
From the Institute of Biochemistry
of the University of Lübeck
Director: Prof. Dr. rer. nat. Dr. h. c. Rolf Hilgenfeld

Structural and Functional Studies on Coronavirus
Non-Structural Proteins 7/8 and 5

Dissertation
for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck

from the Department of Natural Sciences

Submitted by

Yibei Xiao
from Shaoyang, China

Lübeck, 2013



First referee: Prof. Dr. rer. nat. Dr. h.c. Rolf Hilgenfeld

Second referee: Prof. Dr. rer. nat. Norbert Tautz

Date of oral examination: 23.01.2014

Approved for printing. Lübeck, 28.01.2014

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Prof. Dr. rer. nat. Dr. h.c. Rolf Hilgenfeld, for providing me the opportunity to study in Luebeck. His wisdom, support, and encouragement enable me to accomplish this subject with great success. His strict attitude toward science and his humorous personality will have profound influence on me. It is of great honor and pleasure to do research under his guidance and supervision.

In addition, this thesis would not have been possible without the help of the current and past colleagues. I would like to show my gratitude to all their advices and friendship.

Special thanks to Prof. Dr. Georg Sczakiel, Dr. Qingjun Ma, Dr. Shuai Chen, Dr. Jeroen R. Mesters, Dr. Rajesh Ponnusamy, and Dr. Yuri Koussov for their generous support at various stages of my work.

I am grateful for the financial supported by the European Commission through its projects VIZIER (contract no. LSHG-CT-2004-511960) and subsequently by its SILVER project (contract no HEALTH-F3-2010-260644).

Finally, I appreciate my parents and my wife for their sacrifice and selfless support. I love you.

TABLE OF CONTENTS

1. Introduction	1
1.1 Coronaviruses	1
1.2 Coronavirus morphology and structural proteins	3
1.3 Coronavirus genome organization	4
1.4 Translation	5
1.5 Transcription and replication	6
1.6 Nonstructural proteins (Nsps) involved in replication and transcription	8
1.7 Objective and achievement of this thesis	14
2. Materials and Methods	18
2.1. Materials	18
2.1.1 Equipment	18
2.1.2 Consumables	20
2.1.3 Media, chemicals, and enzymes	20
2.2. Methods	21
2.2.1 Cloning	21
2.2.2 Expression	22
2.2.3 Protein purification	22
2.2.4 Dynamic light-scattering	23
2.2.5 Crystallization	23
2.2.6 Diffraction-data collection	25
2.2.7 Structure solution, refinement, analysis, and presentation	25
2.2.8 Glutaraldehyde cross-linking	26
2.2.9 Size-exclusion chromatography	26
2.2.10 Small-Angle X-ray Scattering (SAXS)	27
2.2.11 Circular dichroism	27
2.2.12 RdRp activity assay	28
2.2.13 Enzyme kinetic and inhibition assay for HKU4 M ^{pro}	28

3. Results and Discussion	30
3.1 The FCoV Nsp7+8 complex	30
3.1.1 Protein production	30
3.1.2 Crystallization of the FCoV Nsp7+8 complex	31
3.1.3 Structure solution and quality of the structural models	31
3.1.4 Overall architecture of the Nsp7+8 complex	33
3.1.4.1 Structure of Nsp7 in the complex	34
3.1.4.2 Structure of Nsp8 in the complex	36
3.1.4.3 Interactions between components of the heterotrimer	38
3.1.5 The solution state of Nsp7, Nsp8, and the Nsp7+8 complex.	40
3.1.5.1 Glutaraldehyde cross-linking	40
3.1.5.2 Size-exclusion chromatography	41
3.1.5.3 Small-angle X-ray scattering	42
3.1.6 RdRp activity of Nsp8	44
3.1.7 Interaction screening between hexanucleotides and the FCoV Nsp7+8 complex	46
3.2 The HCoV-229E Nsp7+8 complex	48
3.2.1 Protein production	48
3.2.2 Crystallization of the HCoV-229E Nsp7+8 complex	48
3.2.3 Structure solution and quality of the structural models	49
3.2.4 Overall architecture of the asymmetric unit	51
3.2.4.1 Structure of Nsp7 in the complex	53
3.2.4.2 Structure of Nsp8 in the complex	53
3.2.4.3 Interactions between Nsp7 and Nsp8	56
3.2.5 The solution state of the Nsp7-8 oligoprotein and the Nsp7+8 complex	56
3.2.5.1 Concentration dependence of the oligomerization states of the Nsp7-8 oligoprotein and the Nsp7+8 complex	56
3.2.5.2 The length of the hinge loop influences the oligomerization states of the Nsp7-8 oligoprotein and the Nsp7+8 complex	58

3.2.5.3 SDS-PAGE analyses of the glutaraldehyde-crosslinked and the disulfide-linked Nsp7+8 complex	59
3.2.5.4 Dynamic Light-Scattering	61
3.2.6 RdRp activity of the HCoV-229E Nsp8-containing variants	62
3.3 Discussion of the Nsp7+8 complex	64
3.3.1 Domain swapping of the HCoV-229E Nsp7+8 complex	65
3.3.2 Structure of the HCoV-229E Nsp7-8 oligoprotein	66
3.3.3 Function of the Nsp7+8 complex	67
3.3.4 Active-site(s) within Nsp8	70
3.4 HKU4 M ^{pro}	73
3.4.1 Protein production	73
3.4.2 Crystallization of HKU4 M ^{pro}	73
3.4.3 Structure solution and quality of the structural models	74
3.4.4 Crystal structures of HKU4 M ^{pro} and its complex with inhibitor	76
3.4.4.1 Overall structure of HKU4 M ^{pro}	76
3.4.4.2 The substrate-binding site of HKU4 M ^{pro}	77
3.4.4.3 Crystal structure of HKU4 M ^{pro} in complex with the inhibitor SG85	81
3.4.4.4 Comparison with SARS-CoV M ^{pro}	84
3.4.5 Discussion	86
4. References	90

ABBREVIATIONS

CD	circular dichroism
DdDp	DNA-dependent DNA polymerase
DLS	Dymanic light-scattering
DNA	deoxyribonucleic acid
ds	double-stranded
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
GST	glutathione S-transferase
IPTG	Isopropyl β -D-1-thiogalactopyranoside
kDa	kiloDalton
min	minute
μ l	microlitre
μ M	micromolar
mM	millimolar
mRNA	messenger RNA
nm	nanometer
Nsp	nonstructural protein
NS1	nonstructural protein 1
OB	oligonucleotide/oligosaccharide-binding
ORF	open reading frame
PDB	Protein Data bank
PEG	polyethylene glycol
RdRp	RNA-dependent RNA polymerase
RTC	replication/transcription complex
r.m.s	root-mean-square
RNA	ribonucleic acid
rpm	rounds per minute
SAM	S-adenosyl methionine
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPR	surface plasmon resonance
ss	single-stranded
tRNA	transfer RNA
TRS	transcription-regulating sequence
UTR	untranslated region
v/v	volumn/volumn
wt	wild type

1. Introduction

1.1 Coronaviruses

Coronaviruses are genera of positive-stranded RNA viruses belonging to the subfamily *Coronavirinae* within the family *Coronaviridae*, which are grouped together with *Roniviridae*, *Mesoniviridae*, and *Arteriviridae* to form the order *Nidovirales* (Fauquet *et al.*, 2005; Lauber *et al.*, 2012). Previously, coronaviruses were classified into 3 groups according to serological and antigenic cross reactivity. Alphacoronaviruses and betacoronaviruses are found in mammals, whereas gammacoronaviruses are primarily detected in birds. Recently, novel coronaviruses that are genetically distinct from alpha-, beta- and gamma-coronaviruses have been detected in mammals and terrestrial birds (Dong *et al.*, 2007; Woo *et al.*, 2009; Table 1.1). Viruses of this novel lineage have been proposed to form a new group, provisionally named deltacoronavirus (Chu *et al.*, 2011).

Coronaviruses are believed to cause a significant percentage of all common colds in human adults, usually not life-threatening. But in 2003, the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) spread throughout China and 25 other countries, resulting in over 8000 infections with a fatality rate of about 10%. Ten years after the outbreak of SARS, a new coronavirus called Middle-East Respiratory Syndrome coronavirus (MERS-CoV) was identified. Even though limited human-to-human transmission has occurred within several clusters, the fatality rate is apparently higher than 50%. The emerging life-threatening coronaviruses emphasize the importance of understanding the biology, replication, and pathogenesis of these viruses, and there is an urgent need to develop new strategies to control coronavirus infections.

Table 1.1 Division of the four serological coronavirus groups with representatives

Virus	Abbreviation	Host
Alphacoronavirus		
Transmissible Gastroenteritis coronavirus	TGEV	Pig
Feline coronavirus	FCoV	Cat
Canine coronavirus	CCoV	Dog
Human coronavirus 229E	HCoV-229E	Human
Human coronavirus NL63	HCoV-NL63	Human
<i>Miniopterus</i> bat coronavirus HKU8	HKU8	Bat
Porcine epidemic diarrhea virus	PEDV	Pig
Betacoronavirus		
A		
Human coronavirus HKU1	HCoV-HKU1	Human
Human coronavirus OC43	HCoV-OC43	Human
Bovine coronavirus	BCoV	Cow
Murine coronavirus	MHV	Mouse
B		
Severe acute respiratory syndrome coronavirus	SARS-CoV	Human
SARS-like coronavirus HKU3	HKU3	Bat
C		
<i>Tylonycteris</i> bat coronavirus HKU4	HKU4	Bat
<i>Pipistrellus</i> bat coronavirus HKU5	HKU5	Bat
Middle-East respiratory syndrome coronavirus	MERS-CoV	Human
D		
<i>Rousettus</i> bat coronavirus HKU9	HKU9	Bat
Gammacoronavirus		
Infectious bronchitis virus	IBV	Chicken
Turkey coronavirus	TCoV	Turkey
Beluga whale coronavirus SW1	SW1	Whale
Deltacoronavirus		
Bulbul coronavirus HKU11	HKU11	Bulbul
Thrush coronavirus HKU12	HKU12	Thrush
Munia coronavirus HKU13	HKU13	Munia

Coronaviruses also cause a range of diseases in avians, farm animals, and domesticated pets, resulting primarily in upper respiratory and gastrointestinal tracts diseases, and occasionally causing hepatitis, cardiovascular disease, or neurological illness. Some of these infections are associated with significant mortality and are a big

threat to the farming and pet industry. For instance, porcine Transmissible Gastroenteritis Virus (TGEV) infection results in diarrhea in young pigs with a fatality rate close to 100%; feline coronavirus (FCoV) infection can sometimes result in feline infectious peritonitis (FIP) in cats, which is rare (1%) but fatal.

1.2 Coronavirus morphology and structural proteins

The structure of coronavirions is shown in Fig. 1.1. The virions are spherical enveloped particles about 100-160 nm in diameter (Fig. 1.1 (a)). The “crown-like” appearance gave the virus family its name. Structural proteins that contribute to the overall structure of coronaviruses are the spike glycoprotein (S), the envelope protein (E), the membrane glycoprotein (M), and the nucleocapsid phosphoprotein (N). Some coronaviruses (specifically the members of betacoronavirus clade A, also the gammacoronavirus TCoV), have an additional shorter spike-like protein called hemagglutinin-esterase glycoprotein (HE) (de Haan & Rottier, 2005) (Fig. 1.1 (b)).

The M glycoprotein is a lipid-envelope-anchoring protein which interacts with both N and S proteins; it plays an important role in virus assembly and viral RNA packaging into the nucleocapsid. The E protein is also associated with the virus envelope and required for budding; interestingly, it may also function as an ion channel. The N protein interacts with viral genomic RNA to form the nucleocapsid-RNA complex. The N protein contains an RNA-binding domain, which binds to the leader sequence of viral RNA, and it also interacts with M, thereby incorporating the nucleocapsid-RNA complex into the virion. Several studies have suggested that the N protein may play important roles in both viral RNA synthesis and virus assembly (for reviews see Surjit & Lal, 2008; Sawicki *et al.*, 2007; Masters, 2006; Brian & Baric, 2005). The S protein is a large, multifunctional protein which has several important functions. The S protein binds to specific cellular receptors, thereby determining the host-specificity of different coronaviruses. It also induces fusion between viral envelope and host-cell membranes, and occasionally even cell to cell fusion (for

reviews see Hilgenfeld *et al.*, 2008; Masters, 2006).

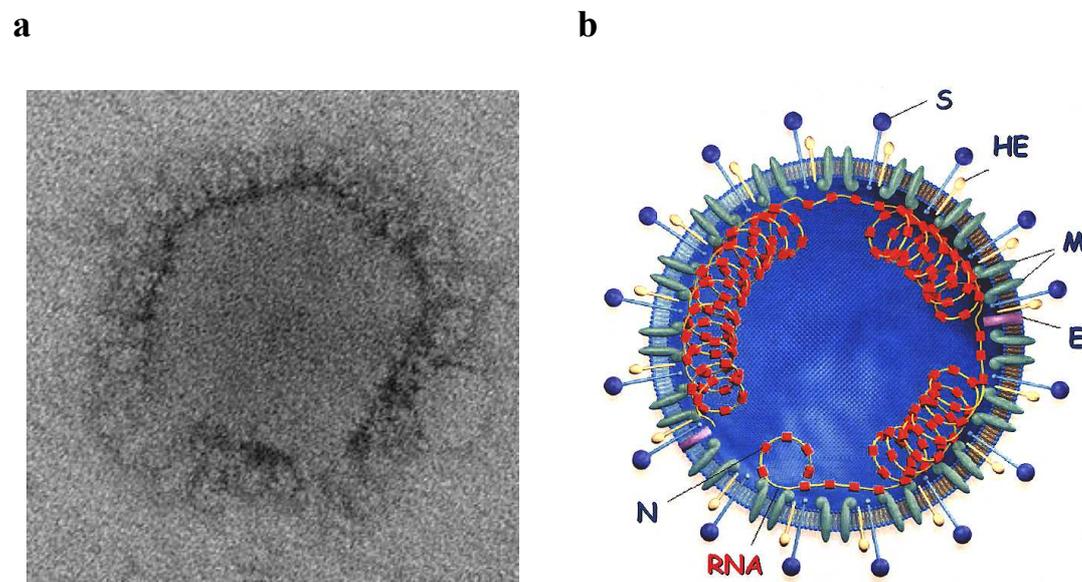


Fig. 1.1 Electron micrograph and systematic representation of coronavirus particles

(a) Negative stain electron micrograph picture of a MERS-CoV particle (<http://www.cdc.gov/coronavirus/mers/photos.html>). (b) Structure representation of the MHV virion (<http://www.nejm.org/doi/full/10.1056/NEJMp030078>). The two figures were last accessed on July 14, 2013.

1.3 Coronavirus genome organization

The RNA genome of coronaviruses is extraordinarily large, with up to 32,000 nucleotides. It is 5'-capped and 3'-polyadenylated. At the 5' end of the genome, there is a leader sequence of about 65 - 98 nucleotides, which is followed by the untranslated region (5'-UTR) and the initiation codon for open reading frame (ORF) 1. At the 3' end, there is the other untranslated region (3'-UTR) following the end of the last ORF and preceding the poly(A) tail. Several secondary structures such as stem loop and pseudoknot are predicted for both UTR regions, which are supposed to play essential functions in genome replication. Many components of the replication/transcription complex (RTC) of CoVs are encoded by ORF1, which covers the 5'-proximal two-thirds of the entire genome. ORF1 is composed of two large open reading frames, ORF1a and ORF1b; the 3' end of ORF1a overlaps with the beginning of ORF1b (for reviews see Sawicki *et al.*, 2007; Masters, 2006; Brian & Baric, 2005).

Secondary structures (stem loops and pseudoknot) and the “slippery” sequence of this overlap region cause a (-1) ribosome frame-shifting at the end of ORF1a; therefore, the ORF is translated into two unusually large polyproteins, pp1a and pp1ab (Namy *et al.*, 2006). Following ORF1, there are another 7 - 14 ORFs, depending on the virus. All the structural and accessory proteins are encoded within the 3'-proximal one-third of the genome. The genes for the major structural proteins are arranged in the following order (5' to 3'): S, E, M, and N (Fig. 1.2).

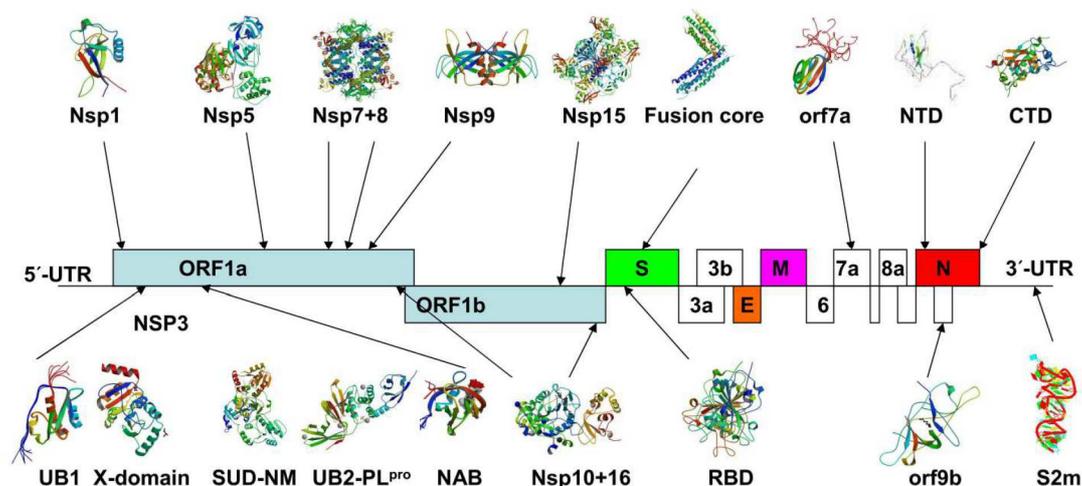


Fig. 1.2 Structure of the RNA genome of SARS-CoV

Three-dimensional structures are depicted for those proteins for which they are available. This figure was used by Hilgenfeld & Peiris, (2013).

1.4 Translation

Coronaviruses do not encapsulate the necessary replicase and transcriptase components within the virus particles, so once the viral genomic RNA is released into the cytoplasm of the infected cell, the virus must first produce its own replicase and transcriptase. Since the viral genome is 5'-capped, translation of the 5'-most ORF (ORF1) is initiated by a cap-dependent ribosomal scanning mechanism. Through the proteolytic activities of cysteine proteases contained within Nsp3 and Nsp5, the primary translation products of ORF1 (pp1a and pp1ab) are processed into intermediate and mature non-structural proteins, ultimately generating up to 15 or 16 Nsps, which form the replication/transcription complex (RTC) together with host-cell factors. The newly formed RTC will then transcribe subgenomic mRNAs, thereby

allowing the translation of the remaining viral proteins (for review see Masters, 2006). Generally, all the viral subgenomic mRNAs are translated into viral proteins, and only the 5'-most ORF of each mRNA can be translated, typically in a cap-dependent manner. But there are also exceptions, for instance, the E protein of MHV is translated from the second ORF of subgenomic mRNA5 (Thiel *et al.*, 1994). Translation of this ORF is cap-independent and there is an internal ribosomal entry site (IRES) preceding the ORF.

1.5 Transcription and replication

Among the RNA viruses, the transcription of coronavirus RNA is unique. First, the large size of the genome requires unusual enzymatic activities, such as an exoribonuclease and an endoribonuclease activity (for reviews see Hilgenfeld *et al.*, 2008; Sawicki *et al.*, 2007), in order to maintain genetic stability. Second, the synthesis of a nested set of subgenomic mRNAs by the discontinuous transcription strategy employed by coronaviruses demands a huge and complicated replication/transcription complex (RTC) (for review see Sawicki *et al.*, 2007).

The replication/transcription complex (RTC) contains multiple enzymatic functions, such as RNA polymerization, modification, and processing. Once the RTC is formed, it will synthesize negative-strand RNAs firstly, which in turn serve as the templates for the transcription of the genomic RNA and a series of subgenomic mRNAs. The coronavirus mRNAs share the same 3' ends but extend differently toward the 5' ends. Therefore, all the mRNAs, except for the smallest one, contain multiple ORFs (for reviews see Sawicki *et al.*, 2007; Lai & Cavanagh, 1997). A common feature of the 5' ends of the mRNAs is the presence of a leader sequence of about 65 to 98 nucleotides, which is derived from the 5' end of the genomic RNA. Following the leader sequence and upstream of the initiating AUG, a short repeated and AU-rich sequence called the transcription-regulating sequence (TRS) is present. In the viral genome, TRS motifs are found at the 3' end of the leader and in front of each ORF (for review see Sawicki *et al.*, 2007).

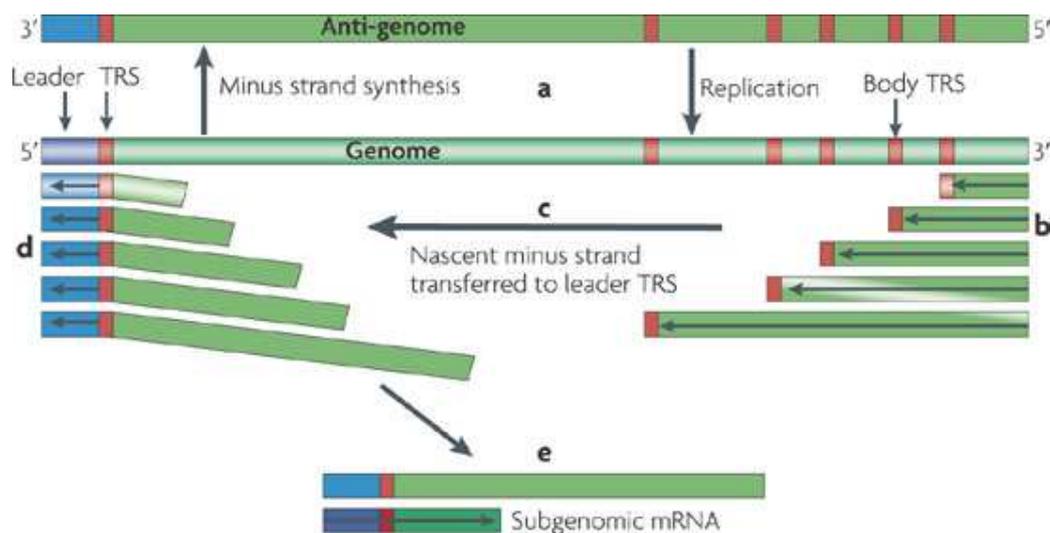


Fig. 1.3 Mechanism of coronavirus replication and transcription

This figure is adopted from Perlman & Netland, (2009).

Two different models have been proposed for the transcription of coronavirus RNA. Both of them support the concept of discontinuous transcription, in spite of controversy over whether discontinuous transcription occurs during positive or negative RNA synthesis. Discontinuous transcription involves two separate transcription steps, one for body-sequence synthesis and the other for leader-sequence synthesis. Most of the existing evidence favors the negative-strand discontinuous transcription model (for a review see Sawicki *et al.*, 2007). In this model, negative-strand RNA synthesis is initiated by the assembly of a functional RTC at the 3' end of a genomic RNA; the RTC will then elongate the nascent negative-strand RNA until the first TRS motif is encountered. A fixed proportion of RTCs will either neglect the TRS motif and continue to elongate the nascent strand, or pause and translocate to the 3' end of the leader sequence, and complete the negative-strand RNA synthesis by copying the 5' end of a genome. This transposition is guided by the annealing of complementary sequences between the nascent negative-strand 3' end TRS motif and the leader TRS motif (for review see Sawicki *et al.*, 2007). The completed negative-strand RNA would then serve as a template for mRNA synthesis (Fig. 1.3).

Replication of the viral genome is presumably achieved by continuous extension of the full-length RNA during both positive- and negative-strand synthesis. However, the “leader switching” experiment supports the idea that coronavirus genomic RNA may also replicate by discontinuous synthesis (Makino & Lai, 1989). The secondary structures in the 5' and 3' UTR regions are supposed to be important for viral RNA replication (Hsue & Masters, 1997; Liu *et al.*, 2007; Züst *et al.*, 2008).

1.6 Nonstructural proteins (Nsps) involved in replication and transcription

In several cases, possible functions of Nsps were derived from their three-dimensional structures as determined by X-ray crystallography or NMR spectroscopy. Many Nsps possess enzymatic activities and are of great importance in replication and transcription, and they are attractive targets for anticoronaviral drug design.

Nsp1

Nsp1 is highly divergent among alphacoronaviruses and betacoronaviruses, whereas it is not even present in gammacoronaviruses. The NMR structure of SARS-CoV Nsp1 revealed a novel β -barrel fold flanked by disordered N- and C-terminal domains (Almeida *et al.*, 2007). Although Nsp1 is not required for viral replication (Züst *et al.*, 2007; Wathelet *et al.*, 2007), it has been suggested to play important role in counteracting cellular immune response through several signaling pathways (Züst *et al.*, 2007; Wathelet *et al.*, 2007; Narayanan *et al.*, 2008; Pfefferle *et al.*, 2011). Furthermore, Nsp1 was shown to bind to 40S ribosomes to inactivate their translation functions and induce host mRNA degradation (Kamitani *et al.*, 2009; Huang *et al.*, 2011; Lokugamage *et al.*, 2012).

Nsp2

Nsp2 is relatively nonconserved across coronaviruses. Deletion of the Nsp2 coding sequence from the SARS-CoV and MHV genomes only attenuates viral growth and RNA synthesis (Graham *et al.*, 2005). Although dispensable for viral replication, Nsp2 of SARS-CoV was confirmed to interact with two host proteins, prohibitin

PHB1 and PHB2, which have been suggested to modulate transcriptional activity by interacting with various transcription factors. Therefore, Nsp2 might be involved in the disruption of intracellular host-signaling during infection (Cornillez-Ty *et al.*, 2009).

Nsp3

As the largest Nsp (180-210 kda) of coronaviruses, Nsp3 is membrane-anchored and contains multiple domains. In the N-terminal part of Nsp3, there is the acidic domain, which contains a large number of Asp and Glu residues. C-terminal to the acidic domain is the papain-like protease PL1^{pro}. Interestingly, the PL1^{pro} is missing in SARS-CoV and is catalytically inactive in IBV (Ziebuhr *et al.*, 2001). This is followed by the X-domain with ADP-ribose-1"-phosphatase (ADRP) activity (Saikatendu *et al.*, 2005; Wojdyla *et al.*, 2009; Piotrowski *et al.*, 2009). However, the X-domain of IBV (strain Beaudette) fails to bind ADP-ribose (Piotrowski *et al.*, 2009). The fold of the X-domain is similar to the macroH2A-like fold, so it is also called macrodomain. Another papain-like protease, PL2^{pro}, is present downstream of the X-domain; together with the PL1^{pro}, it is responsible for cleaving three (IBV: two) Nsps (Nsp1, Nsp2, Nsp3) from the polyprotein (Ziebuhr *et al.*, 2001). Unique for SARS-CoV, in between the X-domain and the PL2^{pro}, there is a domain called "SARS-unique domain" (SUD) which is able to interact with G-quadruplexes (Tan *et al.*, 2007 & 2009). A nucleic-acid binding (NAB) domain has been identified next to the PL2^{pro} (Serrano *et al.*, 2009). At the C-terminal is the highly hydrophobic Y domain anchoring Nsp3 to intracellular membranes.

Nsp4

The crystal structures of Nsp4 C-terminal cytoplasmic domain from MHV and FCoV have been determined, which revealed a novel α/β fold (Xu *et al.*, 2009; Manolaridis *et al.*, 2009). The N-terminal part of the Nsp4 subunit is hydrophobic and is predicted to be membrane embedded. Nsp4, together with Nsp3 and Nsp6, are critical for

double-membrane vesicles (DMVs) formation and modification, as well as replication complex function (Angelini *et al.*, 2013; Gadlage *et al.*, 2010).

Nsp5

Nsp5 is the viral main protease (M^{pro}); it cleaves at least 11 sites in the pp1a and pp1ab polyproteins. Because of its important role in viral polyprotein processing, the M^{pro} is considered the most attractive target for anticoronaviral drug design. Several crystal structures of coronavirus M^{pro} s have been determined, revealing that M^{pro} forms a homodimer that is stabilized by interaction of the N-terminal 7 residues (N-finger) of one protomer with domain II and III of the other protomer. The M^{pro} monomer contains three domains: domains I and II are six-stranded antiparallel β -barrels together forming a chymotrypsin fold, with the substrate-binding pocket located in a cleft between these two domains, while domain III is a globular domain composed of five antiparallel α -helices (Anand *et al.*, 2002 & 2003; Yang *et al.*, 2003).

The M^{pro} employs a cysteine and a histidine residue as the catalytic dyad in its catalytic site, with the histidine acts as a general acid/base and the cysteine performing the nucleophilic attack onto the carbonyl carbon of the scissile bond. The substrate specificities of the coronavirus M^{pro} are well defined, with the proteolytic sites involving large hydrophobic residues (mainly leucine, methionine, and phenylalanine) at the P2 position, conserved glutamine at the P1 position, and small aliphatic residues at the P1' position (Fan *et al.*, 2004; Ziebuhr *et al.*, 1995).

Nsp7 and Nsp8

The crystal structure of the Nsp7:Nsp8 complex of SARS-CoV revealed a hollow, cylindrical hexadecamer composed of eight copies of Nsp8 and eight copies of Nsp7. The channel with a diameter of about 30 Å at the center of this complex suggested that it could encircle double-stranded RNA (dsRNA), in agreement with the observation that it was lined by positively charged amino-acid residues (Zhai *et al.*,

2005). In addition, a heterodimeric complex formed by the Nsp8 C-terminal half and Nsp7 of SARS-CoV was also reported recently. The authors postulated that the truncated complex might be incorporated into the Nsp7+8 supercomplex, thereby interfering with its activity (Li *et al.*, 2008). Furthermore, Imbert *et al.* (2006) described a sequence-specific oligonucleotide-synthesizing activity for Nsp8 of SARS-CoV. In their study, the recombinant Nsp8 carrying an N-terminal His₆-tag was capable of recognizing a specific short sequence (5'-(G/U)CC-3') in the ssRNA coronavirus genome as a template for the synthesis of short oligonucleotides (< 6 residues) in a Mn²⁺-dependent manner, but with a relatively low fidelity. Consequently, Imbert *et al.* proposed that the oligonucleotides synthesized this way might be utilized as primers by the primer-dependent RdRp, Nsp12. According to this proposition, Nsp8 would be the primase of SARS-CoV and Nsp12 its main RNA-dependent RNA polymerase.

Nsp9

The crystal structures of Nsp9 from SARS-CoV and HCoV 229E have been determined and different homodimers were observed (Ponnusamy *et al.*, 2008; Egloff *et al.*, 2004; Sutton *et al.*, 2004). The fold of the Nsp9 monomer is conserved; seven β -strands and one α -helix are arranged into a single compact domain and form an oligonucleotide/oligosaccharide-binding (OB) fold. EMSA and SPR experiments suggest that Nsp9 may function as an ssDNA/RNA-binding protein, and *in-vivo* studies show that dimerization of Nsp9 is essential for efficient viral growth and replication (Miknis *et al.*, 2009; Chen *et al.*, 2009).

Nsp10

Structures for Nsp10 from SARS-CoV have been determined by X-ray crystallography by two groups independently (Su *et al.*, 2006; Joseph *et al.*, 2006). In these two structures, Nsp10 exists in a novel fold either as monomer or dodecamer, but they both possess two different zinc fingers with C-(X)₂-C-(X)₅-H-(X)₆-C and C-(X)₂-C-(X)₇-C-(X)-C motifs. Nsp10 has been proposed to play important roles in

viral RNA synthesis (Donaldson *et al.*, 2007) and polyprotein processing through interaction with the main protease Nsp5 (Donaldson *et al.*, 2007). Recent studies indicate that Nsp10 interacts with both Nsp14 and Nsp16, thereby enhancing their ExoN and RNA Cap 2'-O-methyltransferase activities, respectively (Bouvet *et al.*, 2012; Decroly *et al.*, 2011; Chen *et al.*, 2011).

Nsp12

In addition to the noncanonical primer-independent RdRp Nsp8, it is presumed that replication and transcription of the ~30-kb coronavirus RNA genome are mainly catalyzed by the RdRp activity of Nsp12. In spite of high sequence identity among Nsp12 from different coronaviruses (~ 62 - 73%), the enzyme shares <10% sequence identity with other viral RdRps and RTs. Due to low purification yields of the enzyme, biochemical and structural information on Nsp12 is only limited for SARS-CoV. SARS-CoV Nsp12 is a large protein containing 932 amino-acid residues. Bioinformatical analysis predicted this enzyme to comprise two large domains, the N-terminal domain (residues 1 - 375) and the C-terminal domain (residues 376 - 932). The N-terminal domain of SARS-CoV Nsp12 shares no counterpart with other positive-strand RNA-virus RdRps, therefore its function is still uncharacterized. One speculation could be that the N-terminal domain is involved in interaction with other components within the RTC, e.g. with Nsp8 or Nsp13. The C-terminal domain of Nsp12 can be confidently aligned with the conserved motifs of well-characterized RdRps, thus the RdRp activity is presumed to be located in this domain (Xu *et al.*, 2003).

te Velthuis *et al.* (2010a) have shown that C-terminally hexahistidine-tagged full-length SARS-CoV Nsp12 exhibits both RdRp and DdDp activities in the presence of Mg^{2+} *in vitro*, and its enzymatic activity is primer-dependent for both homo- and heteropolymeric templates. In contrast, the RdRp activity of N-terminally hexahistidine-tagged full-length SARS-CoV Nsp12 used by Ahn *et al.* (2012) is Mn^{2+} -dependent, and it can readily synthesize RNA in both primer-dependent and

primer-independent manners. Zn^{2+} was found to reduce template-binding and to inhibit the elongation step of SARS-CoV Nsp12 RdRp (te Velthuis *et al.*, 2010b).

Nsp13

Nsp13 is a helicase that unwinds double-stranded RNA (dsRNA) or DNA (dsDNA) in a 5'→3' direction, using the energy of nucleotide hydrolysis. The structure of Nsp13 has not been determined yet, but it is predicted to contain two domains; the N-terminal region contains a zinc-binding domain and the C-terminal domain features the canonical sequence motifs of superfamily-1 helicases (Ivanov *et al.*, 2004). Recent studies show that the helicase activity of Nsp13 is enhanced in the presence of Nsp12, suggesting that this protein is involved in formation of the functional replication complex and contributes to the efficiency of viral replication (Adedeji *et al.*, 2012).

Nsp14

Nsp14 is a bifunctional protein. Its N-terminal domain possesses 3'→5' exoribonuclease (ExoN) activity, whereas a (guanine-N7)-methyltransferase (N7-MTase) activity is embedded in the C-terminal domain. The Nsp14 could methylate GpppA-capped RNA at the N7 position of the capping guanosine in a sequence-independent manner. But the N7-MTase domain is not functionally separable from the ExoN domain, because deletions of its N-terminal domain impair the N7-MTase activity (Chen *et al.*, 2009). Interestingly, interaction with Nsp10 promotes around 35-fold larger exoribonuclease activity of Nsp14 while its N7-MTase activity remains the same (Bouvet *et al.*, 2012). The ExoN domain of Nsp14 contains an Asp-Glu-Asp-Asp motif (DEDD), which is conserved in the ExoN superfamily and is capable of hydrolyzing ssRNA and dsRNA. Nsp14-mutated SARS-CoV exhibits an overall 12- to 20- fold increase in mutation frequency compared to the wild-type (Eckerle *et al.*, 2010). Therefore, Nsp14 is proposed to be involved in proofreading, repair, and recombination, which might be important for maintaining the integrity of the large CoV genome (Bouvet *et al.*, 2012).

Nsp15

Nsp15 (NendoU) is a Mn^{2+} -dependent, uridylyate-specific endoribonuclease, which generates 2'–3'-cyclic phosphates upon cleavage (Ricagno *et al.*, 2006). The protein is conserved among nidoviruses but not other RNA viruses. Nsp15 has been reported to form monomers, trimers, and hexamers in solution. Both monomeric and hexameric structures of the SARS-CoV Nsp15 have been determined by X-ray crystallography (Ricagno *et al.*, 2006; Joseph *et al.*, 2007). The monomer of Nsp15 consists of three domains and exhibits a unique fold. Furthermore, this protein is only active when it assembles into a toric hexamer with six potentially active, peripheric catalytic sites, suggesting a role for hexamerization as an allosteric switch (Guarino *et al.*, 2005).

Nsp16

Feline coronavirus (FCoV) Nsp16 alone possesses an S-adenosylmethionine (SAM)-dependent (nucleoside-2'O)-methyltransferase (2'O-MTase) activity that is capable of cap-1 formation (Decroly *et al.*, 2008). In contrast, SARS-CoV Nsp16 requires Nsp10 as an activating partner to execute its 2'-O-MTase activity. The crystal structure of the Nsp10+16 complex reveals that SARS-CoV Nsp16 possesses the canonical scaffold of an MTase and interacts with Nsp10 at an 1:1 ratio. Biochemical assays show that in case of SARS-CoV, Nsp10 is required for Nsp16 to bind both SAM and RNA substrate, and the crystal structure demonstrates that Nsp10 may stabilize the SAM-binding pocket and extend the RNA-binding groove of Nsp16. Remarkably, an N7-methyl guanosine cap (m^7GpppA -RNA, but not m^7GpppG -RNA) is a prerequisite for the activity of Nsp16 (Chen *et al.*, 2011; Decroly *et al.*, 2011).

1.7 Objective and achievement of this thesis

Life-threatening RNA viruses often emerge in an unpredictable manner. An example of an epidemic caused by an emerging virus was the outbreak of SARS-CoV in 2003. The high case/fatality ratio and the devastating economic impact of this virus highlight the importance of research on coronaviruses. In the past ten years, surveillance in wild animals and humans has led to the discovery of many novel

viruses, including coronaviruses. Evidence presented in these studies suggests that exotic wildlife, especially bats, are natural reservoirs of coronaviruses and can be the source of new emerging diseases through interspecies transmission (Smith & Wang, 2013). In addition, many human coronaviruses are supposed to have originated from animal coronaviruses, especially bat coronaviruses. For example, SARS-CoV very likely originated from a bat reservoir and was transmitted to humans via the civet as intermediate (Li *et al.*, 2005). An alphacoronavirus isolated from the North American tricolored bat has been proposed to share common ancestry with HCoV-NL63 (Huynh *et al.*, 2012).

As bats constitute a large reservoir for CoVs with zoonotic potential, our group initiated an antiviral-drug discovery program based on the crystal structures of bat CoV (HKU4, HKU8 and HKU9) main protease. Crystal structures determined by the group were used to design and synthesize broad-spectrum inhibitors that would be available for further development in case of zoonotic transmission of such a virus into the human population. Remarkably, after determination of the crystal structure of bat coronavirus HKU4 M^{pro}, both as the free enzyme and in complex with the broad-spectrum inhibitor SG85 (see chapter 3.4), the new human coronavirus MERS-CoV emerged in the Middle East. This novel human coronavirus exhibits around 80% sequence identity to the bat coronavirus HKU4 (van Boheemen *et al.*, 2012), and its M^{pro} has 81% identity (90% similarity) to its homologue from HKU4. Thus, we can readily use the M^{pro} structure of HKU4 for the design of MERS-CoV M^{pro} inhibitors.

Another goal of this thesis is to elucidate the structure and function of coronavirus Nsp7 and Nsp8, which are key components of the RTC. As mentioned above, the SARS-CoV Nsp7+8 complex forms a hollow, cylindrical hexadecamer composed of eight copies of Nsp8 and eight copies of Nsp7 (Zhai *et al.*, 2005), which is able to synthesize short oligonucleotides (< 6 residues) in a Mn²⁺-dependent manner (Imbert *et al.*, 2006). However, many questions regarding the roles of Nsp7 and Nsp8 in

coronavirus RNA synthesis remain. First, the function of Nsp7 in the primase activity is poorly characterized so far. At present, Nsp7 is only regarded as "mortar" stabilizing the hexadecameric Nsp7+8 complex structure (Zhai *et al.*, 2005), without having significant influence on the synthesis of RNA primers (Imbert *et al.*, 2006). Second, since Nsp8 alone and the Nsp7+8 complex showed equal primase activity in the study by Imbert *et al.*, we still do not know the stoichiometry between the two components *in vivo*, or whether varying stoichiometries may exist at different transcriptional stages. Third, even though the observed Nsp7+8 hexadecamer presents a positive electrostatic channel suitable for dsRNA binding, the mode of action at the molecular level of this non-canonical primase is not understood. Therefore, it is important to investigate Nsp7+8 of coronaviruses other than SARS-CoV, to see whether the hexadecamer occurs universally in all coronaviruses and to hopefully shed light on the molecular mechanism of RNA synthesis by Nsp8 and its complex with Nsp7.

In this thesis, I found that in other coronaviruses, the Nsp7+8 complex possesses different stoichiometries and oligomerization states, and functions as RdRp rather than only primase. I have determined the crystal structure of the feline coronavirus (FCoV) Nsp7+Nsp8 complex, which is a 2:1 heterotrimer containing two copies of the α -helical Nsp7 with conformational differences between them, and one copy of Nsp8 that consists of an α/β domain and a long- α -helix domain. Compared with the Nsp8 of SARS-CoV, the FCoV Nsp8 exhibits primer-independent RdRp activity with much higher processivity. Upon binding with Nsp7, the orientation of the FCoV Nsp8 long- α -helix domain is fixed, potentially leading to higher RdRp activity, and RNA up to template-length can be synthesized. In addition to the FCoV Nsp7+8 complex, I have also determined the crystal structure of the human coronavirus 229E (HCoV-229E) Nsp7+8 complex. Interestingly, this structure reveals a 2:2 heterotetramer. In the heterotetramer, two Nsp8 molecules dimerize with each other through swapping of the C-terminal domain. I further proved that the unprocessed HCoV-229E Nsp7-10 polyprotein is only able to synthesize short oligonucleotides,

while the Nsp7-8 polyprotein or the Nsp7+8 complex are fully processive.

2. Materials and Methods

2.1. Materials

2.1.1 Equipment

Equipment or instruments used for work described in the thesis are listed in Table 2.1.

Table 2.1: Equipments or Instruments and their manufactures

Equipment or Instruments	Manufacturer
- 80 °C Refrigerator storage unit	REVCO – Kendro
- 20 °C/ 4 °C Refrigerator storage unit	LIEBHERR/Siemens
4 °C Cold-room cabin	eco – Coils & Coolers
12 °C Cold-room cabin	VISSMANN
AF20 Ice-machine	Scotsman
Äkta prime FPLC	Amersham Biosciences
Analytical balance - Sartorius BP 210 D	Sartorius
Satorius Universal balance	Sartorius
Avanti J-26xP centrifuge	Beckman Coulter
RC-5B plus centrifuge	DuPont Sorvall
Table-top centrifuge	Heraeus Instruments
Concentrator 5301	Eppendorf
Cryostream Cooler	Oxford Cryosystems
Excella E10 Platform Shaker	New Brunswick Scientific
French Pressure Cell Press	SLM Instruments
GeneAmp PCR System 9700	PE Applied Biosystems
Gene Pulser	Biorad
Ice-box	neoLab
Innova 4230 Refrigerated Incubator Shaker	New Brunswick Scientific
Laminar-flow chamber	Heraeus/GELAIRE
Dynamic light-scattering	RiNA

Membrane-vacuum pump	Vacuubrand
Mighty Small SE 245 Dual Gel Caster	Hofer
Mighty Small II SE 250 Elektrophoresis - chamber	Hofer
Microscope SZX 12	Olympus
Microwave oven	Siemens
Mini-Sub Cell for agarose gel elektrophoresis	Biorad
MR-2002 magnetic stirrer	Heidolph
Petri dishes	Greiner bio-one
pH-meter/titrator TR156	Schott
Phoenix crystallization robot	Art-Robbins Instruments
Pipettes	Eppendorf
Power PAC 300 Voltage adaptor	Biorad
Power pack EPS 3501 XL	Amersham Biosciences
Sonicator	Branson
Spectrophotometer - Cary 50 UV/Vis	Varian
Systec V 150 Autoclave	Systec
Thermomixer comfort	Eppendorf
Hot air oven	Thermo Scientific
Xcalibur PX Ultra X-radiation diffractometer	Oxford Diffraction
Cryo-tools and crystallization materials	Manufacturer
Intelli-Plate-96-well	Art Robbins Instruments
Magnetic basic crystal caps	Hampton Research
Nylon cryo-loops	Hampton Research
22-mm siliconized coverslips	Hampton Research
Vaseline	Hampton Research
crystal storage vials	Hampton Research
cryo-canes	Hampton Research
magnetic crystal wand	Hampton Research
microtools	Hampton Research
Crychem Plates	Hampton Research

Crystal Clear Sealing Tape	Henkel
Cryoloops (litho)	MiTeGen

2.1.2 Consumables

Consumables used for work described in this thesis are listed in Table 2.2.

Table 2.2: Consumables and their manufactures

Material	Manufacturer
Amicon-Ultra centrifuge filter	Millipore
Dialysis tubes Spectra/Por	Spectrum Laboratories
DNA mini kit	Qiagen
Electroporation cuvettes	Biorad
Erlenmeyer flasks	Schott Duran
Filter membranes (0.45 μm /0.22 μm)	Millipore
GST trap High Performance	GE-Healthcare
Screw-capped glass bottles	Schott Duran
HisTrap High Performance/Fast Flow	GE-Healthcare
Hi Load Superdex 75 pg	GE-Healthcare
Hi Load Superdex 200 pg	GE-Healthcare
Tips and tubes	Sarstedt
UV cuvettes (plastic/quartz)	Sarstedt

2.1.3 Media, chemicals, and enzymes

All chemicals and reagents were purchased from Fluka, Merck, Roth, Roche, or Sigma–Aldrich Ltd. Bacterial cultures were grown in 2x YT medium or M9 medium (for the selenomethionyl (SeMet) derivative of FCoV Nsp8), supplemented with 0.1 mg/ml ampicillin. All Tris-HCl buffers were prepared at 4°C. All restriction enzymes were purchased from Fermentas. All PCR enzymes and other necessary DNA isolation and PCR purification kits were purchased from Invitrogen, Roche, Fermentas, or Promega.

2.2. Methods

2.2.1 Cloning

DNA sequences encoding target genes were amplified by PCR using *Pfu* DNA polymerase (Invitrogen) and primers encoding the 5' and 3' ends. To produce HKU4 M^{Pro} with a C-terminal hexahistidine tag, a sequence of six histidines was added onto the reverse primer of HKU4 M^{Pro}. The PCR products were then cloned into the expression vector pGEX-6p-1 (GE-Healthcare) using BamHI and XhoI sites. All the mutants were prepared using single-site mutagenesis. The *Escherichia coli* strain DH5 α (Invitrogen) was used for plasmid amplification. The primers used for cloning are listed in Table 2.3; information on genes is listed in Table 2.4.

Table 2.3: Oligonucleotide primers

Name	Sequence	Restriction enzyme
FCoVNsp8-S	CGCGGATCCGTAGCATCTGCTTATGCTGCTT	BamHI
FCoVNsp8-R	CCGCTCGAGTTACTGCAACTTTGTGGTTCTC	XhoI
FCoVNsp7-S	CGCGGATCCAAACTTACAGAGATGAAGTGT	BamHI
FCoVNsp7-R	CCGCTCGAGTTACTGTAAAATAGTGGTGTCTCA	XhoI
229ENsp8-S	CGCGGATCCAACGACTCCATTTTGCAA	BamHI
229ENsp8-R	CCGCTCGAGTTACTGCAATTTAACGACAC	XhoI
229ENsp7-S	CGCGGATCCGTTTCTACTGTACAGTCTAAATTGACTGATCTT	BamHI
229ENsp7-R	CCGCTCGAGTTATTGCAAAATGGAGTCG	XhoI
229ENsp8T123C-S	TCCGTTATCCCTGCTTGTCTGCAGCCAGGCTC	
229ENsp8T123C-R	GAGCCTGGCTGCAGACAAGCAGGGATAACGGA	
229ENsp8G113C-S	AATATGGCACGTAATTGTGTTGTCCCTCTTTCC	
229ENsp8G113C-R	GGAAAGAGGGACAACAATTACGTGCCATATT	
229ENsp8P133C-S	AGGCTCGTCGTCGTAGTATGCGATCATGATTCATTT	
229ENsp8P133C-R	AAATGAATCATGATCGCACTACTACGACGACGAGCCT	
229ENsp8insert-S	ATCCCTGCTACTTCTGCAACTTCTGCAGCCAGGCTCGTCGTCGTA	
229ENsp8insert-R	TACGACGACGAGCCTGGCTGCAGAAGTTGCAGAAGTAGCAGGGAT	
229ENsp8delet-S	CTTTCCGTTATCCCTGCTGCCAGGCTCGTCGTCGTA	
229ENsp8delet-R	TACGACGACGAGCCTGGCAGCAGGGATAACGGAAAG	
HKU4Nsp5-S	CGCGGATCCAGTGTGCTGCAATCAGGTCTGGTCAAGATGTC	BamHI
HKU4Nsp5-R	CCGCTCGAGTTAATGATGATGATGATGTTGCATGACGACACCCATTA	XhoI

Table 2.4: Genes and expression constructs

Name	Provider	Notes
pET11d-229E710	Prof. John Ziebuhr (Queen's University, Belfast, UK)	C-terminally His ₆ -tagged HCoV-229E Nsp7-10 polyprotein
FCoV-Nsp7 (Vector unknown)	Dr. Bruno Coutard (AFMB, Univ. de la Méditerranée, Marseille, France)	C-terminally His ₆ -tagged FCoV strain FIPV WSU-79/1146
FCoV-Nsp8 (Vector unknown)		N-terminally His ₆ -tagged FCoV strain FIPV WSU-79/1146
HKU4-M ^{pro}	MWG-Biotech	cDNA, codon-optimized and chemically synthesized

2.2.2 Expression

Proteins used in this thesis were produced in a similar way. A sequence-verified clone was transformed into the *Escherichia coli* strain BL21-Gold (DE3). The culture was grown at 37°C in 2x YT medium with vigorous shaking until the OD₆₀₀ reached 0.8 (FCoV Nsp7, 1.3). Expression was induced by adding 0.5 mM IPTG and incubation at 25°C for 16 h. The selenomethionyl (SeMet) derivative of FCoV-Nsp8 was produced by using the method of methionine-biosynthesis pathway inhibition (Doublé, 1997). Cells were harvested by centrifugation at 5000 rpm for 20 min at 4°C.

2.2.3 Protein purification

Purification of GST-tagged protein

Cells were lysed by sonication in buffer containing 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM DTT, and EDTA-free cocktail protease inhibitor (Roche; one tablet per 50 ml buffer). The lysate was cleared by ultracentrifugation at 30,000 rpm for 90 min at 4°C, and the soluble fraction was applied onto a pre-equilibrated 5-ml GSTrap FF column (GE Healthcare). The column was washed with 20 column volumes of lysis buffer and 10 column volumes of PreScission cleavage buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT), followed by loading the PreScission protease (GE Healthcare) onto the column. Cleavage was performed at 4°C overnight, and the eluted samples were further purified by size-exclusion chromatography

(HiLoad 16/60 Superdex 75/200; GE Healthcare) equilibrated with buffer C (10 mM Tris-HCl pH 7.5, 200 mM NaCl).

Purification of His-tagged protein

Cells were lysed by sonication in buffer A containing 50 mM Tris-HCl pH 8.0 and 300 mM NaCl. After ultracentrifugation at 30,000 rpm for 90 min at 4°C, the supernatant was applied onto a pre-equilibrated 5-ml HisTrap HP column (GE Healthcare). After washing with buffer A, the protein was eluted with buffer B (50 mM Tris-HCl pH 7.5, 300 mM NaCl, and 500 mM imidazole). The eluted samples were further purified by size-exclusion chromatography using the same procedure as described above.

Nsp7+8 complex formation and purification

The FCoV Nsp7+8 complex was formed by mixing Nsp8 with a 2-fold molar excess of Nsp7 and passed through the Superdex 75 column in buffer C. Fractions of the FCoV Nsp7+8 complex were concentrated to 15 mg/ml. The HCoV-229E Nsp7+8 complex was produced by adding SARS-CoV M^{pro} to the HCoV-229E Nsp7-8 oligoprotein and cleavage at 12°C overnight. The HCoV-229E Nsp7+8 complex was further purified *via* the Superdex 200 column in buffer C. Fractions of the HCoV-229E Nsp7+8 complex were concentrated to 25 mg/ml.

2.2.4 Dynamic light-scattering

Measurements were performed by using a Spectroscatter 201 (RiNA GmbH, Germany) using 25 µL of protein solution in its storage buffer. The concentration used for all the proteins was around 15 mg/ml. Results were interpreted using the software provided by the manufacturer. Experimental errors were estimated as standard deviations calculated from 10 measurements for each sample.

2.2.5 Crystallization

Crystallization screening was performed using a Phoenix robot (Art Robbins Instruments) with the sitting-drop, vapor-diffusion method at 12°C and drops

containing 0.3 μL of protein solution plus 0.3 μL of reservoir solution. The following commercial screens were used: SaltRxTM, PEG/IonTM, IndexTM, Crystal ScreenTM, and PEGRxTM (all from Hampton Research).

FCoV Nsp7+8 complex

Tiny crystals appeared in a drop containing 2.4 M $(\text{NH}_4)_2\text{HPO}_4$, 0.1 M Tris-HCl pH 8.5. Optimized crystals suitable for diffraction were grown in drops containing 1.5 μL of protein solution and 1.5 μL of reservoir solution (2.24 M $(\text{NH}_4)_2\text{HPO}_4$, 2 mM tris(2-carboxyethyl)phosphine (TCEP), 0.1 M Tris-HCl pH 8.7) at 4°C. Prior to being cryo-cooled in liquid nitrogen, crystals were soaked in reservoir solution supplemented with 20% glycerol.

HCoV-229E Nsp7+8 complex

Crystals appeared in a drop containing 0.2 M calcium chloride, 28% v/v polyethylene glycol 400, 0.1 M HEPES-NaOH pH 7.5. Optimized crystals were grown on hanging-drop plates, drops containing 2 μL of protein solution and 2 μL of precipitant solution (0.2 M calcium chloride, 21% v/v polyethylene glycol 600, 0.1 M Tris-HCl pH 7.8) at 20°C, 1.4 M sodium chloride was used as reservoir solution. Macroseeding was used to optimize crystals suitable for diffraction. Crystals were cryo-cooled in liquid nitrogen; precipitant solution supplemented with 15% glycerol was used as cryoprotectant.

HKU4 M^{pro}

Optimized crystals suitable for diffraction were grown in drops containing 2 μL of protein solution and 2 μL of reservoir solution (6% w/v polyethylene glycol 3350, 2% v/v Tacsimate, 5% v/v 2-propanol, 0.1 M imidazole pH 6.9) at 20°C. Compound SG85 was dissolved in DMSO firstly and then diluted with reservoir solution to 2 mM, SG85-HKU4 M^{pro}-complex crystals were prepared by soaking the crystals in this solution for 48 h. Before cryo-cooling in liquid nitrogen, crystals were soaked in reservoir solution supplemented with 15% glycerol.

2.2.6 Diffraction-data collection

Diffraction-data were collected at 100 K using beamline BL14.2, BESSY (Berlin, Germany), which is equipped with an MX225 CCD detector (Rayonics).

2.2.7 Structure solution, refinement, analysis, and presentation

FCoV Nsp7+8 complex

Indexing, integration, and scaling of data were carried out with XDS (*via* the interface program ixds, developed at BESSY) (Kabsch, 2010). The program XPREP (Bruker) was used to further analyze and prepare datasets for structure solution and refinement. The selenium substructure was determined using SHELXD (Sheldrick, 2010). Initially, the electron density was interpreted automatically using SHELXE (Sheldrick, 2010) and ARP/wARP (Hattne *et al.*, 2008). Thereafter, model refinement and manual rebuilding were performed alternately. For refinement, Refmac5 (Murshudov *et al.*, 1997) was used initially and autoBUSTER (Bricogne *et al.*, 2010) subsequently, whereas model building was carried out employing Coot (Emsley *et al.*, 2010). Molecular figures were prepared using PyMol (Schrödinger).

HCoV-229E Nsp7+8 complex

Diffraction-data were processed using the MOSFLM software (Leslie *et al.*, 2007). For refinement and model rebuilding, autoBUSTER (Global Phasing) and COOT (Emsley *et al.*, 2010) were performed alternatively. The interface analysis was done by using the PISA webserver (http://www.ebi.ac.uk/msd-srv/prot_int/). Other crystallographic analyses were carried out using the CCP4 package (Winn *et al.*, 2011). Molecular figures were prepared using PyMol (Schrödinger).

HKU4 M^{pro}

Diffraction-data were processed using the autoProc software (Vonnrhein *et al.*, 2011). A preliminary model was built using ARP/wARP (Hattne *et al.*, 2008) and refined using Refmac5 (Murshudov *et al.*, 1997). Structure refinement with autoBUSTER (Global Phasing) and model rebuilding with COOT (Emsley *et al.*, 2010) were performed

alternatively. Restraints for the inhibitor were generated by the Grade webserver (Global Phasing). Interactions between the protein and the inhibitor were analyzed using Ligplot (Wallace *et al.*, 1995). The dimerization interface of the HKU4 M^{pro} was analyzed using the PISA webserver (http://www.ebi.ac.uk/msd-srv/prot_int/). Other crystallographic analyses were carried out using the CCP4 package (Winn *et al.*, 2011). Molecular figures were prepared using PyMol (Schrödinger).

2.2.8 Glutaraldehyde cross-linking

Chemical cross-linking was performed for Nsp7, Nsp8, and the Nsp7+8 complex of FCoV. All samples were diluted to 2 mg/ml in 10 mM HEPES pH 7.5, 200 mM NaCl, 5 mM DTT. An aliquot of 25% (v/v) glutaraldehyde was diluted and added to the protein samples to give a final concentration of 0.01% (for Nsp7 and Nsp8) or 0.05% (for the Nsp7+8 complex). The reaction was performed at 4°C for 10 min followed by quenching through the addition of 1.0 M Tris-HCl pH 8.0 (0.5% v/v). Subsequently, an equal volume of loading buffer was added to the reaction mix and the samples were heated at 95°C for 10 min, followed by SDS-PAGE (15% acrylamide gels).

2.2.9 Size-exclusion chromatography

Solution states of Nsp7, Nsp8, Nsp7-8 and the Nsp7+8 complex were analyzed by size-exclusion chromatography (SEC). For FCoV proteins, 1 ml of each protein solution with a concentration of 15 mg/ml was loaded onto a HiLoad 16/60 Superdex 75 column (GE Healthcare). The linear relationship between the gel-phase distribution coefficient (K_{av}) and the average molecular mass (MW) was fitted by $K_{av} = -0.49265 \log MW + 2.51301$. For HCoV-229E proteins, 0.5 ml protein with different concentrations was loaded onto a Superdex 200 10/300 GL column (GE Healthcare). The linear relationship between the gel-phase distribution coefficient (K_{av}) and the average molecular mass (MW) was fitted by $K_{av} = -0.380131 \log MW + 1.01481$. All these experiments were performed in buffer containing 10 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM DTT, and with the detection of absorbance at 280 nm.

2.2.10 Small-Angle X-ray Scattering (SAXS)

SAXS measurements for FCoV Nsp7 and the Nsp7+8 complex were carried out using the EMBL BioSAXS beamline X33 (Roessle *et al.*, 2007; Round *et al.*, 2008) at the DORIS storage ring, DESY (Hamburg, Germany). A photon-counting PILATUS detector with 1 million pixels was used to record the scattered X-rays at a wavelength of 1.54 Å. The sample-to-detector distance was 2.7 m, yielding a maximum recordable momentum transfer ($s = 4\pi \sin\theta/\lambda$) of 0.6 Å⁻¹. A concentration series was analyzed for each sample at 20°C (1.0, 2.0, 4.0, 8.0, and 16.0 mg/ml for Nsp7+8 and 1.2, 2.4, 4.7, and 9.4 mg/ml for Nsp7). The data processing steps were carried out with the help of W. Shang and D. Svergun, EMBL Hamburg Outstation. For either sample, no concentration-dependence of the scattering profile was observed and the SAXS data from the most-concentrated sample were used for further analysis. Standard SAXS data reduction and analysis were carried out using PRIMUS (Kabsch, 2010). The forward scattering $I(0)$ and radius of gyration R_g were evaluated using the Guinier approximation (Guinier, 1939) assuming that at very small angles ($s < 1.3/R_g$), the intensity is represented as $I(s) = I(0) \exp(-(sR_g)^2/3)$; the Guinier plots were linear suggesting monodisperse solutions. The molecular mass of the solutes was evaluated by comparison of the forward scattering $I(0)$ with that from a reference solution of bovine serum albumin (molecular mass 66 kDa; 4.5 mg/ml in 50 mM HEPES, pH 7.5). The maximum size of the protein molecule, D_{\max} , and the pair distance distribution function $p(r)$ were calculated with the program GNOM (Svergun, 1992). The program CRY SOL (Svergun, 1995) was used to calculate the X-ray-scattering profiles with the high-resolution model from the crystal structure (for Nsp7 alone, the corresponding part of the crystal structure of Nsp7+8 was used).

2.2.11 Circular dichroism

CD spectra were measured using a Jasco J-175 spectropolarimeter equipped with a temperature-controlled quartz cell of 0.2 cm length path. Spectra were measured between 190 and 300 nm at 25°C. The molar ellipticity was calculated by using the software K2D3 based on the mean residue mass (Louis-Jeune *et al.*, 2012).

HCoV-229E Nsp7-8 oligoprotein and Nsp7+8 complex were measured at a final concentration of 50 µg/ml in water.

2.2.12 RdRp activity assay

An RNA oligonucleotide representing stem-loop 1 of the 3'-UTR of the FCoV strain FIPV WSU-79/1146 genome (5'-GGCAACCCGAUGUCUAAAACUUGUCUUUCC GAGGAAUUACGGGUCAUCGCGCUGCCUACUCUUGUAC-3', 67 nt) (Riboxx) was used as template in the *de-novo* polymerase reaction. The standard reaction mixture (20 µl) contained 25 mM Tris-HCl pH 8.0, 10 mM KCl, 4 mM MgCl₂, 10 mM DDT, 1 µCi of [α -³²P] GTP, 20 µM each of ATP, CTP, and UTP, 2 µM RNA templates, along with 5 µM of purified protein. The reaction was carried out by incubation at 30°C for 1 h and terminated by addition of EDTA. The products were separated by using denaturing polyacrylamide gel electrophoresis (17.5% acrylamide, 7 M urea in TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.0)). Dried gels were exposed to phosphorimaging screens and scanned with a PhosphorImager (Typhoon, GE Healthcare).

2.2.13 Enzyme kinetic and inhibition assay for HKU4 M^{pro}

The substrate Dabcyl-KTSAVLQ↓SGFRKME-(Edans)-amide (95% purity; Biosyntan GmbH, Berlin, Germany), was used in the fluorescence resonance energy transfer (FRET)-based cleavage assay. The dequenching of the Dabcyl fluorescence due to the cleavage of the substrate was monitored at 490 nm with excitation at 340 nm, using a Cary Eclipse fluorescence spectrophotometer. The experiments were performed in the reaction buffer consisting of 20 mM Tris-HCl pH 7.3, 100 mM NaCl, and 1 mM EDTA. The reaction was initiated by adding different final concentrations of the FRET peptide (10 - 60 µM) to a solution containing M^{pro} (final concentration 0.25 µM). The equilibrium constant K_m was derived by fitting the data to the Michaelis–Menten equation, $V = V_{max} \times [S] / (K_m + [S])$.

Being a Michael acceptor compound, SG85 forms a covalent bond with the Cys148-Sγ of the HKU4 M^{pro}. The binding of this compound is irreversible and it was treated as time-dependent inhibition. Time-dependent inhibition progress curves were fit to a first order exponential (Equation 1) to yield an observed first-order inhibition rate constant (k_{obs}).

$$P = (v_0 / k_{obs})(1 - \exp(-k_{obs} t)) + D \quad (1)$$

$$\frac{1}{k_{obs}} = \frac{1}{k_3} + \frac{K_i}{k_3} (1 + [S]/K_m) \frac{1}{[I]} \quad (2)$$

P is the product fluorescence, v_0 is the initial velocity, t is the time, D is a displacement term to account for the fact that the emission is nonzero at the start of data collection. Values of K_i and k_3 were calculated from plots of $1/k_{obs}$ versus $1/[I]$ according to Equation 2.

The FRET-based assay was also used for the determination of the inhibition constants of the irreversible inhibitor SG85. The K_i and k_3 values were obtained by addition of the enzyme (final concentration 0.25 μ M) to reaction buffer plus 10 μ M FRET substrate and SG85 (five different concentrations ranging from 2.5 μ M to 10 μ M). The program Origin (OriginLab) was used to analyze the data.

3. Results and Discussion

3.1 The FCoV Nsp7+8 complex

Results presented here were published in Y. Xiao *et al.* (Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity). *J. Virol.* 86, 4444-4454. (2012).

3.1.1 Protein production

The quality of FCoV proteins was analyzed by SDS-PAGE under denaturing conditions after size-exclusion chromatography (SEC). As indicated in Fig. 3.1, Nsp8 containing four additional residues (Gly-Pro-Leu-Gly) (GPLG-Nsp8) (Fig. 3.1 (a)) or hexahistidine tag (His₆-Nsp8) (Fig. 3.1 (b)) at the N-terminus exhibited an apparent molecular mass of about 21.5 kDa, while GPLG-Nsp7 (Fig. 3.1 (c)) showed a molecular mass of about 9 kDa. The SDS-PAGE analysis for the Nsp7+8 complex (GPLG-Nsp8 in complex with GPLG-Nsp7) is shown in Fig. 3.1 (d). The purity of all proteins was larger than 90%.

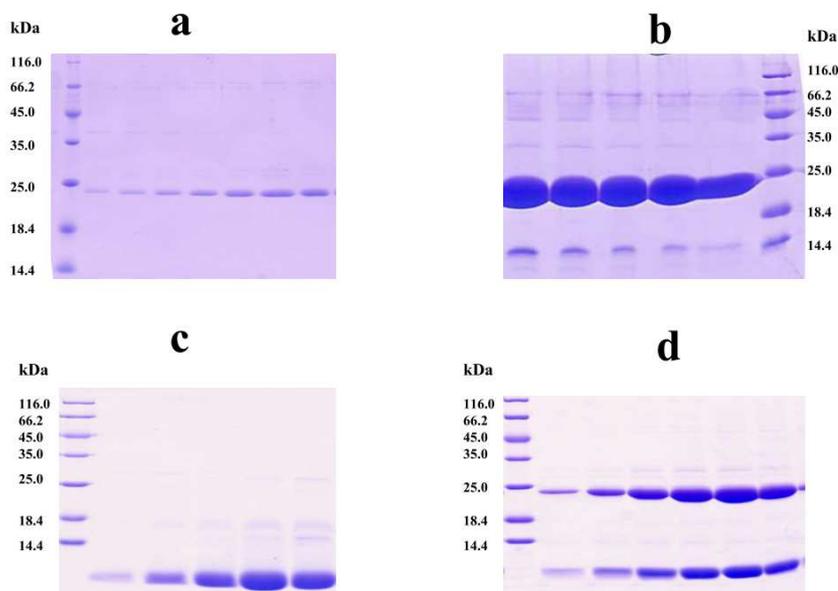


Fig. 3.1 Purification of FCoV Nsp7, Nsp8, and their complex

SDS-PAGE analyses of FCoV Nsp7, Nsp8, and their complex after size-exclusion chromatography (SEC). (a) GPLG-Nsp8. (b) His₆-Nsp8. (c) GPLG-Nsp7. (d) Nsp7+8 complex.

3.1.2 Crystallization of the FCoV Nsp7+8 complex

Tiny crystals appeared in a drop containing 2.4 M $(\text{NH}_4)_2\text{HPO}_4$, 0.1 M Tris-HCl pH 8.5 (Fig. 3.2 (a)). During crystal optimization, it was noticed that the protein was relatively unstable at room temperature, and the crystals always stopped growing after nucleation. Therefore, lower temperature and different additives were tried for optimization. Optimized crystals suitable for diffraction were grown in drops containing 1.5 μL of protein solution and 1.5 μL of reservoir solution (2.24 M $(\text{NH}_4)_2\text{HPO}_4$, 2 mM tris(2-carboxyethyl)phosphine (TCEP), 0.1 M Tris-HCl pH 8.7) at 4°C (Fig. 3.2 (b)).

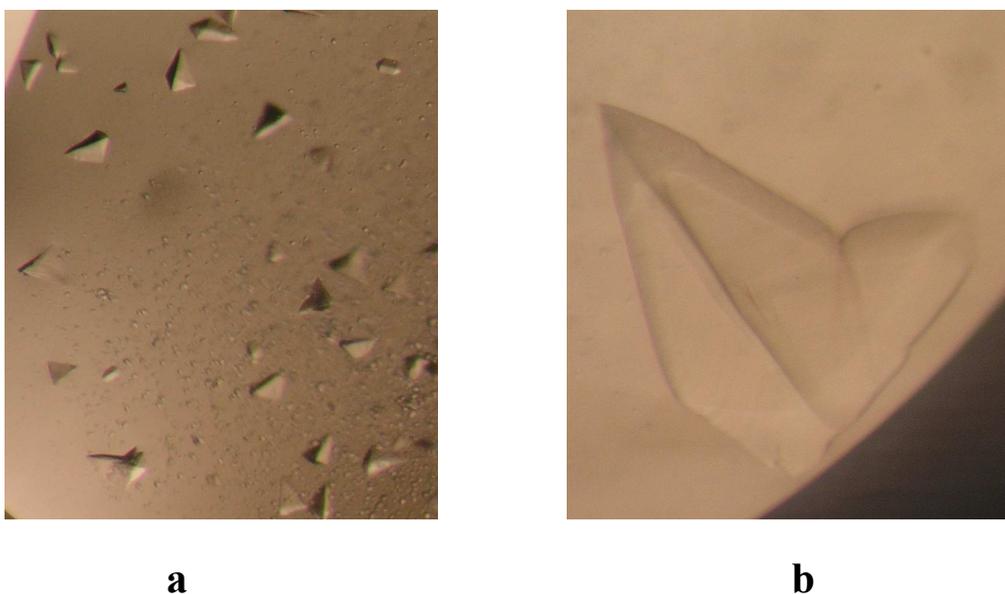


Fig. 3.2 Crystals of the FCoV Nsp7+8 complex

Three-dimensional crystals were obtained after three days. (a) Crystals obtained from initial screening, (b) crystal optimized for diffraction-data collection.

3.1.3 Structure solution and quality of the structural models

The crystal structure of the FCoV Nsp7+8 complex was determined by using the multi-wavelength anomalous diffraction (MAD) method. Phases were initially calculated to 3.5 Å, and then extended to 2.6 Å and improved *via* density modification. Statistics of the datasets are summarized in Table 3.1.

Table 3.1 Diffraction-data collection and refinement statistics

	Nsp7/SeMet-Nsp8 crystal 1	Nsp7/SeMet-Nsp8 crystal 2	
		Peak	Inflection
<i>Data collection statistics</i>			
X-ray source	BL14.2, BESSY, Berlin		
Wavelength (Å)	0.9809	0.97974	0.97990
Space group	C222 ₁		
Unit-cell dimensions (Å)	$a = 121.45, b = 160.30,$ $c = 102.66$	$a = 120.78, b = 161.43,$ $c = 103.25$	
Resolution range (Å)	33.45-2.60 (2.70-2.60)	33.55-3.24 (3.34-3.24)	
Number of unique reflections	31155	16370	16291
Completeness (%)	99.8 (100)	99.8 (99.9)	99.7 (99.3)
Mean I/sigma(I)	20.38 (3.33)	24.74 (4.92)	23.52 (3.62)
Multiplicity	7.23 (7.41)	14.48 (14.99)	14.60 (14.90)
R_{int} (%)	5.77 (62.0)	7.81 (48.4)	8.42 (84.1)
R_{sigma} (%)	3.1 (30.0)	2.84 (19.9)	2.81 (27.08)
Solvent content (%)	61.42		
V_m (Å ³ /Da)	3.19		
<i>Refinement statistics</i>			
Resolution range	33.45-2.60		
R_{work} (%)	20.51		
R_{free} (%)	22.74		
Modeled residues	A: -2-73, 79-192; B: 2-83; C: 0-81; D: -2-191; E: 2-82 F: 0-77		
Protein atoms	5498		
Water molecules	77		
r.m.s. deviation in bond lengths (Å)	0.010		
r.m.s. deviation in bond angles (°)	1.29		
Ramachandran plot (%)			
Most favored	97.3		
Additionally allowed	2.4		
Outliers	0.3		

Values in parentheses represent the highest resolution shell.

$R_{\text{int}} = \frac{\sum ||F_o|^2 - \langle |F_o|^2 \rangle|}{\sum |F_o|^2}$. $|F_o|^2$ are intensities of the reflections in the unmerged data. $\langle |F_o|^2 \rangle$ is the mean intensity for symmetry-equivalents, including Friedel pairs.

$R_{\text{sigma}} = \frac{\sum [\sigma(|F_o|^2)]}{\sum [|F_o|^2]}$. $|F_o|^2$ are intensities of the reflections in the unique data.

$R = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$. $|F_o|$ are amplitudes of the structure factors. R_{work} is the R value for reflections used in refinement, whereas R_{free} is the R value for 5% of the reflections that are selected in thin shells and are not included in the refinement.

3.1.4 Overall architecture of the Nsp7+8 complex

The 2.6-Å FCoV Nsp7+8 complex structure reveals a heterotrimer consisting of two copies of Nsp7 and one copy of Nsp8 (Fig. 3.3 (a)). The shape of the complex resembles a pistol (Nsp8) equipped with a magazine (2 x Nsp7). In the asymmetric unit, two copies of this heterotrimer dock onto each other through interaction between the two Nsp8 molecules, forming a dimer of the 2:1 Nsp7+8 trimer, i.e., a heterohexamer (Fig. 3.3 (b)).

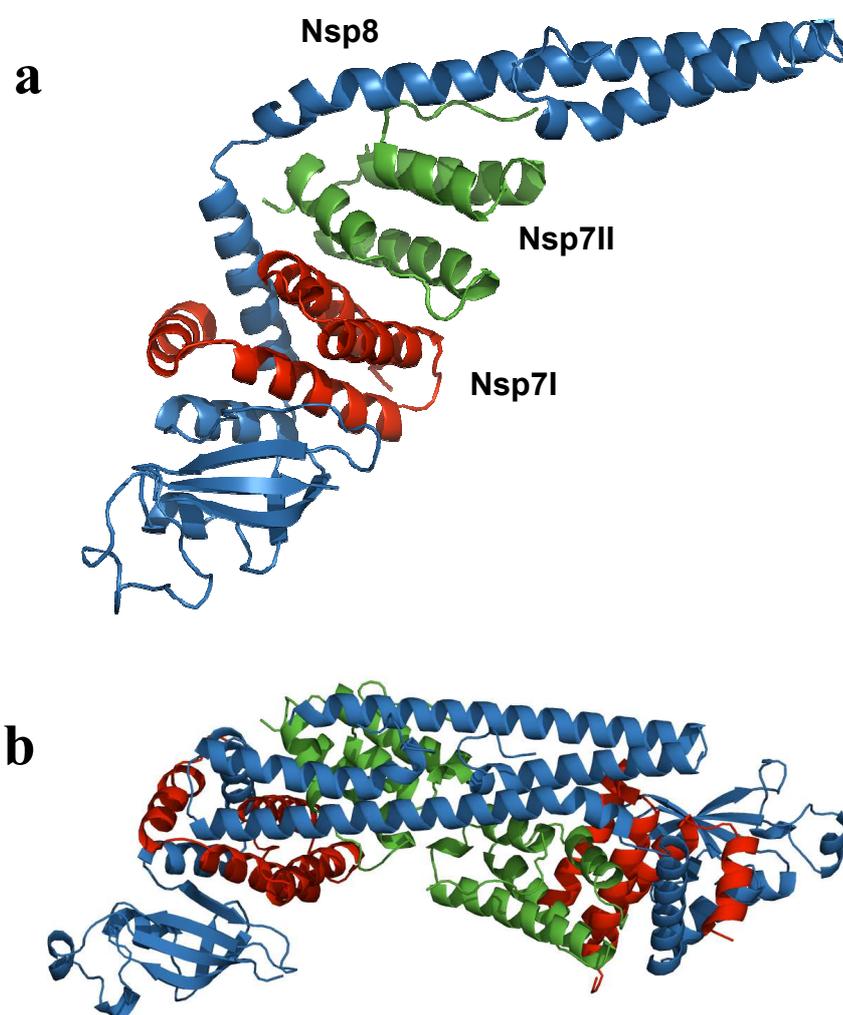


Fig. 3.3 Three-dimensional structure of the FCoV Nsp7+8 complex

(a-b) Overall structure of the Nsp7+8 heterotrimeric (a) and heterohexameric (b) complex: Nsp8, Nsp7I, and Nsp7II are coloured in sky blue, red, and green, respectively.

3.1.4.1 Structure of Nsp7 in the complex

Two different conformations of the Nsp7 monomer are observed within the heterotrimer (Fig. 3.4 (a) and (b)). Both of the two Nsp7 molecules are composed of four α -helices forming an antiparallel helix bundle. Helix H2 (residues 26-41) and helix H3 (residues 45-62) as well as the loops preceding and following them are structurally conserved between the two molecules, with an RMSD (root-mean-square difference) of 2.0 Å (for the 44 C α atoms of segment 21 - 64). In contrast, helix H1 (residues 3-20 for Nsp7I, residues 11-20 for Nsp7II) and helix H4 (residues 65-77 for Nsp7I, residues 69-78 for Nsp7II) are quite different in length (H1 and H4) and in orientation (H4). As seen in Fig. 3.4 (c) and Fig. 3.7, helix H1 of Nsp7I comprises 18 residues while helix H1 of Nsp7II contains only 10 residues. Besides, in Nsp7I, helix H4 is far apart from helix H1 (distance between Val12 (C γ 1) and Leu71 (C δ 2): 12.31 Å), but in Nsp7II, these two helices are much closer to one another (the distance between the two atoms indicated above is only 4.33 Å). The reason for this conformational difference is the insertion of Nsp8 helix N α 3 in between Nsp7 H1 and H4 in molecule Nsp7I but not in Nsp7II (see below). This difference between the two Nsp7 conformations is found identically in both copies of the Nsp7+8 heterotrimer in the asymmetric unit. The RMSD values between the two Nsp7I (for 78 C α atoms) and the two Nsp7II (for 81 C α atoms) are 0.37 Å and 0.34 Å, respectively.

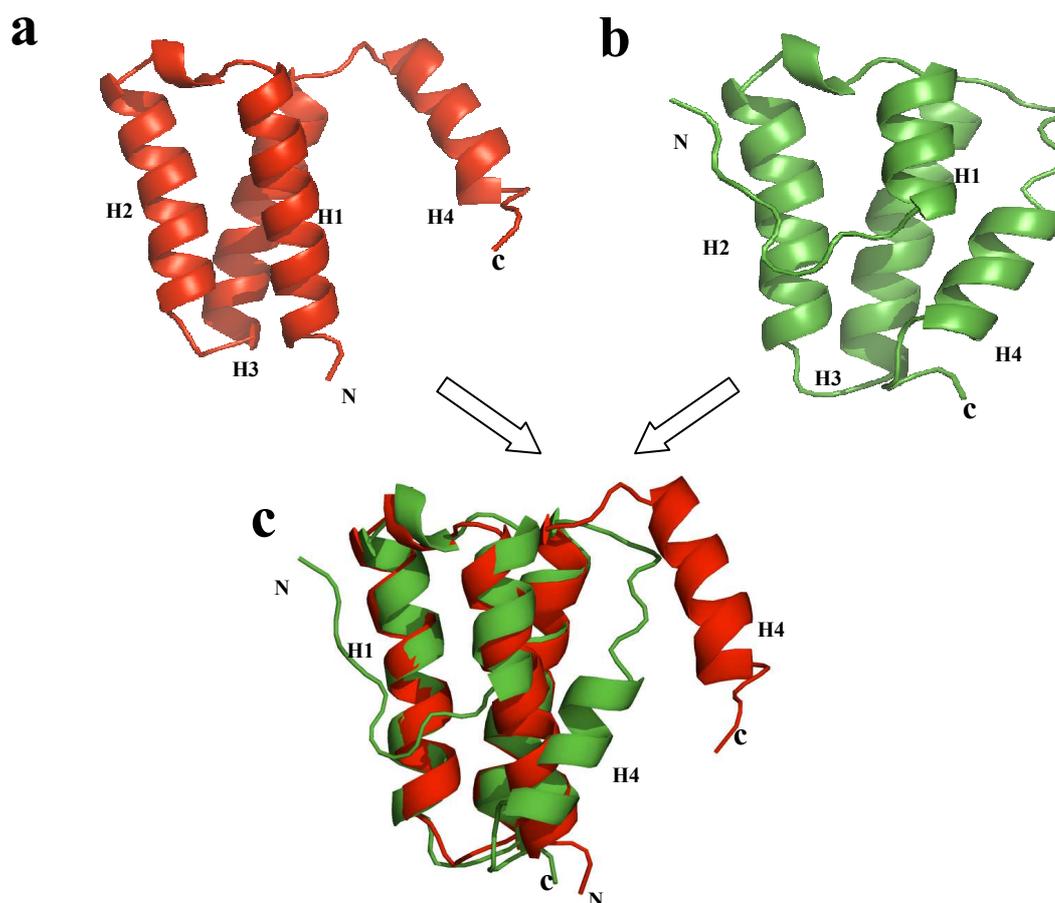


Fig. 3.4 Three-dimensional structure of the FCoV Nsp7I and Nsp7II and their structural comparison

Cartoon illustration of FCoV Nsp7I (a) and Nsp7II (b); (C) Superimposition of FCoV Nsp7I (red) with FCoV Nsp7II (green), with RMSD values (for 44 Ca atoms) of 2.0 Å.

Interestingly, when comparing the structures for Nsp7 within the heterotrimeric complex with the three structures of SARS-CoV Nsp7 with coordinates available in the PDB (2AHM, 2KYS, 1YSY), I noticed that the structure of FCoV Nsp7I is almost the same as that of SARS-CoV Nsp7 in complex with Nsp8 (Zhai *et al.*, 2005) (Fig. 3.5 (a)). The RMSD for 74 Ca atoms between the two molecules is 0.97 Å. On the other hand, FCoV Nsp7II is structurally similar to the isolated SARS-CoV Nsp7 in solution, as determined by NMR spectroscopy at pH 6.5 (Johnson *et al.*, 2010) (Fig. 3.5 (b)), with an RMSD (for 73 Ca atoms) of 2.38 Å. In contrast, the RMSD for 73 Ca atoms between FCoV Nsp7II and that of SARS-CoV determined by NMR

spectroscopy at pH 7.5 (Peti *et al.*, 2005) is 3.9 Å.

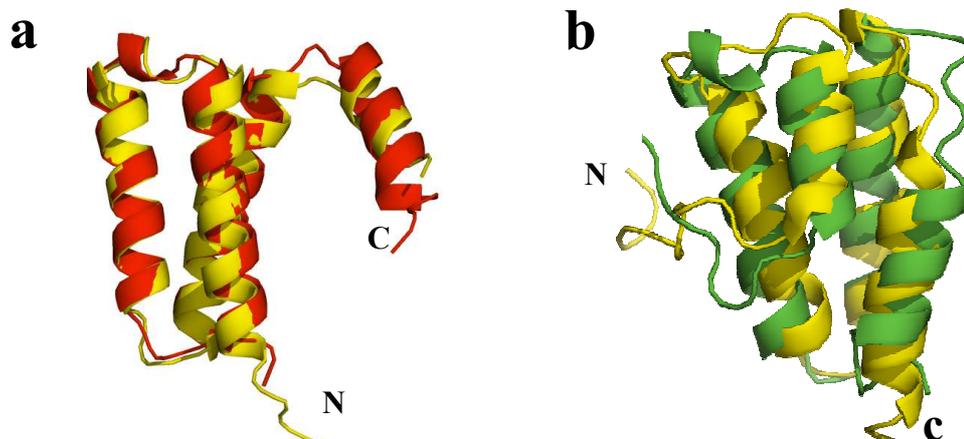


Fig. 3.5 Structural comparison of FCoV Nsp7I and Nsp7II with the corresponding proteins of SARS-CoV

(a) Superimposition of FCoV Nsp7I (red) with SARS-CoV Nsp7 (yellow) in complex with Nsp8, the RMSD (for 74 C α positions) is 0.97 Å. (b) Superimposition of FCoV Nsp7II (green) with the SARS-CoV Nsp7 (yellow) solution structure at pH 6.5; the RMSD (for 73 C α positions) is 2.38 Å.

3.1.4.2 Structure of Nsp8 in the complex

SARS-CoV Nsp8 possesses two different conformations in the hexadecameric complex with Nsp7 (Zhai *et al.*, 2005). In the FCoV Nsp7+8 complex, only one conformation of Nsp8 is observed in the two crystallographically independent copies of the heterotrimer (Fig. 3.6 (a)). Similar to the "golf club" structure of SARS-CoV Nsp8 (Zhai *et al.*, 2005), the Nsp8 of FCoV has a pistol-like structure composed of an N-terminal shaft domain and a C-terminal 'grip' domain (residues 5-112 and 113-191, respectively). The N-terminal shaft domain contains four helices (N α 1, residues 5-28; N α 2, residues 32-76; N α 3, residues 80-98; and N α 4, residues 101-112). Another three α -helices (C α 1, 134-140; C α 2, 168-172; C α 3, 174-179) pack against a five-stranded antiparallel β -sheet (β 1, residues 115-118; β 2, residues 127-132; β 3, residues 146-149; β 4, residues 152-160; and β 5, residues 184-190) form the C-terminal 'grip' domain, which has an α/β fold.

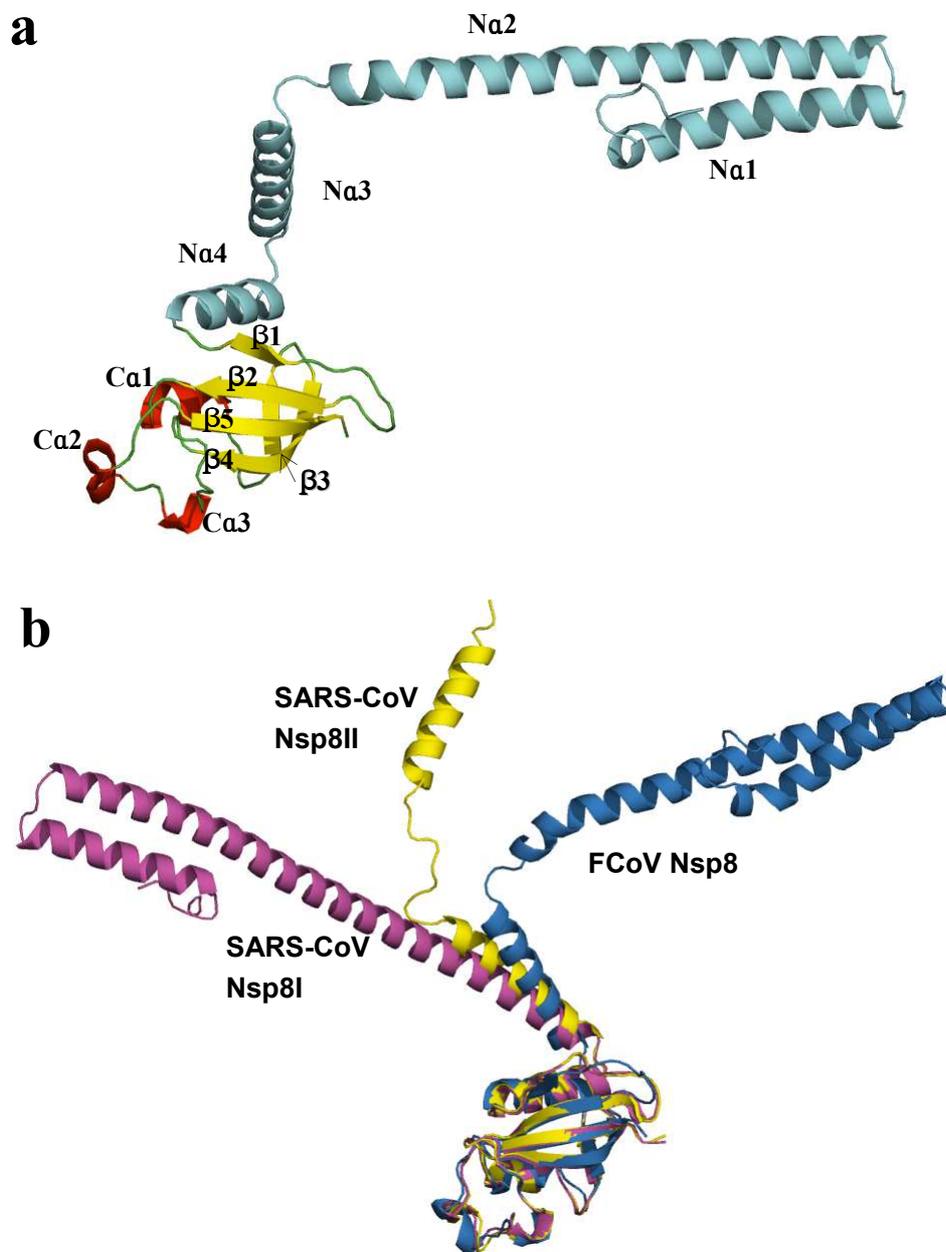


Fig. 3.6 Three-dimensional structure of the FCoV Nsp8 and structural comparison with the corresponding proteins of SARS-CoV

(a) Cartoon illustration of FCoV Nsp8. (b) C-terminal domain superimposition of FCoV Nsp8 (sky blue) with SARS-CoV Nsp8I (magenta) and Nsp8II (yellow), with RMSD values (for 79 C α atoms, residues 113-191) of 1.9 Å and 1.1 Å, respectively.

Superimposition of the FCoV Nsp8 C-terminal 'grip' domain onto the same regions of SARS-CoV Nsp8I and Nsp8II indicates that the C-terminal 'grip' domains are quite conserved (Fig. 3.6 (b)), with RMSD values (for 79 C α atoms) of 1.9 Å and 1.1 Å,

respectively. But the Nsp8 N-terminal shaft domains of the three molecules (SARS-CoV Nsp8I and II, and FCoV Nsp8) display totally different orientations (Fig. 3.6 (b)). Similar to the situation seen for half of the Nsp8 molecules (Nsp8II or "kinked golf club") in the hexadecameric SARS-CoV Nsp7+8 complex (Zhai *et al.*, 2005), the Nsp8 molecules in both copies of our heterotrimeric FCoV Nsp7+8 complex possess a bend (displaying poorly defined electron density) between helices N α 2 and N α 3, hinting at the flexibility of this region. It is likely that the bend provides rotational freedom for the N-terminal shaft domain to swing around the C-terminal 'grip' domain, thereby resulting in the change of the orientation of the N-terminal shaft domain.

3.1.4.3 Interactions between components of the heterotrimer

In the FCoV Nsp7+8 heterotrimeric complex, Nsp8, Nsp7I, and Nsp7II interact predominantly *via* hydrophobic interactions, which are augmented by a few intermolecular hydrogen bonds and salt bridges. The interaction interfaces between Nsp8 and Nsp7I, Nsp8 and Nsp7II, and Nsp7I and Nsp7II feature a total buried surface area of 3,500 Å², 1,250 Å², and 1,310 Å², respectively. Most of the residues involved in the interactions are quite conserved among coronaviruses (Fig. 3.7).

The interaction between Nsp8 and Nsp7I involves two parts. The first interface is located at N α 3 of the Nsp8 N-terminal shaft domain. Residues of Nsp7I helices H1 (Met6, Val12, Leu16) and H4 (Leu71, Tyr75, Phe76) form a hydrophobic core with residues of Nsp8 helix N α 3 (Met87, Leu91, Phe92, Met94, Leu98); a 3.16-Å hydrogen bond is observed between Nsp7I Lys2 N ζ and Nsp8 Leu98 O. The second hydrophobic interface is mainly composed of residues on H3 (Leu49, Leu53, Ile56, Ala57, Leu60) as well as H4 (Ile72) of Nsp7I and residues on N α 4 (Ile 106, Ile107) as well as β 1 (Leu115, Phe116) of Nsp8. The side-chain of Nsp8 Arg111 (N ϵ) forms hydrogen bonds and salt bridges with the side-chains of both Nsp7I Glu73 (O ϵ 2, 3.14 Å) and Glu77 (O ϵ 1, 3.25 Å). The two interfaces are also observed in the hexadecameric complex of SARS-CoV Nsp7+8 (Zhai *et al.*, 2005) and its truncated

heterodimeric form (Li *et al.*, 2010).

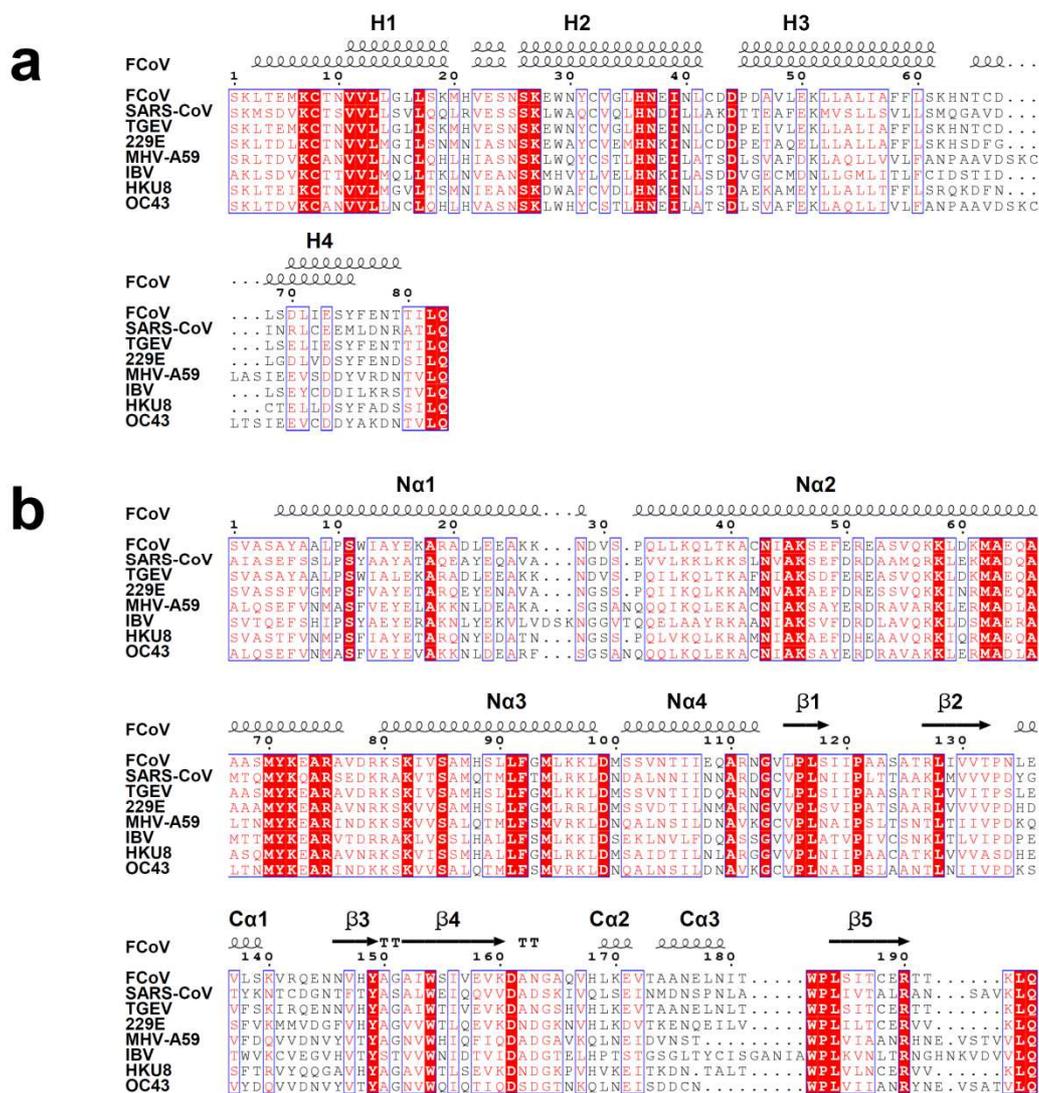


Fig. 3.7 Sequence alignment of FCoV Nsp7 (a) and Nsp8 (b) with homologues from other coronaviruses

The alignments were achieved by using ClustalW2 and the colored figures were generated by ESPrnt2.2. Residues labeled in red are conserved to more than 70% of the sequences and residues boxed in red are completely identical. Secondary structures of FCoV Nsp7 (above, Nsp7II; below, Nsp7I) and Nsp8 are indicated on top of the alignments. All the sequences aligned here are from the National Center for Biotechnology Information (NCBI) database: Feline infectious peritonitis virus (FCoV), DQ010921.1; SARS-CoV, NP_828865 (Nsp7) and NP_828866 (Nsp8); Transmissible Gastroenteritis coronavirus (TGEV), NP_840005.1 (Nsp7) and NP_840006.1 (Nsp8); Human coronavirus 229E (229E), NP_835348.1 (Nsp7) and NP_835349.1 (Nsp8); Mouse Hepatitis virus strain A59 (MHV-A59), NP_740612.1 (Nsp7) and NP_740613.1 (Nsp8); Infectious Bronchitis Virus (IBV), NP_740625.1 (Nsp7) and NP_740626.1 (Nsp7); Bat coronavirus HKU8, YP_001718610.1; Human coronavirus OC43 (OC43), NP_937947

The hydrophobic interface between Nsp7II and Nsp8 is mainly formed between helix H1 of Nsp7II (Val11, Leu14) and helix N α 2 (Leu59, Met62, Ala66) of the Nsp8 N-terminal shaft domain, indicating that the binding of Nsp7II has influence on the orientation of this domain. Besides, Phe76 on H4 of Nsp7II and Ile83, Met87 on N α 3 of Nsp8 also contribute to the hydrophobic interactions between these two polypeptides.

The Nsp7I–Nsp7II interface is formed mostly by hydrophobic interactions between residues on helices H1, H2 of Nsp7I (Cys8, Val11, Val12, Leu14, Leu41) and residues on helices H3, H4 of Nsp7II (Leu49, Leu53, Ile56, Leu60, Ile72). In addition, Ser18 (O γ), Trp29 (N ϵ 1), Asn37 (N δ 2) of Nsp7I form hydrogen bonds with Glu73 (O ϵ 1, 2.46 Å), Asp67 (O δ 1, 2.70 Å), and Cys66 (O, 2.82 Å) of Nsp7II, respectively.

The two copies of the Nsp7+8 heterotrimer present in the asymmetric unit of the crystal form a heterohexamer through interaction between the N-terminal shaft domains of the Nsp8 molecules. To investigate whether the heterohexamer is formed by crystal packing or does also exist in solution, a series of biochemical and biophysical experiments were performed.

3.1.5 The solution state of Nsp7, Nsp8, and the Nsp7+8 complex.

In addition to the question, which quarternary structure of FCoV Nsp7+8 exists in solution (heterotrimer or heterohexamer), the multimeric states of isolated coronavirus Nsp7 and Nsp8 are also poorly described and remain controversial (Johnson *et al.*, 2010; Sutton *et al.*, 2004; Zhai *et al.*, 2005). To study the quarternary structures of Nsp7, Nsp8, and the Nsp7+8 complex in solution, a chemical cross-linking analysis was performed first.

3.1.5.1 Glutaraldehyde cross-linking

After treatment with 0.01% glutaraldehyde, both isolated Nsp7 and Nsp8 displayed cross-linked homodimers (18 kDa for Nsp7, 43 kDa for Nsp8) in addition to their

monomers (Fig. 3.8 (a) and (b)). When the Nsp7+8 complex was cross-linked by 0.05% glutaraldehyde, the following species were detected: Nsp7 monomer (9 kDa), Nsp7 dimer (18 kDa), Nsp8 monomer (21.5 kDa), Nsp7/Nsp8 heterodimer (30.5 kDa), and the Nsp7+8 2:1 heterotrimer (39.5 kDa). Higher-order multimers were not detected (Fig. 3.8 (c)).

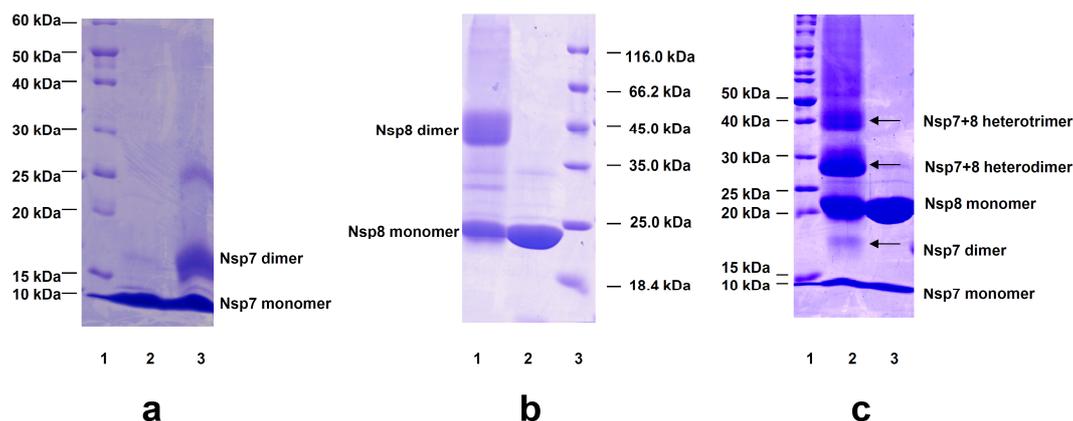


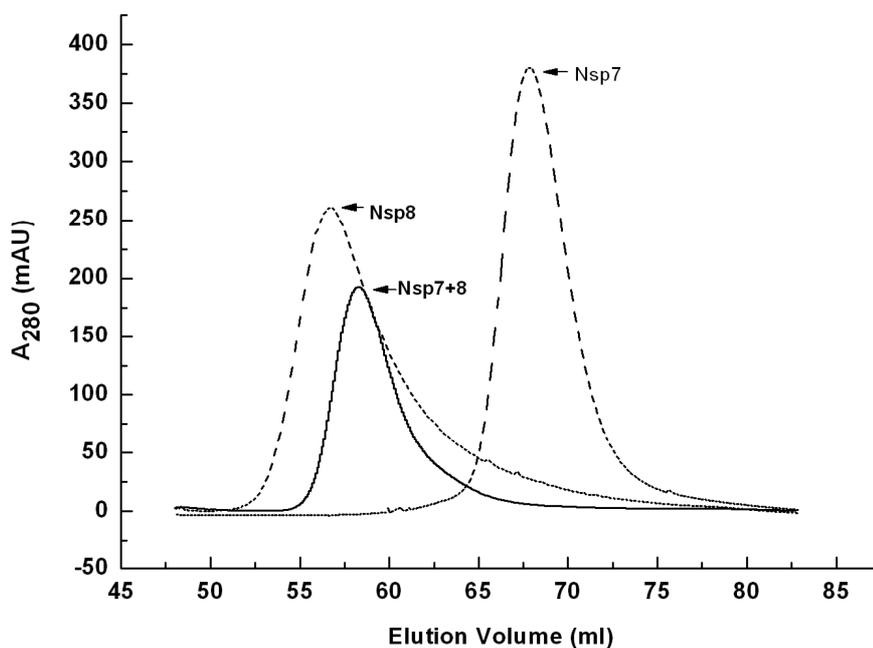
Fig. 3.8 SDS-PAGE analyses of chemically cross-linked Nsp7, Nsp8, and the Nsp7+8 complex

(a) Nsp7: lane 1, MW marker proteins; lane 2, untreated Nsp7; lane 3, Nsp7 cross-linked by 0.01% glutaraldehyde; the monomer (9 kDa) and the dimer (18 kDa) are detected; the weak band around 25 kDa is possibly due to a non-specifically cross-linked Nsp7 trimer. (b) Nsp8: lane 1, Nsp8 cross-linked by 0.01% glutaraldehyde; the monomer (21.5 kDa) and the dimer (43 kDa) are detected; lane 2, untreated Nsp8; lane 3, MW marker proteins. (c) Nsp7+8 complex: lane 1, MW marker proteins; lane 2, Nsp7+8 complex cross-linked by 0.05% glutaraldehyde; lane 3, untreated Nsp7+8 complex. The Nsp7 monomer (9 kDa), Nsp7 dimer (18 kDa), Nsp8 monomer (21.5 kDa), heterodimer (30.5 kDa), and heterotrimer (39.5 kDa) are all observed.

3.1.5.2 Size-exclusion chromatography

Subsequently, size-exclusion chromatography (SEC) was used to determine the solution states of Nsp7, Nsp8, and the Nsp7+8 complex. As shown in Fig. 3.9, the Nsp7+8 complex eluted with a retention volume of 58.23 ml, representing an estimated MW of 43.5 kDa, which corresponds reasonably well to the MW indicated by cross-linking (39.5 kDa). Nsp7 and Nsp8 had a retention volume of 67.86 ml and 56.64 ml, respectively. The estimated MW (24.9 kDa for Nsp7, 47.7 kDa for Nsp8) calculated from the retention volumes also agrees well with the MW of their

cross-linked dimers.



	Nsp7	Nsp8	Nsp7+8 complex
V_e (ml)	67.86	56.64	58.23
K_{av}	0.347	0.208	0.228
Estimated M_r	24921	47676	43491

Fig. 3.9 Size-exclusion chromatography (SEC) of Nsp7, Nsp8, and the Nsp7+8 complex

Estimation of protein solution state based on SEC data is shown in the table. V_e is the elution volume and K_{av} is the gel-phase distribution coefficient. Calibration of the column determined the relationship between K_{av} and MW as $K_{av} = -0.49265 \log MW + 2.51301$. Estimated M_r values from the SEC analyses also indicate that Nsp7 and Nsp8 alone are homodimers in solution, but that their complex is a 2:1 heterotrimer.

3.1.5.3 Small-angle X-ray scattering

In order to investigate the quaternary structures of the FCoV Nsp7+8 complex and of isolated Nsp7 further, small-angle X-ray scattering (SAXS) using synchrotron radiation was applied in collaboration with Dr. Dmitri Svergun, EMBL. The linear Guinier plot of the SAXS profile for the Nsp7+8 complex yielded a radius of gyration (R_g) of $30.0 \pm 0.5 \text{ \AA}$, which agrees well with the R_g of 30.3 \AA derived from the atomic structure of the heterotrimer. In contrast, the calculated R_g for the heterohexamer (dimer of heterotrimers) found in the asymmetric unit of the crystal is 35.0 \AA . The

maximum particle size D_{\max} derived from the experimental data is $105 \pm 5 \text{ \AA}$, which agrees well with the value of 115 \AA for the heterotrimer, whereas the D_{\max} for the asymmetric heterohexamer is about 130 \AA . As shown in Fig. 3.10, a direct comparison of the experimental scattering profile with that calculated from the atomic structure for the trimer yields a good fit with discrepancy $\chi=1.3$ in contrast to the much worse fit with $\chi=6.0$ for the hexamer (the discrepancy χ should be equal to 1.0 for a perfect fit).

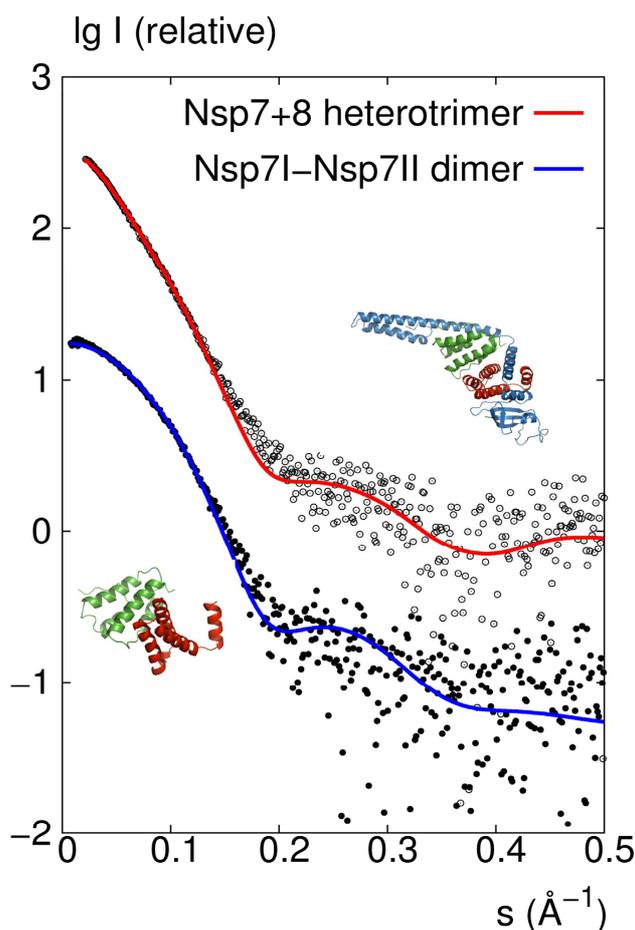


Fig. 3.10 SAXS investigations of the FCoV Nsp7+8 complex and Nsp7 in solution
Comparison of the experimental SAXS profile (empty circles) with a CRY SOL-calculated scattering curve (red) of the heterotrimer model consisting of Nsp7I, Nsp7II, and Nsp8 derived from the crystal structure and comparison of the SAXS data (filled circles) with the calculated scattering curve (blue) of the dimer model consisting of Nsp7I and Nsp7II as found in the crystal structure of the heterotrimer.

The molecular mass of Nsp7 estimated from the Guinier analysis of the SAXS data

was 19 kDa. This value is twice the molecular mass (9 kDa) calculated from the sequence of a monomer, suggesting that the protein forms dimers in solution. There is a good agreement between the experimental SAXS curve and the theoretical curve calculated by CRY SOL from a dimeric model containing the two conformations, Nsp7I and Nsp7II, observed in the crystal structure of the Nsp7+8 heterotrimer (discrepancy $\chi=1.3$; Fig. 3.10).

Taking into account all the evidence from chemical crosslinking, size-exclusion chromatography, and small-angle X-ray scattering, it is clear that the Nsp7+8 complex is heterotrimeric in solution, while the isolated Nsp7 and Nsp8 form homodimers.

3.1.6 RdRp activity of Nsp8

Imbert et al. (Imbert *et al.*, 2006) reported that the N-terminally His₆-tagged Nsp8 of SARS-CoV is a non-canonical RNA-dependent RNA polymerase capable of synthesizing short oligoribonucleotides (< 6 residues), which might be used as primers by the canonical RdRp, Nsp12. In the study, initially, a similar activity was detected for N-terminally His₆-tagged Nsp8 of FCoV (Fig. 3.11 (a)); this protein was able to mainly synthesize short RNA strands of approximately 6 nucleotides. The polymerase activity of FCoV Nsp8 also required the presence of magnesium ions, as activity was non-detectable without Mg²⁺. Surprisingly, however, upon increasing the Mg²⁺ concentration, limited amounts of longer RNA products could also be observed in case of N-terminally His₆-tagged Nsp8 (Fig. 3.11 (a)).

Subsequently, the RNA synthesis activity of the FCoV Nsp8 construct lacking the His₆-tag while carrying 4 additional residues (GPLG) at its N-terminus was tested; this protein was the one used for cocrystallization with Nsp7. Interestingly, compared to His₆-tagged Nsp8, GPLG-Nsp8 showed a much more pronounced ability to synthesize longer RNA, although it still could hardly synthesize RNA of template-length (Fig. 3.11 (b)).

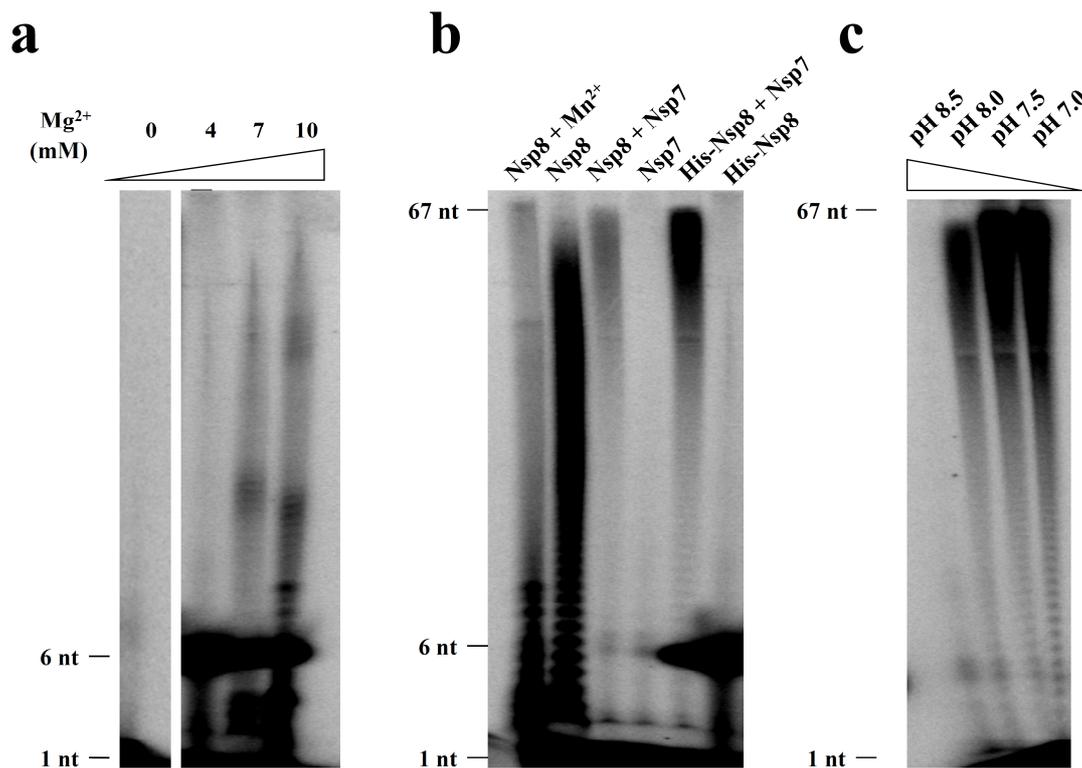


Fig. 3.11 *De-novo* RdRp activities of FCoV Nsp8-containing variants

(a) Mg²⁺ concentration influence on the RdRp processivity of His-Nsp8 (N-terminally His₆-tagged). (b) Side-by-side comparison of the polymerase activities of the Nsp8, Nsp8 + Nsp7 complex, His₆-Nsp8, and His₆-Nsp8 + Nsp7 complex; Nsp7 was used as negative control, Nsp8 with 4 mM Mn²⁺ was also tested. (c) The influence of pH on the Nsp7+8 complex polymerase activity. RNA oligonucleotide representing stem-loop 1 of the 3'-UTR of the FCoV strain FIPV WSU-79/1146 genome was used as template in all RdRp assays.

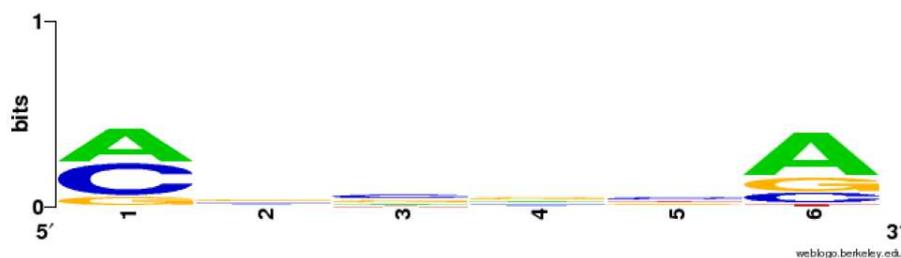
Next, these two Nsp8 variants in complex with Nsp7 were tested. Compared to the two Nsp8 variants alone, both of the two complexes exhibited much higher processivity and were able to synthesize template-length RNA (67 nt; Fig. 3.11 (b)). This observation provides clear evidence for an important role of Nsp7 in the RNA polymerase activity; apparently, it helps Nsp8 reduce abortion during replication and transcription and thus significantly increases its processivity. The influence of pH on the activity of the Nsp7+8 complex was also studied, optimum activity was found at neutral pH, while at pH 8.5, the polymerase activity was almost abolished (Fig. 3.11 (c)). Different from the report on SARS-CoV Nsp8 (Imbert *et al.*, 2006), manganese ions did not enhance the polymerase activity of the FCoV Nsp7+8 complex (data not

shown). On the contrary, the addition of 4 mM Mn^{2+} even had a negative effect on the polymerase activity (Fig. 3.11 (b)). Possibly, Mn^{2+} competes with Mg^{2+} for binding to the coordinating residues.

Oligoribonucleotides representing stem-loop 1 (SL1) and the complete 3'-UTR of SARS-CoV were also used as templates for testing the polymerase activities of the N-terminally His₆-tagged FCoV Nsp8 and the Nsp7+8 complex (data not shown). The identity between the SL1 sequences of SARS-CoV and FCoV is about 50%. Significant changes in the processivity compared to using FCoV SL1 as a template were not observed, implying that Nsp8 does not exhibit strong (if any) preference for template sequences.

3.1.7 Interaction screening between hexanucleotides and the FCoV Nsp7+8 complex

Interactions between the FCoV Nsp7+8 complex and hexadeoxynucleotides were probed using an array that contains all possible hexadeoxynucleotide sequences space (Mescalchin *et al.*, 2011). In this array, 4096 different hexadeoxynucleotides are attached to the chip surface via a non-nucleic acid linker linked to the 3' terminus of the DNA. In the case of the FCoV Nsp7+8 complex, array-binding studies revealed a number of binding oligonucleotides which showed a relatively strong signal on the array (Fig. 3.12). Statistical analysis of these 60 oligonucleotide sequences indicates that at the 5' end of the hexanucleotide, deoxyadenosine and deoxycytidine are preferred, while at the 3' end, deoxyadenosine and deoxyguanosine are preferred. In the positive-strand 3'-UTR sequence of FCoV, sequence homology corresponding to these hexamer sequences can be found. Remarkably, an octamer sequence at the 5' of the poly(A) tail (CACAAAAA) was identified, implying that the FCoV Nsp7+8 complex might interact with this region with the highest binding affinity.



AACACA	ATCTAA	CGGGGG
ACCACA	CACTCA	GGCGGG
CAAAAA	AACATC	GGGGAG
CACAAA	CCCAAC	GGGGCG
CACACA	CAACAC	GGGGGG
ACACTA	CGCCAC	CGGGGT
AACCTA	CACGCC	TGGGGG
AATCCA	ATCGTC	CTAACA
ACCCTA	CCCGGC	CTAACA
ATCCAA	CCTTAC	CTCTTA
ACTGCA	CCTTTC	CCTACA
AGAGCA	AACATG	CTAACA
AGCGCA	ACTCCG	CTCACA
AGGGCA	CCCCGG	CGCATA
AGGGGA	CCGCGG	CGGGCA
ACATCA	AGGGGG	CGTGCA
ACCTCA	AGGGCG	CGGGTA
AAATTA	ATAACT	GGGGGA
AATTTA	AGGGGT	GGGGCA
ACATTA	GGGCGG	GGGGTA

5'GGCAACCCGAUGUCUAAAACUUGUCUUUCCGAGGAAUUACGGGUCAUCGCG
 CUGCCUACUCUUGUACAGAAUGGUAAGCACGUGUAAUAGGAGGUACAAGCA
ACCCUAUUGCAUAAUJAGGAAGUUUAGAUUUUGAUUUUGGCAAUGCUAGAUUUAGU
AAUUUAGAGAAGUUUAAAGAUCGCUAUGACGAGCCAACAAUGGAAGAGCUAA
 CGUCUGGAUCUAGUGAUUGUUUAAAAUGUAAAAUUGUUUGAAAAUUUCCUU
 UUGAUAGUGAU**CACAAAAA**AAAAAAAAAAAAAAAAAAAAA-3'

Fig. 3.12 Sequences of array-bound hexadeoxynucleotides with strongest binding to the FCoV Nsp7+8 complex

The 60 of a total of 4096 hexamers (1.5%) that showed the highest signal on the array, are listed here. On top of the listed sequences is the consensus sequence. Hexamer sequences corresponding to sequences in the positive-strand 3'-UTR of FCoV are colored in red.

3.2 The HCoV-229E Nsp7+8 complex

3.2.1 Protein production

The quality of HCoV-229E proteins was analyzed by SDS-PAGE under denaturing conditions after size-exclusion chromatography (SEC). As seen from Fig. 3.13, the apparent molecular mass of the Nsp7-10 polyprotein is around 57.0 kDa (Fig. 3.13 (a)); while the Nsp7-8 oligoprotein exhibited an apparent molecular mass of about 31.0 kDa (Fig. 3.13 (b)). After cleavage of Nsp7-8 oligoprotein by SARS-CoV M^{pro}, Nsp7 and Nsp8 showed a molecular mass of about 9.0 kDa and 22.0 kDa, respectively (Fig. 3.13 (c)). Nsp7 is also purified separately (Fig. 3.13 (d)), while GST-tagged Nsp8 itself is not soluble. The purity of all proteins was higher than 90%.

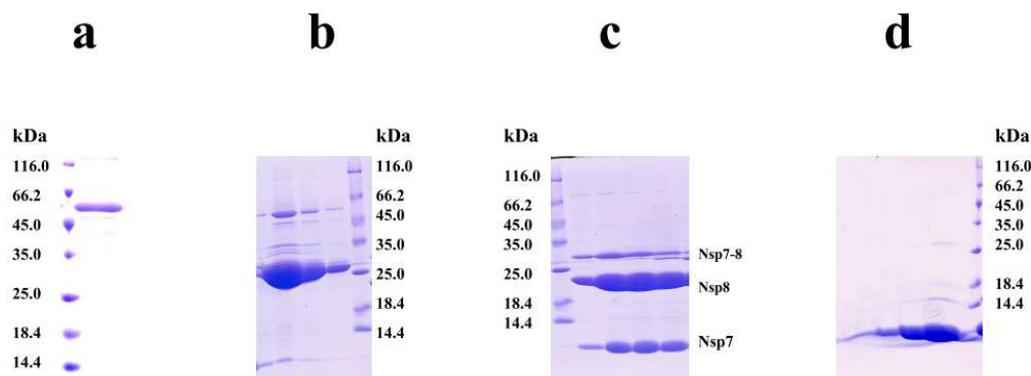


Fig. 3.13 Purification of the HCoV-229E Nsp7-8 oligoprotein and the Nsp7+8 complex.

SDS-PAGE analyses of the HCoV-229E Nsp7-10 polyprotein (a), the Nsp7-8 oligoprotein (b), the Nsp7+8 complex (c), and Nsp7 (d).

3.2.2 Crystallization of the HCoV-229E Nsp7+8 complex

Enormous effort was spent on optimization of the HCoV-229E Nsp7+8 complex crystals for good diffraction. Crystals were first obtained, but could never be repeated in a condition containing 0.2 M calcium chloride, 28% v/v polyethylene glycol 400, 0.1 M HEPES sodium pH 7.5. Crystals were only successfully repeated after replacing HEPES sodium by Tris-HCl, indicating that ionic strength might be crucial for nucleation. The major problem during crystallization trials was controlling crystal nucleation. As seen in Fig. 3.14 (a), normally thousands of small crystals were formed in the drop. This problem was solved by macroseeding in drops containing 2 μ L of protein solution and 2 μ L of precipitant solution (0.2 M calcium chloride, 21% v/v

polyethylene glycol 600, 0.1 M Tris-HCl pH 7.8) at 20°C, 1.4 M sodium chloride was used as reservoir solution. Since big crystals tend to stick tightly to the sitting-drop plates, hanging-drop plates were used instead for the final optimization (Fig. 3.14 (b)).

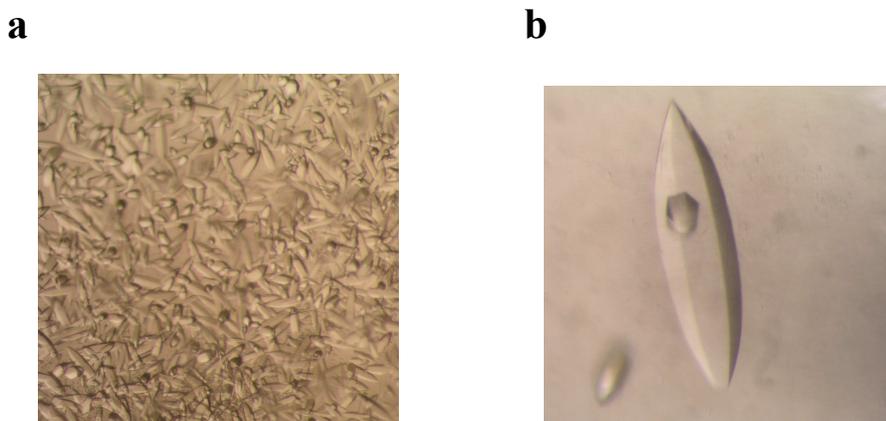


Fig. 3.14 Crystals of the HCoV-229E Nsp7+8 complex

Three-dimensional crystals were obtained after one week. (a) Crystals obtained from initial screening, (b) crystal optimized after several trials.

3.2.3 Structure solution and quality of the structural models

The crystal structure of the HCoV-229E Nsp7+8 complex was determined by molecular replacement. The structure of the SARS-CoV Nsp7+8 complex, truncated to one copy of Nsp7 and one copy of the Nsp8 C-terminal domain (aa 76-191), was used as the initial search model. The crystal of HCoV-229E Nsp7+8 complex displayed space group $P6_522$ with a large unit cell ($a = b = 151.18 \text{ \AA}$, $c = 225.36 \text{ \AA}$). Assuming that the crystal has a solvent content around 50%, five copies of the Nsp7+8 complex would be expected per asymmetric unit according to Matthews coefficient. However, at the beginning, only three copies of the Nsp7+8 complex could be found. At a later stage of refinement, a fourth copy was defined by fixing the three copies of the Nsp7+8 complex, and performing molecular replacement using a refined HCoV-229E Nsp7+8 complex harbouring one copy of Nsp7 and one copy of the Nsp8 C-terminal domain (aa 126-191) as the search model. Finally, a fifth copy of the Nsp7+8 complex was defined by superimposing a refined HCoV-229E Nsp7+8 heterotetramer complex onto the fourth copy of the Nsp7+8 complex. Statistics of the datasets are summarized in Table 3.2.

Table 3.2 Diffraction-data collection and refinement statistics

HCoV-229E Nsp7+8 complex	
Data collection statistics	
X-ray source	BL14.2, BESSY, Berlin
Wavelength (Å)	0.9184
Space group	P6 ₅ 22
Unit-cell dimensions (Å)	$a = b = 151.18, c = 225.36$
Resolution range (Å)	56.60-2.80 (2.95-2.80)
Number of unique reflections	38146 (5451)
Completeness (%)	100 (100)
Mean I/sigma(I)	19.8 (3.0)
Multiplicity	17.9 (17.8)
$R_{merge}(\%)^a$	10.5 (109.0)
$R_{measure}(\%)^b$	10.8 (112.1)
$R_{pim}(\%)^b$	2.5 (26.2)
Solvent content (%)	55.58
V_m (Å ³ /Da)	2.77
Wilson B-value (Å ²)	85.80
Refinement statistics	
Resolution range (Å)	26.74-2.81
$R_{work} / R_{free} (\%)^c$	24.12/26.13
Modeled residues	A: 1-77; B: 50-194; C: 1-77; D: 7-193; E: 1-76; F: 44-191; G: 1-64; H: 71-169, 186-194; I: 1-75; J: 78-153
Protein atoms	7240
r.m.s. d in bond lengths (Å)	0.010
r.m.s. d in bond angles (°)	1.25
Mean B-value (Å ²)	94.46
Mean B-value of each chain (Å ²)	A: 66.36; B: 80.68; C: 76.47; D: 73.44; E: 74.24 F: 87.99; G: 128.43; H: 138.73; I: 127.76; J: 133.69
Ramachandran plot (%)	
Mostly favored	96.2
Additional allowed	3.0
Outliers	0.8

Values in parentheses represent the highest resolution shell.

^a $R_{merge} = \sum_{hkl} \sum_i |I(hkl)_i - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I(hkl)_i$ where $I(hkl)$ is the intensity of reflection hkl and its symmetry-equivalents and $\langle I(hkl) \rangle$ is the average intensity over all equivalent reflections.

^b $R_{measure}$: multiplicity-weighted R_{merge} ; R_{pim} : precision-indicating R_{merge} .

^c $R = \sum ||Fo| - |Fc|| / \sum |Fo|$. $|Fo|$ and $|Fc|$ are amplitudes of the observed and calculated structure factors, respectively. R_{work} is the R value for reflections used in refinement, whereas R_{free} is the R value for 5% of the reflections which are selected randomly and are not included in the refinement.

3.2.4 Overall architecture of the asymmetric unit

In total, five copies of the Nsp7+8 heterodimer (Nsp7: chains A, C, E, G, and I; Nsp8: chains B, D, F, H, and J) were observed in the asymmetric unit of the crystal. Two copies of the Nsp7+8 heterodimer within one asymmetric unit (chains A - D) dimerize with each other through swapping of the Nsp8 C-terminal globular domain (Fig. 3.15 (a)). Another Nsp7+8 heterodimer (copy 3; chains E and F) dimerizes with its symmetry-related copy, also via swapping of the Nsp8 C-terminal globular domain (Fig. 3.15 (b)). The third copy of the Nsp7+8 heterotetramer (chains G - J) is also domain-swapped (Fig. 3.15 (c)). Compared with the other two heterotetramers, the electron density for the third Nsp7+8 heterotetramer is much weaker and the B factors are significantly higher, implying high mobility of this complex molecule. In addition, an area of discontinuous positive Fo-Fc difference density with a maximum sigma value of 4.0 is observed, but I was unable to interpret it in terms of a molecular model. Most probably, this density has to be largely attributed to the flexible N-terminal long-helix domain of the fourth and fifth copies of Nsp8 (chain H and J). Thus, a model of the fourth and fifth copies of the Nsp7+8 heterodimer can be only partially built: residues 65-83 of chain G, residues 1-70 and 159-185 of chain H, residues 1-77 and 155-195 of chain J, are missing in the model. Molecules of the first three copies of the Nsp7+8 heterodimer will be mainly discussed below.

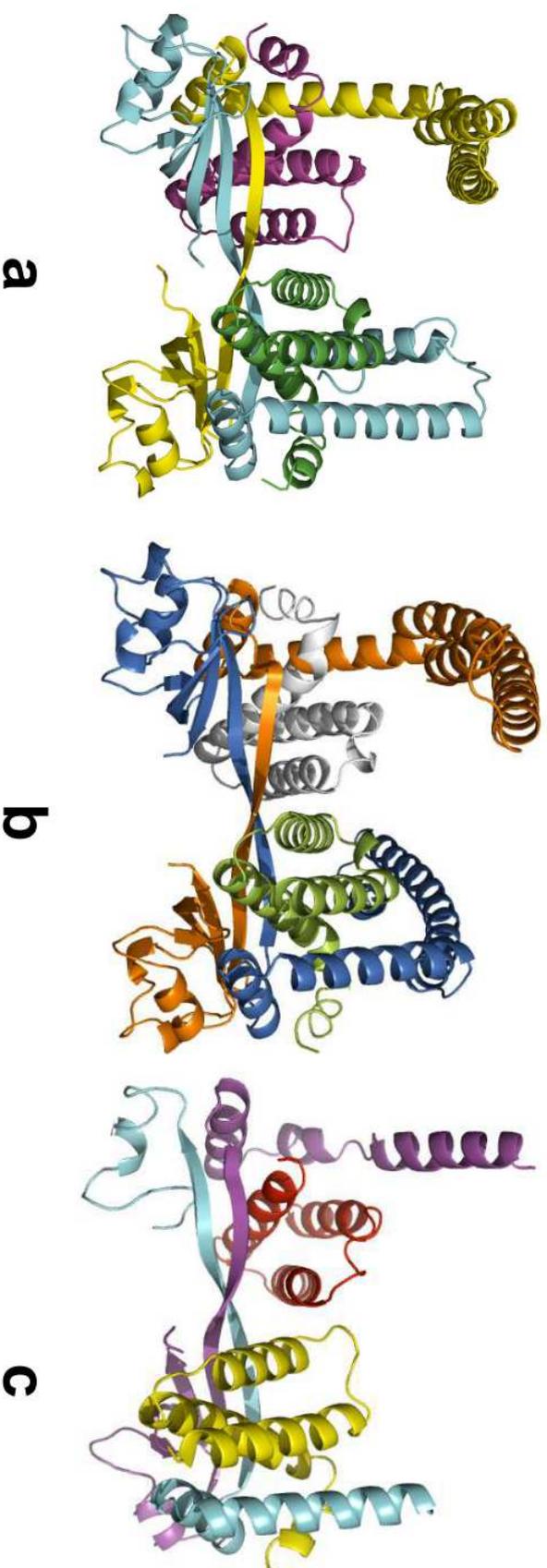


Fig. 3.15 Three-dimensional structure of the HCoV-229E Nsp7+8 complex

(a) Overall structure of the domain-swapped Nsp7+8 heterotetramer I. The two Nsp7s are colored yellow (chain B) and cyan (chain D), the two Nsp8s are colored magenta (chain A) and green (chain C). (b) Structure of the domain-swapped Nsp7+8 heterotetramer II. Nsp7 (chain E) and its symmetry-related chain are colored in white and limon, respectively; Nsp8 (chain F) and its symmetry-related chain are colored in orange and marine, respectively. (c) Structure of the domain-swapped heterotetramer III. The two Nsp7s are colored magenta (chain H) and cyan (chain J), the two Nsp8s are colored red (chain G) and yellow (chain I).

3.2.4.1 Structure of Nsp7 in the complex

Similar to Nsp7 observed in the SARS-CoV Nsp7+8 complex and Nsp7I of the FCoV Nsp7+8 complex, the structure of the HCoV-229E Nsp7 represents a four-helix bundle, which consists of α -helices H1, residues 2-20; H2, residues 26-41; H3, residues 45-62 and H4, residues 69-77 packed in an antiparallel coiled-coil arrangement with a compact hydrophobic core (H1, H2 and H3) in the center (Fig. 3.16 (a)). The three helices of the hydrophobic core stabilize each other through hydrophobic interactions, whereas helix H4 is apart from this core. Superimposition of the first three Nsp7 molecules in the asymmetric unit also indicates a conserved hydrophobic core and some variability in the structure of helix H4 (Fig. 3.16 (b)).

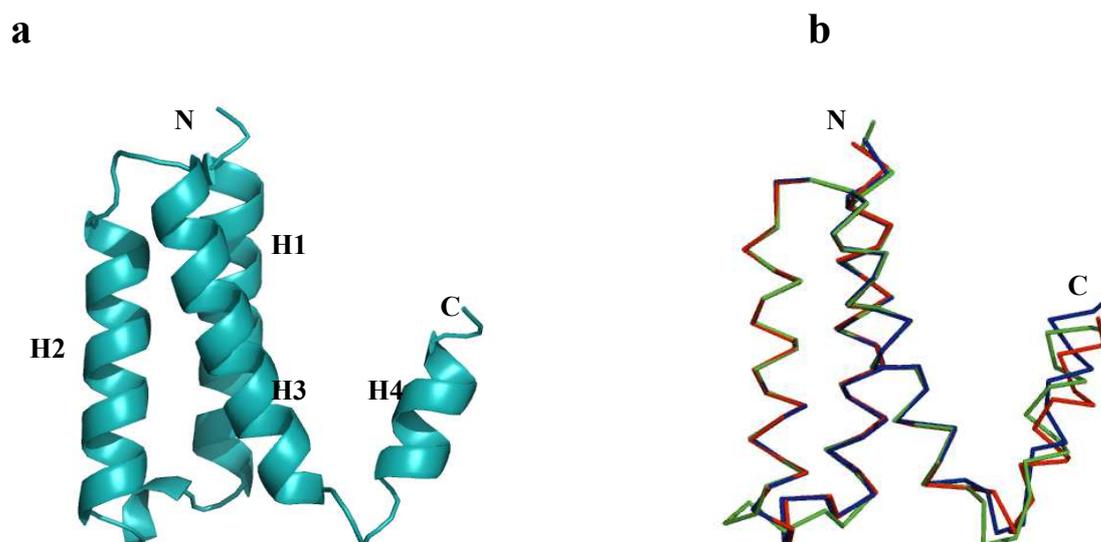


Fig. 3.16 Three-dimensional structure of the HCoV-229E Nsp7

(a) Cartoon illustration of HCoV-229E Nsp7 (chain A). (b) Superimposition of the first three Nsp7 molecules in the asymmetric unit. Chain A, chain C, and chain E are colored in red, green, and blue, respectively.

3.2.4.2 Structure of Nsp8 in the complex

The most striking feature of the HCoV-229E Nsp7+8 complex is that two Nsp8 molecules dimerize with each other through swapping of their C-terminal globular domains. The overall fold of the C-terminal globular domain (residues 126–191) of domain-swapped HCoV-229E Nsp8 (Fig. 3.17 (a)) is similar to that of

non-domain-swapped FCoV Nsp8 (Fig. 3.17 (b)), in which three α -helices are packed against a four-stranded antiparallel β -sheet and thus form an α/β fold.

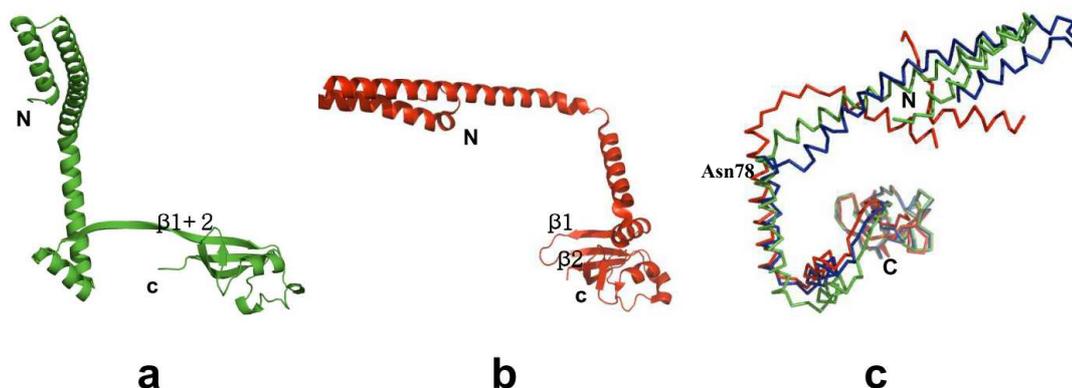


Fig. 3.17 Three-dimensional structure of the HCoV-229E Nsp8

(a) Cartoon illustration of the domain-swapped HCoV-229E Nsp8. (b) Cartoon illustration of the non-domain-swapped FCoV Nsp8. (c) Superimposition of first three Nsp8 molecules in the asymmetric unit. Chain B, chain D, and chain F are colored red, green, and blue, respectively.

The domain swapping occurs *via* the exchange of the second β strands, resulting in the bulk of the C-terminal domain (from residues 123 to the C-terminus) being swapped between the two subunits. Thus the first and the second β strands are integrated into one extended β strand. The newly formed β -strand links the two Nsp8 molecules into a 2-fold symmetric dimer and contributes to the stability of the dimer, with a total buried surface area of approximately 2500 \AA^2 . This dimer interface is stabilized by extensive hydrogen bonding between the main-chain oxygen and amide of the two extended β strands. The electron density corresponding to the hinge loop of the domain-swapped dimer is quite clear, as can be seen from the omit maps (Fig. 3.18).

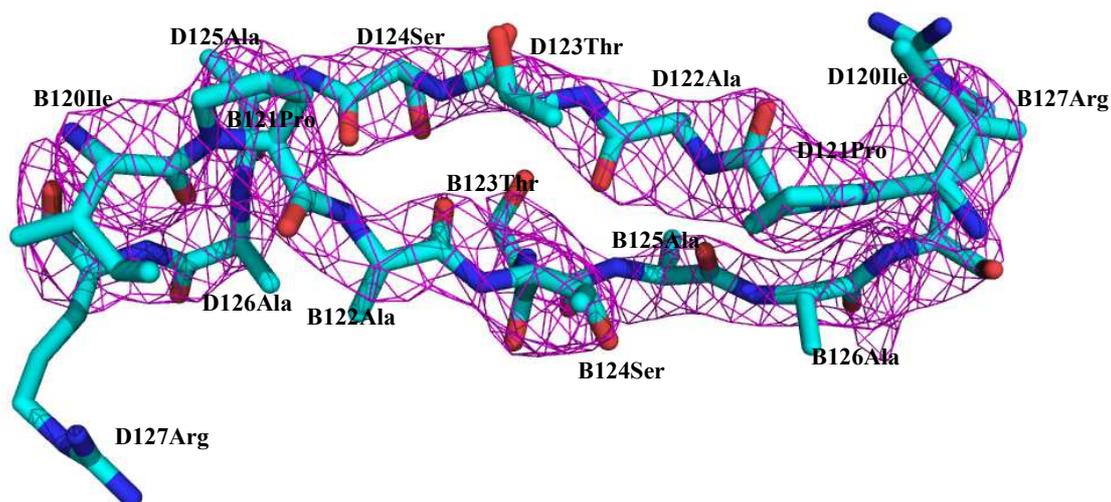


Fig. 3.18 Omit difference map ($F_o - F_c$) for the domain-swapped region 120IPATSAAR127

The map is contoured at a level of 3σ and colored magenta; residues (chain B and chain D) of the domain-swapped region 120IPATSAAR127 are shown as sticks and labeled.

Superimposition of the C-terminal globular domain of domain-swapped molecule B onto the same regions of molecules D and F indicates that the C-terminal globular domains are quite conserved (Fig. 3.17 (c)), with RMSD values (for 79 C α atoms) of 1.9 Å and 1.2 Å, respectively. However, similar to the SARS-CoV Nsp8 which displays two different conformations in the hexadecameric complex with Nsp7, and different from the FCoV Nsp7+8 complex in which only one conformation of Nsp8 is observed, the HCoV-229E Nsp8 N-terminal shaft domains of the first three molecules display some differences both in length and orientation (Fig. 3.17 (c)). The N-terminal first 49 residues of chain B could not be traced from the electron density map; while those of chain F were modeled as Ala due to high B factors and invisibility of side-chains. In contrast, chain D can be well modeled and only the first 6 residues are missing from the electron density map. Circular dichroism (CD) spectra measurements indicated that the Nsp7+8 complex contains more than 80% α -helix. Presumably, the Nsp8 N-terminal shaft domain is quite flexible and it can only be fixed when in contact with other molecules.

In addition, similar to the situation seen for the Nsp8II molecules ("kinked golf club")

in the hexadecameric SARS-CoV Nsp7+8 complex (Zhai *et al.*, 2005), and the Nsp8 molecule in the heterotrimeric FCoV Nsp7+8 complex, the HCoV-229E Nsp8 features a bend (electron density of this region is poorly defined) at a position around Asn78. As seen from Fig. 3.17 (c), the N-terminal shaft domains of the three domain-swapped Nsp8 molecules are swinging around Asn78, hinting at the flexibility of this region.

3.2.4.3 Interactions between Nsp7 and Nsp8

The HCoV-229E Nsp7 interacts with Nsp8 mainly through two hydrophobic interfaces, which are also observed in the SARS-CoV Nsp7+8 complex and the FCoV Nsp7+8 complex (Nsp7I+Nsp8). The first hydrophobic interface is situated at the C-terminal part of the Nsp8 N-terminal shaft domain. The hydrophobic core involves residues on the Nsp8 N-terminal shaft domain (Met87, Leu91, Met94, Leu95, Leu98, Met100) and residues of the Nsp7 helix H1 (Leu6, Thr9, Val12, and Ile16). The second interface is composed of residues of Nsp7 helix H3 (Leu52, Leu53, Leu56, Phe59, Leu60), Nsp7 helix H4 (Leu71, Val72, Tyr75) and residues on Nsp8 (Ile106, Leu107, Val115, and Pro116).

3.2.5 The solution state of the Nsp7-8 oligoprotein and the Nsp7+8 complex

3.2.5.1 Concentration dependence of the oligomerization states of the Nsp7-8 oligoprotein and the Nsp7+8 complex

Size-exclusion chromatography (SEC) was employed to determine the oligomerization states of the Nsp7-8 oligoprotein and the Nsp7+8 complex, in a concentration dependent manner. As shown in Fig. 3.19 (a), the retention volume (V_r) of the Nsp7-8 oligoprotein increased as the protein concentration decreased. At the concentration of 13 mg/ml, the Nsp7-8 oligoprotein eluted with a retention volume of 14.21 ml, corresponding to an estimated MW of 63.5 kDa, indicating that the Nsp7-8 oligoprotein exists mainly as dimer at this concentration. In contrast, at the concentrations of 2.6 mg/ml and 0.52 mg/ml, the retention volume decreased dramatically to 15.26 and 15.66 ml, corresponding to an estimated MW of 41.1 kDa and 34.8 kDa, respectively, clearly suggesting that the monomeric state is dominant at

these concentrations.

Similar size-exclusion chromatography profiles were also observed for the Nsp7+8 complex; the retention volume shifted as the protein concentration changed (Fig. 3.19 (b)). At the highest concentration of 13 mg/ml, the Nsp7+8 complex eluted as a single peak with a retention volume of 13.6 ml, representing an estimated MW of 81.85 kDa, implying that at this concentration, the complex exists almost exclusively as a heterotetramer. As the protein concentration got lower, the retention volumes shifted to 14.09 ml (5 mg/ml) and 14.34 ml (2.6 mg/ml), suggesting a mixture of heterotetramer and heterodimer, with the heterotetramer still being the predominant species. At the lowest concentration (0.52 mg/ml), both heterotetramer (V_r 13.9 ml, MW 72.2 kDa) and heterodimer (V_r 15.11 ml, MW 43.7 kDa) co-exist.

Interestingly, despite of the amino acid sequences being identical, the retention volumes of both the heterotetrameric and the heterodimeric Nsp7+8 complexes are lower than those of the corresponding dimeric and monomeric forms of the Nsp7-8 oligoproteins. Since the retention volume of a protein is directly influenced by its shape, this is indicative of a significant conformational difference between the Nsp7-8 oligoprotein and the Nsp7+8 complex. The estimated MWs of the Nsp7+8 complex heterotetramer (72.2 kDa) and heterodimer (43.7 kDa) are slightly larger than the theoretical MWs (heterotetramer 61 kDa, heterodimer 30.5 kDa). This can be explained on the basis of the structure by the elongated and flexible N-terminal helix of Nsp8. Taken together, these results show that the Nsp7-8 oligoprotein exists in a monomer-dimer equilibrium in solution, and the Nsp7+8 complex exists in a heterodimer-heterotetramer equilibrium in solution.

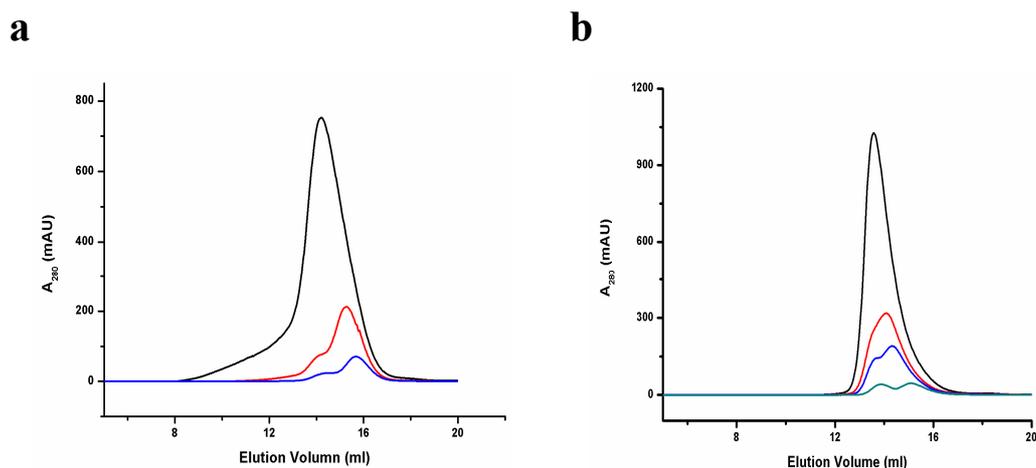


Fig. 3.19 Size-exclusion chromatography (SEC) of the Nsp7-8 oligoprotein and the Nsp7+8 complex

(a) Size-exclusion chromatography profile of the Nsp7-8 oligoprotein; the curves observed at three different protein concentrations (13 mg/ml, 2.6 mg/ml, and 0.52 mg/ml) are colored in black, red, and blue, respectively. (b) Size-exclusion chromatography profile of the Nsp7+8 complex; the curves observed at four different protein concentrations (13 mg/ml, 5 mg/ml, 2.6 mg/ml and 0.52 mg/ml) are colored in black, red, blue, and green, respectively.

3.2.5.2 The length of the hinge loop influences the oligomerization states of the Nsp7-8 oligoprotein and the Nsp7+8 complex

Protein domain swapping can be modulated by manipulating the length of the hinge loop in many cases, even though the mechanism is not clear and the influence appears to be unique to each protein (Gronenborn, 2009). In order to investigate the influence of hinge loop length on the oligomerization states of Nsp7-8 oligoprotein and Nsp7+8 complex, I either deleted the 122-124 Ala-Thr-Ser loop or inserted another Ala-Thr-Ser between 124Ser and 125Ala, at the hinge loop of Nsp8. The size-exclusion chromatography profiles of the Nsp7-8 oligoprotein with hinge-loop deletion or insertion, as well as of the Nsp7+8 complex with hinge-loop deletion or insertion, were compared to that of corresponding wild-type, at a protein concentration of 13 mg/ml. As can be seen from Fig. 3.20 (a), for the Nsp7-8 oligoprotein, the elution volumes of both hinge loop deletion (15.00 ml) and insertion (14.91 ml) were higher than that of the wild-type (14.21 ml). Similarly, the elution volumes of the Nsp7+8 complex hinge-loop deletion (14.06 ml) and insertion (14.17

ml) were also higher than that of the wild-type (13.60 ml, (Fig. 3.20 (b)). Thus, the results suggest that dimer formation of Nsp8 containing variants is influenced by the length of the hinge loop; lengthening or shortening the hinge loop can decrease the protein's propensity of dimerization caused by domain swapping.

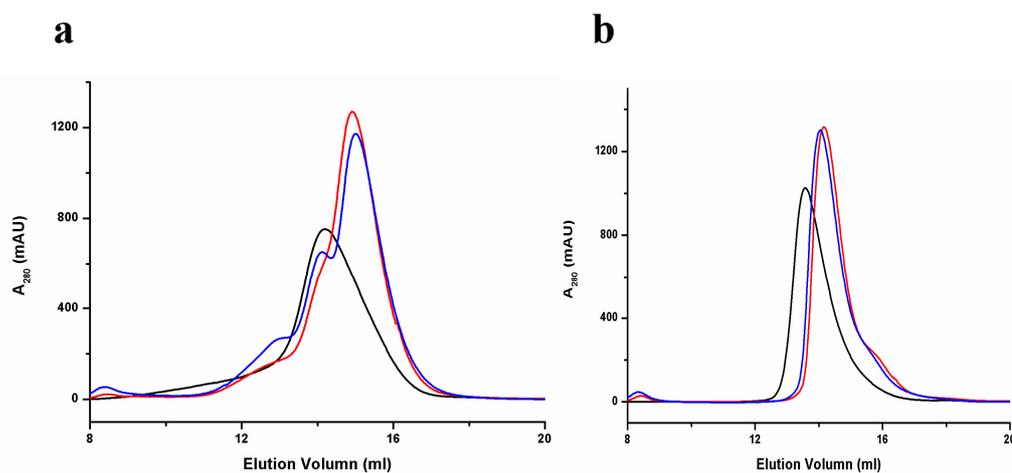


Fig. 3.20 Influence of hinge-loop length on the oligomerization states of the Nsp7-8 oligoprotein and the Nsp7+8 complex by size-exclusion chromatography (a) Size-exclusion chromatography profile of the Nsp7-8 oligoprotein wild-type (black), hinge-loop deletion (blue) and insertion (red). (b) Size-exclusion chromatography profile of the Nsp7+8 complex wild-type (black), hinge-loop deletion (blue) and insertion (red).

3.2.5.3 SDS-PAGE analyses of the glutaraldehyde-crosslinked and the disulfide-linked Nsp7+8 complex

Next, a chemical cross-linking analysis was carried out. After treatment with 0.1% glutaraldehyde, 5 μ g of protein was loaded into each lane for analysis. Bands corresponding to Nsp7 monomer (9 kDa), Nsp8 monomer (22 kDa), the Nsp7+8 heterodimer (31 kDa), and the Nsp8 homodimer (44 kDa) were detected clearly in the SDS-PAGE (Fig. 3.21 (a)). With increasing protein concentration, the bands corresponding to the Nsp7+8 heterodimer and the Nsp8 homodimer increased in intensity. Interestingly, the ratio between the Nsp8 homodimer and Nsp7+8 heterodimer also increased, implying that higher protein concentration increase in the amount of dimerized Nsp8, well in agreement with the conclusion that the oligomerization state of the Nsp7+8 complex is protein concentration-dependent.

In solution, intermolecular disulfide bond is unlikely to form if the cysteine is not located at or near the interaction interface. By investigating the structure of the domain-swapped Nsp7+8 complex, I found that the G113 and P133 residues on the two Nsp8 monomers in the heterotetrameric complex are very close across its dimer interface (the distance between the two C α atoms is 5.99 Å). This dimer interface exists only if the Nsp8 C-terminal domain is swapped (Fig. 3.21 (b)). Thus, in principle, a mixture of the G113C-mutated Nsp7+8 complex and the P133C-mutated Nsp7+8 complex would have the potential to form an intermolecular disulfide bond, if the domain swapping also occurs in solution.

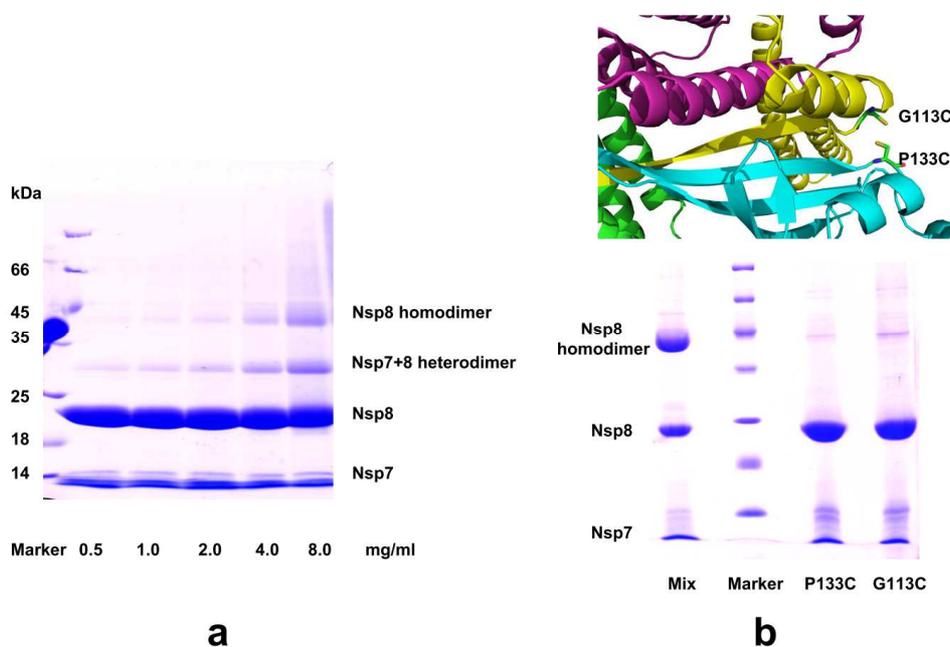


Fig. 3.21 SDS-PAGE analyses of the glutaraldehyde cross-linked, and the disulfide-bond linked Nsp7+8 complex.

(a) Glutaraldehyde cross-linking of the Nsp7+8 complex. Different protein concentrations are indicated below each lane. (b) Above: positions of G113C and P133C are indicated in the domain-swapped Nsp7+8 complex; the two Nsp8 molecules are colored in yellow and cyan respectively; below: comparison of the G113C-mutated Nsp7+8 complex, the P133C-mutated Nsp7+8 complex, and their mixture under non-reducing conditions in 15% SDS-PAGE.

Thus, the Nsp7+8 complex with the G113C or the P133C mutation was purified separately. The G113C- and P133C- mixed Nsp7+8 complex was prepared by mixing the same molar ratio of these two mutated complexes. Since observed from the

size-exclusion chromatography that the heterodimer-heterotetramer equilibrium of this complex is protein concentration-dependent, concentrating and diluting were performed several times to maximize the amount of G113C- and P133C- mixed Nsp7+8 complex. The disulfide-bond formation of the G113C-mutated Nsp7+8 complex, the P133C-mutated Nsp7+8 complex, and their mixture were evaluated under non-reducing conditions by SDS-PAGE (Fig. 3.21 (b)). In both the G113C-mutated Nsp7+8 complex and the P133C-mutated Nsp7+8 complex, Nsp8 migrated as a monomer of 21.5 kDa. However, in the mixture, Nsp8 formed a dimer band corresponding to a MW of about 43 kDa. This dimer is supposed to be disulfide-linked via the 113C and 133C residues of the two Nsp8 molecules. In the wild-type Nsp8, the only cysteine residue is located at position 188, deeply buried inside the C-terminal global-domain and far away from the two mutation sites. This result strongly suggests that the Nsp8 C-terminal global-domain swapping is not an artifact of crystal packing, but does exist in solution as well.

3.2.5.4 Dynamic Light-Scattering

To compare the oligomerization states of the FCoV and HCoV-229E Nsp7+8 complexes, the hydrodynamic radii of these two complexes were measured by Dynamic Light-Scattering in buffer containing 10 mM Tris-HCl pH 7.5 and 200 mM NaCl, at the concentration of 15 mg/ml. As evident from Fig. 3.22 (a), the Nsp7+8 complex of FCoV exhibited a quite monodisperse band with a hydrodynamic radius of about 3.46 ± 0.14 nm. In the case of the HCoV-229E Nsp7+8 complex (Fig. 3.22 (b)), the monodisperse hydrodynamic radius was centered at around 7.90 ± 1.39 nm. Comparison of the hydrodynamic radii of the two Nsp7+8 complexes indicates that the HCoV-229E Nsp7+8 complex is almost double the size of the FCoV Nsp7+8 complex. Again, the result presented here is consistent with what is observed in the crystal structures. The size of the heterotetrameric HCoV-229E Nsp7+8 complex is double the size of the heterotrimeric FCoV Nsp7+8 complex.

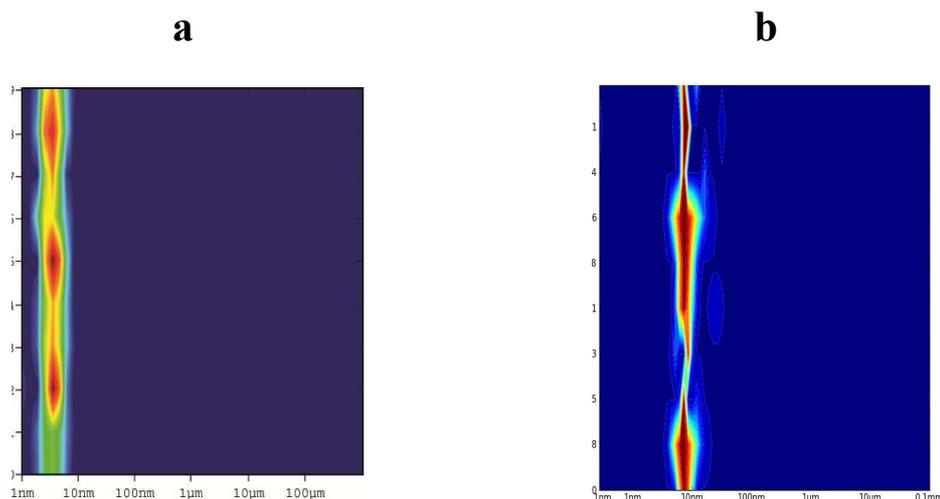


Fig. 3.22 Dynamic Light-Scattering of the FCoV (a) and the HCoV-229E (b) Nsp7+8 complex

The results were interpreted using the software provided by the manufacturer. Experimental errors were estimated as standard deviations calculated from 10 measurements for each sample. The hydrodynamic radius is depicted on the *x-axis*, the time on the *y-axis*.

3.2.6 RdRp activity of the HCoV-229E Nsp8-containing variants

In the FCoV study, I have found that N-terminally His₆-tagged Nsp8 of FCoV was able to mainly synthesize short RNA strands of approximately 6 nucleotides. A similar activity for the HCoV-229E Nsp7-10 polyprotein was also observed (Fig. 3.23 (a)). But different from the report on SARS-CoV Nsp8 (Imbert *et al.*, 2006), manganese ions did not enhance the polymerase activity of the HCoV-229E Nsp7-10 polyprotein (Fig. 3.23 (a)).

I also tested the RdRp activities of the Nsp7-8 oligoprotein and Nsp7+8 complex of HCoV 229E. These two variants exhibited very similar activities and were able to synthesize template-length RNA (67 nt; Fig. 3.23 (b)). This observation is not beyond my expectation. First, these two proteins were of the same composition and exhibited quite similar profiles in size-exclusion chromatography. Second, the Nsp7-8 oligoprotein and the Nsp7+8 complex of SARS-CoV showed comparable primer extension activities (te Velthuis *et al.*, 2012). In addition, I also tested the hinge-loop-inserted and -deleted Nsp7+8 complex. Surprisingly, in spite of similar

size-exclusion chromatography profiles, the RdRp activity of the hinge-loop-deleted Nsp7+8 complex was almost abolished, while that of the hinge-loop-inserted Nsp7+8 complex remained (Fig. 3.23 (c)). Even though the hinge-loop-inserted and -deleted Nsp7+8 complexes share very similar size-exclusion chromatography profiles, the result presented here implied that the hinge-loop (122-124 Ala-Thr-Ser) deletion may have destroyed its structure.

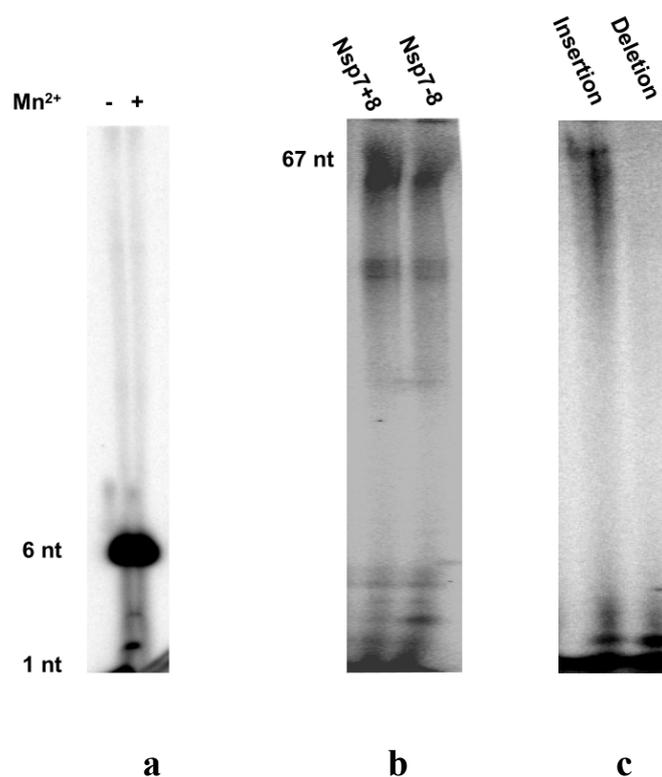


Fig. 3.23 *De-novo* RdRp activities of the HCoV-229E Nsp8-containing variants
 (a) Influence of Mn²⁺ on the RdRp processivity of the HCoV-229E Nsp7-10 polyprotein. (b) Side-by-side comparison of the polymerase activities of the HCoV-229E Nsp7-8 oligoprotein and Nsp7+8 complex. (c) Side-by-side comparison of the polymerase activities of the HCoV-229E Nsp7+8 complex hinge-loop insertion and deletion. An RNA oligonucleotide representing stem-loop 1 of the 3'-UTR of the FCoV strain FIPV WSU-79/1146 genome was used as template in all RdRp assays.

3.3 Discussion of the Nsp7+8 complex

The coronavirus replicase-transcriptase complex (RTC) is involved in different steps of viral RNA synthesis, but its detailed composition and structure is not clear, and how the replicase-transcriptase complex accomplishes these various activities is not understood in any detail. For instance, it is still unclear whether the same replicase-transcriptase complex is engaged in full-length and subgenomic mRNA synthesis or how the switch of RNA synthesis from negative to positive strand is regulated (Sawicki *et al.*, 2005). Biochemical studies have proved that many coronavirus replicase proteins are associated with enzymatic activities. Nsp8 of SARS-CoV has been identified as noncanonical primer-independent RNA polymerase (te Velthuis *et al.*, 2012; Imbert *et al.*, 2006). In addition, Nsp8 is able to interact with many replicase proteins such as Nsp2, Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp12, Nsp13, and Nsp14, suggesting it is a major component within the replicase-transcriptase complex (von Brunn *et al.*, 2007).

Despite high sequence similarity (higher than 40%), I have found that Nsp7+8 complexes of different coronaviruses display great structural diversity. Unlike the SARS-CoV (betacoronavirus) Nsp7+8 hexadecamer complex, in which eight Nsp8 interact with eight Nsp7 (Zhai *et al.*, 2005), the Nsp7+8 complex of FCoV (alphacoronavirus) is a heterotrimer formed by two Nsp7 molecules and one Nsp8. In my study, the Nsp7 and the Nsp8 of FCoV have been mixed at different ratios, but the resulting complex always eluted with a retention volume of around 58.2 ml from the size-exclusion column, corresponding to the heterotrimer. Chemical cross-linking and small-angle X-ray scattering (SAXS) also indicate that the 2:1 stoichiometry of Nsp7:Nsp8 observed in my crystal structure exists in solution as well. In addition, the HCoV-229E (alphacoronavirus) Nsp7+8 complex, which was produced by M^{pro} cleavage of Nsp7-8 oligoprotein, reveals an equilibrium between 1:1 heterodimer and 2:2 heterotetramer. In the 2:2 heterotetrameric complex, two Nsp8 copies form a tight dimer through domain swapping involving the C-terminal globular domains. Mixing the HCoV-229E Nsp7+8 complex with a two-fold excess of Nsp7 did not result in a

2:1 heterotrimer or a 4:2 heterohexamer. Even though these different Nsp7+8 complex structures are only observed from different coronaviruses up to now, I assume that they might exist among all the Nsp7+8 complexes, because most of the residues involved in the complex assembly are conserved. Thereby the structural and stoichiometrical diversity of the Nsp7+8 complex would endow it with multifunctional property in transcription and replication.

3.3.1 Domain swapping of the HCoV-229E Nsp7+8 complex

Domain swapping has been proposed as a possible mechanism for protein oligomerization, misfolding, and functional regulation. When domain swapping occurs, noncovalent interactions of the swapping region are interrupted and replaced by almost identical interactions with the other monomer to form a dimer or oligomer (Liu & Eisenberg, 2002). In the structure, the HCoV-229E Nsp7+8 complex reveals a 2:2 heterotetramer, in which two Nsp8 copies form a tight dimer through domain swapping involving the C-terminal globular domains. Even though the domain swapping does not provide a mechanism for the evolution of functional site(s), I assume it maybe a mechanism of functional regulation. As presented above, Nsp8 is a key component of the RTC and able to interact with many other replicase proteins. In spite of conserved secondary structure, the Nsp7+8 heterotetramer and the Nsp7+8 heterodimer are distinct in oligomerization state. Therefore, the interaction patterns of Nsp8 with other components of the RTC will change upon domain swapping; as a consequence, RTC formation will be modified or totally disrupted.

In domain-swapped structures, the closed monomer and domain-swapped oligomer share the same structures except the hinge loop. Therefore, between the closed monomer and the domain-swapped oligomer, the free energy difference is small but the transition energy barrier is high. However, this energy barrier can be reduced under certain circumstances, such as change of pH, change of temperature, presence of denaturants, and binding of a ligand (Liu & Eisenberg, 2002). Thus, the possibility that Nsp8 of other coronaviruses also posses the domain swapping ability cannot be

excluded, even though they have not been observed so far. Interestingly, the domain swapping of HCoV-229E Nsp7+8 complex is protein concentration-dependent in solution, where a dynamical heterodimer-heterotetramer equilibrium exists. This phenomenon has also been observed in many domain-swapped structures (Rousseau *et al.*, 2001; O'Neill *et al.*, 2001; Liu & Eisenberg, 2002). It has been proposed that under physiological conditions, macromolecular crowding increases local protein concentration and may facilitate 3D domain swapping *in vivo* (Minton 2001; Liu & Eisenberg, 2002). In coronaviruses, RNA synthesis takes place in double-membrane vesicles (DMVs), which provide a membrane-anchored structural framework and may increase the viral RTC component concentration (Knoops *et al.*, 2008). Therefore, a high concentration of Nsp8 in double-membrane vesicles may facilitate 3D domain swapping *in vivo*.

In many cases, modification (mutation, deletion, or insertion) at the hinge loop may have significant effect on domain swapping (Rousseau *et al.*, 2001, Picone *et al.*, 2005, Chen *et al.*, 1999, Bandukwala *et al.*, 2011). Either shortening or elongating the hinge loop of Nsp8 by three amino acids decrease its propensity of dimerization caused by domain swapping. However, these modifications cannot totally disrupt or promote Nsp8 domain swapping, indicating the hinge loop might not be the only determinants for domain swapping. It has been reported recently that the domain swapping of SARS-CoV M^{pro} C-terminal helix $\alpha 1$ is activated by its C-terminal helix $\alpha 5$, emphasizing the importance of non-swapped elements in mediating domain swapping (Kang *et al.*, 2012). It is still not clear how the 3D domain swapping of HCoV-229E Nsp8 is achieved, and further studies are needed to reveal Nsp8 domain swapping mechanism and roles in replicase-transcriptase complex modification.

3.3.2 Structure of the HCoV-229E Nsp7-8 oligoprotein

Rearrangement of the Nsp7 and Nsp8 coding sequences or disruption of the Nsp7-8 cleavage site in MHV-A59 were not permissive for virus replication (Deming *et al.*, 2007), implying the Nsp7-8 oligoprotein and Nsp7+8 complex are both important for

virus replication. Similar to the Nsp7+8 complex, the Nsp7-8 oligoprotein also exists in monomer-dimer equilibrium in solution. Additional Nsp7 cannot form a complex with the Nsp7-8 oligoprotein (data not shown), suggesting that Nsp8 within the Nsp7-8 oligoprotein might already interact with its Nsp7. In addition, it is unlikely that the Nsp7 of one oligoprotein can interact with the Nsp8 of another oligoprotein, otherwise higher-order oligomerization states of the Nsp7-8 oligoprotein would be observed. So most probably, the Nsp7 and Nsp8 within one oligoprotein interact with each other through intramolecular interaction. As indicated in “Results”, the retention volumes of both the heterotetrameric and heterodimeric Nsp7+8 complexes are lower than those of the corresponding dimeric and monomeric states of the Nsp7-8 oligoprotein, suggesting that the overall shape of Nsp7-8 oligoprotein is more compact than that of the Nsp7+8 complex. The SAXS measurements carried out by our collaborators (Falke, Ph.D. thesis, University of Hamburg, 2013) also suggest that both the Nsp7+8 complex and Nsp7-8 oligoprotein are dimerized through the C-terminal part of Nsp8. Interestingly, the overall shape of the Nsp7+8 complex is a bit longer than that of the Nsp7-8 oligoprotein, very likely due to the elongated Nsp8 N-terminal domain of the Nsp7+8 complex. Considering all the evidence presented above, I propose that in one Nsp7-8 oligoprotein, Nsp7 and Nsp8 interact with each other in an intramolecular arrangement. Therefore, the N-terminal long-helix domain of Nsp8 is bent over to its C-terminal domain. Once the linker is cleaved by the M^{pro}, the N-terminal long helix of Nsp8 is released and the Nsp7+8 complex would present a more elongated structure compared with the Nsp7-8 oligoprotein.

3.3.3 Function of the Nsp7+8 complex

I show here that like SARS-CoV, FCoV and HCoV-229E possess a primer-independent RNA polymerase, Nsp8, in addition to the canonical, primer-dependent RNA polymerase Nsp12. Since Nsp8 of SARS-CoV was reported to produce oligoribonucleotides with less than 6 residues, it was postulated that the protein acts as primase and the primers synthesized by Nsp8 would be subsequently utilized by the primer-dependent polymerase Nsp12 for further elongation (Imbert *et*

al., 2006). Distinct from previous observations with SARS-CoV, results presented here demonstrate that both the Nsp8 and the Nsp7+8 complex of FCoV, as well as the Nsp7-8 oligoprotein and Nsp7+8 complex of HCoV 229E are able to synthesize much longer transcripts (up to template-length). Interestingly, quite recently, te Velthuis *et al.* (2012) showed that Nsp8 of SARS-CoV with authentic N-terminus is capable of extending primed RNA templates, implying the possibility that Nsp8 and the Nsp7+8 complex of SARS-CoV may also synthesize longer products. I conclude that Nsp8 is not simply a primase which merely provides primers for Nsp12, but also executes its own polymerase function and associated roles in coronavirus replication and transcription. In addition, considering the fact that the FCoV Nsp7+8 complex exhibits much higher processivity than does Nsp8 alone in *de-novo* RNA synthesis, I consider Nsp7's role to be more than just mortar to stabilize the FCoV Nsp7+8 complex structure. I postulate that Nsp7 could enhance the RNA-binding affinity of the Nsp8 N-terminal shaft domain by stabilizing its orientation and folding, and thereby increasing the polymerase processivity of Nsp8.

In this study, I have also tested the RNA synthesis function of the precursor polyprotein Nsp7-10 of HCoV 229E. Sawicki *et al.* (2005) had proposed that the MHV-A59 precursor polyprotein Nsp4-10 might possess independent functions before being proteolytically processed. Deming *et al.* (2007) further supported this idea by demonstrating that rearrangement of the Nsp7 and Nsp8 coding sequences in MHV-A59 was not permissive for virus replication. Here I demonstrate that the purified Nsp7-10 polyprotein of HCoV 229E, a presumable replicase precursor, can function as RdRp *in vitro*, even though it is only able to produce short RNA oligonucleotides. In comparison, the HCoV-229E Nsp7-8 oligoprotein and Nsp7+8 complex, as well as the mature FCoV Nsp8 and its Nsp7+8 complex, exhibit much stronger polymerase processivity, suggesting that polyprotein processing and replication/transcription complex formation are crucial for the polymerase activity of Nsp8. In addition, the distinct polymerase processivity of different Nsp8 variants implies that at different stages of replication and transcription, different

Nsp8-containing variants may be required as *de-novo* RdRps. The proteolytic processing of the replicase polyproteins and the assembly of the replication complex might be critical for regulating RNA replication and transcription, for instance, switching from negative-strand RNA synthesis to positive-strand RNA synthesis (Brayton *et al.*, 1982). Nsp8 or Nsp8-containing polyprotein might act as primase at the initial step of coronavirus RNA synthesis, but after being cleaved and assembled into the Nsp7+8 complex, the latter might be responsible for further elongation of the primers.

As mentioned above, I noticed that FCoV GPLG-Nsp8 showed much higher polymerase processivity than the N-terminally His₆-tagged Nsp8, indicating that additional residues at the Nsp8 N-terminus might have significant influence on the elongation step during RNA synthesis. It might be argued that the increased processivity is due to the four extra residues (GPLG) remaining at the N-terminus of FCoV Nsp8 from the cloning procedure, although such a gain of function would be extremely unlikely for these four amino-acid residues devoid of functional groups in their side-chains. This has further been clarified by proving that the HCoV-229E Nsp7+8 complex with authentic N-terminus also possesses high polymerase processivity. In addition, the recent report by de Velthuis *et al.* (2012) on a primer-extension activity of SARS-CoV Nsp8 with authentic N-terminus supports my view that the primer-independent polymerase activities of FCoV Nsp8 and the Nsp7+8 complex are not due to a gain of function caused by these four extra residues. Rather, the HCoV-229E Nsp7-10 seems to cause a loss of function in terms of polymerase processivity. However, the HCoV-229E Nsp7-8 oligoprotein, which possesses many additional residues (Nsp7) is fully processive, implying that an authentic N-terminus might not be the only determinant for polymerase processivity. I do not have a good explanation for the apparent influence of residues at the N-terminus of Nsp8, as a structure of free Nsp8 is not available. More structural and functional studies should be performed to address this issue.

The fact that the Nsp7-10 polyprotein of HCoV 229E, which is very unlikely to form the same quaternary structure as the SARS-CoV, the HCoV-229E, or the FCoV Nsp7+8 complex, shows polymerase activity, suggests that the Nsp8 *de-novo* RNA synthesis activity probably does not depend on its multimeric state. This is supported by the observation of polymerase activity for all the Nsp8-containing variants discussed in this thesis (except hinge-loop deleted Nsp7-8 oligoprotein), which showed distinct multimeric states.

3.3.4 Active-site(s) within Nsp8

The crystal structure shows that the N-terminal shaft domain of FCoV Nsp8 is rich in positively charged amino-acid residues on one side (Fig. 3.24 (a)), which could be responsible for RNA binding. A similar charge distribution is also observed in the same region of HCoV-229E Nsp8 (chain D, Fig. 3.24 (b)). In the SARS-CoV Nsp7+8 complex (Zhai *et al.*, 2005), replacing Arg and Lys residues in this region by Ala significantly decreased its nucleic-acid-binding affinity. Previous studies in our group, using 1D-NMR and circular dichroism (CD) spectroscopy, showed that the N-terminal shaft domain of isolated Nsp8 of SARS-CoV is partially disordered in solution (Ponnusamy, 2010), in agreement with the conclusions of Rao and coworkers (Li *et al.*, 2010, Zhai *et al.*, 2005). In the crystal structure of the HCoV-229E Nsp7+8 complex, part of the Nsp8 N-terminal shaft domains cannot be traced from the electron density map. However, the same domain of FCoV Nsp8 is well-defined in the complex structure, with the exception of the bend between helices N α 2 and N α 3. Interestingly, the orientation of the FCoV Nsp8 shaft domain is distinct from that of its counterpart in SARS-CoV. The latter makes the same interactions with Nsp7 as FCoV Nsp8 does with Nsp7I, but the second Nsp7 molecule (with its distinct conformation) is missing in the hexadecameric SARS-CoV complex. It is quite possible that the interaction with two Nsp7 molecules, especially the binding of Nsp7II, not only improves folding of the Nsp8 N-terminal shaft domain, but also has an influence on its orientation.

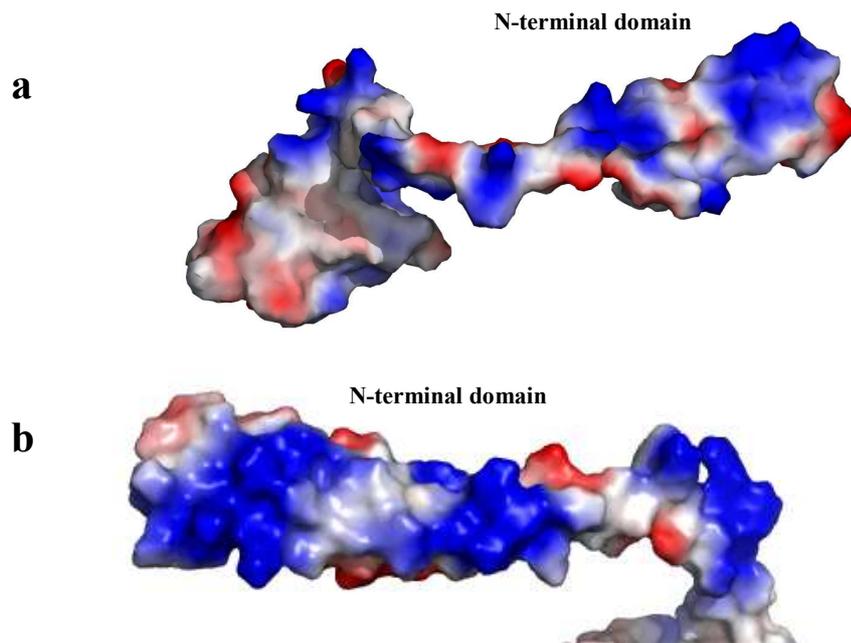


Fig. 3.24 Electrostatic-potential surface of FCoV Nsp8 (a) and HCoV-229E Nsp8 (b)

Red and blue surfaces represent negative and positive potentials ($-10 \text{ k}_B\text{T}$ to $+10 \text{ k}_B\text{T}$), respectively. The electrostatic surface potentials were calculated using the Adaptive Poisson-Boltzmann Solver (APBS) within the PyMOL APBS tools.

However, the molecular mechanism of the RNA polymerase activity of Nsp8 still cannot be derived from the available Nsp7+8 complex structures, as these do not allow the localization of the active site(s), in spite of a large body of mutational data (Imbert *et al.*, 2006, Zhai *et al.*, 2005, te Velhuis *et al.*, 2012). Canonical primases and RNA polymerases usually feature an (Asp/Glu)-X-(Asp/Glu) motif at their active site, which binds to Mg^{2+} or Mn^{2+} (Castro *et al.*, 2007). FCoV Nsp8 comprises a Glu-X-Glu sequence within its helix N α 3, while HCoV-229E possesses Asp-X-Glu at the same place. te Velhuis *et al.* (2012) showed that replacing the Asp residue of this motif by Ala in SARS-CoV Nsp8 abolished its primer-extension activity. But in all the three crystal structures, the two acidic side-chains are oriented towards different sides of the helix. However, I cannot exclude the possibility that the active conformation of Nsp8 could be quite different when in complex with RNA. Indeed, preliminary results from Fluorescence-Resonance-Energy-Transfer measurements of

our collaborators demonstrate that fluorophor-labeled Nsp8 of FCoV undergoes significant conformational changes upon addition of RNA (Seidel, unpublished data). Ralph Baric's group has shown that in MHV, Lys58Ala mutation of Nsp8 (corresponding to Lys58 in FCoV and HCoV-229E) is not viable for virus replication (unpublished data). Considering the fact that a hexahistidine tag at the N-terminal seems to cause a loss of function in terms of polymerase processivity, I postulate that the active-site(s) of Nsp8 might be located at its N-terminal region depicted in Fig. 3.25. Nevertheless, it is important to perform more biochemical and structural studies on the Nsp8-RNA complex in order to reveal the mechanism of this non-canonical RNA-dependent RNA polymerase.

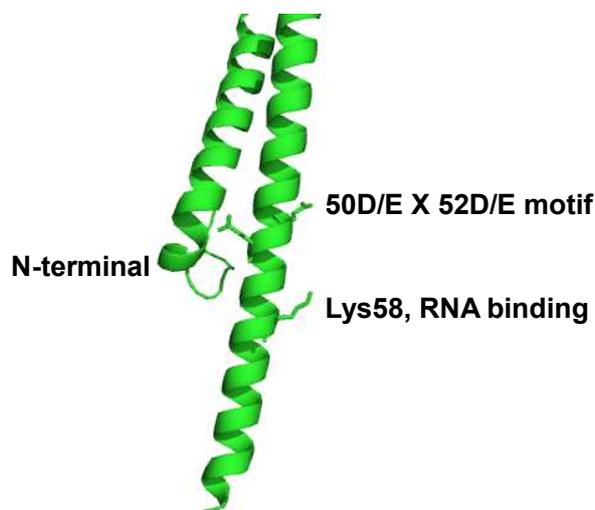


Fig. 3.25 Residues important for RNA polymerase activity of Nsp8

3.4 HKU4 M^{pro}

3.4.1 Protein production

The purity of HKU4 M^{pro} was analyzed by SDS-PAGE under denaturing conditions after size-exclusion chromatography (SEC). The protein exhibited an apparent molecular mass of about 34.0 kDa (Fig. 3.26), and its purity was higher than 95%.

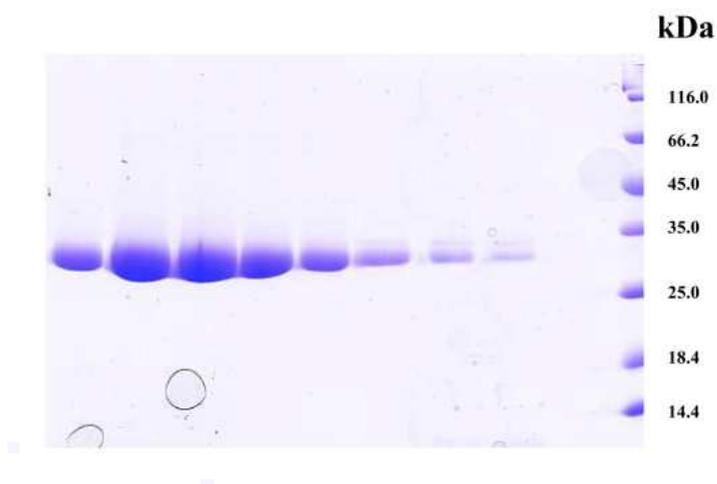


Fig. 3.26 SDS-PAGE analysis of HKU4 M^{pro}

3.4.2 Crystallization of HKU4 M^{pro}

Optimized crystals suitable for diffraction were grown in drops containing 2 μ L of protein solution and 2 μ L of reservoir solution (6% w/v polyethylene glycol 3350, 2% v/v Tacsimate, 5% v/v 2-propanol, 0.1 M imidazole pH 6.9) at 20°C (Fig. 3.27). Compound SG85 was initially dissolved in DMSO and then diluted with reservoir solution to 2 mM. The SG85-HKU4 M^{pro} complex crystals were prepared by soaking crystals of the free enzyme in this solution for 48 h.

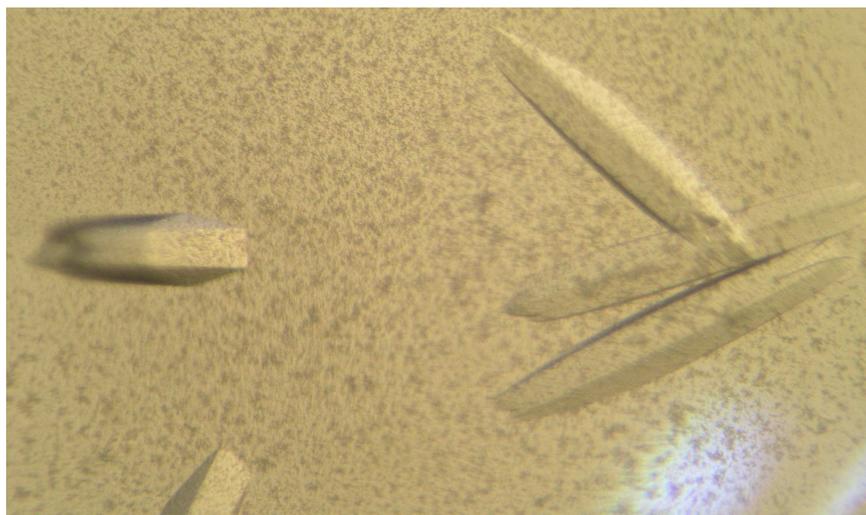


Fig. 3.27 Crystals of HKU4 M^{pro}.

3.4.3 Structure solution and quality of the structural models

The crystal structure of the free HKU4 M^{pro} was determined by molecular replacement, using the structure of its homologue from SARS-CoV (PDB code: 3TNT; Zhu *et al.*, unpublished) as search model. Crystals displayed space group P3₁21 with unit-cell dimensions $a = b = 79.62 \text{ \AA}$, $c = 178.37 \text{ \AA}$, and two monomers in the asymmetric unit. The crystal structure of HKU4 M^{pro} in complex with SG85 was initially refined with the model for the free enzyme. At a later stage, SG85 was modelled into difference density. Statistics of the datasets are summarized in Table 3.3.

Table 3.3. Diffraction-data collection and refinement statistics

	HKU4 M ^{pro}	HKU4 M ^{pro} + SG85
Data collection statistics		
X-ray source	BL14.2, BESSY, Berlin	
Wavelength (Å)	0.9184	
Space group	P3 ₁ 21	
Unit-cell dimensions (Å)	$a = 79.62, c = 178.37$	$a = 79.35, c = 178.67$
Resolution range (Å)	178.37-1.50 (1.52-1.50)	178.67-1.96 (2.00-1.96)
Number of unique reflections	105128 (3283)	47447 (2363)
Completeness (%)	99.6 (95.4)	100.0 (99.9)
Mean I/sigma(I)	32.4 (4.3)	29.9 (4.9)
Multiplicity	11.0 (8.8)	11.1 (11.3)
$R_{merge}(\%)^a$	5.0 (58.3)	5.9 (59.2)
$R_{measure}(\%)^b$	5.2 (61.9)	6.2 (61.9)
$R_{pim}(\%)^b$	1.6 (20.2)	1.8 (18.2)
Solvent content (%)	48	48
V_m (Å ³ /Da)	2.37	2.37
Refinement statistics		
Resolution range (Å)	68.96-1.50	68.72-1.96
$R_{work} / R_{free} (\%)^c$	15.7/17.5	17.1/19.8
Modeled residues	A: 1-306; B: 1-302	A: 1-306; B: 1-302
Protein atoms	9233	9242
Water and other solvent atoms	858	483
Ligand atoms	none	47
r.m.s.d in bond lengths (Å)	0.010	0.010
r.m.s.d in bond angles (°)	1.11	1.09
Ramachandran plot (%)		
Mostly favored	98.5	97.8
Additional allowed	1.5	2.2
Outliers	0	0

Values in parentheses represent the highest resolution shell.

^a $R_{merge} = \frac{\sum_{hkl} \sum_i |I(hkl)_i - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I(hkl)_i}$ where $I(hkl)$ is the intensity of reflection hkl and its symmetry equivalents and $\langle I(hkl) \rangle$ is the average intensity over all equivalent reflections.

^b $R_{measure}$: multiplicity-weighted R_{merge} ; R_{pim} : precision-indicating R_{merge} .

^c $R = \frac{\sum ||Fo| - |Fc||}{\sum |Fo|}$. $|Fo|$ and $|Fc|$ are amplitudes of the observed and calculated structure factors, respectively. R_{work} is the R value for reflections used in refinement, whereas R_{free} is the R value for 5% of the reflections which are selected randomly and are not included in the refinement.

3.4.4 Crystal structures of HKU4 M^{Pro} and its complex with inhibitor

3.4.4.1 Overall structure of HKU4 M^{Pro}

Crystals of the free HKU4 M^{Pro} and its complex with SG85 were isomorphous. The crystal structures were determined at 1.50 Å and 1.96 Å resolution, respectively (see Table 3.3 for diffraction data and crystallographic refinement statistics). One asymmetric unit contains two protease molecules (designated A and B), which form a homodimer with a buried intermolecular surface of ~1300 Å², corresponding to about 9% of the total surface area of the monomer (Fig. 3.28 (a)). The overall r.m.s. difference between the C α positions of the two monomers is 0.6 Å. Residues 70-78, located in a loop region, exhibit a relatively large r.m.s. difference (2.1 Å) for the C α positions, probably due to either the intrinsic flexibility of this region or their locations in different crystallographic environments (Fig. 3.28 (b)). The overall fold of the HKU4 M^{Pro} monomer is similar to that of the corresponding enzymes of other coronaviruses. The HKU4 M^{Pro} monomer consists of three domains; the first two domains are β -barrels harboring the catalytic dyad (Cys148...His41) and the substrate-binding pocket between them, while the third domain is α -helical (Fig. 3.28 (b)). The dimerization interface includes the most important interactions between the amino-terminus ("N-finger", residues 1-7) of one monomer and residues lining the S1 substrate-binding pocket of the other (Glu169, Phe143). These interactions keep the S1 pocket of the other monomer in an open, catalytically competent shape capable of accommodating the P1-Gln residue of the substrate; hence, dimer formation is essential for the proteolytic activity of the mature protease.

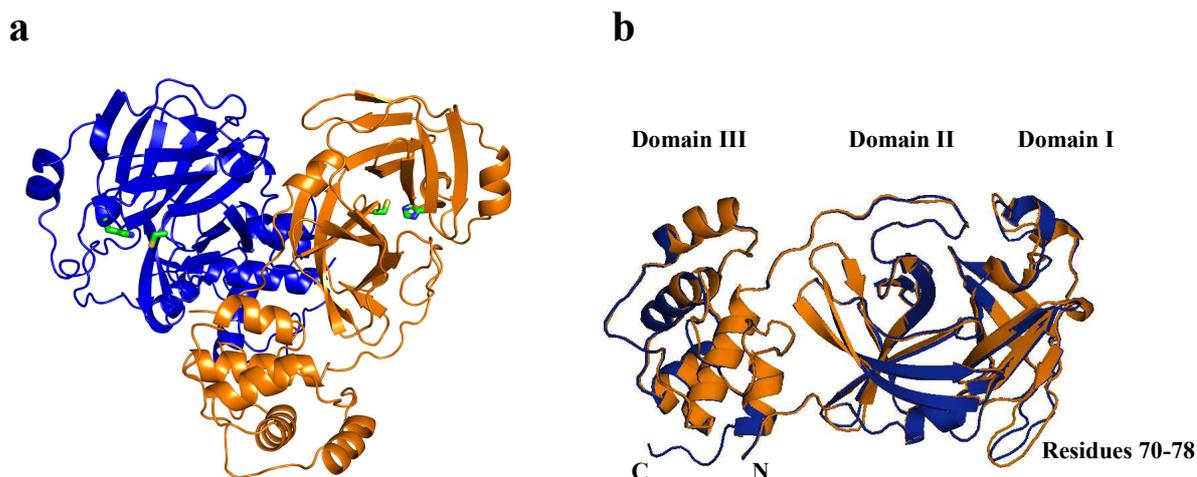


Fig. 3.28 Three-dimensional structure of the HKU4 M^{pro}

(a) Overall structure of the HKU4 M^{pro} dimer, the catalytic cysteine and histidine residues are highlighted. (b) Superimposition of monomer A and B; the three domains are indicated. Monomer A and B are colored in blue and orange, respectively.

3.4.4.2 The substrate-binding site of HKU4 M^{pro}

In the crystal of free HKU4 M^{pro}, the substrate-binding pocket of one monomer (chain B) is exposed to the solvent and thus represents a “free” state; while that of the other (chain A) is occupied by the C-terminal ₃₀₂GVV MQ₃₀₆ peptide from a symmetry-related copy (designated here as A’), resembling a product-bound form (Fig. 3.29). Although a His₅ tag had been attached to the C-terminus of the protease, the electron-density maps clearly shows that this tag has been cleaved off after Gln₃₀₆; thus the C-terminus of the M^{pro} is authentic. Even though the M^{pro} normally prefers small aliphatic residues such as serine or alanine at the S1’ position, there is also exception, e. g. the P1’ residue of the Nsp8/Nsp9 cleavage site is a conserved asparagine. Therefore, the observed unusual Gln-His cleavage may conceivably occur at a slow rate during crystallization.

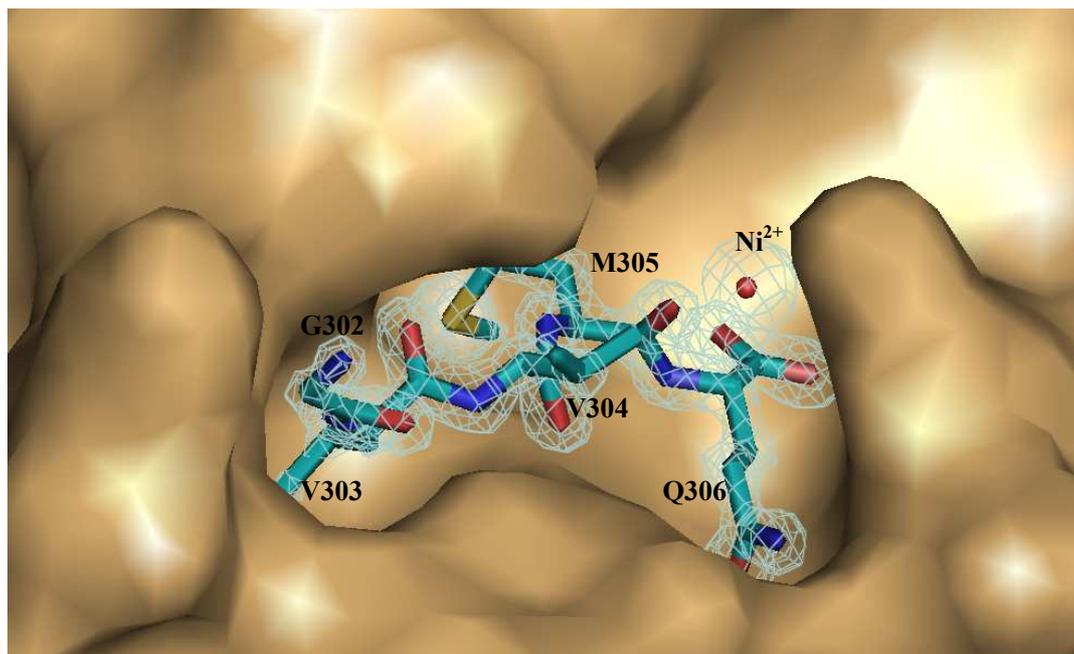


Fig. 3.29 The substrate-binding pocket of HKU4 M^{PrO} monomer A is occupied by the C-terminus of a symmetry-related molecule

The C-terminal $_{302}\text{GVVMQ}_{306}$ peptide of symmetry-related monomer A' inserts into the substrate-binding pocket of monomer A. C-terminal residues of monomer A' are shown in sticks (C: cyan; N: blue; O: red) and labeled. The $2F_o-F_c$ electron-density map is shown for the $_{302}\text{GVVMQ}_{306}$ peptide and around the presumably Ni^{2+} ion; the map is contoured at a level of 1σ .

In addition, there exists a strong, spherical peak (about 30σ) in the difference electron-density map, which is close to the active site of monomer A (Fig. 3.29). Because cysteine and histidine are good ligands for divalent metal ions, I interpreted this peak as a nickel ion, which originated most likely from the purification step using a Ni-NTA column. The HKU4 M^{PrO} is inactive in reaction buffer without EDTA, further supporting that there are divalent metal ions in the protein solution. Even though the interactions of this Ni^{2+} with Cys148-A, His45-A, and the carboxyl group of the C-terminal Gln306-A' might distort the geometry of these residues from their native state, they only have a negligible effect on other residues outside this region. On the other hand, the nickel is important for crystal formation; removal by adding 1 mM EDTA to the crystallization buffer resulted in no crystals at all. A nickel ion has also been placed into a rather weak electron-density peak close to the free active site of monomer B.

The C-terminal ${}_{303}\text{VVMQ}_{306}$ peptide of the neighbouring monomer A' binds to the S4, S2, and S1 subsites of monomer A, hence representing the P4 - P1 residues of an M^{pro} cleavage sequence, and thereby illustrating the product-binding mode.

The S1 subsite, formed by atoms of Phe143mc+sc, Leu144mc, Cys145mc, Ser147sc, Tyr164sc, His166sc, Met168sc, Glu169mc+sc, and His175sc (mc, main chain; sc, side-chain) is a compact space tailor-made to specifically harbour a glutamine residue. P1-Gln306 fills the S1 subsite; its O ϵ 1 atom accepts a hydrogen bond (2.78 Å) from the N ϵ 2 atom of conserved His166, and its N ϵ 2 atom donates two hydrogen bonds to O ϵ 1 of Glu169 (3.19 Å) and the main-chain oxygen of Phe143 (3.05 Å). One of the oxygen atoms of the C-terminal carboxylate of the symmetry-related molecule (residue Gln306) occupies the oxyanion hole, making hydrogen bonds with the peptide NH groups of Cys148 (3.05 Å) and Gly146 (2.85 Å) (Fig. 3.30).

The S2 subsite, constituted by Leu49sc, His41sc, Asp190sc, Gln192sc, Tyr54sc, Met184sc, and Met168sc, is large and capable of accommodating various residue types, such as leucine, phenylalanine, methionine, and valine, all of which occur at the P2 position of M^{pro} cleavage sites. The P2-Met305A' makes close contacts with Met168sc, His41sc, Gln192sc, and Asp190sc, mainly via hydrophobic and van der Waals interactions. Its main-chain oxygen accepts a hydrogen bond from the N ϵ atom of Gln192 (2.96 Å). While the side chain of P3-Val304A' is oriented towards the solvent, its main-chain amide and oxygen form hydrogen bonds with the main-chain oxygen (2.88 Å) and amide (2.88 Å) of Glu169, respectively (Fig. 3.30).

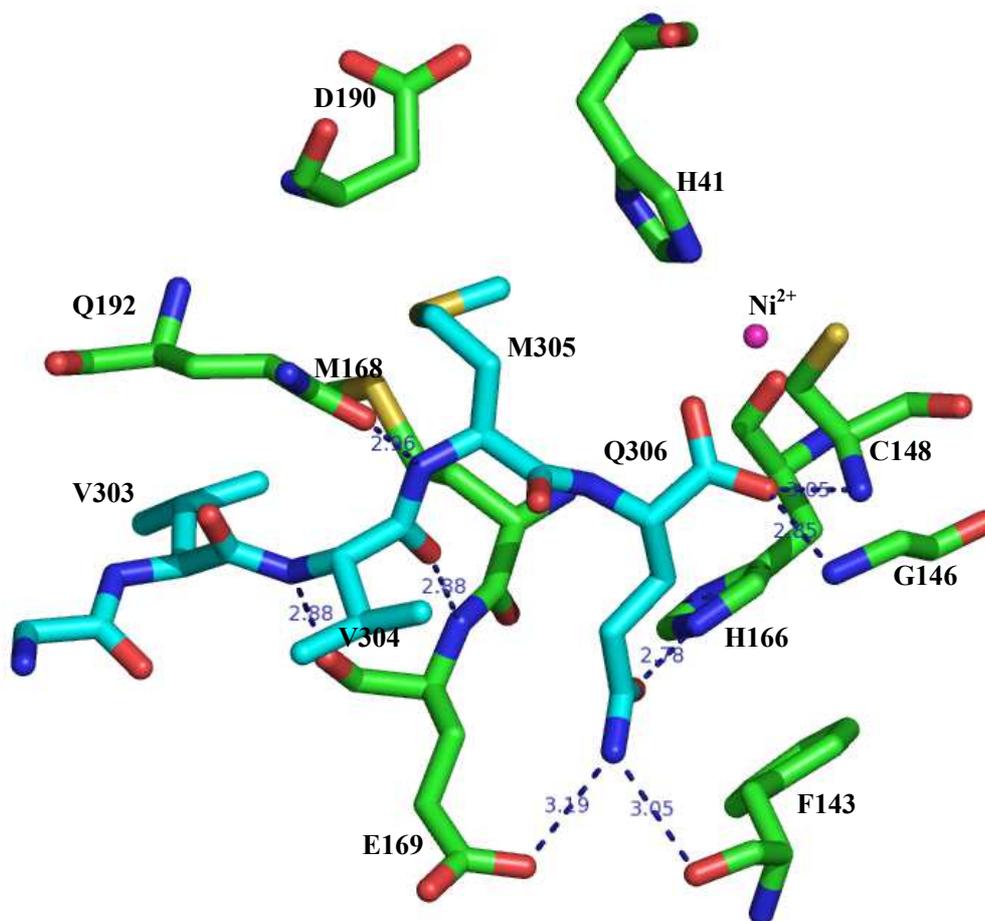


Fig. 3.30 Interactions of monomer A' C-terminal $_{302}\text{GVVMQ}_{306}$ residues with the S4-S1 subsites of monomer A.

Residues involved in the interactions are shown in sticks (C-monomer A': cyan; C-monomer A: green; N: blue; O: red) and labeled. These hydrogen bonds are represented by dashed lines. Distances between atoms are shown in Å.

The S4 subsite, formed by Gln195sc, Glu189mc, Leu170sc, and Met168sc, is rather shallow and connected to the S2 pocket. P4-Val303A' is located at the rim of this pocket, interacting hydrophobically with Met168, Leu170, and the aliphatic part of Gln195.

Compared to the “free” substrate-binding pocket, the S2 site in the product-bound form does not demonstrate significant conformational differences, except that the side-chain of Gln192 moves towards the pocket and forms a hydrogen bond with the main-chain oxygen of the peptide.

3.4.4.3 Crystal structure of HKU4 M^{pro} in complex with the inhibitor SG85

SG85 is a Michael acceptor compound of the formula Cbz-Ser(O-tBu)-Phe-GlnLactam-CH=CH-CO-OEt (Tan *et al.*, 2013). The β -carbon atom of its vinylogous ethyl ester moiety will form a covalent bond with the nucleophilic cysteine of the protease (Fig. 3.31). It is non-toxic (CC_{50} of 190 μ M in Vero cells, and 265 μ M in HuhT7 cells; Tan *et al.*, 2013). The compound irreversibly inhibits HKU4 M^{pro} with inhibition parameters $K_i = 6.78 \pm 0.15 \mu$ M, $k_3 = 0.049 \pm 0.002 \text{ s}^{-1}$, as determined by using a FRET-based assay.

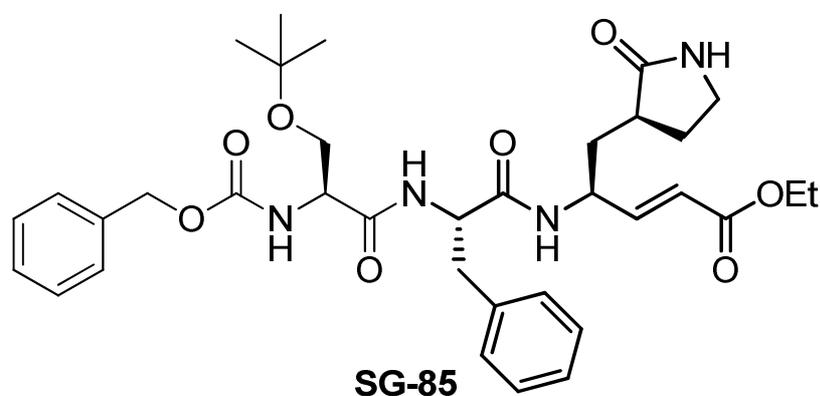


Fig. 3.31 Structural formula of inhibitor SG85

In the crystal structure of HKU4 M^{pro} in complex with SG85, the inhibitor occupies subsites S4 to S1' of molecule B (Fig. 3.32). Molecule A again harbours the C-terminal residues of a symmetry-related A' molecule in its substrate-binding pocket; the conformation of A' residues 303 – 306 is basically the same as in the structure of the “free” HKU4 M^{pro}. No major conformational changes were observed in the SG85 complex structure compared to the free state of the enzyme. However, local conformational changes did occur, such as a movement of the side-chain of Gln192 towards the S2 pocket, enabling its O ϵ 1 to form a hydrogen bond to the main-chain amide of the P2-phenylalanine of SG85, which is also observed in the product-binding mode.

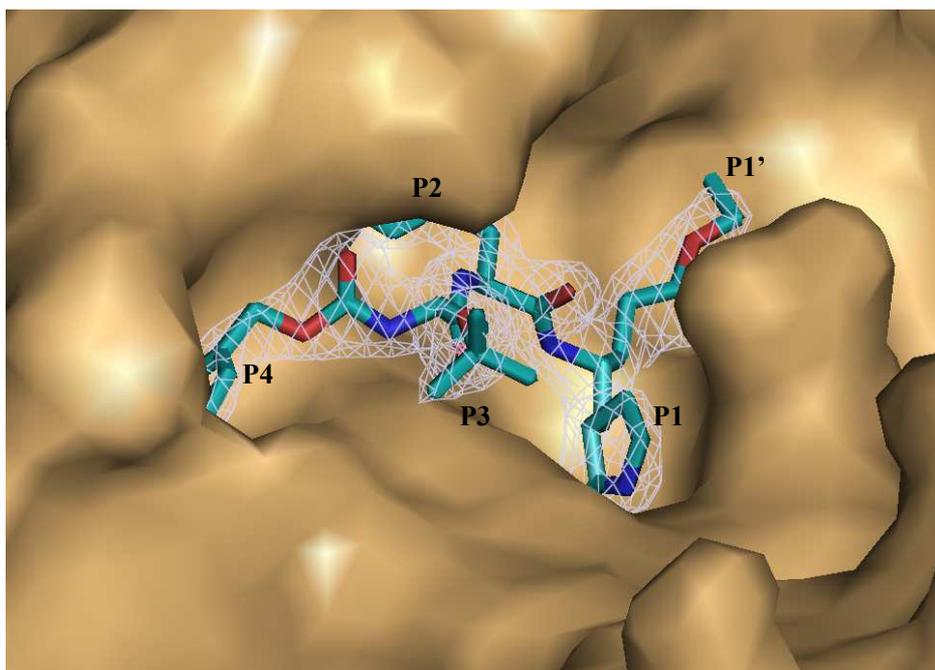


Fig. 3.32 The substrate-binding pocket of HKU4 M^{pro} is occupied by the inhibitor SG85

SG85 binds to the substrate-binding pocket of HKU4 M^{pro} monomer B. SG85 is shown in sticks (C: cyan; N: blue; O: red) and its P1' - P4 residues are labeled. The $2F_o - F_c$ electron-density map is shown for SG85 and contoured at a level of 1σ .

The most important interactions occur between the P1 residue of SG85 and the S1 subsite of the protease. The P1 side-chain fills the S1 pocket; it is a five-membered lactam, which was chosen in this position to replace the P1-Gln residue of the substrate in order to reduce the entropy loss upon binding. The specific hydrogen bonding is retained from the product-bound form: the lactam oxygen atom accepts a hydrogen bond (2.60 Å) from His166sc; the lactam amide donates a bifunctional hydrogen bond to the carboxylate group of Glu169 (3.09 Å) and the main-chain oxygen of Phe143 (3.07 Å). The β -carbon of the vinyl group is covalently bound to the sulphur of the active-site Cys148, rendering an S configuration. The carbonyl oxygen of the ester is located in the oxyanion hole, accepting hydrogen bonds from the amide groups of Cys148 (3.07 Å) and possibly Gly146 (3.43 Å). The ethoxy group only displayed faint electron density, probably due to its weak interaction with the S1' subsite.

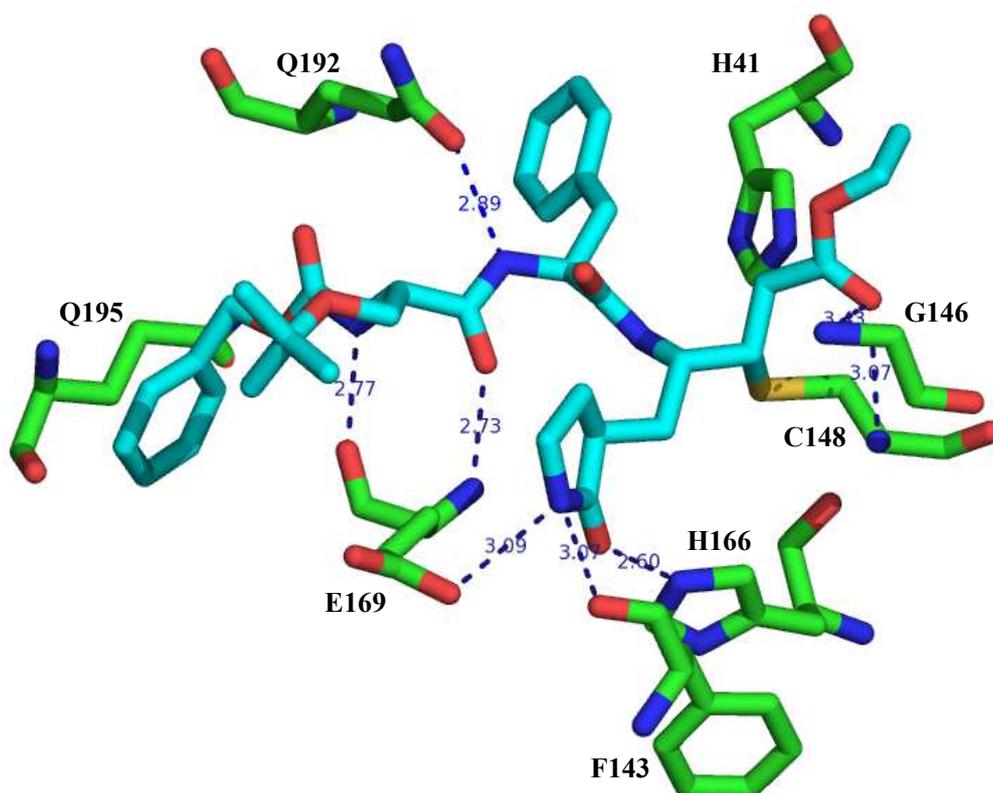


Fig. 3.33 The inhibitor SG85 interacts extensively with HKU4 M^{pro}

Residues involved in the interactions are shown as sticks (C-SG85: cyan; C-HKU4 M^{pro}: green; N: blue; O: red) and labeled. These hydrogen bonds are represented by blue dashed lines. Distances between atoms are shown in Å.

The P2-phenylalanine inserts deeply into the S2 pocket (Fig. 3.33). It interacts with His41sc, Gln192sc, Leu49sc, Met168sc, and Lys191mc via hydrophobic contacts. However, the electron density of the side chain showed a certain degree of discontinuity, implying sub-optimal interactions with the protein. More flexible aliphatic residues such as norleucine might be a better choice in this position. The hydrogen bond between its main-chain amide and O ϵ 1 of Gln192 (2.89 Å) induced an orientation shift in the side chain of the latter, as also seen in the product-bound form (Fig. 3.30). Remarkably, this hydrogen bond might be essential for the SG85 interaction: SG85 cannot inhibit alphacornavirus M^{pro}s (HKU8 and NL63), in which this Gln is replaced by Pro, and unable to make this hydrogen bond (Xiao *et al.*, unpublished; Zhang *et al.*, unpublished).

The side-chain of the P3 residue has little interaction with the protein and is not well defined by electron density. However, the main-chain oxygen and amide form hydrogen bonds with the main-chain amide (2.73 Å) and oxygen atom (2.77 Å) of Glu169, respectively (Fig. 3.33).

The electron density of the benzyl moiety of the capping benzyloxycarbonyl (Cbz) group is weaker than for other parts of the inhibitor. Its phenyl ring interacts with the S4 pocket residues Gln195sc, Leu170sc, and Met168sc mainly via hydrophobic contacts. However, the 4-carbon of the phenyl ring also makes a rather close contact with the main-chain oxygen of Gln195 (2.66 Å), which may impose an adverse effect on the binding.

In conclusion, the P1-GlnLactam of SG85 is the most important contributor to the interactions with the protein; P2, P4, P1' contribute significantly, but there appears to be space for their optimization; P3 makes little contribution to binding but too long a side-chain in this position can lead to cytotoxicity (Tan *et al.*, 2013).

3.4.4.4 Comparison with SARS-CoV M^{pro}

The overall r.m.s.d value for C α atoms between the M^{pro} dimers of HKU4 and SARS-CoV is 1.97 Å (Fig. 3.34). The structure of domain I and domain II (r.m.s.d value for 200 C α atoms is 1.32 Å) is more conserved (except residues 71-76) than that of domain III (r.m.s.d values for 101 C α atoms is 2.82 Å). In the substrate binding sites, significant differences occur around residues 49 (Leu in HKU4 and MERS-CoV; Met in SARS-CoV) and conserved Gln192 (HKU4 numbering), both of which line the S2 specificity pocket.

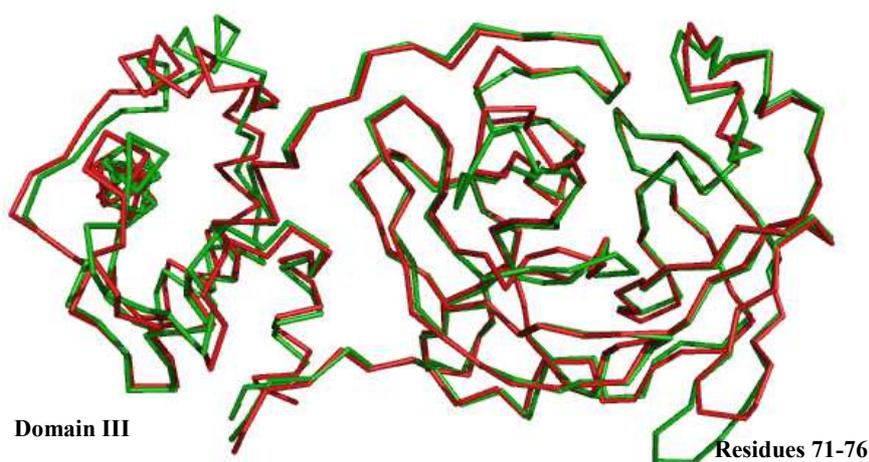


Fig. 3.34 Superimposition of M^{pro} monomers of HKU4 and SARS-CoV

The C α trace of the HKU4 M^{pro} is shown in green ribbons and that of the SARS-CoV M^{pro} (PDB code 1UJ1, Yang *et al.*, 2003) is in red ribbons.

The dimerization interface is also somehow different between the HKU4 and SARS-CoV M^{pro}s. In the SARS-CoV M^{pro}, Arg4 of its "N-finger" form a salt-bridge with Glu290 of domain III of the other monomer, while this Arg is replaced by Val in HKU4 M^{pro}, therefore disrupting the salt-bridge formation. As a compensation, a new salt-bridge is formed between Lys124 (Ser in SARS-CoV) of one monomer and Glu155 (Ile in SARS-CoV) of the other. I conclude that while the overall fold of the main protease is conserved in HKU4 and MERS-CoV when compared to SARS-CoV, this is only partly true for the substrate-binding site and the dimerization interface.

Even though SG85 also binds and inhibits the M^{pro} of SARS-CoV, there are significant differences in its binding mode to the S4 and S1' pockets due to exchanges of amino-acid residues shaping these sites (Fig. 3.35). In SARS-CoV M^{pro}, the SG85 P1' residue bends outside and interacts with its own P3 residue via hydrophobic forces (Zhu *et al.*, unpublished), while it is located right in the S1' pocket in HKU4 M^{pro}. The substitution of Thr25 of SARS-CoV M^{pro} by Met25 increases the hydrophobicity of the S1' pocket of HKU4 M^{pro}. The S4 pocket is apparently larger in HKU4 M^{pro}, and implicitly in MERS-CoV M^{pro} as well, than in SARS-CoV M^{pro}. The Ala194 in the SARS-CoV M^{pro} S4 pocket is substituted by a glutamine (Gln197) in

HKU4 M^{pro}. The side chain of this glutamine forms a hydrogen bond with Ser171 and pushes the loop (194-198) outward, thereby making the S4 pocket larger. However, the residues contributing to differences in the binding modes still remain to be identified by mutagenesis experiments.

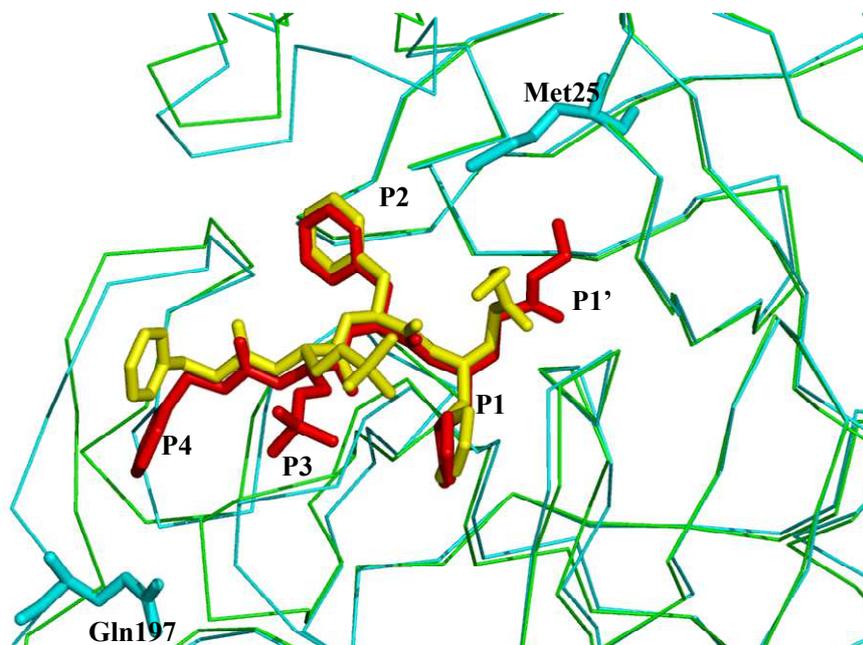


Fig. 3.35 Comparison of HKU4 M^{pro} and SARS-CoV M^{pro} in complex with SG85
The C α trace of the HKU4 M^{pro} is shown in cyan ribbons and that of the SARS-CoV M^{pro} is shown in green; SG85 in complex with HKU4 M^{pro} and SARS-CoV M^{pro} are shown in red and yellow sticks, respectively. The structure of SARS-CoV M^{pro} in complex with SG85 was determined by Zhu *et al.* (unpublished).

3.4.5 Discussion

Viruses arising from animal reservoirs are of increasing threat to humans, and they effectively emerge and re-emerge in the human population. Approximately 75% of all emerging infectious diseases are zoonoses, emphasizing the inexorable link between human and animal health. Transmission of zoonotic pathogens between wildlife species and humans always occur in an unpredictable manner. Therefore, mitigating and tackling a possible exposure of zoonotic pathogens originated from wildlife species is difficult.

To date, over 1230 different bat species have been identified worldwide. These species exhibit a large variety of ecological and behavioral adaptations that have

developed during their evolutionary history of about 70 million years. The tremendous time line of evolution and the gregarious habits of bats seem to have favored bats to carry a large diversity of viral agents, many of which are possible ancestors of human or other animal viruses. The pandemic of SARS aroused a new era of intensive investigations into wildlife reservoirs, including bats. As suggested by previous studies, many human coronaviruses are supposed to have originated from animal coronaviruses, especially bat coronaviruses. For example, SARS-CoV very likely originated from a bat reservoir and was transmitted to humans via the civet as intermediate (Li *et al.*, 2005). An alphacoronavirus isolated from the North American tricolored bat has been proposed to share common ancestry with HCoV-NL63 (Huynh *et al.*, 2012). Remarkably, the newly emerged human coronavirus MERS-CoV, which belongs to the betacoronavirus clade C, is evolutionarily close to bat coronaviruses HKU4 and HKU5.

The emergence of MERS-CoV demonstrates that future zoonotic transmission of bat coronaviruses into the human population is likely to occur. In order to increase the preparedness against such events, our laboratory has embarked on a program focussing on the structure-based design of broad-spectrum antivirals that would also inhibit bat coronaviruses of the various families and clades, in case of zoonotic transmission of such viruses into the human population. The target in these drug design efforts is the M^{pro}, an enzyme that has been shown to be a useful target for anti-coronavirus inhibitors earlier (Anand *et al.*, 2003; Yang *et al.*, 2005; Zhu *et al.*, 2011).

Coronavirus genomes encode two large polyproteins, pp1a and pp1ab, which are then processed by viral proteases to yield the individual components of the large replication/transcription complex. The M^{pro} cleaves these huge substrates at no less than 11 sites out of the 15 processing sites, and the processing is crucial for viral replication and transcription. Therefore, the M^{pro} is one of the major targets for anti-coronaviral inhibitor development. The M^{pro} of TGEV was the first reported

three-dimensional structure of coronavirus (Anand *et al.*, 2002). After the outbreak of SARS-CoV, crystal structures of many different coronavirus M^{pro}s have been determined, i.e., HCoV-229E (Anand *et al.*, 2003), SARS-CoV (Yang *et al.*, 2003), HKU1 (Zhao *et al.*, 2008), IBV (Xue *et al.*, 2008). In general, M^{pro}s of coronaviruses share overall similarity in both fold and substrate binding sites, which forms a basis of designing wide-spectrum M^{pro} inhibitors. On the other hand, there are also significant differences around the substrate binding sites, which should not be neglected in broad-spectrum inhibitor design for best efficiency against each species. For example, SG85 and some of its derivatives are quite active against SARS-CoV M^{pro}, but exhibit very poor inhibition against alpha-CoV M^{pro}s. Therefore, understanding the details of different M^{pro}s is important for broad-spectrum and specific inhibitor design.

In this thesis, I have determined the crystal structure of HKU4 M^{pro}, the first three-dimensional structure from bat coronavirus. This allowed visualization of the substrate-binding sites in three states: free, product-bound, and inhibitor-bound. It can serve as a useful inhibitor design model for this clade. MERS-CoV exhibits around 80% sequence identity to the bat coronavirus HKU4 (van Boheemen *et al.*, 2012). Its M^{pro} has 81% identity (90% similarity) to the homologue from HKU4 (Fig. 3.36); hence, a high similarity can be expected for their 3-D structures as well. The crystal structure of HKU4 M^{pro} in complex with SG85 provides detailed insight into the interactions of the inhibitor with the protein, thereby helping design of improved versions of the former. In particular, the amino-acid residues in the active site and substrate-binding cleft, which are in contact with the inhibitor SG85, are 100% conserved between the two viruses. Consequently, SG85 inhibits MERS-CoV with EC₅₀ of 5 μ M in a cell-based assay (carried out by D. Muth, University of Bonn Medical Centre).

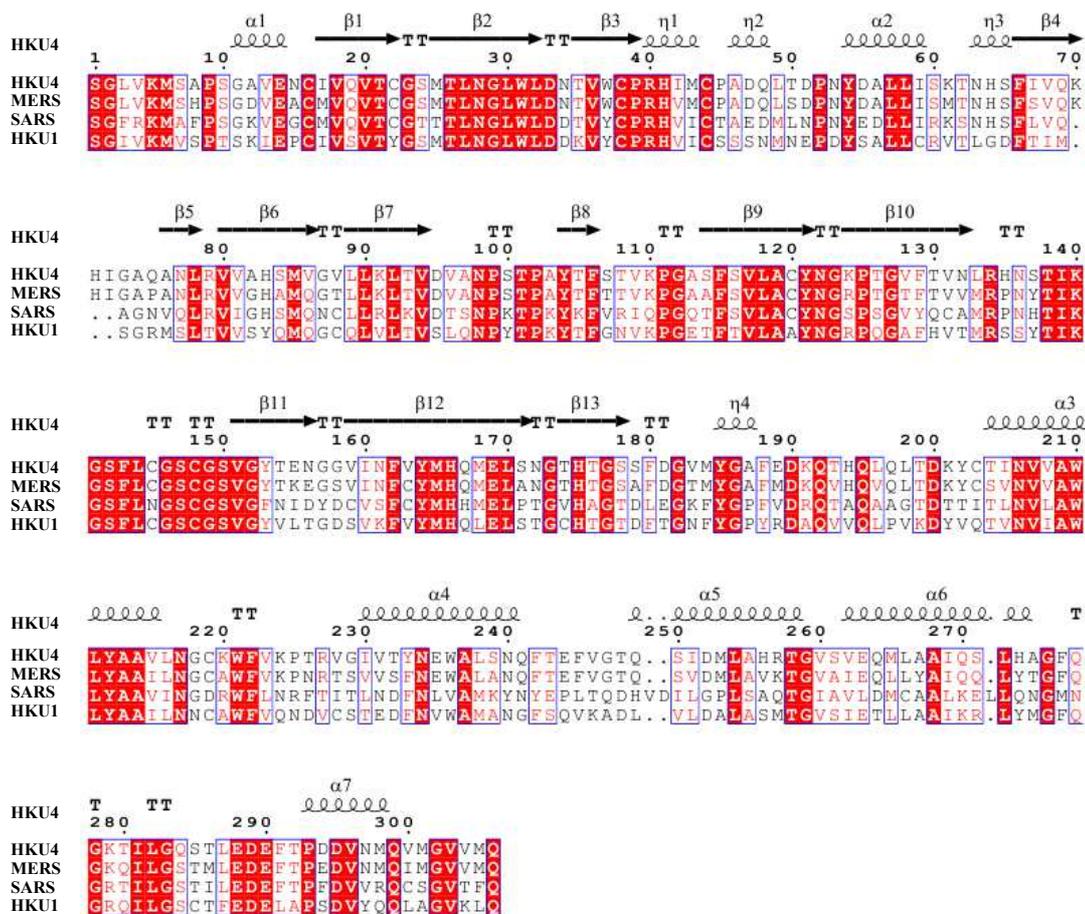


Fig. 3.36 Sequence alignment of the HKU4 M^{pro} with homologues from other coronaviruses

The alignments were achieved by using ClustalW2 and the colored figures were generated by ESPript2.2. Residues labeled in red are conserved to more than 70% and residues boxed in red are completely identical. The secondary structure of HKU4 M^{pro} is indicated on top of the alignments. All sequences aligned here are from the National Center for Biotechnology Information (NCBI) database.

Until now, 130 cases of human infection with this new coronavirus have been reported. Although human-to-human transmission still appears to be limited, the high case-fatality ratio (~45%) makes this virus a dangerous thread. It remains to be seen whether the new coronavirus MERS-CoV will create problems similar to the SARS outbreak of 2003. But in any case, it is advisable to get prepared for a similar scenario; developing anti-MERS inhibitors is hence an urgent task.

4. References

- Adedeji, A.O., Marchand, B., Te Velhuis, A.J., Snijder, E.J., Weiss, S., Eoff, R.L., Singh, K. & Sarafianos, S.G. (2012): Mechanism of nucleic acid unwinding by SARS-CoV helicase. *PLoS One* **7**, e36521.
- Ahn, D.G., Choi, J.K., Taylor, D.R. & Oh, J.W. (2012): Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Arch. Virol.* **157**, 2095-2104.
- Almeida, M.S., Johnson, M.A., Herrmann, T., Geralt, M. & Wüthrich, K. (2007): Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. *J. Virol.* **81**, 3151-3161.
- 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 (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763-1767.
- Angelini, M.M., Akhlaghpour, M., Neuman, B.W. & Buchmeier, M.J. (2013): Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *MBio* **4**, e00524-13.
- Bandukwala, H.S., Wu, Y., Feuerer, M., Chen, Y., Barboza, B., Ghosh, S., Stroud, J.C., Benoist, C., Mathis, D., Rao, A. & Chen, L. (2011): Structure of a domain-swapped FOXP3 dimer on DNA and its function in regulatory T cells. *Immunity* **34**, 479-491.
- Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B. & Decroly, E. (2012): RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl. Acad. Sci. USA* **109**, 9372-9377.
- Brayton, P.R., Lai, M.M., Patton, C.D. & Stohlman, S.A. (1982): Characterization of two RNA polymerase activities induced by mouse hepatitis virus. *J. Virol.* **42**, 847-853.
- Brian, D.A. & Baric, R.S. (2005): Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* **287**, 1-30.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, P., Roversi, P., Sharff, A., Smart, O., Vornrhein, C. & Womack, T. (2010): BUSTER version 2.11.0. Cambridge, United Kingdom: Global Phasing Ltd.

Castro, C., Smidansky, E., Maksimchuk, K.R., Arnold, J.J., Korneeva, V.S., Gotte, M., Konigsberg, W. & Cameron, C.E. (2007): Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc. Natl. Acad. Sci. USA* **104**, 4267-4272.

Chen, B., Fang, S., Tam, J.P. & Liu, D.X. (2009): Formation of stable homodimer via the C-terminal alpha-helical domain of coronavirus nonstructural protein 9 is critical for its function in viral replication. *Virology* **383**, 328-337.

Chen, Y., Cai, H., Pan, J., Xiang, N., Tien, P., Ahola, T. & Guo, D. (2009): Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc. Natl. Acad. Sci. USA* **106**, 3484-3489.

Chen, Y., Su, C., Ke, M., Jin, X., Xu, L., Zhang, Z., Wu, A., Sun, Y., Yang, Z., Tien, P., Ahola, T., Liang, Y., Liu, X. & Guo, D. (2011): Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS Pathog.* **7**, e1002294.

Chen, Y.W., Stott, K. & Perutz, M.F. (1999): Crystal structure of a dimeric chymotrypsin inhibitor 2 mutant containing an inserted glutamine repeat. *Proc. Natl. Acad. Sci. USA* **96**, 1257-1261.

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.

Cornillez-Ty, C.T., Liao, L., Yates, J.R.III., Kuhn, P. & Buchmeier, M.J. (2009): Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *J. Virol.* **83**, 10314-10318.

Decroly, E., Imbert, I., Coutard, B., Bouvet, M., Selisko, B., Alvarez, K., Gorbalenya, A.E., Snijder E.J. & Canard, B. (2008): Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J. Virol.* **82**, 8071-8084.

Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., Gluais, L., Papageorgiou, N., Sharff, A., Bricogne, G., Ortiz-Lombardia, M., Lescar, J. & Canard, B. (2011): Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog.* **7**, e1002059.

de Haan, C.A., Li, Z., te Lintelo, E., Bosch, B.J., Haijema, B.J & Rottier, P.J. (2005): Murine coronavirus with an extended host range uses heparan sulfate as an entry receptor. *J. Virol.* **79**, 14451-14456.

Deming, D.J., Graham, R.L., Denison, M.R. & Baric, R.S. (2007): Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. *J. Virol.* **81**, 10280-10291.

Donaldson, E.F., Graham, R.L., Sims, A.C., Denison, M.R. & Baric, R.S. (2007): Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in polyprotein processing. *J. Virol.* **81**, 7086-7098.

Donaldson, E.F., Sims, A.C., Graham, R.L., Denison, M.R. & Baric, R.S. (2007): Murine hepatitis virus replicase protein nsp10 is a critical regulator of viral RNA synthesis. *J. Virol.* **81**, 6356-6368.

Dong, B.Q., Liu, W., Fan, X.H., Vijaykrishna, D., Tang, X.C., Gao, F., Li, L.F., Li, G.J., Zhang, J.X., Yang, L.Q., Poon, L.L., Zhang, S.Y., 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.

Doublié, S. (1997): Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523-530.

Eckerle, L.D., Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R.L., Baric, R.S., Stockwell, T.B., Spiro, D.J. & Denison, M.R. (2010): Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.* **6**, e1000896.

Egloff, M.P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J., Gorbalenya, A.E., Cambillau, C. & Canard, B. (2004): The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc. Natl. Acad. Sci. USA* **101**, 3792-3796.

Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. (2010): Features and development of COOT. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501.

Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J & Lai, L. (2004): Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.* **279**, 1637-1642.

Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U. & Ball, L.A. (eds.): Virus Taxonomy, VIIIth Report of the ICTV. Elsevier/Academic Press, London, 2005.

Gadlage, M.J., Sparks, J.S., Beachboard, D.C., Cox, R.G., Doyle, J.D., Stobart, C.C. & Denison, M.R. (2010): Murine hepatitis virus nonstructural protein 4 regulates virus-induced membrane modifications and replication complex function. *J. Virol.* **84**, 280-290.

Graham, R.L., Sims, A.C., Brockway, S.M., Baric, R.S. & Denison, M.R. (2005): The nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral replication. *J. Virol.* **79**, 13399-13411.

Gronenborn, A.M. (2009): Protein acrobatics in pairs – Dimerization via domain swapping. *Curr. Opin. Struct. Biol.* **19**, 39-49.

Guarino, L.A., Bhardwaj, K., Dong, W., Sun, J., Holzenburg, A. & Kao, C. (2005): Mutational analysis of the SARS virus Nsp15 endoribonuclease: identification of residues affecting hexamer formation. *J. Mol. Biol.* **353**, 1106-1117.

Guinier, A. (1939): La diffraction des rayons X aux très petits angles: Application a l'étude de phenomenes ultramicroscopiques. *Ann. Phys. (Paris)* **12**, 161-237.

Hattne, J. & Lamzin, V.S. (2008): Pattern-recognition-based detection of planar objects in three-dimensional electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 834-842.

Hilgenfeld, R., Tan, J., Chen, S., Shen, X. & Jiang, H. (2008): Structural proteomics of emerging viruses: The examples of SARS-CoV and other coronaviruses. In Sussman, J. & Silman, I. (eds), *Structural Proteomics and Its Impact on the Life Sciences*, WorldScientific, Singapore. pp. 361-433.

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

Hsue, B. & Masters, P.S. (1997): A bulged stem-loop structure in the 3' untranslated region of the genome of the coronavirus mouse hepatitis virus is essential for replication. *J. Virol.* **71**, 7567-7578.

Huang, C., Lokugamage, K.G., Rozovics, J.M., Narayanan, K., Semler, B.L. & Makino, S. (2011): SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. *PLoS Pathog.* **7**, e1002433.

Huynh, J., Li, S., Yount, B., Smith, A., Sturges, L., Olsen, J.C., Nagel, J., Johnson,

-
- J.B., Agnihothram, S., Gates, J.E., Frieman, M.B., Baric, R.S. & Donaldson, E.F. (2012): Evidence supporting a zoonotic origin of human coronavirus strain NL63. *J. Virol.* **86**, 12816-12825.
- Imbert, I., Guillemot, J. C., Bourhis, J. M., Bussetta, C., Coutard, B., Egloff, M. P., Ferron, F., Gorbalenya, A. E. & Canard, B. (2006): A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J.* **25**, 4933-4942.
- Ivanov, K.A., Thiel, V., Dobbe, J.C., van der Meer, Y., Snijder, E.J. & Ziebuhr, J. (2004): Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J. Virol.* **78**, 5619-5632.
- Johnson, M.A., Jaudzems, K. & Wüthrich, K. (2010): NMR structure of the SARS-CoV nonstructural protein 7 in solution at pH 6.5. *J. Mol. Biol.* **402**, 619-628.
- Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Brooun, A., Griffith, M., Moy, K., Yadav, M.K., Velasquez, J., Buchmeier, M.J., Stevens, R.C. & Kuhn, P. (2006): Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *J. Virol.* **80**, 7894-7901.
- Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Buchmeier, M.J., Stevens, R.C. & Kuhn, P. (2007): Crystal structure of a monomeric form of severe acute respiratory syndrome coronavirus endonuclease nsp15 suggests a role for hexamerization as an allosteric switch. *J. Virol.* **81**, 6700-6708.
- Kabsch, W. (2010): XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132.
- Kamitani, W., Huang, C., Narayanan, K., Lokugamage, K.G. & Makino, S. (2009): A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat. Struct. Mol. Biol.* **16**, 1134-1140.
- Kang, X., Zhong, N., Zou, P., Zhang, S., Jin, C. & Xia, B. (2012): Foldon unfolding mediates the interconversion between M(pro)-C monomer and 3D domain-swapped dimer. *Proc. Natl. Acad. Sci. USA* **109**, 14900-14905.
- Knoops, K., Kikkert, M., Worm, S.H., Zevenhoven-Dobbe, J.C., van der Meer, Y., Koster, A.J., Mommaas, A.M. & Snijder, E.J. (2008): SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol.* **6**, e226.
- Lai, M.M.C. & Cavanagh, D. (1997): The molecular biology of coronaviruses. *Adv. Virus Rev.* **48**, 1-100.

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.

Leslie, A.G.W. & Powell, H.R. (2007): Processing diffraction data with Mosflm. In Read, R.J. & Sussman, L.J. (eds), *Evolving Methods for Macromolecular Crystallography*, Springer, Netherlands. pp. 41-51.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B.T., Zhang, S. & Wang, L.F. (2005): Bats are natural reservoirs of SARS-Like coronaviruses. *Science* **310**, 676-679.

Li, S., Zhao, Q., Zhang, Y., Zhang, Y., Bartlam, M., Li, X. & Rao, Z. (2010): New nsp8 isoform suggests mechanism for tuning viral RNA synthesis. *Protein & Cell.* **1**, 198-204.

Liu, Y. & Eisenberg, D. (2002): 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**, 1285-1299.

Liu, P., Li, L., Millership, J.J., Kang, H., Leibowitz, J.L. & Giedroc, D.P. (2007): A U-turn motif-containing stem-loop in the coronavirus 5' untranslated region plays a functional role in replication. *RNA* **13**, 763-780.

Lokugamage, K.G., Narayanan, K., Huang, C. & Makino, S. (2012): Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation. *J. Virol.* **86**, 13598-13608.

Louis-Jeune, C., Andrade-Navarro, M.A. & Perez-Iratxeta, C. (2012): Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins* **80**, 374-381.

Makino, S. & Lai, M.M. (1989): High-frequency leader sequence switching during coronavirus defective interfering RNA replication. *J. Virol.* **63**, 5285-5292.

Manolaridis, I., Wojdyla, J.A., Panjekar, S., Snijder, E.J., Gorbalenya, A.E., Berglind, H., Nordlund, P., Coutard, B. & Tucker, P.A. (2009): Structure of the C-terminal domain of nsp4 from feline coronavirus. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 839-846.

Masters, P.S. (2006): The molecular biology of coronaviruses. *Adv. Virus. Res.* **66**, 193-292.

Mescalchin, A., Wünsche, W. & Sczakiel, G. (2011): Specific recognition of proteins by array-bound hexanucleotides. *Angew. Chem. Int. Ed. Engl.* **50**, 1052-1054.

Miknis, Z.J., Donaldson, E.F., Umland, T.C., Rimmer, R.A., Baric, R.S. & Schultz, L.W. (2009): Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *J. Virol.* **83**, 3007-3018.

Minton, A.P. (2001): The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* **276**, 10577-10580.

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

Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J. & Brierley, I. (2006): A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244-247.

Narayanan, K., Huang, C., Lokugamage, K., Kamitani, W., Ikegami, T., Tseng, C.T. & Makino, S. (2008): Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. *J. Virol.* **82**, 4471-4479.

O'Neill, J.W., Kim, D.E., Johnsen, K., Baker, D. & Zhang, K.Y. (2001): Single-site mutations induce 3D domain swapping in the B1 domain of protein L from *Peptostreptococcus magnus*. *Structure* **9**, 1017-1027.

Perlman, S. & Netland, J. (2009): Coronaviruses post-SARS: update on replication and pathogenesis. *Nature Reviews Microbiology* **7**, 439-450.

Peti, W., Johnson, M. A., Herrmann, T., Neuman, B. W., Buchmeier, M. J., Nelson, M., Joseph, J., Page, R., Stevens, R.C., Kuhn, P. & Wüthrich, K. (2005): Structural genomics of the severe acute respiratory syndrome coronavirus: nuclear magnetic resonance structure of the protein nsP7. *J. Virol.* **79**, 12905-12913.

Pfefferle, S., Schöpf, J., Kögl, M., Friedel, C.C., Müller, M.A., Carbajo-Lozoya, J., Stellberger, T., von Dall'Armi, E., Herzog, P., Kallies, S., Niemeyer, D., Ditt, V., Kuri, T., Züst, R., Pumpor, K., Hilgenfeld, R., Schwarz, F., Zimmer, R., Steffen, I., Weber, F., Thiel, V., Herrler, G., Thiel, H.J., Schwegmann-Wessels, C., Pöhlmann, S., Haas, J., Drosten, C. & von Brunn, A. (2011): The SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. *PLoS Pathog.* **7**, e1002331.

Picone, D., Di Fiore, A., Ercole, C., Franzese, M., Sica, F., Tomaselli, S. & Mazzarella L. (2005): The role of the hinge loop in domain swapping. The special case of bovine seminal ribonuclease. *J. Biol. Chem.* **280**, 13771-137788.

Piotrowski, Y., Hansen, G., Boomaars-van der Zanden, A.L., Snijder, E.J., Gorbalenya, A.E. & Hilgenfeld, R. (2009): Crystal structures of the X-domains of a Group-1 and a Group-3 coronavirus reveal that ADP-ribose-binding may not be a conserved property. *Protein Sci.* **18**, 6-16.

Ponnusamy, R., Moll, R., Weimar, T., Mesters, J.R. & Hilgenfeld, R. (2008): Variable oligomerization modes in coronavirus non-structural protein 9. *J. Mol. Biol.* **383**, 1081-1096.

Ponnusamy, R. (2010): Crystallographic and biochemical investigations on coronavirus replication proteins - non-structural proteins 8 and 9. Ph.D. thesis, University of Lübeck, Lübeck, Germany.

Ricagno, S., Egloff, M.P., Ulferts, R., Coutard, B., Nurizzo, D., Campanacci, V., Cambillau, C., Ziebuhr, J. & Canard, B. (2006): Crystal structure and mechanistic determinants of SARS coronavirus nonstructural protein 15 define an endoribonuclease family. *Proc. Natl. Acad. Sci. USA* **103**, 11892-11897.

Roessle, M.W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrman, T., Konarev, P., Round, A., Fiedler, S., Hermes, C. & Svergun, D.I. (2007): Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *J. Appl. Cryst.* **40 (Supplement)**, 190-194.

Round, A.R., Franke, D., Moritz, S., Huchler, R., Fritsche, M., Malthan, D., Klaering, R., Svergun, D.I. & Roessle, M.W. (2008): Automated sample-changing robot for solution scattering experiments at the EMBL Hamburg SAXS station X33. *J. Appl. Cryst.* **41**, 913-917.

Rousseau, F., Schymkowitz, J.W., Wilkinson, H.R. & Itzhaki, L.S. (2001): Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proc. Natl. Acad. Sci. USA* **98**, 5596-5601.

Saikatendu, K.S., Joseph, J.S., Subramanian, V., Clayton, T., Griffith, M., Moy, K., Velasquez, J., Neuman, B.W., Buchmeier, M.J., Stevens, R.C. & Kuhn, P. (2005): Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1"-phosphate dephosphorylation by a conserved domain of nsP3. *Structure* **13**, 1665-1675.

Sawicki, S.G., Sawicki, D. L., Younker, D., Meyer, Y., Thiel, V., Stokes, H. & Siddell, S. G. (2005): Functional and genetic analysis of coronavirus replicase-transcriptase proteins. *PLoS Pathog.* **1**, e39.

Sawicki, S.G., Sawicki, D.L. & Siddell, S.G. (2007): A contemporary view of coronavirus transcription. *J. Virol.* **81**, 20-29.

Serrano, P., Johnson, M.A., Chatterjee, A., Neuman, B.W., Joseph, J.S., Buchmeier, M.J., Kuhn, P. & Wüthrich, K. (2009): Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. *J. Virol.* **83**, 12998-13008.

Sheldrick, G. M. (2010): Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 479-485.

Smith, I. & Wang, L.F. (2013): Bats and their virome: an important source of emerging viruses capable of infecting humans. *Curr. Opin. Virol.* **3**, 84-91.

Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X.C., Bartlam, M. & Rao, Z. (2006): Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10. *J. Virol.* **80**, 7902-7908.

Surjit, M. & Lal, S.K. (2008): The SARS-CoV nucleocapsid protein: a protein with multifarious activities. *Infect. Genet. Evol.* **8**, 397-405.

Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., Siddell, S., Poon, L.L., Diprose, J., Alderton, D., Walsh, M., Grimes, J.M. & Stuart, D.I. (2004): The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure* **12**, 341-353.

Svergun, D.I. (1992): Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495-503.

Svergun, D.I., Barberato, C. & Koch, M.H.J. (1995): CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768-773.

Tan, J., Kusov, Y., Mutschall, D., Tech, S., Nagarajan, K., Hilgenfeld, R. & Schmidt, C.L. (2007): The "SARS-unique domain" (SUD) of SARS coronavirus is an oligo(G)-binding protein. *Biochem Biophys Res Commun.* **364**, 877-882.

Tan, J., Vornrhein, C., Smart, O.S., Bricogne, G., Bollati, M., Kusov, Y., Hansen, G., Mesters, J.R., Schmidt, C.L. & Hilgenfeld, R. (2009): The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes.

PLoS Pathog. **5**, e1000428.

Tan, J., George, S., Kusov, Y., Perbandt, M., Anemüller, 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.

te Velthuis, A.J., Arnold, J.J., Cameron, C.E., van den Worm, S.H. & Snijder, E.J. (2010a): The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Res.* **38**, 203-214.

te Velthuis, A.J., van den Worm, S.H., Sims, A.C., Baric, R.S., Snijder, E.J. & van Hemert, M.J. (2010b): Zn²⁺ inhibits coronavirus and arterivirus RNA polymerase activity *in vitro* and zinc ionophores block the replication of these viruses in cell culture. *PLoS Pathog.* **6**, e1001176.

te Velthuis AJ, van den Worm, S.H. & Snijder, E.J. (2012): The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both *de novo* initiation and primer extension. *Nucleic Acids Res.* **40**, 1737-1747.

Thiel, V. & Siddell.S.J. (1994): Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. *J. Gen. Virol.* **75**, 3041-3046.

van Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T.M., Raj, V.S., Zaki, A.M., Osterhaus, A.D., Haagmans, B.L., Gorbalenya, A.E., Snijder, E.J. & Fouchier, R.A. (2012): Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio* **3**, e00473-12.

von Brunn, A., Teepe, C., Simpson, J.C., Pepperkok, R., Friedel, C.C., Zimmer, R., Rhonda, R., Baric, R. & Haas, J. (2007): Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFome. *PLoS ONE* **2**, e459.

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T. & Bricogne, G. (2011): Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293-302.

Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995): LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127-134.

Wathelet, M.G., Orr, M., Frieman, M.B. & Baric, R.S. (2007): Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. *J. Virol.* **81**, 11620-11633.

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., Snijder, E.J., Gorbalenya, A.E., Coutard, B., Piotrowski, Y., Hilgenfeld, R. & Tucker, P.A. (2009): Structure of the X (ADRP) domain of nsp3 from feline coronavirus. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1292-1300.

Woo, P.C., Lau, S.K., Lam, C.S., Lai, K.K., Huang, Y., Lee, P., Luk, G.S., Dyrting, K.C., Chan, K.H. & 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.

Xiao, Y., Ma, Q., Restle, T., Shang, W., Svergun, D.I., Ponnusamy, R., Sczakiel, G. & Hilgenfeld, R. (2012): Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. *J. Virol.* **86**, 4444-4454.

Xu, X., Liu, Y., Weiss, S., Arnold, E., Sarafianos, S.G. & Ding, J. (2003): Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res.* **31**, 7117-7130.

Xu, X., Lou, Z., Ma, Y., Chen, X., Yang, Z., Tong, X., Zhao, Q., Xu, Y., Deng, H., Bartlam, M. & Rao, Z. (2009): Crystal structure of the C-terminal cytoplasmic domain of non-structural protein 4 from mouse hepatitis virus A59. *PLoS One* **4**, e6217.

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.

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.

Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M. & Rao, Z. (2005): Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nat. Struct. Mol. Biol.* **12**, 980-986.

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.

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., Herold, J & Siddell, S.G. (1995): Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *J. Virol.* **69**, 4331-4338.

Ziebuhr, J., Thiel, V. & Gorbalenya, A.E. (2001): The autocatalytic release of a putative virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond. *J. Biol. Chem.* **276**, 33220-33232.

Züst, R., Cervantes-Barragán, L., Kuri, T., Blakqori, G., Weber, F., Ludewig, B. & Thiel, V. (2007): Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines. *PLoS Pathog.* **3**, e109.

Züst, R., Miller, T.B., Goebel, S.J., Thiel, V. & Masters, P.S. (2008): Genetic interactions between an essential 3' cis-acting RNA pseudoknot, replicase gene products, and the extreme 3' end of the mouse coronavirus genome. *J. Virol.* **82**, 1214-1228.