



UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

University of Wollongong  
**Research Online**

---

Faculty of Science, Medicine and Health - Papers

Faculty of Science, Medicine and Health

---

2013

# Structure based design towards the identification of novel binding sites and inhibitors for the chikungunya virus envelope proteins

Adel A. Rashad

*University of Wollongong*, ar917@uowmail.edu.au

Paul A. Keller

*University of Wollongong*, keller@uow.edu.au

---

## Publication Details

Rashad, A. A. & Keller, P. A. (2013). Structure based design towards the identification of novel binding sites and inhibitors for the chikungunya virus envelope proteins. *Journal of Molecular Graphics and Modelling*, 44 241-252.

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

---

# Structure based design towards the identification of novel binding sites and inhibitors for the chikungunya virus envelope proteins

## Abstract

Chikungunya virus is an emerging arbovirus that is widespread in tropical regions and is spreading quickly to temperate climates with recent epidemics in Africa, Asia, Europe and the Americas. It is having an increasingly major impact on humans with potentially life-threatening and debilitating arthritis. Thus far, neither vaccines nor medications are available to treat or control the virus and therefore, the development of medicinal chemistry is a vital and immediate issue that needs to be addressed. The viral envelope proteins play a major role during infection through mediation of binding and fusion with the infected cell surfaces. The possible binding target sites of the chikungunya virus envelope proteins have not previously been investigated; we describe here for the first time the identification of novel sites for potential binding on the chikungunya glycoprotein complexes and the identification of possible antagonists for these sites through virtual screening using two successive docking scores; FRED docking for fast precise screening, with the top hits then subjected to a ranking scoring using the AUTODOCK algorithm. Both the immature and the mature forms of the chikungunya envelope proteins were included in the study to increase the probability of finding positive and reliable hits. Some small molecules have been identified as good in silico chikungunya virus envelope proteins inhibitors and these could be good templates for drug design targeting this virus.

## Keywords

chikungunya, virus, envelope, proteins, towards, identification, novel, binding, sites, structure, design, inhibitors, CMMB

## Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

## Publication Details

Rashad, A. A. & Keller, P. A. (2013). Structure based design towards the identification of novel binding sites and inhibitors for the chikungunya virus envelope proteins. *Journal of Molecular Graphics and Modelling*, 44 241-252.

# Structure Based Design towards the Identification of Novel Binding Sites and Inhibitors for the Chikungunya Virus Envelop Proteins

Adel A. Rashad and Paul A. Keller\*

Centre for Medicinal Chemistry, School of Chemistry, University of Wollongong, Wollongong, Australia, 2522

\* Correspondence to:

Paul A. Keller:

keller@uow.edu.au

Tele: +61 2 4221 4692

Fax: +61 2 4221 4287

## ABSTRACT

Chikungunya virus is an emerging arbovirus that is widespread in tropical regions and is spreading quickly to temperate climates with recent epidemics in Africa, Asia, Europe and the Americas. It is having an increasingly major impact on humans with potentially life-threatening and debilitating arthritis. Thus far, neither vaccines nor medications are available to treat or control the virus and therefore, the development of medicinal chemistry is a vital and immediate issue that needs to be addressed. The viral envelope proteins play a major role during infection through mediation of binding and fusion with the infected cell surfaces. The possible binding target sites of the chikungunya virus envelope proteins have not previously been investigated; we describe here for the first time the identification of novel sites for potential binding on the chikungunya glycoprotein complexes and the identification of possible antagonists for these sites through virtual screening using two successive docking scores; FRED docking for fast precise screening, with the top hits then subjected to a ranking scoring using the AUTODOCK algorithm. Both the immature and the mature forms of the chikungunya envelope proteins were included in the study to increase the probability of finding positive and reliable hits. Some small molecules have been identified as good *in silico* chikungunya virus envelope proteins inhibitors and these could be good templates for drug design targeting this virus.

**Keywords** Alphaviruses; Chikungunya virus; Envelope proteins; Virtual screening

## 1. Introduction

Chikungunya virus (CHIKV) is an emerging mosquito-borne arthrogenic member of the *alphavirus* genus (family *Togaviridae*) that has caused widespread outbreaks of debilitating human disease in the past five years [1]. Chikungunya fever (CHIKF) caused by the virus was first described in 1952 [2], and currently has been identified in

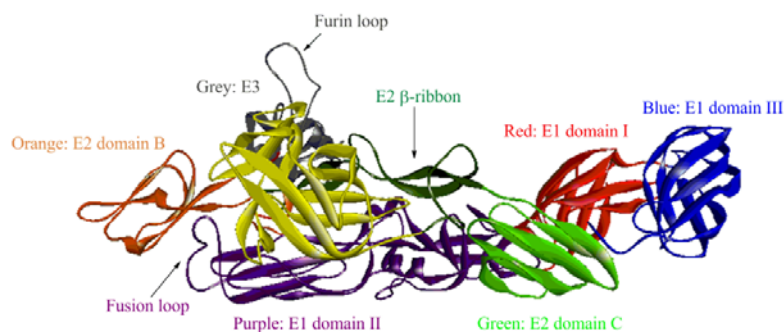
nearly 40 countries. In 2008 it was listed as a US National Institute of Allergy and Infectious Diseases (NIAID) category C priority pathogen because of the high morbidity and mortality rates and major health impact [3, 4].

The symptoms of chikungunya fever infection generally start 4–7 days after the mosquito bite. Infection usually presents in two phases; the first is acute, while the second stage is persistent (chronic), causing disabling polyarthritis [5]. Acute infection lasts 1–10 days and is characterized by a painful polyarthralgia, high fever, asthenia (weakness), headache, vomiting, rash, and myalgia [6]. The persistent chronic stage of CHIKF is characterized by polyarthralgia that can last from weeks to years beyond the acute stage [7]. Neurological disorders including encephalitis, myelopathy, peripheral neuropathy, myeloneuropathy and myopathy have also been reported [8].

The CHIKV genome is approximately 11.8 Kb in size and consists of a single stranded, positive sense RNA genome with two open reading frames (ORFs) [9], one in the 5' end which encodes two polyproteins, the precursors of the non-structural proteins. The second ORF at the 3' end encodes the structural proteins, the capsid (C), envelope glycoproteins E1 and E2 and two small cleavage products (E3, 6K). Similar to other members of the *alphaviruses*, the CHIKV starts the life cycle by entering the target host cells by pH dependent endocytosis via a receptor mediated interaction [10]. A recent study identified prohibitin1 (PHB1) as a microglial cell expressed CHIKV binding protein [11].

After entering the cell, the endosome acidic environment triggers conformational changes in the viral envelope complex made of E1 and E2 proteins, resulting in dissociation of the E2-E1 heterodimers, and the formation of E1 homotrimers. The E1 trimer inserts into the target cell membrane via its hydrophobic fusion peptide (fusion loop) and refolds to form a hairpin-like structure. Exposure of the E1 fusion peptide leads to releasing of the nucleocapsid into the host cell cytoplasm [12],[13]. During the replication cycle inside the host cell, the capsid protein is released, and the pE2 and E1 glycoproteins are translated in the Golgi and are moved to the plasma membrane, where pE2 is cleaved by furin-like protease activity into E2 and E3 [14].

Glycoprotein E2 is responsible for receptor binding whereas E1 is responsible for membrane fusion [4]. E3 contains the 64-amino-terminal residues of p62 and mediates the correct folding of pE2 and its subsequent association with E1 [15]. E3 also protects the E2-E1 heterodimer from premature fusion with cellular membranes [16]. Furin maturation of p62 into E3 and E2 during transport to the cell surface primes the spikes for subsequent fusogenic activation for cell entry. Mature virions bud at the plasma membrane via interactions between E2 and genome-containing viral nucleocapsids present in the cytoplasm [17], ready for infecting new cells. The crystal structures of both the immature and the mature glycoprotein complexes have recently been solved [17], (Fig. 1).



**Fig. 1** Crystal structure of the immature envelope glycoprotein complex of Chikungunya. Generated from the pdb file: 3N40 [17].

E1 is folded into three  $\beta$ -sheet rich domains (I, II and III). E2 is an all  $\beta$  protein belonging to the immunoglobulin superfamily, with three domains A, B and C. Domain B is at the membrane upper end and domain C is towards the viral membrane: Domain A is at the centre while domain C binds to the adjacent domain II of E1. The long  $\beta$  – ribbon of E2 makes most of the connection with E3. Furin loop is E2-E3 junction in the immature complex; this junction contains a functional proprotein convertase motif which is cleaved by the cellular proteases; furin-like proprotein convertases during the maturation of the glycoproteins [14]. The amino acid His60 in this junction is the critical residue that determines the spectrum of furin and furin-like convertases that process E2-E3 glycoprotein complex [18]. The U shaped fusion loop of E1 is inserted in a groove between E2 domains A and B being stabilized by hydrogen bonds with E2 histidine side chains [17]. In the neutral pH, E3 maintains E2 domain B in an orientation with respect to domain A in such way that it creates the groove accommodating the E1 fusion loop, protecting the virus from premature fusion with other cellular membranes [17],[19]. Some residues in domain B of E2 are believed to be associated with cell recognition [16]. The fusogenic activity of the E1 fusion peptide is highly dependent on pH change. The histidine residues of E2 are believed to be involved as the pH sensor for the activation of the fusion protein at lower pH [17] due to the increased probability of histidines to become positively charged at lower pH values, based on the fact that the imidazole ring of the histidine residue is the only amino acid side chain whose apparent dissociation constant from protons (pKa) falls within the physiological range. Within the E1 fusion peptide sequence, the glycine residue (Gly91) is critical for the fusion process. Also, it was found that one histidine residue at E1 230, which is located outside of the fusion sequence, is also critical for the fusion [20].

During the chikungunya fever, some limited symptomatic treatments including corticosteroids may be used in cases of debilitating chronic CHIKV infection [21],[22], and only in the last 24 months have efforts for development of therapeutics been reported such as arbidol [23], mycophenolic acid [24], daphnane-type diterpenoids [25], harringtonine [26], purines and  $\beta$ -lactams based inhibitors [27], and the immunostimulant polycytidylic acid [Poly

(I:C)] [28]. Despite all these efforts, neither a selective antiviral drug nor a vaccine has been approved for use thus far. Further, no directed drug design programs have been implemented using structure-based design principles and the newly reported CHIKV protein X-ray crystal structures.

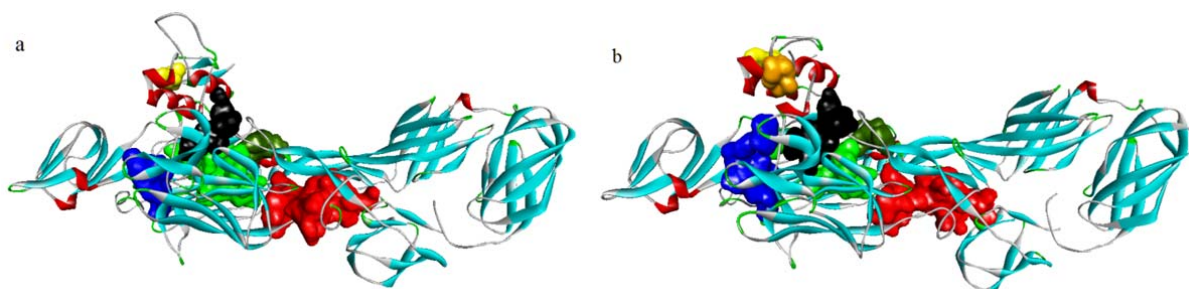
Blocking the *in vitro* CHIKV infection in the host cells targeting the envelope proteins has been demonstrated by blocking the intracellular furin-mediated cleavage of viral envelope glycoproteins (E2E3 or p62 precursors). This blocking was achieved by an irreversible furin-inhibiting peptide which significantly reduce the processing of E3E2 CHIKV glycoproteins. This led to the formation of immature viral particles and impaired viral spreading through other uninfected cells [18]. This reflects the importance of considering the envelope glycoproteins as an attractive target for selective drug development.

The usage of the three-dimensional structure of the target proteins (crystal structures) in the virtual screening (*in silico*) of chemical libraries has been a powerful approach to identify lead compounds with some successful examples in a number of systems [29, 30]. Such structure based drug design techniques, including the identification of new binding sites and virtual screening searches, have been successfully used for the identification of lead compounds for the dengue virus envelop protein (E protein) [31, 32]. Dengue virus is also an arbovirus and is transmitted by the same vector mosquito of the CHIKV. Herein, we report for the first time the novel binding sites in the CHIKV envelope glycoproteins that can be used as sites for inhibitors that could alter the function of the envelope proteins and consequently, inhibit the virus fusion function. To increase the chances of possible hits, we examined both the immature and the mature glycoprotein crystal structures for possible binding sites. Two sites were chosen that were common in both the immature and the mature proteins based on their locations and functions. We then used virtual screening combining two different docking algorithms with a number of chemical databases to identify suitable compounds predicted to bind in these sites. FRED (fast rigid exhaustive docking) was used for fast and precise screening using multiple scoring functions, followed by a re-docking ranking of the top hits using AUTODOCK scoring function. This led to the identification of favoured hits that have suitable binding profiles to the CHIKV glycoproteins. This hypothesis represents a new strategy for inhibiting this particular virus by targeting the envelope proteins which will lead to impaired protein function and thus inhibiting the virus, and will help the further synthetic development and optimization of selective inhibitors, as previously and successfully achieved for the dengue virus envelop protein inhibitors [32, 33].

## **2. Materials and methods**

### **2.1. Identification of novel binding sites**

Both the crystal structure of the immature complex (pdb file: 3N40 [17]) and the mature complex (pdb file: 3N42 [17]) were used. Binding sites within the receptors were detected using the Discovery Studio 3.5 software (Accelrys Software Inc.: San Diego, CA, 2012). The algorithm is based on a grid search and "eraser" algorithm which derives binding sites from cavities in the structure of the receptor. The binding site found is displayed as a set of points. The volume of each cavity is defined as the product of number of site points and the cube of the grid spacing. Six main sites were detected in both the immature and the mature crystal structures and only one site were detected in the mature crystal structure that is not present in the immature form (Fig. 2). Table 1 shows the identified sites with their characters. Suitable cavities were then checked further based on functionality, presence of hydrophobic residues, presence of charged residues and solvent accessibility.



**Fig. 2** Crystal structure of (a) the immature glycoproteins (generated from file pdb: 3N40) and (b) the mature glycoproteins (generated from the file pdb: 3N42) showing the identified binding cavities as solid filled surface.

Table 1 The identified receptor cavities in the immature (3N40) and mature (3N42) crystal structures, grid coordinates x, y and z, cavity volumes, points count and location for each cavity site.

		3N40	3N42	Location
		(immature structure)	(mature structure)	
Site 1 (red)	x, y, z	-15.381, -1.269, 16.434	-15.687, 2.019, -19.939	Between E1
	Volume	687.25	651.375	domain II and E2
	Points count	5498	5211	domain C
Site 2 (light green)	x, y, z	-30.631, 17.481, 33.684	-33.937, -18.731, -31.939	Between E1
	Volume	395	357.375	domain II and the
	Points count	3160	2859	β-ribbon of E2
Site 3 (dark green)	x, y, z	-30.631, 4.981, 37.934	-33.437, -6.731, -33.189	Adjacent to site 2
	Volume	157.625	156.125	

	Points count	1261	1249	
<b>Site 4 (blue)</b>	x, y, z	-38.131, 31.481, 24.934	-42.937, -28.731, -22.939	Behind the fusion
	Volume	126.25	183.875	loop, between E3,
	Points count	1010	1471	E2 domain B, E2 domain A
<b>Site 5 (black)</b>	x, y, z	-44.631, 14.731, 23.184	-44.437, -14.731, -23.439	between the $\beta$ -
	Volume	93.125	124	ribbon of E2 and
	Points count	745	992	E3
<b>Site 6 (yellow)</b>	x, y, z	-57.631, 16.731, 36.184	-16.187, -18.231, -36.439	Within E3 cavity
	Volume	29.5	20.5	
	Points count	236	164	
<b>Site 7 (orange)</b>	x, y, z		-59.187, -15.731, -26.189	Replacing the furin
	Volume	Does not exist	22.25	loop
	Points count		178	

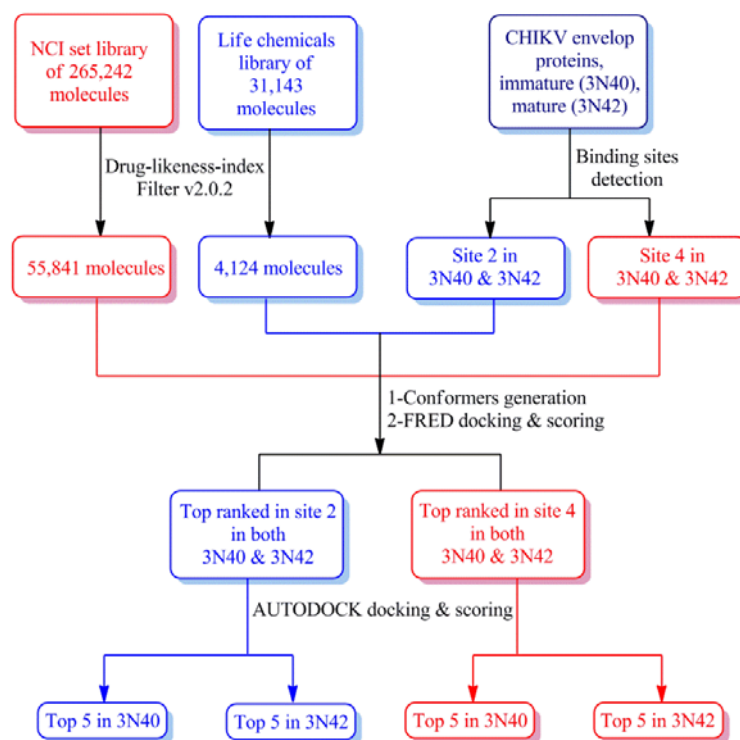
## 2.2. Virtual screening with the CHIKV envelope proteins

Two chemical compounds libraries were used; The NCI set library of 265,242 compounds and the Life chemicals protein-protein interactions inhibitors library of 31,143 compounds. The databases were filtered with the drug-likeness-index and a limited range for the Molecular Weight  $\leq 500$ , calculated octanol-water partition coefficient ( $\text{clogP} \leq 5$ ), and hydrogen bond donors, and acceptors (OH's and NH's  $\leq 5$ ; N's and O's  $\leq 10$ ) [34], using Filter v2.0.2 (OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>), producing 55,841 compounds from the NCI library and 4,124 compounds from the Life Chemicals library. Fast exhaustive virtual screening was performed using FRED v2.2.5 (OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>), FRED is a fast and effective docking application whose performance is significantly more reliable, i.e. lower variance, than most other programs [35, 36]. FRED performs a systematic, exhaustive, non-stochastic examination of all possible poses within the protein active site, filters for shape complementarity [37] and pharmacophoric features before selecting and optimizing poses using the Chemgauss scoring function. Omega2 (Systematic high-throughput conformer generation, OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>) [38, 39], was used to generate multiple conformers for each compound in the database libraries using the default settings. Omega2 takes into account the flexibility of a molecule by

generating all representative conformers. For the NCI library, 2,312,012 conformers were generated, and 334,064 conformers were generated from the Life Chemicals compounds. The work-flow diagram is shown in Fig. 3. The life chemical library was screened on site 2 (light green colour in Fig. 2) in both the immature and the mature glycoproteins. The NCI set compounds were screened on site 4 (blue colour in Fig. 2) of the two envelope protein forms. The binding sites were prepared for docking using Fred receptor setup software (OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>). The grid boxes were determined based on the x, y and z co-ordinates given in Table 1. For site 2 in the 3N40 receptor, the box size was set to 6153 Å<sup>3</sup> and was assigned an inner contour of 99 Å<sup>3</sup> and an outer contour of 1886 Å<sup>3</sup>. Site 4 in the 3N40 receptor has a box size of 6580 Å<sup>3</sup> and was assigned an inner contour of 116 Å<sup>3</sup> and an outer contour of 1071 Å<sup>3</sup>. Site 2 in the 3N42 receptor has a box size of 7578 Å<sup>3</sup> and was assigned an inner contour of 66 Å<sup>3</sup> and an outer contour of 1816 Å<sup>3</sup>. Site 4 in the 3N42 receptor was assigned a box size of 6482 Å<sup>3</sup>, an inner contour of 45 Å<sup>3</sup> and an outer contour of 1547 Å<sup>3</sup>. No constraints were enabled in any of the prepared receptors. During the docking calculations, both chemgauss3 and shapegauss scoring functions were enabled. After the docking calculations, the poses returned were scored and ranked with a Gaussian shape function independently by the five available scoring functions (PLP, Chemgauss3, Chemscore, OEChemscore, and Screenscore) and by a consensus of all. The top ranked poses from the exhaustive docking were then optimized using systematic solid body optimization by chemgauss3. VIDA v4.2.0 (OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>) was used to visualise the docked poses within the receptor active site, and to inspect the critical interacting residues in each pocket with the individual docked poses. Top 20 hits were then recorded for each of the four sites (data not shown).

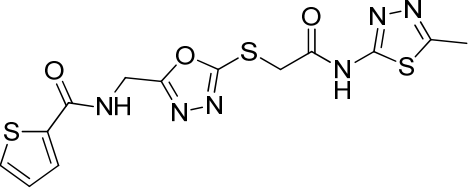
The top 20 docked poses ranked in each of the four binding sites were then extracted as pdb files, and were processed with AutoDock Tools 1.5.6rc3 (ADT) graphical interface [40]. The Gasteiger charges were calculated and the nonpolar hydrogen atoms were merged, torsion angles were defined, they were then saved as pdbqt files for Autodock calculations. Crystal structures (3N40, 3N42) were used by AutoDock Tools 1.5.6rc3 to setup the receptor binding sites. The grid box co-ordinates in each site were determined based on the co-ordinates in Table 1. The grid box size was set to 46 x 46 x 46 points in x, y, and a z direction in each of the four sites and a grid spacing of 0.375 Å was used. AutoGrid 4.2 algorithm was used to evaluate the binding energies between the inhibitors and the enzyme and to generate the energy maps for the docking run. Fifty runs were generated by using Autodock 4.2 Lamarckian genetic algorithm [40] for the searches. Cluster analysis was performed on docked results, with a root-mean-square tolerance of 2.0 Å, the docked poses were ranked according to the binding energies and ligand efficiencies, and finally the five lowest energy poses (Tables 2-5) were selected as the resultant complexes with the **proteins**. The complexes were then

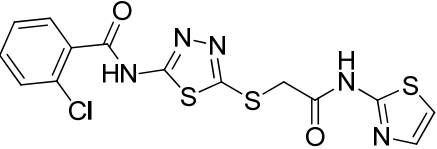
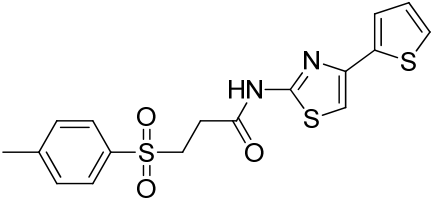
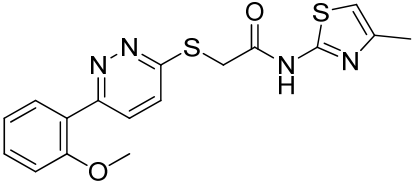
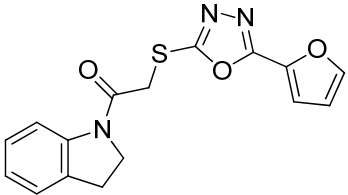
typed with the CHARMM forcefield with Discovery Studio 3.5 software (Accelrys Software Inc.: San Diego, CA, 2012) to relax the obtained poses within the protein pockets, and visualized. Fred and Autodock are powerful tools for the preliminary identification of hits [41], and have previously been used together successfully for the *in silico* identification of potential inhibitors [42]. Compounds are commercially available and have drug like qualities and also can be accessed through chemical syntheses for further optimization process.



**Fig. 3** Work-flow diagram of the virtual screening procedures used for CHIKV envelope proteins.

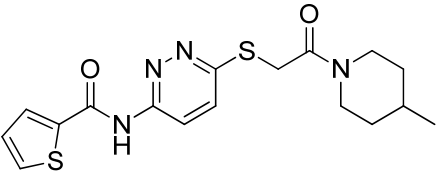
Table 2 Top 5 hits identified for site 2 using the *immature glycoprotein receptor* (3N40), showing the molecular weights, calculated logP (clogP), predicted binding energies, inhibitory constants ( $K_i$ ) and the interaction residues.

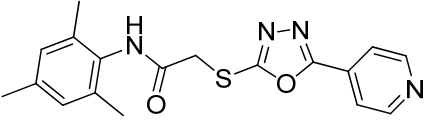
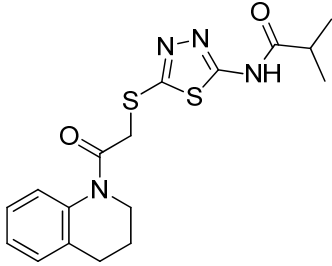
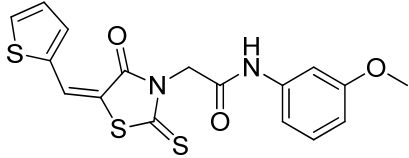
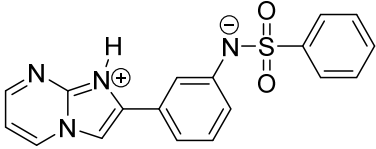
Compound	Molecular weight	clogP*	Binding energy Kcal/mol	Predicted $K_i$ (nM)	Interaction residues
1 	396.47	0.31±0.89	-10.06	42.15	E1 Lys52 E1 Ile55 E1 Thr53 E2 Tyr301

						E2 Arg100
						E1 Lys52
						E1 Ile55
2		411.91	3.43±0.66	-9.43	121.87	E1 Thr53
						E2 Tyr301
						E2 Arg100
3		392.52	3.60±0.43	-9.36	138.19	E1 Lys52
						E1 Ile55
						E2 Tyr301
						E2 Glu232
4		372.46	2.93±0.44	-9.18	187.59	E1 Lys52
						E1 Ile55
						E2 Tyr301
						E2 Glu232
						E2 Arg100
5		327.36	3.47±0.71	-8.99	255.48	E1 Lys52
						E1 Ile55
						E2 Tyr301
						E2 Arg100

\*Calculated using ACDLabs v.12.0 (ACD/Labs, Toronto, Canada).

Table 3 Top 5 hits identified for site 2 using the *mature glycoprotein receptor* (3N42), showing the molecular weights, calculated logP (clogP), predicted binding energies, inhibitory constants ( $K_i$ ) and the interaction residues.

	Compound	Molecular weight	clogP*	Binding energy <i>Kcal/mol</i>	Predicted $K_i$ (nM)	Interaction residues <sup>#</sup>
1		376.50	2.37±0.60	-9.98	48.38	E1 Lys52
						E1 Thr53
						E1 Ile55
						E2 Arg36

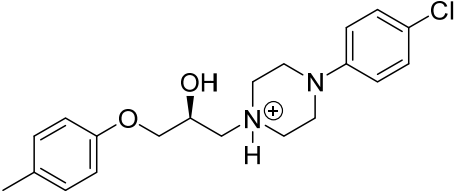
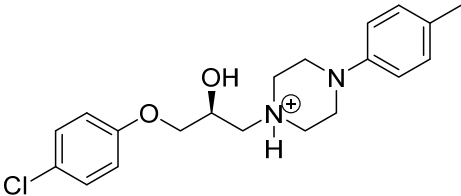
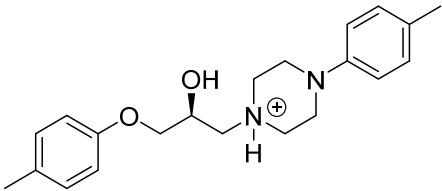
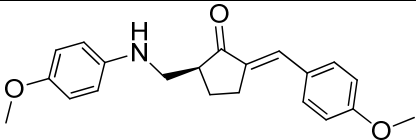
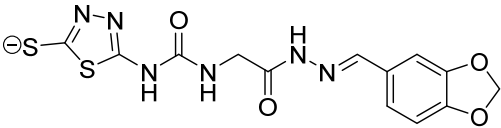
						E2 Glu168
						E1 Lys52
						E1 Ile55
2		354.43	4.12±0.64	-9.71	75.78	E2 Arg36
						E2 Glu168
						E2 Tyr237
						E1 Lys52
						E1 Ile55
3		376.50	3.90±0.72	-9.36	138.07	E2 Tyr237
						E1 Lys52
						E1 Thr53
4		390.50	3.03±0.75	-9.26	163.4	E1 Ile55
						E2 Arg36
						E2 Tyr237
						E1 Lys52
						E1 Thr53
5		350.39	2.83±0.90	-9.17	190.7	E1 Ile55
						E2 Arg36
						E2 Tyr237

\*Calculated using ACDLabs v.12.0 (ACD/Labs, Toronto, Canada).

#Numbers of E2 residues in the mature form are different than the corresponding residues in the immature form.

Table 4 Top 5 hits identified for site 4 using the *immature glycoprotein receptor* (3N40), showing the molecular weights, calculated logP (clogP), predicted binding energies, inhibitory constants ( $K_i$ ) and the interaction residues.

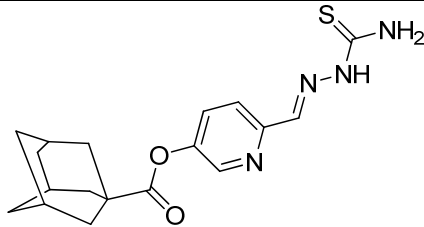
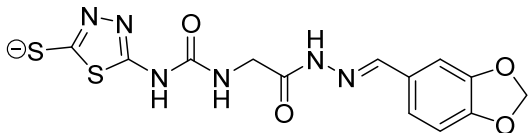
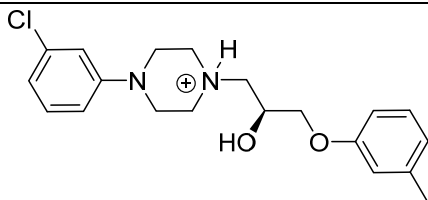
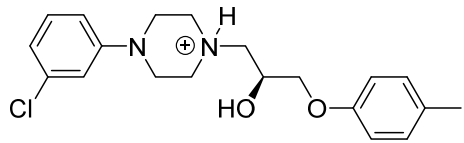
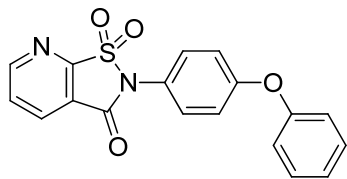
Compound	Molecular weight	clogP*	Binding energy <i>Kcal/mol</i>	Predicted $K_i$ (nM)	Interaction residues
----------	------------------	--------	-----------------------------------	-------------------------	----------------------

1		361.89	4.76±0.54	-11.30	5.18	E1 Val229
						E2 His82
						E2 His93
						E2 Leu80
						E2 Leu305
2		361.89	4.71±0.48	-11.23	5.91	E1 Val229
						E2 His82
						E2 His93
						E2 Leu80
						E2 Leu305
3		341.47	4.49±0.47	-11.20	6.19	E1 Val229
						E2 His82
						E2 His93
						E2 Leu80
						E2 Leu305
4		337.41	3.42±0.34	-10.69	14.49	E2 His82
						E2 His93
						E2 Leu80
5		379.39	1.83±0.90	-10.45	21.98	E1 Phe87
						E2 His82
						E2 His93
						E2 Ser91
						E2 Leu80
						E2 Leu305

\*Calculated using ACDLabs v.12.0 (ACD/Labs, Toronto, Canada).

Table 5 Top 5 hits identified for site 4 using the *mature glycoprotein receptor* (3N42), showing the molecular weights, calculated logP (clogP), predicted binding energies, inhibitory constants ( $K_i$ ) and the interaction residues.

Compound	Molecular weight	clogP*	Binding energy	Predicted $K_i$ (nM)	Interaction residues <sup>#</sup>
----------	------------------	--------	----------------	----------------------	-----------------------------------

Kcal/mol						
1		358.46	3.64±0.39	-10.91	10.03	E1 Thr228
						E1 Gly229
						E2 His18
						E2 His29
2		379.39	1.83±0.90	-10.25	30.49	E1 Phe87
						E2 His18
						E2 His29
						E2 Ser27
						E2 Leu16
						E2 leu241
3		361.89	4.98±0.48	-10.00	46.61	E1 Val229
						E2 His18
						E2 His29
4		361.89	4.98±0.48	-9.98	48.35	E1 Val229
						E2 His18
						E2 His29
5		352.36	2.47±1.22	-9.88	57.61	E1 Trp89
						E2 His18
						E2 His29
						E2 Leu16

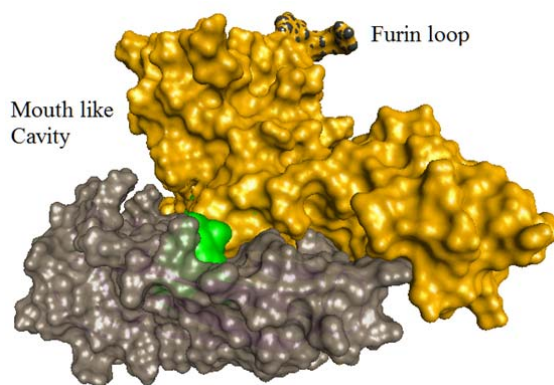
\*Calculated using ACDLabs v.12.0 (ACD/Labs, Toronto, Canada).

#Numbers of E2 residues in the mature form are different than the corresponding residues in the immature form.

### 3. Results and discussion

The essential role of the CHIKV envelope protein in the fusion process, its location on the surface of the mature virus (spikes) and the availability of the crystal structures make it a suitable target for structure-based drug design. The CHIKV glycoprotein exists in two forms, the immature form and the mature form. The immature form represents the

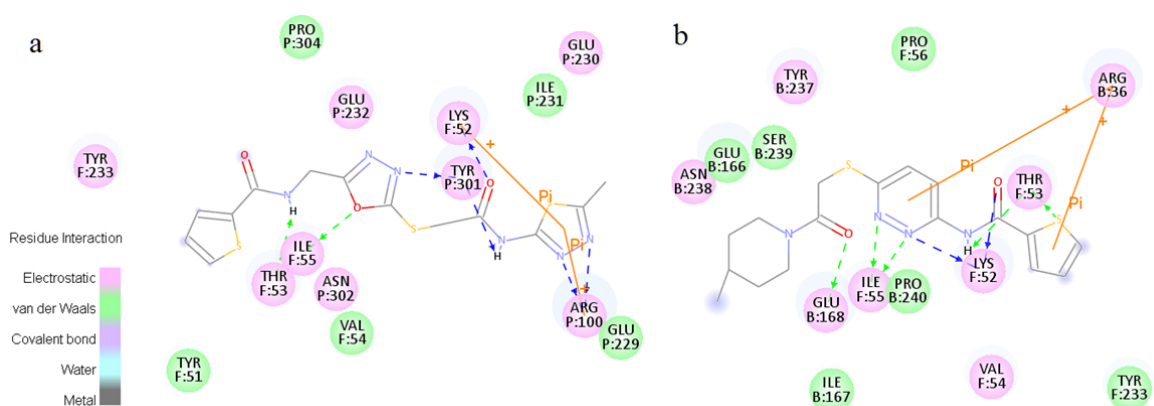
early stages of the envelope protein after the replication cycle, translated in the endoplasmic reticulum and processed in the Golgi for maturation, moved to the plasma membrane, where it is cleaved by furin-like protease activity in the host infected cell into E2 and E3 [14]. The furin cleavage occurs at the furin loop, which represents the junction between E2-E3. The difference between the two crystal structures is the removal of the furin susceptible peptide motif which results in slight changes in the volumes of the predicted binding sites. We searched for possible binding sites within both the immature and the mature crystal structures. Six common sites were detected in both structures (Table 1). Among the detected sites, site 2 (light green, Fig. 2) and site 4 (blue, Fig. 2) were interesting. Site 2 represents a surface cavity that lies between the E1 domain II and E2  $\beta$ -ribbon that connects E2 domain A to E2 domain C; also extends downwards as a channel between E1 domain II and E2 domain A. E2 domains A and B move relative to each other in the pre- and post-fusion structures. Therefore, small molecules that bind to this site may stabilize the E1-E2 heterodimer and prevent their dissociation during the fusion process. An additional hypothesis is that it may also stabilize the orientation of E2 domain A with respect to domain B in a way that inhibits the exposure of the fusion peptide in conditions of low pH in the endosome, preventing the fusion process. Moreover, being a groove in this area looking like the enzyme mouth (Fig. 4), bound small molecules in this site might act as indirect allosteric inhibitors for the furin susceptible peptide motif, and therefore, might impair the cleavage step by the furin proteases. The indirect allosteric inhibition mechanism might be through the inhibition of the interaction between the CHIKV envelop protein, and hence the furin susceptible motif (furin loop), and the acting protease (therefore, the Life Chemicals protein-protein inhibitors library was used here), or through trapping the glycoprotein conformation in one inactive form (relative to the furin cleavage step), which does not interact with the acting proteases. The site 2 volume in the immature form is 9.5% bigger than that in the mature form. Therefore, both structures were included in the virtual screening study in an attempt to find positive hits for this site.



**Fig. 4** Surface representation of the location of site 2 (green colour), parts of E2 and E3 are shown in orange colour where the furin loop takes a greyish orange colour at the top, E1 domain II surface is shown in grey colour. Site 2 is located in a mouth like cavity that might interact during the furin cleavage. Generated from the pdb file: 3N40 [17].

Site 2 makes close contact with residues from E1 and E2. The E1 residues are: Glu50-Val60, Val229-Pro237; the E2 residues are: Ala97-Arg102 (corresponds to Ala33-Arg38 in the mature form), Gln300-Arg308 (Gln236-Arg244 in the mature form). Hydrogen bonding within these residues involve E1 Lys52, Thr53, Ile55, Val231, His230, and E2 Tyr301 (Tyr237 in the mature form), Glu232 (Glu168 in the mature form), and Arg100 (Arg36 in the mature form). Val54, Lys52, Arg100, Ile167 are able to form other types of strong noncovalent molecular interactions. Generally, valine, alanine and proline amino acids within this pocket are also able to participate in the hydrophobic interactions.

Surprisingly, we could not find a common hit ligand that fits in site 2 in both the immature and the mature forms. However, inspection of the top 5 docked poses in each site reveals that they have the common sequence: heterocycle-S-CH<sub>2</sub>-CO-N, the amidic nitrogen in this sequence might be NH, and also can be a part of another ring system, Fig. 5 shows the 2D representations of the top docked poses in site 2 for both the immature and the mature forms of the envelope glycoproteins. The presence of an electron rich system results in strong noncovalent molecular interactions, *e.g.* the  $\pi$ -cation interaction between E2 Arg100 (Arg36 in the mature form) and E1 Lys52. The heterocyclic ring adjacent to the sulphur in most of the top ranked poses can accept H-bonding with E1 Lys52, Ile55, and Thr53. Being able to bind to residues in both E1 and E2, the ligands identified for this site are most likely to confirm our hypothesis and stabilize the E1-E2 heterodimer and prevent the dissociation.



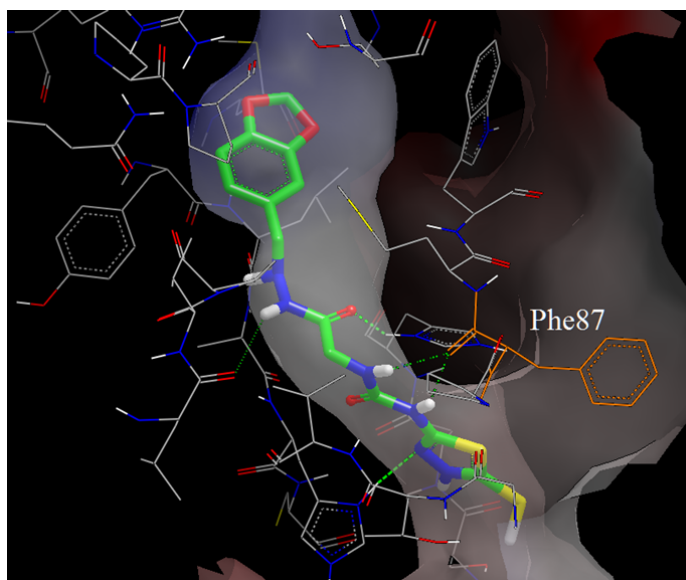
**Fig. 5** 2D representation of the docked poses within the binding site 2. (a) Top ranked pose (number 1 in Table 2) within the *immature glycoprotein* complex site 2. (b) Top ranked pose (number 1 in Table 3) within the *mature*

*glycoprotein* binding site 2. H-bonds are shown in green and blue dashes, while  $\pi$  interactions are shown as orange lines.

Site 4 (blue, Fig. 2) can be described as a narrow channel extending just behind the fusion loop and surrounded by both E2 domains A and B. Comparison between the two sites in the immature and the mature forms indicates that the site volume is 31.3% bigger in the mature form (Fig. 2), which is sufficiently significant to indicate the changes occurring after maturation (cleavage of the furin loop), which looks like an umbrella above this site in the immature form. Small molecules binding to this narrow channel will have significant effects; this might not only freeze the relative movement of E2 domains A and B, but might also freeze the fusion loop through stabilizing interactions, and consequently, prevent the exposure of the fusion loop. The fusion loop is stabilized by the histidine residues of E2, which act as the pH sensors for the activation of the fusion protein at lower pH [17] where the histidine residues become protonated. This site cavity lies in contact with several histidine residues of E2. Therefore, blocking this site may also impair the pH sensor activation mechanism. All the identified hits were found to bind to both E1 and E2 residues, involving the histidine residues of E2, moreover, two hits were found to bind to the fusion loop amino acids, confirming the ability of freezing the fusogenic activity of the envelope proteins.

Site 4 forms close contact with the E1 fusion loop residues Pro86-Gly91, E1 Gly227-His230. The fusion loop Gly91 and His230 (outside of the fusion sequence) were found to be critical for fusion [20]. This emphasises the importance of our hypothesis that binding to this site will impair the fusion process. From E2, residues Arg77-His82 (Arg13-His18 in the mature form), Ser91-Val96 (Ser27-Val32 in the mature form) and residues Leu305-Ala310 (Leu241-Ala246 in the mature form) form close contacts with the binding site.

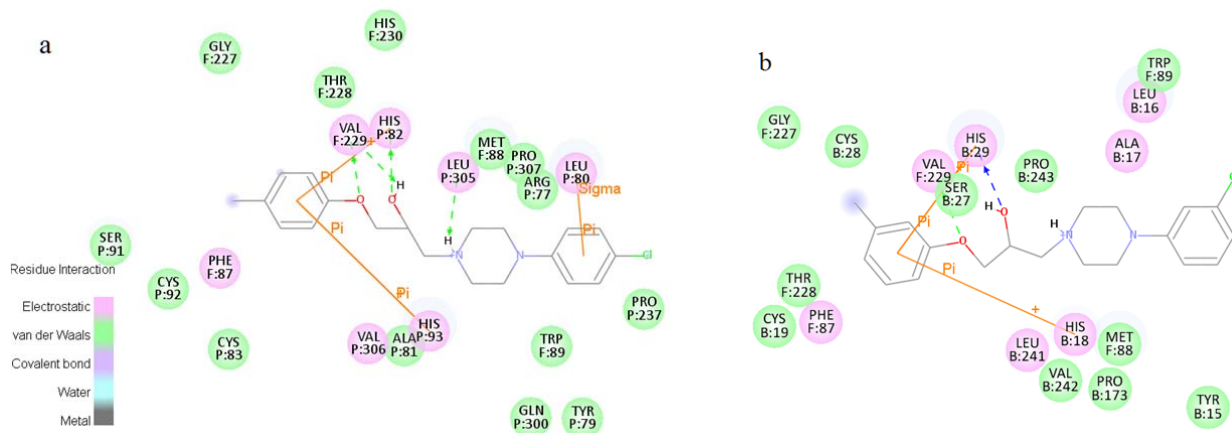
Interestingly, a common hit ligand was found in both the blue sites (ranked 5 in table 4, and ranked 2 in table 5). It shows the same interactions within the two binding pockets and more importantly, it forms two H-bonds (2.05, 2.13 Å) with E1 fusion loop amino acid Phe87, 3 amino acids away from Gly91, the critical [20] residue for the fusion process. It also binds to E2 His82, His93 (His18, His29 in the mature form) via H-bonding and  $\pi$ -cation interactions (Fig. 6). Moreover, the predicated binding affinity and inhibitory constant ( $K_i$ , in the nanomolar range), along with the clogP value of 1.8 (Tables 4, 5), make it an attractive candidate for developing anti-chikungunya drug targeting the envelope proteins.



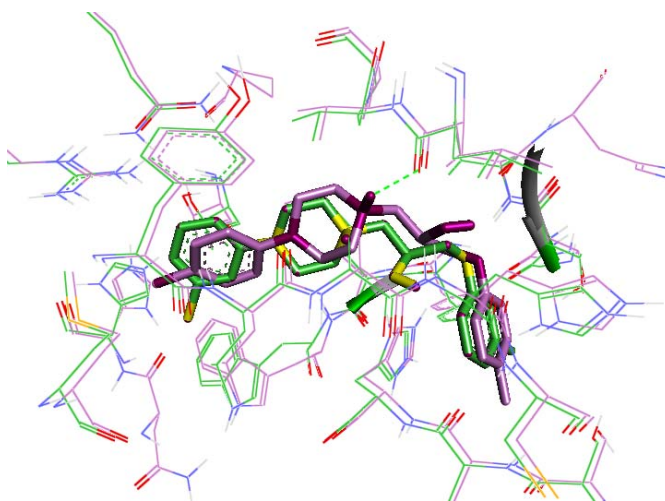
**Fig. 6** 3D representation of the predicted docking pose of compound 2 (Table 5) within site 4 binding pocket (transparent surface) of the *mature* chikungunya envelope protein showing the H-bonding (green dashes) with the fusion loop amino acid Phe87 (shown in orange).

A further interesting observation is the presence of same chiral skeleton; the (*S*)-1-(2-hydroxy-3-phenoxypropyl)-4-phenylpiperazin-1-ium in a series of compounds (1, 2 and 3 in Table 4) and (3, 4 in Table 5). The compounds only differ in the substituents on the terminal phenyl rings. The chirality of this series indicates the selectivity of the compounds and reflects the importance of the stereochemistry in designing inhibitors for this site. The enantiomers of these compounds (within the library) did not pass the first FRED virtual screening (data not shown). Within the immature narrow binding pocket (site 4), these series was able to form H-bonds with the E1 Val229, E2 His82, E2 Leu305. However, in case of the 31.3% bigger pocket of the mature site 4, this series was able to achieve the H-bonding with E1 Val229, E2 His29 whereas it failed to form H-bonds with the E2 Leu241, but was still able to achieve the  $\pi$ -stacking interaction with E2 His18 and His29 (Fig. 7). The importance of this stereo-selectivity in inhibiting the envelop protein, was also noticed recently in the inhibitors of the dengue virus envelop proteins mediated fusion, where compounds with certain stereochemistry of the OH group (the (*S*) enantiomers) were shown to have special effects on the activities [43]. The importance of the (*S*) configuration of the compounds (Fig. 7) can be referred to the ability of the OH groups of the compounds to achieve H-bonds with the E2 histidine residues (His82 of the immature form and His29 of the mature form), whereas these H-bonds might not be possible with the other enantiomers. Superimposition of the two compounds (number 1 in Table 4 and number 3 in Table 5) within the binding pockets (site 4 in both enzyme forms) revealed that the positioning of the docked poses were very similar, with slight

changes in the orientation of the hydroxyl groups and the central piperazinium moiety (Fig. 8). This superimposition not only indicates the reliability of the interactions of this class of compounds with the residues within this site in the two forms of the enzyme, but also confirms our hypothesis that this series might be developed as selective CHIKV envelope protein inhibitors.

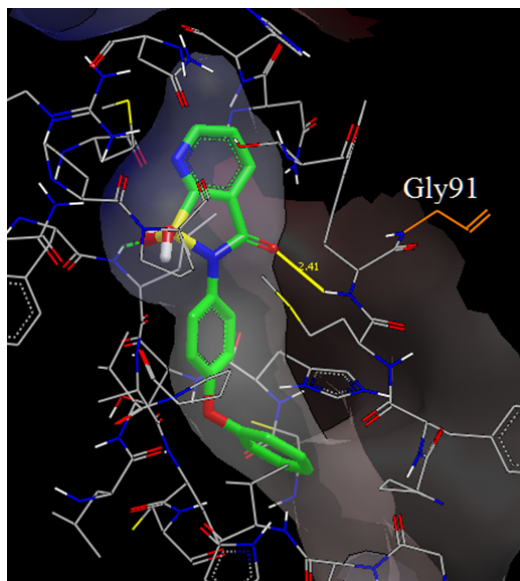


**Fig. 7** 2D representation of the docked poses within the binding site 4. (a) Top ranked pose (number 1 in Table 4) within the *immature glycoprotein* complex site 4. (b) Top ranked pose (number 3 in Table 5) within the *mature glycoprotein* binding site 4. H-Bonds are shown in green and blue dashes, while  $\pi$  interactions are shown as orange lines.



**Fig. 8** Superimposition of compound 1 (in Table 4) within the *immature glycoprotein* site 4 (compound and residues are shown in violet colour) and compound 3 (in Table 5) within the *mature glycoprotein* binding site 4 (compound and pocket residues are shown in green colour). Slight differences can be seen for the orientation of the hydroxyl groups and the central piperazinium ring.

Although ranked 5 in Table 5, this ligand shows extraordinary H-bonding (2.41 Å) with the fusion peptide amino acid Trp89, just one amino acid away from Gly91, the critical [20] amino acid for the fusion process (Fig.9). The pose is also stabilized inside the pocket by the interactions with the E2 His18, His29 and Leu16 (Table 5). This also emphasises the possibility of inhibiting the fusion process through designing ligands for this pocket.



**Fig. 9** Pose 5 (Table 5) within the mature glycoprotein binding site 4 (transparent surface), showing the 2.41 Å H-bond interaction (yellow line) with Trp89. Gly91 is shown in orange.

#### 4. Conclusions

Thus far, the CHIKV envelope protein has not been investigated as a possible target for the drug design against the virus. Therefore, we have investigated for the first time the possible binding target sites within the immature and the mature forms of the CHIKV envelope proteins. We managed to identify two sites that look critical to the protein functions; mainly the fusion process, based on the functionality and the location of the sites. We also have run a virtual screening on the two sites in both forms of the enzymes to increase the chances of finding reliable positive hits. Five hits for each site in both forms of the CHIKV envelope proteins were identified revealing some important features for further developing antagonists for these proteins. To test our hypothesis, the identified hits need to be evaluated against the CHIKV, which is currently under investigation. Our study represents a good template for designing selective inhibitors for the chikungunya virus envelope proteins via *in silico* and *in vitro* optimization process. Our hypothesis might also be a useful tool for inhibiting other *alphaviruses* such as Sindbis virus and Semliki Forest virus as well as other fusion mediated viruses.

## References

- [1] Her, Z., Kam, Y.W., Lin, R.T., Ng, L.F. Chikungunya: a bending reality. *Microbes Infect.* 2009, 11, 1165-76.
- [2] Robinson, M.C. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952-53. I. Clinical features. *Transactions of the Royal Society of Tropical Medicine and Hygiene.* 1955, 49, 28-32.
- [3] Powers, A.M., Logue, C.H. Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *J Gen Virol.* 2007, 88, 2363-77.
- [4] Schwartz, O., Albert, M.L. Biology and pathogenesis of chikungunya virus. *Nat Rev Microbiol.* 2010, 8, 491-500.
- [5] Ziegler, S.A., Lu, L., da Rosa, A.P., Xiao, S.Y., Tesh, R.B. An animal model for studying the pathogenesis of chikungunya virus infection. *Am J Trop Med Hyg.* 2008, 79, 133-9.
- [6] Queyriaux, B., Simon, F., Grandadam, M., Michel, R., Tolou, H., Boutin, J.P. Clinical burden of chikungunya virus infection. *Lancet Infect Dis.* 2008, 8, 2-3.
- [7] 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.
- [8] Chandak, N.H., Kashyap, R.S., Kabra, D., Karandikar, P., Saha, S.S., Morey, S.H., et al. Neurological complications of Chikungunya virus infection. *Neurol India.* 2009, 57, 177-80.
- [9] Singh, S.K., Unni, S.K. Chikungunya virus: host pathogen interaction. *Rev Med Virol.* 2011.
- [10] Sourisseau, M., Schilte, C., Casartelli, N., Trouillet, C., Guivel-Benhassine, F., Rudnicka, D., et al. Characterization of reemerging chikungunya virus. *PLoS Pathog.* 2007, 3, e89.
- [11] Wintachai, P., Wikan, N., Kuadkitkan, A., Jaimipuk, T., Ubol, S., Pulmanusahakul, R., et al. Identification of prohibitin as a Chikungunya virus receptor protein. *J Med Virol.* 2012, 84, 1757-70.
- [12] Kielian, M., Rey, F.A. Virus membrane-fusion proteins: more than one way to make a hairpin. *Nature Reviews Microbiology.* 2006, 4, 67-76.
- [13] Marsh, M., Helenius, A. Virus entry: open sesame. *Cell.* 2006, 124, 729-40.
- [14] Tang, B.L. The cell biology of Chikungunya virus infection. *Cell Microbiol.* 2012, 14, 1354-63.
- [15] Salminen, A., Wahlberg, J.M., Lobigs, M., Liljestrom, P., Garoff, H. Membrane fusion process of Semliki Forest virus. II: Cleavage-dependent reorganization of the spike protein complex controls virus entry. *J Cell Biol.* 1992, 116, 349-57.
- [16] Li, L., Jose, J., Xiang, Y., Kuhn, R.J., Rossmann, M.G. Structural changes of envelope proteins during alphavirus fusion. *Nature.* 2010, 468, 705-8.

- [17] Voss, J.E., Vaney, M.C., Duquerroy, S., Vonnheim, C., Girard-Blanc, C., Crublet, E., et al. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature*. 2010, 468, 709-12.
- [18] Ozden, S., Lucas-Hourani, M., Ceccaldi, P.E., Basak, A., Valentine, M., Benjannet, S., et al. Inhibition of Chikungunya virus infection in cultured human muscle cells by furin inhibitors: impairment of the maturation of the E2 surface glycoprotein. *The Journal of biological chemistry*. 2008, 283, 21899-908.
- [19] Lobigs, M., Zhao, H.X., Garoff, H. Function of Semliki Forest virus E3 peptide in virus assembly: replacement of E3 with an artificial signal peptide abolishes spike heterodimerization and surface expression of E1. *J Virol*. 1990, 64, 4346-55.
- [20] Kuo, S.C., Chen, Y.J., Wang, Y.M., Tsui, P.Y., Kuo, M.D., Wu, T.Y., et al. Cell-based analysis of Chikungunya virus E1 protein in membrane fusion. *J Biomed Sci*. 2012, 19, 44.
- [21] 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, 111-7.
- [22] de Lamballerie, X., Ninove, L., Charrel, R.N. Antiviral treatment of chikungunya virus infection. *Infect Disord Drug Targets*. 2009, 9, 101-4.
- [23] Delogu, I., Pastorino, B., Baronti, C., Nougairede, A., Bonnet, E., de Lamballerie, X. In vitro antiviral activity of arbidol against Chikungunya virus and characteristics of a selected resistant mutant. *Antiviral Res*. 2011, 90, 99-107.
- [24] 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.
- [25] Zhang, L., Luo, R.H., Wang, F., Jiang, M.Y., Dong, Z.J., Yang, L.M., et al. Highly functionalized daphnane diterpenoids from *Trigonostemon thyrsoideum*. *Org Lett*. 2010, 12, 152-5.
- [26] Kaur, P., Thiruchelvan, M., Lee, R.C., Chen, H., Chen, K.C., Ng, M.L., et al. Inhibition of chikungunya virus replication by harringtonine, a novel antiviral that suppresses viral protein expression. *Antimicrob Agents Ch*. 2013, 57, 155-67.
- [27] D'Hooghe, M., Mollet, K., De Vreese, R., Jonckers, T.H., Dams, G., De Kimpe, N. Design, synthesis, and antiviral evaluation of purine-beta-lactam and purine-aminopropanol hybrids. *J Med Chem*. 2012, 55, 5637-41.
- [28] Li, Y.G., Siripanyaphinyo, U., Tumkosit, U., Noranate, N., A, A.N., Pan, Y., et al. Poly (I:C), an agonist of toll-like receptor-3, inhibits replication of the Chikungunya virus in BEAS-2B cells. *Virol J*. 2012, 9, 114.

- [29] McInnes, C. Virtual screening strategies in drug discovery. *Curr Opin Chem Biol.* 2007, 11, 494-502.
- [30] Kirchmair, J., Distinto, S., Schuster, D., Spitzer, G., Langer, T., Wolber, G. Enhancing drug discovery through in silico screening: Strategies to increase true positives retrieval rates. *Curr Med Chem.* 2008, 15, 2040-53.
- [31] Yennamalli, R., Subbarao, N., Kampmann, T., McGeary, R.P., Young, P.R., Kobe, B. Identification of novel target sites and an inhibitor of the dengue virus E protein. *J Comput Aided Mol Des.* 2009, 23, 333-41.
- [32] Zhou, Z.G., Khaliq, M., Suk, J.E., Patkar, C., Li, L., Kuhn, R.J., et al. Antiviral Compounds Discovered by Virtual Screening of Small-Molecule Libraries against Dengue Virus E Protein. *Acs Chem Biol.* 2008, 3, 765-75.
- [33] Li, Z., Khaliq, M., Zhou, Z.G., Post, C.B., Kuhn, R.J., Cushman, M. Design, synthesis, and biological evaluation of antiviral agents targeting flavivirus envelope proteins. *J Med Chem.* 2008, 51, 4660-71.
- [34] 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.
- [35] McGann, M. FRED pose prediction and virtual screening accuracy. *J Chem Inf Model.* 2011, 51, 578-96.
- [36] McGaughey, G.B., Sheridan, R.P., Bayly, C.I., Culberson, J.C., Kreatsoulas, C., Lindsley, S., et al. Comparison of topological, shape, and docking methods in virtual screening. *J Chem Inf Model.* 2007, 47, 1504-19.
- [37] McGann, M.R., Almond, H.R., Nicholls, A., Grant, J.A., Brown, F.K. Gaussian docking functions. *Biopolymers.* 2003, 68, 76-90.
- [38] Bostrom, J., Greenwood, J.R., Gottfries, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J Mol Graph Model.* 2003, 21, 449-62.
- [39] Hawkins, P.C., Nicholls, A. Conformer generation with OMEGA: learning from the data set and the analysis of failures. *J Chem Inf Model.* 2012, 52, 2919-36.
- [40] Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* 2009, 30, 2785-91.
- [41] 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, 935-49.
- [42] Zaheer ul, H., Halim, S.A., Uddin, R., Madura, J.D. Benchmarking docking and scoring protocol for the identification of potential acetylcholinesterase inhibitors. *Journal of Molecular Graphics and Modelling.* 2010, 28, 870-82.
- [43] Mayhoub, A.S., Khaliq, M., Kuhn, R.J., Cushman, M. Design, Synthesis, and Biological Evaluation of Thiazoles Targeting Flavivirus Envelope Proteins. *J Med Chem.* 2011, 54, 1704-14.

