UNIVERSITY
OF WOLLONGONG
AUSTRALIA

University of Wollongong
Research Online

Faculty of Science, Medicine and Health - Papers:
part A

Faculty of Science, Medicine and Health

2017

Molecular docking studies to explore potential binding pockets and inhibitors for chikungunya virus envelope glycoproteins

Phuong Nguyen

University of Wollongong, University of Medicine and Pharmacy at Hochiminh City, ptn919@uowmail.edu.au

Haibo Yu

University of Wollongong, hyu@uow.edu.au

Paul A. Keller

University of Wollongong, keller@uow.edu.au

Publication Details

Nguyen, P. T. V., Yu, H. & Keller, P. A. (2018). Molecular docking studies to explore potential binding pockets and inhibitors for chikungunya virus envelope glycoproteins. *Interdisciplinary Sciences: Computational Life Sciences*, 10 (3), 515-524.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Molecular docking studies to explore potential binding pockets and inhibitors for chikungunya virus envelope glycoproteins

Abstract

The chikungunya virus (CHIKV) envelope glycoproteins are considered important potential targets for anti-CHIKV drug discovery due to their crucial roles in virus attachment and virus entry. In this study, using two available crystal structures of the immature and mature forms of envelope glycoproteins, virtual screenings based on blind dockings and focused dockings were carried out to identify potential binding pockets and hit compounds for the virus. The chemical library database of compounds, NCI Diversity Set II, was used in these docking studies. In addition to reproducing previously reported examples, new binding pockets were identified, e.g., Pocket 2 in the 3N40, and Pocket 2 and Pocket 3 in the 3N42. Convergences in conformational sampling in docking using AutoDock Vina were evaluated. An analysis of docking results was carried out to understand interactions of the envelope glycoproteins complexes. Some key residues for interactions, for example Gly91 and His230, are identified as possessing important roles in the fusion process.

Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details

Nguyen, P. T. V., Yu, H. & Keller, P. A. (2018). Molecular docking studies to explore potential binding pockets and inhibitors for chikungunya virus envelope glycoproteins. *Interdisciplinary Sciences: Computational Life Sciences*, 10 (3), 515-524.

Molecular docking studies to explore potential binding pockets and inhibitors for Chikungunya virus envelope glycoproteins

Phuong T. V. Nguyen^{#a,#b}, Haibo Yu^{#a*} and Paul A. Keller^{#a*}

^{#a} School of Chemistry, University of Wollongong, Wollongong, 2522, Australia

^{#b} Faculty of Pharmacy, University of Medicine and Pharmacy at Hochiminh City, 41 Dinh Tien Hoang, District 1, Vietnam

Corresponding author email

hyu@uow.edu.au +61 2 4221 4235 (Tele) +61 2 4221 4287 (Fax) (HY)

keller@uow.edu.au +61 2 4221 4692 (Tele) +61 2 4221 4287 (Fax) (PK)

Abstract

The Chikungunya virus (CHIKV) envelope glycoproteins are considered important potential targets for anti-CHIKV drug discovery due to their crucial roles in virus attachment and virus entry. In this study, using two available crystal structures of the immature and mature forms of envelope glycoproteins, virtual screenings based on blind dockings and focused dockings were carried out to identify potential binding pockets and hit compounds for the virus. The chemical library database of compounds, NCI Diversity Set II, was used in these docking studies. In addition to reproducing previously reported examples, new binding pockets were identified, e.g. Pocket 2 in the 3N40, Pocket 2 and Pocket 3 in the 3N42. Convergences in conformational sampling in docking using AutoDock Vina were evaluated. An analysis of docking results was carried out to understand interactions of the envelope glycoproteins complexes. Some key residues for interactions, for example Gly91 and His230, are identified as possessing important roles in the fusion process.

Keywords: molecular docking, virtual screening, envelope glycoproteins, chikungunya virus, binding pockets, hit compounds.

1. Introduction

Chikungunya virus (CHIKV) is currently a world-wide threat to human health in nearly 55 different countries. The virus was first identified in 1952 [1] and spread from Africa to Asia, the Indian Ocean islands, European countries, Australia, and the Americas [2]. In 2008, it was listed in the Category C priority pathogen group by the National Institute of Allergy and Infectious Diseases due to its morbidity and mortality [1,2]. After bitten by mosquitoes, a patient suffers symptoms such as high fever, headache, rash, vomiting, especially myalgia and polyarthralgia [2]. The acute symptoms can disappear after a few weeks depending on treatment but intense pain in joints can persist for months to years, with lasting effects on patients. Despite efforts over the years, there is currently no vaccines or effective drugs to combat the CHIKV virus.

CHIKV is a member of an Alphavirus genus in the *Togaviridae* family [3]. Like other alphavirus, it is a single-stranded, positive-sense RNA virus with an icosahedral spherical structure (diameter approximately 60-70 nm) that comprises 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 [4]. Additionally, virus particles contain 80 spikes that make glycoprotein shells with 240 copies of each of two glycoproteins, E1 and E2 heterodimers [3]. The CHIKV genome, which is approximately 11.8 kb, is divided into two open reading frames, encoding four non-structural proteins (nsP1, nsP2, nsP3 and nsP4) and five structural proteins - the capsid C, envelope glycoproteins (EG) E1 and E2 and two small peptides (E3 and 6K) [5]. 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 polyprotein by an autoprotease and signalase [4].

The viral entry process is a receptor-mediated endocytosis in clathrin coated vesicles, which is controlled by two viral EG, 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 E3-E2-E1 homotrimers [6]. 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 in the membrane fusion. The E2 contributes to the receptor binding and protects the loop at neutral pH. During replication, the capsid is released and the precursors of structural proteins are processed in the Golgi complex and then moved to the plasma membrane.

Some remarkable studies have focused on structural and functional characterization of the EG as well as the mechanism of neutralization with an interpretation of cryo-electron microscopy structure of the CHIKV-like particles [3,6,7]. 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 [1,8]. 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 [8]. Secondly, Val178 and Ala226 are less conserved residues whose changes do not lose E1 fusion capacity but depend on pH and cholesterol [8]. 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 [9]. Moreover, the study on CHIKV E2 mutants identified the acid-sensitive region in the E2 (amino acids at the 229, 231, 232, 233, and 234 positions) [10]. 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 initiation of viral fusion and entry [10]. Also, the region 229 to 234 was found to be crucial for viral replication partly due to its participation in inducing pH-dependent conformational changes [10]. Additionally, the E2 glycoprotein interacts with the host receptor protein at the 216 residue which is involved in initiating infection [11].

Until now there have been limited studies to identify potential inhibitors for CHIKV drug discovery, and these have focused mostly on the non-structural proteins [12-16]. Envelope glycoproteins are an attractive target through blocking virus entry or virus attachment [4]. Preliminary, computational investigation of the EG structures (the immature and mature forms) [17] has previously been explored [18], however a more rigorous exploration based on blind docking was deemed essential to the full characterization of potential target sites. In virtual screening, blind dockings covering the entire protein, followed by focused dockings to a specific pocket of protein, have been shown to be a robust strategy to detect potential binding pockets, and inhibitors for CHIKV. The convergence of docking in AutoDock Vina was also investigated to get better sampling for each EG. Interaction analysis of hit compounds and these EG provide useful information for CHIKV drug design.

2. Material and Methods

Virtual screening based on blind docking and focused dockings

Blind dockings were carried out based on docking ligands into a protein target using AutoDock Vina (version 1.5.4) [19], with procedures documented in previous studies [18]. In brief, in key steps in the dockings, a protein or a receptor was maintained rigid, while ligands were fully flexible. Proteins were prepared as follows: Two X-ray crystal structures of envelope glycoproteins, an immature form (PDB id: 3N40) and a mature form (PDB id: 3N42), were downloaded from the Protein Data Bank. Energy minimization of the protein was performed using the Accelrys Discovery Studio (DS) 2.0 software with CHARMM force field to relax the structure and remove the steric overlap [20]. The steepest-descent algorithm (3000 steps) was applied. The minimized structure was then prepared for docking using AutoDock Tools (version 1.5.4). Polar hydrogen atoms were added. The database library of ligands from the Open Chemical Repository Collection, National Cancer Institute (NCI) Diversity Set II was used for docking and virtual screening.

The dockings were carried out in two steps. First, a blind docking was undertaken using boxes of size $108 \times 126 \times 62 \text{ \AA}$ and $84 \times 126 \times 74 \text{ \AA}$ for 3N40 and 3N42, respectively, covering the whole proteins. Potential binding sites were revealed by the location of ligands binding to the target. The subsequent focused docking utilized a smaller box ($20 \times 20 \times 20 \text{ \AA}$), centred around the potential binding site of interest. The convergence in sampling or the search exhaustiveness (E) was verified by varying from the default value $E=8$ to $E=16, 32, 128, 256$, and 1024 in docking with AutoDock Vina. Moreover,

identification of potential binding sites was investigated and compared with the previous published results using a receptor cavities tool in Accelrys DS 3.5 [20] and the MetaPocket program [21].

Exploring potential binding sites using other methods

Accelrys DS 3.5 package [20] was applied to explore potential binding sites of the both mature and immature complexes of envelope glycoproteins using receptor cavities tools. Based on a grid search and “eraser” algorithm, the program defines where a binding site is. The binding sites were then displayed as a set of points (point count) and the volume of each of cavity was calculated as the product of the number of site points and the cube of the grid spacing.

Using the MetaPocket program [20], with a protein structure, there are 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-score 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 the protein surface.

Molecular dynamics simulations of some 3N40-NCI_293778 and 3N42-NCI_293778 complexes

In order to further characterize the interactions and stabilities between 3N40 or 3N42 and their potential inhibitors, explicit solvent molecular dynamics simulations were carried out with the NAMD package [22]. The protein atoms were treated with the CHARMM36 force field [23], and the corresponding parameters for the ligands were generated with AmberTools [24]. The systems were solvated in a cubic box of TIP3P water molecules and neutralized by sodium counterions to achieve the physiological ionic concentration of 0.15 M with NaCl. The total number of atoms (including protein, ligand, water, and counterions) in the two systems were 140,000 and 148,000 respectively. All simulations were performed under periodic boundary conditions at a temperature of 298.15 K and a pressure of 1 atm. Temperature coupling was maintained with the Langevin algorithm and the Langevin Piston Nose-Hoover method was used to keep the pressure constant. Van der Waals forces were assigned a cut-off distance of 12.0 Å with a smoothening function between 10.0 and 12.0 Å, while electrostatic interactions were calculated using the Particle Mesh Ewald method [25]. Covalent bonds involving hydrogen had their rigidity maintained by the RATTLE algorithm [26]. The integration time step was set to 1.0 fs, 2.0 fs, and 4.0 fs for bonded, non-bonded, and long rang electrostatic interactions, respectively. The systems were minimized and then equilibrium simulations with weak harmonic restraints on the heavy atoms were performed for 0.5 ns. The production runs were continued for 5.5 ns. The trajectories were saved every 1 ps. The resulting trajectories were analyzed using the CHARMM [23] and VMD 1.9 programs [27].

3. Results and Discussion

Comparison of two CHIKV envelope glycoproteins 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) with their similarities and differences was necessary. In Fig. 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, domain C) and E3. In addition, 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, present in the E1 of the immature 3N40 but not present in the mature 3N42 [17]. This loop becomes disordered in the cleavage process. Additionally, furin maturation of p62 into E3 and E2 will prime the spikes for fusogenic activation during cell entry. Moreover, dissociation of the 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 interests.

Fig 1. Structure of the envelope glycoproteins complexes: (a) The immature structure (PDB id: 3N40); (b) The mature structure (PDB id: 3N42), generated from the file pdb: 3N40 and 3N42, respectively. These structures are similar with the only difference in the furin loop.

Identification of potential binding pockets for two complexes of the envelope glycoproteins

As mentioned in the Methods section, initial, blind dockings used AutoDock Vina to study two structures of the EG, the immature (PDB id: 3N40) and mature forms (PDB id: 3N42) (results in Table S1). An increase in the value of a search exhaustiveness parameter (E) from the default value of E=8 to E=16, 32, 128, 256, and 1024 was used to investigate the sampling convergence in Vina (results in Table S2 and S3 for the two EG). Taken together, there are two binding pockets in total in the immature structure (3N40_Pocket 1 and 3N40_Pocket 2) and four binding pockets (3N42_Pocket 1, 3N42_Pocket 2, 3N42_Pocket 3 and 3N42_Pocket 4) in the mature structure 3N42. 3N40_Pocket 1 and 3N40_Pocket 2 are similar with 3N42_Pocket 1 and 3N42_Pocket2 in terms of locations in their respective structures. Considering the convergence was achieved when the sampling E parameter was set to 16, the following discussions focus on the results obtained with E = 16 (See **Sampling convergence in Autodock Vina**).

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 both the immature and mature structures, Pocket 1 was between E2 domain A, E2 domain C, and E1 domain II which was a favourable place for most of the ligands. Pocket 2 was located at the E2 β -ribbon. Pocket 1 corresponded to the combined Site 1 and Site 2, identified in a previous study [18]. In the mature structure, Pocket 3 was in the region 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 and together with Pocket 3 were two new binding pockets in the 3N42, while Pocket 1 and Pocket 4 have been reported previously [18]. Most of the pockets included residues from both E1 and E2. The locations of the pockets along with the top hit compounds, are illustrated in Fig. 2, and the residues making up each pocket are listed in Table S5.

Other programs such as MetaPocket [20] or using receptor cavities tools in Accelrys Discovery Studio [19], can search for potential binding sites. The location of binding sites identified were qualitatively similar. In particular, the MetaPocket program showed better correlations with blind docking. It resulted in the identification of the four pockets mentioned above, in the EG structures. Additional pockets were also found in two structures such as in the protein 3N40, pockets located between E1 domain II-E3 and between E1 domain I-E1 domain III or pockets found between E2 domain A-E3 and between E2 domain B-E2 domain A in the 3N42 (Table S7-S8). Table 1 compares the binding pockets proposed by the different methods. No potential binding site was detected using MetaPocket. In particular, some sites identified by using Accelrys Discovery Studio were too small to accommodate ligands, so they could not be potential pockets.

Fig. 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 1. 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 the previous study [18].

Receptor cavities tools ^a	Blind dockings		Previous study[18]
	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	Not detected
Behind the fusion loop, between E1 domain II, E2 domain B, E2 domain A	Not detected	Pocket 4	Site 4
Between E1 domain I, E1 domain III, and E2 domain C	Not detected	Pocket 3	Not detected

^aWith Accelrys Discovery Studio program

Identification of potential binding pockets with other methods

Of the binding sites, Pocket 1 was the largest (volume 352 Å³ in the 3N40 and 621 Å³ in the 3N42), accommodating most ligands, while the other pockets were narrower and shallower, illustrated by the volumes in Table S7-S8 for the 3N40 and 3N42 (Pocket 2 in the 3N40 and 3N42 were 123 Å³ and 156 Å³). The Accelrys program also considered small cavities, thereby highlighting some pockets (Table S7-S8) not found by docking of drug molecules for the NCI Diversity Set II. For example, there are two noticeable small sites in the immature structure at the same location of Pocket 3 and Pocket 4 in the mature structure which were not considered as the binding pockets as their volumes were too small to accommodate molecules in the database (Site 3 was 42 Å³ and Site 4 was 26 Å³). Pocket 4 near the fusion loop was large in the mature form (523 Å³), enabling drug molecules to bind easily to this pocket with significant effects, including blocking the relative movement of E2 domains A and B, and freezing the fusion loop by stabilizing interactions. These findings were in accordance with previous results [18].

Identification and analysis of interactions of hit compounds for envelope glycoproteins

The top ten hit compounds and their binding affinities for each pocket were obtained from both blind dockings and focused dockings. These are listed in Table S1, S2, S3 and S4 with their chemical structures in Table S5. The binding affinities of most were between -10 to -13 kcal/mol indicating favorable binding affinities between the ligands and the envelope glycoprotein complexes. The results showed that focused dockings might introduce new hit compounds. Docking analysis revealed some key residues for interactions between both the 3N40 and 3N42 structures and the ligands are listed in Table 2. Most of resulting hits and pockets are hydrophobic, indicating that this is likely an important role in stabilizing the

binding. The hits contained aromatic moieties e.g. benzimidazole, quinoline, quinoxaline, oxazole, and thiophene, which interacted with the aromatic rings of protein residues Tyr, Phe and His in the pockets by π - π stacking or a π -network to stabilize and maintain the interaction of protein and ligands (illustrated in Fig. 3 and 4). For example, in pocket 1 of the immature structure, ligand NCI_61610 formed π - π stacking with His191, ligands NCI_84100_a, NCI_116702, NCI_156219_b and NCI_227186_a interacted with Tyr51 and Tyr301 through π - π stacking or π - π network (Fig. 3). For Pocket 1 and Pocket 4, the key residues in Table 2 were in close agreement with the published findings [18]. 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 bonds. 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.

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 stronger interactions with E1 or E2 through the formation of hydrogen bonds, which may stabilize the E1-E2 heterodimer and prevent dissociation [18]. 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 [8]. In particular, His230 was reported to be important in stabilization of the fusion loop [8].

Table 2. Key residues for interactions in both envelope glycoprotein structures at each of their binding sites.

Binding pocket	Immature structure (3N40)	Mature structure (3N42)
Pocket 1	E1 residues: Tyr51, Thr53, Tyr233, Ser238	E1 residues: Glu50, Tyr51, Lys52, Thr53, Ile55, Tyr242, Ser238
	E2 residues: Arg100, Ile101, Asn103, Pro192, Tyr301	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

Fig. 3. Hydrogen bonding analysis of compounds with the immature structure at its pocket 1 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 orange).

3D molecular model of the active site of the enzyme. The ligand is shown in green sticks, and the residues His230, Tyr233, Tyr51, Glu50, and Phe129 are shown in grey sticks. Dashed green lines indicate hydrogen bonds between the ligand and the residues.

Fig. 4. Hydrogen bonding analysis of compound with the mature structure at its pocket 1: (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 orange).

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 in Table S2 and Table S3, which shows that the exhaustiveness parameter affected the searching process for ligand conformations in docking. Analysis of docking results revealed that the blind dockings with an increase in the exhaustiveness allowed new binding pockets, e.g. in the immature structure, Pocket 2 appeared in docking for E = 16-1024, but was not present for E=8. In the mature structure, 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, for example NCI_293778, NCI_61610, NCI_37553, NCI_156219_b, NCI_84100, NCI_227186_b 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. Hence, E=16 was found as the smallest parameter to achieve the sampling convergence in this case. Sampling convergence was also investigated based on molecular dynamics simulations to take into account protein flexibility and the solvent effects. However, such simulations required significantly more computing-time which is beyond the scope of the current work considering complexes with envelope glycoproteins (a relative large protein requires substantial simulations to ensure a converged sampling).

Stabilities of the 3N40-NCI_293778 and 3N42- NCI_293778 complexes

The stabilities of the docked complexes were subjected to equilibrium molecular dynamics simulations. The room-mean-square deviations (RMSD) for the heavy-atoms of the ligand in the MD sampled structures are fluctuating around 1.5-2.5 Å with respect to the docked complex.

The key interactions (mainly hydrophobic) are preserved throughout the simulations. These indicate that both complexes are stable within the period of molecular dynamics simulation.

Fig 5. (a) The heavy-atom root-mean-square deviations for the ligand from the initial docked structure for the 3N40-NCI_293778 (black) and 3N42-NCI_293778 complex (red). (b) The superposition of the NCI_293778 in complex with 3N40 (Goodsell: the docked complex; Opaque: the structure at 5.5 ns). (c) The superposition of the NCI_293778 in complex with 3N42 (Goodsell: the docked complex; Opaque: the structure at 5.5 ns).

Conclusion

Until now, very few studies have been reported in the literature about Chikungunya inhibitor design targeting envelope glycoproteins. Particularly, little information is known about the potential binding pockets and the possible binding modes, which our current work is aiming to address. Taking advantage of the available envelope glycoproteins structures, virtual screening based on blind dockings and focused dockings explored the potential binding pockets and inhibitors for both the immature and mature of envelope glycoproteins. Promising hit compounds were identified for two complexes of the envelope glycoproteins. 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 [18]. Some new hits were obtained from focused docking which showed a good combination of blind dockings and focused dockings in molecular approach. The key residues involved in stabilizing the complex or participating in the fusion process were identified. These interactions of hits obtained from docking results with the protein functions of these glycoproteins provided a promising approach for drug discovery and drug design to block the virus entry and virus attachment by targeting these proteins. This

study also supported the current docking protocol utilising Autodock Vina as a robust strategy with sufficient accuracy and to identify potential inhibitors and understand the binding modes. Both computational and experimental verifications with *in vivo* assays are currently either planned or in progress to test the inhibitors reported here and those targeting nsP2 [16] and nsP3 [13].

Acknowledgements

PTVN was supported by a UOW-Vietnamese Government Scholarship (VIED-MOET). H.Y. is the recipient of an Australian Research Council Future Fellowship (Project number FT110100034). This research was undertaken with the assistance of resources provided at the University of Wollongong High Performance Computing Cluster.

References

1. Rashad AA, Mahalingam S, Keller PA (2014) Chikungunya Virus: Emerging Targets and New Opportunities for Medicinal Chemistry. *J Med Chem* 57:1147-1166. doi:10.1021/jm400460d
2. Thiboutot M, Kannan S, Kawalekar O, Shedlock D, Khan A, Sarangan G, Srikanth P, Weiner D, K. M (2010) Chikungunya: a potentially emerging epidemic? *Plos Negl Trop Dis* 4 (4):1-8. doi:10.1371/journal.pntd.0000623
3. Gould EA, 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 (2010) Understanding the alphaviruses: Recent research on important emerging pathogens and progress towards their control. *Antiviral Res* 87 (2):111-124. doi:10.1016/j.antiviral.2009.07.007
4. Sun S, Xiang Y, Akahata W, Holdaway H, Pal P, Zhang X, Diamond MS, Nabel GJ, Rossmann MG (2013) Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *eLife* 2 (2):e00435. doi:10.7554/eLife.00435
5. Khan AH, Morita K, Parquet MdC, Hasebe F, Mathenge EGM, Igarashi A (2002) Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. *J Gen Virol* 83 (12):3075-3084
6. Kielian M, Chanel-Vos C, Liao M (2010) Alphavirus entry and membrane fusion. *Viruses* 2 (4):796-825. doi:10.3390/v2040796
7. Li L, Jose J, Xiang Y, Kuhn RJ, Rossmann MG (2010) Structural changes of envelope proteins during alphavirus fusion. *Nature* 468 (7324):705-708. doi:10.1038/nature09546
8. Kuo S-C, Chen Y-J, Wang Y-M, Tsui P-Y, Kuo M-D, Wu T-Y, Lo SJ (2012) Cell-based analysis of Chikungunya virus E1 protein in membrane fusion. *J Biomed Sci* 19 (1):44-44. doi:10.1186/1423-0127-19-44
9. Sudeep AB, Parashar D (2008) Chikungunya: an overview. *J BioSci* 33 (4):443
10. Akahata W, Nabel GJ (2012) A specific domain of the Chikungunya virus E2 protein regulates particle formation in human cells: implications for alphavirus vaccine design. *J Virol* 86 (16):8879-8883. doi:10.1128/jvi.00370-12
11. Asnet Mary J, Paramasivan R, Tyagi BK, Surender M, Shenbagarathai R (2013) Identification of structural motifs in the E2 glycoprotein of Chikungunya involved in virus-host interaction. *J Biomol Struct Dyn* 31 (10):1077-1085. doi:10.1080/07391102.2012.721496
12. Rungrotmongkol T, Nunthaboot N, Malaisree M, Kaiyawet N, Yotmanee P, Meeprasert A, Hannongbua S (2010) 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* 29 (3):347-353. doi:10.1016/j.jmgm.2010.09.010
13. Nguyen PTV, Yu H, Keller PA (2014) Discovery of in silico hits targeting the nsP3 macro domain of chikungunya virus. *Journal of molecular modeling* 20 (5):1-12. doi:10.1007/s00894-014-2216-6

14. Singh KD, Kirubakaran P, Nagarajan S, Sakthiah S, Muthusamy K, Velmurgan D, Jeyakanthan J (2011) Homology modeling, molecular dynamics, e-pharmacophore mapping and docking study of Chikungunya virus nsP2 protease. *J Mol Model* 18 (1):39-51. doi:10.1007/s00894-011-1018-3
15. Bassetto M, Silvestri R, Tron GC, Neyts J, Leyssen P, Brancale A, De Burghgraeve T, Delang L, Massarotti A, Coluccia A, Zonta N, Gatti V, Colombano G, Sorba G (2013) Computer-aided identification, design and synthesis of a novel series of compounds with selective antiviral activity against chikungunya virus. *Antiviral Res* 98 (1):12-18. doi:10.1016/j.antiviral.2013.01.002
16. Nguyen TVP, Yu H, Keller PA (2015) Identification of chikungunya virus nsP2 protease inhibitors using structure-based approaches. *J Mol Graph Model* 57:1-8
17. Voss J, Vaney MC, Duquerroy S, Vonnrhein C, Ginard-Blanc C, Crublet E, Thompson A, Bricogne G, Rey FA (2010) Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 468 (7324):709-712. doi:10.1038/nature09555
18. Rashad AA, Keller PA (2013) Structure based design towards the identification of novel binding sites and inhibitors for the chikungunya virus envelope proteins. *J Mol Graph Model* 44:241-252. doi:10.1016/j.jmgm.2013.07.001
19. Trott O, Olson AJ (2010) 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* 31 (2):455-461. doi:10.1002/jcc.21334
20. Accelrys Software Inc (2013) Discovery Studio Modeling Environment. 4.0 edn. Accelrys Software Inc., San Diego
21. Huang B (2009) MetaPocket: a meta approach to improve protein ligand binding site prediction. *Omics* 13 (4):325-330. doi:10.1089/omi.2009.0045
22. Phillips JC, Schulten K, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26 (16):1781-1802. doi:10.1002/jcc.20289
23. Huang J, MacKerell AD (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J Comput Chem* 34 (25):2135-2145. doi:10.1002/jcc.23354
24. Case DA, Woods RJ, Cheatham rTE, Darden T, Gohlke H, Luo R, Merz JKM, Onufriev A, Simmerling C, Wang B (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26 (16):1668-1688. doi:10.1002/jcc.20290
25. Ulrich E, Lalith P, Max LB, Tom D, Hsing L, Lee GP (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103 (19):8577-8593. doi:10.1063/1.470117
26. Miyamoto S, Kollman PA (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* 13 (8):952-962. doi:10.1002/jcc.540130805
27. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph Modell* 14 (1):33-38. doi:10.1016/0263-7855(96)00018-5