, 39-51.
352. Bora, L., Homology modeling and docking to potential novel inhibitor for Chikungunya (37997) protein nsP2 protease. *J Proteomics Bioinform.* **2012**, 5, 054-059.
353. Bassetto, M.; Silvestri, R.; Tron, G. C.; Neyts, J.; Leyssen, P.; Brancale, A.; De Burghgraeve, T.; Delang, L.; Massarotti, A.; Coluccia, A.; Zonta, N.; Gatti, V.; Colombano, G.; Sorba, G., Computer-aided identification, design and synthesis of a novel series of compounds with selective antiviral activity against chikungunya virus. *Antivir. Res.* **2013**, 98, 12-18.
354. Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M., MUSTANG: a multiple structural alignment algorithm. *Proteins* **2006**, 64, 559-574.
355. Russo, A. T.; White, M. A.; Watowich, S. J., The Crystal Structure of the Venezuelan Equine Encephalitis Alphavirus nsP2 Protease. *Structure* **2006**, 14, 1449-1458.
356. Kielian, M.; Chanel-Vos, C.; Liao, M., Alphavirus entry and membrane fusion. *Viruses* **2010**, 2, 796-825.
357. Li, L.; Jose, J.; Xiang, Y.; Kuhn, R. J.; Rossmann, M. G., Structural changes of envelope proteins during alphavirus fusion. *Nature* **2010**, 468, 705-708.
358. Kuo, S.-C.; Chen, Y.-J.; Wang, Y.-M.; Tsui, P.-Y.; Kuo, M.-D.; Wu, T.-Y.; Lo, S. J., Cell-based analysis of Chikungunya virus E1 protein in membrane fusion. *J. BioMed. Sci.* **2012**, 19, 44-44.
359. Akahata, W.; Nabel, G. J., A specific domain of the Chikungunya virus E2 protein regulates particle formation in human cells: implications for alphavirus vaccine design. *J. Virol.* **2012**, 86, 8879-8883.
360. Asnet Mary, J.; Paramasivan, R.; Tyagi, B. K.; Surender, M.; Shenbagarathai, R., Identification of structural motifs in the E2 glycoprotein of Chikungunya involved in virus-host interaction. *J. Biomol. Struct. Dyn.* **2013**, 31, 1077-1085.
361. Nguyen, P. T. V.; Yu, H.; Keller, P. A., Discovery of in silico hits targeting the nsP3 macro domain of chikungunya virus. *J. Mol. Model.* **2014**, 20, 1-12.
362. Rashad, A. A.; Keller, P. A., Structure based design towards the identification of novel binding sites and inhibitors for the chikungunya virus envelope proteins. *J. Mol. Graph. Model.* **2013**, 44, 241-252.
363. Lu, P.; Hontecillas, R.; Horne, W. T.; Carbo, A.; Viladomiu, M.; Pedragosa, M.; Bevan, D. R.; Lewis, S. N.; Bassaganya-Riera, J., Computational modeling-based discovery of novel classes of anti-inflammatory drugs that target lanthionine synthetase C-like protein 2. *PloS One* **2012**, 7, e34643.

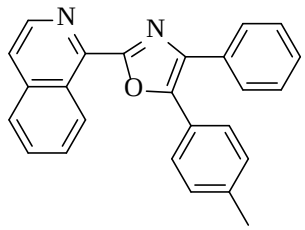
APPENDICES

Appendix 1. Chemical structures of top hit compounds in all dockings for the nsP3 macrodomain, nsP2 protease, and envelope glycoprotein complexes.

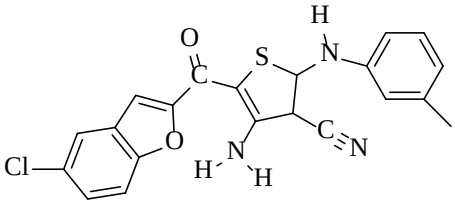
NCI_25457



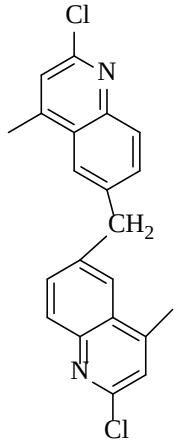
NCI_116702



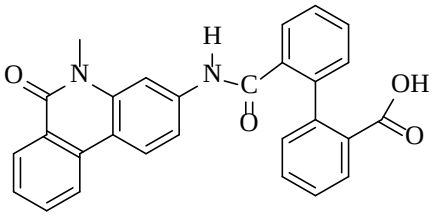
NCI_309892



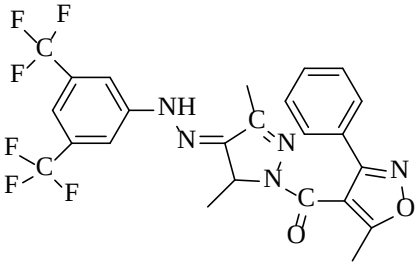
NCI_109451



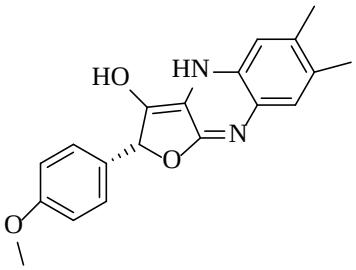
NCI_127133



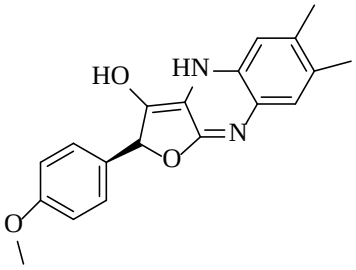
NCI_328101



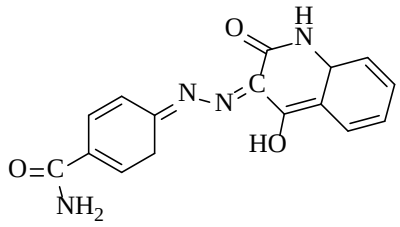
NCI_372275_a



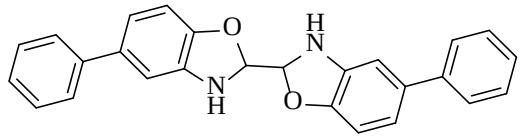
NCI_372275_b



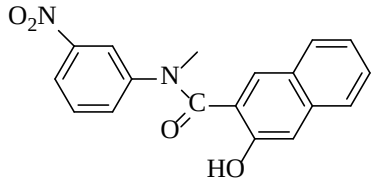
NCI_45545



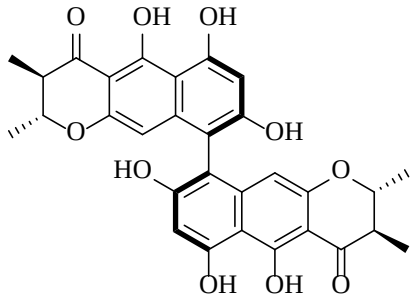
NCI_84100_b



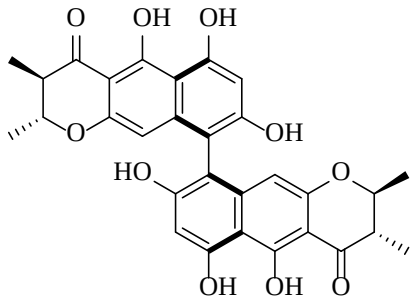
NCI_37168



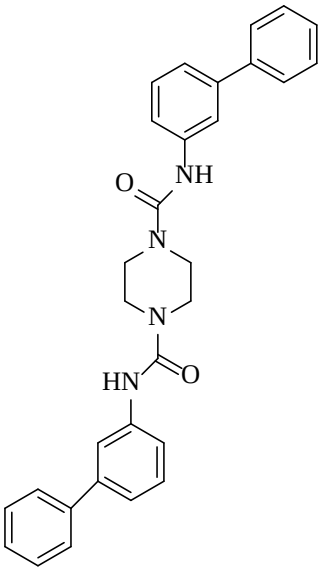
NCI_345647_a



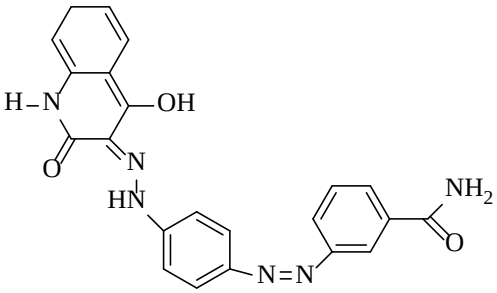
NCI_345647_b



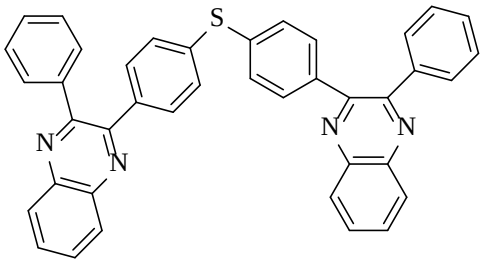
NCI_37553



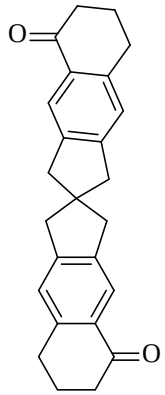
NCI_58052



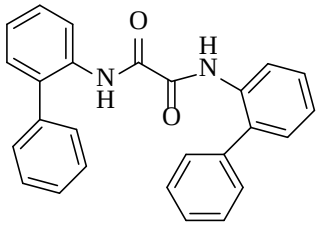
NCI_293778



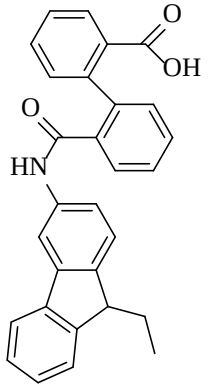
NCI_670283



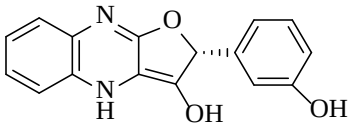
NCI_338042



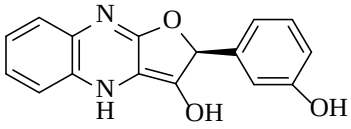
NCI_132232



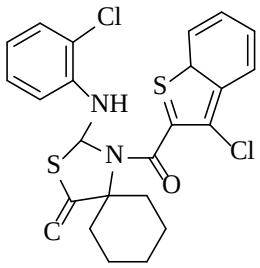
NCI_372499_a



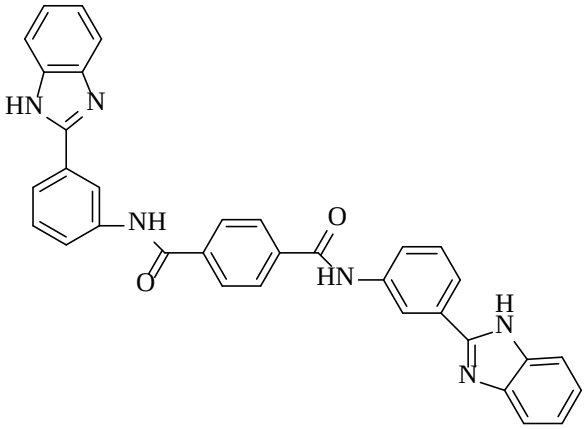
NCI_372499_b



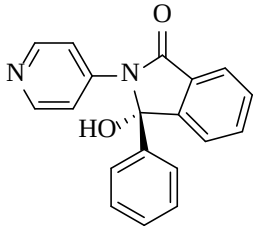
NCI_310326



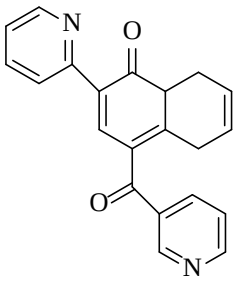
NCI_61610



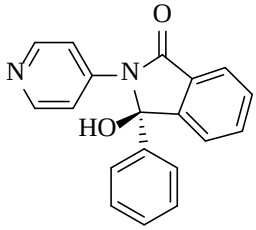
NCI_69359_a



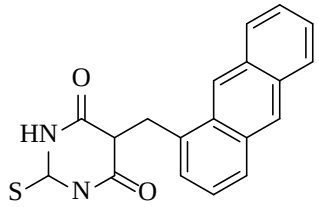
NCI_324623



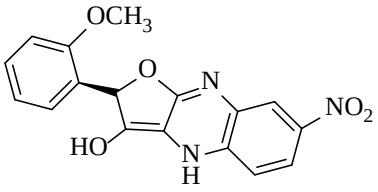
NCI_69359_b



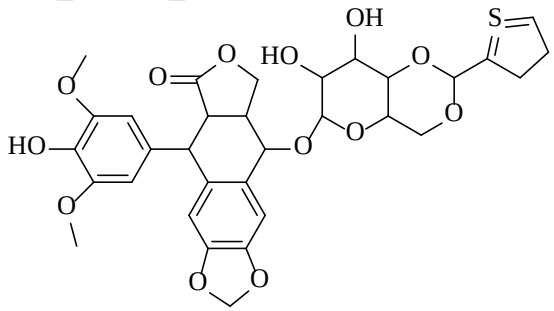
NCI_90737



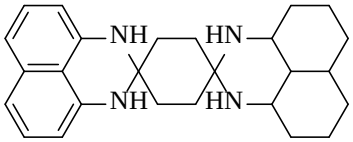
NCI_372287_b



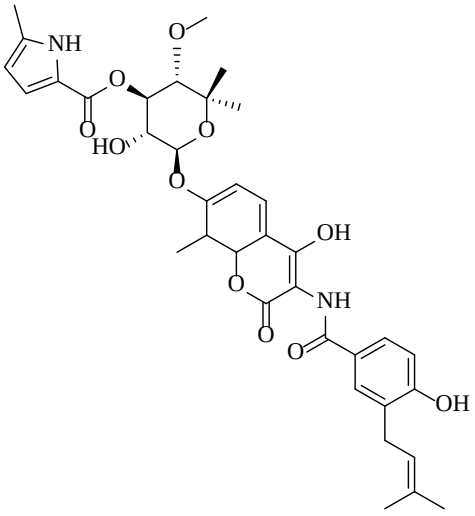
NCI_122819_a



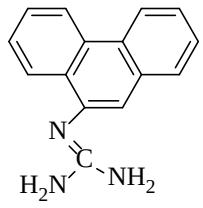
NCI_97920



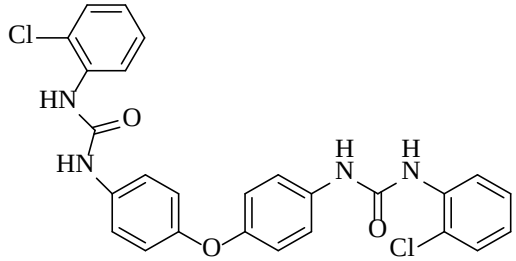
NCI_227186_a



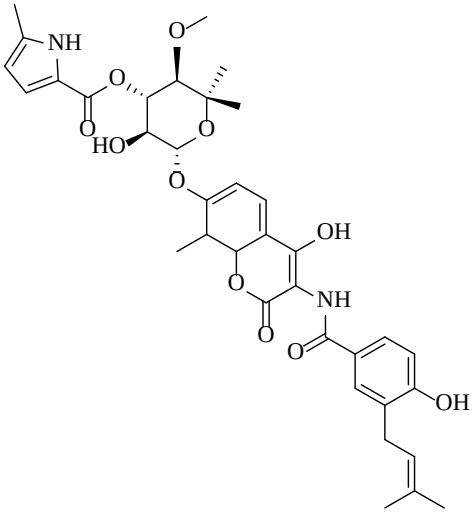
NCI_400976



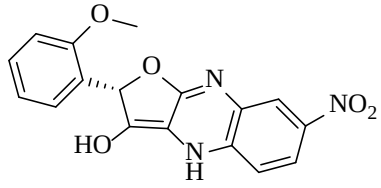
NCI_80731



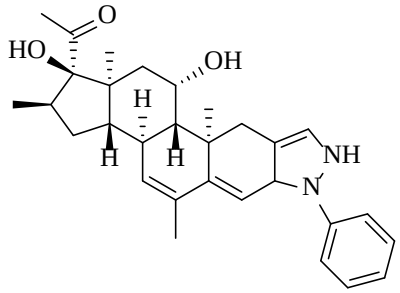
NCI_227186_b



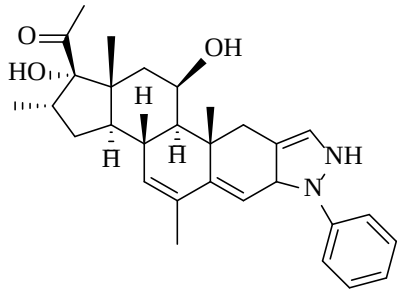
NCI_372287_a



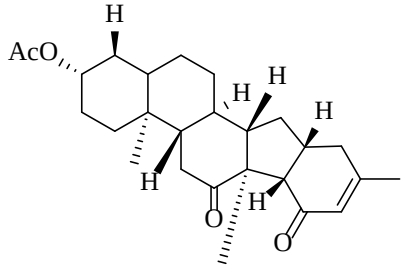
NCI_80997_a



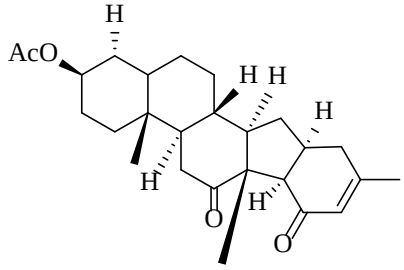
NCI_80997_b



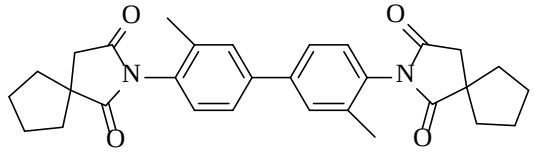
NCI_59620_a



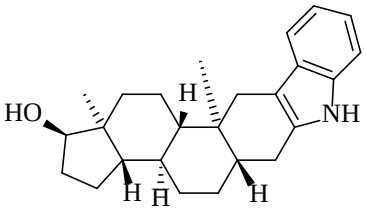
NCI_59620_b



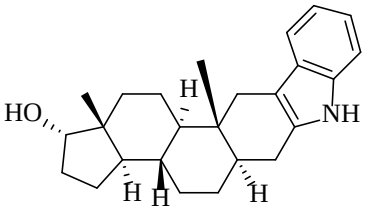
NCI_308835



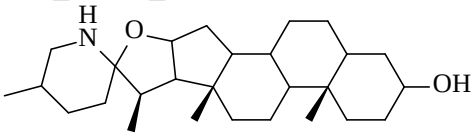
NCI_60785_a



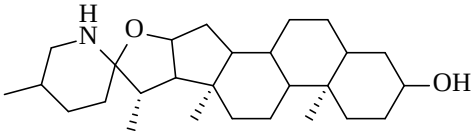
NCI_60785_b



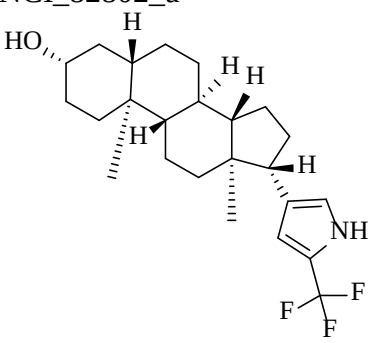
NCI_27592_a



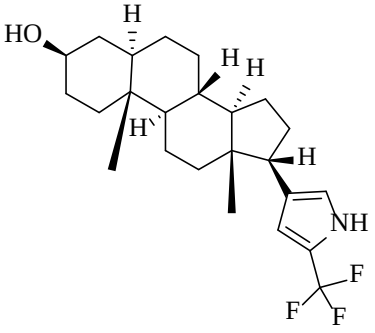
NCI_27592_b



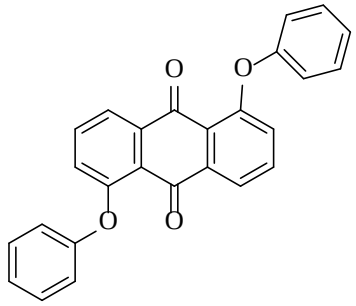
NCI_82802_a



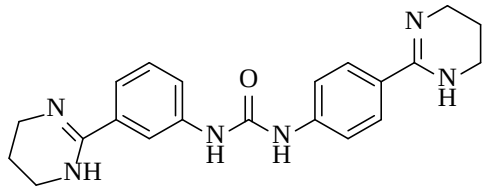
NCI_82802_b



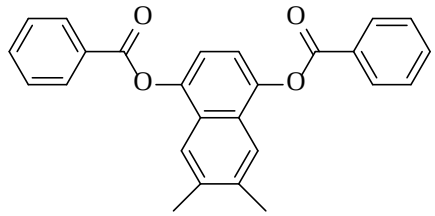
NCI_37627



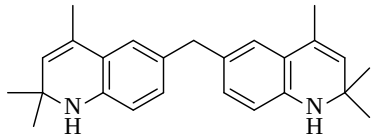
NCI_63680



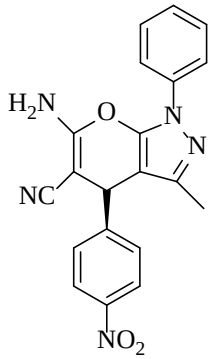
NCI_219894



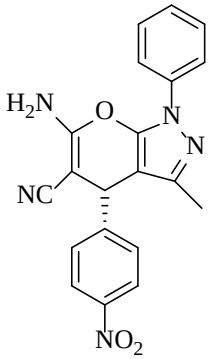
NCI_217697 ($C_{25}H_{30}N_2$)



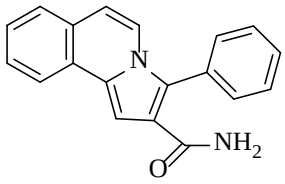
NCI_298892_a ($C_{20}H_{18}N_5O_3$)



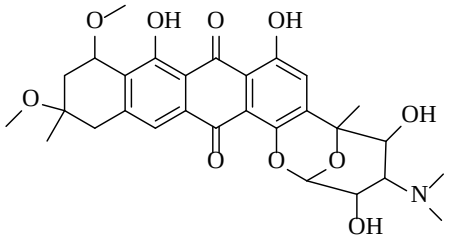
NCI_298892_b ($C_{20}H_{18}N_5O_3$)



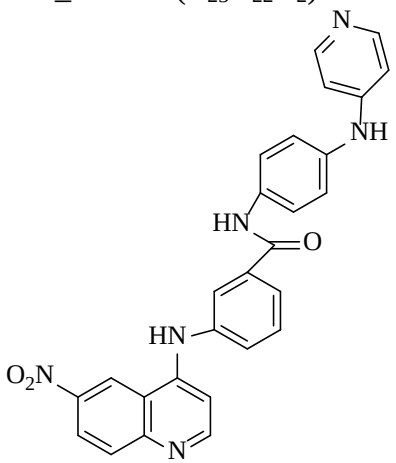
NCI_93427 ($C_{19}H_{14}N_2O$)



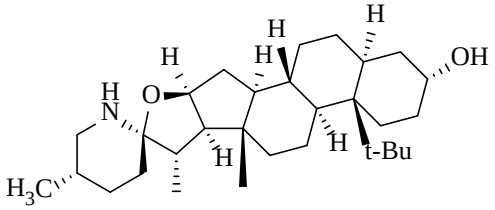
NCI_269148



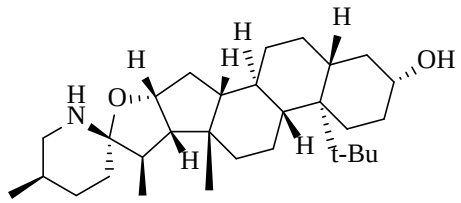
NCI_146771 ($C_{23}H_{22}N_2$)



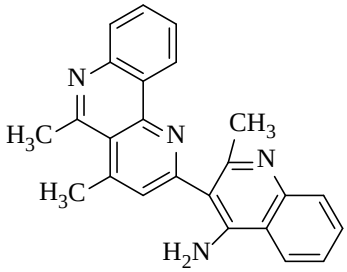
NCI_7524_a ($C_{19}H_{18}N_2O_3$)



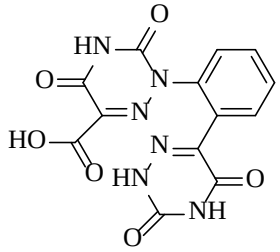
NCI_7524_b



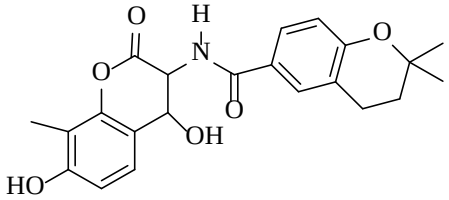
NCI_345845



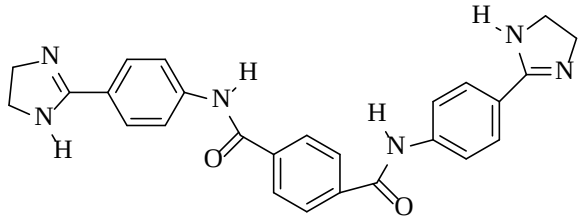
NCI_362639



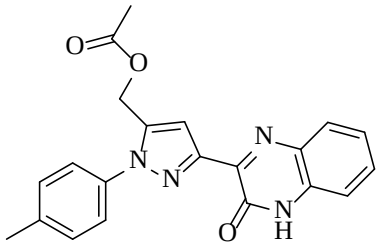
NCI_5157



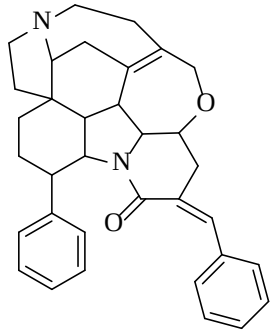
NCI_67436



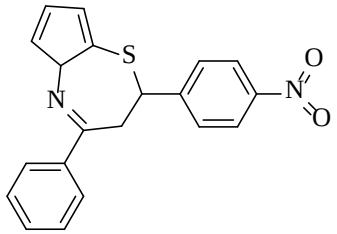
NCI_339161



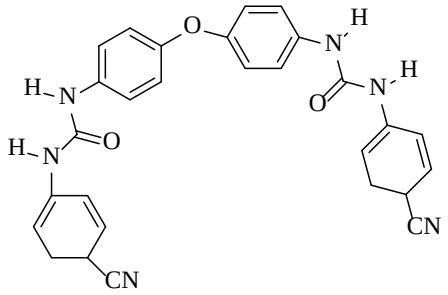
NCI_121868_a



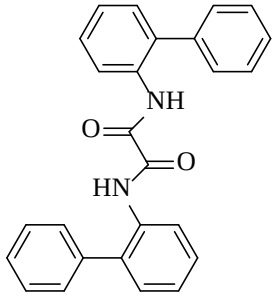
NCI_270738



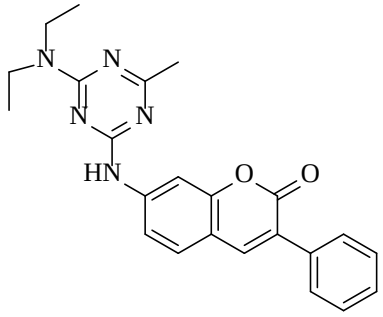
NCI_80734



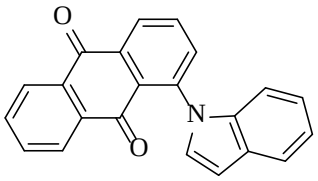
NCI_338042



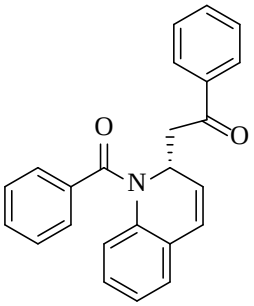
NCI_338963



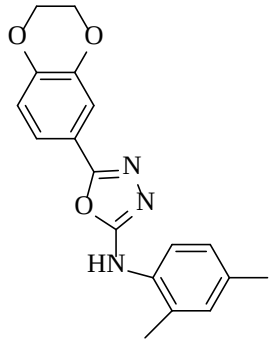
NCI_80313



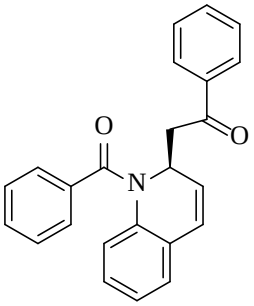
NCI_116709_a



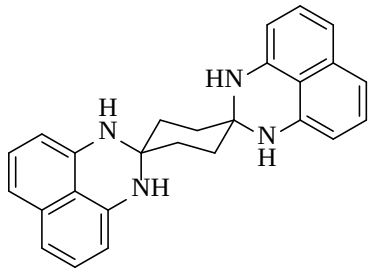
NCI_326182



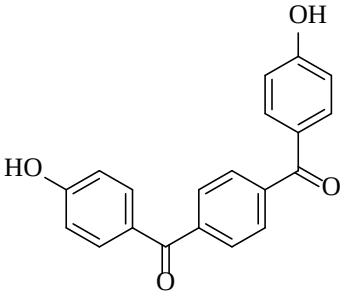
NCI_116709_b



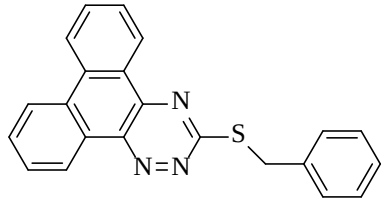
NCI_97920



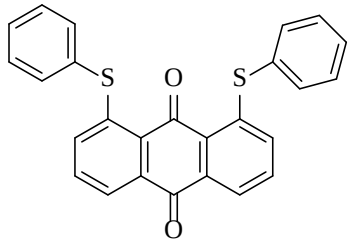
NCI_326757



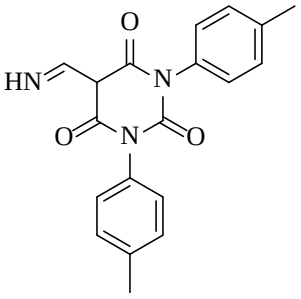
NCI_327702



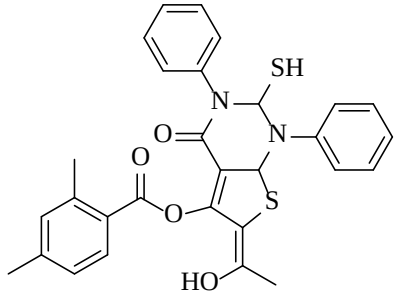
NCI_156516



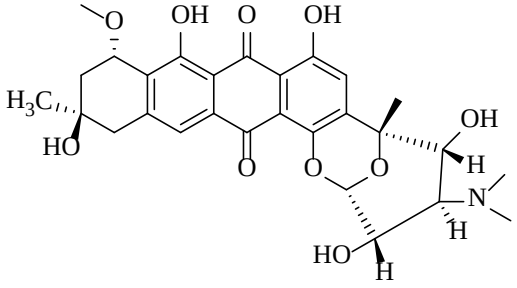
NCI_154585



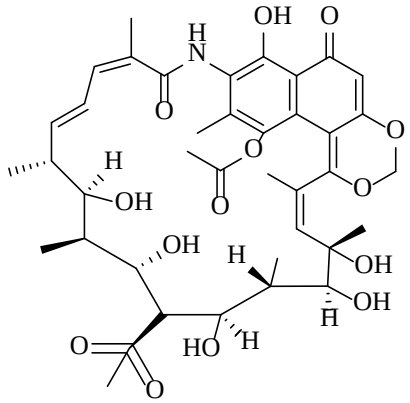
NCI_671136



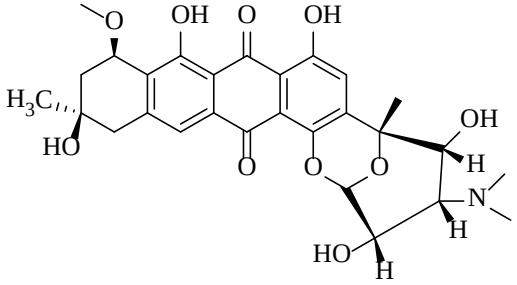
NCI_269148_a



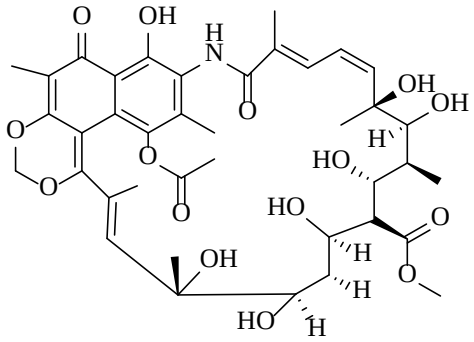
NCI_19990_a



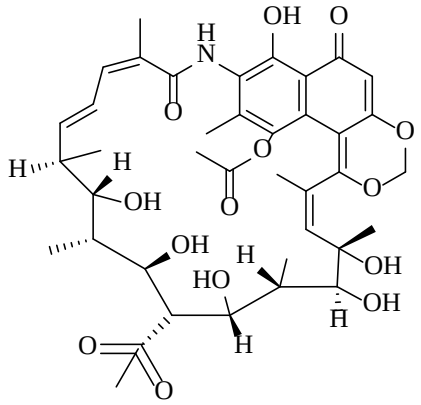
NCI_269148_b



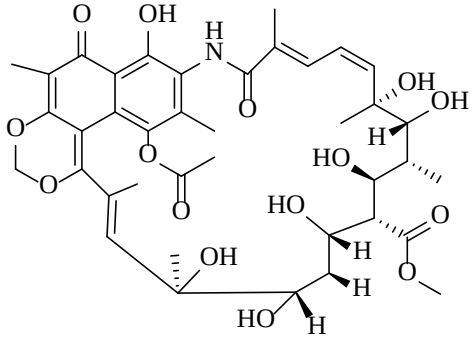
NCI_156219_a



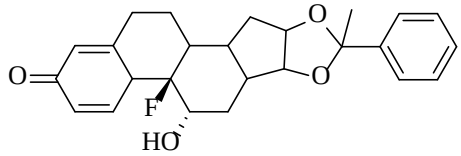
NCI_19990_b



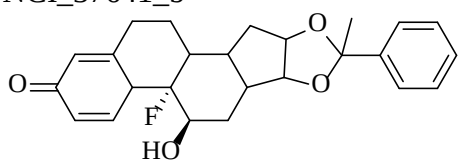
NCI_156219_b



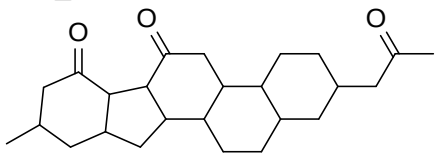
NCI_37641_a



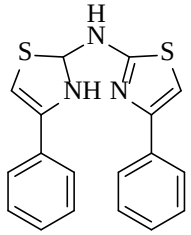
NCI_37641_b



NCI_59620



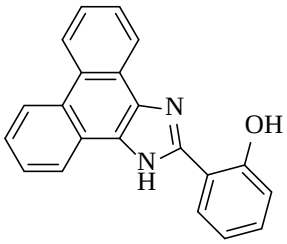
NCI_55691



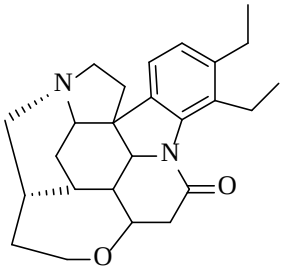
Br.

Br,

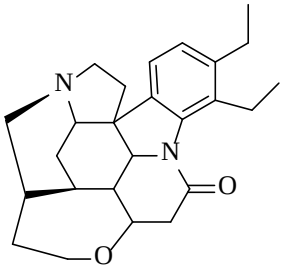
NCI_332670



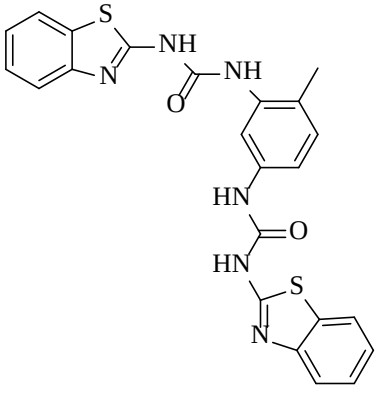
NCI_3391_a



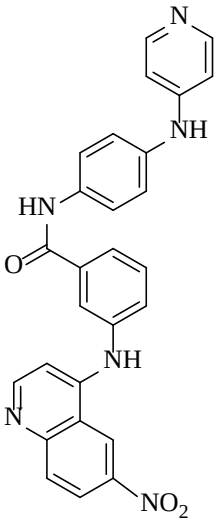
NCI_3391_b



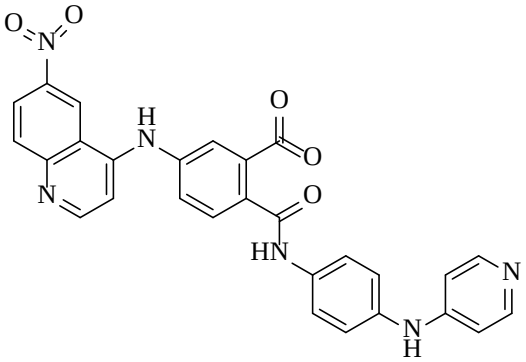
NCI_319990 ($C_{23}H_{18}N_6O_2S_2$)



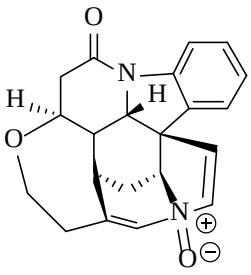
NCI_146771



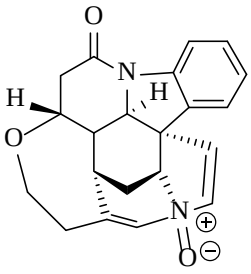
NCI_202386



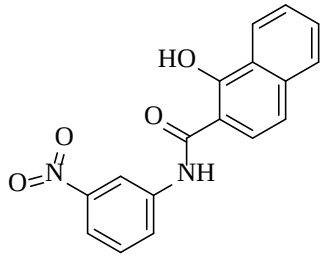
NCI_24951_a



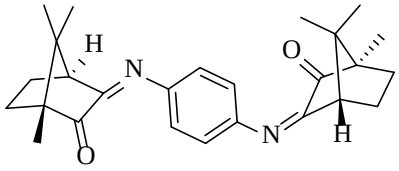
NCI_24951_b



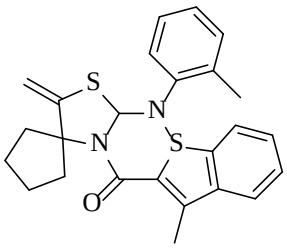
NCI_50648



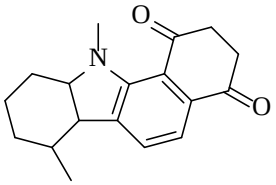
NCI_30663_a



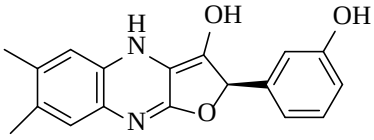
NCI_310326



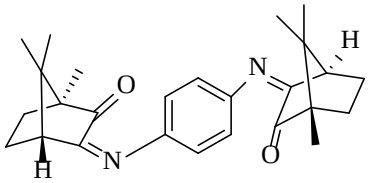
NCI_645330



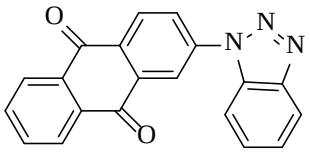
NCI_372499_a



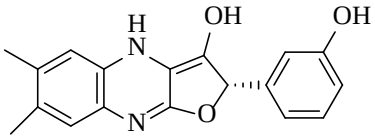
NCI_30663_b



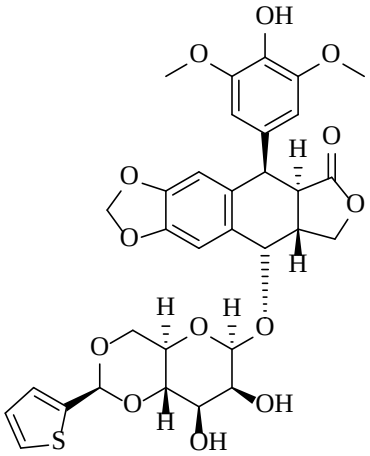
NCI_80313



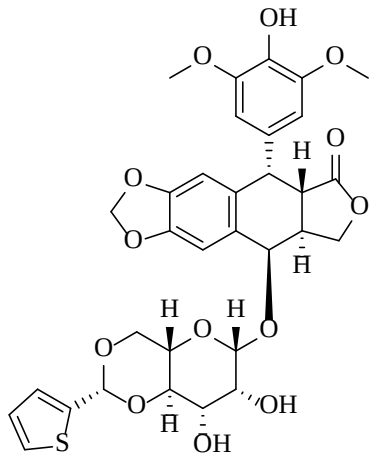
NCI_372499_b



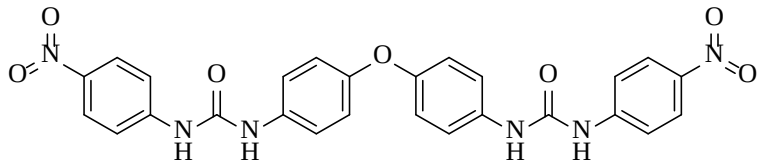
NCI_122819_a



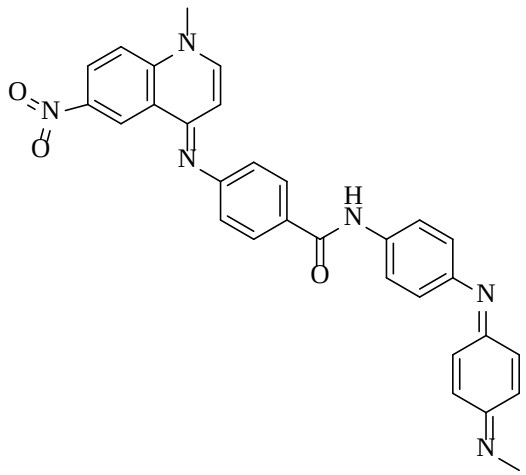
NCI_122819_b



NCI_80735



NCI_260594



Appendix 2. Interactions analysis for each ligand in the VST1 for the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_25457	-10.8	HBs: Val113 (HN)-O π-π interaction with Trp148
2. NCI_116702	-10.7	HBs: Val113 (HN)-O π-π interaction with Tyr114
3. NCI_309892	-10.3	HBs: H-Thr (111) π-system with Tyr114
4. NCI_109451	-10.2	π-system with Tyr114
5. NCI_127133	-10.2	π-system with Trp148
6. NCI_328101	-10.2	HBs: Val113 (HN)-N, Arg144 (HE)-O π-system with Trp148
7. NCI_372275_a	-10.2	π-system with Trp148
8. NCI_45545	-10.2	HBs: Val113 (HN)-N, H-Asp31 (O)
9. NCI_84100_b	-10.2	HBs: Val113 (HN)-O π-system with Tyr114
10. NCI_37168	-10.1	π-system with Trp148

Appendix 3. Interactions analysis for each ligand in the VST2 for the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_34567_a	-10.9	Hydrophobic contacts
2. NCI_37553	-10.9	π-π interaction with Trp148
3. NCI_25457	-10.8	Hydrophobic contacts
4. NCI_116702	-10.7	Hydrophobic contacts
5. NCI_58052	-10.6	Hydrophobic contacts
6. NCI_127133	-10.5	Hydrophobic contacts
7. NCI_293778	-10.5	Hydrophobic contacts
8. NCI_670283	-10.5	Hydrophobic contacts
9. NCI_328101	-10.4	Hydrophobic contacts
10. NCI_372499_b	-10.3	π-system with Trp148

Appendix 4. Interactions analysis for each ligand in the VST3 for the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_61610	-11.1	HBs: Gly112 (HN)-O π-π interaction with Trp148, Tyr114
2. NCI_293778	-11.0	Hydrophobic contacts
3. NCI_345647_a	-10.9	HBs: Gly112 (HN)-O Thr111 (HG1)-O
4. NCI_25457	-10.8	π-π interaction with Trp148
5. NCI_58052	-10.6	HBs: Gly112 (HN)-N Thr111 (HG1)-N
6. NCI_127133	-10.5	π-π interaction with Trp148, Tyr114
7. NCI_372499_b	-10.3	HBs: Arg144 (HN)-O Asp145 (HN)-O π-π interaction with Trp148
8. NCI_37553	-10.3	HBs: Asp31 (HN)-O π-π interaction with Trp148
9. NCI_309892	-10.2	HBs: H-Thr111 (O)
10. NCI_37168	-10.2	π-π interaction with Trp148

Appendix 5. Interactions analysis for each ligand in the VST4 for the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_61610	-11.1	HBs: Val113 (HN)-O π-π interaction with Trp148
2. NCI_37553	-11.0	π-π interaction with Trp148
3. NCI_345647_a	-10.9	HBs: Gly112 (HN)-O Thr111 (HN)-O H-Thr111 (OG1)
4. NCI_25457	-10.8	HBs: Val113 (HN)-O π-π interaction with Trp148, Tyr114
5. NCI_293778	-10.8	HBs: Val113 (HN)-N
6. NCI_127133	-10.7	π-π interaction with Trp148, Tyr114
7. NCI_116702	-10.6	HBs: Val113 (HN)-O π-π interaction with Tyr114, π-system with Trp148
8. NCI_58052	-10.6	HBs: Gly112 (HN)-N, Thr111 (HN)-N Asp31 (HN)-O
9. NCI_670283	-10.5	Hydrophobic contacts
10. NCI_324623	-10.3	HBs: Thr111 (HG1)-O

Appendix 6. Interactions analysis for each ligand in the VST5 for the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_127133	-8.3	HBs: H-Asp133 (O), H-Thr135 (O)
2. NCI_293778	-8.2	π-system with His-1
3. NCI_338042	-7.6	Hydrophobic contacts
4. NCI_132232	-7.5	HBs: H-Asp133 (O)
5. NCI_310326	-7.5	Hydrophobic contacts
6. NCI_328101	-7.4	HBs: Tyr4 (HH)-F, His-1(HT1)-O
7. NCI_69359_a	-7.3	Hydrophobic contacts
8. NCI_90737	-7.3	Hydrophobic contacts
9. NCI_122819_a	-7.2	HBs: Tyr4 (HN)-O
10. NCI_400976	-7.2	Hydrophobic contacts

Appendix 7. Interactions analysis for each ligand in the VST6 for the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_670283	-10.6	Hydrophobic contacts
2. NCI_319990	-10.2	HBs: Tyr144 (HH)-O
3. NCI_80731	-10.1	HBs: Tyr144 (HH)-O, H-Asp31 (O)
4. NCI_84100_b	-10.1	HBs: H-Tyr114 (OH)
5. NCI_372287_a	-10.0	HBs: Thr111 (HN)-O, Asp31 (HN)-O
6. NCI_84100_a	-9.9	HBs: H-Tyr114 (OH)
7. NCI_97920	-9.6	HBs: H-Asp31 (O)
8. NCI_58052	-9.5	HBs: H-Asp31 (O), Asn75 (HD22)-O
9. NCI_227186_a	-9.4	Hydrophobic contacts
10. NCI_293778	-9.4	HBs: Asp31 (HN)-N

Appendix 8. SASA's values (Å^2) in the simulations of the apo nsP3 and its complexes with different ligands: A: Ligand NCI_61610, B: Ligand NCI_25457, C: Ligand NCI_345647, D: Ligand NCI_670283, E: Ligand NCI_127133, F: Ligand NCI_37168vst1, G: Ligand NCI_372499vst2, H: Ligand NCI_37168vst3, I: Ligand NCI_324623vst4, J: Ligand NCI_400976vst5, K: Ligand NCI_293778vst6.

	nsP3	**nsP3-A**	**nsP3-B**	**nsP3-C**	**nsP3-D**	**nsP3-E**	**nsP3-F**	**nsP3-G**	**nsP3-H**	**nsP3-I**	**nsP3-J**	**nsP3-K**
Ala22	31.3	36.6	35.6		45.0		36.9		42.3	41.4		38.4
Pro25	75.5	57.8										65.9
Leu28	102.6	108.2			100.8							113.1
Val33	68.8	81.7	57.4	52.9	79.7		58.2	71.9	75.3	53.7		
Pro107	25.9	20.4	25.2		24.1		30.6		28.4	21.9		
Val113	129.8	146.5	138.8	130.8	142.2		138.2		137.8	147.7		140.4
Tyr114	70.2	63.4	69.3	63.8	87.4		63.5		55.8	54.8		49.7
Trp148	84.0	88.4	90.2	95.5				95.5	82.9			
Ile11	12.6			25.8								
Ala36	34.0			30.0								
Ala1	74.5					81.2					80.6	
Pro2	37.9					56.4					62.3	
Tyr4	39.2					41.7					50.2	
Ala23	10.5						13.4		14.9			10.6
Arg144	172.0						185.7	172.0				
Phe45	14.6									27.0		
Phe129	17.8										25.3	
Arg159	129.1										145.9	
Arg26	202.8											202.5
Thr111	24.3						25.5					

Appendix 9. Interactions analysis for each ligand at Pocket 1 of nsP3 in VST-5ns[a] in the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-11.6	Π-stacking with Tyr114
2. NCI_84100_b	-10.7	Π-stacking with Tyr114
3. NCI_84100_a	-10.6	Π-stacking with Tyr114
4. NCI_80997_b	-10.5	Hydrophobic contacts
5. NCI_37553	-10.4	Hydrophobic contacts
6. NCI_61610	-10.3	Hydrophobic contacts
7. NCI_670283	-10.2	Hydrophobic contacts
8. NCI_59620_a	-10.2	Hydrophobic contacts
9. NCI_308835	-10.2	HBs: Ser110(HN)-O
10. NCI_60785	-10.2	HBs: H-Leu108(O)

Appendix 10. Interactions analysis for each ligand at Pocket 1 of nsP3 in VST-10ns[a] in the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_37553	-11.1	HBs: Asn24(HD22)-O
2. NCI_293778	-11.1	HBs: Val33(HN)-N
3. NCI_60785_a	-11.0	HBs: H-Thr111(OG1)
4. NCI_59620_a	-10.8	HBs: Thr111(HN)-O, Leu108(HN)-O, Asn24(HD22)-O
5. NCI_27592_a	-10.7	Hydrophobic contacts
6. NCI_670283	-10.7	Hydrophobic contacts
7. NCI_82802_a	-10.7	HBs: Ile11(HN)-F
8. NCI_328101	-10.5	HBs: Ala22(HN)-F, Ala22(HN)-F, Asn24(HD22)-O
9. NCI_59620_a	-10.5	HBs: Thr111(HN)-O, Leu108(HN)-O, Asn24(HD22)-O
10. NCI_308835	-10.5	Hydrophobic contacts

Appendix 11. Interactions analysis for each ligand at Pocket 1 of nsP3 in VST-15ns[a] in the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-10.5	Hydrophobic contacts
2. NCI_308835	-10.3	Hydrophobic contacts
3. NCI_37553	-10.3	HBs: Asn24(HD22)-O
4. NCI_97920	-10.0	HBs: H-Asp31(O)
5. NCI_84100_b	-9.9	Hydrophobic contacts
6. NCI_59620_a	-9.8	HBs: Gly112(HN)-O, Asn24(HD22)-O
7. NCI_84100_a	-9.8	Hydrophobic contacts
8. NCI_37627	-9.8	HBs: Gly112(HN)-O
9. NCI_60785_b	-9.7	Hydrophobic contacts
10. NCI_25457	-9.7	HBs: Asp31 (HN)-N

Appendix 12. Interactions analysis for each ligand at Pocket 1 of nsP3 in VST-20ns[a] in the nsP3.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_37553	-12.1	HBs: Asn24(HD22)-O
2. NCI_61610	-11.9	HBs: Asn24(HD22)-O
3. NCI_670283	-11.7	HBs: Asp31(HN)-O
4. NCI_293778	-11.3	Hydrophobic contacts
5. NCI_60785_a	-11.1	HBs: H-Leu108(O)
6. NCI_63680	-11.1	Hydrophobic contacts
7. NCI_82802_a	-11.0	HBs: H-Thr111(OG1)
8. NCI_219894	-10.8	HBs: Leu108(HN)-O, Ser110(HN)-O
9. NCI_80997_b	-10.7	HBs: H-Leu108(O), Asn24(HD22)-O
10. NCI_328101	-10.7	HBss: Gly70(HN)-F

Appendix 13. Interactions analysis for each ligand in the nsP2 in VST1[a].

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_61610	-10.6	HBs: Tyr1177 (HH)-N
2. NCI_293778	-10.3	Hydrophobic contacts
3. NCI_116702	-9.2	HBs: Cys1290 (HG)-N
4. NCI_37553	-9.2	HBs: H-Tyr1079 (OH)
5. NCI_84100_a	-9.2	HBs: H-Trp1161(O), Lys1165(HZ2)-O
6. NCI_84100_b	-9.1	HBs: H-Trp1161(O), Lys1165(HZ2)-O
7. NCI_25457	-9.0	HBs: Cys1290(HG)-N
8. NCI_670283	-10.2	HBs: Lys1045(HZ1)-O, Lys1239(HZ1)-O
9. NCI_97920	-9.0	Hydrophobic contacts
10. NCI_58052	-8.8	HBs: Asn1167(HN)-O

Appendix 14. Interactions analysis for each ligand in the nsP2 in VST2[b].

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-10.3	π -system with Trp1084
2. NCI_61610	-10.1	π -system with His1151, Phe1225
3. NCI_670283	-9.7	π -system with Tyr1078
4. NCI_116702	-9.2	Hydrophobic contacts
5. NCI_217697	-9.2	Hydrophobic contacts
6. NCI_84100_a	-9.2	π -system with Tyr1177
7. NCI_84100_b	-9.2	π -system with Trp1084
8. NCI_298892_b	-9.1	HBs: His1222(HN)-O, Lys1239(HZ1)-N
9. NCI_37553	-9.1	Hydrophobic contacts
10. NCI_25457	-9.0	HBs: Cys1290(HG)-N

Appendix 15. Interactions analysis for each ligand in the nsP2 in VST3[a] .

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_217697	-9.3	Hydrophobic contacts
2. NCI_60783	-9.3	HBs: Lys1045(HZ1)-O, Lys1239(HZ1)-O
3. NCI_298892_b	-9.2	HBs: Lys1045(HZ1)-N, Lys1239(HZ1)-N, His1222(HN)-O
4. NCI_328101	-9.2	Hydrophobic contacts
5. NCI_25457	-9.1	Hydrophobic contacts
6. NCI_116702	-9.0	HBs: Lys1045(HZ1)-N, Lys1239(HZ1)-N
7. NCI_80997_a	-9.0	Hydrophobic contacts
8. NCI_109451	-8.9	Hydrophobic contacts
9. NCI_93427	-8.9	Hydrophobic contacts
10. NCI_127133	-8.8	Hydrophobic contacts

Appendix 16. Interactions analysis for each ligand in the nsP2 in VST4[b].

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-9.8	H-bonds: Lys1239(HZ1)-N
2. NCI_37553	-9.7	H-bonds: Lys1045(HZ1)-O
3. NCI_328101	-9.3	Hydrophobic contacts
4. NCI_670283	-9.3	HBs: Lys1239(HZ1)-O, Lys1045(HZ1)-O
5. NCI_25457	-9.2	Hydrophobic contacts
6. NCI_298892_b	-9.2	HBs: Lys1239(HZ1)-N, Lys1045(HZ1)-N, His1222(HN)-O
7. NCI_61610	-9.2	HBs: Lys1239(HZ1)-N, Lys1045(HZ1)-N
8. NCI_84100_a	-9.2	π -system with Tyr1177
9. NCI_84100_b	-9.2	π -system with Tyr1177
10. NCI_217697	-9.1	Hydrophobic contacts

Appendix 17. Interactions analysis for each ligand in the nsP2 in VST5[c].

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-9.8	π -stacking with Tyr1177
2. NCI_37553	-9.4	π -stacking with Tyr1177
3. NCI_269148_a	-8.5	HBs: Lys1155(HZ3)-O, His1222(HD1)-O
4. NCI_146771	-8.3	HBs: Gln1232(HE22)-O, Thr1295(HG1)-O, H-Thr1292(O)
5. NCI_227186_b	-8.3	HBs: Tyr1177(HH)-O. Ala1180(HN)-O, H-Glu1160(OE2), H-Glu1160(OE1)
6. NCI_308835	-8.3	Hydrophobic contacts
7. NCI_7524_b	-8.3	HBs: Lys1155(HZ3)-O
8. NCI_227186_a	-8.2	HBs: Phe1225(HE22)-O, Tyr1177(HH)-O
9. NCI_328101	-8.2	HBs: Gln1232(HE22)-O, Phe1225(HN)-N, Tyr1177(HH)-F
10. NCI_362639	-8.2	HBs: Phe1225(HN)-O, Thr1295(HG1)-O

Appendix 18. Interactions analysis for each ligand in the nsP2 in VST6[d].

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_37553	-9.6	Hydrophobic contacts
2. NCI_319990	-9.4	HBs: Asn1167(HN)-N
3. NCI_116702	-9.2	HBs: Cys1290(HG)-N
4. NCI_61610	-9.2	Hydrophobic contacts
5. NCI_25457	-9.0	HBs: Cys190(HG)-N
6. NCI_293778	-9.0	Hydrophobic contacts
7. NCI_97920	-9.0	Hydrophobic contacts
8. NCI_67436	-8.9	HBs: H-Ser1256(OG)
9. NCI_80997_b	-8.9	Hydrophobic contacts
10. NCI_84100_b	-8.9	Hydrophobic contacts

Appendix 19. Interactions analysis for each ligand in the nsP2 in VST7[e].

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-10.4	Hydrophobic contacts
2. NCI_670283	-9.7	Hydrophobic contacts
3. NCI_121868_a	-9.3	HBs: Trp1084(HE1)-O
4. NCI_37553	-9.2	Hydrophobic contacts
5. NCI_61610	-9.2	HBs: Tyr1079(HH)-O
6. NCI_319990	-8.9	HBs: Trp1084(HE1)-O, H-Leu1205(O), Tyr1079(HH)-O
7. NCI_328101	-8.9	HBs: Ser1048(HG)-F, Lys1091(HZ3)-F
8. NCI_67436	-8.9	HBs: H-Tyr1079(OH)
9. NCI_227186_a	-8.8	HBs: H-Leu1205(O)
10. NCI_122819_b	-8.7	HBs: Trp1084(HE1)-O, Lys1091(HZ3)-O

Appendix 20. Interactions analysis for each ligand in the nsP2 in VST8[f].

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-10.2	Hydrophobic contacts
2. NCI_80734	-9.2	HBs: H—Glu1270(OE1), H-Glu1270(OE2), Phe1070(HN)-N, Arg1271(HH21)-O
3. NCI_61610	-9.1	HBs: Arg1142(HH21)-O, H-Thr1313(OG1)
4. NCI_37553	-9.0	Hydrophobic contacts
5. NCI_80735	-8.7	HBs: H-Glu1270(OE1), Arg1271(HH21)-O
6. NCI_80997_a	-8.4	HBs: H-Glu1270(OE1)
7. NCI_345845	-8.1	HBs: H-Ser1067(O), H-Ala1099(O)
8. NCI_5157	-8.1	HBs: H-Asp1066(O), Ag1142(HH21)-O
9. NCI_58052	-8.1	HBs: H-Asp1066(O), Ag1140(HH21)-O
10. NCI_67436	-8.1	Hydrophobic contacts

Appendix 21. Sequences alignment of three crystal structures, namely the nsP2 protease of CHIKV (PDB id: 3TRK), the nsP2 protease of VEEV (PDB id: 2HWK) and the nsP2 protein of SINV (PDB id: 4GUA). The catalytic cysteine and histidine are highlighted with *.

```
                          *
3TRK.pdb          1  N-AFQNKANVCWAKSLVPILETAGIKLNDRQWSQIIQAFKEDKAYSPEVALNEICTR-YG   58
2HWK.pdb          1  -DVFQNKANVCWAKALVPVLKTAGIDMTTEQWN-TVDYFETDKAHSAEIVLNQLCVRFFG   58
4GUA.pdb          1  -NPFSCKTNVCWAKALEPILATAGIVLTGCQWSELFPQFADDKPHSAIYALDVICIKFFG   59
                        FqnKaNVCWAKaLvPiL TAGI lt          F DKa hSa eaL niC r fG

                                            *
3TRK.pdb         59  VDLDSGLFSKPLVSVYYAD-------NHWDNRPG-GKFGFNPEAASILERKYPFTKGKWN  110
2HWK.pdb         59  LDLDSGLFSAPTVPLSIRN-------NHWDNSPSPNMYGLNKEVVRQLSRRYPQLPRAVA  111
4GUA.pdb         60  MDLTSGLFSKQSIPLTYHPADSARPVAHWDNSPGTRKYGYDHAIAAELSRRFPVFQ-LAG  118
                      DLdSGLFSkp vpl y         nHWDNsPg  kyG n e a  LsRryP

3TRK.pdb        111  INKQICVTTRRIEDFNPTTNIIPVNRRLPHSLVAEHRPVKG-EREWLVNKINGHHVLLVS  169
2HWK.pdb        112  TGRVYDMNTGTLRNYDPRINLVPVNRRLPHALVLHHNEHPQSDFSSFVSKLKGRTVLVVG  171
4GUA.pdb        119  KGTQLDLQTGRTRVISAQHNLVPVNRNLPHALVPEYKEKQPGPVKKFLNQFKHHSVLVVS  178
                      g q d  Tgr r   p  NlvPVNRrLPHaLV eh e        fvnk kgh VLvVs

3TRK.pdb        170  GYNLALPTKRVTWVAPLGVRGADYTYNLELGLPATLGRYDLVVINIHTPFRIHHYQQCVD  229
2HWK.pdb        172  E-KLSVPGKMVDWLS--DRPEATFRARLDLGIPGDVPKYDIIFVNVRTPYKYHHYQQCED  228
4GUA.pdb        179  EEKIEAPRKRIEWIAPIGIAGADKNYNLAFGFPP-QARYDLVFINIGTKYRNHHFQQCED  237
                      ekl  P Krv W a  g  gAdyn  Ll G P    rYDlvfiNi TpyrHHy QQCeD

3TRK.pdb        230  HA-KL-QLGGDSLRLLKPGGSLLIRAYGYADRTSERVICVLGRKFRSSRALKPPCVTSNT  287
2HWK.pdb        229  HAIKLSMLTKKACLHLNPGGTCVSIGYGYADRASESIIGAIARQFKFSRVCKPKSSLEET  288
4GUA.pdb        238  HAATLKTLSRSALNCLNPGGTLVVKSYGYADRNSEDVVTALARKFVRVSAARPDCVSSNT  297
                      HA kL  L   al  LnPGGtlv   YGYADR SE vi alaRkF   srakP cv snT

3TRK.pdb        288  E-FFLFSNFDNGR-RNFTTH-VNNQLNAAF-----------------------------  314
2HWK.pdb        289  EVLFVFIGYDRKART-HNPYKLSSTLTNIYTGS---------------------------  320
4GUA.pdb        298  EMYLIFRQLDNSRTRQFTPHHLNCVISSVYEGTRDGVGAAPSYRTKRENIADCQEEAVVN  357
                      E   fF   Dn r  rftph ln  l   y

3TRK.pdb             ------------------------------------------------------------
2HWK.pdb             ------------------------------------------------------------
4GUA.pdb        358  AANPLGRPGEGVCRAIYKRWPTSFTDSATETGTARMTVCLGKKVIHAVGPDFRKHPEAEA  417

3TRK.pdb             ------------------------------------------------------------
2HWK.pdb             ------------------------------------------------------------
4GUA.pdb        418  LKLLQNAYHAVADLVNEHNIKSVAIPLLSTGIYAAGKDRLEVSLNCLTTALDRTDADVTI  477

3TRK.pdb             ------------------------------------------------------------
2HWK.pdb             ------------------------------------------------------------
4GUA.pdb        478  YCLDKKWKERIDAALQLKESVTELKDEDMEIDDELVWIHPDSCLKGRKGFSTTKGKLYSY  537

3TRK.pdb             ------------------------------------------------------------
2HWK.pdb             ------------------------------------------------------------
4GUA.pdb        538  FEGTKFHQAAKDMAEIKVLFPNDQESNEQLCAYILGETMEAIREKCPVDHNPSSSPPKTL  597

3TRK.pdb             ------------------------------------------------------------
2HWK.pdb             ------------------------------------------------------------
4GUA.pdb        598  PCLCMYAMTPERVHRLRSNNVKEVTVCSSTPLPKHKIKNVQKVQCTKVVLFNPHTPAFVP  657

3TRK.pdb             -----
2HWK.pdb             -----
4GUA.pdb        658  ARKYI  662
```

Appendix 22. SASA's values (Å^2) in the simulations of the apo nsP2 and its complexes with different ligands: A: Ligand NCI_61610, B: Ligand NCI_217697, C: Ligand NCI_37553, D: Ligand NCI_293778vst4, E: Ligand NCI_293778vst5, F: Ligand NCI_293778vst2, G: Ligand NCI_293778vst7, H: Ligand NCI_293778vst8, I: Ligand NCI_217697w, J: Ligand NCI_25457w, K: Ligand NCI_362639, L: Ligand NCI_84100_b, M: Ligand NCI_67436.

	nsP3	**nsP3-A**	**nsP3-B**	**nsP3-C**	**nsP3-D**	**nsP3-E**	**nsP3-F**	**nsP3-G**	**nsP3-H**	**nsP3-I**	**nsP3-J**	**nsP3-K**	**nsP3-L**	**nsP3-M**
Lys1045	89.6	90.6	89.8		115.4					96.5				
Gly1176	42.6				40.0									
Pro1191	43.4	59.0	45.8											
Val1194	46.1	50.3	45.8											
Leu1203	21.5	27.2	22.6		34.3									
Glu1204	35.9	69.9			110.1					67.7				
Ile1221	27.5	46.4	36.5		38.1					41.2				
Met1238	17.3	29.3	30.6		44.5		24.4			31.0				
Lys1239	49.1	68.8	90.2		85.3					75.1				
Met1242	**46.2**	51.6			56.7									
Leu1243	2.6	11.0	10.8		16.3					13.5				
Tyr1047	24.7		35.4				21.4			25.8				
Trp1161	140.8			137.0							144.6		142.3	
Leu1162	25.1			27.9							29.3		24.9	
Lys1165	170.7			168.6							175.2		182.3	
Ile1166	24.8			26.1									22.0	
Ser1282	57.8			53.6										
Ser1283	33.6			19.1										
Arg1284	123.8			132.7										
Leu1286	37.5			39.7									33.9	
Pro1288	22.4			20.2									25.9	

Pro1289	77.6	101.4							100.4		95.2
Leu1300	20.5	32.0									25.6
Ser1302	27.6	31.6									
Ala1046	24.9		54.4		42.7	43.1					
Hsd1222	95.2		95.9	64.3				99.4			
Thr1223	11.4		32.4								
Pro1224	94.1		117.5	106.3							
Hsd1236	8.1		14.6								
Met1242	**46.2**		56.7								
Lys1155	205.1			210.6							
Gly1156	50.0			66.7							
Glu1157	90.5			87.5					94.0		90.1
Arg1158	152.8			152.4							
Met1159	10.8			64.0							
Gly1176	42.6			38.3							
Tyr1177	115.5			84.6							
Leu1179	-15.3			21.5							
Thr1292	73.8			42.6						51.9	
Ser1048	21.2				27.5						
Tyr1078	88.6				79.5						
Tyr1079	114.2				129.2	129.8					
Asn1082	121.9				138.2						
Trp1084	59.4				63.6	49.2					
Leu1205	109.2				110.9	129.2					
Val1077	6.9					11.3					
Ile1102	103.9						105.2				
Leu1103	25.8						31.8				

Pro1191	43.2	43.2		
Leu1192	116.0	115.5		
Gly1193	67.0	76.2		
Val1194	46.6	46.1		
Ser1293	5.4		5.8	
Asp1066	93.8			79.3
Ser1067	28.6			37.5
Asn1140	63.4			69.0
Arg1141	46.0			47.8
Arg1142	211.6			216
Pro1144	38.7			34.0
His1145	11.7			10.9
Glu1270	45.6			37.6
Arg1271	127.2			122.0
Cys1274	40.8			31.7
Val1275	67.4			72.2

Appendix 23. Interactions analysis for each ligand in the nsP2 in VST-10ns.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-8.6	Hydrophobic contacts
2. NCI_37553	-8.5	Hydrophobic contacts
3. NCI_59620	-8.4	HBs: Thr308(HN)-O
4. NCI_7524_b	-8.3	Hydrophobic contacts
5. NCI_308835	-8.2	Hydrophobic contacts
6. NCI_227186_a	-8.0	HBs: Thr308(HN)-O, His309(HE2)-O, Arg137(HN)-O
7. NCI_328101	-8.0	HBs: Thr308(HN)-N, Thr308(HN)-O
8. NCI_339161	-7.8	HBs: Thr308(HG1)-O
9. NCI_270738_b	-7.8	HBs: Asn135(HD22)-O
10. NCI_58052	-7.8	HBs: N-Thr308(OG1), Asn135(HD22)-O

Appendix 24. Interactions analysis for each ligand in the nsP2 in VST-20ns.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_345647_a	-9.0	HBs: Lys101(HZ1)-O, H-Asp59(OD2), Arg136(HH12)-O
2. NCI_338042	-8.8	Hydrophobic contacts
3. NCI_293778	-8.7	HBs: Thr308(HG1)-N
4. NCI_67436	-8.3	HBs: H-Asn135(O), Arg137(HH11)-O
5. NCI_80997_a	-8.3	HBs: Lys101(HZ1)-O
6. NCI_338963	-8.2	Hydrophobic contacts
7. NCI_319990	-8.2	HBs: His309(HD1)-N
8. NCI_25457	-8.0	Hydrophobic contacts
9. NCI_670283	-7.9	Hydrophobic contacts
10. NCI_326182	-7.9	Hydrophobic contacts

Appendix 25. Interactions analysis for each ligand in the nsP2 in VST-30ns.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-8.3	Hydrophobic contacts
2. NCI_3045647_a	-8.0	Hydrophobic contacts
3. NCI_97920	-8.0	Hydrophobic contacts
4. NCI_327702	-7.9	HBs: Gln236(HE22)-N
5. NCI_156516	-7.9	Hydrophobic contacts
6. NCI_80313	-7.8	HBs: Gln236(HE22)-O, Gln236(HE21)-N
7. NCI_116709_b	-7.8	Hydrophobic contacts
8. NCI_37553	-7.8	HBs: Arg266(HH12)-O
9. NCI_326757	-7.6	HBs: H-Val270(O)
10. NCI_154585	-7.6	HBs: H-Gln236(O)

Appendix 26. Interactions analysis for each ligand in the nsP2 in VST-40ns.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-8.5	Hydrophobic contacts
2. NCI_61610	-7.9	Hydrophobic contacts
3. NCI_122819_b	-7.7	HBs: H-Thr307(OG1), Thr308(HN)-O
4. NCI_671136	-7.6	HBs: Thr308(HN)-O
5. NCI_19990_a	-7.6	HBs: H-Thr307(OG1), His309(HD1)-O, H-Glu265(OE2), Arg137(HN)-O
6. NCI_37641_b	-7.5	HBs: Thr308(HN)-O, H-Thr307(OG1), His309(HN)-O
7. NCI_269148_b	-7.4	HBs: H-Arg137(O), Thr308(HN)-O, H-Thr307(OG1), H-Asn305(OD1), Arg273(HH22)-O
8. NCI_319990	-7.4	HBs: Arg137(HN)-N, H-Arg137(O), H-Thr137(OG1)
9. NCI_156219_b	-7.4	HBs: H-Thr302(OG1), Thr308(HN)-O, H-Thr308(OG1), Asn312(HD21)-O, His309(HN)-O
10. NCI_59620_b	-7.3	HBs: Thr308(HG1)-O

Appendix 27. Interactions analysis for each ligand in the nsP2 in VST-50ns.

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_61610	-8.2	HBs: H-Thr308(OG1)
2. NCI_84100_a	-7.2	HBs: Thr308(HN)-O, H-Thr308(OG1)
3. NCI_146771	-7.2	Hydrophobic contacts
4. NCI_84100_b	-7.1	HBs: Thr308(HN)-O, H-Thr308(OG1)
5. NCI_25457	-7.0	Hydrophobic contacts
6. NCI_116702	-7.0	Hydrophobic contacts
7. NCI_37553	-7.0	Hydrophobic contacts
8. NCI_80731	-7.0	HBs: His309(HD1)-O, H-Glu265(OE1)
9. NCI_35691	-7.0	HBs: H-Leu138(O)
10. NCI_332670	-7.0	HBs: Asn312(HD21)-O

Appendix 28. Binding sites identified in the immature form (PDB id: 3N40) using receptor cavities tools in Accelerys Discovery Studio software. The binding site detected is based on X, Y, Z coordinates; volume; and a set of points (point count).

Pocket location	X	Y	Z	Volume ($Å^3$)	Point count
Between E1 domain II and E2 domain C (Pocket 1)	-25.6	21.1	24.9	131.8	1054
Between E1 domain II and the E2 β-ribbon (Pocket 1)	-13.8	-4.7	19.7	352.3	2818
At the E2 β-ribbon (Pocket 2)	-22.4	-0.2	8.9	123.5	988
Behind the fusion loop, between E3, E2 domain A, E2 domain B (Pocket 4)	-9.6	-7.3	8.2	26.8	214
Between E1 domain I-domain III and E2 domain C (Pocket 3)	-11.1	-58.4	19.7	42.3	338
Between E1 domain II-E3	-34.1	21.1	39.4	49.0	392
Between E1 domain I-E1 domain III	-11.1	-58.4	19.7	42.3	338
Between E1 domain II-near E2 β-ribbon	-26.6	-4.4	25.2	38.8	310
Between E1 domain II-E1 domain I	-6.9	-23.9	33.7	36.1	289
Between E1 domain I and E2 domain C	-2.9	-25.2	19.4	32.8	262
Between E1 domain I-E1 domain III	-4.9	-37.4	12.9	31.8	254
Between E1 domain II-E2 domain A and E3	-33.1	33.3	25.2	27.6	221
Between E2 domain B-E3	-54.4	32.1	29.4	27.1	217
Between E2 domain C	-9.6	-7.3	8.2	26.8	214
Between E2 domain A and E2 β-ribbon	-17.6	7.3	8.2	23.4	187

Appendix 29. Binding sites identified in the mature form (PDB id: 3N42) using receptor cavities tools in Accelerys Discovery Studio software. The binding site detected is based on X, Y, Z coordinates; volume; and a set of points (point count).

Pocket location	X	Y	Z	Volume ($Å^3$)	Point count
Between E1 domain II and E2 domain C (Pocket 1)	-16.7	2.0	-19.9	621.2	4970
At the E2 β-ribbon (Pocket 2)	-19.7	-2.9	-9.9	156.0	1248
Behind the fusion loop, between E3, E2 domain A, E2 domain B (Pocket 4)	-33.4	-14.2	-34.9	523.0	4184
Between E1 domain I-domain III and E2 domain C (Pocket 3)	-6.4	24.3	-26.2	129.6	1037
Between E1 domain II, E2 β-ribbon	-16.7	2.0	-19.9	621.2	4970
E2 domain A	-43.4	-28.7	-22.9	143.5	1148
Between E1 domain I-E1 domain III	-6.4	37.8	-13.4	139.7	1118
Between E1 domain I-E1 domain II and E2 domain C	-6.4	24.8	-26.2	129.6	1037
Between E2 domain A-E3	-46.7	-14.7	-22.9	100.5	804
Between E2 domain B-E2 domain A	-53.2	-31.2	-27.7	35.8	283
E1 domain I	-16.2	51.3	-22.4	25.4	203
E2 domain C	-10.7	7.8	-0.9	25.2	202
E2 domain A	-47.2	-12.9	-12.2	23.5	188
E3	-47.2	-12.9	-12.2	23.5	178

Appendix 30. Interactions analysis for each ligand in blind docking in the 3N40

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.2	π-system with Tyr51
2. NCI_61610	-12.2	HBs: H-Glu99(O)
3. NCI_37553	-11.7	Hydrophobic contacts
4. NCI_156219_b	-11.5	HBs: H-Ile101(O)
5. NCI_84100_b	-10.9	HBs: H-Tyr233(OH)
6. NCI_84100_a	-10.9	HBs: H-Tyr233(OH)
7. NCI_116702	-10.8	HBs: H-Tyr233(OH)
8. NCI_19990_a	-10.8	Hydrophobic contacts
9. NCI_227186_a	-10.7	HBs: Ser238(HN)-O, H-Thr53(O), Tyr301(HH)-O, H-Tyr301(OH)
10. NCI_227186_b	-10.6	HBs: Ile55(HN)-O, Thr53(HN)-O, H-Tyr51(O)

Appendix 31. Interactions analysis for each ligand in blind docking in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.7	π-system with Tyr129
2. NCI_37553	-11.4	Hydrophobic contacts
3. NCI_61610	-11.2	HBs: H-Pro128(O), Ile55(HN)-O
4. NCI_156219_b	-11.1	HBs: H-Glu50(OE2)
5. NCI_84100_a	-11.0	π-system with Tyr242, Pro237
6. NCI_227186_b	-10.9	HBs: Ser238(HN)-O, H-G;u50(OE2)
7. NCI_84100_b	-10.9	HBs: H-Asp43(OD2)
8. NCI_7524_a	-10.9	HBs: H-Glu50(OE2), Thr53(HN)-O, Arg36(HH12)-O, Lys52(HZ1)-O
9. NCI_116702	-10.8	π-stacking with Pro133
10. NCI_308835	-10.8	Hydrophobic contacts

Appendix 32. Interactions analysis for each ligand in VST3[a] in the 3N40

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.1	π-system with Tyr51
2. NCI_61610	-11.9	Hydrophobic contacts
3. NCI_37553	-11.7	Hydrophobic contacts
4. NCI_156219_b	-11.5	HBs: H-Ile101(O)
5. NCI_84100_b	-10.9	π-system with Pro216
6. NCI_84100_a	-10.9	HBs: H-Tyr233(OH)
7. NCI_116702	-10.8	HBs: Arg100(HE)-N
8. NCI_19990_a	-10.8	Hydrophobic contacts
9. NCI_227186_a	-10.7	HBs: Ser238(HN)-O, H-Thr53(O), H-Asn103(OD1)
10. NCI_227186_b	-10.7	HBs: Ser238(HN)-O

Appendix 33. Interactions analysis for each ligand in VST4[b] in the 3N40

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.1	π-system with Tyr233
2. NCI_61610	-12.3	HBs: H-Glu99(O)
3. NCI_156219_b	-11.5	Hydrophobic contacts
4. NCI_308835	-11.0	Hydrophobic contacts
5. NCI_84100_a	-10.9	HBs: H-Tyr233(OH)
6. NCI_84100_b	-10.9	HBs: H-Tyr233(OH)
7. NCI_37553	-10.8	Hydrophobic contacts
8. NCI_116702	-10.8	HBs: Arg100(HE)-N
9. NCI_67436	-10.6	Hydrophobic contacts
10. NCI_227186_b	-10.6	HBs: Ser238(HN)-O

Appendix 34. Interactions analysis for each ligand in VST5[c] in the 3N40

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.1	Hydrophobic contacts
2. NCI_61610	-12.3	HBs: H-Glu99(O)
3. NCI_37553	-11.7	Hydrophobic contacts
4. NCI_156219_b	-11.5	HBs: H-Pro192(O), H-Tyr51(O)
5. NCI_260594	-11.4	Hydrophobic contacts
6. NCI_308835	-11.1	Hydrophobic contacts
7. NCI_84100_a	-11.0	Hydrophobic contacts
8. NCI_84100_b	-10.9	Hydrophobic contacts
9. NCI_227186_a	-10.9	HBs: Tyr301(HH)-O, Asn103(HD22)-O
10. NCI_19990_a	-10.8	HBs: H-Tyr233(OH)

Appendix 35. Interactions analysis for each ligand in VST6[d] in the 3N40

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.1	Hydrophobic contacts
2. NCI_61610	-12.3	HBs: H-Glu99(O)
3. NCI_37553	-11.7	Hydrophobic contacts
4. NCI_156219_b	-11.5	HBs: H-Pro192(O), H-Ile101(O)
5. NCI_260594	-11.4	Hydrophobic contacts
6. NCI_308835	-11.4	Hydrophobic contacts
7. NCI_84100_a	-11.0	Hydrophobic contacts
8. NCI_84100_b	-10.9	Hydrophobic contacts
9. NCI_227186_a	-10.9	HBs: H-Thr53(O), Ser238(HN)-O, H-Asn103(OD1)
10. NCI_19990_a	-10.8	Hydrophobic contacts

Appendix 36. Interactions analysis for each ligand in VST7[e] in the 3N40

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.1	Hydrophobic contacts
2. NCI_61610	-12.3	HBs: H-Glu99(O)
3. NCI_37553	-11.7	Hydrophobic contacts
4. NCI_156219_b	-11.5	Hydrophobic contacts
5. NCI_260594	-11.5	Hydrophobic contacts
6. NCI_308835	-11.4	Hydrophobic contacts
7. NCI_84100_a	-11.0	Hydrophobic contacts
8. NCI_84100_b	-11.0	HBs: H-Glu50(OE2)
9. NCI_227186_a	-10.9	HBs: H-Thr53(O), Ser238(HN)-O, Tyr301(HH)-O, Asn103(HD22)-O
10. NCI_19990_a	-10.9	Hydrophobic contacts

Appendix 37. Interactions analysis for each ligand in VST8[f] in the 3N40

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.0	Hydrophobic contacts
2. NCI_37553	-11.7	Hydrophobic contacts
3. NCI_37553	-11.5	Hydrophobic contacts
4. NCI_156219_b	-11.4	Hydrophobic contacts
5. NCI_260594	-11.4	Hydrophobic contacts
6. NCI_308835	-10.9	HBs: Asn103(HD22)-O, Ser238(HN)-O, Tyr301(HH)-O
7. NCI_84100_a	-10.8	HBs: H-Tyr233(OH), H-Thr53(O)
8. NCI_84100_b	-10.8	HBs: Arg247(HH21)-O
9. NCI_227186_a	-10.8	HBs: Arg100(HE)-N
10. NCI_19990_a	-10.7	HBs: Arg168(HH21)-O

Appendix 38. Interactions analysis for each ligand in VST3[a] in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.7	π-system with Pro128, Pro129
2. NCI_37553	-11.5	Hydrophobic contacts
3. NCI_61610	-11.3	HBs: H-Pro128(O), Ile55(HN)-O
4. NCI_156219_b	-11.1	HBs: H-Glu50(OE2)
5. NCI_84100_a	-11.0	Hydrophobic contacts
6. NCI_7524_a	-10.9	HBs: Ser238(HN)-O
7. NCI_84100_b	-10.9	HBs: H-Asp43(OD2)
8. NCI_25457	-10.8	HBs: Thr53(HN)-O, Arg36(HH12)-O, Lys52(HZ1)-O
9. NCI_116702	-10.8	Hydrophobic contacts
10. NCI_122819	-10.7	Hydrophobic contacts

Appendix 39. Interactions analysis for each ligand in VST4[b] in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.7	Hydrophobic contacts
2. NCI_37553	-11.5	Hydrophobic contacts
3. NCI_61610	-11.4	π-stacking with Tyr390
4. NCI_156219_b	-11.1	HBs: H-Tyr51(O)
5. NCI_227186_b	-10.9	HBs: Ser238(HN)-O
6. NCI_84100_a	-10.9	Hydrophobic contacts
7. NCI_7524_a	-10.9	HBs: Arg36(HH11)-O, Thr53(HN)-O
8. NCI_84100_b	-10.9	Hydrophobic contacts
9. NCI_308835	-10.8	Hydrophobic contacts
10. NCI_116702	-10.8	Hydrophobic contacts

Appendix 40. Interactions analysis for each ligand in VST5[c] in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.7	Hydrophobic contacts
2. NCI_61610	-11.5	π-stacking with Tyr390
3. NCI_37553	-11.4	Hydrophobic contacts
4. NCI_156219_b	-11.1	HBs: H-Pro128(O), π-system with His127
5. NCI_84100_a	-11.0	Hydrophobic contacts
6. NCI_227186_b	-10.9	HBs: Ser238(HN)-O, H-Glu50(OE2)
7. NCI_7524_a	-10.9	HBs: Arg36(HH11)-O, Thr53(HN)-O, Lys52(HZ1)-O
8. NCI_84100_b	-10.9	HBs: H-Asp43(OD2)
9. NCI_308835	-10.8	Hydrophobic contacts
10. NCI_25457	-10.8	Hydrophobic contacts

Appendix 41. Interactions analysis for each ligand in VST6[d] in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.7	Hydrophobic contacts
2. NCI_61610	-11.6	Hydrophobic contacts
3. NCI_37553	-11.4	Hydrophobic contacts
4. NCI_156219_b	-11.1	HBs: H-Glu50(OE2)
5. NCI_84100_a	-11.0	Hydrophobic contacts
6. NCI_84100_b	-10.9	HBs: Arg36(HH11)-O, Thr53(HN)-O, Lys52(HZ1)-O, H-Tyr51(OH)
7. NCI_7524_a	-10.9	HBs: Ser238(HN)-O
8. NCI_227186_b	-10.9	HBs: H-Asp43(OD2)
9. NCI_116702	-10.8	HBs: H-Glu50(OE2)
10. NCI_308835	-10.8	Hydrophobic contacts

Appendix 42. Interactions analysis for each ligand in VST7[e] in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.7	Hydrophobic contacts
2. NCI_61610	-11.6	Hydrophobic contacts
3. NCI_37553	-11.5	Hydrophobic contacts
4. NCI_156219_b	-11.1	HBs: H-Pro128(O)
5. NCI_84100_a	-11.0	Hydrophobic contacts
6. NCI_84100_b	-10.9	HBs: H-Asp43(OD2)
7. NCI_7524_a	-10.9	HBs: Arg36(HH11)-O, Thr53(HN)-O, H-Tyr51(OH)
8. NCI_227186_b	-10.9	HBs: Ser238(HN)-O, H-G;u50(OE2)
9. NCI_116702	-10.8	Hydrophobic contacts
10. NCI_308835	-10.8	Hydrophobic contacts

Appendix 43. Interactions analysis for each ligand in VST8[f] in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.7	Hydrophobic contacts
2. NCI_37553	-11.5	Hydrophobic contacts
3. NCI_156219_b	-11.1	Hydrophobic contacts
4. NCI_25457	-10.8	Hydrophobic contacts
5. NCI_121868	-10.8	Hydrophobic contacts
6. NCI_308835	-10.8	Hydrophobic contacts
7. NCI_116702	-10.8	Hydrophobic contacts
8. NCI_122819	-10.7	HBs: Val231(HN)-O, Ile37(HN)-O, His230(HD1)-O
9. NCI_156219	-10.7	HBs: Ile37(HN)-O, His230(HD1)-O
10. NCI_227186_b	-10.7	HBs: H-Ile55(O), Ser238(HN)-O

Appendix 44. Interactions analysis for each ligand in VST[1a] in focused docking in the 3N40

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.1	Hydrophobic contacts
2. NCI_37553	-11.4	Hydrophobic contacts
3. NCI_156219_b	-11.3	HBs: H-Ile101(O)
4. NCI_19990_a	-10.9	HBs: H-Thr53(O)
5. NCI_84100_b	-10.8	Hydrophobic contacts
6. NCI_84100_a	-10.8	Hydrophobic contacts
7. NCI_116702	-10.7	HBs: Arg100(HE)-N
8. NCI_60785_b	-10.5	Hydrophobic contacts
9. NCI_308835	-10.5	Hydrophobic contacts
10. NCI_227186_b	-10.4	HBs: H-Thr53(O)

Appendix 45. Interactions analysis for each ligand in VST[2a] in focused docking in the 3N40

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-11.4	HBs: Ile200(HN)-S
2. NCI_37641_b	-10.7	HBs: Arg168(HH21)-O
3. NCI_37553	-10.0	HBs: Arg168(HH22)-O
4. NCI_328101	-10.0	HBs: Arg168(HH21)-F
5. NCI_116702	-9.9	Hydrophobic contacts
6. NCI_37641_a	-9.9	HBs: H-Pro197(O), Arg168(HH21)-O
7. NCI_121868_a	-9.8	Hydrophobic contacts
8. NCI_308835	-9.7	HBs: Arg168(HH21)-O
9. NCI_24951_a	-9.5	HBs: O-Pro197(O)
10. NCI_127133	-9.5	Hydrophobic contacts

Appendix 46. Interactions analysis for each ligand in VST[1b] in focused docking in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_293778	-13.7	Hydrophobic contacts
2. NCI_37553	-11.4	Hydrophobic contacts
3. NCI_61610	-11.3	HBs: H-Glu35(O)
4. NCI_25457	-10.8	Hydrophobic contacts
5. NCI_116702	-10.8	Hydrophobic contacts
6. NCI_308835	-10.7	Hydrophobic contacts
7. NCI_97920	-10.6	Hydrophobic contacts
8. NCI_7524_b	-10.4	Hydrophobic contacts
9. NCI_671136	-10.3	Hydrophobic contacts
10. NCI_84100_a	-10.2	Hydrophobic contacts

Appendix 47. Interactions analysis for each ligand in VST[2b] in focused docking in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_61610	-11.6	π-π interaction with Phe141
2. NCI_37553	-11.1	HBs: Arg144(HE)-O, H-Cys266(SG)
3. NCI_308835	-11.1	HBs: Arg104(HH12)-O
4. NCI_84100_b	-10.9	HBs: H-Asp43(OD2)
5. NCI_84100_a	-10.8	Hydrophobic contacts
6. NCI_82802	-10.7	Hydrophobic contacts
7. NCI_670283	-10.5	Hydrophobic contacts
8. NCI_7524_b	-10.4	HBs: Arg144(HE)-O
9. NCI_293778	-10.4	HBs: His142(HE2)-N
10. NCI_328101	-10.3	HBs: Arg267(HN)-F

Appendix 48. Interactions analysis for each ligand in VST[3b] in focused docking in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_61610	-11.5	π-stacking with Phe312, Tyr15
2. NCI_293778	-11.0	π-stacking with Phe312, Tyr15
3. NCI_37553	-9.9	HBs: H-Gly12(O)
4. NCI_670283	-9.4	Hydrophobic contacts
5. NCI_332670	-9.4	Hydrophobic contacts
6. NCI_328101	-9.3	HBs: Lys16(HZ3)-F
7. NCI_60785_a	-9.2	Hydrophobic contacts
8. NCI_50648	-9.2	Hydrophobic contacts
9. NCI_60785_b	-9.2	Hydrophobic contacts
10. NCI_308835	-9.1	Hydrophobic contacts

Appendix 49. Interactions analysis for each ligand in VST[4b] in focused docking in the 3N42

Compound	Binding affinity (kcal/mol)	Interactions
1. NCI_121868_a	-10.7	HBs: Lys61(HZ2)-O
2. NCI_24951_b	-10.6	Hydrophobic contacts
3. NCI_24951_a	-10.2	HBs: Lys200(HZ2)-O
4. NCI_332186	-9.7	HBs: H-Tyr93(O)
5. NCI_332670	-9.6	Hydrophobic contacts
6. NCI_3391_a	-9.6	HBs: Lys61(HZ2)-O
7. NCI_80313	-9.5	HBs: Lys200(HZ2)-N
8. NCI_310326	-9.3	Hydrophobic contacts
9. NCI_645330	-9.3	Hydrophobic contacts
10. NCI_97920	-9.2	Hydrophobic contacts