Chem Biol Drug Des 2015; **86:** 1518–1527

Research Article

Discovery of Novel Peptidomimetics as Irreversible CHIKV NsP2 Protease Inhibitors Using Quantum Mechanical-Based Ligand Descriptors

Eman M. El-labbad[1], Mohammed A. H. Ismail[1], Dalal A. Abou El Ella[1], Marawan Ahmed[2], Feng Wang[3], Khaled H. Barakat[2,4,5] and Khaled A. M. Abouzid[1,*]

[1]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Ain Shams University, Abbassia, Cairo 11566, Egypt
[2]Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB, Canada
[3]Molecular Model Discovery Laboratory, Department of Chemistry and Biotechnology, Faculty of Science, Engineering and Technology, Swinburne University of Technology, Melbourne, Vic. 3122, Australia
[4]LiKaShing Institute of Virology, University of Alberta, Edmonton, AB, Canada
[5]LiKaShing Applied Virology Institute, University of Alberta, Edmonton, AB, Canada
**Corresponding author: Khaled A. M. Abouzid, khaled.abouzid@pharma.asu.edu.eg*

Chikungunya virus (CHIKV) is a mosquito-borne alphavirus. Recent outbreaks of CHIKV infections have been reported in Asia, Africa, and Europe. The symptoms of CHIKV infection include fever, headache, nausea, vomiting, myalgia, rash, and chronic persistent arthralgia. To date, no vaccines or selective antiviral drugs against this important emerging virus have been reported. In this study, the design, synthesis, and antiviral activity screening of new topographical peptidomimetics revealed three potential prototype agents 3a, 4b, and 5d showing 93–100% maximum inhibition of CHIKV replication in cell-based assay having EC_{90} of 8.76–9.57 μg/mL. Intensive molecular modeling studies including covalent docking, lowest unoccupied molecular orbital energies, and the atomic condensed Fukui functions calculations strongly suggested the covalent binding of peptidomimetics 3a, 4b, and 5d to CHIKV nsP2 protease leading to permanent enzyme inactivation via Michael adduct formation between α/β-unsaturated ketone functionality in our designed peptidomimetics and active site catalytic cysteine1013. Furthermore, small molecular weight peptidomimetics 3a and 4b satisfied the Lipinski rule of five for drug-likeness and showed promising intestinal absorption and aqueous solubility via computational ADMET studies making them promising hits for further optimization.

Key words: ADMET, atomic condensed Fukui functions, Chikungunya virus, covalent docking, lowest unoccupied molecular orbital energies, peptidomimetics

Received 22 April 2015, revised 25 June 2015 and accepted for publication 10 July 2015

Chikungunya virus (CHIKV) is a mosquito-borne alphavirus transmitted to humans primarily by *Aedes aegypti* and *Aedes albopictus* (1). After the first isolation of CHIKV in Tanzania in 1953 (2), a few localized outbreaks involving relatively few individuals in Africa and Asia were reported (3). Starting from 2005, CHIKV infection unexpectedly arose as a major epidemic human pathogenic infection throughout tropical regions such as central and southern Africa, India, Indonesia, and Malaysia. CHIKV is now frequently being transmitted into other temperate regions of Asia, Europe, Australia, and America (4,5). Clinical manifestations of CHIKV infection include fever, headache, nausea, vomiting, myalgia, rash, and chronic persistent arthralgia similar to rheumatoid arthritis arthralgia (6,7). Researches were carried out to test the efficacy of some antiviral drug candidates against CHIKV (8–12). But up-to-date, no vaccines or a selective antiviral drug against this potential emerging epidemic infection has been reported.

CHIKV genome consists of a linear, positive-sense, single-stranded RNA molecule of about 11.8 kb (13). The non-structural proteins, required for viral replication, are encoded in the 5′ two-thirds of the genome, while the structural genes are within the 3′ one-third (1). There are four non-structural genes, designated nsP1–4, which are necessary for viral replication; each has specific function during the replication process. The nsP1 protein is required for initiation of synthesis of minus-strand RNA and also functions as a methyltransferase to cap the genomic and subgenomic RNAs during transcription (14). The nsP2 protein is made of three domains, the first containing helicase, RNA triphosphatase, and nucleoside triphosphatase activities (15,16) whereas the second and third domains are a papaine-like protease and a non-functional methyltransferase, respectively (17). The functions, roles, and activities of the nsP3 protein are less well understood. It may be involved in the transcription process at an early

 doi: 10.1111/cbdd.12621

stage of the infection (18). The nsP4 protein contains the RNA-dependent RNA polymerase, involved in genome replication and transcription (19). Alphavirus nsP2 protease domain is considered an interesting antiviral target because it is essential for alphavirus replication. NsP2 is cysteine protease plays crucial role in the cleavage and polyprocessing of the non-structural polyprotein. The inhibition of nsP2 protease will lead to prevention of the cleavage of viral polyprotein chain, halting the viral replication. Recently, interactions between CHIKV nsP3 or nsP4 proteins with the host stress-pathway chaperone HSP-90 protein has been reported (20). Furthermore, nsP2 protease exhibits some degree of sequence specificity among alphaviruses making it an attractive broad spectrum antiviral target.

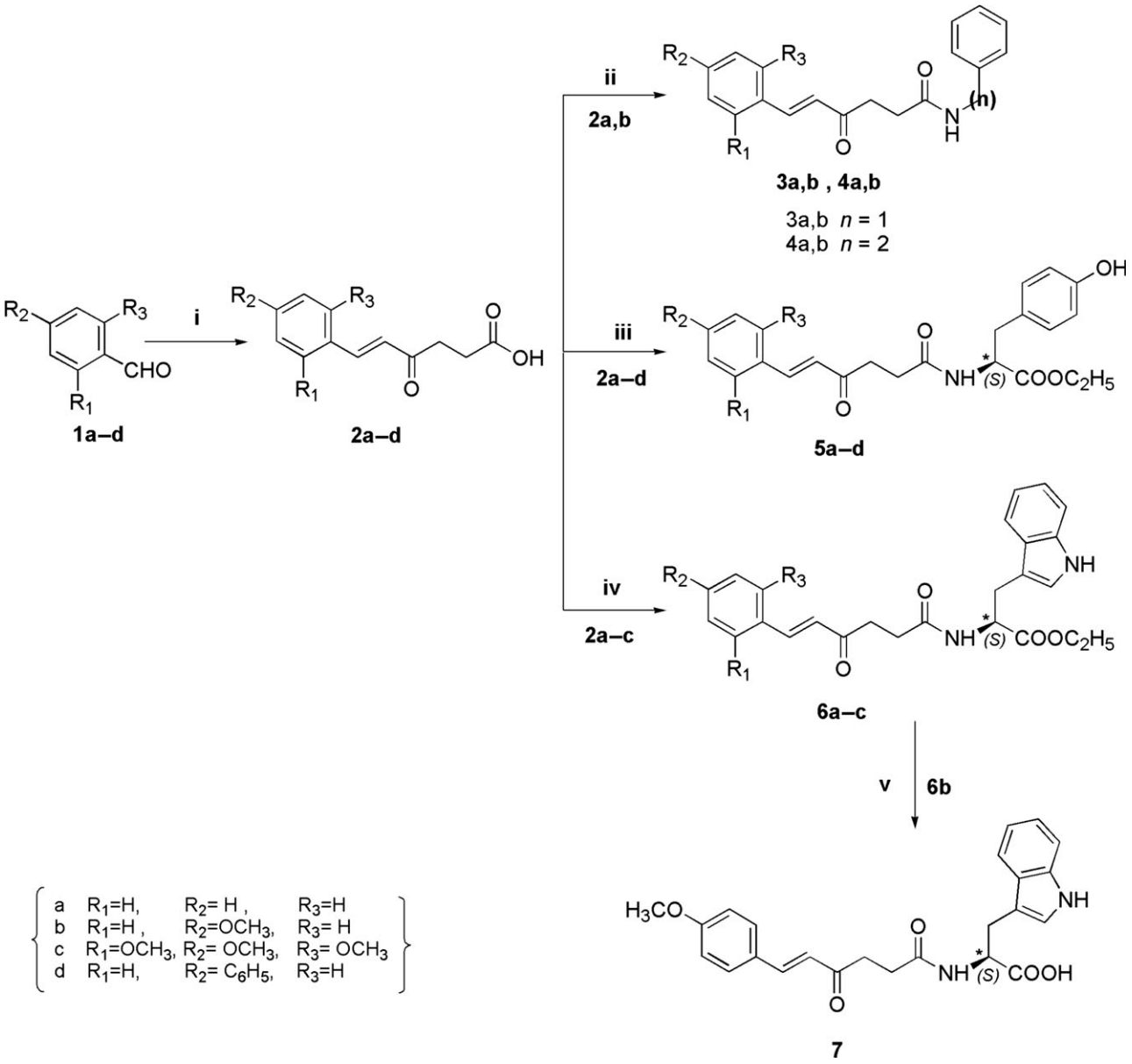
This study focused on nsP2 protease as target for designing new small peptidomimetics as prototype covalent inhibitors of Chikungunya virus nsP2 protease enzyme. CHIKV polynucleotide sequence of non-structural polyprotein (nsP; uniprot database; accession number: Q8JUX6) was investigated and showed that nsP polyprotein has 2474 amino acid and contains four proteins, nsP1 (1–535), nsP2 (536–1333), nsP3 (1334–1863), and nsP4 (1864–2474). The catalytic active site for nsP2 is composed of cysteine1013 and histidine1083. NsP2 protease hydrolyzes peptide bonds at three different sites, the first is the junction between nsP1 and nsP2 (535–536), the second site is nsP2–nsP3 junction (1333–1334), and the third site is nsP3–nsP4 junction (1863–1864) (21). Biochemical studies of the alphavirus nsP2 protease showed considerable differences in cleavage efficiency at the nsP12, nsP23, and nsP34 polyprotein cleavage sites (22–24) with the substrate preference of both full-length nsP2 and the isolated protease domain being nsP34 > nsP12 ≫ nsP23 (24). Even though 15 residues flanking the scissile bond were required for optimal activity, controlled biochemical studies showed that only substrate residues P1′–P4 specifically impacted substrate recognition by the nsP2 protease (22). Therefore, nsP34 tetrapeptide natural substrate at P1′–P3 residues Ala1861-Gly1862-Gly1863-Tyr1864 (AGGY) (25) was selected as prototype lead compound for designing topographical peptidomimetic inhibitors. This was achieved by performing the several rational design strategies (Figure S1A). Tyrosine1864 at P1′ residue was replaced with scaffold (A). Scaffold (A) aryl/heteroaryl groups were designed to mimic interaction pattern of phenolic group of tyrosine1864 at P1′ residue while α/β-unsaturated ketone functionality was adopted to replace scissile amide bond in substrate for possible reaction with the active site cysteine1013 forming a Michael adduct causing an irreversible alkylation of this residue leading to permanent enzyme inactivation (26). Glycine1863 at P1 residue was replaced with pseudopeptide composed of ketone group of α/β-unsaturated ketone attached to two carbon atom linker. Scaffold (B) was designed as dipeptide analogue of glycine1862 and alaine1861 at P2–P3 residues having additional aryl/heteroaryl 2 groups to investigate electrostatic and hydrophobic properties of binding site. The designed inhibitors were superimposed or aligned with AGGY tetrapeptide using FIELDALIGN® software (Cresset group, Cambridgeshire, UK). The alignment pattern was matched with the rational design strategy explained earlier (Figure S1B,C). All designed compounds showed reasonable field points, similarity index and shape similarity with the AGGY tetrapeptide natural substrate. The idea behind this type of alignment is that two molecules which both perfectly aligned to each other based on electric field properties tend to make similar interactions with the binding sites.

The designed compounds were synthesized according to Schemes 1–3. Evaluation for antiviral activity in a CHIKV virus-cell-based assay was carried out according to previously reported method (10). Our study suggested that these inhibitors may act as covalent inhibitors for the CHIKV nsP2 protease based on the fact that the α/β-unsaturated ketone functionality adopted to replace the scissile amide bond in the natural substrate (AGGY tetrapeptide) had been previously reported in many covalent cysteine protease inhibitors. The α/β-unsaturated ketone functionality tend to form Michael adduct with active cysteine in the binding site leading to irreversible alkylation of this residue and permanent enzyme inactivation (26). Furthermore, extensive molecular modeling studies including covalent docking study, lowest unoccupied molecular orbital (LUMO) energies and the atomic condensed Fukui functions calculations were performed. These calculations had been previously reported in determining the reactivity in biological systems and studying the binding of wide range of covalent inhibitors (27), including cysteine protease inhibitors (28).

Methods and Materials

Chemistry

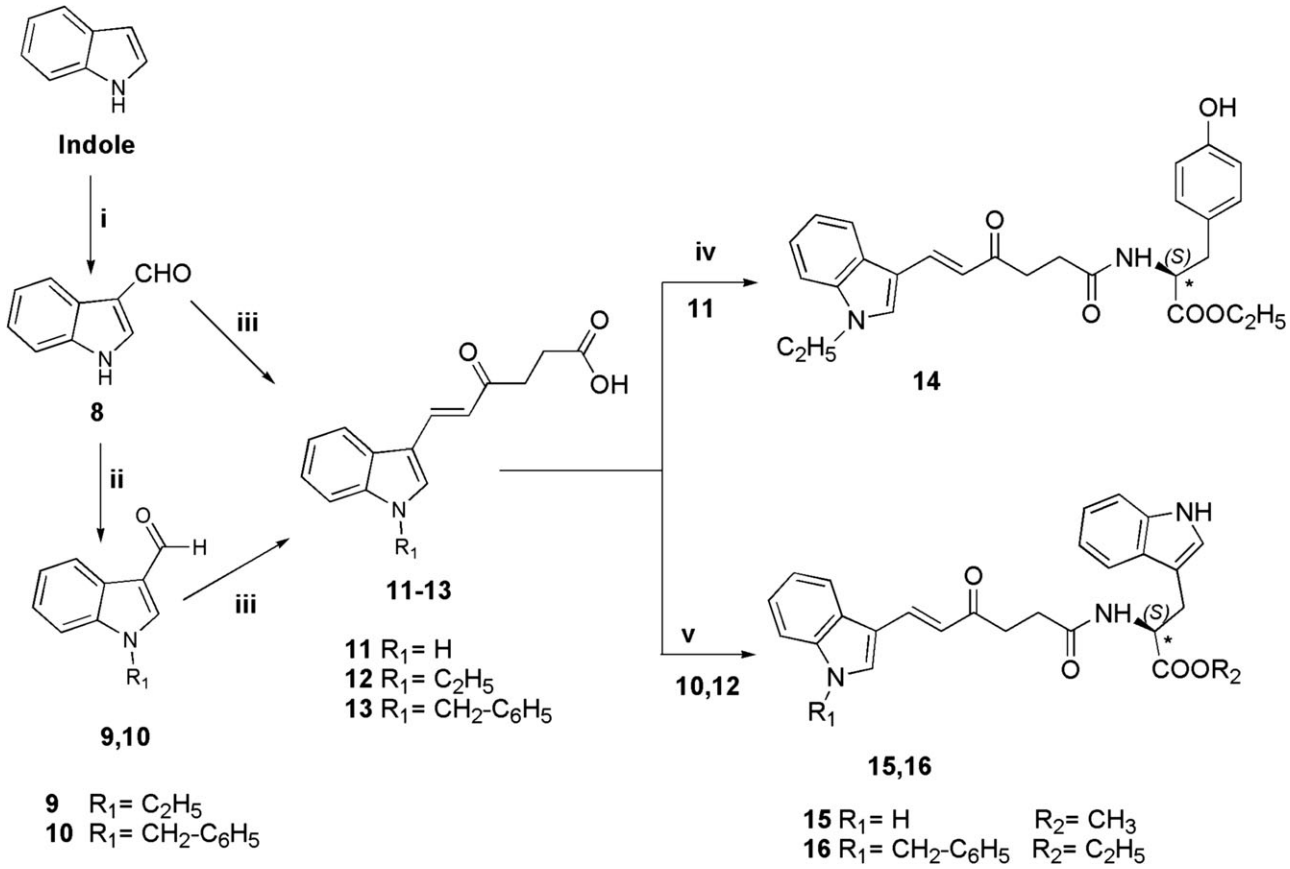
The designed compounds were synthesized according to Schemes 1–3. All the starting materials were purchased from Sigma-Aldrich (Stockholm, Sweden). Melting points were determined on Stuart Scientific apparatus and are uncorrected. Reactions were monitored using thin-layer chromatography (TLC), performed on 0.255 mm silica gel plates, with visualization under UV light (254 nm). FT-IR spectra were recorded on a Perkin–Elmer spectrophotometer. ^{1}H-NMR spectra were recorded on a Perkin–Elmer 300 MHz spectrometer using $CDCl_3$ or DMSO-d_6 as a solvent; chemical shifts (δ) were reported in parts per million (ppm) downfield from TMS; multiplicities are abbreviated as follows: s, singlet; d, doublet; q, quartet; m, multiplet; dd, doublet of doublet; br, broad. MS spectra were recorded on Finnigan Mat SSQ 7000 (70 eV) mass spectrometer and Triple Quadrupole LC/Ms/Ms mass spectrometer API 200 (AB Sciex Instrument). Elemental analyses were performed at the Micro-analytical Centre, Cairo University. Detailed experimental procedures are explained in supplementary information.



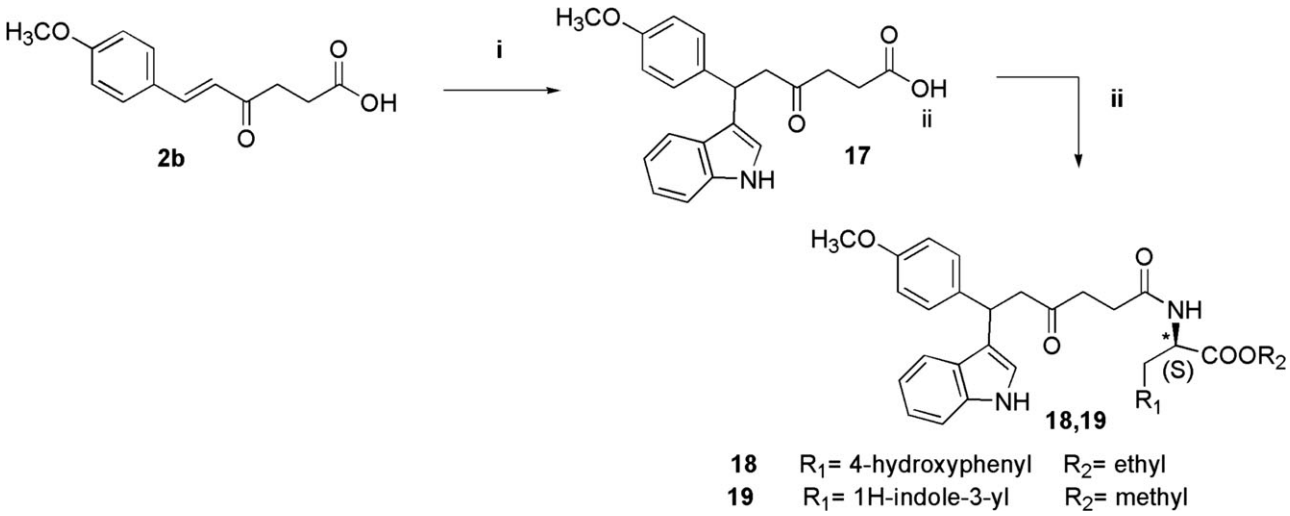
Scheme 1: Synthesis of target compounds. (i) Levulinic acid/morpholine/Gl.acetic acid/dry benzene, reflux for 8 h; (ii) Benzylamine or phenylethylamine/Et_3N/ethyl chloroformate/CH_2Cl_2, stirring 10–12 h; (iii) L-tyrosine ethyl ester/Et_3N/ethyl chloroformate/CH_2Cl_2, stirring 10–12 h; (iv) L-tryptophan ethyl ester/Et_3N/ethyl chloroformate/CH_2Cl2, stirring 10–12 h; and (v) LiOH/MeOH/H_2O, 50 °C for 4 h.

Virus-cell-based antiviral assay

Throughout the experiment, *Vero* (African green monkey kidney) was used as cell line for Chikungunya virus strain 899. In a 96-well microtiter plate (Falcon, BD), serial dilutions of compounds were prepared in assay medium [MEM Rega3 (cat. no. 19993013; Invitrogen) supplemented with 2% FCS (Integro), 5 mL of 200 mM L-glutamine, and 5 mL of 7.5% sodium bicarbonate]. Fifty microliters of a 4× virus dilution in assay medium was added, followed by 50 μL of a cell suspension. This suspension, with a cell density of 25 000 cells/50 μL, was prepared from a *Vero* cell line subcultured in cell growth medium (MEM Rega3 supplemented with 10% FCS, 5 mL of L-glutamine, and 5 mL of sodium bicarbonate) at a ratio of 1:4 and grown for 7 days in 150 cm^2 tissue culture flasks Techno Plastic Products). The assay plates were returned to the incubator for 6–7 days (37 °C, 5% CO_2, and 95–99% relative humidity), a time at which maximal virus-induced cell death or cytopathic effect (CPE) is observed in untreated, infected controls. Subsequently, the assay medium was aspirated, replaced with 75 μL of a 5% MTS (Promega) solution in phenol red-free medium, and incubated for 1.5 h. Absorbance was measured at a wavelength of 498 nm (Safire2; Tecan); optical densities (OD values) reached 0.6–0.8 for the untreated, uninfected controls). Raw data were converted to percentage of



Scheme 2: Synthesis of target compounds. (i) $POCl_3$/DMF; (ii) Ethyl bromide or benzyl chloride/K_2CO_3/CH_3COCH_3; (iii) Levulinic acid/Morpholine/Gl.acetic acid/dry benzene/reflux 8 h; (iv) L-tyrosine ethyl ester/Et_3N/ethyl chloroformate/CH_2Cl_2/stirred 10–12 h; and (v) L-tryptophan methyl or ethyl ester/Et_3N/ethyl chloroformate/CH_2Cl_2/stirred 10–12 h.



Scheme 3: Synthesis of target compounds. (i) Indole/I_2 & ethanol stirring at RT for 6 h and (ii) L-tyrosine ethyl ester or L-tryptophan methyl ester/Et_3N/ethyl chloroformate/CH_2Cl_2, stirring 10–12 h.

controls, and the EC_{50} (50% effective concentration or concentration that is calculated to inhibit virus-induced cell death by 50%) and CC_{50} (50% antimetabolic concentration or concentration that is calculated to inhibit the overall cell metabolism by 50%) were derived from the dose–response curves. All assay conditions producing an

antiviral effect exceeding 50% were checked microscopically for minor signs of a CPE or adverse effects on the host cell (i.e. altered cell morphology). A compound is only considered to elicit a selective antiviral effect on virus replication when, following microscopic quality control, at least at one concentration of compound no CPE nor any adverse effect is observed (image resembling untreated, uninfected cells) (10).

Covalent docking

The crystal structure of the Chikungunya virus nsP2 protease was downloaded from the protein data bank (PDB ID: 3TRK). The protein preparation wizard in Maestro (29,30) was used to repair the protein, adding missing residues and saturate with hydrogens, similar to what have been described in our previous studies (31,32). The geometries of the ligands were optimized at the RM1 (33) level of theory as implemented in the SCHRODINGER package (34). RM1 optimized ligands were used for the primary non-covalent docking with Glide (data not shown) (35,36). Docking has been carried out using the Glide standard precision mode (Glide-SP) with the default settings as implemented in SCHRODINGER. The ultimate goal of this non-covalent docking step was to explore the accessibility of the binding pocket.

Based on previous studies (37,38) docking was restricted to binding site residues asparagine1011, cysteine1013, tyrosine1079, asparagine1082, histidine1083, tryptophan1084, and leucine1205. For the covalent docking, the CovalentDock server (39) was used. The CovalentDock method is well parameterized against wide range of covalently bound inhibitors and achieves excellent results in reproducing the native poses of several covalent drug–protein complexes (39). The method also gives superior results, estimated by calculating the root mean square deviation (RMSD) of the reproduced complexes from the native complexes (average RMSD = 1.68), over other commonly used covalent docking methods such as Gold (average RMSD = 3.69) or Autodock (average RMSD = 2.49 Å) (39).

Quantum mechanical-based ligand descriptors

From the basis of conceptual density functional theory (40,41) and the frontier molecular orbital (FMO) theory (42), Michael addition reaction can be treated as a nucleophilic attack of the highest occupied molecular orbital (HOMO) electrons pair from the Michael donor (S^{-1} of cysteine1013) to the LUMO to form the covalent enzyme–inhibitor complex. In this regard, the LUMO energies and the atomic condensed Fukui functions (43) of the Michael acceptor species (the inhibitors) are two good metrics to explain the reactivity and hence the inhibitory activity of the irreversible covalent inhibitors. Particularly, the atomic condensed Fukui function has several applications in determining the reactivity in biological systems and studying the binding of wide range of covalent inhibitors (27), including cysteine protease inhibitors (28).

The 'response of molecular fragment' approach (44) for calculating the atomic condensed Fukui function evaluates the atomic condensed Fukui function of a given atom (k) in a molecule as follows:

$$f_k^+ = P_k(N+1) - P_k(N)$$

and

$$f_k^- = P_k(N) - P_k(N-1)$$

where f^+ is the susceptibility of atom k to nucleophilic attack; f^- is the susceptibility of atom k to electrophilic attack; P is the atomic electronic population; and N is the total number of electrons of the system.
The ligand were subjected to a more robust geometry optimization at the B3LYP/6-31G* level and a single-point calculations at the B3LYP/6-31 + G* level of theory to calculate the LUMO energies and extract the natural bond orbital (NBO) populations using the GAUSSIAN package (27). For cysteine1013, the active cysteine, we assumed a negatively charged species (cysteine anion) and the calculations were carried out at the same level of theory and basis set as the studied ligands. The presence of both polarized and diffused function on the basis was selected to give consistent results for all studied molecular species (neutral and anions).

Results and Discussion

Chemistry

The designed compounds were synthesized according to Schemes 1–3. The synthesis of (*E*)-6-Aryl-4-oxohex-5-enoic acid **2a–d**, Scheme 1, was adopted according general procedure previously reported (45,46). This involved δ-condensation of the appropriate aldehyde **1a–d** with levulinic acid using catalytic amounts of morpholine and glacial acetic acid. The same procedure was adopted in Scheme 2 to prepare unreported intermediates (*E*)-6-(1H-indol-3-yl)-4-oxo-hex-5-enoic acid **11** and (*E*)-6-(substituted indol-3-yl)-4-oxo-hex-5-enoic acid **12**, **13**. These intermediates were prepared after condensation of 1-indole-3-carbaldehyde **8** or 1-alkyl-1H-indole-3-carbaldehyde **9**, **10** with levulinic acid. The general synthesis of final compounds *(E)*-N-phenylalkyl-6-aryl-4-oxohex-5-enoic acid amides **3a**, **b**, **4a**, **b** Scheme 1, *(S, E)*-3-substituted-2-[6-aryl-4-oxo-hex-5-enoylamino]-propionic acid ethyl ester **5a–d**, **6a–c**, Scheme 1 and *(S, 5E)*-3-(substituted)-2-[6-(indol-3-yl)-4-oxo-hex-5-enoylamino] propionic acid alkyl esters **14–16**, Scheme 2 was carried out through formation of amide bond between corresponding acids **2a–d**, **11–13** with different amines (viz; benzylamine and phenylethylamine) or with amino acid esters (viz, L-tryptophan methyl/ethyl ester and L-tyrosine ethyl ester), using ethyl chlorofor-

mate and triethylamine, adopting the mixed anhydrides method (47). Elucidation of the structures of the target amides **5a–d**, **6a–c**, and **14–16** was proved by different analytical and spectral data. The IR spectra showed the absence of the broad hydroxyl group of the carboxylic function at $\tilde{\nu} \approx 3200–2800\ cm^{-1}$ and the presence of the carboxylic C=O function at shorter wave number at $\tilde{\nu} \approx 1630–1600\ cm^{-1}$ corresponding to the formed amidic C=O and appearance of amidic NH at $\tilde{\nu} \approx 3320–3200\ cm^{-1}$. The MS of compounds showed molecular ions peaks matched with their molecular weights. ^{1}H-NMR data showed the vinylic hydrogens at $\delta \approx 6.6$ and 7.6 ppm, with J constant = 16 Hz, proving that the stereochemistry of such alkene group was the *(E)* isomer (48). Final compounds **18** and **19** were prepared as illustrated in Scheme 3. The unreported Michael adduct *(R S)*-6-(1H-Indol-3-yl)-6-(4-methoxyphenyl)-4-oxo-hexanoic acid **17** was prepared via Michael addition of indole to *(5E)*-6-(4-methoxyphenyl)-4-oxo-hex-5-enoic acid **2b** in non-stereo-specific manner. A modified Wang *et al.* (49) procedure was performed using iodine as catalyst. This procedure successfully produced the titled compound **17** with 73% yield. The structure of the Michael adduct **17** was proved by the different analytical and spectral data. The IR spectra of **17** showed the absence of the C=C at $\acute{\upsilon} \approx 1490\ cm^{-1}$. The MS showed its molecular ion peaks at 351.4 amu. ^{1}H NMR data were consistent with the structure as indicated by the absence of the vinyl hydrogens in its ^{1}H NMR spectrum and the presence of new duplets and triplet signals at δ 3.2 and 4.8 ppm, respectively, corresponding to the –CH–CH_2–CO moiety. Amide bond between carboxylic group of Michael adduct *(R S)*-6-(1H-indol-3-yl)-6-(4-methoxyphenyl)-4-oxo-hexanoic acid **17** and amino group of L-tyrosine ethyl ester or L-tryptophan methyl ester was formed to give final compounds *(6RS,2S)*-3-(4-hydroxyphenyl)-2-[6-(1H-indol-3-yl)-6-(4-methoxyphenyl)-4-oxo-hexanoylamino]-propionic acid ethyl ester **18** and *(6RS,2S)*-3-(1H-indol-3-yl)-2-[6-(1H-indol-3-yl)-6-(4-methoxyphenyl)-4-oxo-hexanoylamino]-propionic acid methyl ester **19**. Based on different analytical and spectra data, the structures of the amides **18** and **19** were proved. The FT-IR spectra of **18** and **19** showed the disappearance of the broad stretching vibration band of hydroxyl group of the carboxylic function at $\tilde{\nu}$ = 3300–2900 cm^{-1} with the shifting of their carbonyl groups vibration to shorter wave number at $\tilde{\nu} \approx 1650\ cm^{-1}$. The MS of compounds **18** and **19** showed matched molecular ion peaks with their molecular weights, at 542.40 and 549.00, respectively.

The covalent binding mode of the new inhibitors

Designing covalent inhibitors for CHIKV cysteine nsP2 protease is a challenging task due to several reasons; first, the designed inhibitor should be capable of accessing the binding site, which is a tunnel-shaped pocket near the protein surface (Figure S2). Substrate access to this pocket is restricted by a β-hairpin at residues alanine1080–histidine1083, as has been shown previously (38). The second challenge is that the inhibitor should be able to react with the activated cysteine residue to form the Michael adduct causing an irreversible alkylation of this residue. In that case, the activated cysteine residue (anion) will be the Michael donor, and the irreversible inhibitor will be the Michael acceptor that covalently binds the protein causing a permanent enzyme inactivation (26). Figure S3 illustrates the proposed mechanism of action of the designed inhibitors. The β-carbon of the active α/β-unsaturated ketone functional group acts as an electrophile (Michael acceptor, E^+) that can be attacked by a nucleophile (Michael donor, Nu^-), which is the active cysteine1013 anion. Table 1 gives the *in vitro* measured inhibitory activity of the synthesized inhibitors. As can be seen in the table, the most active inhibitor identified *in vitro* is compound **3a**, with a maximum inhibition of 100% at 68.2 μg/mL and an EC_{90} of 16.4 μg/mL. Compound **4b** achieves a maximum inhibition of 100% at a slightly lower concentration than compound **3a**, and is given by 59.3 μg/mL and an EC_{90} of 16.6 μg/mL. From the **5a–d** series, compound **5d** achieves a maximum inhibition of 93.8% at a concentration of 42.4 μg/mL and an EC_{90} of 18.7 μg/mL. Our results show that small molecule inhibitors achieve a better success in inhibiting the nsP2 activity, perhaps due to their superior accessibility to the binding pocket.

To determine how the new synthesized inhibitors could covalently bind the nsP2 protease, the CovalentDock web server was used (39). The online tool is able to identify the active β-carbon from the α/β-unsaturated group that acts as the Michael acceptor; however, the sequence number of the active cysteine residue must be supplied by the user; thus, a primary knowledge of the binding site is important. The corresponding two-dimensional (2D) and the three-dimensional (3D) figures of the inhibitor–enzyme covalent complexes formed after *in silico* covalent docking study using CovalentDock web server are given in (Figure S4) for the aforementioned compounds (**3a**, **4b** and **5d**). As we can see in the figure, all ligands are able to bind the cysteine1013 residue covalently and partially protruding inside the binding pocket. The extra phenyl group in compound **5d** seems to facilitate optimal orientation of the active β-carbon so that the Michael addition reaction can take place. This might partially explain the higher activity of compound **5d**, which achieves a maximum inhibition of 93.8%, compared to similar inhibitors in the same series that achieves a much lower maximum inhibition of 16% (compound **5a**), 33.7% (compound **5b**) and 39.3% (compound **5c**). The ΔG values of the inhibitors estimated with CovalentDock do not give strong correlation with *in vitro* measured inhibitory activities and will not be discussed further.

Based on the fact that classical molecular modeling techniques such as covalent docking do not include bond breaking or bond formation, a more accurate method to estimate the expected potency of the inhibitors had been

Compound no.	Max, % inhibition	Conc. (μg/mL)	EC_{50}[a] (μg/mL)	EC_{90}[b] (μg/mL)	ΔG[c] (kcal/mol)	$f^{+CMichaeld}$	LUMO energy (kcal/mol)[d]
3a	100	68.2	8.76	16.4	−8.92	0.152	−59.457
3b	1.36	12.4	>100	–	–	0.147	−47.251
4a	81	65.1	10.9	–	–	0.137	−52.592
4b	100	59.3	8.94	16.6	−7.85	0.143	−46.869
5a	16	10.1	>100	–	–	0.131	−52.133
5b	33.70	235	>100	–	–	0.134	−47.986
5c	39.30	206	>100	–	–	0.145	−39.207
5d	93.80	42.4	9.57	18.7	−9.48	0.103	−53.596
6a	−1.91	1.91	>100	–	–	0.131	−52.918
6b	75.10	44.6	11.8	–	–	0.158	−46.398
6c	ND	ND	ND	–	–	0.129	−39.627
7	2.40	1.9	>100	–	–	0.161	−49.128
14	86.50	216	52.3	–	–	0.141	−44.177
15	35.30	1.8	<0.8	–	–	0.131	−44.051
16	0	183	>100	–	–	0.101	−43.348
18	0.52	1.47	>100	–	–	–	–
19	8.32	181	>100	–	–	–	–

Table 1: Antiviral evaluation in a Chikungunya virus-cell-based assay together with the estimated free energy of binding (ΔG) from CovalentDock, the f^{+} of the Michael acceptor active carbon (denoted as $f^{+CMichael}$) and the LUMO energies, the EC_{90}, and ΔG values are given only for the three most active inhibitors identifies from the cell-based assay

LUMO, lowest unoccupied molecular orbital; ND, not determined.
[a]Concentration at which 50% inhibition of virus replication is observed.
[b]Concentration at which 90% inhibition of virus replication is observed.
[c]ΔG values are obtained from the CovalentDock webserver.
[d]At the B3LYP/6-31 + G* level of theory.

carried out using quantum mechanical-based ligand descriptors. From the basis of conceptual density functional theory (40,41) and the FMO theory (42), Michael addition reaction can be treated as a nucleophilic attack of the HOMO electrons pair from the Michael donor (S^{-1} of cysteine1013) to the LUMO to form the covalent enzyme–inhibitor complex. In this regard, the LUMO energies and the atomic condensed Fukui functions (43) of the Michael acceptor species (the inhibitors) are two good metrics to explain the reactivity and hence the inhibitory activity of the irreversible covalent inhibitors. The atomic condensed Fukui function f^{+} of the Michael acceptor (active β-carbon, $C^{Michael}$) on the corresponding inhibitor is given in Table 1, the table also gives the LUMO orbital energies of the corresponding inhibitors. The f^{+} value of the active Michael acceptor carbon will be denoted as f^{+CM}. As we can see in the table, atomic condensed Fukui function f^{+CM} achieves an excellent correlation with the *in vitro* measured inhibitory activities. With the exception of **5d** in the **5a–d** series, the value of f^{+CM} in any close series of ligands is always higher in the more active inhibitor, another strong proof for the proposed mode of action of the synthesized series of compounds. For example, the f $^{+CM}$ value of in compound **3a** is 0.152, which is higher than that in **3b** (0.147). Also, in the case of the **4a/4b** pair of ligands, the f^{+CM} is in favor of **4b** (f^{+CM} 0.137, Max. Inhibition 100% and EC_{50} = 8.94 μg/mL) over **4a** (f^{+CM} 0.143, Max. Inhibition 81% and EC_{50} = 10.9 μg/mL). In the **5a–c** series, the *in vitro* results are perfectly explained by the f^{+CM} values, that is the least active inhibitor from this series, compound **5a**, has the lowest f^{+CM} (0.131). Also, compound **5c** (Max % inhibition = 39.30%) is slightly more active than compound **5b** (Max % inhibition = 33.70%), which is also reflected by their f^{+CM} values, such that the f^{+CM} value in **5c** is given by 0.145, which is slightly higher than in **5b** where it is given by 0.134. The additional phenyl ring in compound **5d** results in higher *in vitro* activity (Max. Inhibition 93.8% and EC_{50} = 9.57 μg/mL). Similar results are obtained in the **6a–c** series where compound **6b** achieves the highest *in vitro* activity (Max. Inhibition 75.1% and EC_{50} = 11.8 μg/mL) and the highest f^{+CM} as well (f^{+CM} = 0.158), and also in compounds **14**, **15,** and **16**. Please note that we did not consider calculating the f^{+CM} or LUMO energies for compounds **18** and **19** as they do not have an active Michael acceptor site.

The values of the LUMO energies of the inhibitors calculated at the B3LYP/6-31 + G* level of theory are also listed in Table 1. The LUMO energy values show that two inhibitors from the most active inhibitors (**3a** and **5d**) indeed have the lowest LUMO energies (LUMO-**3a** = −59.457 kcal/mol and LUMO-**5d** = −53.596 kcal/mol). In the present case, LUMO energies can provide good enrichment; however, the LUMO energies represent only a static picture of the actual wave function and thus cannot be considered a true good metric in situations in which a chemical reaction takes place, that is wave function reorganization. Nevertheless, the LUMO energies can be considered as a primary metric especially at the early next phase of development in these new series of inhibitors to screen large number of compounds where calculating the f^{+CM} values will be very time consuming. We also observed that the LUMO energies calculated using the faster RM1 semi-empirical method achieves similar performance. It is noted, however, that the two parameters used to assess the potency of the compounds. That

is, LUMO energies and f^{+CM} cannot accurately reproduce the trend for very different molecular scaffolds. Nevertheless, they can be valuable metrics to predict the potency for focused compound libraries or for library enrichment.

To investigate the orbital distribution of the FMO, the orbital characters of the frontier interacting orbitals, that is HOMO of cysteine1013 and LUMO of the inhibitors were identified. Figure S5 displays the orbital characters of the HOMO orbital of cysteine1013 together with the LUMO orbitals of the three most active inhibitors identified *in vitro*, that is **3a**, **4b,** and **5d**. As can be seen in the figure, the HOMO orbital of cysteine1013 is dominated by exclusive contribution from the S^{-1} anion. On the other hand, the LUMO orbitals of the given inhibitors has a much more diffused character with a major contribution from the π-electrons on the active double bond and with other contribution from the surrounding group but with different phases and amplitudes.

In silico *human intestinal absorption & aqueous solubility study*

To assess the druggability of the new series of inhibitors, *in silico* druggability assessment has been carried out. Small molecular weight peptidomimetics **3a**, **4b,** and **5d**, satisfied the Lipinski rule of five (RO5) for drug-likeness compared to tetrapeptide natural substrate AGGY Table S1. The RO5 deals with orally active compounds and defines four simple physicochemical parameter ranges (MWT $\leq$ 500, log p $\leq$ 5, H-bond donors $\leq$ 5, H-bond acceptors $\leq$ 10) associated with 90% of orally active drugs that have achieved phase II clinical status (50). Further investigation was carried out using Accelrys Discovery Studio software ADMET module to predict human intestinal absorption after oral administration & aqueous solubility. Accelrys Discovery Studio software ADMET human intestinal absorption model was developed using 182 compounds in the training set, with descriptors that include AlogP98 and 2D polar surface area (PSA_2D) (51). There are four prediction levels for ADMET absorption ranging from 0 to 3 described as the following: (0) for good absorption, (1) for moderate absorption, (2) for poor absorption, and (3) for very poor absorption. The second ADMET model used was ADMET aqueous solubility model. This model uses linear regression to predict the solubility of compounds in water at 25 °C. The model was generated using a dataset containing 775 compounds (molecular weight between 50 and 800) with training set compounds cover numerous classes, including alkanes, alkenes, alkynes, halogens, amines, alcohols, N-containing compounds, ketones, aldehydes, and organic acids (52). ADMET solubility model can be described by six levels ranging from 0 to 5 described as following: (0) indicated compounds may have extremely low solubility, (1) indicates insoluble compounds or compounds with very low but possible solubility, (2) indicates compound will have low solubility, (3) indicates compound may have good solubility, (4) indicates that this compound may have optimal solubility, and (5) indicates that this compound is too soluble. The analysis of ADMET model results (Table S1) revealed that all biologically active peptidomimetics **3a**, **4a**–**b**, **5d**, **6b,** and **14** showed marked improvement of the predictive intestinal absorption level (good moderate absorption) compared to very poor absorption level of tetrapeptide AGGY. Although this was on the expense of aqueous solubility which was decreased compared to tetrapeptide AGGY, still compound **3a** and **4a**–**b** showed good aqueous solubility while compounds **5d**, **6b,** and **14** were still soluble but with low predictive aqueous solubility.

Conclusions

In summary, new series of inhibitors that covalently bind and inactivate the Chikungunya viral nsP2 protease have been designed, synthesized, and *in vitro* tested. Several inhibitors from the new series show micromolar inhibitory activities, which are strongly, correlated with the reactivity of the Michael acceptor active carbons. The reactivity of the Michael acceptor active carbon has been quantified by calculating some reactivity indices, such as the atomic condensed Fukui functions and LUMO orbital energies and characters. Based on the recent success of the substrate envelope hypothesis in designing several protease inhibitors (53,54), small active inhibitors from the new series (**3a** and **4b**) represent excellent hits for further optimization, as they will be less prone to viral resistance due to binding site residues' mutations. This is in contrast to large size active inhibitors, such as **5d** that partially protrudes outside the binding pocket.

Acknowledgments

The authors are grateful to Dr. Pieter Leyssen at Laboratory for Virology and Experimental Chemotherapy, Rega Institute for Medical Research (KU Leuven), Leuven, Belgium for performing CHIKV virus-cell-based assay.

References

1. Powers A.M. (2008) Togaviruses: Alphaviruses. In: Editors-in-Chief Mahy B.W.J., van Regenmortel M.H.V., editors. Encyclopedia of Virology, 3rd edn. Oxford: Academic Press; p. 96–100.
2. Ross R.W. (1956) The Newala epidemic. III. The virus: isolation, pathogenic properties. J Hyg (Lond);54:177–191.
3. Pialoux G., Gaüzère B.-A., Jauréguiberry S., Strobel M. (2007) Chikungunya, an epidemic arbovirosis. Lancet Infect Dis;7:319–327.
4. Parola P., De Lamballerie X., Jourdan J., Rovery C., Vaillant V., Minodier P., Brouqui P., Flahault A., Raoult D., Charrel R. (2006) Novel chikungunya virus variant in travelers returning from Indian Ocean islands. Emerg Infect Dis;12:1493–1499.

5. Rezza G., Nicoletti L., Angelini R., Romi R., Finarelli A.C., Panning M., Cordioli P. *et al.* (2007) Infection with chikungunya virus in Italy: an outbreak in a temperate region. Lancet;370:1840–1846.
6. Chopra A., Anuradha V., Lagoo-Joshi V., Kunjir V., Salvi S., Saluja M. (2008) Chikungunya virus aches and pains: an emerging challenge. Arthritis Rheum;58:2921–2922.
7. Waymouth H.E., Zoutman D.E., Towheed T.E. (2013) Chikungunya-related arthritis: case report and review of the literature. Semin Arthritis Rheum;43:273–278.
8. Leyssen P., Litaudon M., Guillemot J.-C., Rasoanaivo P., Smadja J., Gurib-Fakim A., Canard B., Gueritte F. (2011) PHYTOCHIK: biodiversity as a source of selective inhibitors of CHIKV replication. Antiviral Res;90:A36–A37.
9. Bourjot M., Leyssen P., Eydoux C., Guillemot J.-C., Canard B., Rasoanaivo P., Guéritte F., Litaudon M. (2012) Chemical constituents of Anacolosa pervilleana and their antiviral activities. Fitoterapia;83:1076–1080.
10. Bourjot M., Delang L., Nguyen V.H., Neyts J., Guéritte F., Leyssen P., Litaudon M. (2012) Prostratin and 12-O-tetradecanoylphorbol 13-acetate are potent and selective inhibitors of Chikungunya virus replication. J Nat Prod;75:2183–2187.
11. Bourjot M., Leyssen P., Eydoux C., Guillemot J.-C., Canard B., Rasoanaivo P., Guéritte F., Litaudon M. (2012) Flacourtosides A-F, phenolic glycosides isolated from flacourtia ramontchi. J Nat Prod;75:752–758.
12. Masrinoul P., Puiprom O., Tanaka A., Kuwahara M., Chaichana P., Ikuta K., Ramasoota P., Okabayashi T. (2014) Monoclonal antibody targeting chikungunya virus envelope 1 protein inhibits virus release. Virology;464–465:111–117.
13. Khan A.H., Morita K., del Carmen Parquet M., Hasebe F., Mathenge E.G.M., Igarashi A. (2002) Complete nucleotide sequence of chikungunya virus and evidence for an internal polyadenylation site. J Gen Virol;83:3075–3084.
14. Ahola T., Kääriäinen L. (1995) Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. Proc Natl Acad Sci;92:507–511.
15. Wengler G., Wengler G. (1993) The NS 3 nonstructural protein of flaviviruses contains an RNA triphosphatase activity. Virology;197:265–273.
16. Rikkonen M., Peränen J., Kääriäinen L. (1994) ATPase and GTPase activities associated with Semliki Forest virus nonstructural protein nsP2. J Virol;68:5804–5810.
17. Strauss E.G., De Groot R.J., Levinson R., Strauss J.H. (1992) Identification of the active site residues in the nsP2 proteinase of sindbis virus. Virology;191:932–940.
18. Wang Y.F., Sawicki S.G., Sawicki D.L. (1994) Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. J Virol;68:6466–6475.
19. Rubach J.K., Wasik B.R., Rupp J.C., Kuhn R.J., Hardy R.W., Smith J.L. (2009) Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. Virology;384:201–208.
20. Rathore A.P.S., Haystead T., Das P.K., Merits A., Ng M.-L., Vasudevan S.G. (2014) Chikungunya virus nsP3 & nsP4 interacts with HSP-90 to promote virus replication: HSP-90 inhibitors reduce CHIKV infection and inflammation in vivo. Antiviral Res;103:7–16.
21. Strauss J.H., Strauss E.G. (1994) The alphaviruses: gene expression, replication, and evolution. Microbiol Rev;58:491–562.
22. Lulla A., Lulla V., Tints K., Ahola T., Merits A. (2006) Molecular determinants of substrate specificity for Semliki Forest virus nonstructural protease. J Virol;80:5413–5422.
23. Vasiljeva L., Merits A., Golubtsov A., Sizemskaja V., Kääriäinen L., Ahola T. (2003) Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. J Biol Chem;278:41636–41645.
24. Vasiljeva L., Valmu L., Kääriäinen L., Merits A. (2001) Site-specific protease activity of the carboxyl-terminal domain of Semliki Forest virus replicase protein nsP2. J Biol Chem;276:30786–30793.
25. Pastorino B.A.M., Peyrefitte C.N., Almeras L., Grandadam M., Rolland D., Tolou H.J., Bessaud M. (2008) Expression and biochemical characterization of nsP2 cysteine protease of Chikungunya virus. Virus Res;131:293–298.
26. Santos M.M.M., Moreira R. (2007) Michael acceptors as cysteine protease inhibitors. Mini Rev Med Chem;7:1040–1050.
27. Faver J., Merz K.M. Jr (2010) Utility of the hard/soft acid–base principle via the Fukui function in biological systems. J Chem Theory Comput;6:548–559.
28. Shah F., Mukherjee P., Gut J., Legac J., Rosenthal P.J., Tekwani B.L., Avery M.A. (2011) Identification of novel malarial cysteine protease inhibitors using structure-based virtual screening of a focused cysteine protease inhibitor library. J Chem Inf Model;51:852–864.
29. Wodicka L.M., Ciceri P., Davis M.I., Hunt J.P., Floyd M., Salerno S., Hua X.H., Ford J.M., Armstrong R.C., Zarrinkar P.P., Treiber D.K. (2010) Activation state-dependent binding of small molecule kinase inhibitors: structural insights from biochemistry. Chem Biol;17:1241–1249.
30. Madhavi Sastry G., Adzhigirey M., Day T., Annabhimoju R., Sherman W. (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des;27:221–234.
31. Ahmed M., Sadek M.M., Abouzid K.A., Wang F. (2013) In silico design: extended molecular dynamic simulations of a new series of dually acting inhibitors against EGFR and HER2. J Mol Graph Model;44:220–231.
32. Ahmed M., Sadek M.M., Serrya R.A., Kafafy A.H.N., Abouzid K.A., Wang F. (2013) Assessment of new anti-HER2 ligands using combined docking, QM/MM scoring and MD simulation. J Mol Graph Model;40:91–98.
33. Rocha G.B., Freire R.O., Simas A.M., Stewart J.J.P. (2006) RM1: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. J Comput Chem;27:1101–1111.
34. Bochevarov A.D., Harder E., Hughes T.F., Greenwood J.R., Braden D.A., Philipp D.M., Rinaldo D., Halls

M.D., Zhang J., Friesner R.A. (2013) Jaguar: a high-performance quantum chemistry software program with strengths in life and materials sciences. Int J Quantum Chem;113:2110–2142.
35. Keledjian K., Garrison J.B., Kyprianou N. (2005) Doxazosin inhibits human vascular endothelial cell adhesion, migration, and invasion. J Cell Biochem;94:374–388.
36. Friesner R.A., Murphy R.B., Repasky M.P., Frye L.L., Greenwood J.R., Halgren T.A., Sanschagrin P.C., Mainz D.T. (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. J Med Chem;49:6177–6196.
37. Bassetto M., De Burghgraeve T., Delang L., Massarotti A., Coluccia A., Zonta N., Gatti V., Colombano G., Sorba G., Silvestri R., Tron G.C., Neyts J., Leyssen P., Brancale A. (2013) Computer-aided identification, design and synthesis of a novel series of compounds with selective antiviral activity against chikungunya virus. Antiviral Res;98:12–18.
38. Russo A.T., Malmstrom R.D., White M.A., Watowich S.J. (2010) Structural basis for substrate specificity of alphavirus nsP2 proteases. J Mol Graph Model;29:46–53.
39. Ouyang X., Zhou S., Su C.T.T., Ge Z., Li R., Kwoh C.K. (2013) CovalentDock: automated covalent docking with parameterized covalent linkage energy estimation and molecular geometry constraints. J Comput Chem;34:326–336.
40. Liu S.B. (2009) Conceptual density functional theory and some recent developments. Acta Phys Chim Sin;25:590–600.
41. Roos G., Geerlings P., Messens J. (2009) Enzymatic catalysis: the emerging role of conceptual density functional theory. J Phys Chem B;113:13465–13475.
42. Li Y., Evans J.N.S. (1995) The Fukui function: a key concept linking frontier molecular orbital theory and the hard–soft-acid–base principle. J Am Chem Soc;117:7756–7759.
43. Fukui K. (1982) Role of Frontier orbitals in chemical reactions. Science;218:747–754.
44. Bultinck P., Fias S., Van Alsenoy C., Ayers P.W., Carbó-Dorca R. (2007) Critical thoughts on computing atom condensed Fukui functions. J Chem Phys;127:034102–034111.
45. Zaheer S.H., Kacker I.K., Rao N.S. (1956) The condensation of levulinic acid with aromatic aldehydes. Chem Ber;89:351–354.
46. Abouzid K., Frohberg P., Lehmann J., Decker M. (2007) 6-Aryl-4-oxohexanoic acids: synthesis, effects on eicosanoid biosynthesis, and anti-inflammatory in vivo-activities. Med Chem;3:433–438.
47. Ismail M., Lehmann J., Abou El Ella D., Albohy A., Abouzid K. (2009) Lonazolac analogues: molecular modeling, synthesis, and in vivo anti-inflammatory activity. Med Chem Res;18:725–744.
48. Silverstein R.M., Webster F.X., Kiemie D. (2005) Spectrometric Identification of Organic Compounds, 7th edn. Hoboken: John Wiley & Sons, Inc.
49. Wang S.-Y., Ji S.-J., Loh T.-P. (2003) The Michael addition of indole to α, β-unsaturated ketones catalyzed by iodine at room temperature. Synlett;15:2377–2379.
50. Lipinski C.A. (2004) Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol;1:337–341.
51. Egan W.J., Merz K.M., Baldwin J.J. (2000) Prediction of drug absorption using multivariate statistics. J Med Chem;43:3867–3877.
52. Cheng A., Merz K.M. (2003) Prediction of aqueous solubility of a diverse set of compounds using quantitative structure–property relationships. J Med Chem;46:3572–3580.
53. Nalam M.N.L., Ali A., Altman M.D., Reddy G.S.K.K., Chellappan S., Kairys V., Ozen A., Gilson M.K., Tridor B., Rana T.M., Schiffer C.A. (2010) Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. J Virol;84:5368–5378.
54. Shen Y., Altman M.D., Ali A., Nalam M.N.L., Cao H., Rana T.M., Schiffer C.A., Tidor B. (2013) Testing the substrate-envelope hypothesis with designed pairs of compounds. ACS Chem Biol;8:2433–2441.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Materials and methods.

Figure S1. Design strategy of new CHIKV nsP2 protease peptidomimetic covalent inhibitors.

Figure S2. Binding site of nsP2 CHIKV protease, (A) the tunnel shaped binding site of nsP2 CHIKV protease, (B) binding site residues in ball and stick representation. The β-hairpin structure designated is thought to restrict the access of substrates.

Figure S3. The proposed mechanism of action of the designed inhibitor as illustrated by the Michael addition reaction between cys1013 and compound **3a**.

Figure S4. Covalent docking of the three most active inhibitors defined *in vitro* (**3a**, **4b** and **5d**) with the nsP2 enzyme.

Figure S5. The orbital distribution of the frontier molecular orbitals involved in the Michael addition reaction, that is HOMO of Cys1013 and the LUMO orbitals of **3a**, **4b** and **5d**. Orbital are calculated at the B3LYP/6-31 + G*.

Table S1. Principal parameters calculated by ADMET module.