# Proteases of Emerging Viruses as Targets for Broad-Spectrum Inhibitors

Linlin Zhang

From the Institute of Biochemistry

of the University of Lübeck

Director: Prof. Dr. rer. nat. Dr. h.c. Rolf Hilgenfeld

# Proteases of Emerging Viruses as Targets for Broad-Spectrum Inhibitors

Dissertation in Fulfillment of the

Requirements for the Doctoral Degree

of the University of Lübeck

From the Department of Natural Sciences

Submitted by

**Linlin Zhang**

from Heze, Shandong, China

Lübeck, 2017

First referee: Prof. Dr. Rolf Hilgenfeld

Second referee: Prof. Dr. Thomas H. Peters

Date of oral examination: 3 February, 2017

Approved for printing: 7 February, 2017

## Declaration

I declare that I have written this dissertation completely on my own. Furthermore, I confirm that no other sources have been used than those specified in the dissertation itself.

This disseration, in same or similar form, has not been submitted to any other doctoral degree committee yet.

(Linlin Zhang)

**Results achieved in this thesis are described in the following manuscripts and publications (in alphabetical order according to first author; §: equal contribution):**

Kouretova J., M. Hammamy Z., Epp A., Hardes K., Kallis S., **Zhang L.**, Hilgenfeld R., Bartenschlager R., Steinmetzer T. The effects of NS2B-NS3 protease and furin inhibition on West Nile and Dengue virus replication. J. Enzyme Inhib. Med. Chem., in revision (2017). (*I developed a new preparation method for DENV NS2B-NS3$^{pro}$.*)

Lei J., Hansen G., Nitsche C., Klein D. C., **Zhang L.**, Hilgenfeld R. Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor. Science 353, 503-505 (2016). (*I determined the kinetics of the of ZIKV NS2B-NS3$^{pro}$ wild type and variants, and the inhibitory activity of cn-716 against ZIKV NS2B-NS3$^{pro}$.*)

Lin, D.§, **Zhang L.**§, Kusov Y., Nian Y., Ma Q., Wang J., Leyssen P., Lanko K., Neyts J., de Wilde A., Snijder J. E., Liu H., Hilgenfeld R. Alpha-ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication. J. Med. Chem., to be submitted (2017). (*I determined the inhibitory activities of alpha-ketoamides against coronavirus proteases and partially against enterovirus proteases, and elucidated the crystal structures of coronavirus proteases in complex with respective α-ketoamides.*)

Nitsche C., **Zhang L.**, Weigel F. L., Schilz J., Graf D., Bartenschlager R. Hilgenfeld R., Klein C. D. Peptide-boronic acid inhibitors of flaviviral proteases: medicinal chemistry and structural biology. J. Med. Chem. 60, 511-516 (2017). (*I determined the crystal structure of WNV protease in complex with a boronate inhibitor.*)

Rut W., **Zhang L.,** Kasperkiewicz P., Poreba M., Hilgenfeld R., Drąg M. Extended substrate specificity and first potent irreversible inhibitor/ activity-based probe design for Zika virus NS2B-NS3 protease. Antiviral Res. 139, 88-94 (2017). (*I prepared the ZIKV NS2B-NS3 protease.*)

Schöne T., Grimm L., Sakai N., **Zhang L.**, Hilgenfeld R., and Peters T. STD NMR experiments identify a structural motif with novel exosite activity against West Nile Virus NS2B-NS3$^{pro}$ protease. To be submitted (2017). (*I prepared the WNV NS2B-NS3$^{pro}$ and performed crystallization trials together with Schöne T.*)

**Zhang L.**, Gül Y., Hilgenfeld R. Zika virus NS2B-NS3 protease is a hyperactive enzyme. J. Virol., to be submitted (2017).

**Zhang L.**, Zhu L., Hilgenfeld R. The main protease of human alphacoronavirus NL63 has a much more stringent substrate specificity than that of the betacoronavirus SARS-CoV. To be submitted (2017).

**In addition, publication related to research in this thesis:**

Sun L., Meijer A., Froeyen M., **Zhang L.**, Thibaut J. H., Baggen J., George S., Vernachio J., van Kuppeveld J. F., Leyssen P., Hilgenfeld R., Neyts J. & Delang L. Antiviral activity of broad-spectrum and enterovirus-specific inhibitors against clinical isolates of enterovirus D68. Antimicrob. Agents Chemother. 59, 7782-7785 **(2015)**.

## Abbreviations and symbols

| | |
|---|---|
| $3CL^{pro}$ | 3C-like protease |
| $3C^{pro}$ | 3C protease |
| A | adenine |
| ACD | Available Chemicals Directory |
| *ap* | antiperiplanar |
| BPTI | bovine pancreatic trypsin inhibitor |
| C | capsid protein |
| CMK | chloromethyl ketone |
| CVA16 | Coxsackievirus A16 |
| CVB3 | Coxsackievirus B3 |
| DENV | Dengue virus |
| DHF/DSS | Dengue hemorrhagic fever and shock syndrome |
| DLS | dynamic light-scattering |
| DMSO | dimethyl sulfoxide |
| DTT | dithiothreitol |
| E | envelope protein |
| EDTA | ethylenediamine tetraacetic acid |
| ER | endoplasmic reticulum |
| EV-A71 | Enterovirus A71 |
| EV-D68 | Enterovirus D68 |
| FDA | Food and Drug Administration |
| FRET | Fluorescence Resonance Energy Transfer |
| $G_4SG_4$ | $Gly_4$-Ser-$Gly_4$ |
| GBS | Guillain-Barré syndrome |
| GlnLactam | glutamine lactam |
| HCoV NL63 | human coronavirus NL63 |
| HCV | Hepatitis C Virus |
| HE | hemagglutinin-esterase |
| HIV | human immunodeficiency virus |
| hNPC | human neural progenitor cells |
| IPTG | isopropyl-β-D-thiogalactopyranoside |
| JEV | Japanese encephalitis virus |
| kD | kilo Dalton |
| M | membrane glycoprotein |
| MDDCs | monocyte-derived dendritic cells |
| MERS-CoV | Middle East respiratory syndrome coronavirus |
| mg | milligram |
| min | minute |
| mM | millimolar |
| MPD | 2-methyl-2,4-pentanediol |
| $M^{pro}$ | main protease |
| mRNA | messenger RNA |

| | |
|---|---|
| N | nucleocapsid phosphoprotein |
| NCR | noncoding regions |
| nm | nanometer |
| NS2B | nonstructural protein 2B |
| NS2B-NS3$^{pro}$ | NS2B-NS3 protease |
| NS3 | nonstructural protein 3 |
| Nsp | nonstructural protein |
| nt | nucleotides |
| ORF | open reading frame |
| PCR | polymerase chain reaction |
| PDB | Protein Data Bank |
| PEG | polyethylene glycol |
| PL$^{pro}$ | papain-like protease |
| prM | precursor of M protein |
| RFU | relative fluorescence units |
| rmsd | root-mean-square deviation |
| RNA | ribonucleic acid |
| RTC | replication/transcription complex |
| S | spike glycoprotein |
| SAR | structure-activity relationships |
| SARS-CoV | severe acute respiratory syndrome coronavirus |
| *sc* | synclinal |
| SD | standard deviation |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SS-dimer | disulfide-bonded dimer |
| STING | stimulator of interferon genes |
| TBEV | tick-borne encephalitis virus |
| TCEP | tris(2-carboxyethyl) phosphine |
| UTR | untranslated region |
| v/v | volume/volume |
| WHO | World Health Organization |
| WNV | West Nile virus |
| wt | wild-type |
| YFV | Yellow fever virus |
| ZIKV | Zika virus |
| µg | microgram |
| µl | microlitre |
| µM | micromolar |

# Table of contents

## Abstract

Human coronaviruses (CoVs) are pathogens causing respiratory illness with mild to severe symptoms. The two emerging viruses, severe acute respiratory syndrome coronavirus (SARS-CoV), and Middle East respiratory syndrome coronavirus (MERS-CoV), attracted the world's attention towards the coronaviruses. However, in spite of the high mortality and morbidity caused by these viruses, there are no approved drugs available for clinical treatment. All the coronaviruses require the main protease ($M^{pro}$) to process the viral polyprotein into the many components of the replication/transcription complex (RTC) at no less than 11 cleavage sites. This makes the $M^{pro}$ an attractive target for drug design against coronaviruses. The $M^{pro}$ is also called 3C-like protease ($3CL^{pro}$), because it has a chymotrypsin-like fold similar to the 3C protease of enteroviruses. In addition, the two types of proteases share similar cleavage preferences, in particular for the P1 position conserved as Gln in almost all the cleavage sites, thereby offering opportunities to design broad-spectrum drugs against corona- and enteroviruses. Herein, the crystal structure of human coronavirus NL63 (HCoV-NL63) $M^{pro}$ was determined, revealing major differences between the S2 pockets of the $M^{pro}$s from alpha- and betacoronavirus. Subsequently, a mutant of HCoV-NL63 $M^{pro}$ was generated to mimic the S2 pocket of SARS-CoV $M^{pro}$, resulting in good inhibitory activity against this variant of two lead compounds (inactive against the wild-type HCoV-NL63 $M^{pro}$) with a phenyl group at P2. Concerning the differences of the S2 pocket in proteases from alpha- and betacoronavirus, as well as enteroviruses, a series of α-ketoamides was tested focusing on the modification of the P2 residue. The best- compromise α-ketoamide with a cyclopentylmethyl at the P2 position showed micromolar-range $IC_{50}$ values against several proteases from alpha- and betacoronaviruses, as well as enteroviruses.

Within this study, six crystal structures of $M^{pro}$s or $3C^{pro}$ in complex with different α-ketoamides were determined to elucidate the binding modes.

In addition, the inhibitory activities of a series of aldehydes were determined against SARS-CoV $M^{pro}$. However, no obvious structure-activity relationship (SAR) could be detected against SARS-CoV $M^{pro}$ due to the highly electrophilic aldehyde warhead. Two crystal structures of SARS-CoV $M^{pro}$ in complex with two aldehydes were determined to elucidate the binding details.

Dengue virus (DENV), West Nile virus (WNV) as well as the newly emerging Zika virus (ZIKV) belong to the genus *flavivirus*. Infection by ZIKV not only results in self-limited illness with rash, fever, myalgia symptoms, but can also cause microcephaly in newborns and Guillain-Barré syndrome. Similar to the situation with coronaviruses, there are also no effective approved drugs. The proteolytic function of the N-terminal third of the NS3 protein, along with a hydrophilic part of the NS2B cofactor, is essential for viral replication and maturation of the virions. Therefore, the NS2B/NS3 protease is an attractive drug target. In this part of the work, a method for preparing stable and highly pure DENV2 NS2B-NS3 protease (NS2B-$NS3^{pro}$) was established. Subsequently, the crystal structure of DENV2 NS2B-$NS3^{pro}$ was determined by using this method for the protein preparation. Next, a complex structure of WNV NS2B-$NS3^{pro}$ with a drug-like, capped dipeptidyl boronic-acid inhibitor (**cn-716**) provides a basis for further inhibitor design against DENV, ZIKV and other re-emerging agents. This boronate inhibitor exhibits a nanomolar-range $K_i$ value against the ZIKV NS2B-$NS3^{pro}$, so that the compound could be developed into a broad-spectrum therapeutic for flaviviruses. Moreover, the ZIKV NS2B-$NS3^{pro}$ exhibits hyper-activity against the canonical substrate Bz-Nle-Lys-Lys-Arg-AMC, which is partially due to Asp83* in the S2 pocket forming a salt-bridge with the substrate P2 residue. Finally, determination of the cleavage activities against Fluorescence-Resonance-Energy-Transfer (FRET) substrates designed to mimic the native and putative cleavage sites in the polyprotein as well as the possible cleavage site of flavivirus proteases in STING (stimulator of interferon genes) revealed the cleavage pattern of ZIKV NS2B-NS3$^{pro}$, and a possible pathway for suppression of the host innate immunity.

## Zusammenfassung

Coronaviren sind pathogene Erreger, die milde bis schwere Atemwegserkrankungen hervorrufen. Nach dem erstmaligen Auftreten zweier hochvirulenter Coronaviren, des Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) sowie des Middle East Respiratory Syndrome Coronavirus (MERS-CoV), rückten diese in den Fokus der Forschung. Dennoch sind bis heute noch keine antiviralen Präparate zugelassen. Die Main Protease ($M^{pro}$) stellt ein essentielles Protein aller Coronaviren dar. Sie prozessiert das Polyprotein an insgesamt 11 Schnittstellen in die jeweiligen Bestandteile des Replikations-/ Transkriptionskomplexes. Infolgedessen stellt die $M^{pro}$ ein ausgezeichnetes Zielprotein für die Entwicklung von Hemmstoffen gegen Coronaviren dar.

Aufgrund der vergleichbaren Faltung zu 3C-Proteasen der Enteroviren wird die $M^{pro}$ auch als 3C-artige Protease ($3CL^{pro}$) bezeichnet; sie weist darüber hinaus eine vergleichbare Spaltungsspezifität auf. Dies wird vor allem in Position P1 deutlich, da Gln in nahezu allen Spaltstellen konserviert ist. Aus den genannten Gründen können Breitbandinhibitoren gegen Corona- und Enteroviren entwickelt werden. In der vorliegenden Arbeit wurde die Proteinstruktur der HCoV-NL63 $M^{pro}$ aufgeklärt. Dabei wurden die wichtigsten strukturellen Unterschiede der S2-Tasche der Main Protease von Alpha- und Betacoronaviren untersucht.

Darüber hinaus wurde eine Variante der HCoV-NL63 $M^{pro}$ erzeugt, um die S2-Tasche der SARS $M^{pro}$ nachzubilden. Dies führte zu einer großen Steigerung der inhibitorischen Wirkung zweier Leitstrukturen, die eine Phenylgruppe in P2 besitzen und zuvor keine Aktivität gegen die HCoV-NL63 $M^{pro}$ aufwiesen. Des Weiteren wurde eine Reihe von $\alpha$-Ketoamiden getestet, um die inhibitorische Wirkung sowohl mit Modifikationen der Position P2 als auch mit strukturellen Unterschieden

der S2-Tasche in Verbindung zu bringen. Als stärkster Inhibitor erwies sich dabei ein $\alpha$-Ketoamide mit einer Cyclopentylmethyl-Gruppe in der P2-Position. Dieser Inhibitor wies eine $IC_{50}$ im mikromolaren Bereich auf und war gegen verschiedene Proteasen sowohl von Alpha- und Betacoronaviren als auch Enteroviren wirksam. Im Rahmen dieser Arbeit konnten die Strukturen von insgesamt 6 Komplexen der $M^{pro}$ oder $3C^{pro}$ mit verschiedenen $\alpha$-Ketoamiden aufgeklärt werden. Aus den gewonnen Informationen konnten die Bindungsmoden der Inhibitoren an das aktive Zentrum bestimmt werden.

Darüber hinaus wurden die inhibitorischen Aktivitäten keiner Reihe von Aldehyden gegen SARS-CoV $M^{pro}$ bestimmt. Allerdings konnte eine offensichtliche Struktur-Aktivität-Beziehung (SAR) aufgrund der stark elektrophilen Aldehyd-Gruppe gegen die SARS-CoV $M^{pro}$ nicht gefunden werden. Zwei Kristallstrukturen von SARS-CoV $M^{pro}$ in Komplex mit Aldehyden wurden aufgeklärt, um die Bindungsdetails zu visualisieren.

Dengue Virus (DENV), West Nil Virus (WNV) und das neu aufgetretene Zikavirus (ZIKV) gehören zur Gattung der Flaviviren. Das ZIKV verursacht selbstlimitierende Infektionen mit Symptomen wie Fieber, Ausschlägen und Myalgie. Zusätzlich können ZIKV-Infektionen eine Mikrozephalie von Neugeborenen sowie das Guillain-Barré-Syndrom verursachen. Wie bei Coronaviren sind bis heute keine effektiven antiviralen Präparate zugelassen, so dass nur eine unterstützende Behandlung möglich ist. Die proteolytische Funktion des NS3-Proteins entfaltet sich mit Hilfe des hydrophilen Bereichs des NS2B-Cofaktors und ist essentiell für die virale Replikation und die Reifung des Virions. Folglich stellt die NS2B/NS3-Protease ein sehr gutes Zielprotein zur Entwicklung von Therapeutika gegen Flaviviren dar.

Teil dieser Arbeit war die Etablierung der Präparation stabiler DENV2-NS2B/NS3 Protease (NS2B-$NS3^{pro}$) mit hoher Reinheit. Dies ermöglichte die Aufklärung der Proteinstruktur der DENV2-NS2B--NS3$^{pro}$. Darüber hinaus wurde erstmals die Struktur eines Komplexes der WNV-NS2B-NS3$^{pro}$ mit einem Dipeptidylboronsäure-Inhibitor (cn-716) aufgeklärt. Die resultierenden Ergebnisse lieferten eine gute Basis, um weitere Inhibitoren gegen DENV, ZIKV und verwandte Viren zu entwickeln. Der verwendete Boronatinhibitor weist $K_i$ -Werte im nanomolaren Bereich für die ZIKV-NS2B-NS3$^{pro}$ auf und ist folglich als Ausgangspunkt für Breitbandinhibitoren gegen Flaviviren sehr gut geeignet.

Darüber hinaus wurde eine außerordentlich hohe Aktivität der ZIKV-NS2B-NS3$^{pro}$ für das viel verwendete Substrat Bz-Nle-Lys-Lys-Arg-AMC ermittelt. Eine mögliche Ursache hierfür ist die Ausbildung einer Salzbrücke zwischen Asp83* der S2-Tasche mit dem P2-Rest des Substrates.

Schließlich wurde die Aktivität der Protease für verschiedene FRET-Substrate getestet. Die verwendeten Substrate enthielten sowohl mutmaßliche Schnittstelle des viralen Polyproteins als auch eine mögliche Schnittstelle im STING-Protein (stimulator of interferon genes), die durch die ZIKV-NS2B-NS3-Protease geschnitten werden könnten. Letztere Spaltung könnte den Mechanismus zur Unterdrückung der angeborenen Immunantwort des Wirtsorganismus durch das Virus erklären.

# 1 Introduction

## 1a Coronavirus part

### 1a.1 Coronaviruses

Coronaviruses have first been isolated almost 80 years ago, from chickens with infectious bronchitis (Beaudette et al., 1937), and later from pigs with transmissible gastroenteritis (Doyle et al., 1946), and mice suffering from severe hepatitis and neurological disorders (Cheever et al., 1949; Perlman, 1998). However, as long as several decades later, these viruses were only grouped together due to their widely spaced, club-shaped spikes extending from the virion surface, as revealed by electron microscopy. These viruses are morphologically totally different from the ortho- and paramyxoviruses (Siddell et al., 1983). The bundles of spikes surrounding the virion surface, giving the virus particle, the appearance of a solar corona, led to the new virus group being named coronavirus (Tyrell et al., 1968).

Coronaviruses are enveloped and positive-stranded RNA viruses belonging to the subfamily *Coronavirinae* of the family *Coronaviridae*, which is the largest group within the *Nidovirales*, an order also including the A*rteriviridae* and *Roniviridae* families (Fauquet et al., 2005; Lauber et al., 2012). Originally, coronaviruses were sorted into three groups according to serological relationships and, subsequently, on the basis of phylogenetic clustering (van Regenmortel et al., 2000). As more and more novel coronaviruses have been identified, the subfamily was divided into four genera: alpha-, beta-, gamma-, and deltacoronavirus (Table 1a.1; King et al., 2012; Masters & Perlman, 2013). Almost all of the alpha- and betacoronaviruses infect mammals, whereas gammacoronaviruses are mostly isolated from avian hosts. Recently, a novel genus called *deltacoronavirus* has been detected in mammals and terrestrial birds (Dong et al., 2007; Woo et al., 2009; Chu et al., 2011).

**Table 1a.1 The four coronavirus genera with representatives**

| Virus | Abbreviation | GenBank codes | Host |
|---|---|---|---|
| **Genus *Alphacoronavirus*** | | | |
| Transmissible gastroenteritis virus | TGEV | AJ271965 | Pig |
| Feline coronavirus type I | FeCoV I | EU186072 | Cat |
| Feline coronavirus type II, Feline infectious peritonitis virus | FeCoV II, FIPV | AY994055 | Cat |
| Canine coronavirus | CCoV | GQ477367 | Dog |
| Human coronavirus 229E | HCoV-229E | AGT21344.1 | Man |
| Human coronavirus NL63 | HCoV-NL63 | ABE97129.1 | Man |
| Porcine epidemic diarrhea virus | PEDV | AF353511 | Pig |
| *Rhinolophus* bat coronavirus HKU2 | *Rh*-BatCoV HKU2 | EF203067 | Bat |
| *Scotophilus* bat coronavirus 512 | *Sc*-BatCoV 512 | DQ648858 | Bat |
| *Miniopterus* bat coronavirus 1 | *Mi*-BatCoV 1 | EU420138 | Bat |
| *Miniopterus* bat coronavirus HKU8 | *Mi*-BatCoV HKU8 | EU420139 | Bat |
| **Genus *Betacoronavirus*** | | | |
| **Lineage A** | | | |
| Human coronavirus HKU1 | HCoV-HKU1 | AGW27879.1 | Man |
| Mouse hepatitis virus | MHV | AY700211 | Mouse |
| Rat coronavirus | RCoV | FJ938068 | Rat |
| Bovine coronavirus | BCoV | U00735 | Cow |
| Human coronavirus OC43 | HCoV-OC43 | AY903460 | Man |
| **Lineage B** | | | |
| Human severe acute respiratory syndrome coronavirus | SARS-CoV | AGT21317.1 | Man |
| Severe acute respiratory syndrome–related *Rhinolophus* bat | SARSr-*Rh*-Bat CoV HKU3 | DQ022305 | Bat |

| | | | |
|---|---|---|---|
| coronavirus HKU3 | | | |
| Severe acute respiratory syndrome–related *Rhinolophus* bat coronavirus Rp3 | SARSr-*Rh*-Bat CoV Rp3 | DQ071615 | Bat |
| Severe acute respiratory syndrome-like *Rhinolophus sinicus* bat coronavirus-RsSHC014 | bat SL-CoV-RsSHC014 | KC881005 | Bat |
| Severe acute respiratory syndrome-like *Rhinolophus sinicus* bat coronavirus-Rs3367 | bat SL-CoV-Rs3367 | KC881006 | Bat |
| Severe acute respiratory syndrome-like *Rhinolophus sinicus* bat coronavirus-WIVI | bat SL-CoV-WIV1 | KC881007 | Bat |
| **Lineage C** | | | |
| *Tylonycteris* bat coronavirus HKU4 | *Ty*-BatCoV HKU4 | EF065505 | Bat |
| *Pipistrellus* bat coronavirus HKU5 | *Pi*-BatCoV HKU5 | EF065509 | Bat |
| Middle-East respiratory syndrome coronavirus | MERS-CoV | AMO03400.1 | Man |
| **Lineage D** | | | |
| *Rousettus* bat coronavirus HKU9 | *Ro*-BatCoV HKU9 | EF065513 | Bat |
| **Genus *Gammacoronavirus*** | | | |
| Infectious bronchitis virus | IBV | AJ311317 | Chicken |
| Turkey coronavirus | TuCoV | EU022526 | Turkey |
| Beluga whale coronavirus SW1 | SW1 | EU111742 | Whale |
| **Genus *Deltacoronavirus*** | | | |
| Bulbul coronavirus HKU11 | HKU11 | FJ376619 | Bulbul |
| Thrush coronavirus HKU12 | HKU12 | NC011549 | Thrush |
| Munia coronavirus HKU13 | HKU13 | NC011550 | Munia |

**Note:** Only one representative GenBank code number is given for each species; in many cases, multiple genomic sequences for a given virus are available.

Previously, coronaviruses were studied mainly as agents of upper respiratory tract disease with low mortality. However, in 2003, the outbreak of SARS, caused by a new coronavirus named SARS-CoV, with a fatality rate of about 10% turned the coronaviruses into a research hotspot (Drosten et al., 2003; Ksiazek et al., 2003; Kuiken et al., 2003; Peiris et al., 2003). Ten years later, a novel human coronavirus, MERS-CoV with 35% fatality was identified in patients from Saudi Arabia and other countries of the Arab peninsula (Zaki et al., 2012). Infection with MERS-CoV causes fever, cough, shortness of breath, pneumonia, gastrointestinal symptoms including diarrhea, and often renal failure (Eckerle et al., 2013; Hilgenfeld & Peiris, 2013). So far, according to the latest update by the World Health Organization (WHO) on 27th October 2016, there were 1806 cases confirmed globally, with 643 deaths (http://www.who.int/emergencies/mers-cov/en/; last accessed: 27th Oct, 2016). The emergence of these life-threatening novel viruses highlights the importance and urgency of developing effective and broad-spectrum antiviral drugs.

### 1a.2 Virion morphology and structural proteins

Coronavirus virions are enveloped and roughly spherical particles with a diameter ranging from 50 to 200 nm. An electron micrograph of a MERS-CoV particle is shown in Fig. 1a.1A. As schematically shown in Fig. 1a.1B, the overall structure of coronaviruses compromises 4 structural proteins: the spike glycoprotein (S), the envelope protein (E), the membrane glycoprotein (M), and the nucleocapsid phosphoprotein (N). The long spike proteins are present on all coronaviruses, whereas the short spikes consisting of the hemagglutinin-esterase (HE) are only present in some coronaviruses, especially in members of betacoronavirus clade A and also the gammacoronavirus TCoV (Dea &Tijssen, 1988).

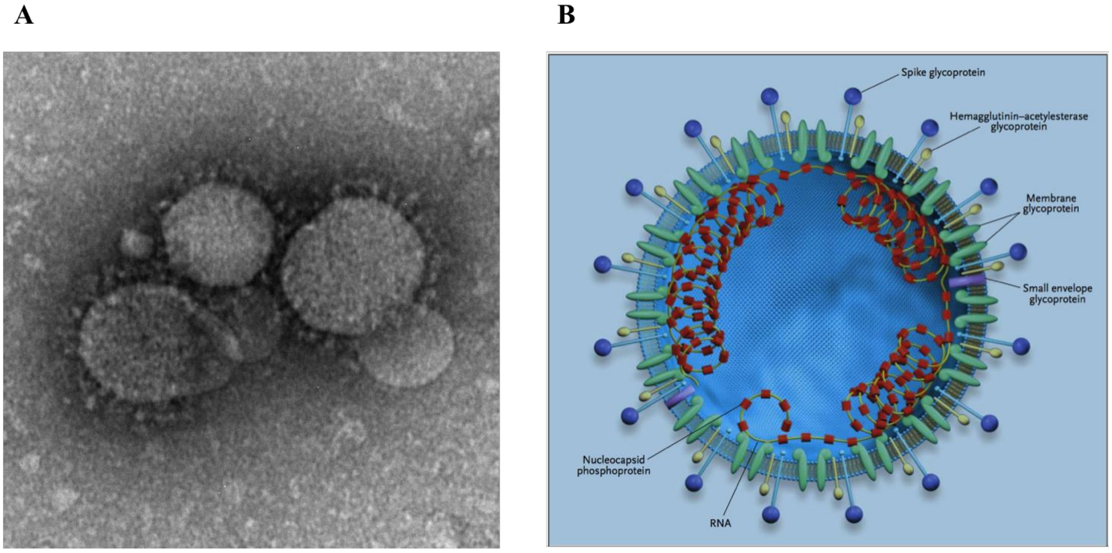


**Fig. 1a.1 Coronavirus structure.**
**(A)** Electron micrograph image of MERS-CoV visualized by negative staining (http://www.cdc.gov/coronavirus/mers/photos.html; last accessed: 27th Oct, 2016)
**(B)** Schematic presentation of the coronavirus particle (http://www.nejm.org/doi/-full/10.1056/NEJMp030078; last accessed: 27th Oct, 2016).

The S glycoprotein forms the large, petal-shaped spikes on the surface of the virion and has multiple functions. Some portions of the S protein are responsible for binding to specific receptors on the surface of susceptible cells (Li, 2016). The small membrane-embedded E protein plays an important role in virion assembly. In addition, studies have shown that the E protein has multiple functions beyond virus assembly, i.e. ion channel activity *in vitro*, and inhibition of the host cell stress response (Torres et al., 2015; Ruch & Machamer, 2012). The M glycoprotein, the most abundant structural protein in coronaviruses, gives the virion envelope its shape. It plays an important role in virus assembly, and is also critical for RNA packaging into nucleocapsids, by specifically interacting with the viral RNA packaging signal (de Haan et al., 1999; Kuo & Masters, 2002). The phosphoprotein N interacts with viral genomic RNA, integrating it into the viral nucleocapsid to help the formation of replication complexes and virus assembly (McBride et al., 2014).

## 1a.3 Genome organization of coronaviruses

With a size ranging from 26 to 32 kb, the coronavirus genome is the largest of all RNA viruses, including the segmented-genome RNA viruses. The genome is 5′-terminally capped and polyadenylated at the 3’-terminus. The positive-stranded RNA molecule can directly act as messenger RNA (mRNA). In addition to this, the coronavirus RNA also can serve as the template for RNA replication and transcription, and as the substrate for packaging into progeny. At the 5’ end of the genome, there is an untranslated region (UTR) of 200 to 400 nucleotides following a 65- to 98-nucleotide sequence termed the leader RNA. Similarly, another UTR of 200 to 400 nucleotides is located at the 3’ end, followed by the poly-adenine (A) tail. Both of these two UTRs are important for RNA replication and transcription. Divided into two open reading frames (ORF1a and ORF1b), ORF1, comprising two thirds of the genome, encodes many components of the replication/transcription complex of CoVs. Translation of ORF1b proteins depends on a ribosomal frame-shifting signal located at the junction of ORF1a and ORF1b. These ORFs are translated as a polyprotein precursor, which will be processed by two proteases into mature proteins with different functions. Most of the coronaviruses with the exception of IBV, SARS-CoV, and MERS-CoV, encode two papain-like proteases ($PL^{pro}$), as parts of nonstructural protein 3 (Nsp3). $PL^{pro}$s cleave at three sites in the N-terminal one-third of the polyprotein precursor to release the mature Nsp1, Nsp2, and Nsp3 (Fig. 1a.2). The other protease encoded by ORF1a is a chymotrypsin-like protease, which is called 3C-like protease or main protease, due to its crucial role of processing the polyprotein precursor at no less than 11 cleavage sites (Fig. 1a.2). All the structural and accessory proteins are encoded by the other 7 to 14 ORFs following the ORF1 within the 3'-proximal one-third of the genome (for reviews see: Brian & Baric, 2005; Liu et al., 2014; Plant et al., 2010; Ziebuhr, 2005).

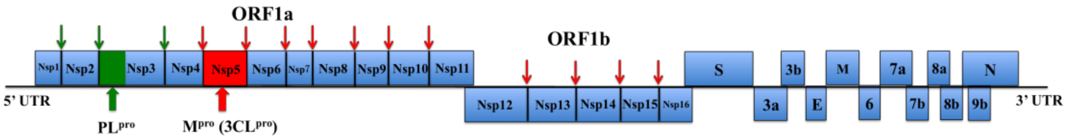


**Fig. 1a.2 Schematic representation of the SARS-CoV RNA genome.** The green and red arrows illustrate the cleavage sites of the PL$^{pro}$ and M$^{pro}$, respectively.

## 1a.4 The M$^{pro}$ of coronaviruses

### 1a.4.1 Overall structure of the M$^{pro}$ (free enzyme)

In 2002, the first crystal structure of a coronavirus protein, the TGEV M$^{pro}$, was determined by Prof. Rolf Hilgenfeld's group (Anand et al., 2002; PDB entry 1LVO). Thereafter, several crystal structures of inhibitor-free coronavirus M$^{pro}$s have been reported (Anand et al., 2003; PDB entry 1P9S; Hilgenfeld, 2014; Ma, Xiao et al., unpublished data; PDB entry 2YNA; Ho et al., 2015; PDB entry 5C3N; Needle et al., 2015; PDB entry 4WMD, 4WME, and 4WMF; Tan et al., 2005; PDN entries 2BX3 and 2BX4; Xue et al., 2008; PDB entry 2Q6D; Yang et al., 2003; PDB entry 1UJ1, 1UK2, and 1UK3; Zhao et al., 2008; PDB entry 3D23). The crystal structure of the SARS-CoV enzyme revealed that the M$^{pro}$ forms a homodimer with two protomers oriented to each other at almost a right angle (Fig. 1a.3A). Each protomer includes 3 domains, the N-terminal domain I consisting of residues 8-101, domain II made up from residues 102 to 184, and the C-terminal domain III formed by 201-306. Domains I and II together adopt a chymotrypsin-like fold, each with an antiparallel β-barrel. Domain III is composed of five α-helices arranged into a largely globular cluster. The substrate-binding site is located in the cleft between the domains I and II.

Coronavirus M$^{pro}$s share similar features with the enterovirus 3C$^{pro}$, therefore they are also called 3CL$^{pro}$s. The 3C$^{pro}$s from Coxsackievirus B3

(CVB3, Tan et al., unpublished; PDB entry 3ZYD), enterovirus D68 (EV-D68, Tan et al., 2013; PDB entry 3ZV8), and enterovirus-A71 (EV-A71, Lu et al., 2011; PDB entry 3SJK) also have a chymotrypsin-like fold, which is a typical feature for all of the reported structures of $M^{pro}$s and $3C^{pro}$s. The $3C^{pro}$s are responsible for processing the enterovirus polyprotein. Different to coronavirus $M^{pro}$s, $3C^{pro}$s lack the third α-helical domain involved in the dimerization, as they act as a monomer (Fig. 1a.3B). In addition, the $3C^{pro}$s have a classic Cys…His…Glu/Asp catalytic triad, while the $M^{pro}$s only feature a catalytic Cys…His dyad. However, regardless of the differences between these two groups of proteases, they share a very important common cleavage preference: the almost absolute requirement for Gln at the P1 position. This offers a great chance to develop broad-spectrum drugs against coronaviruses and enteroviruses.

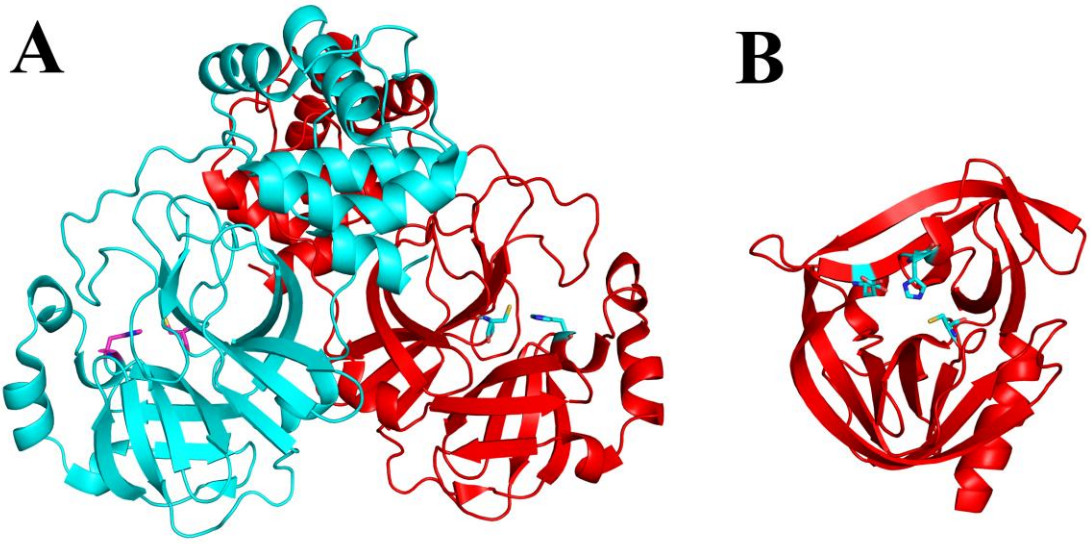


**Fig. 1a.3 Crystal structures of SARS-CoV $M^{pro}$ (Tan et al., 2005; PDB entry 2BX4) and CVB3 $3C^{pro}$ (Tan et al., unpublished data; PDB entry 3ZYD). The structures are on the same scale.**

**(A)** The SARS-CoV $M^{pro}$ is a homodimer, with each monomer comprising three domains. The two protomers are represented in red and blue, respectively. Catalytic residues are indicated by sticks.

**(B)** The structure of the CVB3 $3C^{pro}$ resembles the N-terminal two domains of the SARS-CoV $M^{pro}$, and the monomer is represented in red. Catalytic residues are indicated by sticks.

### 1a.4.2 Substrate-binding site and amino-acid preference at the cleavage site of $M^{pro}$

In the coronavirus $M^{pro}$s, the cysteine of the Cys…His catalytic dyad (Hilgenfeld, 2014) performs the nucleophilic attack onto the carbonyl carbon of the scissile bond and the histidine acts as general acid/base catalyst. A water molecule occupies the position of the third component of the catalytic triad, usually an aspartic acid residue in the serine proteases. The covalent tetrahedral intermediate between the cysteine and the substrate is formed with the assistance of the so-called ‘oxyanion hole’ when the nucleophilic cysteine attacks the carbonyl carbon of the scissile bond of the substrate. The oxyanion hole is comprised of the main-chain amides of Gly143 and Cys145 in SARS-CoV $M^{pro}$.

According to the analysis of the polyprotein cleavage sites of $M^{pro}$s (Table 1a.2), the P1 position is dominantly occupied by Gln, which indicates that the glutamine residue is an almost absolute requirement. In all the coronavirus $M^{pro}$s, the imidazole ring of conserved His163 is in the suitable position to interact with the P1 glutamine. In addition to this, two other conserved amino-acid residues, Phe140 (or Phe 139) and Glu166 (or Glu165) are also critical for the S1 pocket. The other structure elements involved in the S1 pocket are the main-chain atoms of Pro51, Met165, Glu166, and His172 in SARS-CoV $M^{pro}$ (in HCoV-229E $M^{pro}$, they are Ile51, Leu164, Glu165, and His171; Anand et al., 2003).

The hydrophobic S2 pocket is formed by the side-chains of Met49, Pro52, Tyr154, Met165, and Gln189 in SARS-CoV $M^{pro}$. Furthermore, the main-chain atoms of residues 187-189 also contribute to the S2 subsite. Hence, the substrate specificity for the P2 residue is for hydrophobic side-chains, mainly leucine, but also methionine or phenylalanine.

In the structure of the coronavirus $M^{pro}$, no obvious S3 pocket can be detected. This agrees with to the complex structure of TGEV $M^{pro}$ with a substrate analogue, chloromethyl ketone (CMK) inhibitor (PDB entry: 1P9U), where the P3 side-chain is oriented towards to the bulk solvent (Anand et al., 2003). Therefore, there is no specificity for the P3 of the substrate, as can also be seen in Table 1a.2.

The S4 pocket of the SARS-CoV $M^{pro}$ is constituted of Met165, Leu167, Gln189, and Ala191 (Leu164, Leu166, Ser189, and Gln191, as well as main-chain atoms of Ser189 in HCoV-229E $M^{pro}$), offering a congested cavity with a preference for small amino acids such as Ser, Thr, Val, or Pro.

The S1' pocket is formed by Leu27, His41, and Ala46 (Leu27, His41, and Thr47 in the HCoV-229E $M^{pro}$ S1' pocket), and small amino acids such as Ser, Ala, Gly, or even Cys can be easily accommodated in this pocket.

**Table 1a.2 Cleavage sites in the polyproteins of M$^{pro}$s from different CoVs**

| Cleavage site | Alphacoronavirus | | Betacoronavirus | | | |
|---|---|---|---|---|---|---|
| | **HCoV NL63** | **HCoV 229E** | **HKU1** | **SARS CoV** | **MERS CoV** | **HKU9** |
| Nsp4-Nsp5 | NSTLQ↓SGLKK | GSTLQ↓AGLRK | TSFLQ↓SGIVK | SAVLQ↓SGFRK | SGVLQ↓SGLVK | SAALQ↓AGLTR |
| Nsp5-Nsp6 | GVNLQ↓SGKVI | GVNLQ↓SGKTT | GVKLQ↓SKTKR | GVTFQ↓GKFKK | GVVMQ↓SGVRK | GVKLQ↓GKFQS |
| Nsp6-Nsp7 | ISTVQ↓SKLTD | VSTVQ↓SKLTD | VSQIQ↓SKLTD | VATVQ↓SKMSD | VAAMQ↓SKLTD | VSTIQ↓SNMTD |
| Nsp7-Nsp8 | SSTLQ↓SVASS | DSILQ↓SVASS | STVLQ↓ALQSE | RATLQ↓AIASE | PSVLQ↓ATLSEF | NSVLQ↓AVASE |
| Nsp8-Nsp9 | VVKLQ↓NNEIM | VVKLQ↓NNEIM | NAVMQ↓NNELM | AVKLQ↓NNE LS | AVKLQ↓NNEIK | PVKLQ↓NNELM |
| Nsp9-Nsp10 | TIRLQ↓AGKQT | TVRLQ↓AGKQT | TIRLQ↓AGVAT | TVRLQ↓AGNAT | TVRLQ↓AGSNT | TVRLH↓AGSAT |
| Nsp10-Nsp11 | RTTIQ↓SVDIS | RTAIQ↓SFDSS | SVAVQ↓SKDLN | EPLMQ↓SADAS | AALPQ↓SKDSN | EINLQ↓ARDEC |
| Nsp12-Nsp13 | STILQ↓AAGLC | STVLQ↓AAGLC | SAVMQ↓SVGAC | HTVLQ↓AVGAC | PTTLQ↓AVGSC | SATLQ↓SVGTC |
| Nsp13-Nsp14 | HADLH↓SSQVC | MTDLQ↓SESSC | LPRLH↓CTTNL | VATLQ↓AENVT | KLQSQ↓IVTGL | KPKLQ↓SEVTV |
| Nsp14-Nsp15 | ETNLQ↓SLENI | DVNLQ↓GLENI | FTTLQ↓SLENV | FTRLQ↓SLENV | FTKVQ↓GLENI | FVKLQ↓SLENV |
| Nsp15-Nsp16 | YPQLQ↓SAEWK | YPQLQ↓SAEWK | YPKMQ↓ATNDW | YPKLQ↓ASQAW | YPRLQ↓ASADW | YPQLQ↓SKQDW |

The cleavage sites are indicated by arrows.

Noncanonical amino acids at the P1 position are indicated in red.

The GenBank codes for the strains of the CoVs are as follows: HCoV NL63: ABE97129.1; HCoV 229E: AGT21344.1; HKU1: AGW27879.1; SARS-CoV: AGT21317.1; MERS-CoV: AMO03400.1; HKU9: ABN10910.1

### 1a.5 Peptidic or peptidomimetic inhibitors

In the past ten years, many inhibitors have been reported against the coronavirus $M^{pro}$s, as a result of intensive research efforts triggered by the outbreak of SARS in 2003. Moreover, a large number of complex structures have been determined of these inhibitors with the $M^{pro}$s. Although such a big amount of inhibitors have been tested and published, none of them have been licensed or approved for the treatment of coronavirus infection (Hilgenfeld & Peiris, 2013). As a result, we were insufficiently prepared when MERS emerged in 2012. This emphasized again that it is absolutely necessary to develop non-toxic broad-spectrum compounds against these viruses. Different types of warheads have been applied to achieve good inhibitory activity against the CoVs, including the α-ketoamides (Kim et al., 2012, 2016; Prior et al., 2013), α, β-unsaturated esters/Michael acceptors (Ghosh et al., 2005, 2007; Lee et al., 2009; Shie et al., 2005; Yang et al., 2005), aldehydes (Akaji et al., 2011; Al-Gharabli et al., 2006; Schmidt et al., 2008; Zhu et al., 2011), α,β-epoxyketones (Lee et al., 2005, 2007; Goetz et al., 2007), halomethylketones (Anand et al., 2003; Bacha et al., 2008, Yang et al., 2003), nitriles (Chuck et al., 2013), and phthalhydrazide ketones (Yin et al., 2007; Zhang et al., 2007).

Some of the inhibitors with warheads such as halomethylketones are too active to be developed into drugs, because this would lead to considerable side effects (Hilgenfeld, 2014). While the nitrile group is not particularly electrophilic in contrast to the former warhead, it is quite robust and not readily metabolized (Fleming et al., 2010).

As in this part of work, the studies were focused on the α-ketoamides and aldehydes, these two types of compounds will be introduced in detail. α-Ketoamides have been successfully developed into drugs, e.g. telaprivir and boceprivir for the treatment of Hepatitis C virus (HCV) infection.

These compounds are peptidomimetic inhibitors of the HCV NS3/NS4A serine protease (Lin et al., 2006; Romano et al., 2012; Venkatraman et al., 2006). However, the α-ketoamides were only occasionally described as inhibitors for viral cysteine proteases (see Discussion for details).

### 1a.5.1 Aldehydes

Peptide aldehydes have been developed as inhibitors for several classes of proteolytic enzymes, including the serine proteases (Barabas et al., 1993; McConnel et al., 1993), cysteine proteases (Graybill et al., 1994; McConnel et al., 1993; Thornberry et al., 1992), and aspartic proteases (Sarubbi et al., 1993; Fehrentz et al., 1984). The aldehyde warhead can react with the cysteine proteases to yield a hemithioacetal mimicking the structure of the transition state of the peptidlytic reaction. They can reversibly bind to the active site sulfhydryl of cysteine proteases.

A set of 1230 natural and non-natural amino acids from the available chemicals directory (ACD) as well as from the Sigma-Aldrich catalogue were docked into the S1, S2, and S4 subsites of the crystal structure of SARS-CoV M$^{pro}$ (Al-Gharabli et al., 2006). According to the docking results, glutamine received a higher score for the S1 pocket, amino acids with an aromatic ring were superior to Leu for the S2 pocket, and for the S4 pocket, uncharged polar amino acids were preferred. Based on these results, a set of peptidic aldehydes was synthesized and the inhibitory activity was determined against the SARS-CoV M$^{pro}$. Among the 36 aldehydes, four showed micromolar-range IC$_{50}$ values (Table 1a.3). Interestingly, two of the inhibitors with a Ser or Asp at the P2 position showed good inhibitory activity. In order to understand the binding mode of these inhibitors, several complex structures were determined by our former colleague (Zhu et al., 2011), demonstrating that the P1-Glu or -Phe binds to the hydrophilic S1 pocket, as well as the P2 -Ser or -Asp are located in the hydrophobic S2 subsite instead of pointing out into the solvent as assumed before (Schmidt et al., 2008). This unusual tolerance of special amino acids with respect to the P1 and P2 positions is due to the highly electrophilic character of the aldehyde warhead, which can overcome the stringent substrate specificity of the P1 and P2 pockets (Fig. 1a.4).

Aksji et al. (2011) reported a number of peptidic aldehyde compounds as SARS-CoV $M^{pro}$ inhibitors. According to their study, aldehydes were found to be more effective than compounds with the irreversible Michael acceptor warhead. A substrate-based aldehyde (Table 1a.3) was used as the starting point, different side chains of the P1, P2, and P4 sites were investigated. For the P1 position, an imidazole ring proved more effective compared with the aliphatic *iso*-butyl, *tert*-butyl, aromatic, and aliphatic rings. For the P2 site, a cyclohexane is superior to the iso-butyl or phenol group. A shortening of the penta- to tetrapeptide aldehyde will lead to reduction of the inhibitory activity by a factor of 4. This reduction can be reverted via replacement of the Ser at the P2 position by Thr. Finally, two of the inhibitors (compounds 31 and 35; Table 1a.3) showed low nanomolar-range $IC_{50}$ values against the SARS-CoV $M^{pro}$.

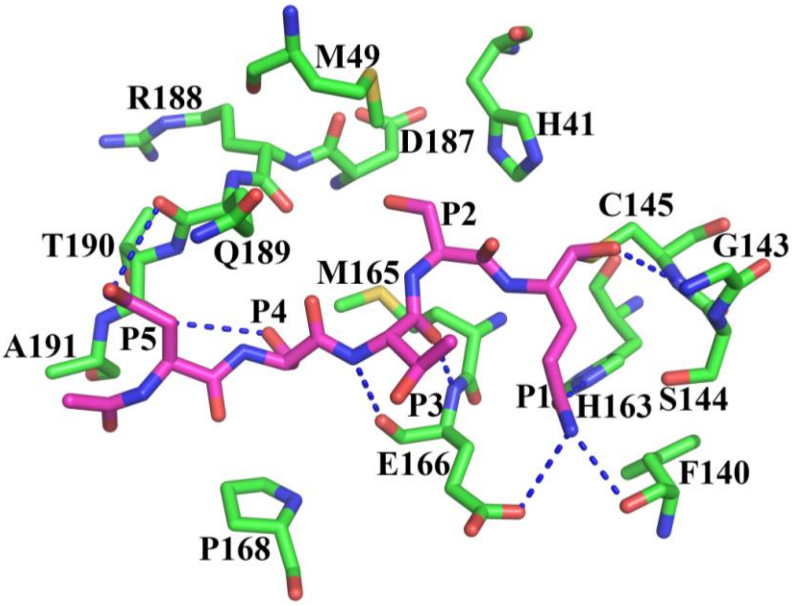
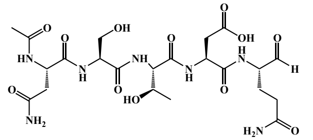
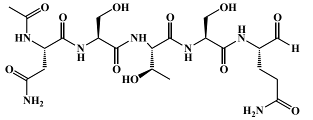
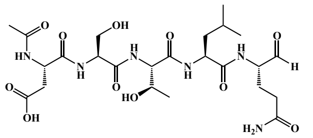
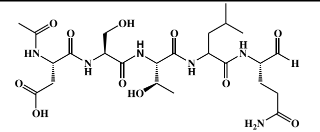
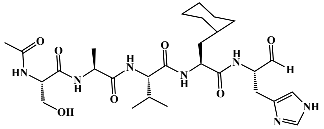
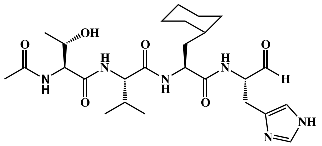


**Fig. 1a.4 Complex structure of SARS-CoV M$^{pro}$ with aldehyde Ac-NSTSQ-H.** Detailed interactions of the peptidomimetic aldehyde (Zhu et al., 2011; PDB entry 3SNC) (pink carbon atoms) with the target protease (green carbon atoms). Hydrogen bonds are depicted as blue dashed lines. Note that the non-canonical P2-Ser side-chain is embedded in the S2 subsite, without engaging in specific interactions.

**Table 1a.3 Formulae, inhibitory activities, toxicities and PDB entries of representative aldehydes**

| Compound No. | Formula | $IC_{50}$ (µM) against SARS-CoV $M^{pro}$ | PDB entry |
|---|---|---|---|
| AcNSTDQ-H<br>Al-Gharabli et al., 2006 |  | 20 | nd |
| AcNSTSQ-H<br>Al-Gharabli et al., 2006<br>Zhu et al., 2011 |  | 7.5 | 3SNC |
| AcDSTLQ-H<br>Al-Gharabli et al., 2006 |  | 10 | nd |
| AcESTLQ-H<br>Al-Gharabli et al., 2006<br>Zhu et al., 2011 |  | 7.5 | 3SND |
| 31<br>Aksji et al. (2011) |  | 0.065 | nd |
| 35<br>Aksji et al. (2011) |  | 0.098 | 3ATW |

nd: Data not done in the literature

### 1a.6 Objectives of this thesis

So far, 13 years have passed since the epidemic of SARS in 2003, but there is not a single approved drug available against coronaviruses (Hilgenfeld & Peiris, 2013). The reason is the lack of a market for the development of drugs against coronaviruses, as SARS was essentially eliminated within a few months. As a result, we were insufficiently prepared when MERS emerged in 2012 (Zaki et al., 2012). Although the registered cases of MERS (1806 cases confirmed globally, with 643 deaths; http://www.who.int/emergencies/mers-cov/en/; last accessed: 27th Oct, 2016) are not as high as with SARS (about 8000 cases in 2003), the number of infected cases and deaths is still slowly growing since the discovery of MERS-CoV in 2012. In addition, the fatality rate of MERS (about 35%) is much higher than for SARS-CoV (about 10%). In addition to the threat posted by MERS-CoV to man, some newly identified SARS-like coronaviruses from bats use the same receptor as SARS-CoV and are still circulating in bats in China (Ge et al., 2013). Public health is under threat not only from the known, but also from the unknown.

If the demand of drugs targeting CoVs from genus *betacoronavirus* is limited, a possible solution is to design broad-spectrum drugs also against members of the genus *alphacoronavirus*, i.e. HCoV-NL63 (van der Hoek et al., 2004) and HCoV-229E (Hamre & Procknow, 1966), as well as against the large genus *Enterovirus* in the *picornavirus* family (Norder et al., 2011). These members of the alphacoronaviruses mostly cause common flu, but are much more widespread than SARS-CoV or MERS-CoV. The genus *enterovirus* includes agents such as EV-D68, EV-A71, Coxsackievirus A16 (CVA16), and CVB3. Infection by these viruses can lead to mild to serious illness; i.e. EV-D68 can cause flu-like symptoms, and also be involved in rare cases with polio-like symptoms (Oermann et al., 2015; Greninger et al., 2015); CVB3 can trigger symptoms ranging from mild gastrointestinal distress to myocardic inflammation (Massilamany et al., 2014); as well as EV-A71 and CVA16 often associate with hand, foot and mouth disease (Xing et al., 2014). Enteroviruses cause clinical disease much more frequently than coronaviruses, thus antiviral drugs against both families should be commercially viable.

To date, many three-dimensional structures of M$^{pro}$s from betacoronaviruses, e.g. SARS-CoV (Hilgenfeld 2014; Tan et al., 2005; PDB entries 2BX3 and 2BX4; Yang et al., 2003; PDB entries 1UJ1, 1UK2, and 1UK3), MERS-CoV (Ho et al., 2015; PDB entry 5C3N; Needle et al., 2015; PDB entries 4WMD, 4WME, and 4WMF), HKU4 (Hilgenfeld, 2014; Ma, Xiao et al., unpublished data; PDB entry 2YNA), alpha-coronaviruses, i.e. TGEV (Anand et al., 2002; PDB entry 1LVO), and HCoV-229E (Anand et al., 2003; PDB entry 1P9S), and 3C$^{pro}$s from enteroviruses, i.e. CVB3 (Tan et al., unpublished; PDB entry 3ZYD), EV-D68 (Tan et al., 2013; PDB entry 3ZV8), and EV-A71 (Lu et al., 2011; PDB entry 3SJK) were determined. As mentioned in 1a.4.1, proteases from both families share a similar chymotrypsin-like fold and cleavage preference, in particular with the strong specificity for Gln at P1 position, enabling the design of broad-spectrum drugs against both coronaviruses and enteroviruses.

However, the major differences between the M$^{pro}$s from the genera of *alpha-* and *betacoronavirus* were not well characterized. One of the goals of this thesis was therefore to determine the crystal structure of the free enzyme of HCoV-NL63 M$^{pro}$. Furthermore, the inhibitory activities of two lead compounds were to be determined against the HCoV-NL63 M$^{pro}$. The structural and activity relationships were to be characterized by multiple studies.

Although several α-ketoamides have occasionally been reported, there was no systematic study and no complex structure available of the compounds with $M^{pro}$s or $3C^{pro}$s. In addition, the aldehydes were also not extensively optimized and even not any aldehyde had been characterized with a canonical γ-lactam at the P1 position.

A series of peptidomimetic α-ketoamides was designed by Prof. Rolf Hilgenfeld and synthesized by Dr. Daizong Lin. In this thesis, the inhibitory activities were evaluated against different coronavirus $M^{pro}$s and $3C^{pro}$s. Compounds with broad-spectrum activities were afforded by studying the structure-activity relationships (SAR) with respect to the P2 site. Moreover, five complex structures were determined between α-ketoamides with both SARS-CoV and HCoV-NL63 $M^{pro}$s to illustrate the binding modes.

In addition, a number of peptidomimetic aldehydes were described. The optimization focused mainly on the P2 and P3 positions. The SAR was studied against the SARS-CoV $M^{pro}$, from which two complex structures were determined.

## 1b Flavivirus part

### 1b.1 Flaviviruses

Yellow fever virus (YFV) was the first human virus to be discovered over one century ago by Walter Reed, who demonstrated that yellow fever could be transferred via a filterable agent in the serum of an infected individual. In addition, this agent could be transmitted to humans by mosquito bites (Strode, 1951). It is now appreciated that YFV belongs to the genus *Flavivirus* (from the Latin flavus, “yellow”). The *Flavivirus* genus belongs to the *Flaviviridae* family, which also includes the genera: *Pestivirus* and *Hepacivirus*, and *Pegivirus* (King et al., 2012).

The *Flavivirus* genus contains more than 70 viruses. Most of these viruses are transmitted from hematophagous arthropods to vertebrate hosts and are therefore considered to be dual-host viruses (Blitvich & Firth, 2015). These viruses are further differentiated into mosquito- and tick-borne viruses. In particular, the mosquito-borne viruses including dengue virus (DENV), YFV, Japanese encephalitis virus (JEV), West Nile virus (WNV), and the newly emerging Zika virus (ZIKV) are human pathogens of global significance. Different from the others, dengue virus comprises four serotypes (DENV1, DENV2, DENV3, and DENV4) (Fig. 1b.1) (Lindenbach et al., 2007; Mackenzie et al., 2004). Infection by flaviviruses can cause symptoms ranging from mild fever and malaise, to fatal encephalitis and haemorrhagic fever. In particular, the four serotypes of dengue virus transmitted by *Aedes (A.)* mosquitoes, primarily *A. aegypti*, are responsible for the highest rates of disease and mortality. Despite the generally mild disease upon DENV infection, some individuals develop severe symptoms, i.e. dengue hemorrhagic fever and dengue shock syndrome (DHF/DSS) (Lim et al., 2013; Mukhopadhyay et al., 2005). The World Health Organization estimated that each year, around 50 to 100 million people get infected with DENV in the tropics and subtropics (http://www.who.int; last accessed: 27th Oct, 2016). Recently, according to a new estimation, the annual number of infection is almost three-fold higher than the previous estimate (Bhatt et al., 2013).

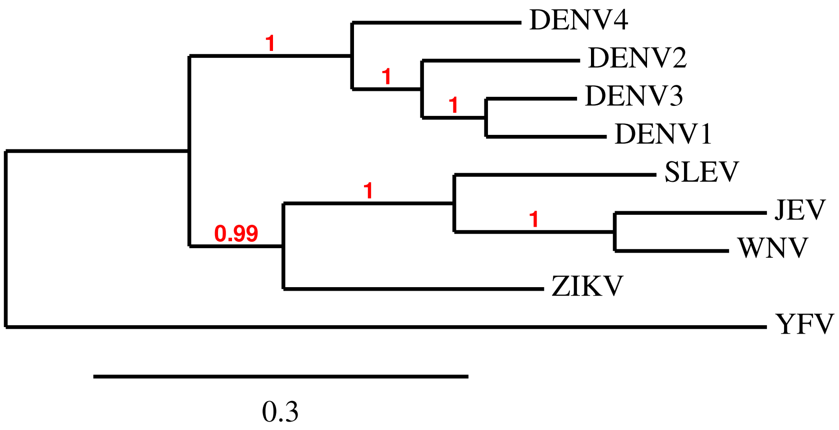


**Fig 1b.1 Phylogenetic analysis of NS3 amino-acid sequences of flaviviruses.** (A) Maximum-likelihood trees were constructed using MrBayes software (Huelsenbeck & Ronquist, 2001) with the substitution of WAG (protein) model (Whelan & Goldman, 2001). The Phylogenetic tree is shown in TreeDyn viewer (Chevenet et al., 2006). All the amino-acid sequences of the NS3 protein were downloaded from the GenBank entries with the following codes ZIKV: KU729217.2; WNV: P06935.2; DENV1: AB519681.1; DENV2: EU920850.1; DENV3: GU189648.1; DENV4: GU289913.1; SLEV: KM267635.1; YFV: AY968065.1; JEV: EF623988.1.

Previously, ZIKV was described only as causing sporadic human infections in Africa and Asia, and considered as a mild human pathogen (Bearcroft et al., 1956; Hayes, 2009; Simpson, 1964). However, since the 2013 outbreak in French Polynesiaa French overseas territory located in the South Pacific (Aubry et al., 2015), as well as the dramatic epidemic in the Americas from 2014 and still ongoing (Heukelbach et al., 2016), more studies were performed on this virus. In addition to the major infection route via mosquito bites (*Aedes* species), sexual interaction, blood transfusion, and placental transmission to the fetus are also routes of transmission (D'Ortenzio et al., 2016; Musso et al., 2015). ZIKV infection has recently been found associated with neurological disorders including disseminated encephalomyelitis (Ferreira, 2016) and Guillain-Barré syndrome (GBS) (Brasil et al., 2016; Cao-Lormeau et al.,

2016) in a significant percentage of infected individuals. Moreover, prenatal ZIKV infection is related to a drastic increase of microcephaly cases in fetuses and newborns (Rasmussen et al., 2016). ZIKV was specifically detected in fetal microcephaly brain tissue in Brazil (Mlakar et al., 2016). Furthermore, the ZIKV has also been identified in two mothers' amniotic fluid, who delivered babies with microcephaly (Calvet et al., 2016). Since 1971 till today, several studies on ZIKV in animal models were performed, from which the causative relationship between ZIKV infection and neurological disorders was confirmed, and ZIKV infection of human neural progenitor cells (hNPC) was demonstrated (Bell et al., 1971; Lazear et al., 2016; Tang et al., 2016; Zmurko et al., 2016).

Currently, there are mainly supportive options for the treatment of flavivirus infection, but only a few FDA-approved vaccines and no drugs are available, with the exception of the live-attenuated YFV vaccine (YFV strain 17D) (Pulendran, 2009; Strode, 1951), inactivated tick-borne encephalitis virus (TBEV) and JEV (Pugachev et al., 2005), and a live recombinant tetravalent dengue vaccine (Dengvaxia, Sanofi-Pasteur). Therefore, there is an urgent need for the development of antiviral drugs for the treatment of flavivirus infections.

### 1b.2 Flavivirus virion and structural proteins

The mature flavivirus particles are enveloped virions with diameter of ~50 nm (Westaway et al, 1985) that is composed of a single, positive-strand RNA of ~11-kb packaged by a capsid protein (C) (Rice et al., 1986) (Fig. 1b.2). There are two proteins located on the surface of the virus particle, i.e. the envelope glycoprotein (E) and the membrane protein (M) (Brinton, 1986).

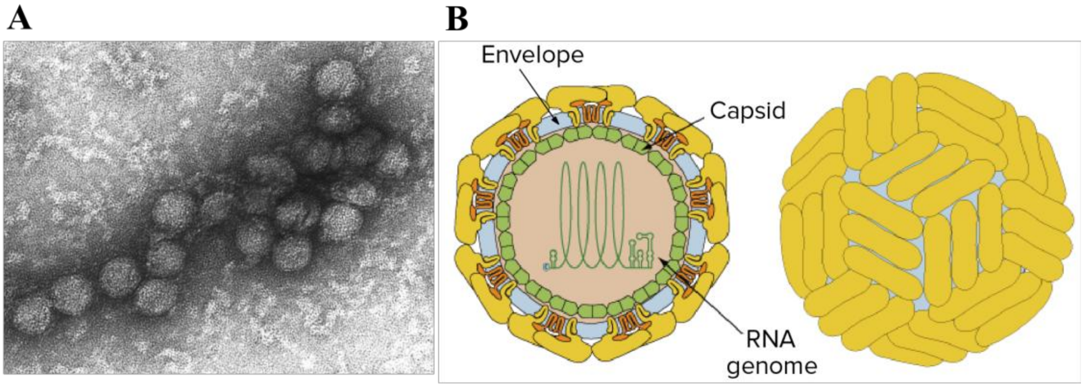


**Fig. 1b.2 Flavivirus structure. A:** An electron micrograph of the West Nile virus (http://www.cdc.gov/vhf/virus-families/flaviviridae.html, last accessed: 21th Oct, 2016) **B:** Schematic representation of the flavivirus particle (http://expasy.org/viral-zone/all_by_species/24.html, last accessed: 21th Oct, 2016)

## Capsid protein

The C protein is a ~11-kD, highly basic protein, as both N- and C-termini are rich in positively-charged amino acids for RNA-binding, connected by an internal hydrophobic region for the membrane interaction (Ma et al., 2004). For the nascent C protein (anchC), there is a hydrophobic anchor attached at the C terminal part serving as a signal peptide for endoplasmic reticulum (ER) translocation of the glycoprotein precursor of M protein (prM). Subsequently, this hydrophobic region is cleaved by the viral protease to yield the mature C protein (Lobigs, 1993). The mature C folds into a homo-dimer with proper distribution of the positively charged and hydrophobic residues for RNA and membrane binding, respectively (Jones et al., 2003).

## Membrane glycoprotein

The prM (~26 kD) is a precursor of the M protein, which is translocated into the ER membrane by the C-terminal hydrophobic region (serving as signal peptide) of the C protein. The cleavage between the signal peptide and prM by the host signalase only occurs after the generation of the mature C by viral protease processing (Lobigs, 1993). In addition, the signalase cleavage rate is influenced by the expression level of the E gene

(Lorenz et al., 2002). Therefore, the cleavage by the signalase is not regulated by the viral serine protease to decrease the production of empty virions and lower levels of infectious virus in the early stage. Thus, this mechanism serves to regulate the structural protein processing and virus production until the viral serine protease reaches sufficiently high levels in the later stage (Lobigs et al., 2004). In the immature virion, the prM folds rapidly and forms a hetero-dimer with the E protein to prevent E undergoing rearrangement to the fusogenic form at acidic condition, during the transit through the secretory pathway (Konishi & Mason, 1993; Lorenz et al., 2002). Afterwards, the prM is cleaved into pr and M fragments by furin or a related enzyme located on the Golgi, resulting in the conversion from immature to mature virion in the secretory pathway. After the cleavage, the hetero-dimer between prM and E will disassociate, and the pr fragment is released, while M forms a homodimer (Stadler et al., 1997).

### E protein

E is a type-II membrane protein with a molecular mass of ~53 kD, which mediates receptor binding and membrane fusion of the flavivirus (Nowak & Wengler, 1987; Stiasny & Heinz, 2006). The E protein is composed of three domains and forms a homodimer with head-to-tail configuration lying parallel to the lipid bilayer in a mature flavivirus virion (Rey et al., 1995). At low pH, the E protein will dissociate into monomers and then rearrange into trimers to expose the putative fusion peptide towards the membrane of the host cells (Allison et al., 1995; Bressanelli et al., 2004; Modis et al., 2004; Stiasny et al., 2004). The different carbohydrate moieties on the surface of different flaviviruses might regulate the receptor specificity (Stiasny & Heinz, 2006).

### 1b.3 Genome organization of flaviviruses

The flavivirus genome consists of a ~11-kb, single-open-reading-frame, positive-strand RNA. Similar to other positive-strand viruses, the genome possesses three functions: serving as the template for replication, as the mRNA for translation of the polyprotein, and as the substrate for packing into the virion. The genome has a 5’ cap ($m^7$GppAmG), but lacks a 3’-poly (A) tail, unlike cellular mRNA (Cleaves & Dubin, 1979; Wengler et al., 1978). In addition, there are two noncoding regions (NCR) of about 100 nucleotides (nt), and 400 to 700 nt located at the 5’ and 3’ termini, respectively (Markoff, 2003).

The single ORF is translated into a long polyprotein, which is processed by both host cell proteases (signalase and furin) and viral protease into at least 10 proteins. The order of the proteins encoded by the ORF from 5’to 3’ is: the three structural proteins (C, prM/M, and E), and 7 non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (Rice et al., 1985). The cellular signalase is responsible for cleavages between C-prM, prM-E, and E-NS1, and furin for the junction between the pr and M fragments of prM (Ruiz-Linares et al., 1989; Stadler et al., 1997). The viral protease processes junctions in the polyprotein: within C protein and signal peptide, NS2A-NS2B, NS2B-NS3, NS3-NS4A, NS4B-NS5 (Chambers et al., 1991). The schematic topology of the flavivirus polyprotein is shown in Fig. 1b.3.

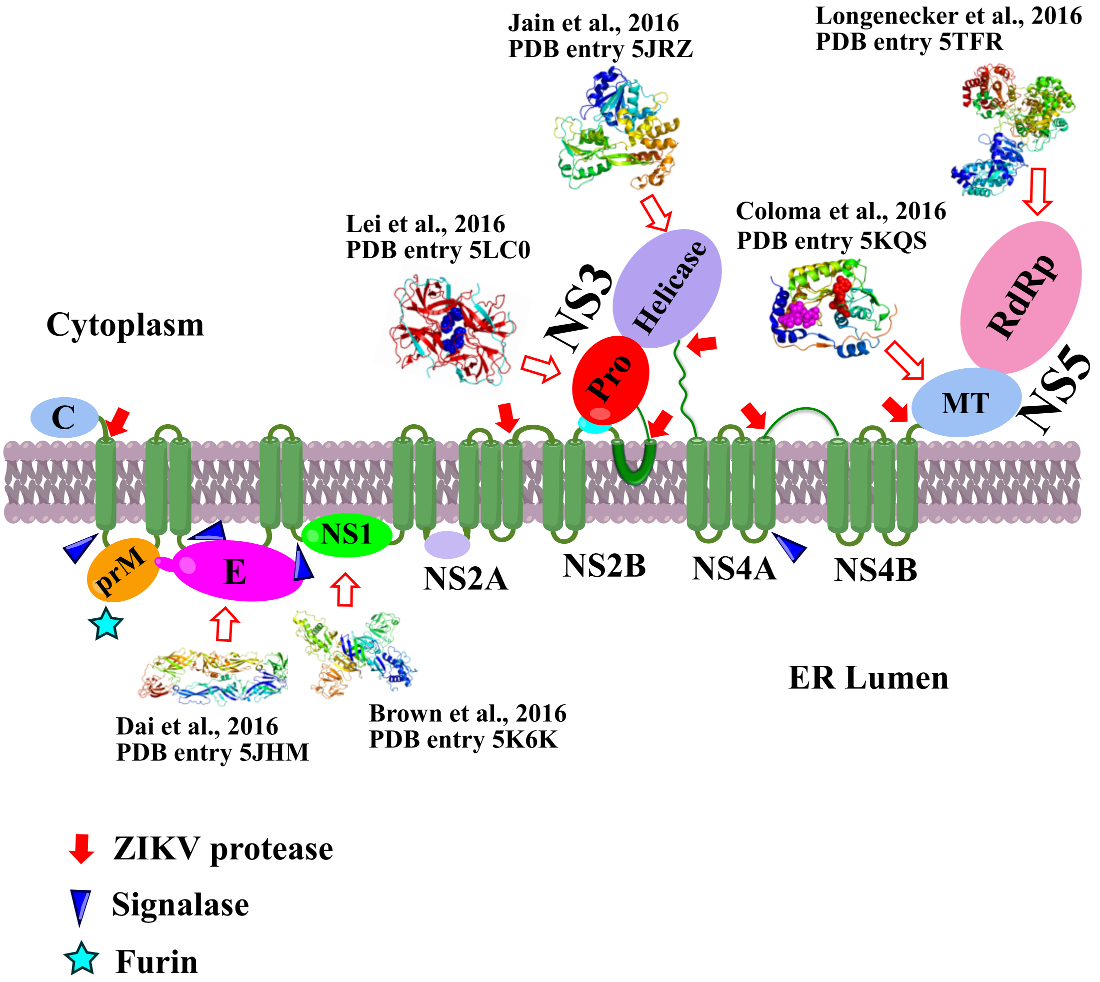


**Fig. 1b.3 Topology of the ZIKV polyprotein**

Polyprotein processing takes place at the membrane of the ER. Two host-cell proteases, furin and signalase, are involved in polyprotein processing on the lumenal side of the ER, whereas the viral NS2B/NS3 protease performs the cleavages on the cytoplasmic side. The cleavage sites of known protease are indicated by different symbols.

## 1b.4 Flavivirus protease

The flavivirus protease is composed of about 180 amino-acid residues from the N-terminal part of the NS3 protein ($NS3^{pro}$). It is a trypsin-like serine protease with a Ser-His-Asp catalytic triad (Bazan & Fletterick, 1989; Chambers et al., 1990a). The activity and solubility of the protease is dependent on a hydrophilic central domain of NS2B (amino-acid residues 49 - 95) protein as cofactor (Falgout et al., 1991; 1993).

It is hard to obtain soluble $NS3^{pro}$ alone from both recombinant protein production or via refolding. Young and colleagues designed a $Gly_4$-Ser-

$Gly_4$ ($G_4SG_4$) linker, where the NS2B cofactor was connected to the $NS3^{pro}$ as a fusion protein (Leung et al., 2001). This design greatly promoted studies on the NS2B-$NS3^{pro}$. Besides the fusion protein, an unlinked protease and the cofactor including the 50 and 180 amino-acid residues of NS2B and NS3, respectively, was generated in the active form by using a coexpression system (Kim et al., 2013). Subsequently, a number of crystal structures were determined as free enzyme form or in complex with inhibitors (Table 1b.1).

## Table 1b.1 Crystal and NMR structures of flavivirus proteases

| Virus | Protein | Conformation | PDB Entry | Ligands | Dimer | References |
|---|---|---|---|---|---|---|
| ZIKV | NS2B-NS3$^{pro}$ | Closed | 5LC0 | Boronate inhibitor | Yes | Lei et al., 2016 |
| DENV1 | NS2B-NS3$^{pro}$ | Open | 3L6P | No | Yes | Chandramouli et al., 2010 |
| DENV1 | NS2B-NS3$^{pro}$ (S135A) | Open | 3LKW | No | Yes | Chandramouli et al., 2010 |
| DENV2 | NS2B-NS3$^{pro}$ | Open | 2FOM | No | No | Erbel et al., 2006 |
| DENV2 | NS2B-NS3$^{pro}$ | Closed | 2M9P (NMR) | Bz-Nle-KR-$CF_3$ | No | Gibbs et al., 2014, unpublished |
| DENV2 | NS2B-NS3$^{pro}$ | Closed | 2M9Q (NMR) | Bz-Nle-KR-$CF_3$ | No | Gibbs et al., 2014, unpublished |
| DENV2 | NS2B-NS3$^{pro}$ | Open (pH 8.5) | 4M9M | No | Yes | Yildiz et al., 2013 |
| DENV2 | NS2B-NS3$^{pro}$ | Open (pH 5.5) | 4M9K | No | Yes | Yildiz et al., 2013 |
| DENV2 | NS2B-NS3$^{pro}$ (A125C) | Open (pH 8.5) | 4M9F | No | Yes | Yildiz et al., 2013 |
| DENV2 | NS2B-NS3$^{pro}$ (A125C) | Open (pH 5.5) | 4M9I | No | Yes | Yildiz et al., 2013 |
| DENV2 | NS2B-NS3$^{pro}$ (A125C) | Open | 4M9T | DTNB | Yes | Yildiz et al., 2013 |
| DENV3 | NS2B-NS3$^{pro}$ | Closed | 3U1I | Bz-Nle-KRR-H | Yes | Noble et al., 2012 |
| DENV3 | NS2B-NS3$^{pro}$ | Closed | 3U1J | Aprotinin | No | Noble et al., 2012 |
| DENV4 | NS2B-NS3 (full length) | -- | 2WHX | No | No | Luo et al., 2010 |
| DENV4 | NS2B-NS3 (full length) | -- | 2WZQ | No | No | Luo et al., 2010 |
| DENV4 | NS2B-NS3 (full length) | -- | 2VBC | No | No | Luo et al., 2008 |
| WNV | NS2B-NS3$^{pro}$ | Closed | 2FP7 | Bz-Nle-KRR-H | No | Erbel et al., 2006 |
| WNV | NS2B-NS3$^{pro}$ | Closed | 2IJO | Aprotinin | No | Aleshin et al., 2007 |
| WNV | NS2B-NS3$^{pro}$ (H51A) | Open | 2GGV | No | Yes | Aleshin et al., 2007 |
| WNV | NS2B-NS3$^{pro}$ | Closed | 3E90 | 2-naphthoyl-KKR-H | No | Robin et al., 2009 |

| | | | | | | |
|---|---|---|---|---|---|---|
| WNV | NS2B-NS3$^{pro}$ | Closed | 2YOL | 3,4-dichlorophenyl-acetyl-KK-GCMA | Yes | Hammamy et al., 2013 |
| JEV | NS2B-NS3$^{pro}$ | -- | 4R8T | No | Yes | Weinert et al., 2015 |
| MVEV | NS2B-NS3$^{pro}$ | -- | 2WV9 | No | No | Assenberg et al., 2009 |

-- There is no open or close conformation to be defined due to short length of NS2B.

### 1b.4.1 Overall structure of NS2B-NS3$^{pro}$

The structure of the NS3$^{pro}$ part is conserved between the flavivirus proteases, possessing a chymotrypsin-like fold with the catalytic triad located between the two β-barrels, each of which is formed by six β-strands. Crystal structures of WNV and DENV NS2B-NS3$^{pro}$ free enzymes adopt an "open" conformation, where the NS3 core is well folded and the NS2B part inserts the first β-strand into the N-terminal β-barrel of NS3; and the C-terminal part of NS2B is orientated away from the NS3 core (Erbel et al., 2006; PDB entry 2FOM; Fig. 1b.4A). However, in the complex structures with a ligand, the NS2B wraps around the NS3 core and directly interacts with the ligands' P2 and P3 substituents, forming a "closed" conformation (Noble et al., 2012; PDB entry 3U1I; Fig. 1b.4B).

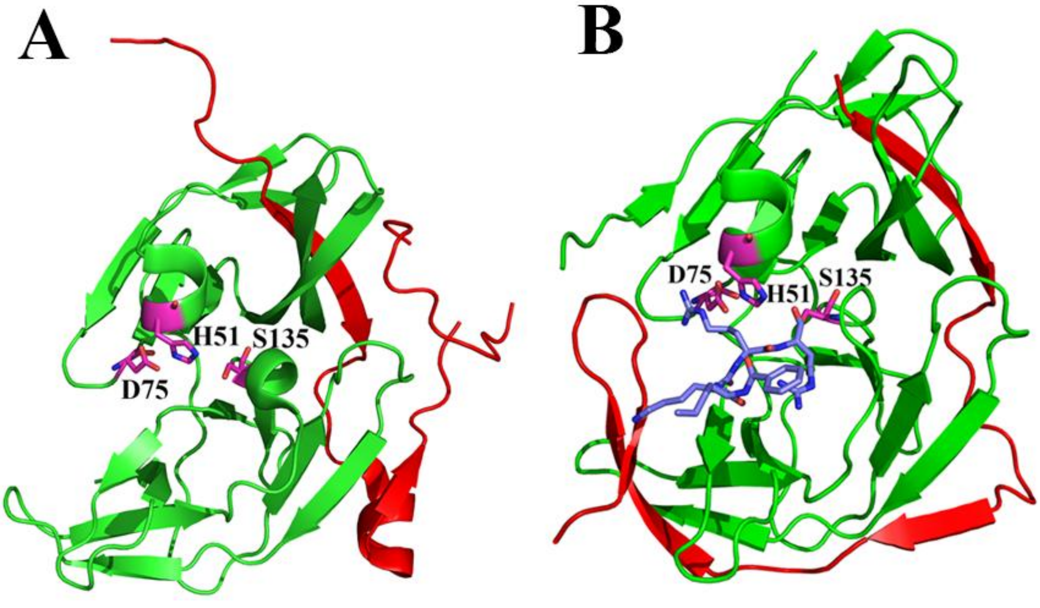


**Fig. 1b.4 Crystal structures of DENV NS2B-NS3$^{pro}$.**
**(A)** Crystal structure of DENV2 NS2B-NS3$^{pro}$ free enzyme in open conformation (Erbel et al., 2006; PDB entry 2FOM).
**(B)** Crystal structure of DENV3 NS2B-NS3$^{pro}$ in complex with an aldehyde inhibitor (Bz-Nle-KRR-H) in closed conformation (Noble et al., 2012; PDB entry 3U1I). The NS2B part is shown in red, NS3 part in green, and inhibitor in purple. Catalytic residues are indicated by sticks.

### 1b.4.2 Cleavage preferences

The flavivirus NS3$^{pro}$ prefers two basic amino acids at the P1 and P2 (Arg-Arg, Lys-Arg, Arg-Lys; occasionally, cleavage after Gln-Arg occurs, e.g. in DENV2 and DENV4) sites in the viral polyproteins (Chambers et al., 1990b; Table 1b.2). The P1' and P2' are usually occupied by amino acids with short side-chains, most commonly Gly, Ala, or Ser. Li et al (2005) studied the cleavage preference of NS2B-NS3$^{pro}$ of the four DENV serotypes by using positional scanning based on tetra- and octapeptide libraries. It was demonstrated that the NS2B-NS3 enzymes of the four DENV serotypes share very similar substrate preferences at both prime- and nonprime sites. For the nonprime sites, a strong specificity for Arg at the P1 position was identified, whereas the preference order for P2 was Arg > Thr > Gln/Asn/Lys, Lys > Arg > Asn for P3, and Nle > Leu > Lys >X (X represents any amino acid) for the P4 position. For the prime sites, the enzyme preferred small and polar amino acids at P1' and P3' (Ser), whilst P2' and P4' showed minimal activity. However, the preference of the WNV NS2B-NS3$^{pro}$ is different from DENV, with Lys more favored than Arg at P2 (Shiryaev et al., 2007). This different preference at the P2 position is due to the replacement of Ser or Thr83* (the star indicates an amino-acid residue from NS2B) in DENV by Asn84* in WNV (Chappell et al., 2006; Fig. 1b.5).

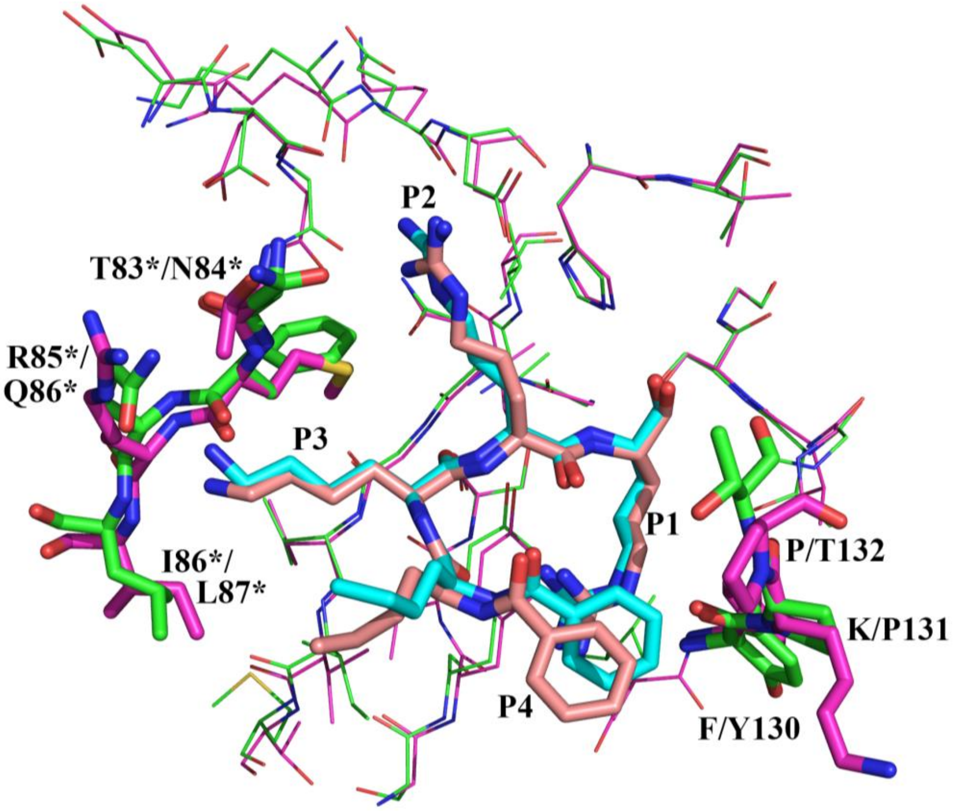


**Fig. 1b.5 Comparison of DENV and WNV protease structures.** Superimposition of the complex structures of DENV3 and WNV NS2B-NS3$^{pro}$ in complex with an aldehyde inhibitor (Bz-Nle-KRR-H) (DENV3: Noble et al., 2012; PDB entry 3U1I; WNV: Erbel et al., 2006; PDB entry 2FP7). The DENV NS2B-NS3$^{pro}$ is shown in magenta, inhibitor in orange, WNV NS2B-NS3$^{pro}$ in green, and inhibitor in blue. Amino acids that affect the specificity of the S2-S3 and S1'-S2' pockets are indicated by sticks, and the others are indicated by lines.

**Table 1b.2 Cleavage sites in the polyproteins of NS2B/NS3$^{pro}$ from different flaviviruses**

| Virus | Capsid | NS2A-NS2B | NS2B-NS3 | NS3-NS4A | NS4A-2K Peptide | NS4B-NS5 |
|---|---|---|---|---|---|---|
| ZIKV | EKKRR↓GADTS | RSGKR↓SWPPS | KTGKR↓SGALW | AAGKR↓GAAFG | PEKQR↓SPQDN | LVKRR↓GGGTG |
| WNV | KQKKR↓GGTAG | PNRKR↓GWPAT | QYTKR↓GGVLW | ASGKR↓SQIGL | PEKQR↓SQTDN | PGLKR↓GGAKG |
| DENV1 | NRRKR↓SVTML | I WGRK↓SWPLN | KKKQR↓SGVLW | AAGRR↓SVSGD | PDRQR↓TPQDN | GGGRR↓GTGAK |
| DENV2 | NRRRR↓SAGMI | TSKKR↓SWPLN | VKKQR↓AGVLW | AAGRK↓SLTLN | PEKQR↓TPQDN | TNARR↓GTGNI |
| DENV3 | NKRKK↓TSLCL | TLKRR↓SWPLN | KQTQR↓SGVLW | AAGRK↓SIALD | PEKQR↓TPQDN | GTGKR↓GTGSQ |
| DENV4 | NGRKR↓STITL | GASRR↓SWPLN | VKTQR↓SGALW | ASGRK↓SITLD | PEKQR↓TPQDN | QTPRR↓GTGTT |
| YFV | SRKRR↓SHDVL | IFGRR↓SIPVN | RGARR↓SGDVL | AEGRR↓GAAEV | PGQQR↓SIQDN | KTGRR↓GSANG |
| JEV | KQNKR↓GGNEG | PNKKR↓GWPAT | KTTKR↓GGVFW | AAGKR↓SAVSF | PEKQR↓SQTDN | PSLKR↓GRPGG |

The cleavage sites are indicated by arrows (↓).

The GenBank codes for the strains of the flaviviruses shown are ZIKV: KU729217.2; WNV: P06935.2; DENV1: P33478.1; DENV2: P29990.1; DENV3: P27915.1; DENV4: P09866.2; JEV: P19110.1; YFV: P03314.1.

## 1b.5 Flavivirus protease inhibitors

As protease inhibitors have been successfully developed into clinically available drugs against human immunodeficiency virus (HIV) (De Clercq, 2009) and Hepatitis C virus (HCV; Wyles, 2013), it is plausible that protease inhibitors might also be applied in the clinic treatment of flavivirus infections. A great number of inhibitors have been reported with both peptidomimetic and non-peptidic skeletons, either binding to the catalytic site or interrupting the interaction between NS2B and the NS3 protease domain (for reviews see: Lim et al., 2013; Luo et al., 2015; Noble & Shi, 2012).

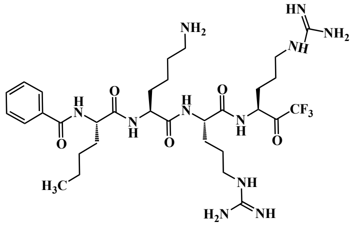
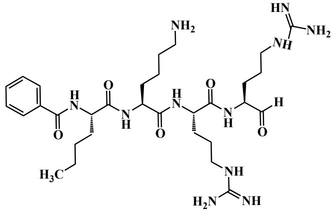
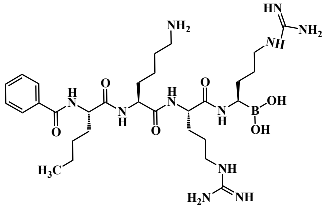
### 1b.5.1 Peptidomimetic compounds

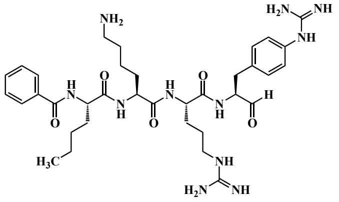
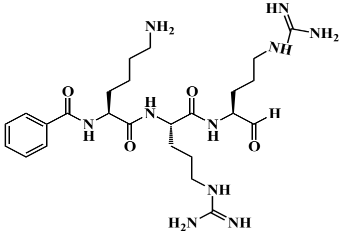
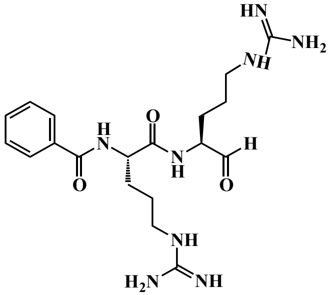
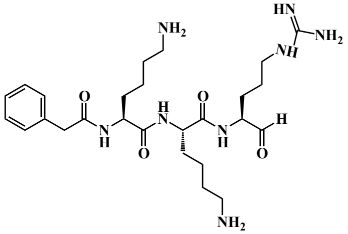
Usually, peptidic inhibitors posses electrophilic warheads that form a covalent bond with the serine of the catalytic triad. In addition, most of the reported compounds comprise at least one basic amino acid.

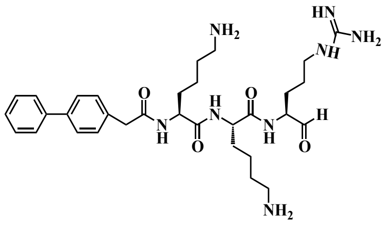
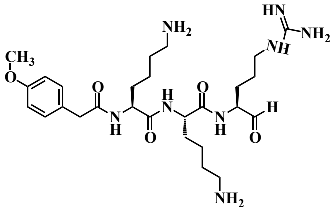
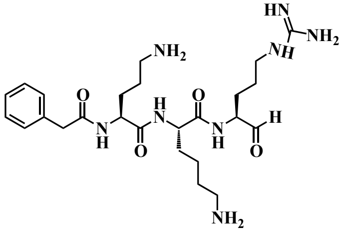
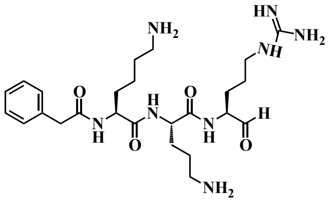
Aprotinin, also called bovine pancreatic trypsin inhibitor (BPTI), a small protein composed of 58 residues, is a highly potent inhibitor of both WNV and DENV2 protease ($K_i$ = 0.16 and 0.026 µM; Mueller et al., 2007). In addition, the complex structures of both WNV and DENV have been determined, indicating a binding model for further inhibitor design (Aleshin et al., 2007; Noble et al., 2012; Table 1b.1).

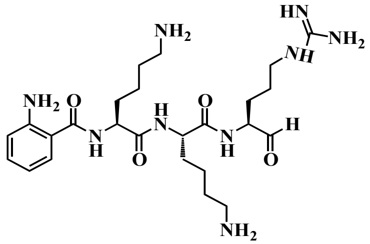
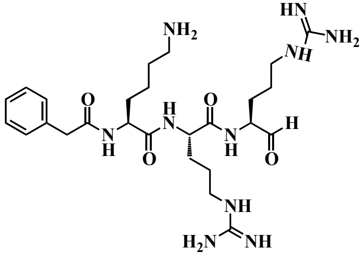
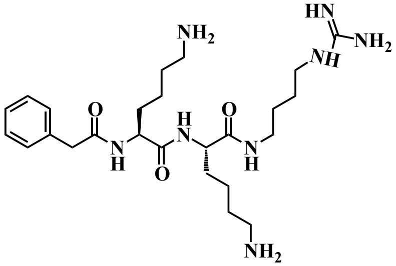
Based on two high-affinity non-prime-side substrates: Bz-Nle Lys-Arg-Arg-ACMC ($K_m$ = 12 µM) and Bz-Nle-Lys-Thr-Arg-ACMC ($K_m$ = 34 µM) identified by Li et al. (2005), various warheads were investigated against the DENV2 NS2B-NS3$^{pro}$ by Yin et al. (2006a). Among these warheads, the carboxylic acid, trifuoroacetysulfonamide, and α-ketoamide were totally inactive ($K_i$ > 500 µM), the simple amide ($K_i$ = 127.5 µM) or the α-hydroxyamide ($K_i$ = 178 µM) only showed very poor activity at high concentration. α-keto heterocycle moieties such as α-keto benzoxazole ($K_i$ = 82.9 µM) or α-ketothiazole ($K_i$ = 42.8 µM) only showed mediocre activities. Interestingly, a trifluoromethyl-ketone (compound 18, $K_i$ = 0.85 μM; Table 1b.3), an aldehyde (compound 7, $K_i$ = 5.8 μM; Table 1b.3), and a boronic acid (compound 21, $K_i$ = 0.043 μM; Table 1b.3) possessed good inhibitory activity. In particular, the boronic acid warhead demonstrated an inhibitory activity of a factor of at least 10 higher compared to other ones. Subsequently, they (Yin et al., 2006b) carried out a systematic SAR study against DENV2 NS2B-NS3$^{pro}$ based on the peptidic aldehyde compound 7 (Table 1b.3) with the substitution of amino acids by Ala, Phe, Lys, Pro, also in *D* configuration or with methylated main-chain amide. It was found that the P2-Arg played a crucial role in inhibitor binding to the enzyme. Furthermore, an inhibitor (compound 23, $K_i$ = 2.8 μM; Table 1b.3) was successfully identified at the P1 position with a (p-guanidinyl)Phe residue. The tri- and di-peptidic (inhibitor 21 and 22, Table 1b.3) aldehydes afforded low micromolar-range activity, and complex structures between the former and DENV3 NS2B-NS3$^{pro}$ (Noble et al., 2012; PDB entry 3U1I) and WNV NS2B-NS3$^{pro}$ (Erbel et al., 2006; PDB entry 2FP7) were determined to illustrate the binding mode.

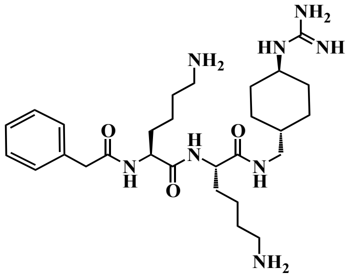
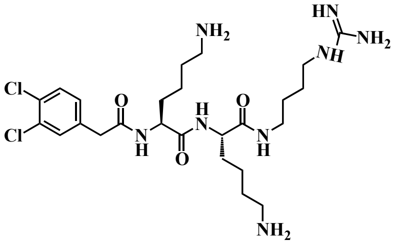
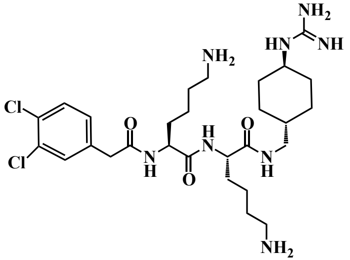
**Table 1b.3 Formulae, inhibitory activities, and PDB entries of flavivirus protease inhibitors**

| Compound | Formula | Inhibitory activity (µM) | Target | PDB entry |
|---|---|---|---|---|
| 18<br>Yin et al., 2006b |  | $K_i = 0.85$ | DENV2 NS2B-NS3$^{pro}$ | N/A |
| 7<br>Yin et al., 2006b |  | $K_i = 5.8$ | DENV2 NS2B-NS3$^{pro}$ | N/A |
| 21<br>Yin et al., 2006b |  | $K_i = 0.043$ | DENV2 NS2B-NS3$^{pro}$ | N/A |

| | | | | |
|---|---|---|---|---|
| 23<br>Yin et al., 2006a |  | $K_i = 2.8$ | DENV2 NS2B-NS3$^{pro}$ | N/A |
| 21<br>Yin et al., 2006a |  | $K_i = 1.5$ | DENV2 NS2B-NS3$^{pro}$ | 3U1I (DENV3)<br>2FP7 (WNV) |
| 22<br>Yin et al., 2006a |  | $K_i = 12.0$ | DENV2 NS2B-NS3$^{pro}$ | N/A |
| 3<br>Stoermer et al. 2008 |  | $K_i = 0.009$ | WNV NS2B-NS3$^{pro}$ | N/A |

| | | | | |
|---|---|---|---|---|
| 10<br>Stoermer et al., 2008<br>Schüller et al., 2011 |  | $K_i = 0.006$ | WNV NS2B-NS3$^{\text{pro}}$<br>(Stoermer et al. 2008) | N/A |
| | | $K_i = 0.056 \pm 0.004$ | WNV NS2B-NS3$^{\text{pro}}$<br>(Schüller et al., 2011) | |
| | | $K_i = 12.2 \pm 0.38$ | DENV2 NS2B-NS3$^{\text{pro}}$<br>(Schüller et al., 2011) | |
| 21<br>Stoermer et al. 2008 |  | $K_i = 0.011$ | WNV NS2B-NS3$^{\text{pro}}$ | N/A |
| 30<br>Stoermer et al. 2008 |  | $K_i = 0.046$ | WNV NS2B-NS3$^{\text{pro}}$ | N/A |
| 34<br>Stoermer et al. 2008 |  | $K_i = 0.013$ | WNV NS2B-NS3$^{\text{pro}}$ | N/A |

|  |  |  |  |  |
|---|---|---|---|---|
| 28<br>Stoermer et al. 2008 |  | $K_i = 0.16$ | WNV NS2B-NS3$^{pro}$ | N/A |
| 2<br>Schüller et al., 2011 |  | $K_i = 0.39 \pm 0.21$ | WNV NS2B-NS3$^{pro}$ | N/A |
|  |  | $K_i = 6.7 \pm 1.1$ | DENV2 NS2B-NS3$^{pro}$ |  |
| 4<br>Hammamy et al., 2013 |  | $K_i = 3.9$ | WNV NS2B-NS3$^{pro}$ | N/A |

| | | | | |
|---|---|---|---|---|
| 17<br>Hammamy et al., 2013 |  | $K_i = 1.2$ | WNV NS2B-NS3$^{pro}$ | N/A |
| 37<br>Hammamy et al., 2013 |  | $K_i = 0.4$ | WNV NS2B-NS3$^{pro}$ | N/A |
| 40<br>Hammamy et al., 2013 |  | $K_i = 0.13$ | WNV NS2B-NS3$^{pro}$ | 2YOL |

N/A: Data not available in the literature

Stoermer et al. (2008) reported several tripeptide aldehydes (X-KKR-H) with different groups capping the N-terminal, from which the flexible ones conferred more potent inhibition against WNV NS2B-NS3$^{pro}$, such as phenylacetyl (compound 3, $K_i$ = 9 nM; Table 1b.3), 4-phenyl-phenylacetyl (compound 10, $K_i$ = 6 nM; Table 1b.3), and 3-methoxyl-phenylacetyl (compound 21, $K_i$ = 11 nM; Table 1b.3). In addition, the replacement of Arg or Lys at the P2 or P3 position by the unnatural amino acid ornithine (Orn) resulted still in high inhibitory potency (compound 34, $K_i$ = 13 nM, compound 30, $K_i$ = 46 nM; Table 1b.3). They demonstrated that the fluorescent compound 28 ($K_i$ = 0.16 μM; Table 1b.3) could be taken up by the COS-1 cell via a charge interaction with the cell surface heparin sulfate. Moreover, the antiviral activity of compound 3 against WNV (Kunjin strain) was tested in Vero cells with an $EC_{50}$ of 1.6 μM.

Though the peptidic aldehydes have been studied against DENV and WNV separately, Schüller et al. (2011) generated a systematic SAR study against both WNV and DENV protease. A series of 17 compounds with the sequence X-KRR-H and X-KKR-H were tested, where X was a varying capping group including a large range of physical properties such as polarity, length, flexibility and bulk. The data suggested that DENV prefers Arg while WNV prefers Lys at the P2 position. The compound 11 (4-phenyl-phenylacetyl-KKR, identical to compound 10 in Stoemer et al., 2008; Table 1b.3) showed the best activity against WNV NS2B-NS3 ($IC_{50}$ = 0.056 μM). On the other hand, inhibitors with smaller aromatic (benzoyl and phenylacetyl) or aliphatic (acetyl, propionyl cyclopropionyl, and trifluoroacetyl) caps showed less potent inhibitory activity by a factor of 15 compared to compound 11. A polar cap is disfavored by the WNV NS2B-NS3$^{pro}$. The most potent inhibitor against DENV NS2B-NS3$^{pro}$ was phenylacetyl-KRR-H (compound 2, $IC_{50}$ = 6.7 μM; Table 1b.3). In summary, all the peptidic aldehydes were several-fold more potent against WNV than against DENV protease. In addition, several modifications (acetyl, benzoyl, *p*-anisoyl, and benzyl) were attached to the P2 side-chain, based on the inhibitor benzoyl-KKR-H. These modified inhibitors were totally inactive against DENV ($IC_{50}$ > 300 μM), and showed drastically decreased activity by 19-fold against WNV protease. Therefore, the S2 pockets in both DENV and WNV proteases are limited in depth, especially so in the DENV protease.

In addition to the covalently binding inhibitors, Hammamy et al. (2013) reported peptidic compounds lacking an electrophilic warhead, some of which still possessed low-micromolar inhibitory activities against WNV protease *in vitro*. The P1 position was investigated based on the inhibitor PhAc-Lys-Lys-X, where X represents the modified substituents. The agmatine (compound 4) and *trans*-(4-guanidino)cyclohexylmethyl-amide (GCMA, compound 17) substituents showed most potent inhibitory activities with $K_i$ values of 3.9 and 1.2 μM (Table 1b.3), respectively. Moreover, these arginine analogues are stable, and not easily cleaved by the protease compared with the original P1-Arg moiety. Furthermore, the P4 was modified based on the formula of X-Lys-Lys-agmatine. Among all the modified P4 groups, the dichloro-substituted (2,5-, 3,4-, and 2,3-) phenylacetyl group showed most potent inhibitory activities with $K_i$ values of 0.6, 0.4, and 0.45 μM (compound 37; Table 1b.3), respectively. Therefore, the combination of GCMA at P1 and dichloro-substituted phenylacetyl group at the P4 position offered three inhibitors (compound 40; Table 1b.3) with most potent and equivalent inhibitory activities ($K_i$ = 0.14, 0.13, and 0.12). The crystal structure of WNV NS2B-NS3$^{pro}$ in complex with compound 40 was determined; the P4 group was seen to interact with the P1 via a hydrogen bond, forming a horseshoe-like conformation of the inhibitor (Fig. 1b.6). This offered a starting point for the design of cyclic compounds against the flavivirus proteases.

In addition to the peptidic inhibitors, several studies have identified some non-peptidic inhibitors, which are summarized by recent reviews (Lim et al., 2013; Luo et al., 2015).

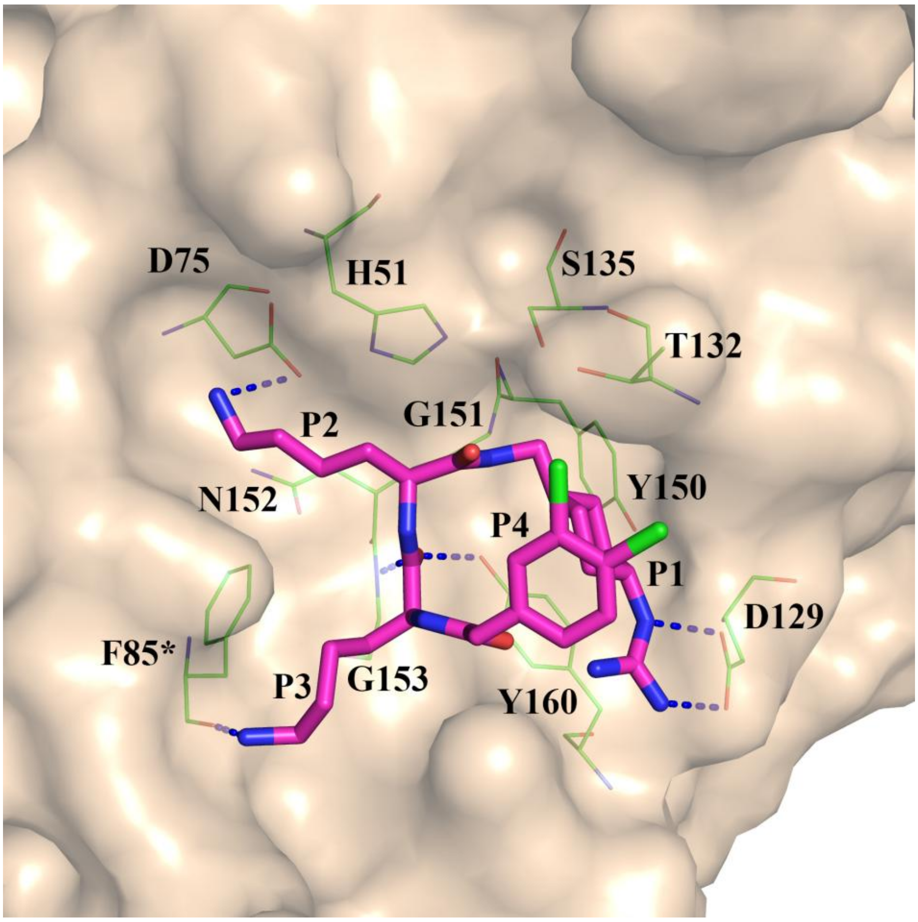


**Fig. 1b.6 Fit of the inhibitor (3,4-dichloro)phenylacetyl-KK-GCMA (Hammamy et al., 2013; PDB entry 2YOL) to WNV NS2B-NS3$^{pro}$ (wheat transparent surface) as revealed by X-ray crystallography of the complex.** Selected side-chains of the protease are shown by sticks with green carbon atoms. The compounds are indicated by sticks in magenta. Hydrogen bonds are depicted as blue dashed lines.

### 1b.6 Objectives of this part

Although several complex structures have been determined of flavivirus proteases with inhibitors, most of the inhibitors contain Arg or Lys at the P1 and P2 positions. Herein, the crystal structure of WNN NS2B-NS3$^{pro}$ in complex with a drug-like boronate (cn-716) inhibitor was determined at 1.50 Å resolution.

The peptidolytic activity and cleavage preference of the newly emerging Zika virus NS2B-NS3 protease was not characterized at all so far. In this thesis, the kinetic parameters were determined as well as the cleavage preference characterized with respect to the native cleavage sites in the polyprotein by using both AMC and Dabcyl/Edans-paired substrates. The inhibitory activity of aprotinin and boronate inhibitor was determined against the ZIKV NS2B-NS3$^{pro}$.

# 2 Material and methods

## 2.1 Material

### 2.1.1 Equipment

The equipment or instruments used in this thesis are listed in Table 2.1.

**Table 2.1 Laboratory equipment, instruments, and their manufactures**

| Instrument | Manufacturer |
|---|---|
| - 20 °C / 4 °C Refrigerator storage unit | Liebherr/Siemens |
| - 80 °C Refrigerator storage unit | REVCO – Kendro |
| AF20 Ice-machine | Scotsman |
| Agarose gel electrophoresis system | Bio-Rad |
| Analytical balance - Sartorius BP 210 D | Sartorius |
| Avanti J-26 XP Centrifuge | Beckman Coulter |
| Biophotometer ($OD_{600}$) | Eppendorf |
| Centrifuge Sigma 3-18K | Sigma |
| Concentrator 5301 | Eppendorf |
| Cryo-Express CX100 Dry shipper | Taylor-Wharton |
| Cryostream Cooler Oxford Cryosystems | Oxford Cryosystems |
| Crystallization robot “Phoenix” | Art-Robbins Instruments |
| Dynamic light-scattering | RiNA |
| Flx800 Fluorescence spectrophotometer | BioTeK |
| Innova 4230 refrigerated incubator shaker | New Brunswick Scientific |
| Microscope SZX 12 | Olympus |
| Multi-channel electronic pipet (Pipetman concept 8 × 300) | Gilson |
| PAGE gel electrophoresis system | Hoefer |
| PCR cycler | Biometra |
| pH-meter | Schott |
| Pipettes | Eppendorf |
| Protein purification systems Äkta FPLC/Prime | GE Healthcare |
| Satorius Universal balance | Sartorius |
| Single-channel electronic pipet (Pipetman concept C10) | Gilson |
| Sonificator W-250D | Branson |
| Synergy ultrapure water system | Merck Millipore |
| Systec V 150 Autoclave | Systec |
| Thermomixer | Eppendorf |
| Ultracentrifuge Centrikon T-2070 | Kontron |
| UV/Vis spectrophotometer Cary 50 | Agilent |
| Xcalibur PX Ultra X-radiation diffractometer | Oxford Diffraction |

### 2.1.2 Consumables

The consumables used in this thesis are listed in Table 2.2.

**Table 2.2 Columns, consumables, and their Manufacturers**

| Material | Manufacturer |
|---|---|
| AccuPrime *Pfx* polymerase kit | Invitrogen |
| Amicon-Ultra centrifuge filter | Millipore |
| Brand Plates 96-well Pure Grade | Brand GmbH |
| Cryo-canes | Hampton Research |
| Cryschem M plates | Hampton Research |
| Crystal Clear Sealing Tape | Hampton Research |
| Crystal storage vials | Hampton Research |
| Crystallization screens | Hampton Research |
| Dialysis bag | Spectrum Laboratories |
| Erlenmeyer flasks | Schott Duran |
| Filter membranes (0.45 µm/0.22 µm) | Millipore |
| GeneJet gel extraction kit | Thermo Fisher Scientific |
| GeneJet plasmid miniprep kit | Thermo Fisher Scientific |
| GeneRuler 1 kb DNA Ladder | Fermentas |
| GSTrap FF (5 ml) | GE Healthcare |
| HiLoad Superdex 200 prep grade | GE Healthcare |
| HiLoad Superdex 75 prep grade | GE Healthcare |
| HisTrap FF (5 ml) | GE Healthcare |
| HiTrap Q FF (5 ml) | GE Healthcare |
| Intelli-Plate-96-well | Art Robbins Instruments |
| Magnetic basic crystal caps | Hampton Research |
| Magnetic crystal wand | Hampton Research |
| Microtools | Hampton Research |
| Nylon Cryo-loops | Hampton Research |
| Rapid DNA ligation kit | Thermo Fisher Scientific |
| Screw-capped glass bottles | Schott Duran |
| Tips and tubes | Sarstedt |
| Unstained protein molecular weight marker (#SM0431) | Fermentas |
| UV cuvettes (plastic/quartz) | Sarstedt |

### 2.1.3 Chemicals

The chemicals used in this thesis are listed in Table 2.3.

**Table 2.3 Chemicals used in this thesis**

| Chemical | Manufacturer |
|---|---|
| 2-Methyl-2,4-pentanediol (MPD) | Fluka |
| Acrylamid | Sigma |
| Agar | Gerbu Biotechnik |
| Agarose | Biozym |
| Ammonium chloride | Sigma-Aldrich |
| Ammonium sulfate | Merck |
| Ampicillin sodium | Gerbu Biotechnik |
| APS | Merck |
| Bis-Tris | Sigma-Aldrich |
| CHAPS | Sigma |
| Chloramphenicol | Gerbu Biotechnik |
| Dimethylsulfoxid (DMSO) | Merck |
| Dithiothreitol (DTT) | Gerbu Biotechnik |
| Ethylenediaminetetraacetic acid (EDTA) | Gerbu Biotechnik |
| Glycerol | Gerbu Biotechnik |
| HEPES | Roth |
| Imidazole | Merck |
| Isopropyl-β-D-thiogalactopyranoside (IPTG) | Gerbu Biotechnik |
| Lithium sulfate monohydrate | Sigma-Aldrich |
| MES | Gerbu Biotechnik |
| $MgCl_2$ | Roth |
| Polyethylene glycol (PEG) 200, 1000, 3350, 6000 | Sigma-Aldrich |
| Sodium chloride | Roth |
| Sodium citrate tribasic dihydrate | Sigma-Aldrich |
| Sodium Formate | Merck |
| TEMED | Serva |
| Tris (tris(hydroxymethyl)aminomethane) | Gerbu Biotechnik |
| Tris (2-carboxyethyl) phosphine (TCEP) | Sigma |
| TY-Medium | Roth |

## 2.2. Methods

All the pH values were adjusted at room temperature.

### 2.2a Coronavirus part

### 2.2a.1 Recombinant production and purification of SARS-CoV M$^{pro}$

The recombinant production and purification of SARS-CoV M$^{pro}$ with authentic N and C termini was described in detail previously (Xue et al., 2007; Verschueren et al., 2008; Zhu et al., 2011).

### 2.2a.2 Cloning of HCoV-NL63 M$^{pro}$ wild-type (wt) and mutants

**Table 2.4 Primers used for the cloning**

| **Primer** | **Sequence (5' to 3')** |
|---|---|
| HCoV-NL63 M$^{pro}$ (wt) forward | CGCGGATCCAATTCCACCTTACAATCT |
| HCoV-NL63 M$^{pro}$ (wt) reverse | CCGCTCGAGTTAGTGGTGGTGGTGGTGGT<br>TAACACCATACATCTGCTTC |
| HCoV-NL63 M$^{pro}$ (+M49) forward | CCATCAACCACTATGGTTCTTATTGATTA |
| HCoV-NL63 M$^{pro}$ (+M49) reverse | TAATCAATAAGAACCATAGTGGTTGATGG |
| HCoV-NL63 M$^{pro}$ (P189Q) forward | GATGACCAACAGAGTTTGCAAG |
| HCoV-NL63 M$^{pro}$ (P189Q) reverse | CTTGCAAACTCTGTTGGTCATC |

Restriction sites and mutated codons are indicated in red.

The HCoV-NL63 M$^{pro}$ (GeneBank code: FJ211861.1) comprises 303 amino-acid residues, corresponding to Ser2940 – Gln3242 of pp1a. For simplicity, the Ser2940 was renumbered to Ser1. The DNA sequence encoding HCoV-NL63 M$^{pro}$ was amplified by the polymerase chain reaction (PCR) with the HCoV-NL63 M$^{pro}$ (wt) forward and reverse primers. The resulting PCR product was digested by the restriction enzymes BamHI and XhoI generating the sticky ends, which were ligated into the expression vector pGEX-6P-1 using the same restriction sites. For the single mutant clones (HCoV-NL63 M$^{pro}$ (+M49, P189Q), site-directed mutagenesis was done by using *Pfx* DNA polymerase (Invitrogen) with corresponding primers. For the double mutant clone (HCoV-NL63 M$^{pro}$ (+M49, P190Q), site-directed mutagenesis was performed by using the gene sequence of HCoV-NL63 M$^{pro}$ (+M49) as the template, and P189Q mutant primers for the PCR amplification. All constructs were verified by sequencing (MWG Biotech). All the HCoV-NL63 $M^{pro}$ clones were designed with an autocleavage site of the $M^{pro}$ itself to generate the authentic N-termini and a C-terminal penta-histidine tag ($His_5$) without protease cleavage site.

### 2.2a.3 Recombinant production and purification of HCoV-NL63 $M^{pro}$ (wt and mutants)

The sequence-verified recombinant plasmids of HCoV-NL63 $M^{pro}$ were transformed to the *Escherichia coli* strain BL21-*Gold* (DE3) (Novagen). Transformed cells were incubated at 37°C overnight in 2×YT medium with ampicillin (100 $\mu g \cdot ml^{-1}$). Next day, the pre-culture was used to inoculate into the large-scale medium growing at 37°C with vigorous shaking until the $OD_{600}$ reached 0.6-0.8. Afterwards, the overexpression of the $M^{pro}$ gene was induced by adding 0.2 mM IPTG for 10 h at 25°C. Subsequently, the cells were harvested by centrifugation at 7300 x *g* at 4°C for 30 min. The cells were stored at -80°C or lysed for immediate purification.

All the purification procedures were performed at 4°C. Bacterial cells were suspended with 60 ml buffer A (20 mM Tris-HCl, 150 mM NaCl, pH 7.8) and lysed by sonication on ice. The lysate was separated by centrifugation at 48,000 *g* and 4°C for 1h to get a clear supernatant. The supernatant was loaded onto a 5-ml HisTrap™ FF column (GE Healthcare) equilibrated with buffer A. The column was washed with buffer A plus 20 mM imidazole until the curve was close to the baseline; this was followed by eluting with a linear gradient of buffer B (20 mM Tris-HCl, 150 mM NaCl, 500 mM imidazole, pH 7.8). Fractions containing the target protein were collected, concentrated by using an Amicon YM10 membrane (EMD Millipore) and further purified by gel filtration (HiLoad™ 16/60 Superdex 200 column, GE Healthcare) equilibrated with buffer A.

### 2.2a.4 Dynamic light-scattering measurement of HCoV-NL63 M$^{pro}$ (wt)

Measurements were performed by using a Spectroscatter 201 (RiNA GmbH) with 25 μl of HCoV-NL63 M$^{pro}$ in the buffer (20 mM Tris-HCl, 150 mM NaCl, pH 7.8). The concentration used for the protease was around 10 mg·ml$^{-1}$. Results were interpreted using the software provided by the manufacturer. Experimental errors were estimated as standard deviations calculated from 10 measurements for each sample.

### 2.2a.5 Crystallization of SARS-CoV M$^{pro}$ with α-ketoamides

Using an Amicon YM10 membrane (EMD Millipore), the purified SARS-CoV M$^{pro}$ was concentrated to 21 mg·ml$^{-1}$ in buffer A (20 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.5). Crystallization was performed by equilibrating 1 μl protein (mixed with 1 μl reservoir) against 500 μl reservoir containing 6 - 8% PEG 6000, 0.1 M MES (pH 6.0), at 20°C using the vapor-diffusion sitting-drop method. α-Ketoamide compounds (**11a** (DC401903) and **11s** (DZL10)), were dissolved in 100% DMSO at 50 mM and 200 mM stock concentrations, respectively. A crystal of the free enzyme was soaked in cryo-protectant buffer containing 20% MPD, 6% PEG 6000, 0.1 M MES, 7.5 mM **11a** (DC401903), pH 6.0, for 2 h at 20°C. Another set of free-enzyme crystals was soaked in another cryo-protectant buffer with 6% PEG 6000, 5% MPD, 0.1 M MES, 15% glycerol, 10 mM **11s** (DZL10), pH 6.0, for 2 h. Subsequently, crystals were fished and flash-cooled in liquid nitrogen prior to data collection.

### 2.2a.6 Crystallization of SARS-CoV M$^{pro}$ with aldehydes

Crystals of SARS-CoV M$^{pro}$ with aldehydes (**DC401923** and **DC401954**) were obtained using cocrystallization. The purified SARS-CoV M$^{pro}$ in buffer (20 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.5) was concentrated to 23 mg·ml$^{-1}$ and incubated with 2.5 mM

**DC401923** or **DC401954** at 20°C for 4 h. Afterwards, cocrystallizaiton was set up using the vapor-diffusion sitting-drop method at 20°C, equilibrating 1 µl protein (mixed with 1 µl reservoir) against 500 µl reservoir containing 4 - 10% PEG 6000, 0.1 M MES, pH 5.5-6.25. Crystals were incubated in the cryo-protectant containing 6% PEG 6000, 5% MPD, 0.1 M MES, 2.5 mM **DC401923** or **DC401954**, 15% glycerol, pH 6.0, and fished to be flash-cooled in liquid nitrogen prior to data collection.

### 2.2a.7 Crystallization of variants and complexes of HCoV NL63 $M^{pro}$

The wild-type HCoV-NL63 $M^{pro}$ was concentrated to 45 $mg \cdot ml^{-1}$ in buffer (20 mM Tris–HCl, 150 mM NaCl, 2 mM DTT, pH 7.8,). A basic crystallization screen was performed by using a Phoenix robot (Art Robbins) employing the sitting-drop vapor-diffusion method. The drop was composed of 0.25 µl protein and 0.25 µl reservoir to equilibrate against 75µl reservoir solution. The following commercially available kits were used: PEG/Ion™ 1 & 2 Screen, Sigma™, Index™, SaltRx™, and PEG Rx™ 1 & 2 (Hampton Research) for the basic crystallization screen. Within one day, tiny crystals were detected in the presence of 0.1 M lithium sulfate monohydrate, 0.1 M sodium citrate tribasic dihydrate, 20% PEG 1000, pH 5.5. Subsequently, the optimization was performed with equilibration of 1 µl protein-inhibitor solution mixed with 1 µl reservoir containing 0.1 M lithium sulfate monohydrate, 0.1 M sodium citrate tribasic dihydrate, 15% to 25% PEG 1000, pH 4.0 to 6.0, against 500 µl reservoir solution at 18°C. Optimized crystals were protected with cryo-buffer composed of 0.1 M sodium citrate tribasic dihydrate, 25% PEG 1000, 15% glycerol, pH 6.0. Subsequently, the crystals were flash-cooled with liquid nitrogen.

### 2.2a.8 Crystallization of variants of HCoV-NL63 M$^{pro}$ in complex with ligands

All the compounds were dissolved in 100% DMSO. Crystals of HCoV-NL63 M$^{pro}$ with **11a**, and **SG85** were obtained using cocrystallization. The concentrated HCoV-NL63 M$^{pro}$ (45 mg·ml$^{-1}$) was incubated with 5 mM **11a** or **SG85** for 4 h at 20°C, followed by setting up crystallization using the vapor diffusion sitting-drop method at 20°C with equilibration of 1 µl protein (mixed with 1 µl reservoir) against 500 µl reservoir composed of 0.1 M lithium sulfate monohydrate, 0.1 M sodium citrate tribasic dihydrate, 25% PEG 1,000, pH 6.0. The crystals were protected by a cryo-buffer containing 0.1 M lithium sulfate monohydrate, 0.1 µM sodium citrate tribasic dihydrate, 25% PEG 1,000, 15% glycerol, 2 mM **11a** or **SG85**, pH 6.0 and flash-cooled in liquid nitrogen.

Crystals of HCoV-NL63 M$^{pro}$ with **11n** (DC401908) and **11f** (DC401918), and HCoV-NL63 M$^{pro}$ (+M49, P190Q) with **11a** and **SG85** were generated by using the soaking method. Several free enzyme crystals from both wt or double mutant were soaked in the cryo-protectant buffer with 0.1 M lithium sulfate monohydrate, 0.1 M sodium citrate tribasic dihydrate, 25% PEG 1,000, 15% glycerol, 5 mM **11a**, **11f**, **11n**, or **SG85**, pH 6.0. Subsequently, the soaked crystals were flash-cooled in liquid nitrogen.

### 2.2a.9 Diffraction data collection, structure elucidation and refinement

Diffraction data from the crystals of the HCoV-NL63 M$^{pro}$ (wt) and the SARS-CoV M$^{pro}$ in complex with **11a** were collected at 100 K at synchrotron beamline PXI-X06SA (PSI, Villigen, Switzerland) using a Pilatus 6M detector (DECTRIS). A diffraction data set from the SARS-CoV M$^{pro}$ crystal with compound **11s** was collected at 100 K at beamline P11 of PETRA III (DESY, Hamburg, Germany), using the same type of detector. Diffraction data sets of SARS-CoV $M^{pro}$ with two aldehydes (**DC401923** and **DC401954**), and HCoV-NL63 $M^{pro}$ (wt and +M49, P190Q mutant) complex structures were collected at synchrotron beamline BL14.2 of BESSY II (Berlin, Germany), using an MX225 CCD detector (Rayonics).

All data sets were processed using the program XDSAPP (Krug et al., 2012) and scaled by SCALA from the CCP4 suite (Evans, 2011; Winn et al., 2011). The structure of SARS-CoV $M^{pro}$ with **11a** and **11s** was determined by molecular replacement with the structure of the complex between SARS-CoV $M^{pro}$ and **SG85** (PDB entry 3TNT; Zhu et al., unpublished) as search model, employing the MOLREP program (also from the CCP4 suite; Vagin & Teplyakov, 2010; Winn et al., 2011).

The structure of HCoV-NL63 $M^{pro}$ free enzyme was determined by the molecular replacement method, also using MOLREP (Vagin & Teplyakov, 2010) from the CCP4 suite (Winn et al., 2011), with the structure of HCoV-229E $M^{pro}$ (Anand et al., 2003; PDB entry 1P9S) as the search model. The complex structures of HCoV-NL63 $M^{pro}$ (wt) with **11a**, **11f**, **11n**, or **SG85**, and of mutant (+M49, P190Q) with **11a** or **SG85** were also determined with MOLREP, using the structure of the free enzyme as a search model. Geometric restraints for the compounds in the complex structures were generated by using JLIGAND (Lebedev et al., 2012; Winn et al., 2011) and built into $F_o$-$F_c$ difference density using the COOT software (Emsley et al., 2010). Refinement of the structures was performed with REFMAC5 version 5.8.0131 (Murshudov et al., 1997, 2011; Winn et al. 2011).

### 2.2a.10 Inhibitory activity assay for α-ketoamides, aldehydes, and Micheal acceptor compound (SG85) against CoV M$^{pro}$.

A buffer containing 20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.3, was used for all enzymatic assays. The substrate with the cleavage sites of M$^{pro}$ (indicated by the arrow, ↓), Dabcyl-KTSAV-LQ↓SGFRKME(Edans)-NH2 (95% purity; Biosyntan), were employed in the FRET-based cleavage assay, using a 96-well microtiter plate. The dequenching of the Dabcyl fluorescence due to the cleavage of the substrate as catalyzed by the proteases was monitored at 460 nm with excitation at 360 nm, using a Flx800 fluorescence spectrophotometer (BioTek). Curves of relative fluorescence units (RFU) against substrate concentration were linear for all substrates up to beyond 50 μM, excluding an influence of the inner-filter effect. Stock solutions of the compounds were prepared by dissolving them in 100% DMSO. The UV absorption of **11a** was found to be negligible at $\lambda = 340$ nm, so that no interference with the FRET signal through the inner-filter effect was to be expected. For the determination of the $IC_{50}$ of α-ketoamides or **SG85**, 0.5 μM SARS-CoV or HCoV-NL63 M$^{pro}$ (wt or mutants) were separately incubated with the compounds at various concentrations (0 to 100 μM) in reaction buffer at 37°C for 10 min. For the measurement of the $IC_{50}$ of aldehydes, 0.5 μM SARS-CoV M$^{pro}$ were incubated with the compounds at various concentrations (0 to 33 μM) in reaction buffer at 37°C for 10 min. Afterwards, the reaction was initiated by adding FRET peptide substrate at 20 μM final concentration (final volume: 50 μl). The $IC_{50}$ value was determined by using the GraphPad Prism 6.0 software (GraphPad). Measurements of enzymatic activity were performed in triplicate and are presented as the mean ± standard deviations (SD).

All enzymatic assays were performed in buffer containing 20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 1 mM TCEP, pH 7.3, using a 96-well microtiter plate. All the enzymes were freshly prepared. The same substrate as for SARS-CoV $M^{pro}$, Dabcyl-KTSAVLQ↓SGFRKME (Edans)-$NH_2$ (95% purity; Biosyntan GmbH), containing the cleavage site of the $M^{pro}$ as shown by arrow, was used in the FRET-based cleavage assay. The dequenching of the Dabcyl fluorescence due to the cleavage of the substrate as catalyzed by the HCoV-NL63 $M^{pro}$ was monitored at 460 nm with excitation at 340 nm, using a Flx800 fluorescence spectrophotometer (BioTek). The reaction was initiated by adding different final concentrations of the FRET peptide (10 to 200 µM) to a solution containing HCoV-NL63 $M^{pro}$ (0.5 µM final concentration) in a final volume of 50 µl at 37°C.

### 2.2a.11 Kinetic assays of variants of HCoV-NL63 $M^{pro}$

All kinetic assays were performed in buffer containing 20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 1 mM TCEP, pH 7.3, using a 96-well microtiter plate. All enzymes were freshly prepared. The same substrate as applied for the compound inhibition assay was used here. The reaction was initiated by adding different final concentrations of the FRET peptide (10 to 200 µM) to a solution containing HCoV-NL63 $M^{pro}$ (0.5 µM final concentration) in a final volume of 50 ul at 37°C. Initial velocities were determined from the linear section of the curve. As the saturation state for the curves for the wild-type and single M49 insertion of HCoV-NL63 $M^{pro}$ enzymes were achieved, the kinetic constants ($V_{max}$ and $K_m$) were derived by fitting the data to the Michaelis–Menten equation, $V = V_{max} * [S]/(K_m + [S])$. Then $k_{cat}$ was calculated according to the eqution $k_{cat} = V_{max}/[E]$. Since no saturation could be observed for the curves of HCoV-NL63 $M^{pro}$ (P189Q and +M49, P190Q), the data were fitted to the equation $v / [E]_{tot.} = k_{app} [S]$, where $k_{app}$ approximates $k_{cat}/K_m$ (Barretto et al., 2005; Lei et al., 2014; Wojdyla et al., 2010). A calibration curve was generated by measuring the saturated fluorescence of the peptidyl-Edans released from the substrate digested by HCoV-NL63 $M^{pro}$ (wt) in reaction buffer at concentrations ranging from 0 to 200 uM. The $K_m$ and $V_{max}$ values were determined by using the GraphPad Prism 6.0 software (GraphPad).

### 2.2b Flavivirus part

#### 2.2b.1 Recombinant production, purification, crystallization, and diffraction data collection of DENV2 NS2B-NS3$^{pro}$

The expression construct for the DENV2 NS2B-NS3$^{pro}$ generated by our former colleague Dr. Naoki Sakai was the same as described in the literature (D'Arcy et al., 2006), consisting of a 47-amino-acid hydrophilic core sequence of NS2B (amino-residues acid 1394–1440) linked via a Gly$_4$-Ser-Gly$_4$ (G$_4$SG$_4$) linker to the N-terminal 185 amino-residues acid of NS3 (amino-residues acid 1476–1660) in vector pET15b (Novagen).

The BL21-*Codon Plus* (DE3) RIL (Novagen) competent cells transformed with the expression construct for DENV2 NS2B-NS3$^{pro}$ was cultured in 100 ml 2 x YT medium supplied with 100 $\mu g \cdot ml^{-1}$ ampicillin and 34 $\mu g \cdot ml^{-1}$ chloramphenicol at 37°C overnight. The next day, 20 ml pre-culture was inoculated into 2 L 2 x YT medium with 100 $mg \cdot ml^{-1}$ and 34 $\mu g \cdot ml^{-1}$ chloramphenicol, incubating at 37°C until the OD$_{600}$ reached 0.6 - 0.8. After cooling the cultures to 20°C, 0.5 mM IPTG was added to induce the expression of the recombinant gene constructor for 16 h. The cells were harvested by centrifugation (7300 x *g*) at 4°C for 30 min, and resuspended in 50 ml buffer A (20 mM Bis-Tris, 500 mM NaCl, pH 6.3). Subsequently, the cells were lysed by sonication on ice, followed by separation of supernatant using centrifugation (48,000 x *g*) at 4°C for 1 h. The supernatant was loaded onto the pre-equilibrated 5-ml HisTrap$^{TM}$ FF (GE Healthcare) column with buffer A, and washed with four column volumes of buffer A with 20 mM imidazole and gradient eluted with 20 column volumes of buffer B (20 mM Bis-Tris, 500 mM NaCl, 500 mM imidazole pH 6.3). The N-terminal His tag was cleaved off by thrombin, while dialyzing overnight against 2 L buffer A at 4°C. The treated fraction was applied to the same HisTrap$^{TM}$ FF column to remove the cleaved His tag and uncleaved protein. The flow through from the HisTrap$^{TM}$ FF column was concentrated by using an Amicon YM10 membrane (EMD Millipore), and loaded onto a HiLoad$^{TM}$ 16/60 Superdex 75 column (GE Healthcare) equilibrated with buffer A for further purification.

The purified DENV2 NS2B-NS3$^{pro}$ was concentrated to 25 mg·ml$^{-1}$ for crystallization trials using the vapor diffusion method, with 1 µl protein and 1 µl reservoir containing 0.1 M MES, 35-50% PEG 200, pH 6.0-7.0 equilibrating against 500 µl reservoir at 20°C. The crystals were transferred to a cryo-protectant (0.1 M MES, 50% PEG 200, pH 6.2) for a few seconds and flash-cooled in liquid nitrogen. A diffraction data sets from DENV2 NS2B-NS3$^{pro}$ crystals were collected at synchrotron beamline BL14.2 of BESSY II (Berlin, Germany), using an MX225 CCD detector (Rayonics).

### 2.2b.2 Kinetic parameter determination of DENV2 NS2B-NS3$^{pro}$

Enzyme kinetics of the DENV NS2B-NS3$^{pro}$ were determined at 37°C in a reaction buffer containing 10 mM Tris-HCl, 20% glycerol, 1 mM CHAPS, pH 8.5. Initially, the DENV2 NS2B-NS3$^{pro}$ was pipetted into a 96-well plate at final concentrations of 50 nM in 25 µl buffer. Afterwards, the reaction was initiated by adding 25 µl buffer containing the substrate benzoyl-norleucine-lysine-arginine-arginine-7-amino-4-methylcoumarine (Bz-Nle-Lys-Arg-Arg-AMC; Biosyntan) at final concentrations varying from 10 to 200 µM. The fluorescence signal from released AMC was monitored at 460 nm with excitation at 360 nm, using an Flx800 fluorophotometer (BioTek). Initial velocities were determined from the linear section of the curves and the variation of the fluorescence intensity per unit of time (ΔRFU/s) was converted to the amount of the cleaved substrate per unit of time (µM/s) via a calibration curve generated from the fluorescence measurements of free AMC (concentration range: 0.1 - 2.7 µM). As saturation could be achieved, kinetic constants ($V_{max}$ and $K_m$) were derived by fitting the data to the Michaelis-Menten equation, V =

$V_{max} \times [S] / (K_m + [S])$ using the GraphPad Prism 6.0 software (GraphPad). $k_{cat}/K_m$ values were calculated according to the equation $k_{cat}/K_m = V_{max} / ([E] \times K_m)$. Triplicate measurements were taken for each data point and data are presented as mean ± SD.

### 2.2b.3 Recombinant production, purification, crystallization, and diffraction data collection of WNV NS2B-NS3$^{pro}$

Cloning and expression of a DNA construct encoding the WNV NS2B-NS3$^{pro}$ with a covalent $Gly_4$-Ser-$Gly_4$ linker between NS2B and NS3$^{pro}$ was performed as described previously, as was the purification of the protease (Hammamy et al., 2013). The construct comprised 47 residues (49-96) of NS2B and 170 residues (1-170) of NS3; the C-terminal lysine of NS2B was replaced by alanine, in order to avoid autoproteolytic cleavage in the NS2B-NS3$^{pro}$ junction region.

The purified protein in 25 mM Tris-HCl, 5% glycerol, pH 8.5, was concentrated to 49.8 $mg \cdot ml^{-1}$ by using an Amicon YM10 membrane (EMD Millipore), and mixed with **cn-716** (a dipeptide inhibitor with boronic acid warhead, provided by Prof. Christian D. Klein's group from the university of Heidelberg) at 20 mM final concentration. Following overnight-incubation at 4°C, crystallization screens were run using a Phoenix™ robot (Art Robbins), with application of the sitting-drop vapor-diffusion method at 18°C. The drops contained 0.25 µl of protein and 0.25 µl of reservoir solution, to be equilibrated against 75 µl reservoir solution. Five commercially available screening kits were used: MD1-01 & 02 (Molecular Dimensions), SaltRx™, PEG/Ion™ 1 & 2, Index™, and PEG Rx™ 1 & 2 (Hampton Research). Within two days, tiny crystals were detected in the presence of 0.2 M magnesium formate dehydrate, 20% PEG 3350, pH 7.0. Subsequent optimization involved equilibration of 1 µl protein-inhibitor solution mixed with 1 µl mother liquor containing 0.2 M magnesium formate and 20% PEG 3350, pH 7.0, against 500 µl reservoir solution at 18°C. Optimized crystals were protected by a cryo-buffer consisting of 80% reservoir, 20% glycerol, and 20 mM of **cn-716**, and flash-cooled in liquid nitrogen prior to data collection. Diffraction data were collected at 100 K at synchrotron beamline P11 of PETRA III (DESY, Hamburg), using a Pilatus 6M fast detector at $\lambda$ = 0.9919 Å.

### 2.2b.4 Structure elucidation and refinement of DENV2 and WNV NS2B-NS3$^{pro}$ structures

Diffraction data sets of both DENV2 and WNV NS2B-NS3$^{pro}$ structures were processed using the program XDSAPP (Krug et al., 2012) and scaled using SCALA from the CCP4 suite of programs (Evans, 2011; Winn et al., 2011). The structures of DENV2 NS2B-NS3$^{pro}$ and WNV NS2B-NS3$^{pro}$ with **cn-716** were elucidated by the molecular replacement method using MOLREP from the CCP4 suite (Vagin & Teplyakov, 2010; Winn et al., 2011). The structure of DENV NS2B-NS3$^{pro}$ free enzyme (Erbel et al., 2006; PDB entry 2FOM), and the structure of the complex between the WNV NS2B-NS3$^{pro}$ and the inhibitor 3,4-dichlorophenyl-acetyl-Lys-Lys-GCMA (Hammamy et al., 2013; PDB entry 2YOL) were selected as search model for the elucidation of the DENV and WNV NS2B-NS3$^{pro}$ structures, respectively. A molecular model for **cn-716** was generated by using JLIGAND (Lebedev et al., 2012; Winn et al., 2011) and built into clear $F_o$-$F_c$ difference density in the substrate-binding site of the WNV NS2B-NS3$^{pro}$ by using the COOT software (Emsley et al., 2010). Most atoms of the inhibitor were well-defined by difference density at 3$\sigma$ above the mean of this map. The two structures were refined using REFMAC version 5.8.0131 (Murshudov et al., 1997, 2011; Winn et al. 2011).

### 2.2b.5 Recombinant production, and purification of ZIKV NS2B-NS3$^{pro}$

Gene cloning of the ZIKV NS2B-NS3$^{pro}$ was performed by Jian Lei, and described in our publication (Lei et al., 2016).

The ZIKV NS2B-NS3$^{pro}$ clone codes for 47 residues (49*-95*) of NS2B and 170 residues (1-170) of NS3 connected by a Gly$_4$-Ser-Gly$_4$ linker. In order to avoid autoproteolytic cleavage, the C-terminal Arg95* of NS2B was replaced by Ala in the NS2B-NS3$^{pro}$ junction region and Arg29 of NS3 was replaced by Gly. Regardless of these two modifications, the corresponding protein is designated “wt” for simplicity.

The verified NS2B-NS3$^{pro}$ plasmid was transformed into *Escherichia coli* strain BL21-*Gold* (DE3) (Novagen). Transformed cells were grown at 37°C overnight in 100 mL 2 x YT medium with ampicillin (100 $\mu g \cdot ml^{-1}$). The culture was inoculated into 2 L 2 x YT medium the next day. When the $OD_{600}$ of the 2-L culture had reached 0.6 - 0.8, 0.5 mM (IPTG) was added for the induction of the overexpression of the ZIKV NS2B-NS3$^{pro}$ gene at 20°C for 12 h. Subsequently, the E. coli cells was harvested by centrifugation (7277 x *g*) for 30 min at 4°C. Pellets were resuspended in 50 mL buffer A (25 mM Tris-HCl, 5% glycerol, pH 8.5) and lysed by sonication on ice. The lysate was centrifuged (48384 x *g*) for 1 h at 4°C to remove the cell debris. The supernatant was loaded to a 5-ml HisTrap$^{TM}$ FF column (GE Healthcare), followed by washing with buffer A plus 20 mM imidazole until the baseline reached zero. Afterwards, the target protein was eluted with buffer B (25 mM Tris-HCl, 500 mM imidazole, 5% glycerol, pH 8.5) using a linear gradient (0 – 100 % buffer B, 20 column volumes). The target protein was cleaved by thrombin (Sigma-Aldrich) overnight at 4°C to remove the His-tag, leaving four extra residues (GSHM) at the N-terminus of the target protein. The protein was applied to the nickel column again to remove the cleaved

$His_6$ tag and protein with uncleaved His-tag. The His-tag-free protein was loaded onto a HiTrap$^{TM}$ Q FF column (GE Healthcare) and eluted with buffer C (25 mM Tris-HCl, 500 mM NaCl, 5% glycerol, pH 8.5) using 20 column volumes (100 mL) of a linear gradient ranging from 0 to 500 mM NaCl. Subsequently, the target protein was further purified by gelfiltration (HiLoad 16/60 Superdex 75 column, GE Healthcare) using buffer A.

### 2.2b.6 ZIKV NS2B-NS3$^{pro}$ Cys80Ser/Cys143Ser and Asp83*Asn site-directed mutagenesis

The site-directed mutagenesis work of the Cys80Ser/Cys143Ser was performed by Jian Lei and described in our publication (Lei et al., 2016).

The work for generation of the clone, recombinant production, and purification of ZIKV NS2B-NS3$^{pro}$ (Asp83*Asn) will be described in our manuscript (Zhang et al.) and the master thesis of Yasmin Gül, who did the experiment for her master thesis under my supervision in the Institute of Biochemistry.

### 2.2b.7 Determination of enzyme kinetics of ZIKV NS2B-NS3$^{pro}$ and WNV NS2B-NS3$^{pro}$ with a fluorogenic substrate

Enzyme kinetics were determined at 37°C in a reaction buffer containing 10 mM Tris-HCl, 20% glycerol, 1 mM CHAPS, pH 8.5, for the disulfide-bonded dimer (SS-dimer), the Cys80Ser/Cys143Ser double mutant, the Asp83*Asn variant, and for the WNV NS2B-NS3$^{pro}$. In the cases of the wild-type monomer and the Asp83*Asn variant, 1 mM TCEP was used in addition. Initially, the ZIKV or WNV NS2B-NS3$^{pro}$ was pipetted into a 96-well plate at final concentrations of 5 nM and 20 nM, respectively. Afterwards, the reaction was initiated by adding the substrate Bz-Nle-Lys-Lys-Arg-AMC (Biosyntan) dissolved in 50 µl reaction buffer to final concentrations varying from 5 to 320 µM for ZIKV protease, and from 5 to 640 µM for WNV protease. The fluorescence signal from released AMC was monitored at 460 nm with excitation at 360 nm, using an Flx800 fluorophotometer (BioTek). Initial velocities were determined from the linear section of the curves and the variation of the fluorescence intensity per unit of time (ΔRFU/s) was converted to the amount of cleaved substrate per unit of time (μM/s) via a calibration curve generated from the fluorescence measurements of free AMC (concentration range: 0.013 - 9.72 μM). As saturation could be achieved, kinetic constants ($V_{max}$ and $K_m$) were derived by fitting the data to the Michaelis-Menten equation, $V = V_{max} \times [S] / (K_m + [S])$ using the GraphPad Prism 6.0 software (GraphPad). $k_{cat}/K_m$ values were calculated according to the equation $k_{cat}/K_m = V_{max} / ([E] \times K_m)$. Triplicate measurements were taken for each data point and data are presented as mean ± SD.

### 2.2b.8 Inhibition of ZIKV NS2B-NS3$^{pro}$ (monomer) by cn-716

The same substrate and TCEP-containing buffer used for determination of the enzyme kinetics of the ZIKV NS2B-NS3$^{pro}$ (monomer) were also applied for the inhibition assay with **cn-716**. The compound was dissolved in 100% DMSO for preparation of a stock solution. For the determination of the $IC_{50}$, 5 nM protease was incubated for 10 min with different concentrations of the inhibitor varying between 6 nM and 50 μM at 37°C. The reaction was initiated by adding substrate at a concentration of 10 μM to each well at a final volume of 50 μl. The $IC_{50}$ was calculated by using the GraphPad Prism 6.0 software (GraphPad). For the determination of $K_i$, the assay was performed with different final concentrations of compound **cn-716** (0.01, 0.05, 0.25, 1.00, 3.00 μM) and substrate (5, 10, 20, 40, 80, 160, 320 μM). At each compound concentration, 5 nM protease was incubated with the compound for 10 min at 37°C. Subsequently, the reaction was initiated by addition of the corresponding concentration series of substrate described above to a final volume of 50 µl. The $K_i$ was also calculated by using the GraphPad Prism 6.0 software (GraphPad) in the competitive inhibition mode. Experiments were performed in triplicate and data are presented as mean ± SD.

### 2.2b.9 Characterization of ZIKV NS2B-NS3$^{pro}$ kinetics with various FRET-substrates

#### 2.2b.9.1 FRET substrates

The FRET substrates carried a Dabcyl quencher and an Edans donor pair (Dabcyl-KXX(K/R )R↓XXXX-Edans amide, X represents any amino acid, cleavage site indicated by arrow, ↓). The energy emitted from Edans is quenched by Dabcyl, when they are in a relatively close distance (approximately from 10 to 100 Å), resulting in low or no fluorescence signal. However, the fluorescence will increase, when the substrate is cleaved by the protease. The fluorescence intensity can be detected continuously and directly. All the FRET substrates were commercially ordered from Biosyntan (Biosyntan GmbH), and contain the putative cleavage sites of the NS2B-NS3$^{pro}$ in the ZIKV polyprotein. The purity of the substrate was analyzed by HPLC (95% purity). Free Edans was purchased from Biomol (Biomol GmbH) with 99% purity as analyzed by HPLC.

#### 2.2b.9.2 Kinetic parameter assay

The enzyme used was freshly prepared. The buffer containing 10 mM Tris-HCl, 20% glycerol, 1 mM CHAPS, 1 mM TCEP, pH 8.5 was employed for the kinetic assay of ZIKV NS2B-NS3$^{pro}$ wild-type and the D83*N mutant. This buffer omitted lacking TCEP was used for the assays of WNV NS2B-NS3$^{pro}$. Initially, the protease was added to a 96-well plate at a final concentration of 1 µM. In case of the substrates corresponding to the putative intermolecular NS3$^{pro}$ cleavage site and to the stimulator of interferon genes (STING), 50 and 5 µM ZIKV

NS2B-NS3$^{pro}$ were used to reach the saturation state, as cleavage efficiency was low. Also, in the cases of WNV NS2B-NS3$^{pro}$, the catalyzed cleavage of the substrates with the putative cleavage sites between NS2B-NS3$^{pro}$ and NS4B-NS5, 2 μM and 10 μM were applied, respectively. Afterwards, the reaction was initiated by pipetting the substrate to a final volume of 50 μl. A varying range of final substrate concentrations (2.5, 5, 10, 20, 40, 80, 160, 320 μM) were used for the experiments, among which only the reactions achieving the saturation states were chosen for data processing. The fluorescence signal from the Edans product was detected at 460 nm with excitation at 360 nm, using a Flx800 fluorophotometer (BioTek). The fluorescence intensity of free Edans at the final concentration varied from 0 to 5 μM (0, 0.31, 0.63, 1.25, 2.50, 5.00 μM) was measured at identical conditions to generate the calibration curve. The initial velocity ($V_o$) of the Edans and Dabcyl product formation was calculated by fitting the relative fluorescence units (RFU) to the free Edans for the calibration curve. Triplicate experiments were performed for each data point, and the value was presented as mean ± SD.

#### 2.2b.9.3 Inner-filter effect correction on the microtiter plate

After 30 minute monitoring of the fluorescence signal, the end point RFU was taken as f (substrate). Afterwards, 1 μl free Edans (final concentration: 5 μM) was added to each reaction point when the saturation had been achieved, and the fluorescence end-point reading was taken as f (substate + Edans). Simultaneously, a reference RFU was measured with the same concentration of free Edans in 50 μl reaction buffer as f (reference). The inner-filter correction at each substrate concentration was calculated according to the function:

Detected% = (f (substrate + Edans) – f (substrate)) / f (reference)

Finally, the corrected initial velocity of the reaction was calculated as: $V_{corr} = V_o$ / (Detected%)

Kinetic parameters were calculated by fitting the corrected initial velocity to the Michaelis-Menten equation: $V = V_{max} \times [S] / (K_m + [S])$ using nonlinear least-squares regression analysis performed with the GraphPad Prism 6.0 software (GraphPad).

# 3 Results

## 3a Coronavirus part

### 3a.1 Recombinant production of SARS-CoV $M^{pro}$

A considerable amount of His-tagged SARS-CoV $M^{pro}$ was obtained as shown in Fig. 3a.1A. However, after the first round of purification through the nickel column, the purity of the $M^{pro}$ was still not high, as indicated by SDS-PAGE (Fig. 3a.1B). After cleaving off the His-tag and further purification using a ion-exchange column, the purity was higher than 95%, and the protein exhibited a molecular mass of ~33 kDa (Fig. 3a.1A and B).

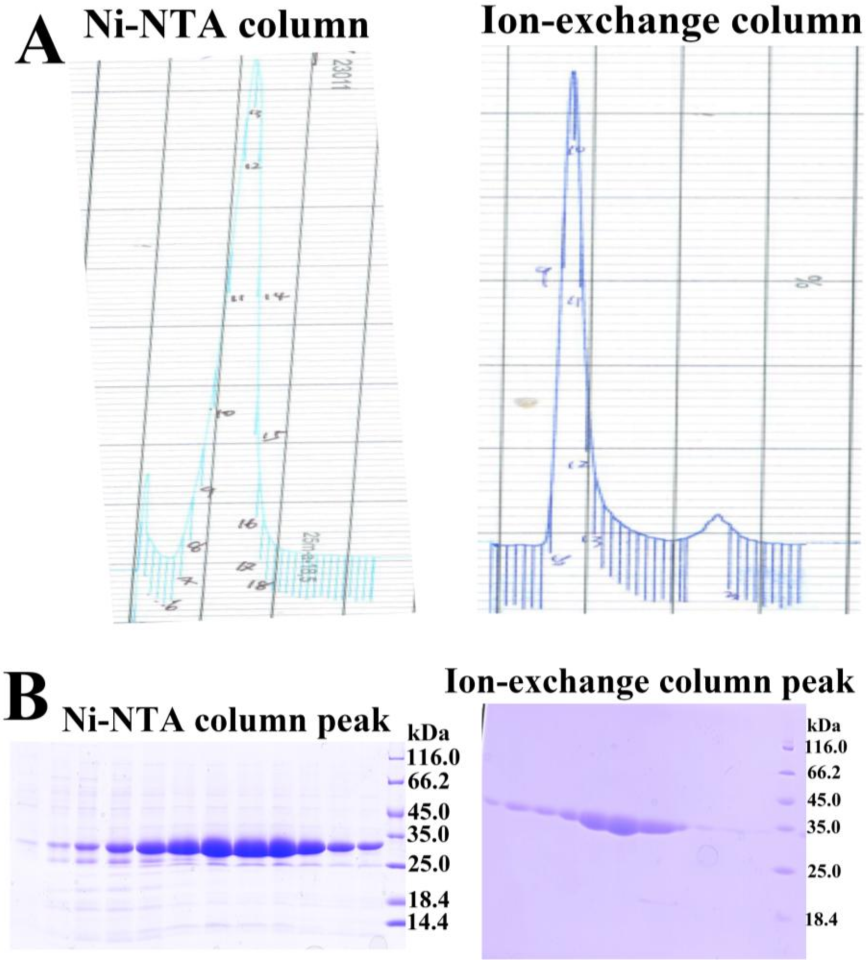


**Fig. 3a.1 Purification of SARS-CoV $M^{pro}$**

**(A)** Purification of SARS-CoV $M^{pro}$ by Ni-NTA column (left) and ion-exchange column (right), respectively.

**(B)** The chromatographic fractions from the Ni-NTA column (left) and ion-exchange column (right) were analyzed by SDS-PAGE.

### 3a.2 Recombinant production of HCoV-NL63 M$^{pro}$

As evident from the chromatographic curve and the SDS-PAGE (Fig. 3a.2A and B), the recombinant production of HCoV-NL63 M$^{pro}$ was at a high level, and a main peak showed up in the diagram for the nickel column. The further purification by the size-exclusion chromatography achieved a protein with improved purity. However, there was high-molecular-weight aggregation detected by SDS-PAGE. DLS displayed a narrow size distribution at 5.26 ± 0.04 nm indicating that the protein was homogeneous in solution (Fig. 3a.2C).

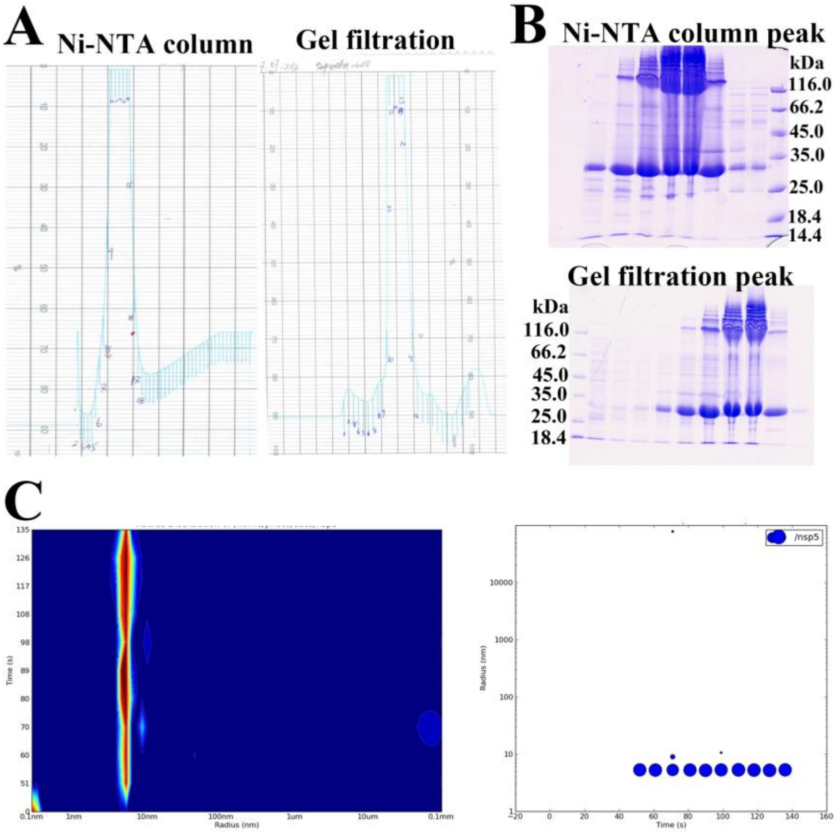


**Fig. 3a.2 Recombinant production of HCoV-NL63 M$^{pro}$**

**(A)** Purification of HCoV-NL63 M$^{pro}$ by Ni-NTA column (left) and gel filtration (right, Superdex$^{TM}$ 200) column, respectively.

**(B)** Chromatographic fractions from the Ni-NTA column (top) and gel filtration (bottom) column were analyzed by SDS-PAGE.

**(C)** DLS analysis of purified HCoV-NL63 M$^{pro}$. Left-side image: the hydrodynamic radius is depicted on the x-axis, and time on the y-axis. Right-side image: the time is depicted on the x-axis, and the hydrodynamic radius on the y-axis.

### 3a.3 Crystallization and structure elucidation of HCoV-NL63 M$^{pro}$

A crystallization hit was initially detected in the basic-screen plate using the kit PEG Rx$^{TM}$ 1 & 2 (primary condition: 0.2 M lithium sulfate monohydrate, 0.1 M sodium citrate tribasic dihydrate pH 5.0, 30% polyethylene glycol monomethyl ether 550). After optimization, crystals suitable for diffraction were obtained from the drops (Fig. 3a.3A).

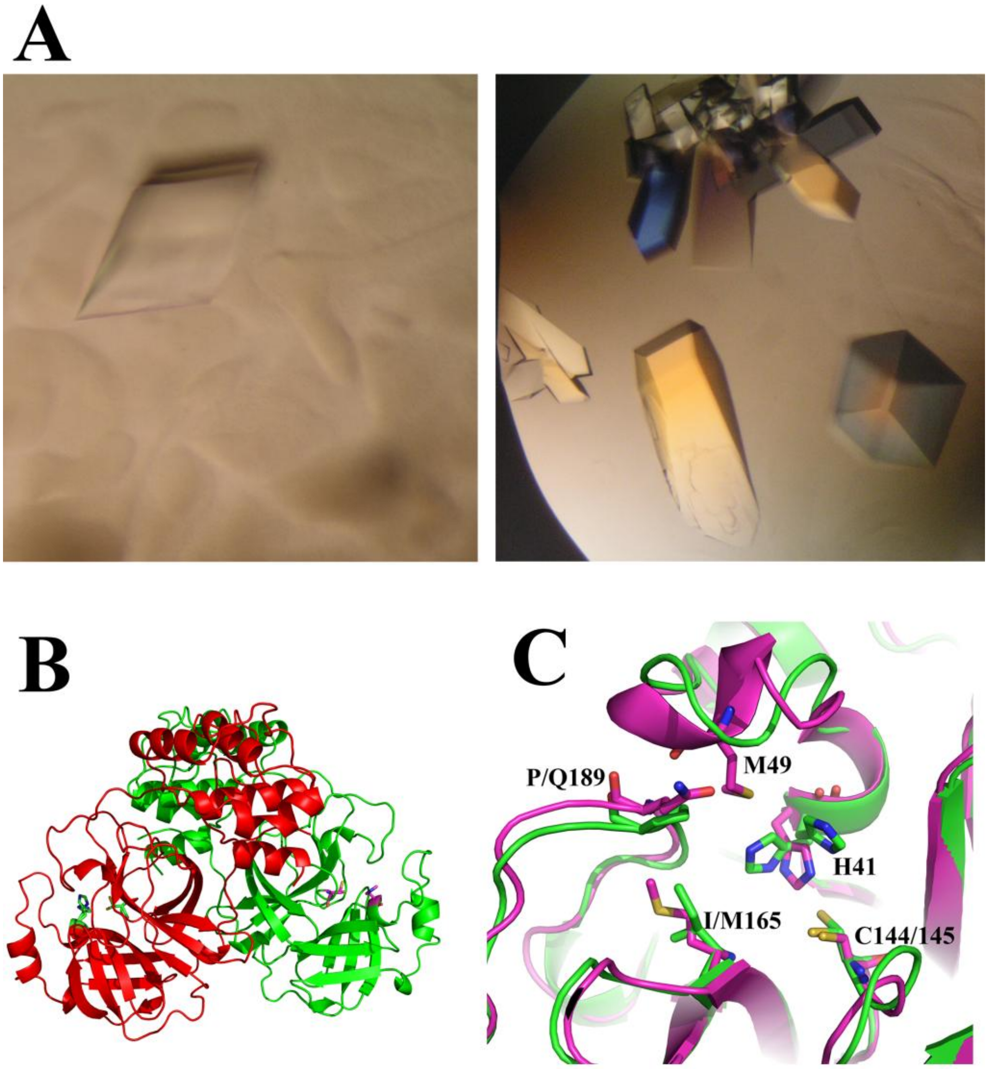


**Fig. 3a.3 Crystals and structure of HCoV-NL63 M$^{pro}$ (wt)**

**(A)** Crystals of HCoV-NL63 M$^{pro}$ (wt) growing under optimized condition.

**(B)** Crystal structure of HCoV-NL63 M$^{pro}$ (wt). The protease is a homodimer, with each monomer comprising three domains. The two protomers are represented in red and green color, respectively. Catalytic residues are indicated by sticks.

**(C)** Superimposition of HCoV-NL63 and SARS-CoV M$^{pro}$ structures (Tan et al., 2005; PDB entry 2BX4). The HCoV-NL63 M$^{pro}$ is shown in green, and SARS-CoV M$^{pro}$ in magenta. Amino-acid side-chains that affect the S2 pocket, and catalytic dyad, are indicated by sticks.

The crystal structure of the HCoV-NL63 $M^{pro}$ was determined at 1.75 Å resolution by molecular replacement, using the structure of HCoV-229E $M^{pro}$ (Anand et al., 2003; PDB entry 1P9S) as search model. The space group was determined as $C222_1$, with unit-cell parameters $a$ = 131.13 Å, $b$ = 214.43 Å, $c$ = 116.83 Å. There are three molecules in the asymmetric unit, of which chain B and C form a homodimer, and chain A interacts with a symmetry-related molecule A' from another asymmetric unit. Diffraction data statistics are listed in Table 3a.1, which will be presented together with the data collected from variants of HCoV-NL63 $M^{pro}$ in part 3a.5.

The structure has a canonical chymotrypsin-like architecture, comprising three domains (Fig. 3a.3B). Domains I (residues 11 to 99) and II (residues 100 to 180) are 6- and 7-stranded β-barrels, and together resemble chymotrypsin. The catalytic dyad is located in the cleft between the two domains. A long loop (181 to 200) connects domain II to the five-α-helix domain III (201 to 303).The root mean square deviations (rmsd) of the Cα atoms of each of the three chains from the structure of SARS-CoV $M^{pro}$ (Tan et al., 2005, PDB entry 2BX4) are 1.42 Å, 1.29 Å, and 1.34 Å, respectively. The oxyanion holes in the three chains are all in the catalytically competent conformation, which is different from the structure of SARS-CoV $M^{pro}$ crystallized at pH 6.0, which has both the active site and substrate-binding pocket collapsed in protomer B. The S1 subsite is conserved and similar to SARS-CoV $M^{pro}$. However, there is a considerable difference in the S2 pocket between the two structures. Firstly, according to the sequence alignment, Met49 is absent in HCoV-NL63 $M^{pro}$, and the Gln in position 189 is replaced by Pro, compared to SARS-CoV $M^{pro}$ (Fig. 3a.4). Moreover, due to the presence of proline in HCoV-NL63 $M^{pro}$, the loop harboring P189 is closer to the catalytic histidine compared to SARS-CoV $M^{pro}$ (Fig. 3a.3C).

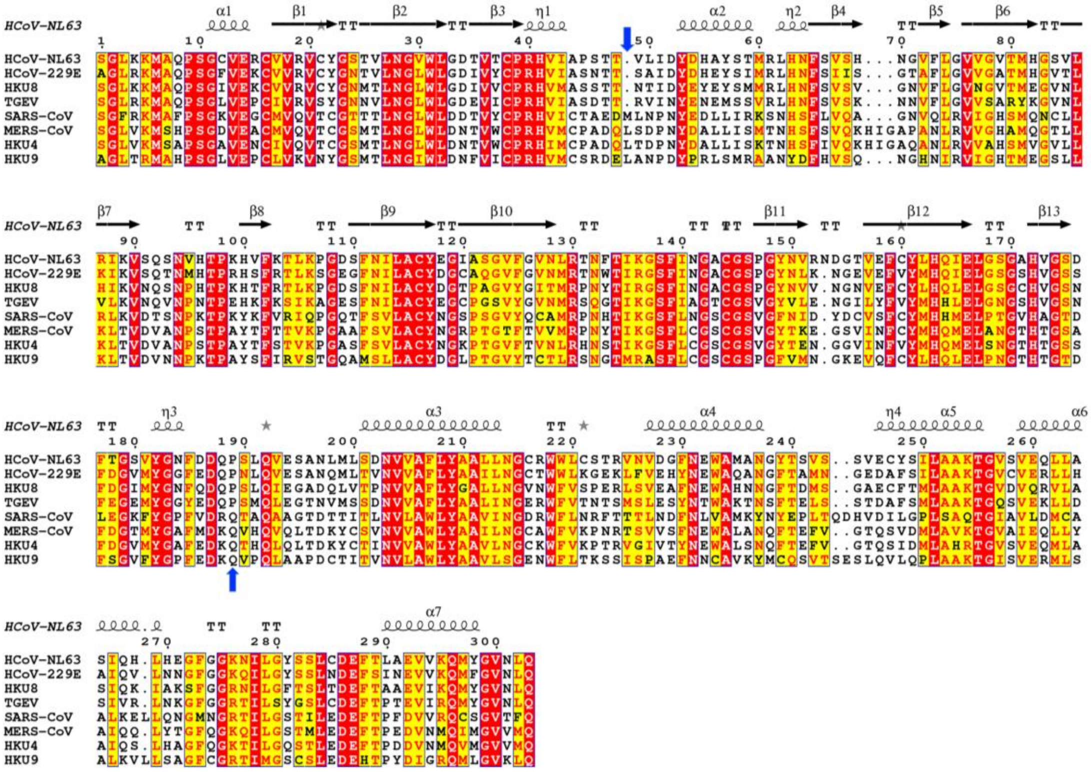


**Fig. 3a.4 Multiple sequence alignment of the main protease of coronaviruses.** The major differences of residues in the S2 pocket are marked by the blue arrows. The GenBank codes for the strains of the CoVs are: HCoV NL63: ABE97129.1; HCoV 229E: AGT21344.1; HKU8: EU420139.1; TGEV: AJ271965.2; SARS CoV: AGT21317.1; MERS CoV: AMO03400.1; HKU4: EF065505.1; HKU9: ABN10910.1) The figure was generated by using the programme ESPript (Gouet et al., 1999).

## 3a.4 Design of a lead α-ketoamide

Herein, a novel α-ketoamide was designed as a broad-spectrum inhibitor of coronavirus $M^{pro}$s and enterovirus $3C^{pro}$s. As the proteases targeted in this study all specifically cleave the peptide bond following a P1-glutamine residue (HCoV-NL63 $M^{pro}$ uniquely also accepts P1 = His at the Nsp13/Nsp14 cleavage site (Pyrc et al., 2007; Wang et al., 2016), we decided to use a 5-membered ring (γ-lactam) derivative of glutamine (GlnLactam) as the P1 residue in this α-ketoamide lead compound. This moiety has been found to be a good mimic of glutamine and enhance the power of the inhibitors by up to 10-fold, most probably because compared to the flexible glutamine side-chain, the more rigid lactam leads to a reduction of the loss of entropy upon binding to the target protease (Dragovich et al., 1999; Tan et al., 2013).

The lead compound to be designed was **11a** (DC401903, RH01), which carries a cinnamoyl N-cap in the P3 position, a benzyl group in P2, the glutamine lactam (GlnLactam) in P1, and benzyl in P1' (Fig. 3a.5A). The inhibitory activity of **11a** was tested against SARS-CoV and HCoV-NL63 $M^{pro}$s, which showed good activity against recombinant SARS-CoV $M^{pro}$ ($IC_{50}$ = 1.95 μM), but was surprisingly completely inactive ($IC_{50}$ > 50 μM) against HCoV-NL63 $M^{pro}$. In addition, our lead Micheal acceptor compound (**SG85**, Tan et al., 2013; Fig. 3a.5B), also with a benzyl group at the P2 position, was inactive against HCoV-NL63 $M^{pro}$ as well.

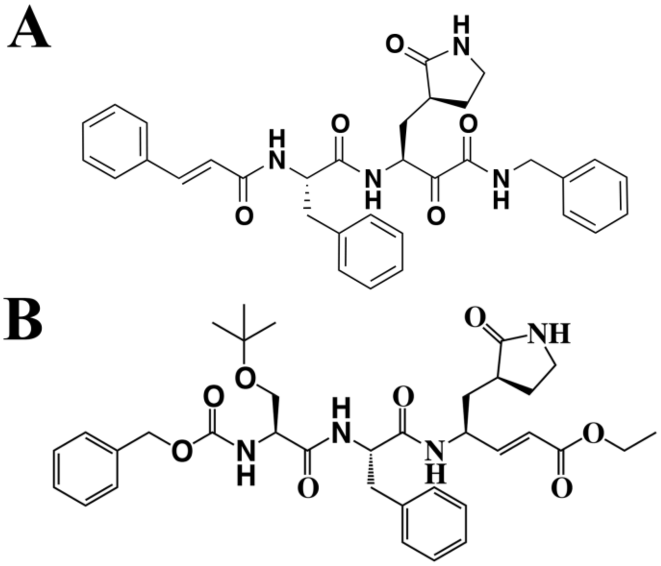


**Fig. 3a.5 Chemical structures of lead compounds 11a (A) and SG85 (B)**

### 3a.5 Structure of HCoV NL63-$M^{pro}$ in complex with 11a and SG85

It is remarkable that I succeeded in obtaining crystal structures of compounds **11a** and **SG85** in complex with the HCoV-NL63 $M^{pro}$, even though they have no inhibitory effect on the activity of the enzyme ($IC_{50}$ > 50 μM), (Fig. 3a.6A and B). Apparently, the compounds are able to bind to this $M^{pro}$ in the absence of peptide substrate, but cannot compete with substrate for the binding site due to their low affinity. A similar observation has been made in a previous study, where the crystal structure of a complex between the inactive Michael-acceptor compound **SG74** and the EV-D68 3C$^{pro}$ was determined (Tan et al., 2013; PDB entry 3ZV9).

The two complex structures of HCoV-NL63 M$^{pro}$ with **11a** and **SG85** revealed that the P2-benzyl group could not fully enter the S2 pocket of the enzyme because of the restricted size of this site (Fig.3a.6A and B). According to the sequence alignment, and results revealed by the crystal structure of HCoV-NL63 M$^{pro}$ free enzyme, the two major differences in the S2 pocket (see above) could be responsible for the low inhibitory activity (Fig. 3a.3C, and 3a.4). Therefore, a double mutant HCoV-NL63 M$^{pro}$ (+M49, P190Q) was generated, mimicking the S2 subsite in SARS-CoV M$^{pro}$ to illustrate the mechanism.

### 3a.6 Kinetic parameters of the variants of HCoV-NL63 M$^{pro}$

First, the kinetic parameters of HCoV-NL63 M$^{pro}$ wt and the (+M49, P190Q) mutant with authentic N-termini were determined using the FRET substrate Dabcyl-KTSAVLQ↓SGFRKME(Edans)-$NH_2$. The $K_m$ value for this substrate and the HCoV-NL63 M$^{pro}$ wt was determined as 40.25 ± 4.98 μM and the $k_{cat}/K_m$ value as 43869.07 ± 4554.24 $M^{-1}\cdot s^{-1}$. This $K_m$ value is in a reasonable range, which is typically between 10 and 50 μM. The $k_{cat}/K_m$ value is almost two-fold higher than the one for SARS-CoV M$^{pro}$ determined by a former colleague with the same substrate and buffer conditions (Zhu et al., 2011). Since no saturation state could be achieved for the HCoV-NL63 M$^{pro}$ (+M49, P190Q) mutant, the apparent $K_{app}$ was calculated as 1279.64 ± 110.79 $M^{-1}\cdot s^{-1}$ to mimic the $k_{cat}/K_m$. Moreover, the inhibitory activity of the two leading compounds against the HCoV-NL63 M$^{pro}$ (+M49, P190Q) mutant was dramatically improved compared with the wt enzyme, namely 5.72 and 6.06 μM respectively, for **11a** and **SG85**.

Next, crystal structures of the HCoV-NL63 $M^{pro}$ (+M49, P190Q) mutant in complexes with the two lead compounds were determined, and the details of the binding sites of the complex structures of the HCoV-NL63 $M^{pro}$ wt and mutant, as well as of the SARS-CoV $M^{pro}$, were compared (Fig. 3a.6).

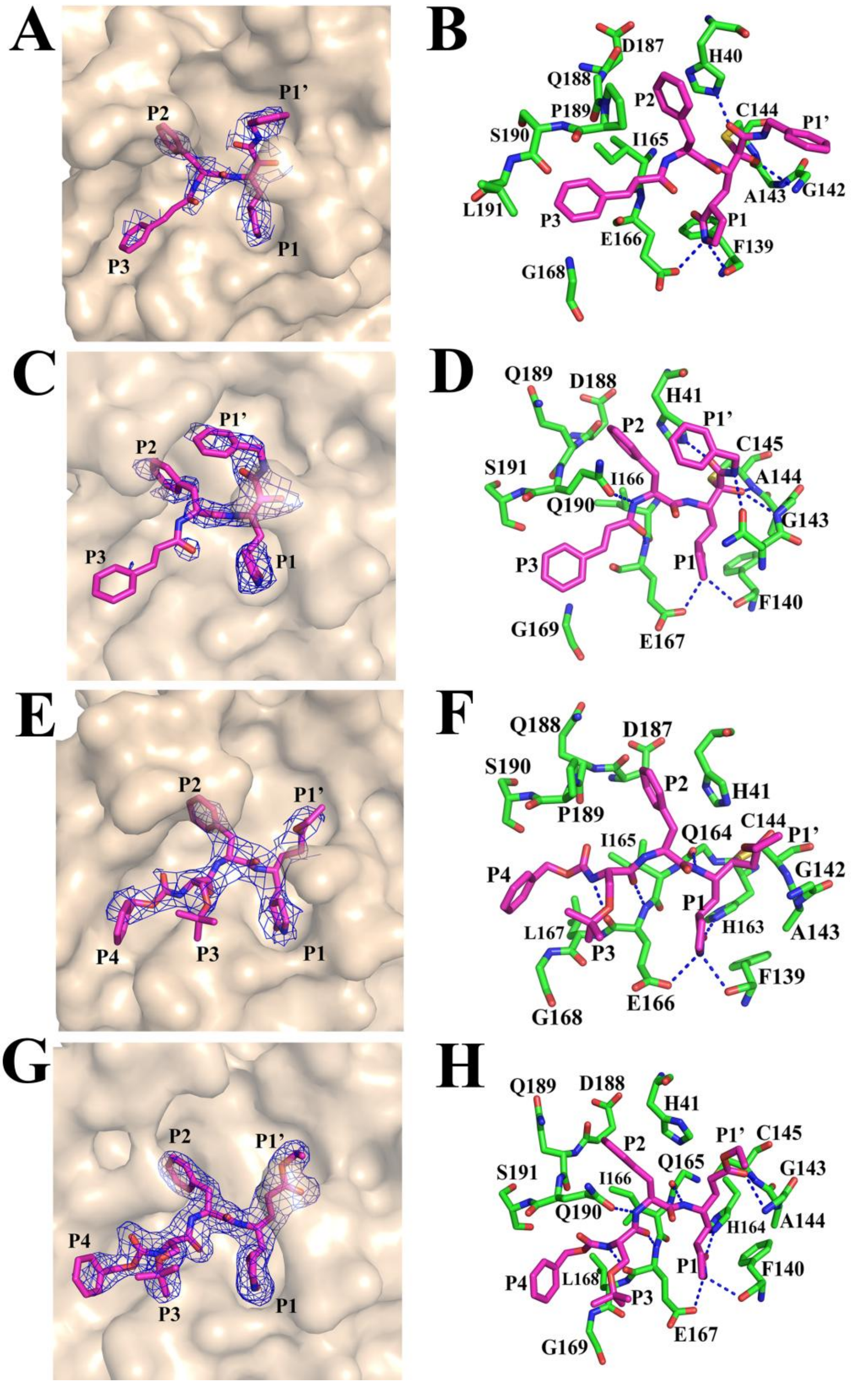

A
P1'
P2
P3
P1
B
D187
H40
Q188
P2
C144
P1'
P189
S190
I165
A143
G142
L191
P3
P1
E166
F139
G168
C
P1'
P2
P3
P1
D
Q189
D188
H41
P2
P1'
C145
S191
I166
A144
Q190
G143
P3
P1
F140
G169
E167
E
P2
P1'
P4
P3
P1
F
Q188
D187
S190
P2
H41
P189
I165
Q164
C144
P1'
P4
G142
P1
L167
H163
A143
P3
F139
G168
E166
G
P2
P1'
P4
P3
P1
H
Q189
D188
H41
P1'
C145
P2
S191
I166
Q165
G143
Q190
H164
A144
P1
F140
P4
L168
P3
G169
E167

**Fig. 3a.6 Crystal structures of HCoV-NL63 M$^{pro}$ (+M49, P190Q)**
**(A, C, E, G)** Fit of **11a** (pink carbon atoms) to HCoV-NL63 M$^{pro}$ (wt) and (+M49, P190Q) mutant (wheat surface), and **SG85** (pink carbon atoms) to HCoV-NL63 HCoV-NL63 M$^{pro}$ (wt) and (+M49, P190Q) mutant (wheat surfaces) as revealed by X-ray crystallography of the complexes. $F_o$ - $F_c$ difference density (contoured at 3σ) for **11a** and **SG85** are indicated in the substrate-binding site of the protease (transparent surface).
**(B, D, F, H)** Detailed interactions of **11a** and **SG85** (pink carbon atoms) with target proteases (green carbon atoms). Residues of the protease and inhibitors are indicted as sticks. Hydrogen bonds are depicted as blue dashed lines. The side-chains of Q188, and A143 in B, Q164 in D, and Q165 in F have been omitted for clarity.

## 3a.7 Comparison of structures of HCoV-NL63 wt and (+M49, P190Q) in complex with 11a

The complex structures of HCoV-NL63 M$^{pro}$ wt and mutant with **11a** are in space group C222$_1$ with three molecules in the asymmetric unit, and C2 for the SARS-CoV M$^{pro}$ complex with one molecule in the asymmetric unite. In the complex structure of HCoV-NL63 M$^{pro}$ wt, **11a** is bound to all 3 chains, but only to chains A and C of the mutant complex structure. The catalytic cysteine attacks the α-C of the α-keto warhead leading to a thiohemiketal product (Fig.3a.6B and D).

### 3a.7.1 S1' subsite

The phenylalanine P1' residue mainly points to the bulk solvent although with different orientation in the structures of SARS-CoV and HCoV-NL63 wt M$^{pro}$. In addition, it makes several hydrophobic interactions with the protease. In both structures of HCoV-NL63 M$^{pro}$ wt and mutant, the carbonyl accepts a H-bond from the Nε of His41.

### 3a.7.2 S1 subsite

As mentioned in the Michael acceptor complex part (see below), the lactam also donates two H-bonds to Glu166 Oε1 and Phe139 CO and the oxygen accepts a H-bond from His163 Nε2. The main-chain amide donates a H-bond to the Gln164 carbonyl group. However, compared to the complex structure of Michael acceptor inhibitors, the α-keto oxygen forms two more hydrogen bonds with the oxyanion-hole amides (Fig.3a.

6B and D).

### 3a.7.3 S3 subsite

In both wt and mutant structures, continuous electron density could not be observed for P3 except for some density for the benzene ring. The orientation of the cinnamoyl P3 is similar between the wt and mutant in HCoV-NL63 $M^{pro}$. The P3 is bound in the extended S3 pocket. The remarkable difference is that the P3 main-chain carbonyl accepts a H-bond from the amide of Glu167, but no such H-bond is formed in the wt complex structure. Importantly, a similar H-bond exists in the complex structure of SARS-CoV $M^{pro}$. The reason for this difference will be extensively discussed in the S2 subsite part (Fig. 3a.6B and D)

## 3a.8 Comparison of structures of HCoV-NL63 wt and (+M49, P190Q) in complex with SG85

The complexes of HCoV-NL63 $M^{pro}$ wt and the (+M49, P190Q) mutant with **SG85** are in space group $C222_1$, with three molecules in one symmetric unit. Overall, the binding of the inhibitor to wt and mutant HCoV-NL63 $M^{pro}$ is similar, but there are important differences in detail. In the complex structure of HCoV-NL63 $M^{pro}$ wt with **SG85**, continuous electron density for **SG85** can be observed in all three chains. However, in the complex structure of mutant HCoV-NL63 $M^{pro}$, electron density can only be found in the A and C chains. Continuous and clear electron density can be found in the complex structure of SARS-CoV $M^{pro}$ (Zhu et al., unpublished; PDB entry 3TNT). In all complex structures, the continuous electron density between the Michael acceptor β-carbon and Cys145 (144)-Sγ indicates the formation of a thioether (Fig. 3a.6F and H).

### 3a.8.1 S1' subsite

In the HCoV-NL63 $M^{pro}$ wt and **SG85** complex structure (chains A and B), the ester carbonyl group is oriented away from the oxyanion hole (amides of Gly142 and Cys144). In chain C, there is one possible 2.47-Å hydrogen bond with the amide of residue Gly142. In the mutant, there appear to be two hydrogen bonds (3.25 ± 0.16 and 3.09 ± 0.16 Å) accepted by the ester oxygen, from the amides of the oxyanion hole. The same applies to the C chain, but both the C-O...H-N (Gly142/Cys144) H-bonds are not ideal in terms of angles (Fig. 3a.6F and H). In the SARS-CoV M$^{pro}$ and **SG85** complex structure, the ester carbonyl accepts one 2.72-Å hydrogen bond from the amide of Gly143 (Zhu et al., unpublished; PDB entry 3TNT).

### 3a.8.2 S1 subsite

Similarly, the P1 lactam side chain of **SG85** is specifically bound to the S1 pocket. The lactam nitrogen donates H-bonds to Glu166 Oε1 and Phe139 CO, in both wt and mutant enzyme. The oxygen of the lactam accepts an H-bond from His163 Nε2. The main-chain amide donates an H-bond to the main-chain carbonyl of Gln164, again both in wt and mutant enzyme. Furthermore, the main-chain amide of the **SG85** donates a H-bond to the Gln165 main-chain carbonyl (Fig. 3a.6F and H). In SARS-CoV M$^{pro}$, the same interactions exist in both side-chain lactam and main-chain amide (Zhu et al., unpublished; PDB entry 3TNT).

### 3a.8.3 S3 subsite

In both wt and mutant of HCoV-NL63 M$^{pro}$ as well as SARS-CoV M$^{pro}$, the P3 main-chain carbonyl and amide of **SG85** interact with the Glu166 main chain via two H-bonds. However, the side-chain of the P3 residue points out into the bulk solvent (Fig. 3a.7F and H).

### 3a.8.4 S4 subsite

In all the structures, the P4 binds mainly through hydrophobic interaction to the extended pocket (Fig. 3a.7F and H).

### 3a.9 The distinct S2 subsite

The hydrophobic S2 pocket of wt HCoV-NL63 $M^{pro}$ wt is comprised of the residues His41, Ile51, Cys144, Ile165, Gln164, Asp187, Q188 (mc: contribution from main chain only), and Pro189. The pocket in SARS-CoV $M^{pro}$ is lined by His41, Met49, Cys145, His164, Met165, Asp187, Arg188 (mc: contribution from main chain only), and Gln189. As mentioned above, there are two important differences in the S2 pocket between HCoV-NL63 and SARS-CoV $M^{pro}$s: Pro189 becomes Gln189, and M49 is inserted in SARS-CoV (Fig. 3a.3C).

In the complex structure of **11a** with SARS-CoV, Gln189 makes no direct interaction with the P2-amide, but this is a water-mediated hydrogen bond. However, in the HCoV-NL63 $M^{pro}$ wt complex structure, there is no direct or indirect H-bond between **11a** and Pro189 (Fig. 3a.6B). In contrast, in the mutant complex structure, there is a 3.09 -Å direct H-bond between the Gln190 side-chain oxygen and the P2 main-chain amide of **11a** (Fig. 3a.6D). Significantly, the H-bond between the Gln190 (189 in SARS-CoV $M^{pro}$) and the main-chain amide leads to a flip of the main chain resulting in a nice H-bond between the P3 main-chain carbonyl and the amide of Glu167. This extra hydrogen bond makes the inhibitor bind more tightly with the enzyme in both the mutant HCoV-NL63 $M^{pro}$ and in SARS-CoV $M^{pro}$. Surprisingly, the P2 phenyl ring is rotated by 90 degree in the SARS-CoV $M^{pro}$ complex structure with **11a**, compared to the orientation in the HCoV-NL63 $M^{pro}$ (wt) structure, where it is parallel to the imidazole ring of His41. This is due to the high flexibility of the S2 pocket in SARS-CoV $M^{pro}$. The interaction details of 11a with SARS-CoV $M^{pro}$ will be discussed in part 3a.10 (Fig. 3a.6B and D).

In the structure of SARS-CoV $M^{pro}$ with **SG85**, the main-chain amide of the P2 residue makes a H-bond with the side-chain carbonyl of Gln189 (Zhu et al., unpublished; PDB entry 3TNT), but in case of HCoV-NL63

$M^{pro}$, this is not possible. In the mutant, the Pro is replaced by Gln, which makes a decent hydrogen bond with the main-chain amide of the P2 residue, just like in SARS-CoV $M^{pro}$ (Fig. 3a.6H). In the HCoV-NL63 free enzyme, the Pro189 φ/ψ torsion angles are -65.1°/130.3° respectively, while in the **SG85** complex structure, they are adapted to -98.5°/46.5°, which might be due to the phenyl group not being suitable for the S2 pocket and therefore leading to non-favorable torsion angles of Pro189 (Morris et al., 1992). Furthermore, the loop (residues: 43-48) became more flexible with discontinuous density after insertion of Met49, which enables the S2 subsite to accommodate the inhibitor more easily.

In general, the low inhibitory activity of the two different lead compounds (**SG85** and **11a**) against HCoV-NL63 $M^{pro}$ is mainly due to the absence of Met49 and replacement of Gln189 by Pro.

Based on the difference of the S2 pocket between HCoV-NL63 and SARS-CoV $M^{pro}$, another α-ketoamide (**11n**) was designed with an *iso*-butyl at the P2 position, which showed good inhibitory activity against both proteases. Moreover, the complex structure was determined with HCoV-NL63 $M^{pro}$. The details will be described in part 3a.10.

Diffraction data collection and structure refinement of the variants of HCoV-NL63 $M^{pro}$ are summarized in Table 3a.1.

**Table 3a.1 Crystallographic data and refinement statistics of variants of HCoV-NL63 $M^{pro}$ structures**

| **Protease** | **HCoV-NL63 $M^{pro}$ (wt)** | | | **HCoV-NL63 $M^{pro}$ (+M49, P190Q)** | |
|---|---|---|---|---|---|
| **Inhibitor** | **Free enzyme** | **11a** | **SG85** | **11a** | **SG85** |
| **Data collection statistics** | | | | | |
| X-ray source | SLS X10SA | BESSY 14.2 | BESSY 14.2 | BESSY 14.2 | BESSY 14.2 |
| Wavelength [Å] | 1.00001 | 0.9184 | 0.91841 | 0.91841 | 0.91841 |
| Space group | $C222_1$ | $C222_1$ | $C222_1$ | $C222_1$ | $C222_1$ |
| Unit cell dimensions [Å] | a = 131.13,<br>b = 214.43,<br>c = 116.83 | a=131.12,<br>b=211.04,<br>c=115.63 | a = 132.38,<br>b = 211.63,<br>c = 115.63 | a = 129.97,<br>b = 213.70,<br>c = 116.05 | a = 128.88,<br>b = 215.02,<br>c = 116.12 |
| Number of protease molecules per asymmetric unit | 3 | 3 | 3 | 3 | 3 |
| Resolution range [Å] [a] | 111.87-1.75<br>(1.84-1.75) | 48.00 - 2.95<br>(3.11-2.95) | 48.38-3.50<br>(3.69-3.50) | 48.53-2.20<br>(2.32-2.20) | 49.91-2.10<br>(2.21-2.10) |
| Number of observations | 1,126,056<br>(165,650) | 188,558 (28,512) | 240,012 (35,007) | 910,716<br>(113,864) | 709,986<br>(103,251) |
| Number of unique reflections | 164,852 (23,921) | 33,470 (4,925) | 21,555 (3,093) | 81,901 (11,763) | 93,922 (13,588) |
| Completeness [%] | 100 (100) | 98.1 (99.8) | 100 (100) | 99.9 (99.4) | 99.9 (99.8) |
| Mean I/σ (I) | 13.5 (2.4) | 13.1 (3.0) | 9.0 (2.2) | 22.8 (2.6) | 22.4 (2.7) |
| Multiplicity | 6.8 (7.0) | 5.6 (5.8) | 11.1 (11.3) | 11.1 (9.7) | 7.6 (7.6) |
| $R_{merge}$ [%] [b] | 0.071 (0.727) | 0.112 (0.569) | 0.253 (0.906) | 0.090 (0.911) | 0.066 (0.814) |
| $R_{pim}$ [%] [c] | 0.029 (0.259) | 0.051 (0.259) | 0.078 (0.278) | 0.028 (0.300) | 0.026 (0.315) |
| $CC_{1/2}$ [d] | 0.999 (0.805) | 0.994 (0.808) | 0.992 (0.895) | 0.999 (0.767) | 0.999 (0.858) |

| **Protease** | **HCoV-NL63 $M^{pro}$ (wt)** | | | **HCoV-NL63 $M^{pro}$ (+M49, P190Q)** | |
|---|---|---|---|---|---|
| **Inhibitor** | **Free enzyme** | **11a** | **SG85** | **11a** | **SG85** |
| **Refinement statistics** | | | | | |
| $R_{cryst}$ [e]/$R_{free}$ [f] [%] | 19.92/22.80 | 17.83/23.97 | 18.91/24.38 | 19.75/22.99 | 17.81/21.56 |
| r.m.s.d in bond lengths (Å) | 0.0228 | 0.0140 | 0.0121 | 0.0196 | 0.0198 |
| r.m.s.d in bond angles (°) | 2.0858 | 1.9082 | 1.9049 | 2.0419 | 0.2156 |
| Average B-factor for all atoms ($Å^2$) | 34.08 | 56.20 | 78.02 | 46.17 | 48.76 |
| Number of protein atoms | 6886 | 6837 | 6819 | 6877 | 6870 |
| Number of inhibitor atoms | 0 | 126 | 141 | 84 | 94 |
| Number of water molecules | 678 | 280 | 71 | 368 | 436 |
| **Ramachandran plot** | | | | | |
| Preferred regions (%) | 97.27 | 95.65 | 94.74 | 97.54 | 96.77 |
| Allowed regions (%) | 2.62 | 4.24 | 5.26 | 2.23 | 2.78 |
| Outlier regions (%) | 0.11 | 0.11 | 0.00 | 0.22 | 0.44 |

a The highest resolution shell is shown in parentheses.

b $R_{merge} = \sum_{hkl}\sum_{i=1}^{n}|I_i(hkl) - \bar{I}(hkl)|/\sum_{hkl}\sum_{i=1}^{n} I_i(hkl)$ (Arndt et al., 1968)

c $R_{pim} = \sum_{hkl}\sqrt{1/(n-1)}\sum_{i=1}^{n}|I_i(hkl) - \bar{I}(hkl)|/\sum_{hkl}\sum_{i=1}^{n} I_i(hkl)$ (Weiss & Hilgenfeld, 1997)

d $CC_{1/2}$ is the correlation coefficient determined by two random half data sets (Karplus & Diederichs, 2012).

e $R_{cryst} = \sum_{hkl}|F_o(hkl) - F_c(hkl)|/\sum_{hkl}|F_o(hkl)|$

f $R_{free}$ was calculated for a test set of reflections (5%) omitted from the refinement.

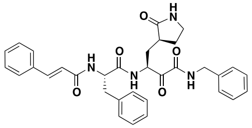
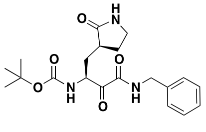
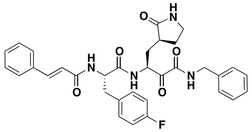
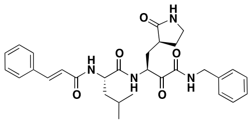
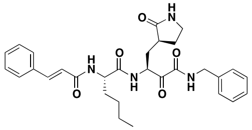
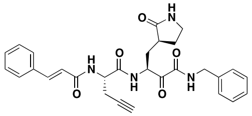
## 3a.10 SAR study on the α-ketoamides, with focus on the P2 group

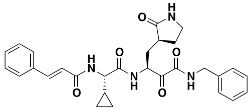
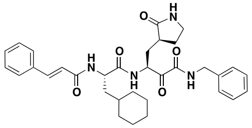
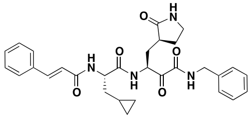
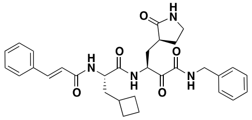
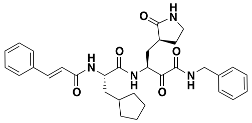
The results on α-ketoamides in this thesis have been described in the following manuscript: Lin, D.§, Zhang L.§, Kusov Y., Nian Y., Ma Q., Wang J., Leyssen P., Lanko K., Neyts J., de Wilde A., Snijder E. J., Liu H.*, Hilgenfeld R.* (2016) Alpha-ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication.

The α-ketoamides were designed by Prof. Rolf Hilgenfeld and synthesized by Dr. Daizong Lin. The inhibitory activity of the α-ketoamides against enterovirus $3C^{pro}$ was determined together by me and Dr. Daizong Lin. The crystal structure of the CVB3 $3C^{pro}$ in complex with **11a** (DC401903) was determined by Dr. Qingjun Ma. I did the *in-vitro* inhibitory activity test of all the α-ketoamides against coronavirus $M^{pro}$s as well as crystallization and determination of the complex structures of SARS-CoV $M^{pro}$ with **11a**, and **11s**, and HCoV-NL63 $M^{pro}$ with **11a**, **11f**, and **11n**.

In this part, the efforts aimed at optimizing the substituents at the P2 positions of the α-ketoamides are described. As mentioned in part 3a.4, the lead compound of the α-ketoamides series was **11a**, which showed good to mediocre activities against recombinant SARS-CoV $M^{pro}$ ($IC_{50}$ = 1.95 μM; for all compounds, see Table 3a.2 for standard deviations), CVB3 $3C^{pro}$ ($IC_{50}$ = 6.6 μM), and EV-A71 $3C^{pro}$ ($IC_{50}$ = 1.2 μM), and was completely inactive ($IC_{50}$ > 50 μM) against HCoV-NL63 $M^{pro}$.

**Table 3a.2 $IC_{50}$ (μM) values of α-ketoamides against viral proteases.**

| Compound No. | Formula | EV-A71 $3C^{pro}$ | CVB3 $3C^{pro}$ | SARS-CoV $M^{pro}$ | HCoV-NL63 $M^{pro}$ |
|---|---|---|---|---|---|
| DC401903 (**11a**) |  | 1.22 ± 0.12 | 6.56 ± 3.10 | 1.95 ± 0.24 | >50 |
| DC401918 (**11f**) |  | >50 | >50 | >50 | >50 |
| DC401909 (**11m**) |  | 2.32 ± 1.21 | 8.69 ± 1.28 | >50 | >50 |
| DC401908 (**11n**) |  | 13.80 ± 4.17 | 3.82 ± 1.21 | 0.33 ± 0.04 | 1.08 ± 0.09 |
| DC401910 (**11o**) |  | 3.24 ± 1.69 | 5.20 ± 0.28 | 8.50 ± 3.71 | >50 |
| DzL11 (**11p**) |  | 8.50 ± 1.81 | 8.93 ± 2.39 | 10.68 ± 7.34 | 17.37 ± 13.97 |

| Compound No. | Formula | EV-A71 3C$^{pro}$ | CVB3 3C$^{pro}$ | SARS-CoV M$^{pro}$ | HCoV-NL63 M$^{pro}$ |
|---|---|---|---|---|---|
| DzL09 (**11q**) |  | >30 | 15.38 ± 4.12 | 6.27 ± 2.87 | 11.50 ± 5.05 |
| DzL08 (**11r**) |  | 1.69 ± 0.47 | 0.95 ± 0.15 | 0.71 ± 0.36 | 12.27 ± 3.56 |
| DzL10 (**11s**) |  | 18.47 ± 4.25 | 4.25 ± 1.40 | 0.24 ± 0.08 | 1.37 ± 0.35 |
| DzL14 (**11t**) |  | 10.81 ± 2.62 | 4.78 ± 1.13 | 1.44 ± 0.40 | 3.43 ± 2.45 |
| DzL15 (**11u**) |  | 4.73 ± 0.94 | 1.93 ± 0.43 | 1.27 ± 0.34 | 5.41 ± 2.31 |

nd-not done

Crystal structures of compound **11a** (DC401903) in complex with SARS-CoV M$^{pro}$, HCoV-NL63 M$^{pro}$, and CVB3 3C$^{pro}$ demonstrate that the $\alpha$-keto-carbon is covalently linked to the active-site Cys (no. 145, 144, and 147, resp.) of the protease (Fig. 3a.7A-C). The resulting thiohemiketal is in the *R* configuration in the SARS-CoV and HCoV-NL63 M$^{pro}$ but in the *S* configuration in the CVB3 3C$^{pro}$ complex. The reason for this difference is that the oxygen atom of the thiohemiketal accepts a hydrogen bond from the catalytic His40 in the CVB3 protease, rather than from the main-chain amides of the oxyanion hole as in the SARS-CoV and HCoV-NL63 enzymes (Fig. 3a.7A-C insets).

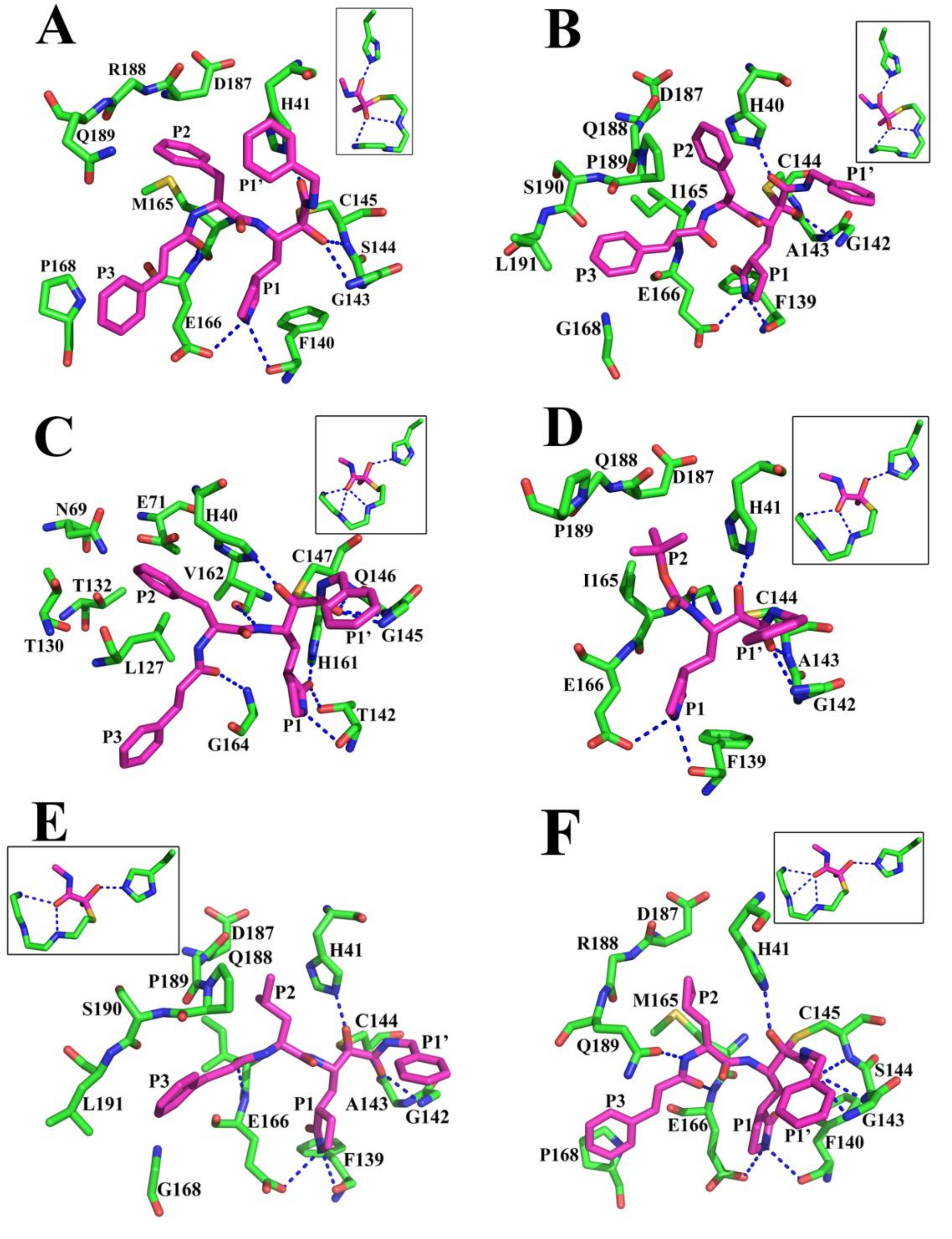


**Fig. 3a.7 Detailed interactions of peptidomimetic α-ketoamides (pink carbon atoms) with target proteases (green carbon atoms).** Hydrogen bonds are depicted as black dashed lines. The inset at the top of the images shows the sterochemistry of the thiohemiketal formed by the nucleophilic attack of the catalytic Cys residue onto the α-keto group.

**(A)** Binding of **11a** to SARS-CoV M$^{pro}$. The thiohemiketal is in the *R* configuration, with its oxygen accepting two hydrogen bonds from the oxyanion-hole amides of Gly143 and Cys145. The amide oxygen accepts an H-bond from His41. The side-chains of Ser144 and Arg188 have been omitted for clarity.

**(B)** The P2-benzyl substituent of **11a** cannot fully enter the S2 pocket of the

HCoV-NL63 M$^{pro}$, which is much smaller and has less plasticity than the corresponding pocket of SARS-CoV M$^{pro}$ (cf. A). The benzyl therefore binds above the pocket in the view shown here; this is probably the reason for the total inactivity ($IC_{50} > 50$ μM) of compound **11a** against HCoV-NL63 M$^{pro}$. The small size of the pocket is due to the replacement of the flexible Gln189 of the SARS-CoV M$^{pro}$ by the more rigid Pro189 in this enzyme. The stereochemistry of the thiohemiketal is *R*. The side-chains of Ala143 and Gln188 have been omitted for clarity.

**(C)** Binding of **11a** to the CVB3 3C$^{pro}$. The stereochemistry of the thiohemiketal is *S*, as the group accepts a hydrogen bonds from His41, whereas the amide keto group accepts three H-bond from the oxyanion hole (residues 145 - 147). The side-chain of Gln146 has been omitted for clarity.

**(D)** The crystal structure of **11f** (DC401918) in complex with HCoV-NL63 M$^{pro}$ shows that this short (inactive) compound lacking a P3 residue has its P2-Boc group inserted into the S2 pocket of the protease. The stereochemistry of the thiohemiketal is *S*. The side-chains of Ala143 and Gln188 have been omitted for clarity.

**(E)** In contrast to P2 = benzyl in **11a**, the isobutyl group of **11n** (DC401908) is small and flexible enough to enter into the narrow S2 pocket of the HCoV-NL63 M$^{pro}$. The thiohemiketal is in the *R* configuration. The side-chains of Ala143 and Gln188 have been omitted for clarity.

**(F)** In spite of its small size, the cyclopropylmethyl side-chain in the P2 position of **11s** can tightly bind to the S2 subsite of the SARS-CoV M$^{pro}$, as this pocket exhibits pronounced plasticity due to the conformational flexibility of Gln189 (see also Fig. 3a.7). The stereochemistry of the thiohemiketal is *S*. The side-chains of Ser144 and Gln188 have been omitted for clarity.

The crystal structures of SARS-CoV M$^{pro}$, HCoV-NL63 M$^{pro}$, and CVB3 3C$^{pro}$ in complex with **11a** (DC401903) revealed a fundamental difference between the S2 pockets of the coronavirus proteases and the enterovirus proteases: The cavities are covered by a "lid" in the former but are open to one side in the latter (Fig. 3a.8B-D). In SARS-CoV M$^{pro}$, the lid is formed by the $3_{10}$ helix 46 - 51 and in HCoV-NL63 M$^{pro}$ by the loop 43-48. Residues from the lid, in particular Met49 in the case of SARS-CoV M$^{pro}$, can thus make hydrophobic interactions with the P2 substitutent of the inhibitor, whereas such interaction is missing in the enterovirus 3C$^{pro}$s. In addition to the lid (43-48), the S2 pocket is lined by the "back-wall" (main-chain atoms of residues 186 and 188 and Cβ atom of Asp187), the side-walls (Gln189, His41), as well as the "floor" (Met165) in SARS-CoV M$^{pro}$. In HCoV-NL63 M$^{pro}$, the corresponding structural elements are main-chain atoms of residues 187 and 188 as well as the Cβ atom of

Asp187 (back-wall), Pro189 and His41 (side-walls), and Ile165 (floor). Finally, in CVB3 3C$^{pro}$, Arg39, Asn69, and Glu71 form the back-wall, residues 127-132 and His40 form the side-walls, and Val 162 constitutes the floor.

In addition, the S2 pocket is of different size in the various proteases. The SARS-CoV enzyme features the largest S2 pocket, with a volume of 252 Å$^3$ embraced by the residues (Gln189, His41) defining the side-walls of the pocket in the ligand-free enzyme, as calculated by using Chimera (Pettersen et al., 2004), followed by the CVB3 3C$^{pro}$ S2 pocket with about 180 Å$^3$ (space between Thr130 and His40). The HCoV-NL63 M$^{pro}$ has by far the smallest S2 pocket of the three enzymes, with a free space of only 45 Å$^3$ between Pro189 and His41, according to Chimera.

In agreement with these observations, a good fit is observed between the P2 benzyl group and the S2 subsite of the SARS-CoV M$^{pro}$ as well as of the CVB3 3C$^{pro}$ (Fig. 3a.7A and C). In contrast, the crystal structure of the complex between **11a** and HCoV-NL63 M$^{pro}$, against which the compound is inactive, demonstrates that the P2-benzyl group cannot fully enter the S2 pocket of the enzyme because of the restricted size of this site (Fig. 3a.7B). Thus, the properties of the target proteases with respect to the S2 pocket were defined at this point as "small" and "covered by a lid" for HCoV-NL63 M$^{pro}$, "large" and "covered" for SARS-CoV M$^{pro}$, and "large" and "open" for CVB3 3C$^{pro}$. Through comparison with crystal structures of other proteases of the same virus genus (HCoV-229E M$^{pro}$ for alphacoronaviruses (Anand et al., 2003; PDB entry 1P9S); HKU4 M$^{pro}$ for betacoronaviruses (Ma, Xiao et al., unpublished; PDB entry 2YNA); and EV-A71 3C$^{pro}$ for enteroviruses (Lu et al., 2011; PDB entry 3SGK), it was ensured that the conclusions drawn from the template structures were valid for other family members as well.

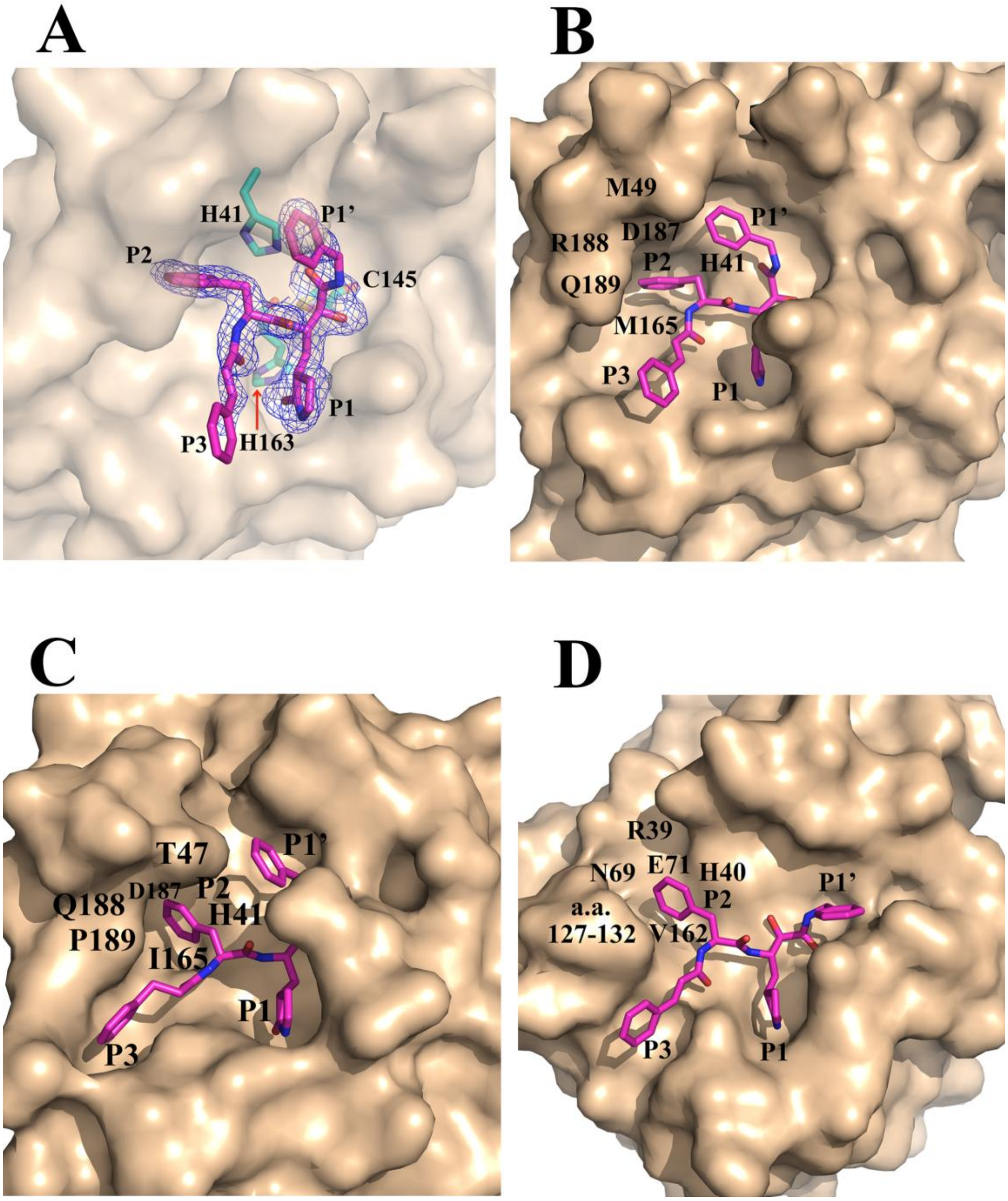


**Fig. 3a.8 Fit of compound 11a (pink carbon atoms) to the target proteases (wheat surfaces) as revealed by X-ray crystallography of the complexes.**

**(A)** $F_o$ - $F_c$ difference density (contoured at $3\sigma$) for **11a** in the substrate-binding site of the SARS-CoV $M^{pro}$ (transparent surface). Selected side-chains of the protease are shown with green carbon atoms. **(B)** Another view of **11a** in the substrate- binding site of the SARS-CoV $M^{pro}$. Note the "lid" formed by residue M49 and its neighbors above the S2 pocket.

**(C)** **11a** in the substrate-binding site of HCoV-NL63 $M^{pro}$. Because of the restricted size of the S2 pocket, the P2 benzyl group of the compound cannot enter deeply into this site. Note that the S2 pocket is also covered by a "lid" centered around T47.

**(D)** **11a** in the substrate-binding site of the CVB3 $3C^{pro}$. The S2 site is large and not covered by a "lid".

To explore the sensitivity of the S2 pocket towards a polar substituent in the *para* position of the benzyl group, compound **11m** (DC401909) was synthesized carrying a 4-fluorobenzyl group in P2. This substitution abolished almost all activity against the SARS-CoV M$^{pro}$ (IC$_{50}$ > 50 µM), and the compound proved inactive against HCoV-NL63 M$^{pro}$ as well, whereas IC$_{50}$ values were 2.3 µM against the EV-A71 3C$^{pro}$ and 8.7 µM against CVB3 3C$^{pro}$. From this, it was concluded that the introduction of the polar fluorine atom is not compatible with the geometry of the S2 pocket of SARS-CoV M$^{pro}$, whereas the fluorine can accept a hydrogen bond from Arg39 in EV-A71 3C$^{pro}$ (Lu et al., 2011) and probably also CVB3 3C$^{pro}$. In SARS-CoV M$^{pro}$, however, the carbonyl groups of residues 186 and 188 might lead to a repulsion of the fluorinated benzyl group.

As the P2-benzyl group of **11a** was apparently too large to fit into the S2 pocket of the HCoV-NL63 M$^{pro}$, it was replaced by isobutyl in **11n** (DC401908). This resulted in improved activities against SARS-CoV M$^{pro}$ (IC$_{50}$ = 0.33 µM) and in a very good activity against HCoV-NL63 M$^{pro}$ (IC$_{50}$ = 1.08 µM, compare with the inactive **11a**). For EV-A71 3C$^{pro}$, however, the activity decreased to IC$_{50}$ = 13.8 µM, different from CVB3 3C$^{pro}$, where IC$_{50}$ was 3.8 µM. The interpretation of this result is that the smaller P2-isobutyl substituent of **11n** can still interact with the "lid" (in particular, Met49) of the SARS-CoV M$^{pro}$ S2 site, but is unable to reach the "back-wall" of the EV-A71 3C$^{pro}$ pocket and thus, in the absence of a "lid", cannot generate sufficient enthalpy of binding. We will see from examples to follow that this trend persists among all inhibitors with a smaller P2 substituent: Even though the SARS-CoV M$^{pro}$ S2 pocket has a larger volume than that of the enterovirus 3C$^{pro}$, the enzyme can be efficiently inhibited by compounds carrying a small P2 residue that makes hydrophobic interactions with the lid (Met49) and floor (Met165) residues.

I managed to obtain crystals of **11n** (DC401908) in complex with the M$^{pro}$ of HCoV NL63 and found the P2 isobutyl group to be well-embedded in the S2 pocket (Fig. 3a.7E). This is not only a consequence of the smaller size of the isobutyl group compared to the benzyl group, but also of its larger conformational flexibility, which allows a better fit to the binding site.

When we replaced the P2-isobutyl residue of **11n** (DC401908) by n-butyl in **11o** (DC401910), the activities were as follows: IC$_{50}$ = 8.5 μM for SARS-CoV M$^{pro}$, totally inactive (IC$_{50}$ > 50 μM) against HCoV-NL63 M$^{pro}$, IC$_{50}$ = 3.2 μM for EV-A71 3C$^{pro}$, and 5.2 μM for CVB3 3C$^{pro}$. The decreased activity in case of SARS-CoV M$^{pro}$ and the total inactivity against HCoV-NL63 M$^{pro}$ indicate that the n-butyl chain is too long for the S2 pocket of these proteases, whereas the slight improvement against EV-A71 3C$^{pro}$ and CVB3 3C$^{pro}$ is probably a consequence of the extra space that is available to long and flexible substituents because of the lack of a lid covering the enterovirus 3C$^{pro}$ pocket.

As the n-butyl substituent in P2 of **11o** was obviously too long, we next synthesized a derivative with the shorter propargyl (ethinylmethyl) as the P2 residue (compound **11p** (DZL11)). This led to very mediocre activities against all tested proteases. Using cyclopropyl as the P2 residue (compound **11q**, DZL09), the IC$_{50}$ values were even higher against most of the proteases tested. Obviously, the P2 side-chain requires a methylene group in the β-position in order to provide the necessary flexibility for the substituent to be embedded in the S2 pocket.

Having realized that in addition to size, flexibility of the P2 substituent may be an important factor influencing inhibitory activity, we introduced flexibility into the phenyl ring of **11a** by reducing it. The cyclohexylmethyl derivative **11r** (DZL08) exhibited IC$_{50}$ = 0.7 μM against SARS-CoV M$^{pro}$, 12.3 μM against HCoV-NL63 M$^{pro}$, 1.7 μM against EV-A71 3C$^{pro}$, and 0.9 μM against CVB3 3C$^{pro}$. Thus, the replacement of the phenyl group by the cyclohexyl group led to a significant improvement of the inhibitory activity against the recombinant SARS-CoV M$^{pro}$ and to a dramatic improvement in case of CVB3 3C$^{pro}$. Even for the HCoV-NL63 M$^{pro}$, against which **11a** was completely inactive, greatly improved albeit still weak activity was observed (Table 3a.2).

At this point, we decided to systematically vary the size of the ring system in P2. The next substituent to be tried was cyclopropylmethyl (compound **11s** (DZL10), which showed good activities against SARS-CoV M$^{pro}$ (IC$_{50}$ = 0.24 μM) and HCoV-NL63 M$^{pro}$ (1.4 μM), but poor values against EV-A71 3C$^{pro}$ (IC$_{50}$ = 18.5 μM) and CVB3 3C$^{pro}$ (IC$_{50}$ = 4.3 μM) (Table 3a.2).

I next analyzed the crystal structure of the complex between SARS-CoV M$^{pro}$ and compound **11s** (DZL10) (Fig. 3a.7F). The cyclopropylmethyl substituent was found to be incorporated deeply into the S2 pocket, making hydrophobic interactions with Met49 (the lid), Met165 (the floor), and the Cβ of Asp187 (the back-wall). In spite of the small size of the P2 substituent, this is possible because the S2 pocket of SARS-CoV M$^{pro}$ is flexible enough to "shrink" and enclose the P2 moiety tightly. This plasticity is expressed in a conformational change of residue Gln189, both in the main chain and in the side-chain. The main-chain conformational change is connected with a flip of the peptide between Gln189 and Thr190. The $\chi 1$ torsion angle of this side-chain changes from roughly antiperiplanar (*ap*) to (-)-synclinal (*-sc*). The conformational variability of Gln189 has been noted before, both in Molecular Dynamics simulations (Tan et al., 2005) and in other crystal structures (Zhu et al., 2011). As a consequence of these changes, the side-chain oxygen of Gln189 can accept a 2.54-Å hydrogen bond from the main-chain NH moiety of the P2 residue (Fig. 3a.9).

Affinity of **11s** (DZL10) for the S2 pocket of HCoV-NL63 M$^{pro}$ is good because of an almost ideal match of size and not requiring conformational changes, which this enzyme would not be able to undergo because of the replacement of the flexible Gln189 by the more rigid Pro. On the other hand, docking of the same compound into the crystal structure of the CVB3 3C$^{pro}$ revealed that the cyclopropylmethyl moiety was probably unable to generate sufficient Free Energy of binding because of the missing lid and the large size of the S2 pocket in the enterovirus 3C$^{pro}$, thereby explaining the poor inhibitory activity of **11s** against these targets.

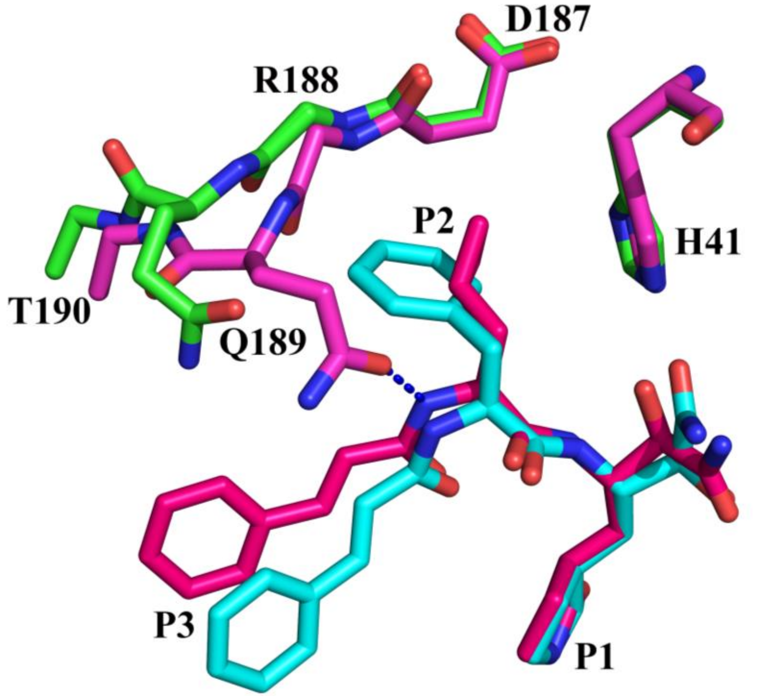


**Fig. 3a.9 A pronounced plasticity of the S2 pocket of SARS-CoV M$^{pro}$ is revealed by a comparison of the geometry of the subsite in the complexes with 11a (DC401903; P2 = benzyl; inhibitor cyan, protein green) and 11s (DZL10; P2 = cyclopropylmethyl; inhibitor red, protein pink).** The main differences here concern the main chain around Gln189 as well as the side-chain of this flexible residue, conformational change of which allows the S2 pocket to "shrink" and adapt to the small size of the P2-substituent in **11s**. This change also enables the formation of a hydrogen bond between the main-chain amide of the P2 residue and the side-chain oxygen of **11s**. The side-chains of Arg188 and Thr190 as well as the P1' substituents of the inhibitors have been omitted for clarity.

We next introduced cyclobutylmethyl in the P2 position (compound **11t**, DZL14) and obtained the following results: $IC_{50}$ = 1.4 μM for SARS-CoV M$^{pro}$, 3.4 μM for HCoV-NL63 M$^{pro}$, 10.8 μM for EV-A71 3C$^{pro}$, and 4.8 μM for CVB3 3C$^{pro}$ (Table 3a.2). Obviously, this substituent was still a bit too small for the enterovirus proteases, so as the next step, we tested P2 = cyclopentylmethyl (compound **11u**, DZL15). This turned out to be the most interesting compound so far with respect to broad-spectrum activity, i.e. acceptable $IC_{50}$ values against all tested enzymes: 1.3 μM against SARS-CoV M$^{pro}$, 5.4 μM against HCoV-NL63 M$^{pro}$, 4.7 μM against EV-A71 3C$^{pro}$, and 1.9 μM against CVB3 3C$^{pro}$

(Table 3a.2).

**11u** (DZL15) appeared so far the best compromise compound, yet for each of the individual viral enzymes, the following compounds proved superior: P2 = cyclopropylmethyl (compound **11s**, DZL10) for SARS-CoV $M^{pro}$, P2 = isobutyl (compound **11n** (DC401908) and P2 = cyclopropylmethyl (**11s**, DZL10) for HCoV-NL63 $M^{pro}$, **11a** (DC401903) and **11r** (DZL08) for EV-A71 $3C^{pro}$, and **11r** (DZL08) for CVB3 $3C^{pro}$. However, if we relax the condition that the universal inhibitor should show good activity against HCoV-NL63 $M^{pro}$, then the best compound is **11r** (DZL08, P2 = cyclohexylmethyl), with submicromolar $IC_{50}$ values against CVB3 $3C^{pro}$ and SARS-CoV $M^{pro}$, and $IC_{50}$ = 1.7 μM against EV-A71 $3C^{pro}$ (Table 3a.2).

Diffraction data collection and structure refinement of the complex structures are summarized in Table 3a.3.

**Table 3a.3 Crystallographic data and refinement statistics of viral proteases in complex with α-ketoamides**

| **Protease** | **SARS-CoV M$^{pro}$** | | **HCoV-NL63 M$^{pro}$** | |
|---|---|---|---|---|
| **Compound** | **11a (RH01)** | **11s (DZL10)** | **11n (DC401908)** | **11f (DC401918)** |
| **Data collection statistics** | | | | |
| X-Ray source | SLS X10SA | DESY P11 | BESSY 14.2 | BESSY 14.2 |
| Wavelength | 1.0000 | 0.9919 | 0.9184 | 0.9184 |
| Space group | C2<br>(cell choice 1) | C2<br>(cell choice 3, I121) | $C222_1$ | $C222_1$ |
| Unit cell dimensions [Å] | a = 108.2,<br>b = 82.28,<br>c = 53.44 | a=53.50,<br>b=81.87,<br>c=109.03 | a=133.86,<br>b=211.33,<br>c=118.32 | a=133.65,<br>b=211.29,<br>c=118.26 |
| Unit cell dimensions [º] | $\alpha = \gamma = 90$,<br>$\beta = 103.95$ | $\alpha = \gamma = 90$,<br>$\beta = 104.05$ | $\alpha = \beta = \gamma = 90$ | $\alpha = \beta = \gamma = 90$ |
| Number of protease molecules per asymmetric unit | 1 | 1 | 3 | 3 |
| Resolution range[Å] | 64.77 - 1.62<br>(1.71 - 1.62) | 43.83 – 2.20<br>(2.32 - 2.20) | 48.24 – 2.30<br>(2.42 - 2.30) | 44.28 – 2.35<br>(2.48 – 2.35) |
| Number of observations | 386,723 (56,621) | 74,079 (10,007) | 37,2476 (54,396) | 301,104 (37,670) |
| Number of unique reflections | 56,530 (8,160) | 22,847 (3,241) | 74,507 (10,793) | 68,423 (9,796) |
| Completeness [%] | 98.2 (97.4) | 98.4 (96.5) | 99.9 (100.0) | 98.2 (97.2) |
| Mean I/σ (I) | 24.5 (3.4) | 13.1 (3.9) | 20.4 (3.3) | 11.5 (3.0) |
| Multiplicity | 6.8 (7.0) | 3.2 (3.1) | 5.0 (5.0) | 4.4 (3.8) |
| $R_{merge}$ [%] | 0.034 (0.490) | 0.053 (0.289) | 0.060 (0.536) | 0.076 (0.458) |
| $R_{pim}$ [%] | 0.014 (0.198) | 0.034 (0.197) | 0.030 (0.262) | 0.038 (0.250) |
| $CC_{1/2}$ | 1.000 (0.960) | 0.997 (0.930) | 0.999 (0.896) | 0.996 (0.873) |

| **Protease** | **SARS-CoV $M^{pro}$** | | **HCoV-NL63 $M^{pro}$** | |
|---|---|---|---|---|
| **Compound** | **11a (RH01)** | **11s (DZL10)** | **11n (DC401908)** | **11f (DC401918)** |
| **Refinement statistics** | | | | |
| r.m.s.d in bond lengths (Å) | 0.0279 | 0.0194 | 0.0196 | 0.0190 |
| r.m.s.d in bond angles (°) | 2.5065 | 2.0662 | 2.0503 | 2.0480 |
| Average B-factor for all atoms (Å$^2$) | 33.979 | 39.731 | 51.924 | 44.608 |
| Number of protein atoms | 2395 | 2379 | 6807 | 6797 |
| Number of inhibitor atoms | 42 | 39 | 117 | 84 |
| Number of water molecules | 398 | 143 | 531 | 259 |
| **Ramachandran plot** | | | | |
| Preferred regions (%) | 97.99 | 96.32 | 97.20 | 97.08 |
| Allowed regions (%) | 2.01 | 3.34 | 2.58 | 2.58 |
| Outlier regions (%) | 0.00 | 0.33 | 0.22 | 0.34 |

## 3a.11 Inhibitory activity of aldehydes against SARS-CoV M$^{pro}$

The aldehydes were designed by the group of Prof. Hong Liu (Shanghai Institute of Materia Medica, Chinese Academy of Sciences, SIMM), and were synthesized by Hailong Liu, who worked in Prof. Hong Liu's group, and Prof. Rolf Hilgenfeld's group later (Institute of Biochemistry, University of Lübeck).

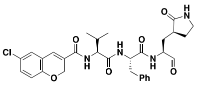
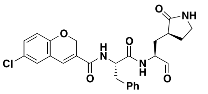
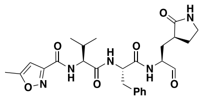
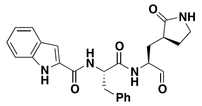
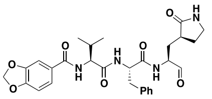
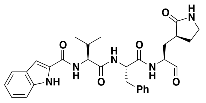
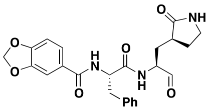
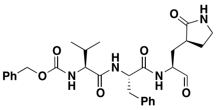
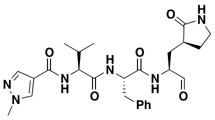
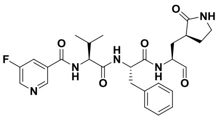
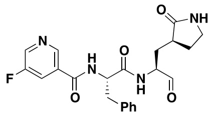
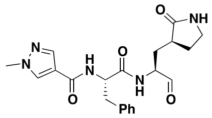
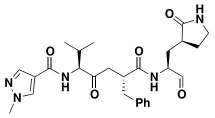
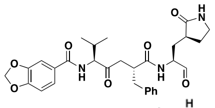
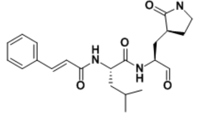
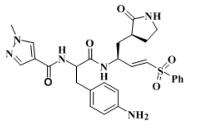
### 3a.11.1 P1 substituent

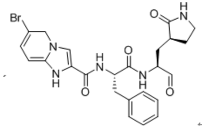
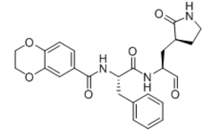
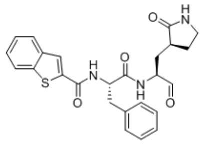
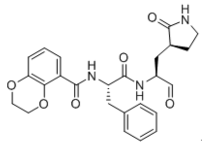
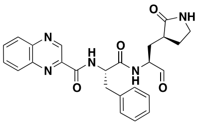
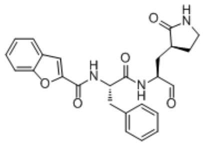
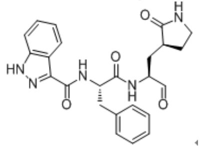
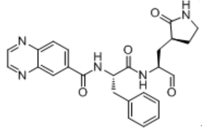
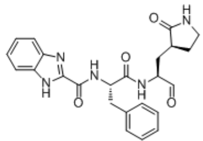
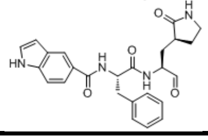
As the SARS-CoV M$^{pro}$ specifically recognizes glutamine at the P1 position, a 5-membered ring (γ-lactam) derivative of glutamine was selected as the P1 residue in all the aldehydes presented here. The advantage of this moiety has been described in the α-ketoamides part.

### 3a.11.2 P2 substituent

As described in the α-ketoamide part, the SARS-CoV M$^{pro}$ posseses alarge, "covered", and flexible S2 pocket, so the phenyl group could be properly accommodated in this subsite. Having learned from the α-ketoamides that proper length of the P2 is an important factor influencing inhibitory activity, we introduced the benzyl group at the P2 position in most of the aldehydes, except for **DC401937** with an isobutyl. **DC401937** exhibited an IC$_{50}$ (0.32 μM; Table 3a.4) similar to compounds with the benzyl group at P2 position. In addition, the main-chain peptide bond at the P2 position was replaced by a C-C bond in two of the aldehydes (**DC401933** and **DC401934**), both of which lost inhibitory activity against the SARS-CoV M$^{pro}$ (IC$_{50}$ > 25 μM, Table 3a.4).

Table 3a.4 $IC_{50}$ of aldehydes against SARS-CoV $M^{pro}$

| Compound No. | Formula | $IC_{50}$ (μM) |
|---|---|---|
| DC401921 |  | 7.40 ± 1.18 |
| DC401922 |  | 0.75 ± 0.21 |
| DC401923 |  | 0.03 ± 0.012 |
| DC401924 |  | 0.70 ± 0.17 |
| DC401925 |  | 0.05 ± 0.017 |
| DC401926 |  | 0.10 ± 0.15 |
| DC401927 |  | 0.60 ± 0.31 |
| DC401928 |  | 0.035±0.009 |
| DC401929 |  | 0.20 ± 0.18 |
| DC401930 |  | 0.45 ± 0.24 |
| DC401931 |  | 1.05 ± 0.33 |
| DC401932 |  | 1.10 ± 0.27 |
| DC401933 |  | >25 |
| DC401934 |  | >25 |
| DC401937 |  | 0.32 ± 0.03 |
| DC401941 |  | >100 |

| Compound No. | Formula | $IC_{50}$ (µM) |
|---|---|---|
| DC401950 |  | 0.78 ± 0.12 |
| DC401951 |  | 0.56 ± 0.21 |
| DC401952 |  | 1.69 ± 0.96 |
| DC401953 |  | 4.64 ± 1.75 |
| DC401954 |  | 1.13 ± 0.36 |
| DC401955 |  | 5.40 ± 2.88 |
| DC401956 |  | 2.62 ± 0.61 |
| DC401957 |  | 4.21 ± 1.36 |
| DC401958 |  | 0.111 ± 0.002 |
| DC401964 |  | 0.79 ± 0.08 |

### 3a.11.3 P3 substituent

A series of tripeptide aldehydes with various heterocyclic-carbonyl groups at the P3 position showed sub-micromolear inhibitory activity against SARS-CoV $M^{pro}$ (Table 3a.4), i.e. 6-chloro-2*H*-chromene-3-carbonyl (**DC401922**, $IC_{50}$ = 0.75 μM), 1*H*-indole-2-carbonyl (**DC401924**, $IC_{50}$ = 0.70 μM), 3a, 7a-dihydrobenzo[*d*][1,3]dioxole-5-carbonyl (**DC-401927**, $IC_{50}$ = 0.60 μM), cinnamoyl (**DC401937**, $IC_{50}$ = 0.32 μM), 6-bromoimidazo[1,2-a]pyridine-2-carbonyl (**DC401950**, $IC_{50}$ = 0.78 μM), 2,3-dihydrobenzo[b][1,4]dioxine-6-carbonyl (**DC401951**, $IC_{50}$ = 0.56 μM), 1*H*-benzo[*d*]imidazole-2-carbonyl (**DC401958**, $IC_{50}$ = 0.11 μM), 1*H*-indole-5-carbonyl (**DC401964**, $IC_{50}$ = 0.79 μM). Compounds with P3 = 3-fluoropyridine-3-carbonyl (**DC401931**, $IC_{50}$ = 1.05 μM), 1-methyl-1-*H*-pyrazole-4-carbonyl (**DC401932**, $IC_{50}$ = 1.10 μM), benzo[b]thiophene-2-carbonyl (**DC401952**, $IC_{50}$ = 1.69 μM), 2,3-dihydrobenzo[b][1,4]-dioxine-5-carbonyl (**DC401953**, $IC_{50}$ = 4.64 μM), quinoxaline-2-carbonyl (**DC401954**, $IC_{50}$ = 1.13 μM), = benzofuran-2-carbonyl (**DC401955**, $IC_{50}$ = 5.40 μM), 1*H*-indazole-3-carbonyl (**DC401956**, $IC_{50}$ = 2.62 μM), quinoxaline-6-carbonyl (**DC401957**, $IC_{50}$ = 4.21 μM) showed micromolar-range $IC_{50}$ against SARS-CoV $M^{pro}$. Compared to previous pentapepti- de aldehyde inhibitors (Zhu et al., 2011), the aldehydes presented herein showed superior inhibitory activity against SARS-CoV $M^{pro}$ Since no apparent SAR could be deduced with respect to the P3 group, I speculated that specificity of the P3 group is overruled by the highly electrophilic character of the aldehyde warhead.

### 3a.11.4 P4 substituent

A number of tetrapeptide aldehydes with isobutyl as P3 group, and different heterocyclic-carbonyl moieties at the P4 position, were tested against SARS-CoV $M^{pro}$ (Table 3a.4). Three of them showed excellent activity, with the P4 = 5-methylisoxazole-3-carbonyl (**DC401923**, $IC_{50}$ =

0.03 μM), benzo[*d*][1,3]dioxole-5-carbonyl (**DC401925**, $IC_{50}$ = 0.05 μM), and benzyl acetate (**DC401928**, $IC_{50}$ = 0.035 μM), respectively. P4 = 1*H*-indole-2-carbonyl (**DC401926**, $IC_{50}$ = 0.10 μM), 1-methyl-1*H*-pyra- zole-4-carbonyl (**DC401929**, $IC_{50}$ = 0.20 μM), 5-fluoropyrid ine-3-carbonyl (**DC401930**, $IC_{50}$ = 0.45 μM) showed submicromolar activity. Only one compound (**DC401921**) with a 6-chloro-2*H*-chrom- ene-3-carbonyl at the P4 position showed less potent activity ($IC_{50}$ =7.40 μM).

### 3a.12 Crystal structures of SARS-CoV M$^{pro}$ in complex with DC401923 and DC401954

The crystal structures of SARS-CoV M$^{pro}$ in complex with **DC401923** and **DC401954** were determined at 1.70 and 1.85 Å, respectively. The warhead of both aldehydes forms a hemithioacetal (*S* configuration) with the Cys145 of the catalytic dyad of the protease (Fig. 3a.10A and B, insets). The hydroxyl group of the thiohemiacetal accepts two hydrogen bonds from the amides of Cys145 and Gly143 in the oxyanion hole. The canonical 5-membered ring (γ-lactam) derivative of glutamine (GlnLactam) always shares a similar binding mode involving three hydrogen bonds with His163, Glu166, and Phe140 of the main protease. The main-chain amide of the P1 residue donates a hydrogen bond to the carbonyl of His164 in the M$^{pro}$. The P2 phenyl group from both of the aldehydes neatly accommodates in the S2 pocket, mediated by hydrophobic interactions with Cα, Cβ, and carbonyl carbon of Asp187, Cα and the carbonyl carbon of Arg188, Cα of Gln189, as well as the side-chains of Met49 and Met165 (Fig. 3a.10).

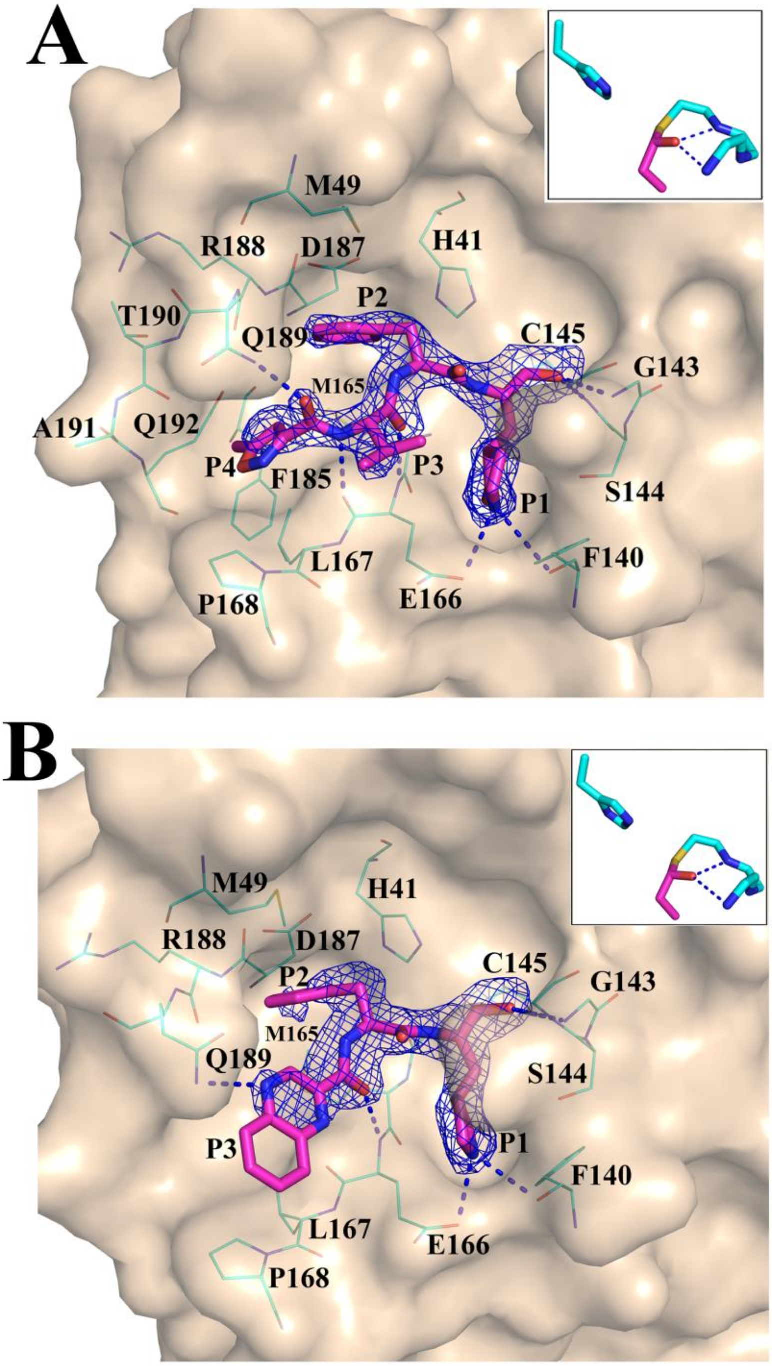


**Fig. 3a.10 Crystal structures of SARS-CoV M$^{pro}$ in complex with aldehyde compounds.** Fit of compound aldehydes (pink carbon atoms) to the SARS-CoV M$^{pro}$ (wheat surfaces) as revealed by X-ray crystallography of the complexes. $F_o$ - $F_c$ difference density (contoured at $3\sigma$) for aldehydes in the substrate-binding site of the SARS-CoV M$^{pro}$ (transparent surface). Detailed interactions of aldehydes (pink carbon atoms) with SARS-CoV M$^{pro}$ (cyan carbon atoms) are indicated as sticks and lines, respectively. Hydrogen bonds are depicted as black dashed lines. The inset at the top of the images shows the sterochemistry of the hemithioacetal formed by the nucleophilc attack of the catalytic Cys residue onto the aldehyde group.

**(A)** Crystal structure of SARS-CoV M$^{pro}$ in complex with with **DC401923**.

**(B)** Crystal structure of SARS-CoV M$^{pro}$ in complex with **DC401954**.

In the complex structure with **DC401954**, the P3 quinoxaline donates two hydrogen bonds to the side-chain of Gln189, and the main-chain carbonyl of Glu163. Moreover, the main chain carbonyl of the P3 residue accepts a hydrogen bond from the main-chain amide of Glu166. In the complex structure with **DC401923**, the P3 side-chain orients itself towards the bulk solvent. The main-chain carbonyl accepts a H-bond from the main-chain amide of Glu166, and the main-chain amide of the P3 moiety donates a hydrogen bonds to the main-chain carbonyl of Glu166. The carbonyl of the P4 group accepts a hydrogen bond from Gln189. The P4 side-chain deeply penetrates into the S4 pocket formed by Met165, Leu167, Phe185, and Gln192. In particular, the tetrapeptidic aldehyde has more hydrophobic interaction from the P4 moiety with the protease, compared to the tripeptidic aldehyde, which might result in a better inhibitory activity. Remarkably, the side-chain of Gln189, involved in a hydrogen bond with the main-chain amide of the P2 moiety in both of the Michael acceptor **SG85** (Zhu et al., unpublished; PDB entry 3TNT) and the α-ketoamide compound **11s** (**DZL10**, Fig 3a.7F) complexes, interacts with the P3 residue in both of the complex structures with aldehydes (**DC401923** and **DC401954**), demonstrating that the S2 pocket in SARS-CoV M$^{pro}$ is highly flexible. The binding details of the two aldehydes with SARS-CoV M$^{pro}$ are indicated in Fig 3a.10.

Diffraction data collection and structure refinement of SARS-CoV M$^{pro}$ in complex with aldehydes are summarized in Table 3a.5.

**Table 3a.5 Crystallographic data collection and refinement statistics of SARS-CoV $M^{pro}$ in complex with aldehydes**

| **Protease** | **SARS-CoV $M^{pro}$** | |
|---|---|---|
| **Inhibitor** | **DC401923** | **DC401954** |
| **Data collection statistics** | | |
| X-Ray source | BESSY 14.2 | BESSY 14.2 |
| Wavelength | 0.91841 | 0.91841 |
| Space group | C2<br>(cell choice 1) | C2<br>(cell choice 3, I121) |
| Unit cell dimensions [Å] | a = 108.71,<br>b = 81.22,<br>c = 53.27 | a = 53.31<br>b = 81.42,<br>c = 108.41 |
| Unit cell dimensions [º] | $\alpha = \gamma = 90$<br>$\beta = 104.12$ | $\alpha = \gamma = 90$<br>$\beta = 103.90$ |
| Number of protease molecules per asymmetric unit | 1 | 1 |
| Resolution range [Å] | 43.58-1.70 (1.79-1.70) | 43.67-1.85 (1.95-1.85) |
| Number of observations | 200,167 (29,897) | 145,328 (21,191) |
| Number of unique reflections | 49,135 (7,169) | 38,168 (5,520) |
| Completeness [%] | 99.5 (99.8) | 99.5 (99.1) |
| Mean $I/\sigma$ (I) | 13.4 (2.2) | 18.8 (2.3) |
| Multiplicity | 4.1 (4.2) | 3.8 (3.8) |
| $R_{merge}$ [%] | 0.052 (0.560) | 0.035 (0.576) |
| $R_{pim}$ [%] | 0.029 (0.314) | 0.041 (0.339) |
| $CC_{1/2}$ | 0.999 (0.920) | 0.999 (0.872) |
| **Refinement statistics** | | |
| $R_{cryst}$ /$R_{free}$ [%] | 22.5/26.94 | 20.72/25.04 |
| r.m.s.d in bond lengths (Å) | 0.0426 | 0.0213 |
| r.m.s.d in bond angles (°) | 3.35 | 2.09 |
| Average B-factor for all atoms ($Å^2$) | 16.30 | 44.44 |
| Number of protein atoms | 2381 | 2384 |
| Number of inhibitor atoms | 41 | 34 |
| Number of water molecules | 171 | 208 |
| **Ramachandran plot** | | |
| Preferred regions (%) | 96.36 | 97.01 |
| Allowed regions (%) | 3.64 | 2.99 |
| Outlier regions (%) | 0.00 | 0.00 |

## 3b Flavivirus part

### 3b.1 DENV NS2B-NS3$^{pro}$

#### 3b.1.1 Recombinant production of stable DENV2 NS2B-NS3$^{pro}$

Initially, the purification of DENV NS2B-NS3$^{pro}$ was performed according to the protocol reported by D'Arcy et al. (2006). The DENV2 NS2B-NS3$^{pro}$ was purified under the optimal buffer condition described in this publication (50 mM NaCl, 20 mM Tris pH 8.5, 5% (v/v) glycerol). However, the cleavage activity of the DENV protease is high under this condition, leading to considerable autolysis during the purification procedure (Fig. 3b.1A). Leung et al. (2001) studied the effect of ionic strength, glycerol and pH on the enzymatic activity of the DENV2 NS2B-NS3$^{pro}$, and found that low pH ($pH < 7$), high ionic strength (concentration of $NaCl > 300$ mM), and no glycerol could tremendously decrease the enzymatic activity of the DENV2 NS2B-NS3$^{pro}$. Thus, a series of buffers including different glycerol concentrations, ionic strength, glycerol, and pH range was tested, from which the buffer consisting of 20 mM Bis-Tris, 500 mM NaCl, pH 6.3 was selected for the purification of DENV2 NS2B-NS3$^{pro}$. Under this buffer condition, the enzymatic activity was low and the protein was not denatured.

As shown in Fig. 3b.1B-F, after two steps of purification by Ni-NTA column and size-exclusion chromatography, high-purity (>95%) and stable DENV NS2B-NS3$^{pro}$ could be obtained. Afterwards, the protein was adapted to the buffer optimized for the enzymatic assay (20 mM Tis-HCl, 20% glycerol, 1 mM CHAPS, pH 8.5), and the kinetic parameters were determined against the standard substrate Bz-Nle-Lys-Arg-Arg-AMC for DENV protease, with a $k_{cat}/K_m$ value ($33{,}000 \pm 10{,}000$ $s^{-1}M^{-1}$) similar to results from the literature (Erbel et al., 2006), with a $K_m$ value of $147.10 \pm 36.73$ μM. Crystallization was set up by using the high-purity DENV2 NS2B-NS3$^{pro}$. Optimized crystals were observed in the drop with 0.1 M MES, 50% PEG 200, pH 6.0. The crystals are shown in Fig. 3b.1G.

### 3b.1.2 Structure elucidation of DENV2 NS2B-NS3$^{pro}$

The structure was determined by the molecular replacement method with the structure of DENV2 NS2B-NS3$^{pro}$ as search model (Erbel et al., 2006; PDB entry 2FOM) at 1.75 Å, in space group $C222_1$, and with one molecule per asymmetric unit. The structure indicated that the NS3 domain adopted a chymotrypsin-like fold with two β-barrels, each including six β-strands, and the catalytic triad (Ser135-His51-Asp75) located at the cleft between the two β-barrels. The first β-strand (residue 51-57) of NS2B, participating in the first β-barrel of the NS3$^{pro}$ domain, is important for the stabilization of the NS3$^{pro}$ fold. The residual part of NS2B does not wrap around the NS3, and is not involved in forming the substrate-binding site; this is characteristic of the "open" conformation (Fig. 3b.1H). The r.m.s. deviation between this structure and the reported DENV2 NS2B-NS3$^{pro}$ (Erbel et al., 2006; PDB entry 2FOM) is 0.56 Å (for main-chain atoms). Diffraction data collection and final model refinement statistics are summarized in Table 3b.1, and will be presented together with the WNV NS2B-NS3$^{pro}$ structure in part 3b.2.

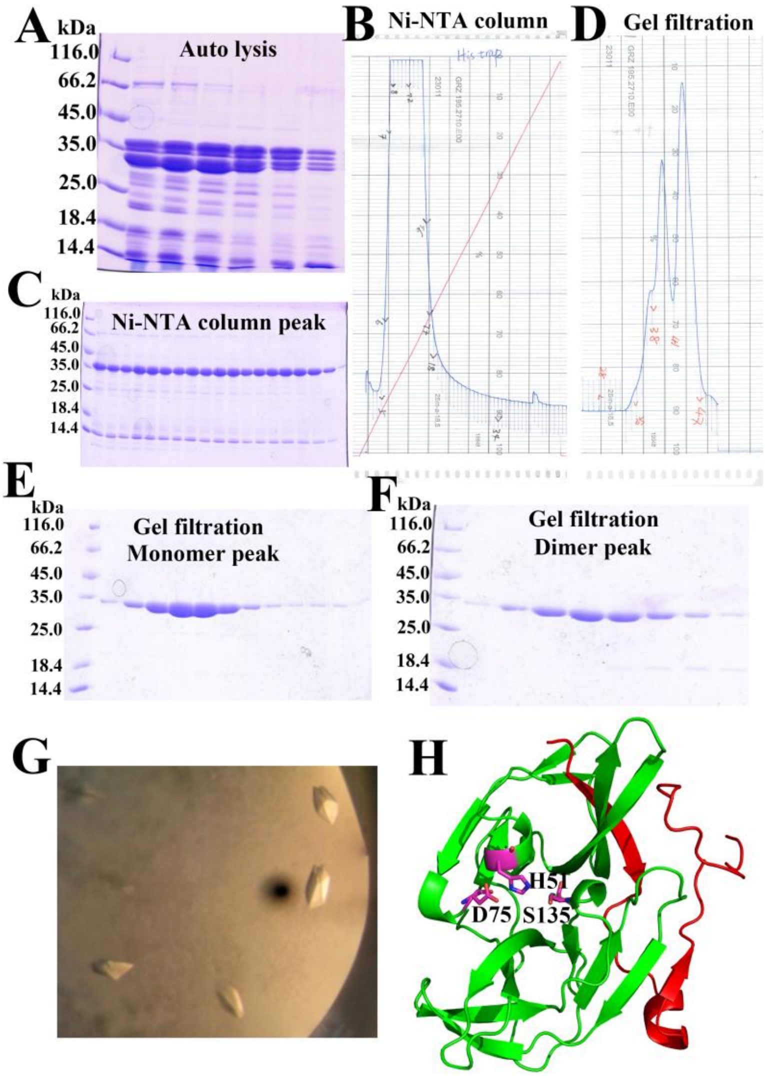


**Fig. 3b.1 Purification, crystals, and structure of DENV2 NS2B-NS3$^{pro}$**

**(A)** Autolysis of DENV2 NS2B-NS3$^{pro}$ during purification in the buffer 50 mM NaCl, 20 mM Tris pH 8.5, 5%(v/v) glycerol revealed by SDS-PAGE. Similar to the result in Leung et al (2001), the free DENV2 NS2B-NS3$^{pro}$ protease migrates more slowly than expected from its molecular mass (~26 kD), probably due to the pronounced non-spherical shape of the protease in the open conformation.

**(B and C)** Purification of DENV2 NS2B-NS3$^{pro}$ using a Ni-NTA column; the corresponding chromatographic fractions were analyzed by SDS-PAGE.

**(D)** Purification of DENV2 NS2B-NS3$^{pro}$ by gel filtration (Superdex$^{TM}$ 75); two peaks corresponding to the monomer and dimer were detected.

**(E and F)** The chromatographic fractions corresponding to the monomer and dimer peak were analyzed by SDS-PAGE, indicating high purity and non-autolysis of the DENV2 NS2B-NS3$^{pro}$ during purification in the buffer 20 mM Bis-Tris, 500 mM NaCl, pH 6.3.

**(G)** Crystals of DENV2 NS2B-NS3$^{pro}$ grown under optimized condition.

**(H)** Crystal structure of DENV2 NS2B-NS3$^{pro}$. The protease is in the open conformation as the NS2B (red color) is not wrapping around the NS3$^{pro}$ (green color). The Gly$_4$-Ser-Gly$_4$ linker between NS2B and NS3$^{pro}$ is not defined by electron density. Catalytic residues are indicated by sticks.

### 3b.2 Structure elucidation of WNV NS2B-NS3$^{pro}$ with a boronic-acid inhibitor (cn-716)

This portion of the work will be published in the following manuscript: Nitsche C., Zhang L., Weigel F. L., Schilz J., Graf D., Bartenschlager R., Hilgenfeld R., Klein C. D. (2016) Peptide-boronic acid inhibitors of flaviviral proteases.
The **cn-716** was synthesized by Dr. Christoph Nitsche and the *in vitro* inhibitory activity of the **cn-716** against WNV NS2B-NS3$^{pro}$ was also determined by colleagues from Prof. Dr. Christian D. Klein's group. The crystallization work of WNV NS2B-NS3$^{pro}$ with **cn-716** was performed by Jonas Schilz, who did his bachelor thesis in the institute of biochemistry under my supervision in the summer semester, 2015. I determined the complex structure of WNV NS2B-NS3$^{pro}$ with **cn-716**.

The capped dipeptide boronic-acid compound **cn-716** (Fig. 3b.2A) was shown to covalently and reversibly bind the NS3 protease. The boron atom of this compound is covalently linked to the side-chain Oγ of the catalytic Ser135. It was found that this compound possesses a good *in-vitro* inhibitory activity against the WNV NS2B-NS3$^{pro}$, with $IC_{50}$ = 0.105 μM and $K_i$ = 0.082 μM.

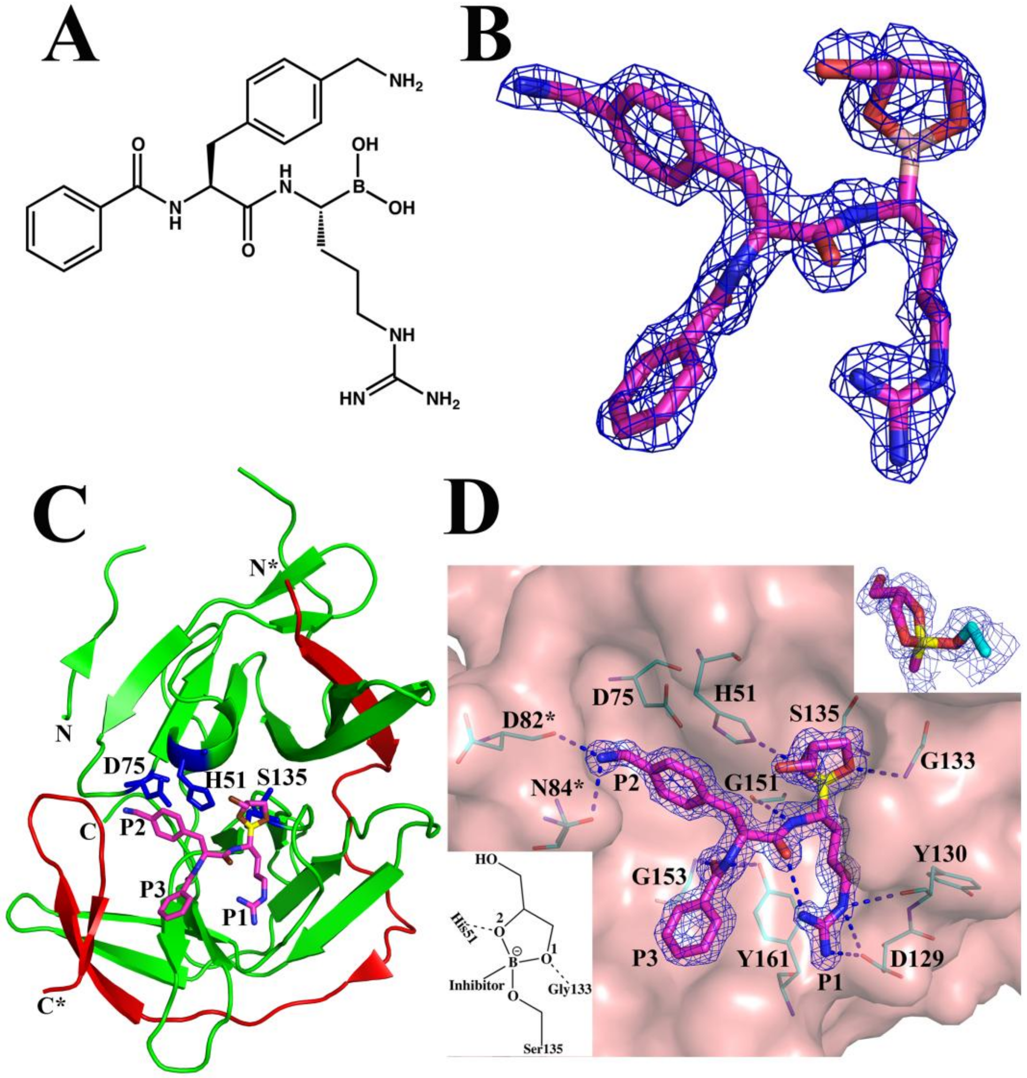


**Fig. 3b.2 Crystal structure of WNV NS2B-NS3$^{pro}$ in complex with cn-716**
**(A)** Chemical formula of **cn-716**.
(B) The Fo-Fc difference density contoured at 3.0 σ is shown for **cn-716,** calculated from a full dataset collected by rotating the crystal by 360°. No continuous electron density could be identified for the boron-carbon bond.
**(C)** Overall structure of the WNV NS2B-NS3$^{pro}$ protease in complex with **cn-716**, (PDB entry 5IDK). A distinctive feature of the closed (active) form of the enzyme as found here is that the NS2B chain (red) is wrapped around the NS3 protease domain (green). The Gly$_4$-Ser-Gly$_4$ linker between NS2B and NS3 is not defined by electron density. Chain termini are labeled N*, C* for NS2B and N, C for NS3. Colors of inhibitor atoms: C, magenta; N, blue; O, dark red, B, yellow. Active-site residues are dark blue and labeled.
**(D)** Compound **cn-716** in the substrate-binding site of WNV NS2B-NS3$^{pro}$ protease. The $F_o$-$F_c$ difference density, calculated from the final coordinates with the inhibitor omitted and contoured at 4σ, is shown as blue grid. Residues of the NS2B chain are indicated by an asterisk (*). H-bonds are shown as dashed lines. Upper right inset: The tetrahedral boronate; Ser135 in green, boron in yellow, oxygens and carbons of the glycerol moiety in red, Fo-Fc difference density as blue grid. Lower left inset: Structure of the glycerol boronate bound to Ser135.

The crystals were obtained from cocrystallization attempts of WNV NS2B-NS3$^{pro}$ with **cn-716**. The complex structure of WNV NS2B-NS3$^{pro}$ with **cn-716** was determined at 1.50 Å resolution, in space group $P2_12_12_1$ with three molecules per asymmetric unit. Initially, the images for a 360-degree rotation of the crystal were processed and scaled for the determination of the structure. During the refinement, continuous $F_o$-$F_c$ difference density was not visible at 3.0 σ level for the covalent bond between the Cα of the P1 residue and the boron atom in all three chains in one asymmetric unit (Fig. 3b.2B). In the PhD thesis of Dr. Cynthia Nichole Fuhrmann (2005, University of California, San Francisco, USA), it was described that the Ser195-boron adduct between α-lytic protease and a peptidyl boronic acid compound adopted alternate conformations as a result of radiation damage at the high X-ray dosage necessary to achieve ultra-high resolution. Herein, I propose that the lack of $F_o$-$F_c$ difference density between the Cα of the P1 group and the boron atom might be a direct result of X-ray irradiation, which stimulated the breakage of the boron-carbon bond within the inhibitor. Therefore, the images for the first 90 degrees were processed separately and scaled for determination of the structure, and clear and continuous $F_o$-$F_c$ different density emerged immediately even at the 4σ level (Fig. 3b.2D). Remarkably, the B factors obtained by using the 90-degree rotation data set for both boron and P1-Cα (Boron: 17.77 ± 0.89, P1-Cα: 16.75 ± 0.16; average and standard deviations calculated from the three copies of the complex in the asymmetric unit) are about 30% lower than those obtained by using the 360-degree data set (Boron: 25.56 ± 0.64, P1-Cα: 23.30 ± 0.62).

The crystal structure of **cn-716** in complex with the WNV NS2B-NS3$^{pro}$ protease revealed the enzyme in the closed (active) conformation, with the NS2B chain wrapped around the NS3 domain (Fig. 3b.2C; PDB entry 5IDK). The backbone of the inhibitor forms only two hydrogen bonds to the protease (P1 NH...Gly151 O, P3 O...Gly153 NH), whereas the P1 and P2 side-chains and the benzoyl cap are neatly accommodated in the corresponding subsites of the enzyme (Fig. 3b.2D). The terminal $NH_2$ group of the P2 side-chain is involved in hydrogen bonds with the main-chain oxygen of Asp82* and with Asn84* Oδ1, both from the NS2B polypeptide chain. These interactions presumably contribute to the stabilization of the closed form of the complex. Remarkably, the phenyl group of the P2 residue interacts with the imidazole of the catalytic His51 through π-π stacking (Fig. 3b.2D distance ~3.5 Å). The Nε atom of the P1 guanidinium group is involved in a 2.97 ± 0.08 Å hydrogen bond with the carbonyl oxygen of Tyr130, and the Nη2 donates a "charged" hydrogen bond to Asp129 Oδ1 (2.97 ± 0.22 Å; average and standard deviation calculated from the 3 monomers of the complex in the asymmetric unit of the crystal). Interestingly, the Nη1 forms a 3.18 ± 0.11 Å hydrogen bond with the carbonyl oxygen of the inhibitor’s P2 residue, drawing the P1 guanidinium group close to the main chain of the compound. Moreover, there may be a weak (parallel) interaction between the guanidinium group and the phenyl group of Tyr161 (Fig. 3b.2D).

The boron atom forms a 1.59 ± 0.01 Å covalent bond with Oγ of the active-site Ser135 (Fig. 3b.2D, upper right inset). During refinement, additional positive $F_o$-$F_c$ density emerged near the two hydroxyl groups of the boronic acid and was unambiguously assigned to a cyclic double ester formed between the boronic acid and atoms O1 and O2 of glycerol, leading to the formation of a 5-membered ring (Fig. 3b.2D, lower left inset). Boronic acids tend to form esters with aliphatic diols, especially if 5- or 6-membered rings can be formed (Türker, 2006), and glycerol was present in my enzyme preparation during purification, crystallization, and cryo-protection of crystals.

Negatively charged and chiral, the tetrahedral boronate unit is a close mimic of the transition state of peptide hydrolysis. The cyclic-ester oxygens (O1 and O2) accept hydrogen bonds from the amide of Gly133 (2.85 ± 0.05 Å; part of the oxyanion hole) and the Nε2 of His51 (2.67 ± 0.04 Å), respectively. The five-membered ring occupies the S1' site, but the hydroxyl group (O3 of glycerol) is oriented towards bulk solvent (Fig. 3b.2D). Diffraction data collection and final model refinement statistics are summarized in Table 3b.1.

**Table 3b.1 Crystallographic data collection and refinement statistics of structures of DENV2 NS2B-NS3$^{pro}$ free enzyme, and WNV NS2B-NS3 $^{pro}$ in complex with cn-716**

| Protease | DENV2 NS2B-NS3$^{pro}$ | WNV NS2B-NS3$^{pro}$ with cn-716 |
|---|---|---|
| **Data collection statistics** | | |
| X-ray source | DESY P11 | DESY P11 |
| Wavelength [Å] | 0.9919 | 0.9919 |
| Space group | $C222_1$ | $P2_12_12_1$ |
| Unit cell dimensions [Å] | a = 61.07,<br>b = 61.35,<br>c = 113.8 | a = 36.42,<br>b = 96.96,<br>c = 187.86 |
| Number of complex molecules per asymmetric unit | 1 | 3 |
| Resolution range [Å] | 43.28 -1.75 (1.79-1.75) | 46.94 - 1.50 (1.58 - 1.50) |
| Number of observations | 135,230 (6,493) | 350,519 (51,736) |
| Number of unique reflections | 21,796 (1,129) | 105,999 (15,393) |
| Completeness [%] | 99.7 (95.8) | 98.6 (99.3) |
| Mean $I/\sigma(I)$ | 20.7 (2.1) | 15.6 (2.5) |
| Multiplicity | 6.2 (5.8) | 3.3 (3.4) |
| $R_{merge}$ [%] | 0.048 (0.751) | 0.045 (0.557) |
| $R_{pim}$ [%] | 0.031 (0.502) | 0.028 (0.344) |
| $CC_{1/2}$ | 0.999 (0.619) | 0.999 (0.749) |
| **Refinement statistics** | | |
| $R_{cryst}$ /$R_{free}$ [%] | 18.21/22.96 | 15.92 / 18.58 |
| r.m.s. deviation in bond lengths (Å) | 0.0191 | 0.0252 |
| r.m.s. deviation in bond angles (°) | 1.9492 | 2.3429 |
| Average B-factor for all atoms (Å$^2$) | 31.31 | 21.04 |
| Number of protein atoms | 1487 | 4711 |
| Number of inhibitor atoms | 0 | 111 |
| Number of water atoms | 88 | 637 |
| **Ramachandran plot** | | |
| Residues in preferred regions (%) | 98.39 | 96.42 |
| Residues in additionally allowed regions (%) | 1.61 | 3.41 |
| Residues in outlier regions (%) | 0.00 | 0.17 |

## 3b.3 ZIKV NS2B-NS3$^{pro}$ part 1

This portion of the work has been published in part: Lei J., Hansen G., Nitsche C., Klein C. D., Zhang L., Hilgenfeld R. (2016) Crystal structure of Zika virus NS2B-NS3$^{pro}$ protease in complex with a boronate inhibitor. *Science* 353, 503-505.

### 3b.3.1 Recombinant production of ZIKV NS2B-NS3$^{pro}$

A gene construct coding for residues 49-95 of ZIKV NS2B, linked via a Gly$_4$-Ser-Gly$_4$ sequence to the N-terminal part of ZIKV NS3 (1-170; GenBank accession number KU729217.2) was expressed in *E. coli*. As shown in Fig 3b.3A, the recombinant protein bound the nickel column. The purity of the protein was analyzed by SDS-PAGE (Fig. 3b.3B). After digestion of the His$_6$ tag by thrombin, a second-round purification was performed by ion-exchange chromatography (Fig 3b.3A), and the purity is shown in Fig 3b.3B. Finally, the protein was loaded onto a size-exclusion chromatography column for a third-round purification. The chromatography (Fig 3b.3A and B) revealed that the preparation was a mixture of monomer and disulfide-linked dimer (here designated “SS-dimer”), and - to a lesser extent - higher oligomers.

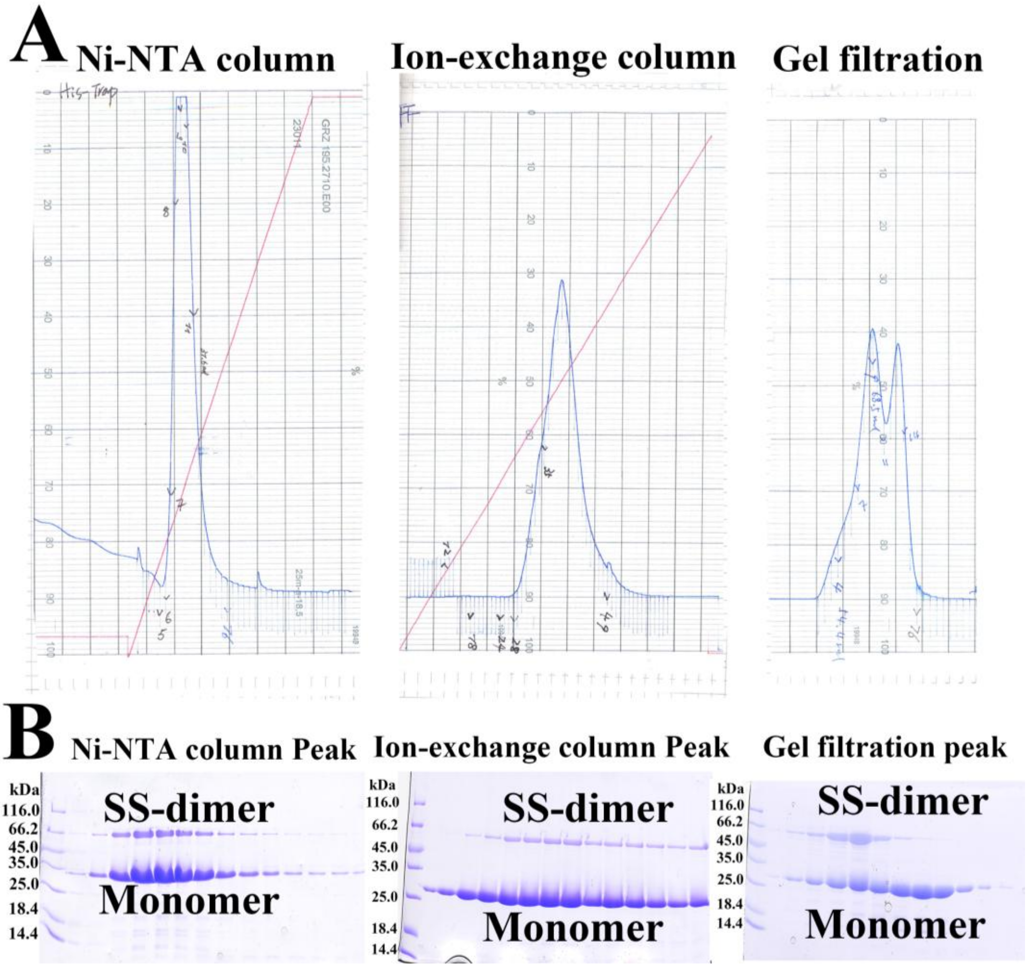


**Fig. 3b.3 Purification of ZIKV NS2B-NS3$^{pro}$**

**(A)** Purification of ZIKV NS2B-NS3$^{pro}$ by using Ni-NTA chromatography, ion-exchange chromatography, and gel filtration (Superdex$^{TM}$ 75). The profile of the gel filtration indicates that the ZIKV NS2B-NS3$^{pro}$ is a mixture of monomer and homodimer in a solution without reductant.

**(B)** The corresponding chromatographic fractions were analyzed by SDS-PAGE. The upper band is in the position for the disulfide-bonded homodimer, and the lower band corresponds to the monomer. Similar to other flavivirus free NS2B-NS3$^{pro}$ proteases (Leung et al., 2001), the free ZIKV protease migrates more slowly than expected from the molecular weight (~24 kD).

### 3b.3.2 Inhibitory activity of cn-716 against ZIKV NS2B-NS3$^{pro}$ (monomer)

The inhibitory activity of **cn-716** against ZIKV NS2B-NS3$^{pro}$ (wt, monomer) was determined in the buffer with glycerol (20%, v/v) and reductant (1 mM TCEP). I determined the *in-vitro* activity of **cn-716** against the ZIKV NS2B-NS3$^{pro}$ wt. Boronic acids tend to form esters with diols and triols, especially if 5- or 6-membered rings can be formed

(Türker, 2006). The complex structure (Lei et al., 2016; see 3b.2) reveals that the boronic-acid moiety forms a cyclic diester (six-membered ring) with glycerol, which was continuously present in our enzyme preparation during purification and crystallization, as well as cryoprotection of crystals (Lei et al., 2016; PDB entry 5LC0; Fig. 3b.4 A). Thus, I speculate that it was a dipeptide boronate instead of the free boronic acid, which reversibly bound to the ZIKV NS2B-NS3$^{pro}$ with $IC_{50} = 0.25 \pm 0.02$ µM and $K_i = 0.040 \pm 0.006$ µM (Fig. 3b.4A and B). In the absence of glycerol, the $IC_{50}$ for the boronic acid inhibitor was nearly unchanged (0.20 ± 0.02 µM; Fig. 3b.4C). We note that through ester formation with larger, more hydrophobic diols or triols, a prodrug might be obtained that will traverse the cellular membrane more readily than free boronic-acid derivatives.

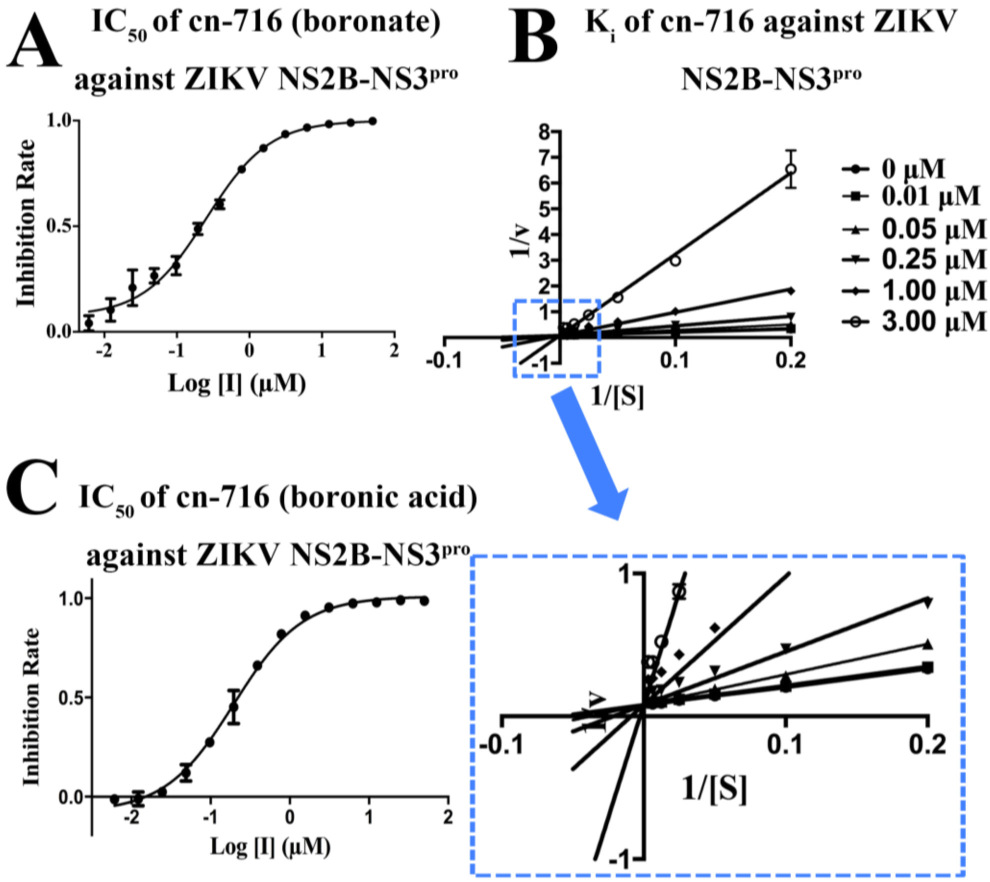


**Fig. 3b.4 $IC_{50}$ and $K_i$ curves of cn-716 against ZIKV NS2B-NS3$^{pro}$.**
**(A)** $IC_{50}$ plot of **cn-716** against ZIKV NS2B-NS3$^{pro}$ wt (monomer) in 10 mM Tris-HCl, 20% glycerol (v/v), 1 mM CHAPS, 1 mM TCEP, pH 8.5.
**(B)** $K_i$ plot of **cn-716** against ZIKV NS2B-NS3$^{pro}$ wt (monomer) in the same buffer. The Lineweaver-Burk plot reveals **cn-716** as a competitive inhibitor.
**(C)** $IC_{50}$ plot of **cn-716** against ZIKV NS2B-NS3$^{pro}$ wt (monomer) in the buffer lacking of glycerol.
[S]: substrate concentration (μM); v: initial reaction rate (μM/min). The curves were generated using the GraphPad Prism 6.0 software (GraphPad).

### 3b.3.3 Kinetic parameters of variants of ZIKV NS2B-NS3$^{pro}$

The SS-dimer and the monomer obtained by reduction with TCEP, as well as the Cys80Ser/Cys143Ser mutant of ZIKV NS2B-NS3$^{pro}$, were hyper-active against the standard flavivirus protease substrate Bz-Nle-Lys-Lys-Arg-AMC, with a low $K_m$ and a $k_{cat}/K_m$ more than 20-times higher than for the WNV enzyme (Table 3b.2).

**Table 3b.2 Kinetic parameters of variants of ZIKV NS2B-NS3$^{pro}$, in comparison to a similar WNV NS2B-NS3$^{pro}$ construct.**

| Protease | $k_{cat}$ ($s^{-1}$) | $K_m$ (µM) | $k_{cat}/K_m$ ($s^{-1}M^{-1}$) |
|---|---|---|---|
| **ZIKV NS2B-NS3$^{pro}$** | | | |
| Monomer (wt) | 44.6 ± 1.0 | 18.3 ± 1.6 | 2,440,444 ± 215,000 |
| SS-dimer (wt) | 28.5 ± 0.6 | 5.9 ± 0.5 | 4,850,000 ± 429,000 |
| Cys80Ser/Cys143Ser | 28.8 ± 0.5 | 5.1 ± 0.5 | 5,620,000 ± 546,000 |
| Asp83*Asn (monomer) | 38.5 ± 1.4 | 35.3 ± 4.2 | 1,091,000 ± 136,000 |
| **WNV NS2B-NS3$^{pro}$** | | | |
| wt (my preparation) | 8.7 ± 0.1 | 77.4 ± 3.6 | 112,000 ± 5000 |
| wt (Erbel et al., 2006) | -- | -- | 37,000 ± 7000 |
| **DENV2 NS2B-NS3$^{pro}$** | | | |
| wt (Erbel et al., 2006) | -- | -- | 30,000 ± 7000 |

Data are for the cleavage of the flavivirus protease substrate Bz-Nle-Lys-Lys-Arg-AMC. “Monomer (wt)” (See Material and Method part on ‘recombinant production and purification of ZIKV NS2B-NS3$^{pro}$ for details, including definition of “wt”) and “SS-dimer” indicate enzyme preparations corresponding to the monomer (in the presence of TCEP) and the SS-dimer fraction from gel permeation chromatography. The kinetic parameters for the ZIKV protease with Asp83* replaced by Asn are also included. “WNV NS2B-NS3$^{pro}$ (wt)” is our recombinant preparation of the WNV protease. For comparison, the $k_{cat}/K_m$ values for WNV and DENV-2 NS2B-NS3$^{pro}$ and with the substrate Bz-Nle-Lys-Arg-Arg-AMC reported in Erbel et al. (2006) are given. (Dashed line indicates not reported). All values in this table are obtained at pH 8.5.

Most probably protonated, the amino group of the 4-aminomethyl-phenylalanyl residue in the P2 position forms a salt-bridge with Asp83* of the NS2B polypeptide (Lei et al., 2016; PDB entry 5LC0; Fig. 3b.5) In addition, according to the sequence alignment, Asp83* is Asn in WNV and Ser or Thr in DENV 1-4 NS2B-NS3$^{pro}$ (Fig. 3b.6), i.e., unable to form an ion-pair interaction with the P2 residue of the inhibitor or the substrate, Bz-Nle-Lys-Lys-Arg-AMC. Therefore, Asp83* might account for the low $K_m$ and super-high activity of the ZIKV NS2B-NS3$^{pro}$. We therefore produced the Asp83*Asn variant of ZIKV NS2B-NS3$^{pro}$ and determined the kinetic parameters. The Asp83*Asn mutation leads to a ~2-fold increase of $K_m$ and a reduction of $k_{cat}/K_m$ by 55%, compared to the wt enzyme (Table 3b.2). The Asp residue in this position provides an at least partial explanation for the significantly lower $K_m$ and hence, the much higher $k_{cat}/K_m$ of ZIKV protease compared to the WNV and DENV enzymes (Table 3b.2).

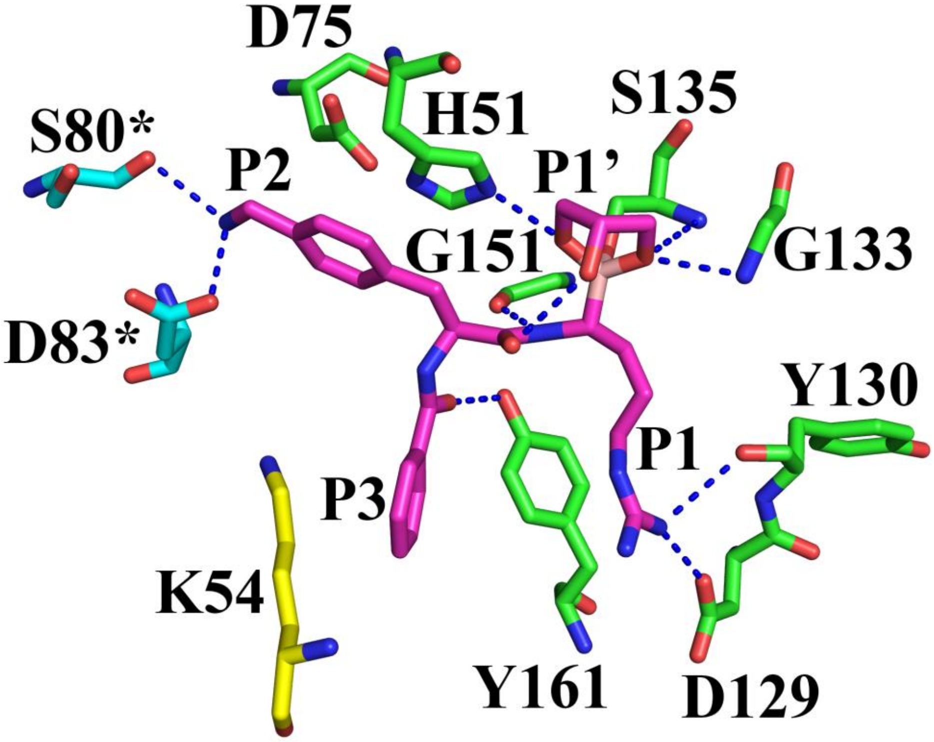
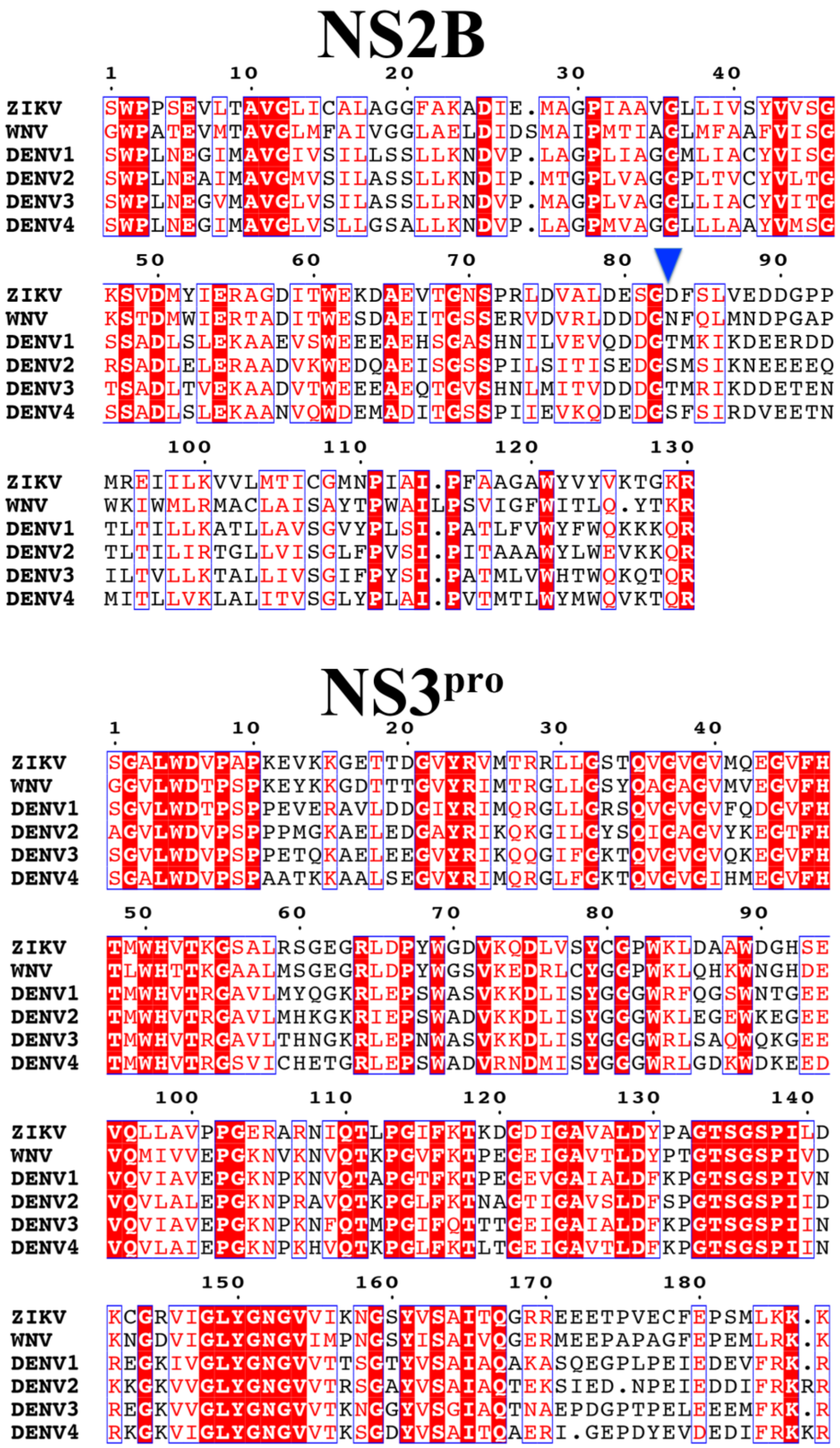


**Fig. 3b.5 Crystal structure of ZIKV NS2B-NS3$^{pro}$ in complex with cn-716.** Detailed interactions of **cn-716** with ZIKV NS2B-NS3$^{pro}$. Residues of the protease and inhibitor are indicted as sticks. Hydrogen bonds are depicted as blue dashed lines. Residues from NS2B are marked with an asterisk (*). The inhibitor (**cn-716**) is indicated in magenta, NS3 in green, NS2B in cyan, and Lys54 from the other molecule in the "tight dimer" is represented in yellow.



```
NS2B

         1        10        20         30        40
ZIKV     SWPPSEVLTAVGLICALAGGFAKADIE.MAGPIAAVGLLIVSYVVSG
WNV      GWPATEVMTAVGLMFAIVGGLAELDIDSMAIPMTIAGLMFAAFVISG
DENV1    SWPLNEGIMAVGIVSILLSSLLKNDVP.LAGPLIAGGMLIACYVISG
DENV2    SWPLNEAIMAVGMVSILASSLLKNDIP.MTGPLVAGGPLTVCYVLTG
DENV3    SWPLNEGVMAVGLVSILASSLLRNDVP.MAGPLVAGGLLIACYVITG
DENV4    SWPLNEGIMAVGLVSLLGSALLKNDVP.LAGPMVAGGLLLAAYVMSG

           50        60        70        80        90
ZIKV     KSVDMYIERAGDITWEKDAEVTGNSPRLDVALDESGDFSLVEDDGPP
WNV      KSTDMWIERTADITWESDAEITGSSERVDVRLDDDGNFQLMNDPGAP
DENV1    SSADLSLEKAAEVSWEEEAEHSGASHNILVEVQDDGTMKIKDEERDD
DENV2    RSADLELERAADVKWEDQAEISGSSPILSITISEDGSMSIKNEEEEQ
DENV3    TSADLTVEKAADVTWEEEAEQTGVSHNLMITVDDDGTMRIKDDETEN
DENV4    SSADLSLEKAANVQWDEMADITGSSPIIEVKQDEDGSFSIRDVEETN

             100       110       120       130
ZIKV     MREIILKVVLMTICGMNPIAI.PFAAGAWYVYVKTGKR
WNV      WKIWMLRMACLAISAYTPWAILPSVIGFWITLQ.YTKR
DENV1    TLTILLKATLLAVSGVYPLSI.PATLFVWYFWQKKKQR
DENV2    TLTILIRTGLLVISGLFPVSI.PITAAAWYLWEVKKQR
DENV3    ILTVLLKTALLIVSGIFPYSI.PATMLVWHTWQKQTQR
DENV4    MITLLVKLALITVSGLYPLAI.PVTMTLWYMWQVKTQR

NS3pro

         1        10        20        30        40
ZIKV     SGALWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQVGVGVMQEGVFH
WNV      GGVLWDTPSPKEYKKGDTTTGVYRIMTRGLLGSYQAGAGVMVEGVFH
DENV1    SGVLWDTPSPPEVERAVLDDGIYRIMQRGLLGRSQVGVGVFQDGVFH
DENV2    AGVLWDVPSPPPMGKAELEDGAYRIKQKGILGYSQIGAGVYKEGTFH
DENV3    SGVLWDVPSPPETQKAELEEGVYRIKQQGIFGKTQVGVGVQKEGVFH
DENV4    SGALWDVPSPAATKKAALSEGVYRIMQRGLFGKTQVGVGIHMEGVFH

          50        60        70        80        90
ZIKV     TMWHVTKGSALRSGEGRLDPYWGDVKQDLVSYCGPWKLDAAWDGHSE
WNV      TLWHTTKGAALMSGEGRLDPYWGSVKEDRLCYGGPWKLQHKWNGHDE
DENV1    TMWHVTRGAVLMYQGKRLEPSWASVKKDLISYGGGWRFQGSWNTGEE
DENV2    TMWHVTRGAVLMHKGKRIEPSWADVKKDLISYGGGWKLEGEWKEGEE
DENV3    TMWHVTRGAVLTHNGKRLEPNWASVKKDLISYGGGWRLSAQWQKGEE
DENV4    TMWHVTRGSVICHETGRLEPSWADVRNDMISYGGGWRLGDKWDKEED

           100       110       120       130       140
ZIKV     VQLLAVPPGERARNIQTLPGIFKTKDGDIGAVALDYPAGTSGSPILD
WNV      VQMIVVEPGKNVKNVQTKPGVFKTPEGEIGAVTLDYPTGTSGSPIVD
DENV1    VQVIAVEPGKNPKNVQTAPGTFKTPEGEVGAIALDFKPGTSGSPIVN
DENV2    VQVLALEPGKNPRAVQTKPGLFKTNAGTIGAVSLDFSPGTSGSPIID
DENV3    VQVIAVEPGKNPKNFQTMPGIFQTTTGEIGAIALDFKPGTSGSPIIN
DENV4    VQVLAIEPGKNPKHVQTKPGLFKTLTGEIGAVTLDFKPGTSGSPIIN

             150       160       170       180
ZIKV     KCGRVIGLYGNGVVIKNGSYVSAITQGRREEETPVECFEPSMLKK.K
WNV      KNGDVIGLYGNGVIMPNGSYISAIVQGERMEEPAPAGFEPEMLRK.K
DENV1    REGKIVGLYGNGVVTTSGTYVSAIAQAKASQEGPLPEIEDEVFRK.R
DENV2    KKGKVVGLYGNGVVTRSGAYVSAIAQTEKSIED.NPEIEDDIFRKRR
DENV3    REGKVVGLYGNGVVTKNGGYVSGIAQTNAEPDGPTPELEEEMFKK.R
DENV4    RKGKVIGLYGNGVVTKSGDYVSAITQAERI.GEPDYEVDEDIFRKKR
```

**Fig. 3b.6 Multiple sequence alignment of the NS2B/NS3 proteases of ZIKV, WNV as well as DENV 1-4.** Residues at position 83 of NS2B that are involved in inhibitor binding are indicated by the blue triangle. The GenBank codes for the flavivirus strains shown here are as follow: ZIKV: KU729217.2; WNV: P06935.2; DENV1: P33478.1; DENV2: P29990.1; DENV3: P27915.1; DENV4: P09866.2. The figure was generated with the program ESPript (Gouet et al., 1999).

## 3b.4 ZIKV NS2B-NS3$^{pro}$ part 2: characterization of cleavage properties of NS2B-NS3$^{pro}$ with FRET substrates

This part of the work has been described in the following manuscript: Zhang L., Gül Y., Hilgenfeld R. Zika virus NS2B-NS3 protease is a hyperactive enzyme, to be submitted (2016).

In order to investigate the cleavage properties of the ZIKV NS2B-NS3$^{pro}$ against the native or putative cleavage sites in the polyprotein, 9 FRET substrates carrying a Dabcyl quencher and an Edans donor pair spanning from the P4 to the P4’ sites (Dabcyl-KXX(K/R)R↓XXXX-Edans-$NH_2$, X represents any amino acid) were ordered from Biosyntan (Biosyntan GmbH). Among the 9 FRET substrates, five were designed according to the putative cleavage sites in the polyprotein precursor (capsid, NS2A/NS2B, NS2B/NS3, NS3/NS4A, and NS4B/NS5), and the other four corresponded to the cleavage site of furin at prM/M, as well as the putative cleavage site inside the NS3 protease, the putative cleavage site between the NS3 protease and helicase, and the flavivirus protease cleavage site in the signalling protein STING.

### 3b.4.1 Inner-filter-effect correction

The inner-filter effect is the phenomenon that the measured relative fluorescence units (RFU) are not proportional to the concentration of the substrate due to the high quencher density (Lakowicz, 1983). At high concentration of substrate, the quencher group may absorb the fluorescence signal emitted from the neighboring free fluorescent group. In this study, I used the method reported by Liu et al. (1999) for correction of the inner-filter effect.

The inner-filter corrections of the FRET substrate 8 (NS4B/NS5) is shown as an example in Table 3b.3. The absorbed RFU of the free Edans was measured at different concentrations of each substrate applied for determination of the kinetic parameters. Data in Table 3b.3 indicate that the inner-filter effect is small at low concentration of the substrate (2.5-5 μM), as the RFU of 5 μM free Edans was only reduced by less than 10%. However, the effect can be very large at a higher substrate concentration, e. g. the fluorescence signal of the free Edans was absorbed to an extent of around 70% when the substrate was at 160 μM. Different concentrations of substrate were tested using the Michaelis-Menten equation for the calculation of the $K_m$ and $k_{cat}$ parameters. Results in Table 3b.4 show that the non-corrected $K_m$ and $k_{cat}$ value were 1.18 μM and 8.62 $min^{-1}$, respectively. However, the $K_m$ and $k_{cat}$ values corrected for the inner-filter effect were 31.06 μM and 6.07 $min^{-1}$, respectively. The non-corrected $k_{cat}/K_m$ (2279.4 $s^{-1}M^{-1}$) was decreased by to 30% compared to the corrected value (3257.1 $s^{-1}M^{-1}$). Therefore, the inner-filter effect led to an earlier saturation state in the enzymatic assay, resulting in a smaller value of the enzymatic parameters $K_m$ and $k_{cat}$.

### Table 3b.3 Inner-filter effect correction

| Substrate conc. (μM) | f(S) | f(S+Edans) | Δf | Detected% | $V_o$ (μM/min) | $V_{corr.}$ (μM/min) |
|---|---|---|---|---|---|---|
| 160 | 35169 | 36743 | 1574 | 33.82 | 1.79 | 5.29 |
| 80 | 23252 | 25610 | 2358 | 50.66 | 2.05 | 4.05 |
| 40 | 14217 | 17222 | 3005 | 64.56 | 2.58 | 3.99 |
| 20 | 8058 | 11856 | 3798 | 81.60 | 2.19 | 2.69 |
| 10 | 4231 | 8152 | 3921 | 84.23 | 1.37 | 1.63 |
| 5 | 2263 | 6473 | 4210 | 90.44 | 0.42 | 0.47 |
| 2.5 | 1383 | 5693 | 4310 | 92.60 | 0.014 | 0.015 |
| 0 | 241 | 4896 | 4655<br>f (free Edans) | 100 | -- | -- |

Substrate sequence: Dabcyl-KVKRR↓GGGTE(Edans)-$NH_2$ (NS4B/NS5)
Inner filter effect correction formula: Detected% = Δf / f(free Edans)
$V_{corr.} = V_o$ / Detected%

### Table 3b.4 Influence of the inner-filter effect on the kinetic parameters of the ZIKV NS2B-NS3$^{pro}$-catalyzed cleavage of FRET substrate corresponding to the NS4B/NS5 cleavage site.

| Cleavage site | $k_{cat}$ ($min^{-1}$) | $K_m$ (μM) | $k_{cat}/K_m$ ($s^{-1}M^{-1}$) | Sequence |
|---|---|---|---|---|
| NS4B/NS5 (corrected)* | 6.07 ± 0.36 | 31.06 ± 5.095 | 3257.1 ± 568.1 | Dabcyl-KPLRR↓GALLE(Edans)-$NH_2$ |
| NS4B/NS5 (non-corrected)# | 1.18 ± 0.11 | 8.62 ± 3.24 | 2279.4 ± 883.41 | Dabcyl-KPLRR↓GALLE(Edans)-$NH_2$ |

* The inner-filter correction was applied for the calculation of the kinetic parameters.
# The inner-filter correction was not applied for the calculation of the kinetic parameters.

### 3b.4.2 Characterization of the kinetic parameters with different substrates

The kinetic parameters were screened with 9 substrates, and the results are shown in Table 3b.5 and Fig. 3b.7. The order of the turn-over value ($k_{cat}$) of substrates corresponding to the presumable cleavage sites in the polyprotein precursor is NS2A/NS2B (11.29 $min^{-1}$) > NS3/NS4A (8.63 $min^{-1}$) > NS4B/NS5 (6.07 $min^{-1}$) ≈ NS2B/NS3 (6.04 $min^{-1}$) > capsid (3.16 $min^{-1}$). On the other side, the $K_m$ value varies differently from the order of $k_{cat}$: NS2A/NS2B (81.84 μM) > NS3/NS4A (37.09 μM) > NS2B/NS3 (32.10 μM) ≈ NS4B/NS5 (31.06 μM) > capsid (19.89 μM). Four out of five presumable cleavage sites possess the Gly-Lys-Arg and one (NS4B-NS5) has a Lys-Arg-Arg motif for the P3-P1 sequence, and similar to proteases of other flaviviruses (YFV, WNV, and DENV; Li et al., 2005; Rice et al., 1985; Wengler et al., 1991), dibasic residues at the P1-P2 sites. Although there are slight differences of the $K_m$ between the NS2B/NS3, NS3/NS4A, and NS4B/NS5 sites, the values are comparable. However, the $K_m$ of NS2A/NS2B is around 2.5 times higher than the former; this might be due to the replacement of the small hydrophobic amino acid (Gly, Ala, or Val) by a Trp at the P2' site, which might reduce the binding affinity. A basic amino acid (Lys) at the P4 position of the capsid substrate is unique compared to the other native substrates, and results in a much smaller $K_m$ and a higher binding affinity even overcome the possible reducing of the binding affinity due to the Asp at the P2' site. Comprehensively considering $K_m$ and $k_{cat}$, the protease cleaves the NS3/NS4A substrate most efficiently ($k_{cat}/K_m$ = 3878.0 $s^{-1}M^{-1}$), followed by NS4B/NS5 ($k_{cat}/K_m$ = 3257.0 $s^{-1}M^{-1}$). The processing efficiency of the other three presumable substrates (capsid, NS2B/NS3, and NS2A/NS2B) is lower than the two former ones, and varies slightly with the $k_{cat}/K_m$ value being 2647.9, 2421.5, and 2299.2 $s^{-1}M^{-1}$, respectively (Table 3b.5, and Fig. 3b.7).

**Table 3b.5 Kinetic parameters of ZIKV NS2B-NS3$^{pro}$ (wt, monomer) against FRET substrates derived from cleavage sites in the polyprotein.**

| Cleavage site | $k_{cat}$ ($min^{-1}$) | $K_m$ (μM) | $k_{cat}/K_m$ or $K_{app}$ ($s^{-1}M^{-1}$) | Percentage | Sequence |
|---|---|---|---|---|---|
| **Capsid** | 3.16 ± 0.16 | 19.89 ± 3.09 | 2647.9 ± 432.6 | 68.28% | Dabcyl-KKKRR↓GADTE(Edans)-$NH_2$ |
| **prM/M (Furin)** | 8.79 ± 0.50 | 38.97 ± 5.82 | 3759.3 ± 600.8 | 96.94% | Dabcyl-KRSRR↓AVTLE(Edans)-$NH_2$ |
| **NS2A/NS2B** | 11.29 ± 0.43 | 81.84 ± 7.94 | 2299.2 ± 239.6 | 59.29% | Dabcyl-KSGKR↓SWPPE(Edans)-$NH_2$ |
| **NS2B/NS3** | 6.04 ± 0.24 | 32.10 ± 4.64 | 2421.5 ± 383.3 | 62.44% | Dabcyl-KTGKR↓SGALE(Edans)-$NH_2$ |
| **Intra-NS3$^{pro}$** | -- | -- | 74.2 ± 12.1* | 1.91% | Dabcyl-KMTRR↓LLGSE(Edans)-$NH_2$ |
| **NS3 (Protease/Helicase)** | 0 | 0 | 0 | 0% | Dabcyl-KQGRR↓EEETE(Edans)-$NH_2$ |
| **NS3/NS4A** | 8.63 ± 0.40 | 37.09 ± 4.59 | 3878.0 ± 512.0 | 100%* | Dabcyl-KAGKR↓GAAFE(Edans)-$NH_2$ |
| **NS4B/NS5** | 6.07 ± 0.36 | 31.06 ± 5.10 | 3257.1 ± 568.1 | 83.99% | Dabcyl-KVKRR↓GGGTE(Edans)-$NH_2$ |
| **STING** | 1.00 ± 0.11 | 33.22 ± 9.82 | 501.7 ± 158.2 | 12.94% | Dabcyl-KPLRR↓GALLE(Edans)-$NH_2$ |

* The $k_{cat}/K_m$ value determined from the substrate with the cleavage site between NS3/NS4A is defined as 100%.

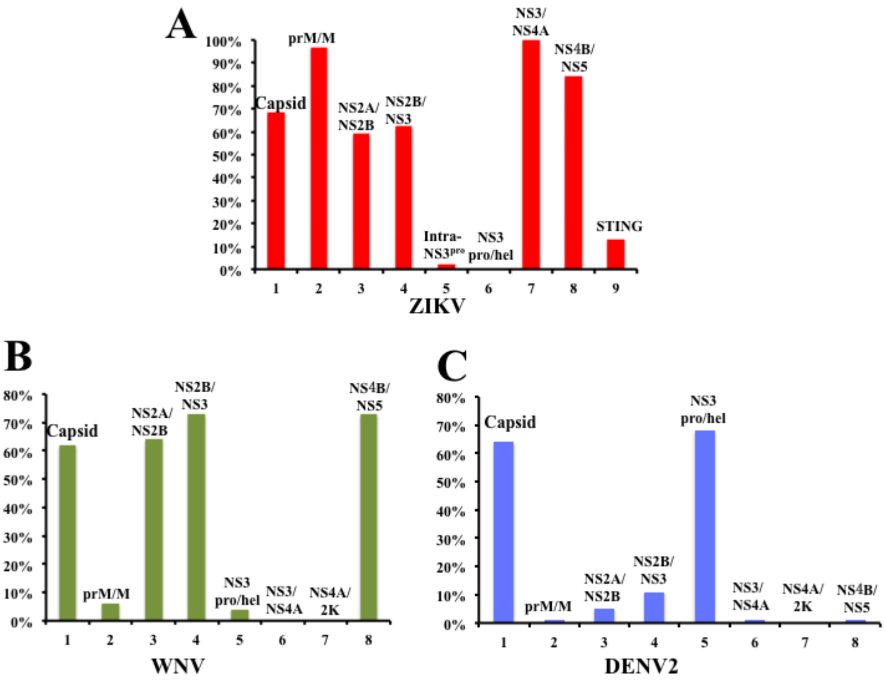


**Fig. 3b.7 Comparison of cleavage site preferences between ZIKV, WNV, and DENV2 NS2B-NS3$^{pro}$**

The cleavage efficiency of ZIKV NS2B-NS3$^{pro}$ turned out to be very different from results on WNV and DENV2 NS2B-NS3$^{pro}$ published by Shiryaev et al. (2007). The most efficient ZIKV NS2B-NS3$^{pro}$ cleavage site is that between NS3 and NS4A, followed by NS4B/NS5, and NS2B/NS3. There is no cleavage of the octapeptide corresonding to the linker between the NS3 protease and the NS3 helicase, in contrast to what Shiryaev et al. reported for the DENV2 enzyme. Moreover, the ZIKV NS2B-NS3$^{pro}$ could efficietly process the octapeptide corresponding to a potential cleavage site in the signaling protein STING.

The histograms of cleavage efficiency of WNV and DENV2 NS2B-NS3$^{pro}$ were prepared based on the values reported by Shiryaev et al (2007).

**(A)** The efficiency of the cleavage (%) of ZIKV NS2B-NS3$^{pro}$ against the substrates corresonding to presumable cleavage sites in ZIKV polyprotein. The $k_{cat}/K_m$ value determined from the substrate with the cleavage site between NS3/NS4A is defined as 100%.

**(B)** The efficiency of the cleavage (%) of WNV NS2B-NS3$^{pro}$ against the substrates corresonding to cleavage sites in the WNV polyprotein. The percentage values are reported by Shiryaev et al. (2007).

**(C)** The efficiency of the cleavage (%) of DENV2 NS2B-NS3$^{pro}$ against the substrates corresonding to cleavage sites in DENV2 polyprotein. The percentage values are reported by Shiryaev et al. (2007).

It is known that furin is involved in the cleavage of the prM/M in the polyprotein precursor (Stadler et al., 1997; Zheng et al., 2010). Surprisingly, $K_m$ (38.97 μM) and $k_{cat}$ (8.79) values for the prM/M substrate were similar to NS3/NS4A, and in particular, the high $k_{cat}/K_m$ (3759 $s^{-1}M^{-1}$) value indicated that the furin substrate could be well recognized and efficiently cleaved by the ZIKV NS2B-NS3$^{pro}$. However, according to the results published by Shiryaev et al. (2007), both WNV and DENV protease could cleave the furin substrate only with a considerably lower rate (Fig .3b.7A-C).

One putative cleavage site inside the NS3$^{pro}$ (MTRR$^{29}$↓L$^{30}$LGS) was also investigated, for which the $K_{app}$ (74.2 $s^{-1}M^{-1}$) was calculated as an approximation to the $k_{cat}/K_m$ parameter, as the cleavage efficiency was relatively low. Apparently, the sensitivity of this substrate is quite low compared to other substrates mentioned above, but a concentration accumulation effect could lead to serious autolysis when a high enzyme concentration was used for crystallization. Therefore, a mutant clone (Arg29Gly) was always employed to prevent autolysis, resulting in the successful crystallization of the ZIKV NS2B-NS3$^{pro}$ in complex with a boronate inhibitor (Lei et al., 2016).

To investigate whether the flexible linker region between the protease and helicase domains in the NS3 protein is cleavable, a substrate corresponding to the putative cleavage site (QGRR$^{170}$↓E$^{171}$EET), was tested. Results demonstrated that this putative cleavage site could not be recognized by the NS2B-NS3$^{pro}$ at all. This might be due to the presence of three consecutive Glu residues at the P1'-P3' sites, which are not comparable with the preference of ZIKV NS2B-NS3$^{pro}$ for small, hydrophilic amino acids at the P1' and P2' sites. This is in agreement with the cleavage efficiency of the putative junction between protease-helicase by the WNV NS2B-NS3$^{pro}$; in contrast, the DENV2 NS2B-NS3$^{pro}$ could digest the substrate corresponding to this junction efficiently (Fig. 3b.7A-C; Shiryaev et al., 2007).

According to my previous result, the catalytic activity of the ZIKV NS2B-NS3$^{pro}$ against the Bz-Nle-Lys-Lys-Arg-AMC was 20-fold higher than that of the WNV NS2B-NS3$^{pro}$. I also tested the kinetic parameters of the two proteases by using the substrate with the prime-side residues, derived from the cleavage site between ZIKV NS2B/NS3 in the polyprotein precursor. Similarly, the ZIKV NS2B-NS3$^{pro}$ exhibited a higher $k_{cat}$ than its WNV part (6.04 vs 0.24 min$^{-1}$), a smaller $K_m$ (32.10 vs 46.45 μM), and a 5.5-fold higher $k_{cat}/K_m$ (2421.5 vs 442.8 $s^{-1}M^{-1}$). Although the difference of the cleavage efficiency between ZIKV and WNV NS2B-NS3$^{pro}$ against the FRET substrate with the NS2B/NS3 cleavage site is not as large as in the case of the AMC substrate, the ZIKV protease is still superior to the WNV protease. As described in part 3b.3.3, Asp83* might account for the low $K_m$ and high activity. Therefore, the catalytic activity of ZIKV NS2B-NS3$^{pro}$ (Asp83*Asn) mutant against the FRET substrate with the NS2B/NS3 cleavage site was determined. The Asp83*Asn mutant leads to a slight increase of the $K_m$, and an about 2-fold decrease of the $k_{cat}$ (3.90 min$^{-1}$), and an approximately 1.4-fold decrease of the $k_{cat}/K_m$ (1749.7 $s^{-1}$ $M^{-1}$). Thus, the obvious decrease of the cleavage efficiency of the Asp83*Asn mutant against the AMC and FRET (NS2B/NS3) substrates revealed that Asp83* at S2 partially afforded the superior activity of the ZIKV NS2B-NS3$^{pro}$ (Table 3b.6).

Moreover, the catalytic activities of WNV NS2B-NS3$^{pro}$ against the FRET substrates derived from its own cleavage sites (NS2B/NS3, NS3/NS4A, NS4B/NS5) in the polyprotein were determined. Results revealed that the catalytic activity of WNV NS2B-NS3$^{pro}$ were much lower on average than ZIKV NS2B-NS3$^{pro}$ (Table 3b.7).

**Table 3b.6 Kinetic parameters of variants of ZIKV NS2B-NS3$^{pro}$ protease, in comparison to a similar WNV NS2B-NS3$^{pro}$ construct.**

| Protease | $k_{cat}$ ($min^{-1}$) | $K_m$ (μM) | $k_{cat}/K_m$ ($s^{-1}M^{-1}$) | Percentage |
|---|---|---|---|---|
| ZIKV NS2B-NS3$^{pro}$ (wt) | 6.04 ± 0.24 | 32.10 ± 4.64 | 2421.5 ± 383.3 | 100% |
| ZIKV NS2B-NS3$^{pro}$ (D83*N) | 3.90 ± 0.27 | 37.12 ± 5.47 | 1749.73 ± 285.02 | 72.25% |
| WNV NS2B-NS3$^{pro}$ | 0.24 ± 0.02 | 46.45 ± 8.87 | 442.8 ± 94.8 | 18.29% |

Data are for the cleavage of the FRET substrate designed from the cleavage site between NS2B/NS3: Dabcyl-KTGKR↓SGALE(Edans)-$NH_2$. The kinetic parameters for the ZIKV protease with Asp83* replaced by Asn are also included. "WNV NS2B-NS3$^{pro}$ (wt)" is my recombinant preparation of the WNV protease. For comparison, the kinetic parameters for WNV NS2B-NS3$^{pro}$ were also determined with the same ZIKV NS2B-NS3$^{pro}$ substrate. All values in this table are obtained at pH 8.5.

**Table 3b.7 Kinetic parameters of WNV NS2B-NS3$^{pro}$ against FRET substrates derived from cleavage sites in its polyprotein and STING.**

| Cleavage site | $k_{cat}$ ($min^{-1}$) | $K_m$ (μM) | $k_{cat}/K_m$ ($s^{-1}M^{-1}$) | Percentage | Sequence |
|---|---|---|---|---|---|
| **NS2B/NS3** | 0.71 ± 0.04 | 9.37 ± 1.91 | 1264.81 ± 265.36 | 52.23% | Dabcyl-KYTKR↓GGVLE(Edans)-$NH_2$ |
| **NS3/NS4A** | 0.38 ± 0.02 | 72.04 ± 9.91 | 87.15 ± 13.23 | 2.25% | Dabcyl-KSGKR↓SQIGE(Edans)-$NH_2$ |
| **NS4B/NS5** | 1.24 ± 0.08 | 20.92 ± 3.95 | 989.48 ± 196.23 | 30.38% | Dabcyl-KGLKR↓GGAKE(Edans)-$NH_2$ |
| **STING** | 0.20 ± 0.02 | 61.19 ±10.47 | 55.54 ± 10.39 | 11.07% | Dabcyl-KPLRR↓GALLE(Edans)-$NH_2$ |

The percentage value was calculated as follows: the $k_{ca}t/K_m$ value of the WNV NS2B-NS3$^{pro}$ was divided by the $k_{ca}t/K_m$ value of ZIKV NS2B-NS3$^{pro}$ against the corresponding FRET substrate.

It has been shown that DENV2 NS2B/NS3$^{pro}$ could counteract the type-I interferon production via cleavage of the stimulator of interferon genes (STING) in human monocyte- derived dendritic cells (MDDCs) (Aguirre et al., 2012). In order to verify our speculation on the protlytic activity of ZIKV NS2B-NS3$^{pro}$ against STING, a FRET substrate corresponding to the putative cleavage site in STING of the ZIKV NS2B-NS3$^{pro}$ (Dabcyl-KPLRR↓GALLE(Edans)-$NH_2$) was ordered from Biosyntan (Biosyntan GmbH). The ZIKV NS2B-NS3$^{pro}$ could efficiently cleave the substrate with $K_m = 33.22 \pm 9.82$ μM, and $k_{cat}/K_m = 501.7 \pm 158.2$ $s^{-1}M^{-1}$ (Table 3b.5). However, the catalytic activity of WNV NS2B-NS3$^{pro}$ against this substrate is much lower compared to ZIKV NS2B-NS3$^{pro}$ ($K_m = 61.19 \pm 10.47$ μM, $k_{cat}/K_m = 55.54 \pm 10.39$ $s^{-1}M^{-1}$; Table 3b.7). The efficient cleavage of the substrate derived from STING suggests the possibility that ZIKV NS2B/NS3$^{pro}$ digests STING *in vivo*. In addition, as ZIKV has also been shown to permissively infect human DCs (Hamel et al., 2015), I deduce that the super-high activity of the ZIKV NS2B-NS3$^{pro}$ could lead to more efficient cleavage of STING, thereby enhancing the suppression of the host innate immune response compared to WNV.

# 4 Discussion

## 4.1 Coronavirus part

All coronaviruses require the proteolytic activity of the Nsp5 protease (also known as $M^{pro}$ or $3CL^{pro}$) to process the polyprotein during viral replication, which makes the $M^{pro}$ an attractive drug target. As the coronavirus $M^{pro}$s have a similar fold to the enterovirus 3C proteases, they are also called $3CL^{pro}$. The $M^{pro}$s resemble the chymotrypsin-like fold of the enteroviral $3C^{pro}$, except for an additional α-helical domain that is involved in the dimerization of the protease. Compounds active against the $M^{pro}$s have the potency to show good activities against the $3C^{pro}$.

In this thesis, the free enzyme crystal structure of HCoV-NL63 $M^{pro}$ was determined. This revealed an overall fold similar to SARS-CoV $M^{pro}$, but with significant differences in the S2 pocket. As a Met is missing at position 49, and Gln was replaced by Pro at position 189 in HCoV-NL63 $M^{pro}$, a more rigid S2 pocket resulted. Thus, the two lead compounds **11a** and **SG85** were inactive against HCoV-NL63 $M^{pro}$. Remarkably, the structures of HCoV-NL63 $M^{pro}$ in complex with **11a** and **SG85** were determined in spite of the poor inhibitory activities, revealing that the large P2 benzyl group of the two compounds could not be properly accommodated in the rigid S2 pocket. Therefore, an HCoV-NL63 $M^{pro}$ mutant (+M49, P189Q) was designed to generate a SARS-CoV $M^{pro}$-like pocket. The inhibitory activities of both compounds against this mutant were improved, and the P2 group was neatly accommodated in the S2 pocket according to the complex structures.

One of the lead compounds tested against the $M^{pro}$ of HCoV NL63 was α-ketoamide compound **11a**. Some α-ketoamides have been successfully developed into drugs, e.g. telaprivir and boceprivir for the treatment of HCV infection. These compounds are peptidomimetic inhibitors of the HCV NS3/NS4 serine protease (Lin et al., 2006; Romano et al., 2012;

Venkatraman et al., 2006). However, the α-ketoamides were only occasionally described as inhibitors for viral cysteine proteases.

A single capped dipeptidyl α-ketoamide (GC-375; Table 4a.1) was reported by Kim et al. (2012), which showed one-digit micromolar $IC_{50}$ values against several corona- and picornavirus proteases. In addition, this compound also showed micromolar-range $EC_{50}$ values in cell culture against coronaviruses (TGEV, HCoV 229E, FIPV, MHV, and BCV), but poor activity against enterovirus A71 (EV-A71). Further, a number of capped dipeptidyl α-ketoamides were reported as inhibitors of the norovirus $3CL^{pro}$ (Mandadapu et al., 2012). These inhibitors carry a P1 5-membered γ-lactam and P2-isobutyl in most cases, and a Cbz group as cap. These authors optimized the P1' substituent. Most of the inhibitors showed micromolar-range activities with both $IC_{50}$ and $EC_{50}$ values against the $3CL^{pro}$ and a norovirus replicon (representative compound 6d in Table 4a.1). However, there was not any data on the inhibitory activity against coronaviruses reported in this publication.

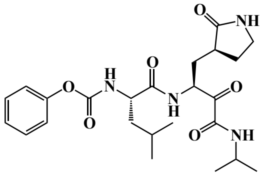
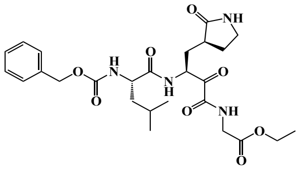
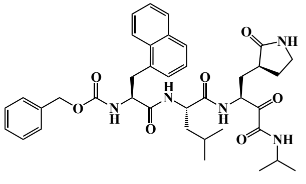
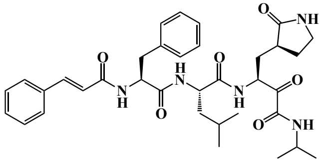
Another single capped tripeptidyl α-ketoamide (compound 6; Table 4a.1) among a series of aldehydes was described by Prior et al. (2013), which exhibited $IC_{50}$ values in the 3-digit nanomolar range against SARS-CoV $M^{pro}$ and HRV $3C^{pro}$, as well as $EC_{50}$ values of 0.5 μM against HCoV 229E and 0.03 μM against HRV18 in cell culture. However, no further optimization nor toxicity studies were performed or mentioned.

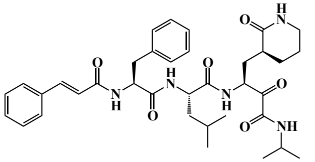
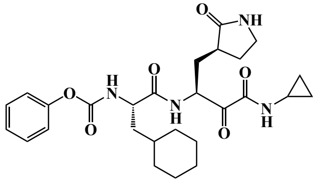
Zeng et al. (2016) reported a series of α-ketoamides against the EV-A71 $3C^{pro}$. They mainly focused on the structure-activity relationship (SAR) investigation of the P1' substituent. They found small alkyl groups to be superior to larger ones. Furthermore, they also replaced the 5-membered γ-lactam by a 6-membered δ-lactam, which led to 2-3 times higher activities (representative compounds 8a and 8f in Table 4a.1). At the same time, Kim et al. (2016) reported another five α-ketoamides with P1' as cyclopropyl, P1 also 5-membered γ-lactam, P2 as Leu, Phe, or cyclohexyl alanine (Cha), and Cbz or Cbz-Cl as P3 group (representative compound 10 in Table 4a.1). These inhibitors showed submicromolar activity against EV-D68 and two HRV strains. However, no data were reported against coronaviruses. Considering the limited data available on α-ketoamides in previous publications, it is of interest to produce broad-spectrum α-ketoamides, and evaluate the SAR against both coronavirus and enterovirus proteases.

Hence, I describe here the structure-based design and assessment of dipeptide α-ketoamides not only targeting the main proteases of alpha- or betacoronaviruses, but also the 3C proteases of enteroviruses. The canonical GlnLactam (γ-lactam) was chosen as P1 substitutent, benzyl in P1', and a cinnamoyl N-cap in the P3 position. Optimization focused on the P2 site.

Remarkably, according to the complex structures of α-ketoamides with SARS-CoV and HCoV-NL63 $M^{pro}$s as well as CVB3 $3C^{pro}$, the α-keto oxygen and the amide oxygen of the warhead (-CO-CO-NH-) accept two hydrogen bonds from the target protease, whereas other warheads such as Michael acceptors (-CH=CH-CO-) and aldehydes (-CH=O) have only one such acceptor. As a consequence, the α-ketoamide warhead is sterically more versatile than Michael acceptors and aldehydes. Alternatively, the amide oxygen of the α-ketoamide warhead could accept a hydrogen bond from either the catalytic His side-chain or the main-chain amides of the oxyanion hole, and *vice versa* for the α-keto oxygen. Depending on the exact interaction, the stereochemistry at the thiohemiketal C atom would be different. A similar difference in case of aldehyde inhibitors was observed previously, where the single hydroxyl group of the thiohemiacetal could interact either with the oxyanion hole or the catalytic His side-chain through hydrogen bonds (Zhu et al., 2011), resulting in a different stereochemistry of the thiohemiacetal carbon. Both α-ketoamides and aldehydes react reversibly with the catalytic nucleophile of proteases, whereas Michael acceptors form irreversible adducts. A big advantage of α-ketoamides over Michael acceptors and aldehydes is that they can probe the primed specificity subsites beyond S1' via easy extension of the inhibitor, although this has so far rarely been explored (e.g., Qian et al., 2008, in case of calpain).

**Table 4a.1 Structure, inhibitory activities, and toxicities of representative α-ketoamides**

| Compound No. | Formula | IC$_{50}$ (μM) against protease | EC$_{50}$ (μM) against virus | CC$_{50}$ (μM) |
|---|---|---|---|---|
| GC375<br>Kim et al., 2012 |  | 4.66 ± 0.19 (SARS-CoV)<br>1.55 ± 0.10 (TGEV) | 0.2 ± 0.1 (TGEV)<br>1.5 ± 0.3 (FIPV)<br>0.25 ± 0.1 (229E)<br>4.5 ± 0.4 MHV | nd |
| 6d<br>Mandadapu et al., 2012 |  | 2.1 (Norovirus) | 0.8 (Norovirus) | nd |
| Compound 6<br>Prior et al. 2013 |  | 0.61 ± 0.2 (SARS-CoV)<br>2.6 ± 1.5 (Norovirus) | 4.2 ± 2.4 (Norovirus) | nd |
| 8a<br>Zeng et al., 2016 |  | 8.75 ± 1.65 (EV-A71) | 6.45 ± 1.78 (EV-A71) | >100 (RD cells) |

| Compound No. | Formula | IC$_{50}$ (µM) against protease | EC$_{50}$ (µM) against virus | CC$_{50}$ (µM) |
|---|---|---|---|---|
| 8f<br>Zeng et al., 2016 |  | 1.34 ± 0.33 (EV-A71) | 1.66 ± 0.45 (EV-A71) | >100 (RD cells) |
| Compound 10<br>Kim et al., 2016 |  | 0.022 ± 0.05 (EV-D68)<br>0.02 ± 0.01 (HRV14) | 0.04 ± 0.02 (EV-D68)<br>0.08 ± 0.01 (HRV1B)<br>0.07 ± 0.01 (HRV51) | >100 (WI-38 cells) |

nd- not done

Crystal structures of HCoV NL63 (as representative of the alphacoronavirus proteases) and SARS-CoV (beta-CoV) $M^{pro}$s as well as the $3C^{pro}$ of Coxsackievirus B3 (enterovirus proteases) in complex with the early lead compound **11a** indicated fundamental different shapes of the S2 pockets in these enzymes. In the SARS-CoV $M^{pro}$, the S2 subsite is a deep hydrophobic pocket that is truly three-dimensional in shape. It consists of the "walls" formed by the polypeptide main chain around residues 186 - 188 as well as by the side-chains of His41 (of the catalytic dyad) and Gln189, as well as the "floor" formed by Met165 and the "lid" by residues 45 - 51, in particular Met49. The two methionine residues of the “lid” and “floor” are involved in important hydrophobic interactions with the P2 substituents of inhibitors. It was described by my former colleagues Zhu et al. (2011) that an aspartic-acid residue in P2 could make polar interactions with the sulfur atoms of these methionines in the SARS-CoV $M^{pro}$. As the pocket features a pronounced plasticity, P2 substituents such as isobutyl (from Leu), which are too small to fill the pocket entirely, can still generate sufficient binding enthalpy. Accordingly, the S2 pocket of SARS-CoV $M^{pro}$ is the most tolerant among the three enzymes investigated here, in terms of versatility of the P2-substituents accepted.

On the other hand, in the S2 pocket of the HCoV-NL63 $M^{pro}$, Gln189 is replaced by proline, resulting in a significant loss of flexibility and a smaller entrance for the inhibitor, because the proline is not as flexible as the side-chain of Gln189 to adjust its conformation according to the steric requirements of the P2 group. In addition, lack of Met49 reduces the possibility of hydrophobic interaction of the protease with the P2 substituent. As a consequence, the large P2-benzyl substituent cannot enter the pocket, whereas the smaller and more flexible isobutyl group of P2-Leu can.

Finally, the S2 pocket in the 3C$^{pro}$s of EV-A71 and CVB3 lacks a lid, and is open towards the top, thereby offering fewer interaction points for P2 substituent of inhibitors. Therefore, the latter must be large enough to reach the "back-wall" of the pocket (formed by Arg39, Asn69, Glu71) in order to create sufficient binding energy. Hence, large aromatic substituents such as benzyl are favored by the enterovirus 3C$^{pro}$s.

In the series of $\alpha$-ketoamides presented in this thesis, different chemical groups with respect to the length, flexibility, and size were systematically investigated at the P2 position based on the features of the S2 pockets in both corona- and enterovirus proteases. When a fluoro substituent was introduced in the *para* position of the benzyl group of the lead compound, **11a**, it showed good activity against the enterovirus 3C$^{pro}$s but complete inactivity against the coronavirus M$^{pro}$s. This is easily explained on the basis of the crystal structures: in the enterovirus 3C$^{pro}$s, the fluorine can accept a hydrogen bond from Arg39 (Lu et al., 2013), whereas in the coronavirus M$^{pro}$s, there would be electrostatic repulsion from the main-chain carbonyls of residues 186 and 188. In agreement with this, rupintrivir (which has P2 = *p*-fluorobenzyl) is a good inhibitor of the enteroviral 3C$^{pro}$s (Binford et al., 2005), but not of the coronaviral main proteases, as it was predicted by Hilgenfeld back in 2003 (Anand et al., 2003).

So far, the best-compromise compound **11u** (P2 = cyclopentylmethyl) has satisfactory broad-spectrum activity against all proteases tested, as it can fulfill different requirements of the S2 pockets (HCoV-NL63 M$^{pro}$: small and covered, SARS-CoV M$^{pro}$: large and covered, and CVB3 3C$^{pro}$: large and open). However, if the demand for equipotency against alphacoronaviruses, betacoronaviruses, and enteroviruses is relaxed, and the alphacoronaviruses are omitted from the list, the best inhibitor is **11r** (DZL08; P2 = cyclohexylmethyl), which exhibits very good inhibitory activity against both SARS-CoV M$^{pro}$ and enterovirus 3C$^{pro}$. Moreover, the performance of several α-ketoamides in the SARS-CoV, EV-A71, and CVB3 replicons was tested by Dr. Yuri Kusov, who detected micro- or submicromolar-range activities for some of them. The data for the anticorona- and enterovirus activities (obtained by our collaborators in Prof. Eric J. Snijder's, and Prof. Johan Neyts's groups) in cell culture for some α-ketoamides correlate well with the inhibitory power of the compound against recombinant proteases *in vitro* as well as in replicon-based assays (Lin, Zhang et al., to be submitted).

In addition to the systematic study of α-ketoamides, the inhibitory activities of a series of aldehydes (these compounds were designed and synthesized by the group of Prof. Hong Liu, Shanghai) against SARS-CoV M$^{pro}$ were studied. The GlnLactam (γ-lactam) was chosen as P1 subsite. In most of the aldehydes, the P2 subsite was a benzyl group except for **DC401937** with an *iso*-butyl at the P2 position, because the S2 pocket in SARS-CoV M$^{pro}$ is flexible enough to accommodate either a large group e.g. phenyl, or a small group such as *iso*-butyl. Different heterocyclic groups were designed and tested at the P3 position.

Compound **DC401937** exhibits the same inhibitory potency against SARS-CoV M$^{pro}$ as the α-ketoamide (**11n**) with identical chemical groups at the P1-P3 position. However, no obvious SAR could be detected against SARS-CoV M$^{pro}$, as 90% of the aldehydes showed micro- or submicromolar $IC_{50}$ values against SARS-CoV M$^{pro}$. Even serine or aspartic-acid residues could be properly accommodated in the hydrophobic S2 pocket, as indicated by crystal structures of SARS-CoV M$^{pro}$ in complex with two aldehydes determined by Zhu et al (2011). These results demonstrated that the highly electrophilic aldehyde warhead could overcome the stringent specificity at the P1 and P2 positions of the substrate or inhibitor. This might also explain the strong inhibitory activities of most tested aldehydes regardless the different heterocyclic groups at the P3 positions.

In two of the complex structures of SARS-CoV M$^{pro}$ with aldehydes, the hydroxyl groups of the thiohemiacetal interact with the oxyanion hole via hydrogen bonds, and the stereochemistry of the thiohemiacetal carbons are both in the *S* configuration. In addition, the fundamental difference between the structures of SARS-CoV M$^{pro}$ in complex with α-ketoamides and aldehydes is that residue Gln189 forms a hydrogen bond with the carbonyl of the P4 (**DC401954**) or the main-chain amide of the P3 (**DC401923**) substituent. In the complex structure with α-ketoamide **11a**, the S2 pocket was expanded by the large benzyl group, so that Gln189 has no direct interaction with the cinnamonyl group. However, in the aldehyde complex structures, the quinoxaline at P3 or the carbonyl of P4 offered possible interaction points for Gln189.

Moreover, several aldehydes showed submicromolar-range $EC_{50}$ values in the SARS-CoV replicon (performed by my colleague Dr. Yuri Kusov) and virus-infected cell culture (tested in Prof. Eric J. Snijder's group). Surprisingly, the two aldehydes tested are not toxic in different cell line experiments, which constitutes a deviation from the dogma that aldehydes are highly toxic.

## 4.2 Flavivirus part

Though the crystal structure of DENV2 NS2B-NS3$^{pro}$ had previously been determined, the challenge of preparing high-purity and stable recombinant protease is still a major issue due to the autolysis of the enzyme (D'Arcy et al., 2005). In this thesis, a method for preparation of highly pure and stable DENV2 NS2B-NS3$^{pro}$ was developed and the catalytic activity against the substrate (Bz-Nle-Lys-Arg-Arg-AMC) of the purified DENV2 NS2B-NS3$^{pro}$ was found to be similar as values reported by Noble et al. (2012). This method will be used for further studies: e.g. *in-vitro* activities of inhibitors against this target protease, and crystallization with compounds.

In this thesis, the structure of WNV NS2B-NS3 protease in complex with a dipeptide-boronate inhibitor (**cn-716**) was determined. This is the highest-resolution structure of a flavivirus protease with a small-molecule drug-like inhibitor, and offers a good model for us and others to optimize the binding affinity and selectivity of flaviviral protease inhibitors. Analysis of the binding details of the inhibitor with the protease provided hints for further optimization. First of all, several backbone atoms of the compound are not engaged in interaction with the protease. Therefore, these groups could be replaced to reduce the peptidic nature and increase the stability of the inhibitor. In particular, a replacement could be made for the P2 main-chain amide by a methylene group or a similar moiety to promote the drug-like character of the compound. Secondly, a hydrogen bond between the P2 main-chain carbonyl and the P1 side-chain indicates the possibility for ring closure to generate a cyclic boronic acid inhibitor. In addition, further modifications in the P2 and P3 regions could offer additional specific interactions with side-chains and possibly with backbone atoms of the target protease. It would be highly attractive to generate additional interactions between the backbone of protease and the inhibitor, because this is expected to reduce the drug-resistance potential, and produce broad-spectrum inhibitors against different flaviviruses.

Moreover, the 5-membered ester formed by the boronic-acid warhead and glycerol manifests that inhibitors can be extended towards the prime-sites of the substrate-binding cleft, which so far remain completely unexploited. The boronate prodrugs could be obtained via ester formation with larger, more hydrophobic diols or triols, and the membrane-traversing ability could also be increased compared to free boronic-acid derivatives.

When comparing the complex structure of ZIKV NS2B-NS3$^{pro}$ with the same compound (Lei et al., 2016, PDB entry 5LC0), the two structures are overall similar, in the closed conformation, and the inhibitor is similarly oriented in the active sites. However, the monomeric WNV protease structure shows no evidence for the formation of a “tight dimer”, indicating a different potential of protease assembly.

In addition, the inhibitory activities of this boronic acid compound were also tested against ZIKV NS2B-NS3$^{pro}$ in absence of glycerol, from where the boronic-acid compound shows almost the same $IC_{50}$ value as the glycerol-modified boronate variant. This implies that the modification of the warhead by hydrophobic diols or triols would not reduce the inhibitory activity, whereas the selectivity and membrane-traversing ability are expected to be improved.

Compared to WNV NS2B-NS3$^{pro}$, ZIKV NS2B-NS3$^{pro}$ exhibits super-high (~21 fold) catalytic activity towards the canonical substrate Bz-Nle-Lys-Lys-Arg-AMC, and also superior (~5.5-fold) activity against the FRET substrate Dabcyl-KTGKR↓SGAL-E(Edans)-$NH_2$. The aspartic-acid residue at position 83 in NS2B partially explained the significantly higher activity, as it is able to form an ion-pair with the positively charged residue in P2, while the Asn or Ser in WNV and

DENV are unable to do so. The D83*N mutant indeed shows a processing ability against both substrates that is reduced by 55% and about 30%, respectively.

Further, the tight dimer in the ZIKV NS2B-NS3$^{pro}$ structure (Lei et al., 2016; PDB entry 5LC0) might be relevant for the substrate binding in a positively cooperative manner. Binding of the substrate to one molecule in the homodimer might result in a lower energy barrier of the other monomer for taking up the next substrate, leading to enhancement of the subsequent binding (Freiburger et al., 2011).

Interestingly, ZIKV NS2B-NS3$^{pro}$ exhibits significant cleavage activity against the substrate Dabcyl-KPLRR↓GALLE(Edans)-$NH_2$, corresponding to the cleavage site of flavivirus proteases in the STING protein. It has been shown that DENV2 NS2B/NS3 could counteract the type-I interferon production via cleavage of STING in human monocyte-derived dendritic cells (Aguirre et al., 2012). As human DCs can also be permissively infected by ZIKV (Hamel et al., 2015), I speculate that efficient cleavage of STING by ZIKV NS2B/NS3$^{pro}$ could also occur in human DCs, thereby resulting in an enhanced suppression of the host innate immunity compared to other flaviviruses.

In this thesis, a screen of the catalytic activity of ZIKV NS2B-NS3$^{pro}$ against multiple FRET substrates corresponding to the native or putative cleavage sites in the polyprotein was performed. All the substrates carry an Arg at the the P1 position, which is identical with the result recently reported by Gruba et al. (2016) that Arg is the only preferred residue at the P1 position. Similarly, native cleavage sites in the WNV polyprotein have Arg at P1 position, whereas Lys occurs as P1 residue in several cleavage sites in the DENV polyprotein, i.e. NS2A/NS2B in DENV1, NS3/NS4A in DENV2-4, capsid in DENV3 (Table 1b.2). This is consistent with the data reported by Li et al. (2005) for the preference of

Arg/Lys at the P1 site. Gruba et al. demonstrated that Lys is more preferred than Arg at the P2 position; indeed, the substrate corresponding to the cleavage site between NS3/NS4A (Dabcyl-KAGKR↓GAAFE-(Edans)-$NH_2$) shows the highest $k_{cat}/K_m$ value according to the results shown in Table 3b.5, and Lys appears more often than Arg in the ZIKV polyprotein cleavage sites (Table 1b.2). Importantly, the cleavage activity of ZIKV against Bz-Nle-Lys-Lys-Arg-AMC is certainly higher than when P2-Lys is replaced by Arg (kinetic parameters were determined by Yasmin Gül, details will be described in her Master thesis). For DENV, the preference was observed to follow the order Arg > Thr > Gln/Asn/Lys by Li et al. (2005), while in WNV, a Lys-Lys combination is preferred at the P3-P2 sites, and the Lys-Arg pattern is also accepted (Ang et al., 2014). According to Gruba et al. (2016), ZIKV protease has a very strong requirement for Lys at the P3 position; however, a substrate with Gly at P3 shows the highest activity in this work, and P3-Gly occurs more often than Lys or Arg in the cleavage sites in the ZIKV polyprotein. DENV and WNV NS3 proteases also prefer Lys more than others at the P3 position (Li et al., 2005; Ang et al., 2014). Certainly, other amino acids, such as: Arg, Asn, or Thr were accordingly acceptable at P3 position in the three enzymes (Ang et al., 2014; Gruba et al., 2016; Li et al., 2005; Table 1b.2). The most intense signal is found for Val at the P4 postion by Gruba et al. (2016). A wide range of amino acids (Lys, Ser, Thr, Ala, Val) is used at the P4 position in native cleavage sites in the ZIKV polyprotein, which means that there is no strong preference at the P4 position in the cleavage site in the polyprotein. Li et al., (2005) found that the preference of DENV protease was in the order of Nle > Leu > Lys > Xaa. Certainly, the super- high catalytic activity against the substrate Bz-Nle-Lys-Lys-Arg-AMC proves that ZIKV NS2B-$NS3^{pro}$ can also perfectly recognize Nle at the P4 site. In addition, among the cleavage sites in the polyprotein of WNV and DENV 1-4, various amino acids were identified at P4, suggesting that the low specificity at the P4 position would not have played a critical role for substrate binding.

The prime site of the ZIKV NS2B/NS3$^{pro}$ was not systematically studied yet. However, the preference of the P1' site can be easily deduced from the cleavage sites in the polyprotein (Table 1b.2), as small residues such as Gly, Ser are intensively preferred. This also applies to the WNV and DENV cleavage sites. For the P2' position, most cases have Gly or Ala, but large amino acids (Trp, cleavage sites between NS2A/NS2B in ZIKV, WNV, DENV1-4; Gln in cleavage sites between NS3/NS4A, and NS4A/2K in WNV; Table 1b.2) occasionally occur at this site. Moreover, the FRET substrate with a Trp at the P2' site, designed based on the cleavage site between NS2A/NS2B in ZIKV, exhibits similar $k_{cat}/K_m$ values as those with small residues (capsid, NS2B/NS3). In the ZIKV polyprotein cleavage sites, most of the amino acids at the P3' position are Gly or Ala, except for an Asp occupying the position in Capsid. On the other hand, the preference at the P3' position is not restrained in both WNV and DENV according to the cleavage sites in polyproteins (Table 1b.2). There is no apparent specificity at the P4' postion in ZIKV, WNV, and DENV, which means that the contribution of amino acids beyond P4' might be negligible. Furthermore, Li et al. (2005) described that the prime-site substrate specificity was for small and polar amino acids in P1' and P3', whereas the P2' and P4' substrate positions showed minimal activity when compared with P1' and P3' sites. Niyomrattanakit et al (2006) also demonstrated that small unbranched amino acid (Ser, Ala or Gly) residues are preferred at the P1' position.

The kinetic parameters of ZIKV NS2B-NS3$^{pro}$ against the multiple substrates derived from native and putative cleavage sites in the polyprotein, presented in this thesis constitute a big deviation to results determined for WNV and DENV2 NS2B-NS3$^{pro}$ by Shiryaev et al. (2007). In their results, the cleavage activity was extraordinarily low against several substrates designed from the native cleavage sites (NS3/NA4A in WNV and DENV2, NS4B/NS5 in DENV2), which might due to the distinct analytic method used by these authors. However, the respective results in this thesis agree better with those reported by Khumthong et al. (2003) than Shiryaev et al. (2007).

Taken together, a basis for further enzymatic tests, and substrate based drug design against ZIKV was achieved by this part of work.

## References

Aguirre S., Maestre A. M., Pagni S., Patel J. R., Savage T., Gutman D., Maringer K., Bernal-Rubio D., Shabman R. S., Simon V., Rodriguez-Madoz J. R., Mulder L. C., Barber G. N., Fernandez-Sesma A., 2012. DENV inhibits type I IFN production in infected cells by cleaving human STING. PLoS Pathog. 8, e1002934.

Akaji K., Konno H., Mitsui H., Teruya K., Shimamoto Y., Hattori Y., Ozaki T., Kusunoki M., Sanjoh A., 2011. Structure-based design, synthesis, and evaluation of peptide-mimetic SARS 3CL protease inhibitors. J. Med. Chem. 54, 7962-7973.

Al-Gharabli S. I., Shah S. T., Weik S., Schmidt M. F., Mesters J. R., Kuhn D., Klebe G., Hilgenfeld R., Rademann J., 2006. An efficient method for the synthesis of peptide aldehyde libraries employed in the discovery of reversible SARS coronavirus main protease (SARS-CoV Mpro) inhibitors. Chembiochem 7, 1048-1055.

Aleshin A. E., Shiryaev S. A., Strongin A. Y., Liddington R. C., 2007. Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. Protein Sci. 16, 795-806.

Allison S. L., Schalich J., Stiasny K., Mandl C. W., Kunz C., Heinz F. X., 1995. Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. J. Virol. 69, 695-700.

Anand K., Palm G. J., Mesters J. R., Siddell S. G., Ziebuhr J., Hilgenfeld R., 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J. 21, 3213-3224.

Anand K., Ziebuhr J., Wadhwani P., Mesters J. R., Hilgenfeld R., 2003. Coronavirus main proteinase ($3CL^{pro}$) structure: basis for design of anti-SARS drugs. Science 300, 1763-1767.

Ang M. J., Li Z., Lim H. A., Ng F. M., Then S. W., Wee J. L., Joy J., Hill J., Chia C. S., 2014. A P2 and P3 substrate specificity comparison between the Murray Valley encephalitis and West Nile virus NS2B/NS3 protease using C-terminal agmatine dipeptides. Peptides 52, 49-52.

Arndt U. W., Crowther R. A., Mallett J. F., 1968. A computer-linked cathode-ray tube microdensitometer for x-ray crystallography. J. Sci. Instrum. 1, 510-516.

Assenberg R., Mastrangelo E., Walter T. S., Verma A., Milani M., Owens R. J., Stuart D. I., Grimes J. M., Mancini E. J., 2009. Crystal structure of a novel conformational state of the flavivirus NS3 protein: implications for polyprotein processing and viral replication. J. Virol. 83, 12895-12906.

Aubry M., Finke J., Teissier A., Roche C., Broult J., Paulous S., Despres P., Cao-Lormeau V. M., Musso D., 2015. Seroprevalence of arboviruses among blood donors in French Polynesia, 2011-2013. Int. J. Infect. Dis. 41, 11-12.

Bacha U., Barrila J., Gabelli S. B., Kiso Y., Mario Amzel L., Freire E., 2008. Development of broad-spectrum halomethyl ketone inhibitors against coronavirus main protease $3CL^{pro}$. Chem. Biol. Drug. Des. 72, 34-49.

Barabas E., Szell E., Bajusz S., 1993. Screening for fibrinolysis inhibitory effect of synthetic thrombin inhibitors. Blood Coagul Fibrinolysis 4, 243-248.

Barretto N., Jukneliene D., Ratia K., Chen Z., Mesecar A. D., Baker S. C., 2005. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. J. Virol. 79, 15189-15198.

Bearcroft W. G., 1956. Zika virus infection experimentally induced in a human volunteer. Trans. R. Soc. Trop. Med. Hyg. 50, 442-448.

Beaudette F. R., Hudson C. B., 1937. Cultivation of the virus of infectious bronchitis. J. Am. Vet. Med. Assoc. 90, 51-58.

Bell T. M., Field E. J., Narang H. K., 1971. Zika virus infection of the central nervous system of mice. Arch. Gesamte Virusforsch. 35, 183-193.

Bhatt S., Gething P. W., Brady O. J., Messina J. P., Farlow A. W., Moyes C. L., Drake J. M., Brownstein J. S., Hoen A. G., Sankoh O., Myers M. F., George D. B., Jaenisch T., Wint G. R., Simmons C. P., Scott T. W., Farrar J. J., Hay S. I., 2013. The global distribution and burden of dengue. Nature 496, 504-507.

Binford S. L., Maldonado F., Brothers M. A., Weady P. T., Zalman L. S., Meador J. W., 3rd, Matthews D. A., Patick A. K., 2005. Conservation of amino acids in human rhinovirus 3C protease correlates with broad-spectrum antiviral activity of rupintrivir, a novel human rhinovirus 3C protease inhibitor. Antimicrob. Agents Chemother. 49, 619-626.

Blitvich B. J., Firth A. E., 2015. Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. Viruses 7, 1927-1959.

Brasil P., Sequeira P. C., Freitas A. D., Zogbi H. E., Calvet G. A., de Souza R. V., Siqueira A. M., de Mendonca M. C., Nogueira R. M., de Filippis A. M., Solomon T., 2016. Guillain-Barré syndrome associated with Zika virus infection. Lancet 387, 1482.

Bressanelli S., Stiasny K., Allison S. L., Stura E. A., Duquerroy S., Lescar J., Heinz F. X., Rey F. A., 2004. Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. EMBO J. 23, 728-738.

Brian D. A., Baric R. S., 2005. Coronavirus genome structure and replication. Curr. Top Microbiol. Immunol. 287, 1-30.

Brinton M. A., 1986. Replication of Flaviviruses. The Togaviridae and Flaviviridae, Plenum Press, New York. chapter 11, 327-374.

Brown W. C., Akey D. L., Konwerski J. R., Tarrasch J. T., Skiniotis G., Kuhn R. J., Smith J. L., 2016. Extended surface for membrane association in Zika virus NS1 structure. Nat. Struct. Mol. Biol. 23, 865-867.

Calvet G., Aguiar R. S., Melo A. S., Sampaio S. A., de Filippis I., Fabri A., Araujo E. S., de Sequeira P. C., de Mendonca M. C., de Oliveira L., Tschoeke D. A., Schrago C. G., Thompson F. L., Brasil P., Dos Santos F. B., Nogueira R. M., Tanuri A., de Filippis A. M., 2016. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. Lancet Infect. Dis. 16, 653-660.

Cao-Lormeau V. M., Blake A., Mons S., Lastere S., Roche C., Vanhomwegen J., Dub T., Baudouin L., Teissier A., Larre P., Vial A. L., Decam C., Choumet V., Halstead S. K., Willison H. J., Musset L., Manuguerra J. C., Despres P., Fournier E., Mallet H. P., Musso D., Fontanet A., Neil J., Ghawche F., 2016. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study. Lancet 387, 1531-1539.

Chambers T. J., Weir R. C., Grakoui A., McCourt D. W., Bazan J. F., Fletterick R. J., Rice C. M., 1990a. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. Proc. Natl. Acad. Sci. USA 87, 8898-8902.

Chambers T. J., Hahn C. S., Galler R., Rice C. M., 1990b. Flavivirus genome organization, expression, and replication. Annu. Rev.

Microbiol. 44, 649-688.

Chambers T. J., Grakoui A., Rice C. M., 1991. Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. J. Virol. 65, 6042-6050.

Chandramouli S., Joseph J. S., Daudenarde S., Gatchalian J., Cornillez-Ty C., Kuhn P., 2010. Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family. J. Virol. 84, 3059-3067.

Chappell K. J., Stoermer M. J., Fairlie D. P., Young P. R., 2006. Insights to substrate binding and processing by West Nile Virus NS3 protease through combined modeling, protease mutagenesis, and kinetic studies. J. Biol. Chem. 281, 38448-38458.

Cheever, F. S., Daniels J. B., Pappenheimer A. M., Bailey O. T., 1949. A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin. J. Exp. Med. 90, 181-210.

Chevenet F., Brun C., Bañuls A., Jacq B., Christen R., 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics. 7, 439-447.

Chu D. K., Leung C. Y., Gilbert M., Joyner P. H., Ng E. M., Tse T. M., Guan Y., Peiris J. S., Poon L. L., 2011. Avian coronavirus in wild aquatic birds. J. Virol. 85, 12815-12820.

Chuck C. P., Chen C., Ke Z., Wan D. C., Chow H. F., Wong K. B., 2013. Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases. Eur. J. Med. Chem. 59, 1-6.

Cleaves G. R., Dubin D. T., 1979. Methylation status of intracellular dengue type 240S RNA. Virology 96, 159-165.

Coloma J., Jain R., Rajashankar K. R., Garcia-Sastre A., Aggarwal A. K.,

2016. Structures of NS5 methyltransferase from Zika virus. Cell Rep. 16, 3097-3102.

D'Arcy A., Chaillet M., Schiering N., Villard F., Lim S. P., Lefeuvre P., Erbel P., 2006. Purification and crystallization of dengue and West Nile virus NS2B-NS3$^{pro}$ complexes. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 62, 157-162.

D'Ortenzio E., Matheron S., Yazdanpanah Y., de Lamballerie X., Hubert B., Piorkowski G., Maquart M., Descamps D., Damond F., Leparc-Goffart I., 2016. Evidence of xexual transmission of Zika virus. N. Engl. J. Med. 374, 2195-2198.

Dai L., Song J., Lu X., Deng Y. Q., Musyoki A. M., Cheng H., Zhang Y., Yuan Y., Song H., Haywood J., Xiao H., Yan J., Shi Y., Qin C. F., Qi J., Gao G. F., 2016. Structures of the Zika virus envelope protein and tts complex with a flavivirus broadly protective antibody. Cell Host Microbe. 19, 696-704.

De Clercq E., 2009. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. Int. J. Antimicrob. Agents. 33, 307-320.

de Haan C. A., Smeets M., Vernooij F., Vennema H., Rottier P. J., 1999. Mapping of the coronavirus membrane protein domains involved in interaction with the spike protein. J. Virol. 73, 7441-7452.

Dea S., Tijssen P., 1988. Identification of the structural proteins of turkey enteric coronavirus. Arch. Virol. 99, 173-186.

Dong B., Liu W., Fan X., Vijaykrishna D., Tang X., Gao F., Li L., Li G., Zhang J., Yang L., Poon L., Zhang S., Peiris J. S., Smith G. J., Chen H., Guan Y., 2007. Detection of a novel and highly divergent coronavirus from asian leopard cats and Chinese ferret badgers in Southern China. J. Virol. 81, 6920-6926.

Doyle L. P., Hutchings L. M., 1946. A transmissible gastroenteritis in pigs.

J. Am. Vet. Med. Assoc. 108, 257-259.

Dragovich P. S., Zhou R., Skalitzky D. J., Fuhrman S. A., Patick A. K., Ford C. E., Meador J. W., 3rd, Worland S. T., 1999. Solid-phase synthesis of irreversible human rhinovirus 3C protease inhibitors. Part 1: Optimization of tripeptides incorporating N-terminal amides. Bioorg. Med. Chem. 7, 589-598.

Drosten C., Gunther S., Preiser W., van der Werf S., Brodt H. R., Becker S., Rabenau H., Panning M., Kolesnikova L., Fouchier R. A., Berger A., Burguiere A. M., Cinatl J., Eickmann M., Escriou N., Grywna K., Kramme S., Manuguerra J. C., Muller S., Rickerts V., Sturmer M., Vieth S., Klenk H. D., Osterhaus A. D., Schmitz H., Doerr H. W., 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. 348, 1967-1976.

Eckerle I., Muller M. A., Kallies S., Gotthardt D. N., Drosten C., 2013. In-vitro renal epithelial cell infection reveals a viral kidney tropism as a potential mechanism for acute renal failure during Middle East Respiratory Syndrome (MERS) Coronavirus infection. Virol. J. 10, 359-363.

Emsley P., Lohkamp B., Scott W. G., Cowtan K., 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486-501.

Erbel P., Schiering N., D'Arcy A., Renatus M., Kroemer M., Lim S. P., Yin Z., Keller T. H., Vasudevan S. G., Hommel U., 2006. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. Nat. Struct. Mol. Biol. 13, 372-373.

Evans P. R., 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D Biol. Crystallogr. 67, 282-292.

Falgout B., Pethel M., Zhang Y. M., Lai C. J., 1991. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J. Virol. 65, 2467-2475.

Falgout B., Miller R. H., Lai C. J., 1993. Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. J. Virol. 67, 2034-2042.

Fauquet C. M., Mayo, M. A., Maniloff, J., Desselberger, U., Ball, L.A., (ed.), 2005. Virus taxonomy, eighth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Amsterdam, The Netherlands.

Fehrentz J. A., Heitz A., Castro B., Cazaubon C., Nisato D., 1984. Aldehydic peptides inhibiting renin. FEBS Lett. 167, 273-276.

Ferreira M. L. B., 2016. Neurologic manifestations of arboviruses in the epidemic in pernambuco, Brazil. American Academy of Neurological, 68th Annual Meeting Abstract.

Fleming F. F., Yao L., Ravikumar P. C., Funk L., Shook B. C., 2010. Nitrile-containing pharmaceuticals: efficacious roles of the nitrile pharmacophore. J. Med. Chem. 53, 7902-7917.

Fouchier R. A., Hartwig N. G., Bestebroer T. M., Niemeyer B., de Jong J. C., Simon J. H., Osterhaus A. D., 2004. A previously undescribed coronavirus associated with respiratory disease in humans. Proc. Natl. Acad. Sci. USA 101, 6212-6216.

Freiburger L. A., Baettig O. M., Sprules T., Berghuis A. M., Auclair K., Mittermaier A. K. 2011. Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. Nat. Struct. Mol. Biol. 18, 288-294.

Fuhrmann C. N., 2005. Structural studies of alpha-lytic protease at sub-Angstrom resolution reveal insights into the mechanism of serine protease catalysis and the kinetic stability. University of California, San Francisco.

Ge X., Li J., Yang X., Chmura A., Zhu G., Epstein J. H., Mazet J. K., Hu B., Zhang W., Peng C., Zhang Y., Luo C., Tan B., Wang N., Zhu Y., Crameri G., Zhang S., Wang L., Daszak P., Shi Z., 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503, 535-538.

Ghosh A. K., Xi K., Ratia K., Santarsiero B. D., Fu W., Harcourt B. H., Rota P. A., Baker S. C., Johnson M. E., Mesecar A. D., 2005. Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. J. Med. Chem. 48, 6767-6771.

Ghosh A. K., Xi K., Grum-Tokars V., Xu X., Ratia K., Fu W., Houser K. V., Baker S. C., Johnson M. E., Mesecar A. D., 2007. Structure-based design, synthesis, and biological evaluation of peptidomimetic SARS-CoV 3CL$^{pro}$ inhibitors. Bioorg. Med. Chem. Lett. 17, 5876-5880.

Goetz D. H., Choe Y., Hansell E., Chen Y. T., McDowell M., Jonsson C. B., Roush W. R., McKerrow J., Craik C. S., 2007. Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS coronavirus. Biochemistry 46, 8744-8752.

Gouet P., Courcelle E., Stuart D. I., Métoz F., 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. 15, 305-308.

Graybill T. L., Dolle R. E., Helaszek C. T., Miller R. E., Ator M. A., 1994. Preparation and evaluation of peptidic aspartyl hemiacetals as reversible inhibitors of interleukin-1 beta converting enzyme (ICE). Int. J. Pept. Protein Res. 44, 173-182.

Greninger A. L., Naccache S. N., Messacar K., Clayton A., Yu G., Somasekar S., Federman S., Stryke D., Anderson C., Yagi S., Messenger S., Wadford D., Xia D., Watt J. P., Van Haren K., Dominguez S. R., Glaser C., Aldrovandi G., Chiu C. Y., 2015. A novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the USA (2012-14): a retrospective cohort study. Lancet Infect. Dis. 15, 671-682.

Gruba N., Rodriguez Martinez J. I., Grzywa R., Wysocka M., Skorenski M., Burmistrz M., Lecka M., Lesner A., Sienczyk M., Pyrc K., 2016. Substrate profiling of Zika virus NS2B-NS3 protease. FEBS Lett. 590, 3459-3468.

Hamel R., Dejarnac O., Wichit S., Ekchariyawat P., Neyret A., Luplertlop N., Perera-Lecoin M., Surasombatpattana P., Talignani L., Thomas F., Cao-Lormeau V. M., Choumet V., Briant L., Despres P., Amara A., Yssel H., Misse D., 2015. Biology of Zika virus infection in human skin cells. J. Virol. 89, 8880-8896.

Hammamy M. Z., Haase C., Hammami M., Hilgenfeld R., Steinmetzer T., 2013. Development and characterization of new peptidomimetic inhibitors of the West Nile virus NS2B-NS3 protease. ChemMedChem 8, 231-241.

Hamre D., Procknow J. J., 1966. A new virus isolated from human respiratory tract. Proc. Soc. Exp. Biol. Med. 121, 190-193.

Hayes E. B., 2009. Zika virus outside Africa. Emerg. Infect. Dis. 15, 1347-1350.

Heukelbach J., Alencar C. H., Kelvin A. A., de Oliveira W. K., Pamplona de Goes Cavalcanti L., 2016. Zika virus outbreak in Brazil. J. Infect. Dev. Ctries. 10, 116-120.

Hilgenfeld R., 2014. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. FEBS J. 281, 4085-4096.

Hilgenfeld R., Peiris M., 2013. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. Antiviral Res. 100, 286-295.

Ho B., Cheng S., Shi L., Wang T., Ho K., Chou C., 2015. Critical assessment of the important residues involved in the dimerization and catalysis of MERS coronavirus main protease. PloS One 10, e0144865.

Huelsenbeck J. P., Ronquist F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Jain R., Coloma J., Garcia-Sastre A., Aggarwal A. K., 2016. Structure of the NS3 helicase from Zika virus. Nat. Struct. Mol. Biol. 23, 752-754.

Jones C. T., Ma L., Burgner J. W., Groesch T. D., Post C. B., Kuhn R. J., 2003. Flavivirus capsid is a dimeric alpha-helical protein. J. Virol. 77, 7143-7149.

Karplus P. A., Diederichs K., 2012. Linking crystallographic model and data quality. Science 336, 1030-1033.

Khumthong R., Niyomrattanakit P., Chanprapaph S., Angsuthanasombat C., Panyim S., Katzenmeier G., 2003. Steady-state cleavage kinetics for dengue virus type 2 NS2B-NS3$^{pro}$ serine protease with synthetic peptides. Protein Pept. Lett. 10, 19-26.

Kim Y., Lovell S., Tiew K. C., Mandadapu S. R., Alliston K. R., Battaile K. P., Groutas W. C., Chang K. O., 2012. Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. J. Virol. 86, 11754-11762.

Kim Y. M., Gayen S., Kang C., Joy J., Huang Q., Chen A. S., Wee J. L., Ang M. J., Lim H. A., Hung A. W., Li R., Noble C. G., Lee le T., Yip A., Wang Q. Y., Chia C. S., Hill J., Shi P. Y., Keller T. H., 2013. NMR analysis of a novel enzymatically active unlinked dengue

NS2B-NS3 protease complex. J. Biol. Chem. 288, 12891-12900.

Kim Y., Kankanamalage A. C., Damalanka V. C., Weerawarna P. M., Groutas W. C., Chang K. O., 2016. Potent inhibition of enterovirus D68 and human rhinoviruses by dipeptidyl aldehydes and alpha-ketoamides. Antiviral Res. 125, 84-91.

King A.M.Q., Adams M.J., Carstens E.B., Lefkowitz E.J., (ed.), 2012. Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Amsterdam, The Netherlands.

Konishi E., Mason P. W., 1993. Proper maturation of the Japanese encephalitis virus envelope glycoprotein requires cosynthesis with the premembrane protein. J. Virol. 67, 1672-1675.

Krug M., Weiss M. S., Heinemann U., Mueller U., 2012. XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS. Journal of Applied Crystallography 45, 568-572.

Ksiazek T. G., Erdman D., Goldsmith C. S., Zaki S. R., Peret T., Emery S., Tong S., Urbani C., Comer J. A., Lim W., Rollin P. E., Dowell S. F., Ling A., Humphrey C. D., Shieh W., Guarner J., Paddock C. D., Rota P., Fields B., DeRisi J., Yang J., Cox N., Hughes J. M., LeDuc J. W., Bellini W. J., Anderson L. J., SARS Working Group, 2003. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. 348, 1953-1966.

Kuiken T., Fouchier R. A., Schutten M., Rimmelzwaan G. F., van Amerongen G., van Riel D., Laman J. D., de Jong T., van Doornum G., Lim W., Ling A. E., Chan P. K., Tam J. S., Zambon M. C., Gopal R., Drosten C., van der Werf S., Escriou N., Manuguerra J. C., Stohr K., Peiris J. S., Osterhaus A. D., 2003. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. Lancet 362, 263-270.

Kuo L., Masters P. S., 2002. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. J. Virol. 76, 4987-4999.

Lakowicz J. R., 1983. Principles of fluorescence spectroscopy. Plenum Press, New York.

Lauber C., Ziebuhr J., Junglen S., Drosten C., Zirkel F., Nga P. T., Morita K., Snijder E. J., Gorbalenya A. E., 2012. Mesoniviridae: a proposed new family in the order Nidovirales formed by a single species of mosquito-borne viruses. Arch. Virol. 157, 1623-1628.

Lazear H. M., Govero J., Smith A. M., Platt D. J., Fernandez E., Miner J. J., Diamond M. S., 2016. A mouse model of Zika virus pathogenesis. cell host microbe. 19, 720-730.

Lebedev A. A., Young P., Isupov M. N., Moroz O. V., Vagin A. A., Murshudov G. N., 2012. JLigand: a graphical tool for the CCP4 template-restraint library. Acta Crystallogr. D Biol. Crystallogr. 68, 431-440.

Lee C. C., Kuo C. J., Ko T. P., Hsu M. F., Tsui Y. C., Chang S. C., Yang S., Chen S. J., Chen H. C., Hsu M. C., Shih S. R., Liang P. H., Wang A. H., 2009. Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds. J. Biol. Chem. 284, 7646-7655.

Lee T. W., Cherney M. M., Huitema C., Liu J., James K. E., Powers J. C., Eltis L. D., James M. N., 2005. Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. J. Mol. Biol. 353, 1137-1151.

Lee T. W., Cherney M. M., Liu J., James K. E., Powers J. C., Eltis L. D., James M. N. G., 2007. Crystal structures reveal an induced-fit binding of a substrate-like aza-peptide epoxide to SARS coronavirus main peptidase. J. Mol. Biol. 366, 916-932.

Lei J., Mesters J. R., Drosten C., Anemuller S., Ma Q., Hilgenfeld R., 2014. Crystal structure of the papain-like protease of MERS coronavirus reveals unusual, potentially druggable active-site features. Antiviral Res. 109, 72-82.

Lei J., Hansen G., Nitsche C., Klein C. D., Zhang L., Hilgenfeld R., 2016. Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor. Science 353, 503-505.

Leung D., Schroder K., White H., Fang N. X., Stoermer M. J., Abbenante G., Martin J. L., Young P. R., Fairlie D. P., 2001. Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors. J. Biol. Chem. 276, 45762-45771.

Li F., 2016. Structure, function, and evolution of coronavirus spike proteins. Annu. Rev. Virol. 3, 237-261.

Li J., Lim S. P., Beer D., Patel V., Wen D., Tumanut C., Tully D. C., Williams J. A., Jiricek J., Priestle J. P., Harris J. L., Vasudevan S. G., 2005. Functional profiling of recombinant NS3 proteases from all four serotypes of dengue virus using tetrapeptide and octapeptide substrate libraries. J. Biol. Chem. 280, 28766-28774.

Lim S. P., Wang Q. Y., Noble C. G., Chen Y. L., Dong H., Zou B., Yokokawa F., Nilar S., Smith P., Beer D., Lescar J., Shi P. Y., 2013. Ten years of dengue drug discovery: progress and prospects. Antiviral Res. 100, 500-519.

Lin K., Perni R. B., Kwong A. D., Lin C., 2006. VX-950, a novel hepatitis C virus (HCV) NS3-4A protease inhibitor, exhibits potent antiviral activities in HCV replicon cells. Antimicrob. Agents Chemother 50, 1813-1822.

Lindenbach B. D., Thiel H. J., Rice C. M., 2007 *Flaviviridae*: the viruses and their replication. Fields Virology fifth edition, Lippincott

Williams & Wilkins, chapter 33, 1101-1152.

Liu D., Fung T., Chong K., Shukla A., Hilgenfeld R., 2014. Accessory proteins of SARS-CoV and other coronaviruses. Antiviral Res. 109, 97-109.

Liu Y. Y., Kati W., Chen C. M., Tripathi R., Molla A., Kohlbrenner W., 1999. Use of a fluorescence plate reader for measuring kinetic parameters with inner filter effect correction. Analytical Biochemistry 267, 331-335.

Lobigs M., 1993. Flavivirus premembrane protein cleavage and spike heterodimer secretion require the function of the viral proteinase NS3. Proc. Natl. Acad. Sci. USA 90, 6218-6222.

Lobigs M., Lee E., 2004. Inefficient signalase cleavage promotes efficient nucleocapsid incorporation into budding flavivirus membranes. J. Virol. 78, 178-186.

Lorenz I. C., Allison S. L., Heinz F. X., Helenius A., 2002. Folding and dimerization of tick-borne encephalitis virus envelope proteins prM and E in the endoplasmic reticulum. J. Virol. 76, 5480-5491.

Lu G., Qi J., Chen Z., Xu X., Gao F., Lin D., Qian W., Liu H., Jiang H., Yan J., Gao G. F., 2011. Enterovirus 71 and coxsackievirus A16 3C proteases: binding to rupintrivir and their substrates and anti-hand, foot, and mouth disease virus drug design. J. Virol. 85, 10319-10331.

Luo D., Xu T., Hunke C., Gruber G., Vasudevan S. G., Lescar J., 2008. Crystal structure of the NS3 protease-helicase from dengue virus. J. Virol. 82, 173-183.

Luo D., Wei N., Doan D. N., Paradkar P. N., Chong Y., Davidson A. D., Kotaka M., Lescar J., Vasudevan S. G., 2010. Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications. J. Biol.

Chem. 285, 18817-18827.

Luo D., Vasudevan S. G., Lescar J., 2015. The flavivirus NS2B-NS3 protease-helicase as a target for antiviral drug development. Antiviral Res. 118, 148-158.

Ma L., Jones C. T., Groesch T. D., Kuhn R. J., Post C. B., 2004. Solution structure of dengue virus capsid protein reveals another fold. Proc. Natl. Acad. Sci. USA 101, 3414-3419.

Mackenzie J. S., Gubler D. J., Petersen L. R., 2004. Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. Nat. Med. 10, S98-109.

Mandadapu S. R., Weerawarna P. M., Gunnam M. R., Alliston K. R., Lushington G. H., Kim Y., Chang K. O., Groutas W. C., 2012. Potent inhibition of norovirus 3CL protease by peptidyl alpha-ketoamides and alpha-ketoheterocycles. Bioorg. Med. Chem. Lett. 22, 4820-4826.

Markoff L., 2003. 5'- and 3'-noncoding regions in flavivirus RNA. Adv. Virus Res. 59, 177-228.

Massilamany C, Gangaplara A, Reddy J., 2014. Intricacies of cardiac damage in coxsackievirus B3 infection: implications for therapy. Int. J. Cardiol. 177, 330 - 339.

Masters P. S., Perlman S., 2013 Coronaviridae. Fields virology sixth edition, Wolters Kluwer, Lippincott Williams & Wilkins (LWW), chapter 28, 825-857.

McBride R., van Zyl M., Fielding B. C., 2014. The coronavirus nucleocapsid is a multifunctional protein. Viruses 6, 2991-3018.

McConnell R. M., York J. L., Frizzell D., Ezell C., 1993. Inhibition studies of some serine and thiol proteinases by new leupeptin analogues. J. Med. Chem. 36, 1084-1089.

Mlakar J., Korva M., Tul N., Popovic M., Poljsak-Prijatelj M., Mraz J., Kolenc M., Resman Rus K., Vesnaver Vipotnik T., Fabjan Vodusek V., Vizjak A., Pizem J., Petrovec M., Avsic Zupanc T., 2016. Zika virus associated with microcephaly. N. Engl. J. Med. 374, 951-958.

Modis Y., Ogata S., Clements D., Harrison S. C., 2004. Structure of the dengue virus envelope protein after membrane fusion. Nature 427, 313-319.

Morris A. L., MacArthur M. W., Hutchinson E. G., Thornton J. M., 1992. Stereochemical quality of protein structure coordinates. Proteins 12, 345-364.

Mueller N. H., Yon C., Ganesh V. K., Padmanabhan R., 2007. Characterization of the West Nile virus protease substrate specificity and inhibitors. Int. J. Biochem. Cell Biol. 39, 606-614.

Mukhopadhyay S., Kuhn R. J., Rossmann M. G., 2005. A structural perspective of the flavivirus life cycle. Nat. Rev. Microbiol. 3, 13-22.

Murshudov G. N., Vagin A. A., Dodson E. J., 1997. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240-255.

Murshudov G. N., Skubak P., Lebedev A. A., Pannu N. S., Steiner R. A., Nicholls R. A., Winn M. D., Long F., Vagin A. A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. 67, 355-367.

Musso D., Roche C., Robin E., Nhan T., Teissier A., Cao-Lormeau V. M., 2015. Potential sexual transmission of Zika virus. Emerg. Infect. Dis. 21, 359-361.

Needle D., Lountos G. T., Waugh D. S., 2015. Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity. Acta Crystallogr. D Biol. Crystallogr. 71, 1102-1111.

Niyomrattanakit P., Yahorava S., Mutule I., Mutulis F., Petrovska R., Prusis P., Katzenmeier G., Wikberg J. E., 2006. Probing the substrate specificity of the dengue virus type 2 NS3 serine protease by using internally quenched fluorescent peptides. Biochem. J. 397, 203-211.

Noble C. G., Seh C. C., Chao A. T., Shi P. Y., 2012. Ligand-bound structures of the dengue virus protease reveal the active conformation. J. Virol. 86, 438-446.

Noble C. G., Shi P. Y., 2012. Structural biology of dengue virus enzymes: towards rational design of therapeutics. Antiviral Res. 96, 115-126.

Norder H., De Palma A. M., Selisko B., Costenaro L., Papageorgiou N., Arnan C., Coutard B., Lantez V., De Lamballerie X., Baronti C., Sola M., Tan J., Neyts J., Canard B., Coll M., Gorbalenya A. E., Hilgenfeld R., 2011. Picornavirus non-structural proteins as targets for new anti-virals with broad activity. Antiviral Res. 89, 204-218.

Nowak T., Wengler G., 1987. Analysis of disulfides present in the membrane proteins of the West Nile flavivirus. Viology 156, 127-137.

Oermann C. M., Schuster J. E., Conners G. P., Newland J. G., Selvarangan R., Jackson M. A., 2015. Enterovirus D68. A focused review and clinical highlights from the 2014 U.S. Outbreak. Ann. Am. Thorac. Soc. 12, 775-781.

Peiris J. S., Lai S., Poon L., Guan Y., Yam L., Lim W., Nicholls J., Yee W., Yan W., Cheung M., Cheng V., Chan K., Tsang D., Yung R., Ng T. K., Yuen K. Y., 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 361, 1319-1325.

Perlman S., 1998. Pathogenesis of coronavirus-induced infections. Review of pathological and immunological aspects. Adv. Exp. Med. Biol. 440, 503-513.

Petersen L. R., Marfin A. A., 2002. West Nile virus: a primer for the clinician. Ann. Intern. Med. 137, 173-179.

Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C., Ferrin T. E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605-1612.

Plant E. P., Rakauskaite R., Taylor D. R., Dinman J. D., 2010. Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. J. Virol. 84, 4330-4340.

Prior A. M., Kim Y., Weerasekara S., Moroze M., Alliston K. R., Uy R. A., Groutas W. C., Chang K. O., Hua D. H., 2013. Design, synthesis, and bioevaluation of viral 3C and 3C-like protease inhibitors. Bioorg. Med. Chem. Lett. 23, 6317-6320.

Pugachev K. V., Guirakhoo F., Monath T. P., 2005. New developments in flavivirus vaccines with special attention to yellow fever. Curr. Opin. Infect. Dis. 18, 387-394.

Pulendran B., 2009. Learning immunology from the yellow fever vaccine: innate immunity to systems vaccinology. Nat. Rev. Immunol. 9, 741-747.

Pyrc K., Berkhout B., van der Hoek L., 2007. The novel human coronaviruses NL63 and HKU1. J. Virol. 81, 3051-3057.

Qian J., Cuerrier D., Davies P. L., Li Z., Powers J. C., Campbell R. L., 2008. Cocrystal structures of primed side-extending alpha-ketoamide inhibitors reveal novel calpain-inhibitor aromatic interactions. J. Med. Chem. 51, 5264-5270.

Rasmussen S. A., Jamieson D. J., Honein M. A., Petersen L. R., 2016. Zika virus and birth defects--reviewing the evidence for causality. N. Engl. J. Med. 374, 1981-1987.

Rey F. A., Heinz F. X., Mandl C., Kunz C., Harrison S. C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 A resolution. Nature 375, 291-298.

Rice C. M., Lenches E. M., Eddy S. R., Shin S. J., Sheets R. L., Strauss J. H., 1985. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science 229, 726-733.

Rice, C. M., Strauss, E. G., Strauss, J. H., 1986. Structure of the flavivirus genome. The Togaviridae and Flaviviridae, Plenum, New York, chapter 10, 279-326.

Robin G., Chappell K., Stoermer M. J., Hu S. H., Young P. R., Fairlie D. P., Martin J. L., 2009. Structure of West Nile virus NS3 protease: ligand stabilization of the catalytic conformation. J. Mol. Biol. 385, 1568-1577.

Romano K. P., Ali A., Aydin C., Soumana D., Ozen A., Deveau L. M., Silver C., Cao H., Newton A., Petropoulos C. J., Huang W., Schiffer C. A., 2012. The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors. PLoS Pathog. 8, e1002832.

Ruch T. R., Machamer C. E., 2012. The coronavirus E protein: assembly and beyond. Viruses 4, 363-382.

Ruiz-Linares A., Cahour A., Despres P., Girard M., Bouloy M., 1989. Processing of yellow fever virus polyprotein: role of cellular proteases in maturation of the structural proteins. J. Virol. 63, 4199-4209.

Sarubbi E., Seneci P. F., Angelastro M. R., Peet N. P., Denaro M., Islam K., 1993. Peptide aldehydes as inhibitors of HIV protease. FEBS Lett. 319, 253-256.

Schmidt M. F., Isidro-Llobet A., Lisurek M., El-Dahshan A., Tan J., Hilgenfeld R., Rademann J., 2008. Sensitized detection of inhibitory fragments and iterative development of non-peptidic protease inhibitors by dynamic ligation screening. Angew. Chem. Int. Ed.

Engl. 47, 3275-3278.

Schüller A., Yin Z., Brian Chia C. S., Doan D. N., Kim H. K., Shang L., Loh T. P., Hill J., Vasudevan S. G., 2011. Tripeptide inhibitors of dengue and West Nile virus NS2B-NS3 protease. Antiviral Res. 92, 96-101.

Shie J. J., Fang J. M., Kuo T. H., Kuo C. J., Liang P. H., Huang H. J., Wu Y. T., Jan J. T., Cheng Y. S., Wong C. H., 2005. Inhibition of the severe acute respiratory syndrome 3CL protease by peptidomimetic alpha,beta-unsaturated esters. Bioorg. Med. Chem. 13, 5240-5252.

Shiryaev S. A., Kozlov I. A., Ratnikov B. I., Smith J. W., Lebl M., Strongin A. Y., 2007. Cleavage preference distinguishes the two-component NS2B-NS3 serine proteinases of Dengue and West Nile viruses. Biochem. J. 401, 743-752.

Siddell S., Wege H., Ter Meulen V., 1983. The biology of coronaviruses. J. Gen. Virol. 64, 761-776.

Simpson D. I., 1964. Zika Virus Infection in Man. Trans. R. Soc. Trop. Med. Hyg. 58, 335-338.

Stadler K., Allison S. L., Schalich J., Heinz F. X., 1997. Proteolytic activation of tick-borne encephalitis virus by furin. J. Virol. 71, 8475-8481.

Stiasny K., Bressanelli S., Lepault J., Rey F. A., Heinz F. X., 2004. Characterization of a membrane-associated trimeric low-pH-induced form of the class II viral fusion protein E from tick-borne encephalitis virus and its crystallization. J. Virol. 78, 3178-3183.

Stiasny K., Heinz X. F., 2006. Flavivirus membrane fusion. J. Gen. Virol. 87, 2755-2766.

Stoermer M. J., Chappell K. J., Liebscher S., Jensen C. M., Gan C. H., Gupta P. K., Xu W. J., Young P. R., Fairlie D. P., 2008. Potent cationic inhibitors of West Nile virus NS2B/NS3 protease with serum stability, cell permeability and antiviral activity. J. Med. Chem. 51, 5714-5721.

Strode G. K., 1951. Yellow Fever. New York: McGraw-Hill.

Tan J., Verschueren K. H., Anand K., Shen J., Yang M., Xu Y., Rao Z., Bigalke J., Heisen B., Mesters J. R., Chen K., Shen X., Jiang H., Hilgenfeld R., 2005. pH-dependent conformational flexibility of the SARS-CoV main proteinase ($M^{pro}$) dimer: molecular dynamics simulations and multiple X-ray structure analyses. J. Mol. Biol. 354, 25-40.

Tan J., George S., Kusov Y., Perbandt M., Anemuller S., Mesters J. R., Norder H., Coutard B., Lacroix C., Leyssen P., Neyts J., Hilgenfeld R., 2013. 3C protease of enterovirus 68: structure-based design of Michael acceptor inhibitors and their broad-spectrum antiviral effects against picornaviruses. J. Virol. 87, 4339-4351.

Tang H., Hammack C., Ogden S. C., Wen Z., Qian X., Li Y., Yao B., Shin J., Zhang F., Lee E. M., Christian K. M., Didier R. A., Jin P., Song H., Ming G. L., 2016. Zika virus infects human cortical neural progenitors and attenuates their growth. Cell Stem Cell 18, 587-590.

Thornberry N. A., Bull H. G., Calaycay J. R., Chapman K. T., Howard A. D., Kostura M. J., Miller D. K., Molineaux S. M., Weidner J. R., Aunins J., Elliston O. K., Ayala J. M., Casano J. F., Chin J., Ding J.-F. G., Egger A. L., Gaffney P. E., Limjuco G., Palyha C. O., Raju. M. S., Rolando M. A., Salley P. J., Yamin T.-T., Lee D. T., Shively E. J., Maccross M., Mumford A. R., Schmidt A. J., Tocci J. M., 1992. A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes. Nature 356, 768-774.

Torres J., Surya W., Li Y., Liu D., 2015. Protein-protein interactions of viroporins in coronaviruses and paramyxoviruses: new targets for antivirals? Viruses 7, 2858-2883.

Türker L., 2006. Some boric acid esters of glycerol - An *ab initio* treatment. Indian J. Chem. 45A, 1339 – 1344.

Tyrell DA A. J., Berry DM. Cunningham CH, Hamre D, Hofstad MS, Mulluci L, McIntosh K., 1968. Coronaviruses. Nature 220, 650-657.

Vagin A., Teplyakov A., 2010. Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. 66, 22-25.

van der Hoek L., Pyrc K., Jebbink M. F., Vermeulen-Oost W., Berkhout R. J. M., Wolthers K. C., Wertheim-van Dillen P. M. E., Kaandorp J., Spaargaren J., Berkhout B., 2004. Identification of a new human coronavirus. Nat. Med. 10, 368-373.

van Regenmortel H. V. M., Fauquet C. M., Bishop D. H.L., Carsten E. B., Estes M. K., Lemon S. M., Maniloff J., Mayo M.A., McGeoch D. J., Pringle C. R., Wickner R. B., 2000. Virus taxonomy, seventh report of the International Committee on Taxonomy of Viruses 1st Edition. Elsevier Academic Press, Amsterdam, The Netherlands.

Venkatraman S., Bogen S. L., Arasappan A., Bennett F., Chen K., Jao E., Liu Y. T., Lovey R., Hendrata S., Huang Y., Pan W., Parekh T., Pinto P., Popov V., Pike R., Ruan S., Santhanam B., Vibulbhan B., Wu W., Yang W., Kong J., Liang X., Wong J., Liu R., Butkiewicz N., Chase R., Hart A., Agrawal S., Ingravallo P., Pichardo J., Kong R., Baroudy B., Malcolm B., Guo Z., Prongay A., Madison V., Broske L., Cui X., Cheng K. C., Hsieh Y., Brisson J. M., Prelusky D., Korfmacher W., White R., Bogdanowich-Knipp S., Pavlovsky A., Bradley P., Saksena A. K., Ganguly A., Piwinski J., Girijavallabhan V., Njoroge F. G., 2006. Discovery of (1R,5S)-N-[3-amino-1-(cyclobutylmethyl) -2,3-dioxopropyl]-3-[2(S)-[[[(1,1-dimethylethyl)amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl]-6,6-dimethyl-3-azabicyclo[3.1.0]hexan-2(S)-carboxamide (SCH 503034), a selective, potent, orally bioavailable hepatitis C virus NS3 protease inhibitor: a potential therapeutic agent for the treatment of hepatitis C infection. J. Med. Chem. 49, 6074-6086.

Verschueren K. H., Pumpor K., Anemuller S., Chen S., Mesters J. R., Hilgenfeld R., 2008. A structural view of the inactivation of the SARS coronavirus main proteinase by benzotriazole esters. Chem. Biol. 15, 597-606.

Wang F., Chen C., Tan W., Yang K., Yang H., 2016. Structure of Main Protease from Human Coronavirus NL63: Insights for wide spectrum anti-coronavirus drug design. Sci. Rep. 6, 22677.

Weinert T., Olieric V., Waltersperger S., Panepucci E., Chen L., Zhang H., Zhou D., Rose J., Ebihara A., Kuramitsu S., Li D., Howe N., Schnapp G., Pautsch A., Bargsten K., Prota A. E., Surana P., Kottur J., Nair D. T., Basilico F., Cecatiello V., Pasqualato S., Boland A., Weichenrieder O., Wang B. C., Steinmetz M. O., Caffrey M., Wang M., 2015. Fast native-SAD phasing for routine macromolecular structure determination. Nat. Methods 12, 131-133.

Weiss M. S., Hilgenfeld R., 1997. On the use of the merging R factor as a quality indicator for X-ray data. Journal of Applied Crystallography 30, 203-205.

Wengler G., Wengler G., Gross H. J., 1978. Studies on virus-specific nucleic acids synthesized in vertebrate and mosquito cells infected with flaviviruses. Virology 89, 423-437.

Wengler G., Czaya G., Farber P. M., Hegemann J. H., 1991. In vitro synthesis of West Nile virus proteins indicates that the amino-terminal segment of the NS3 protein contains the active centre of the protease which cleaves the viral polyprotein after multiple basic amino acids. J. Gen. Virol. 72, 851-858.

Westaway E. G., Brinton M. A., Gaidamovich S. Y., Horzinek M. C., Igarashi A., Kääriäinen L., Lvov D. K., Porterfield J. S., Russell P.

K., Trent D. W., 1985. Flaviviridae. Intervirology 24: 183-192.

Whelan S., Goldman N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691-699.

Winn M. D., Ballard C. C., Cowtan K. D., Dodson E. J., Emsley P., Evans P. R., Keegan R. M., Krissinel E. B., Leslie A. G., McCoy A., McNicholas S. J., Murshudov G. N., Pannu N. S., Potterton E. A., Powell H. R., Read R. J., Vagin A., Wilson K. S., 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. 67, 235-242.

Wojdyla J. A., Manolaridis I., van Kasteren P. B., Kikkert M., Snijder E. J., Gorbalenya A. E., Tucker P. A., 2010. Papain-like protease 1 from transmissible gastroenteritis virus: crystal structure and enzymatic activity toward viral and cellular substrates. J. Virol. 84, 10063-10073.

Woo P. C., Lau S. K., Chu C. M., Chan K. H., Tsoi H. W., Huang Y., Wong B. H., Poon R. W., Cai J. J., Luk W. K., Poon L. L., Wong S. S., Guan Y., Peiris J. S., Yuen K. Y., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. 79, 884-895.

Woo P. C., Lau S. K., Lam C. S., Lai K., Huang Y., Lee P., Luk G. S., Dyrting K. C., Chan K., Yuen K. Y., 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J. Virol. 83, 908-917.

Wyles D. L., 2013. Antiviral resistance and the future landscape of hepatitis C virus infection therapy. J. Infect. Dis. 207 Suppl. 1, S33-39.

Xing W., Liao Q. Viboud C., Zhang J., Sun J., Wu J. T., Chang Z., Liu F., Fang V. J., Zheng Y., Cowling B. J., Varma J. K., Farrar J. J., Leung

G. M., Yu H., 2014. Hand, foot, and mouth disease in China, 2008-12: an epidemiological study. Lancet Infect. Dis. 14, 308-318.

Xue X., Yang H., Shen W., Zhao Q., Li J., Yang K., Chen C., Jin Y., Bartlam M., Rao Z., 2007. Production of authentic SARS-CoV $M^{pro}$ with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. J. Mol. Biol. 366, 965-975.

Xue X., Yu H., Yang H., Xue F., Wu Z., Shen W., Li J., Zhou Z., Ding Y., Zhao Q., Zhang X. C., Liao M., Bartlam M., Rao Z., 2008. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J. Virol. 82, 2515-2527.

Yang H., Yang M., Ding Y., Liu Y., Lou Z., Zhou Z., Sun L., Mo L., Ye S., Pang H., Gao G. F., Anand K., Bartlam M., Hilgenfeld R., Rao Z., 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc. Natl. Acad. Sci. USA 100, 13190-13195.

Yang H., Xie W., Xue X., Yang K., Ma J., Liang W., Zhao Q., Zhou Z., Pei D., Ziebuhr J., Hilgenfeld R., Yuen K. Y., Wong L., Gao G., Chen S., Chen Z., Ma D., Bartlam M., Rao Z., 2005. Design of wide-spectrum inhibitors targeting coronavirus main proteases. PLoS. Biol. 3, e324.

Yildiz M., Ghosh S., Bell J. A., Sherman W., Hardy J. A., 2013. Allosteric inhibition of the NS2B-NS3 protease from dengue virus. ACS Chem. Biol. 8, 2744-2752.

Yin J., Niu C., Cherney M. M., Zhang J., Huitema C., Eltis L. D., Vederas J. C., James M. N., 2007. A mechanistic view of enzyme inhibition and peptide hydrolysis in the active site of the SARS-CoV 3C-like peptidase. J. Mol. Biol. 371, 1060-1074.

Yin Z., Patel S. J., Wang W. L., Wang G., Chan W. L., Rao K. R., Alam J., Jeyaraj D. A., Ngew X., Patel V., Beer D., Lim S. P., Vasudevan S. G., Keller T. H., 2006a. Peptide inhibitors of Dengue virus NS3 protease. Part 1: Warhead. Bioorg. Med. Chem. Lett. 16, 36-39.

Yin Z., Patel S. J., Wang W. L., Chan W. L., Ranga Rao K. R., Wang G., Ngew X., Patel V., Beer D., Knox J. E., Ma N. L., Ehrhardt C., Lim S. P., Vasudevan S. G., Keller T. H., 2006b. Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors. Bioorg. Med. Chem. Lett. 16, 40-43.

Zaki A. M., van Boheemen S., Bestebroer T. M., Osterhaus A. D., Fouchier R. A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. 367, 1814-1820.

Zeng D., Ma Y., Zhang R., Nie Q., Cui Z., Wang Y., Shang L., Yin Z., 2016. Synthesis and structure-activity relationship of alpha-keto amides as enterovirus 71 3C protease inhibitors. Bioorg. Med. Chem. Lett. 26, 1762-1766.

Zhang J., Pettersson H. I., Huitema C., Niu C., Yin J., James M. N., Eltis L. D., Vederas J. C., 2007. Design, synthesis, and evaluation of inhibitors for severe acute respiratory syndrome 3C-like protease based on phthalhydrazide ketones or heteroaromatic esters. J. Med. Chem. 50, 1850-1864.

Zhao Q., Li S., Xue F., Zou Y., Chen C., Bartlam M., Rao Z., 2008. Structure of the main protease from a global infectious human coronavirus, HCoV-HKU1. J. Virol. 82, 8647-8655.

Zheng A., Umashankar M., Kielian M., 2010. In vitro and in vivo studies identify important features of dengue virus pr-E protein interactions. PLoS Pathog. 6, e1001157.

Zhu L., George S., Schmidt M. F., Al-Gharabli S. I., Rademann J., Hilgenfeld R., 2011. Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease.

Antiviral Res. 92, 204-212.

Ziebuhr J. 2005. The coronavirus replicase. Curr. Top Microbiol. Immunol. 287, 57-94.

Zmurko J., Marques R. E., Schols D., Verbeken E., Kaptein S. J., Neyts J., 2016. The viral polymerase inhibitor 7-deaza-2'-C-methyladenosine is a potent inhibitor of in *vitro* Zika virus replication and delays disease progression in a robust mouse infection model. PLoS Negl. Trop. Dis. 10, e0004695.

## Acknowledgements

First of all, I would like to express my deepest and sincere appreciation to my supervisor: Prof. Dr. Rolf Hilgenfeld. Without out his excellent supervision, this work would not have been realized. He is the person who could initiate elucidation of the structures of some viral proteases before an epidemy of the respective viruses occurs. His preciseness, wisdom on science, and perspective prediction of questions could be asked by reviewers, always inspired me a lot. I cannot forget that he always patiently trained me in writing and presentating. Finally, I would like to thank my supervisor again for his support in science and my life in Germany in the past few years.

I appreciate the generous support from current and former colleagues in the Biochemistry Institute, University of Lübeck, Germany. I wish to sincerely acknowledge my former group leader Dr. Naoki Sakai, and Dr. Qingjun Ma for their generous help of various stages of my work. Many thanks are due to my two former colleagues for their encouragements: Dr. Yibei Xiao, and Dr. Zhenggang Han who were also supervised by Prof. Dr. Rolf Hilgenfeld. Prof. Dr. Stefan Anemüller, Dr. Yuri Kusov, and Dr. Jeroen R. Mesters are thanked for the discussion on kinetics of enzymes, inhibitory activities of compounds against coronavirus and enterovirus proteases, and crystallography, respectively. Dipl.-Ing. Hans-Joachim Kraus is thanked for IT support.

Dr. Daizong Lin and Hailong Liu are acknowledged for synthesizing the α-ketoamides and aldehydes, and Prof. Dr. Christian D. Klein is appreciated for kindly offering the boronic acid compounds.

I also would like to thank people, who gave me help either on science or life (Angela Pampel, Ann-Kristin Gebhardt, Antje Lindae, Charlotte Flory, Doris Mutschall, Henning Bischoff, Jonas Schilz, Dr. Ksenia Pumpor, Linda Chen, Dr. Margret Schwab, Raffaele Ciriello, Silke

Schmidtke, Susanne Zoske, Walter Verheyen, Yasmin Gül).

Financial support by the European Commission through its SILVER project (contract HEALTH-F3-2010-260644) and the German Center for Infection Research (DZIF; TTU 01.803) is gratefully acknowledged.

Finally, I would like to present true love and gratitude to my family members: my grandparents, parents, parents-in-law, and two younger sisters.

Special thanks to my husband (Mr. Yan Liu), the most important person in my life; without your selfless support and continuous company, I would not have stayed alone overseas, and finished my PhD thesis.