

Coronaviral Polyprotein *Nsp7-10*:
Proteolytic Processing and Dynamic Interactions within
the Transcriptase/Replicase Complex

Vorgelegt von

Sven Falke,
Hamburg, 2013

Dissertation

eingereicht im Fachbereich Chemie, Universität Hamburg
zur Erlangung des akademischen Grades Doctor rerum naturalium
(*Dr. rer. nat.*)

Die vorliegende Arbeit wurde im Zeitraum von Januar 2010 bis Oktober 2013 in der Arbeitsgruppe von Dr. Lars Redecke im Laboratorium für Strukturbiologie von Infektion und Entzündung der Universitäten Lübeck und Hamburg und am Institut für Biochemie und Molekularbiologie des Fachbereichs Chemie der Universität Hamburg durchgeführt.

1. Gutachter: Prof. C. Betzel

2. Gutachter: Prof. R. Bredehorst

Tag der Disputation: 06.12.2012

Für mich und alle die an mich geglaubt haben

„Strike 2...“

Table of contents

1.	Introduction	1
1.1.	Coronaviruses	1
1.2.	Polyproteins: pp1a and pp1ab.....	4
1.3.	The coronaviral main proteases (EC 3.4.22.69)	6
1.4.	Mature Nsp's and their interaction partners.....	9
1.5.	RNA polymerase activity – basic knowledge about Nsp7, Nsp8 and Nsp12 (and Nsp9)	11
1.6.	RNA capping – interaction of Nsp10 with Nsp14 and Nsp16	13
1.7.	Advanced methods in structural biology: Structure, function and oligomerization.....	15
1.8.	Aims of this project.....	16
2.	Material and methods	18
2.1.	Material and devices	18
2.1.1.	Devices	18
2.1.2.	Plastic consumables	18
2.1.3.	Plasmids and oligonucleotides	19
2.2.	Methods.....	23
2.2.1.	Agarose gel electrophoresis	23
2.2.2.	Polymerase chain reaction (PCR)	23
2.2.3.	Mutagenesis PCR.....	24
2.2.4.	Transformation of <i>E. coli</i>	24
2.2.5.	Plasmid isolation	24
2.2.6.	Heterologous gene overexpression in <i>E. coli</i>	24
2.2.7.	Gene expression and <i>in vivo</i> crystallization experiments in <i>Sf9</i> insect cells	25
2.2.8.	Cell lysis and chromatography.....	25
2.2.9.	Protein quantification	27
2.2.10.	SDS polyacrylamide gel electrophoresis (SDS-PAGE).....	27
2.2.11.	Native gel electrophoresis	28
2.2.12.	Western Blot.....	28
2.2.13.	Quantification of free thiol groups	29
2.2.14.	Quantification of enzymatic activity of M ^{Pro}	29
2.2.15.	Protein processing by M ^{Pro}	29
2.2.16.	T7 transcription, labelling of RNA and filter retention RNA-binding assay.....	30
2.2.17.	RNA-dependent RNA polymerase (RdRp) activity assay	31
2.2.18.	STD-NMR spectroscopy	31
2.2.19.	Surface plasmon resonance (SPR) spectroscopy	32
2.2.20.	Circular dichroism (CD) spectroscopy	33
2.2.21.	Dynamic light scattering	33
2.2.22.	Small-angle X-ray scattering (SAXS).....	34
2.2.23.	Crystallization, data collection and model building.....	35
2.2.24.	Mass spectrometry (MS).....	36

2.2.25.	<i>In silico</i> docking/modelling.....	38
3.	Results and discussion.....	40
3.1.	Sequence analysis and cloning	40
3.2.	HCoV-229E and SARS-CoV pp710	42
3.2.1.	Gene overexpression in <i>E. coli</i> and polyprotein purification	42
3.2.2.	Variable oligomerization of pp710	45
3.2.3.	Structural characterization of pp710	52
3.2.4.	Trials to crystallize pp710.....	71
3.2.5.	Insights into the function of pp710	73
3.3.	SARS-CoV M ^{pro} (Nsp5): Structural properties and monomer-dimer equilibrium.....	76
3.4.	Proteolytic processing of polyprotein Nsp7-10 by SARS-CoV M ^{pro}	79
3.4.1.	Interactions of mature Nsps	84
3.4.2.	Investigation of pp710 processing and cleavage product interactions by native MS	86
3.5.	Interaction of M ^{pro} with substrates lacking glutamine in P1 position.....	91
3.5.1.	Full-length polyproteins	91
3.5.2.	Full-length polyproteins with Q/N substitution in P1 position.....	100
3.5.3.	Inhibition of M ^{pro} by peptides derived from polyprotein recognition sites	104
3.6.	Interaction of M ^{pro} with flavonoid derivatives.....	115
3.7.	Co-crystallization trials with SARS-CoV M ^{pro}	116
3.8.	Purification and oligomerization of SARS-CoV pp79	118
3.8.1.	Dimerization of SARS-CoV pp79	118
3.8.2.	Mutation of a GXXXG motif does not alter dimerization of SARS-CoV pp79 and pp710.....	120
3.9.	Polyprotein Nsp7-8 (pp78)	123
3.9.1.	Oligomerization of HCoV-229E and SARS-CoV pp78	123
3.9.2.	SARS-CoV pp78 binds tRNA with a low μM affinity	126
3.10.	Structural/functional comparison of pre-processed polyproteins of SARS-CoV and HCoV-229E	127
3.10.1.	Conservation of the Nsps' secondary structure during maturation.....	129
3.10.2.	RNA-dependent RNA polymerase (RdRp) activity assay	132
3.10.3.	Identification of binding epitopes of nucleoside derivatives by STD-NMR spectroscopy	134
3.11.	Interaction of Nsp7 and Nsp8	138
3.11.1.	Oligomerization of SARS-CoV Nsp7 and Nsp8.....	138
3.11.2.	Nsp(7+8) in comparison to its precursor pp78 by SAXS	143
3.11.3.	SARS-CoV Nsp7 forms a dimer in solution.....	150
3.11.4.	C-terminal domain of HCoV-229E Nsp8	153
3.12.	HCoV-229E Nsp9.....	159
3.12.1.	DLS and SEC do not suggest a tight interaction of Nsp9 with Nsp(7+8)	162
3.13.	HCoV-229E Nsp10.....	163
3.13.1.	Purification, folding and oligomeric state.....	163
3.13.2.	Crystallization of HCoV-229E Nsp10	166
3.13.3.	Structural characterization in solution	172
3.13.4.	Impact of Nsp10 on the interaction with HCoV-229E Nsp16.....	175

4.	Conclusion.....	179
5.	Summary	181
6.	Zusammenfassung.....	183
7.	References	186
8.	Appendix	201
9.	Risk and safety statements	207
9.1.	Chemicals (GHS classification).....	207
9.2.	Commercial crystallization solutions	211
9.3.	GHS and risk symbols and information about hazard-, risk-, safety- and precaution-statements	212
10.	Poster presentations and publications	216
11.	Acknowledgement	218
12.	Curriculum vitae	219

Abbreviations

2D	two dimensional
Å	Ångström (unit, 10^{-10} m)
aa	amino acid(s)
AHT	anhydrotetracycline
AMP	adenosine monophosphate
Amp ^R	ampicillin resistance
ANSI	American National Standards Institute
approx.	approximately
APS	ammonium persulfate
<i>aqua dest</i> /dH ₂ O	distilled water
FTIR	Fourier transformation infrared spectroscopy
AUC	analytical ultracentrifugation
BCIP	5-brom-4-chlor-3-indoyl phosphate
BCoV	Bovine coronavirus
BLAST	Basic Local Alignment Search Tool
bp	base pair(s)
BSA	bovine serum albumin
c	concentration
°C	degree Celsius
CD	circular dichroism
cDNA	<i>complementary</i> DNA
CIAP	calf intestinal alkaline phosphatase
CID	collision-induced dissociation
cmc	critical micelle concentration
Cm ^R	chloramphenicol resistance
CoV	coronavirus
d	day(s)
D	diffusion coefficient
DDM	n-dodecyl β-D-maltoside
DEPC	diethyl pyrocarbonate
DESY	Deutsches Elektronen Synchrotron (German electron synchrotron)
DLS	dynamic light scattering
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTPs	2'-deoxynucleoside-5'-triphosphate
ds	double strand (DNA/RNA)
DTT	dithiothreitol
<i>E.</i>	<i>Escherichia</i>
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide

EDTA	ethylene diamine tetraacetic acid
EGCG	epi-gallocatechin gallate
EMBL	European Molecular Biology Laboratory
ER	endoplasmic reticulum
ESI	electrospray ionization
<i>et al.</i>	<i>et alii</i>
EtBr	ethidium bromide
EtOH	ethanol
f.c.	final concentration
FCoV	feline coronavirus
FRET	fluorescence resonance energy transfer/Förster resonance energy transfer
g	gram (unit)
g	relative centrifugal force (rcf) as multiples of the gravitational acceleration on earth
GCB	granada crystallization box
GCG	gallocatechin gallate
gRNA	genomic RNA
GST	glutathione S-transferase
h	hour
HCoV	human coronavirus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOAc	acetic acid
Hz	Hertz (unit)
I	intensity
IBV	infectious bronchitis virus
IPTG	isopropyl-β-D-thiogalactopyranoside
k	kilo (multiplied by 10 ³)
K	Boltzmann constant
K	Kelvin (unit)
K _D	dissociation constant
Kan ^R	kanamycin resistance
kb	kilobase(s)
kDa	kilodalton
l	litre (unit)
LB	Luria Bertani
M	mega (multiplied by 10 ⁶)
M	molar
m	milli- (multiplied by 10 ⁻³)
max	maximum
Mb	mega base pair(s)
MCS	multiple cloning site
MERS(-CoV)	middle-east respiratory syndrome (coronavirus)

MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MHV	murine hepatitis virus
min	minutes
MPD	2-methyl-2,4-pentanediol
M ^{pro}	main protease
MS	mass spectrometry
MW	molecular weight (g/mol or Da)
MWCO	molecular weight cut off
n	nano- (multiplied by 10 ⁻⁹)
N _A	Avogadro's constant
NaOAc	sodium acetate
NBT	nitro blue tetrazolium chloride
NCBI	National Centre for Biotechnology Information
n.d.	not determined
NMR	nuclear magnetic resonance
no.	number
NSD	normalized spatial discrepancy
Nsp	non-structural protein
NTA	nitriлотriacetic acid
OD	optical density (e.g. OD ₆₀₀ , at a wavelength of 600 nm)
ORF	open reading frame
p	pico- (multiplied by 10 ⁻¹²)
pp	polyprotein
PAA	polyacrylamide
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCT	pre-crystallization test
PDB	protein Data Bank
PEG	polyethylene glycol
<i>Pfu</i>	<i>Pyrococcus furiosus</i> (polymerase)
PMSF	phenylmethylsulfonyl fluoride
<i>pNA</i>	<i>para</i> -nitroaniline
Q	quadrapole
R _g	radius of gyration
R _h	hydrodynamic radius
RdRp	RNA-dependent RNA polymerase
RMSD	root-mean-square deviation
RNA	ribonucleic acid
rpm	revolutions per minute
RU	response unit

s	scattering vector (SAXS; depending on angle and wavelength of scattered waves)
s	second(s)
S	Svedberg (unit)
SARS	severe acute respiratory syndrome
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulfate
SEC	size-exclusion chromatography
sec.	secondary
sgRNA	subgenomic RNA
SIB	Swiss Institute of Bioinformatics
SLS	static light scattering
<i>spec.</i>	species
SPR	surface plasmon resonance
ss	single strand (DNA/RNA)
STD	saturation transfer difference
struc.	structure
T	temperature [K]
t	time [s]
TAE	Tris-acetate-EDTA
<i>Taq</i>	<i>Thermus aquaticus</i> (polymerase)
TB	Tris-borate
TCEP	Tris(2-carboxyethyl)phosphine
Tet ^R	tetracycline resistance
TE	Tris-EDTA
TEMED	<i>N,N,N',N'</i> -tetramethylethylenediamine
term.	terminus
TEV (protease)	tobacco etch virus (protease)
TFE	2,2,2-trifluoroethanol
theor.	theoretical
TLC	thin layer chromatography
T _m	melting temperature
TMD	transmembrane domain
TOF	time of flight
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
U	unit (enzyme activity)
UTR	untranslated region
UV	ultra violet
V	volt
v/v	volume per volume
vis	visible

vol	volume
w/v	weight per volume
wt	wild type
XFEL	X-ray free electron laser
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
Y2H	yeast-two-hybrid
ZIGE	Zone-interference gel-electrophoresis
β -ME	β -mercaptoethanol
β -OG	n-octyl- β -D-glucopyranoside
μ	micro- (multiplied by 10^{-6})

Abbreviations of nucleotides

A	adenine
C	cytosine
G	guanine
T	thymine
U	uridine

Abbreviations of amino acids

A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartate
E	Glu	glutamate
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophan
Y	Tyr	tyrosine

1. Introduction

1.1. Coronaviruses

The majority of known viruses possesses an RNA genome including most plant viruses, some bacteriophages and many human and animal viruses, e.g. within the order of *Nidovirales*. The taxon *Nidovirales* is characterized by an envelope and (+)-ssRNA as well as a set of *nested* (latin: nidus = nest) subgenomic mRNAs. The taxon is commonly subdivided into the subfamilies *Arteriviridae* (*Arterivirus*, equine arteritis virus), *Roniviridae* (*Okavirus*, gill-associated virus) and *Coronaviridae* [1]. Members of the *Nidovirales* have a chymotrypsin-like protease in common owing a catalytic dyad consisting of cysteine and histidine. For the phylogeny of the *Nidovirales* a conserved order of homologous replicase domains and transcriptional/(post)-translational strategies are the most important criteria, beside the overall polycistronic genome organization [2]. A unique genetic marker used in identification of *Nidovirales* is a uridylyate-specific endonuclease (NendoU) that allows the discrimination from all other RNA viruses [3].

In 1965, the first coronaviruses were identified and subsequently termed human coronavirus OC43 (HCoV-OC43) and human coronavirus 229E (HCoV-229E), responsible for relatively mild respiratory diseases [4]. Coronaviruses, i.e. the subfamily *Coronaviridae*, gained a much higher public and scientific attention in early 2003, due to the epidemic outbreak of a respiratory disease, which likely had its zoonotic origin in southern China in late 2002 (fig. 1). It spread throughout the whole northern hemisphere in an area spanning 29 countries. Due to the high pathogenicity and the relatively high fatality rate of approximately 10% the World Health Organization (WHO) called the virus severe acute respiratory syndrome coronavirus (SARS-CoV). Among elderly people the case-to-fatality rate reached even 50% [5]. Overall, the symptoms are flu-like with dry non-productive dyspnea, lymphopenia, infiltrate on chest radiography, myalgia and fever exceeding 38 °C [5; 6]. From November 2002 to February 2003 some cases of atypical pneumonia appeared, reported in the beginning of 2003 by the Chinese ministry of Health, which was initially thought to be caused by *Chlamydia pneumoniae*. Most probably a Chinese doctor from Guangdong, who treated patients there, spread the disease while he was resident in a hotel in Hong Kong. The search for a closely related virus originating from bats in animals identified one in palm civets (fig. 1A) and racoon dogs that are gourmet food in this region [7]. The close contact to animals could have promoted the virus to quickly jump to human beings. Recently, the on-going emergence of a novel *Betacoronavirus*, called MERS-CoV (middle-east respiratory syndrome coronavirus) was a severe reminder that coronaviruses are still a major global health threat, requiring perspectives for anti-viral drug development and medicinal preparedness in the future. So far 34 out of 64 cases of MERS-CoV have been confirmed fatal [8]. Most cases are related to a period of residence in the Kingdom of Saudi Arabia. The emergence and its consequences are carefully investigated at the moment. The outbreak of SARS-CoV promoted the

discovery of HCoV-NL63 isolated from young hospitalized patients [9]. Beyond human coronaviruses, highly economically relevant animal coronaviruses like bovine coronavirus (BCoV), infectious bronchitis virus (IBV), transmissible gastroenteritis virus (TGEV), and the sporadically arising feline infectious peritonitis virus strain of feline coronavirus (FCoV) causing a fatal systemic disease [10] should not be unrecognized. All coronaviruses known to date are grouped as either *Alpha-*, *Beta-*, *Gamma-* or *Deltacoronavirus* (or genetic group 1-4), whereas SARS-CoV was commonly considered a basal outgroup of the *Betacoronaviruses*, as determined for instance by the sequence of its RdRp Nsp12 [11] (fig. 1C), originating from bat species. HCoV-229E is a very distantly related *Alphacoronavirus* which is, together with HCoV-OC43, estimated to cause 10-35% of human common cold [12] with only a few cases of severe pneumonia commonly affecting immunocompromised patients [13].

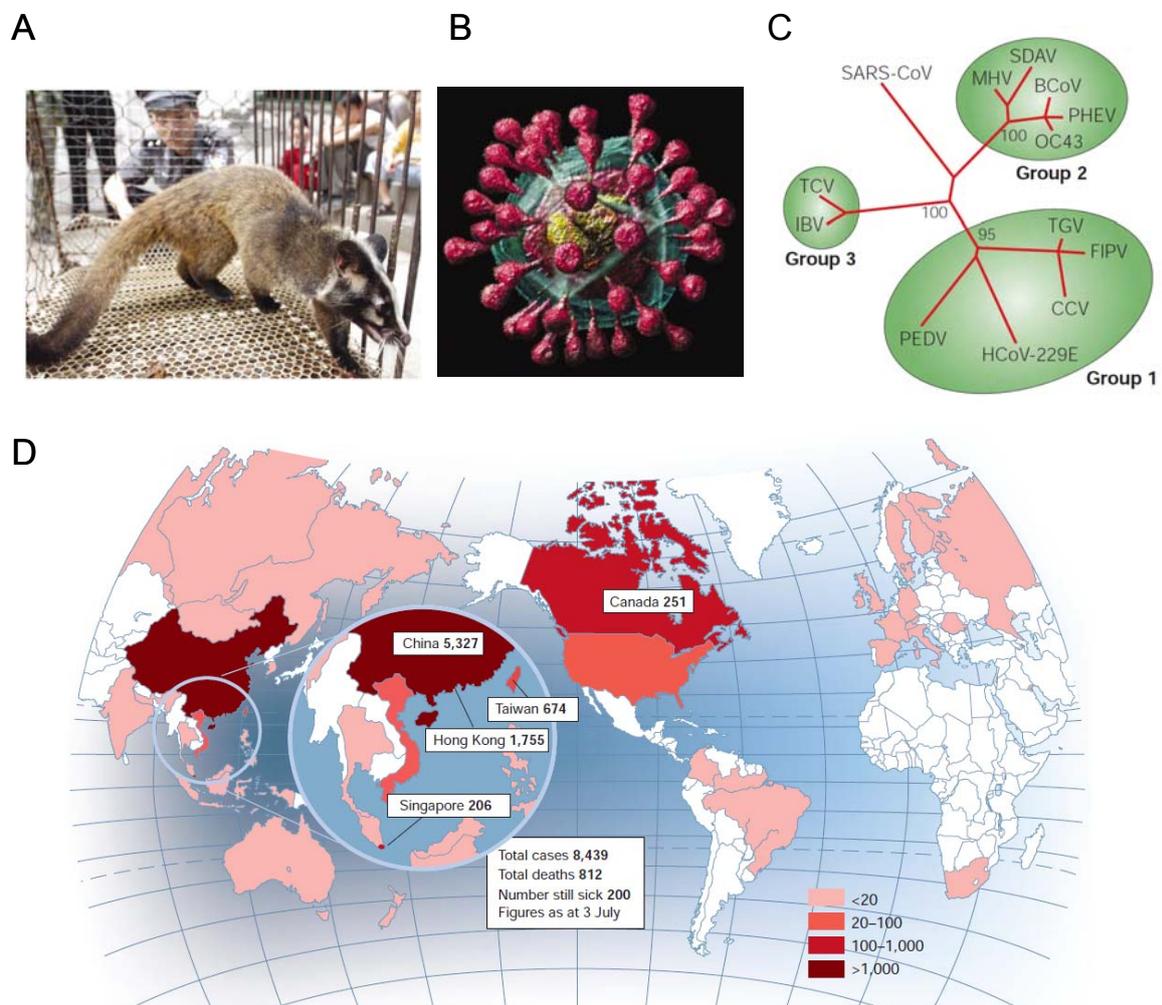


Figure 1. (A) Palm civet are available on Chinese markets (Zhou Guoqiang/ZUMA PRESS; Nature vol. 424, 2003), a putative source of pandemic SARS-CoV; (B) Coloured electron micrograph of a SARS-CoV virion (R. Kightley/SPL; Nature vol. 424, 2003); (C) Initial phylogenetic tree [11] according to the RdRp Nsp12. SARS-CoV was characterized as an outgroup of the *Betacoronavirus* phylum (Group 2); (D) SARS-CoV infections world-wide as reported up to the beginning of July 2003 (WHO; Nature vol. 424, 2003). More than 50% of all cases were reported within China.

1. Introduction

Structurally, *Coronaviridae* are enveloped and contain remarkably large (+)-ssRNA genomes with a 3'-poly(A) tail and a 5'-methylguanosin cap structure [14]. The MHV-A59 (murine hepatitis virus, strain A59) genome consists of 31357 nt. MHV is frequently used as a model organism, primarily because it is maintainable at a low biological safety level. All coronaviruses possess structured *cis*-acting untranslated regions at both ends, essentially involved in viral (continuous and discontinuous) replication [3]. The 5'-UTR (and 3'-UTR) strongly varies in length, whereas even distantly related coronaviruses own a defined pattern of at least five stem loops [15]. It is presumed that the 3'-UTR is similarly heavily involved in replication regulation and carries a hyper variable region (HVR). Interestingly, it contains a highly conserved octanucleotide motif in *Betacoronaviruses* which is assumed to be essential, while the hypervariable region in total was reported to be not essential but to modulate the pathogenicity [16] [see figure 6 for a model of the (-)ssRNA replication initiation]. A few protein interaction partners like the poly(A)-binding protein [17] have been identified. During membrane-associated replication in the cytosol of a host cell minus-strand RNA is formed along with a set of subgenomic (sg) mRNAs. They own a common 3' and 5' sequence and are synthesized by discontinuous replication within 75-90 minutes post infection [18]; their formation is regulated by transcriptional regulatory sequences (TRS). The common 5' end results from base pairing with a six nucleotide leader core sequence. The addition of the leader requires discontinuous replication and a recombination involving nascent (-)-RNA and a leader sequence copy [19]. (-)-RNA acts as a template for those mono- or bicistronic sgRNAs. SARS-CoV provides eight sg mRNAs that give rise to structural and accessory proteins that are discussed below, when translated [20]. Viral proteins involved in transcription and replication are historically called non-structural proteins as they do not participate in structural integrity and in surface determination of the virions. Beside those "replicase proteins" encoded by the two alternative open reading frames ORF1a/ORF1ab, approximately one third of the genome is covered by four major structural proteins (spike (S), envelope (E), membrane (M), nucleoprotein (N)) and interspersing accessory proteins close to the 3' terminus (for review see [11]). In contrast to the other three, the N-protein is located inside the envelope and forms the nucleocapsid along with genomic RNA (gRNA). The preference to specifically incorporate gRNA into nascent particles is accomplished by highly variable *cis*-acting packaging signals requiring a defined RNA secondary structure [21]. The nucleocapsid interacts with the C-terminus of the M protein, as demonstrated for TGEV, to enlarge the nucleocapsid stability [22]. The S protein likely contributes to the infection specificity and to the individual course of the disease, supported by the recently revealed unique interaction of the spike protein (N-terminal receptor binding domain) of the novel MERS-CoV with DPP4 (CD26) at molecular level [23]. The trimeric surface exposed spike protein interacts with a host cell receptor and undergoes a conformational change [24] to facilitate receptor mediated endocytosis that favours a low pH value [25] – angiotensin converting enzyme (ACE2) is the receptor of SARS-CoV. Remarkably, NL63 which is an *Alphacoronavirus* binds to the zinc peptidase ACE2 as well, whereas other *Alphacoronaviruses* commonly bind to the

aminopeptidase CD13. Mechanistic details of the cell entry remain poorly understood. The C-terminal half of the S protein is the membrane anchor. Visually, the spike protein molecules on the surface give rise to the typical corona-like shape (fig.1B); sun corona) of the particles. Additionally, there are accessory proteins, e.g. the ORF6 protein that is discussed below. Some of them have regulatory properties, like influencing the virulence/interferon response [26; 27]. Moreover, 3a and 7a are known to induce apoptosis [28; 29]. At least one transmembrane domain has been suggested for some of the accessory proteins, namely for 3b, 6, 7a and 7b (for review see [11]).

1.2. Polyproteins: pp1a and pp1ab

The coronaviral genome typically encodes 16 non-structural proteins (Nsp), i.e. Nsp1-Nsp16 that were suggested to be involved in the formation of a large hypothetical transcriptase/replicase complex (RTC). As an exception infectious bronchitis virus (IBV) encodes only 15 Nsps. Most probably, the RTC involves host-cell factors as well and is localized at specialized virus-induced lipid bilayer structures in the cytosol, derived from late endosomes or the ER [30].

When (+)ssRNA is released into the cytosol of a host cell, subsequent translation is obligate to obtain the polymerase and other non-structural proteins required for genome replication – most probably ordinary eukaryotic 5'-7-methylguanosine cap-mediated translation. However, the primary translation products of coronaviral ORF1a and ORF1ab are the large polyproteins 1a (pp1a; Nsp1 to Nsp11, \approx 4300 amino acids) and, as a result of a (-1) ribosomal frameshift (RFS), polyprotein 1ab (pp1ab; \approx 7100 amino acids), respectively. The ribosomal frameshift “site” consists of a “slippery” sequence (7 nt) and a short spacer that separates it from an individual downstream frameshift-stimulating pseudoknot [31; 32]. There is still a discussion about the existence of Nsp11 as a protein which is encoded in this region. Due to the low frequency of ribosomal frameshifts the amount of pp1a exceeds that of pp1ab by a factor of three to five. The polyproteins contain several transmembrane domains, e.g. six are predicted for Nsp6, which allow anchoring in ER-derived membranes and protection of the replicase structures [33]. Very recently, mature Nsp3, Nsp4 and Nsp6 were reported to cooperatively induce the formation of specialized double lipid layer vesicles required for the transcription and replication machinery [34]. The membrane spanning Nsp4 and Nsp3 (interacting with Nsp2) had already been suggested before to modify and “specialize” a membrane for replication, studying equine arteritis virus [35] and murine coronavirus Nsp4 [36].

The polyprotein 1ab fragment typically comprising Nsp5 to Nsp16 is sequentially processed by the chymotrypsin-like main protease (Nsp5; M^{pro}, 3CL protease), which releases itself by N-terminal autocleavage. Previously, Nsp1 to Nsp4 are sequentially cleaved by the proteolytic activity of the multi-domain protein Nsp3, with contains one papain-like protease activity in SARS-CoV [37] and

1. Introduction

two (PLP1 and PLP2) in other coronaviruses [20]. The discovery and investigation of the SARS-CoV unique domain (SUD) of Nsp3 that is not encoded by other coronaviruses [38; 39] further promotes the search for unique characteristics [including oligomerization, as discussed below] of Nsps in a certain coronavirus, which could, for instance, explain the high pathogenicity of selected species.

The current model of sequential polyprotein processing consequently results in a certain amount of various polyproteins released in the cytosol of the infected host-cell. The sequence specificity of the main protease is known and results in remarkably different turn-over rates of the respective cleavage site [40]. In a few studies so far pre-polyproteins of different sizes derived from pp1a/1ab were detected *in vivo*. Two large processed products (approximately 240 and 290 kDa in size) of the ORF1a-derived polyprotein of murine hepatitis virus (MHV) were successfully detected [41; 42]. In 1998, several cleavage sites of MHV were predicted and verified as well and a 22 kDa protein fragment of polyprotein 1a (ORF1a) was observed in MHV infected cells using a specific antibody [43]. Furthermore, pulse-chase experiments revealed the existence of initially formed pre-processed fragments of the MHV polyprotein 1a (p28, p72, p250, and p150) that are further processed to form the MHV main protease, p65, p210, p40, p27, and p15. P210 and p40 are predicted to originate from p250. A protein with an approximated mass of 150 kDa may correspond to the fragment Nsp4-Nsp10/11. Interestingly, a fragment of approximately 72 kDa (p72) appears to be unique for the strain JHM of MHV [44; 45]. Similarly, a monoclonal antibody allowed the *in vivo* identification of an HCoV-229E ORF1b-derived processing product of 41 kDa that displayed punctate, perinuclear distribution [46]. All these observations already strongly suggest a potential relevance of native pre-processed polyproteins.

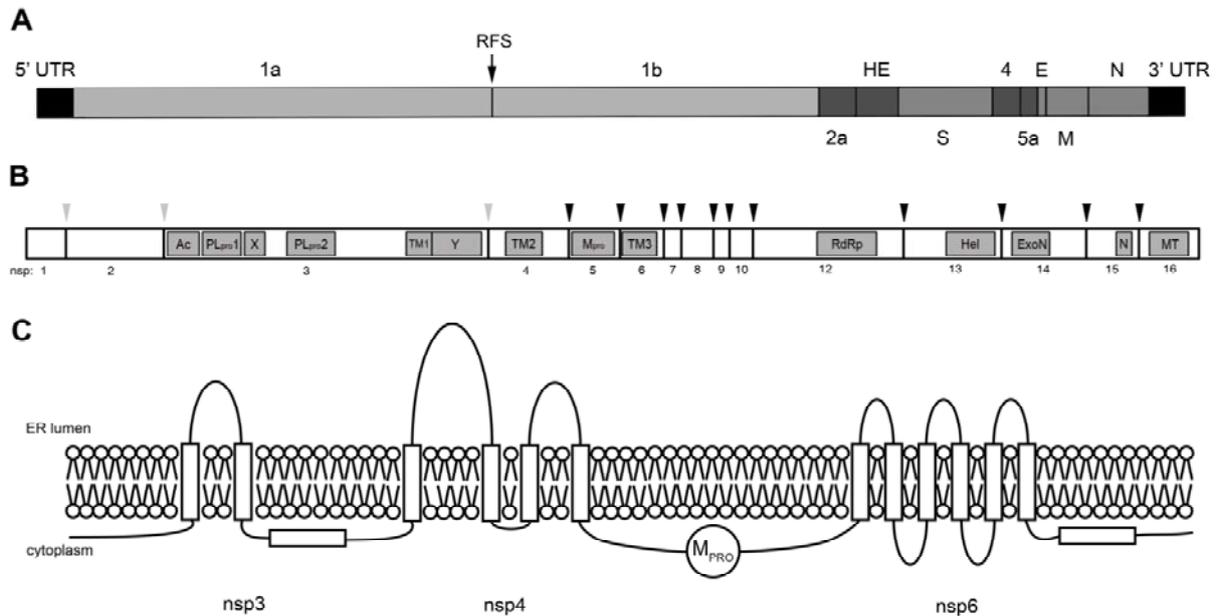


Figure 2. Genome and replicase proteome of coronaviruses. (A) Scheme of the genome of murine hepatitis virus (MHV) strain A59, representing a typical arrangement of open reading frames, and (B) of the translated replicase polyprotein 1ab. (C) Predicted membrane topology explaining the localization of the main protease (M^{pro}) flanked by Nsp4 and Nsp6 in the “pre-cleavage” state [47].

According to a proposed model based on sequence analysis [47] the full-length polyproteins 1a/1ab are anchored by six flanking transmembrane domains (TMD) of Nsp6, along with 4 TMDs of Nsp4 and membrane-spanning Nsp3 (fig. 2) This specific topology of membrane association realizes that both ends of the main protease are exposed to the cytosol, suggesting a *cis*-acting autoproteolytic release of M^{pro} , followed by *trans*-acting cleavage. Interestingly, mutations of Arg4, Glu290 and Arg298 which are critical to form the native dimer of the protease are still capable of N-terminal autocatalytic cleavage [48], even though the dimeric state was seen to be essential for catalytic activity *in vitro* (see section 1.3).

1.3. The coronaviral main proteases (EC 3.4.22.69)

Native SARS M^{pro} forms a catalytically active dimer under a variety of *in vitro* conditions [49-51]. Structurally, it was shown that the dimeric state is essential for catalytic activity, and MD-simulations indicated that, due to an asymmetric fold of both protomers, the two active centres within a dimer cannot be active at the same time [52]. High-resolution structures of the M^{pro} dimer at various pH-values and binding to diverse ligands are available and elucidated the catalytic mechanism and inhibitory potential, e.g. see [53-55]. In addition, the monomeric state was crystallized after stabilization by the mutation R298A [56]. A “super-active” octameric SARS-CoV M^{pro} (pdb code 3IWM), controversially discussed regarding its physiological relevance, represents the third aggregation state that was observed in an M^{pro} crystal [57]. The K_D -value of the dimer formation was

1. Introduction

experimentally determined with highly varying results between 190 nM and 227 μ M [58; 59]. However, different complementary approaches indicated a tendency for the low μ M range [60]. The concentration-dependent dimerization may be hypothesized to provide a mechanism to regulate the polyprotein maturation during infection.

The SARS-CoV M^{pro} enzyme owns a catalytic dyad (C145 and H41) similar to other coronavirus main proteases, reflecting the typical catalytic cycle of a cysteine protease [61]. In contrast to serine proteases that harbour a catalytic serine, the cysteine thiol group is more acidic and the thiolate moiety generated in the first step is consequently a stronger nucleophile. Experimentally approximated pK_a-values for C145 and H41 imply that the free “resting” enzyme prefers the neutral state rather than the zwitterions state [62]. An oxyanion hole is formed after nucleophilic addition, targeting the carbonyl carbon atom of the peptide bond. The optimal pH range of M^{pro} is between pH 7.3 and pH 8.5. This was explained by a pH-dependent cooperative movement of several side chains hindering the substrate-binding and resulting in a collapsed oxyanion hole at pH 6.0 [53]. A single protease protomer consists of the domains I-III (residues 8-101/102-184/201-303), dimerizing via the five α -helix containing domain III. Domains I and II form an anti-parallel β -barrel, the active centre is located at its interface. This chymotrypsin-like fold is also present in other viral proteases [63]. Nonetheless, the cysteine of a catalytic dyad (instead of serine in chymotrypsin), domain III that is unique to viruses and the essential function of the main protease render it as a prominent drug target. The highly similar substrate specificity of different CoV M^{pro} homologues and the high level of sequence identity resulted in much stronger efforts to design M^{pro}-specific inhibitors compared to that on papain-like proteases. Aligning HCoV-229E M^{pro} (pdb code 1P9S) and SARS-CoV M^{pro} (pdb code 1Q2W) results in a low RMSD of 0.889 Å as determined by the PyMOL Molecular Graphics System (version 0.99; Schrödinger, LLC). A glutamine residue is the ubiquitously conserved and essential amino acid in P1 position of the recognition/cleavage site(s) of a substrate, i.e. coronaviral polyproteins. Glutamine in P1 position is common even for proteases of other RNA virus taxa (fig. 3). A similar catalytic activity for histidine at P1 position was indicated by a single study that uses tetrapeptide substrates [64]. In P1' position, a small hydrophobic amino acid is preferred, while only minor preferences have been observed at the other flanking positions, e.g. Leu in P2 position [40; 65] as determined by modified peptide substrates. It has to be mentioned that several (artificial) substrates varying in length and modification were applied for these investigations, which impairs an accurate comparison of the results.

protease	P5	P4	P3	P2	P1	P1'	P2'
SARS-CoV M^{pro}	S	A	V/T/K	L	Q	A/S	G
CVA16 (coxsackievirus)	E	A	L	F	Q	G	P
EV71 (enterovirus)	E	A	V/L/T	L/F	Q	G	P
NV (norwalk virus)	D/E	F/Y	H/Q/E	L	Q	G	P

Figure 3. Alignment of consensus recognition sites of main proteases in the picornavirus-like protease super cluster (adopted from [66]). A strictly conserved glutamine residue in P1 position and a preference for L or F in P2 position is obvious for structurally distantly related proteases in the large picornavirus taxon.

Numerous chemically diverse inhibitors of viral main proteases (3C/3CL protease) have already been reported. Quercetin, epigallocatechin gallate (EGCG), and gallic acid, all representing flavan derivatives from the secondary plant metabolism, displayed IC_{50} -values in the moderately low μ M range with SARS-CoV main protease expressed in *P. pastoris* [67]. The galloyl ester moiety of EGCG and GCG is critical for the inhibition magnitude. Tannic acid consisting of multiple galloyl moieties that form a polyester with a glucose ring was reported to inhibit with an IC_{50} value of 3 μ M as well [68] and is associated with health benefits [69]. Flavonoids are well known inhibitors of a variety of enzymes [70] and more than 4000 of the typically antioxidant flavonoid compounds were identified in nature [71]. Interestingly, a group of related geranylated plant flavan derivatives also displayed inhibitory activity against the papain-like protease of SARS-CoV [37].

The modified tripeptide Z-Val-Leu-Ala(pyrrolidone-3-yl)-2-benzothiazole that displayed a K_i -value of 4.1 nM recently served as a starting point to design effective dipeptide lead compounds covering an epitope that comprises the positions S1' [binding of the benzothiazole moiety] to S3 [72; 73]. The high conservation of the S1 subsite and strong similarities of the S2 and S4 sites, as recently also revealed for the novel MERS-CoV, suggest that many characterized inhibitors, e.g. the peptide-like broad-spectrum inhibitor N3, may be applicable for related newly emerging coronaviruses in the future [74]. A potent antiviral activity against IBV in chicken embryos was already confirmed for N3 [75]. It undergoes a Michael addition reaction, forming a covalent bond between the catalytic cysteine and a vinyl group of the inhibitor. In contrast, another inhibitor AG7088 displays specificity for rhinovirus main proteases and does not inhibit coronaviral main proteases, while analogue compounds can inhibit them [76; 77]. A synthesized macrocyclic inhibitor was proven to inhibit not only the 3CL protease of coronavirus, but also the homologue 3CL proteases of norovirus and enterovirus (CVB3 Nancy strain) with similar IC_{50} -values in the low μ M range. Macrocyclic compounds were associated with beneficial drug-like characteristics, enhanced protease stability and cellular permeability. Furthermore, the binding affinity is enhanced due to the minimization of entropy loss upon binding of such cyclic compounds to its target enzyme [66]. Peptides derived from the C-terminus of the cleavage site ending with a glutamine and modified by an aldehyde group were successfully applied to co-

crystallization trials, revealing the formation of a reversible thiohemiacetal involving the catalytic sulfhydryl group. Interestingly, the property to accommodate a large aspartate residue well in the hydrophobic S2 pocket was observed as well [78]. This is not obvious when analyzing the P2 preference of the protease and a previous crystal structure suggested that such an aspartate residue of a modified peptide inhibitor would face out of the pocket [53]. In summary, a large number of “warheads” has been developed to act as an electrophile attacked by the active site cysteine. Inhibition experiments used epoxides [79], α,β -unsaturated esters [76], nitriles [80], aldehydes [78], benzotriazol esters [81] and others, inhibiting the protease in a wide range of IC_{50}/K_i -values.

1.4. Mature Nsps and their interaction partners

The knowledge about structural and functional properties of individual mature CoV Nsps significantly increased within the recent years. A high-resolution structure of TGEV Nsp1 was, for example, solved by X-ray crystallography [82]. This protein is known to suppress eukaryotic translation at different stages [83]. Likewise, Nsp2 of avian infectious bronchitis virus (IBV) was successfully crystallized, even though its function is completely unknown so far [84]. Fragments of the multi-domain Nsp3 were structurally characterized as well, the FCoV X domain [85] and the SARS-unique domain (SUD) that is binding oligonucleotides forming G-quadruplexes and is discussed to participate in the host cell's response to the viral infection [38; 39]. The identities of many non-structural proteins are widely well established, figure 4 provides a systematic genome/proteome overview. In terms of anti-coronavirus drug development natural compounds capable of inhibiting the helicase Nsp13 were identified in a FRET-based assay [86] comprising an approach to target coronaviruses independent of the proteases.

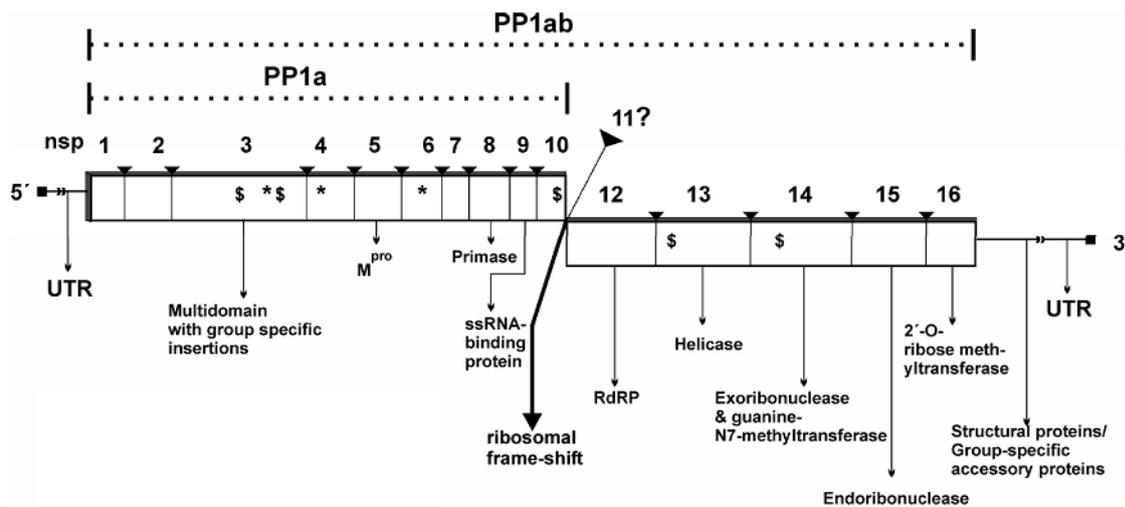


Figure 4. Schematic organization of the coronavirus genome and function of some of the derived mature Nsps (adopted from Dr. R. Ponnusamy, dissertation, 2009). [\$: Cys/his binding site that may allow metal-ion binding; *: predicted transmembrane domain(s)]

Up to now it is highly difficult to elucidate and describe the complexity of the protein-protein-interaction network that involves the mature non-structural proteins as well as most probably their polyprotein precursors as a putative “core” structure of a dynamic RTC. Other coronaviral proteins or, most importantly, a set of host-cell proteins are suggested to be additionally involved. Nevertheless, some efforts were spent on proteome-wide screening of coronaviruses for putative interactions applying *in vivo* two-hybrid methods [87; 88], essentially summarized in figure 5. Previously, some interaction partners of Nsp2, namely Nsp3, Nsp6, Nsp8, Nsp16 and ORF 3a were verified independently by co-immune precipitation (CoIP) [89; 90]. Only for ORF9b itself, already 16 interaction partners within the SARS proteome were identified using Y2H methods, while only four of them were confirmed by CoIP [87]. CoIP was also applied to verify the interaction of the hexameric endoribonuclease Nsp15 [91] with the retinoblastoma tumor suppressor protein which can affect coronavirus infections [92]. In general, a powerful complementary technique to confirm interactions might be the *in vivo* investigation of co-localization of the involved molecules by fluorescence microscopy, presuming that the proteins are labelled. Taken together, although a few specific and reasonable Nsp-Nsp-interactions have been detected and discovered in more detail (see discussion below) most of the putative interactions remain enigmatic or non-specific.

SARS-CoV Nsp8 and the corresponding Nsp(7+8) complex, 5'-(G/U)CC-3' was shown to be a sufficient template for synthesis of a complementary oligonucleotide with low fidelity [94].

Recently, the RdRp activity of the SARS-CoV Nsp(7+8) complex was shown not only to elongate RNA primer, but also to synthesize RNA *de novo* [97]. The polymerase activity and oligomerization seem to depend on the position of the purification tag within Nsp8, indicating that an authentic end or a “mimic” of an incomplete proteolytic processing is critical for the regulation of the Nsp8 polymerase activity. Structurally, the shape of a hexadecameric (8:8) Nsp7-Nsp8 interaction state forming a hollow cylinder core was in discussion to act as a processivity “clamp” binding dsRNA, based on the corresponding structure solved by X-ray crystallography [98]. This suggestion is supported by the dimensions of the central channel as well as an accumulation of positively charged residues around it which was identified to be critical for RNA binding *in vitro* [98]. Despite the high-resolution structures that are available up to now, the detailed mechanisms of RNA-binding and of the specific catalytic activity of the Nsp(7+8) complex remain at least partly enigmatic. For Nsp8 the common motifs “A” and “C” that are usually indicative for RdRp activity are missing [94]. Referring to the traditional nomenclature of polymerases, the motifs A-D within the palm subdomain of RNA polymerases are highly conserved, even among viruses [95]. On the other hand, a (D/E)-X-(D/E) motif that is additionally well-known for DNA-dependent RNA polymerases is conserved at position 50-52 as well as at position 161-163. A replacement of either D161 or D163 by alanine resulted only in minor effects on the activity of SARS-CoV Nsp8, in contrast to the significant effects reported for the mutations D50A, D52A and K58A [97]. Considering a major importance of the residues 50 and 52 for the catalytic activity, a proper substrate-accessible orientation is not obvious, since the residues point in opposite directions of a helix, e.g. in FCoV [96]. In theory, this could be overcome by either conformational flexibility in this region or a dimerization mode to complement such an active centre involving residues of two protomers. Another high-resolution structure of Nsp7 interacting with an N-terminally truncated isoform of Nsp8 suggested a mechanism for the truncated Nsp8 to interfuse into a larger complex to regulate RNA replication [99]. In contrast to Nsp8 the structure of SARS-CoV Nsp7 without its interaction partner was revealed by NMR spectroscopy, indicating remarkable differences of length and orientation of the helices at different pH values [100], which has not yet been associated to a physiological function.

The hypothetical RTC Nsp(7+8) element described above might temporarily associate with postulated oligomers of other Nsps, e.g. Nsp9 [101], and/or of host-cell factors that remain to be identified. Nsp9 is an RNA- and DNA-binding protein lacking nucleotide sequence specificity. K_D -values in the μM range were determined for SARS-CoV and HCoV-229E Nsp9 [102-104]. The fold of Nsp9 from HCoV-229E is related to the OB fold (oligonucleotide/oligosaccharide-binding module, which is common for several proteins that bind single-stranded nucleic acids, e.g. in bacteria [105] and

1. Introduction

in viruses [106]. It was shown that SARS-CoV Nsp9 dimerizes *in vivo* involving a loop with a GXXXG motif. The Nsp9 dimerization was reported to be essential for efficient viral growth [104].

For the polyprotein pp78, an essential unknown function was reported before cleavage of the covalent bond between Nsp7 and Nsp8 by M^{pro} [107] that crucially requires the correct order of Nsp8 encoded in the coronaviral genome. Exchanging the position of MHV Nsp7 with that of Nsp8 resulted in a mutant that was not viable. In contrast, pp910 that carries a mutation that prevents cleavage resulted in a viable phenotype, indication that this specific processing event is not important for viral replication [107]. Recently, an enzymatic function was revealed for the large pre-processed polyprotein comprising Nsp7-Nsp10 (pp710) for the first time. Pp710 of HCoV-229E was shown to have primase activity [96]. Additionally, a functional RTC likely contains further protein interaction with Nsp8. A model of negative strand RNA replication initiation mediated by Nsp8 was postulated by [108] (fig. 6). The accessory protein ORF6 was described as an interacting partner of Nsp8 in yeast two-hybrid experiments, co-immunoprecipitation and co-localization *in vivo*. Therefore, the ORF6 protein is suggested to be a replication-related link between an Nsp and an accessory protein to modulate virulence and accelerate the replication [26; 27; 109]. Even though some of the accessory proteins are poorly understood, the ORF 3b protein could be another link to the non-structural proteins encoded by ORF1a, as an interaction with Nsp8 was suggested by a proteome-wide mammalian two-hybrid screen [88].

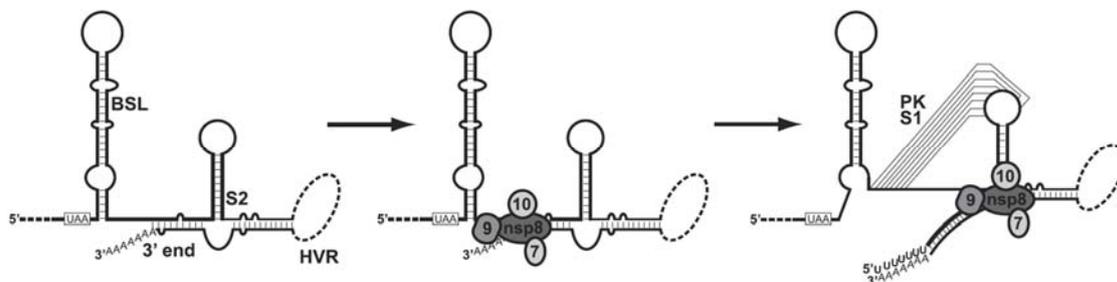


Figure 6. Schematic model of the negative strand RNA replication initiation and elongation according to [108] (modified), precisely regulated by RNA secondary structure and Nsps: In an early stage, a „core unit“ of the replicase complex including Nsp8 is binding to the 3'-UTR of the viral genome close the poly(A) tail and a stem loop structure (BSL). This event results in the formation of a pseudo knot (PKS1) of a basal part of the BSL hairpin S2, while the original base pairing of BSL in this area is disrupted. The new PKS1 structure allows binding of additional (viral-)proteins and allows initiation of negative strand synthesis.

1.6. RNA capping – interaction of Nsp10 with Nsp14 and Nsp16

For sufficient export of mRNA from the nucleus, initiation of translation, and stability a three to four step capping mechanism of the 5'-end is highly conserved in eukaryotes. A similar modification

of coronaviral mRNA was first identified analyzing MHV [110]. For SARS-CoV a conserved 2'-*O*-methyltransferase (Nsp16) was identified, initially by sequence analysis, allowing methylation of the ribose of a terminal adenosine to complete the RNA cap. However, it was shown that the transferase activity requires a rigid interaction with Nsp10 to allow Nsp16 to bind m⁷GpppA-RNA and the co-factor SAH, i.e. demethylated S-adenosyl methionine. Furthermore, the unspecific and relatively weak binding of Nsp10 to various nucleic acids [111] fits the suggestion that Nsp10 can enhance the RNA affinity and prolongs the respective binding groove of Nsp16. In addition, the dependence of the Nsp16 activity on bivalent cations, preferably Mg²⁺, likely provides another regulatory mechanism that is not usual for methyltransferases in general [112]. A core structure of SARS-CoV Nsp10 (amino acid 42-120) was shown to be sufficient for activating Nsp16, which is significantly larger than required to obtain the entire interaction affinity. In contrast, short peptides derived from Nsp10 were reported to have an inhibitory effect [113]. This effect was shown for GGASCCLYCRCH (position 69-80) and FGGASCCLYCRCHIDHPNPKGFCDLKGKY (position 68-96), which are both examples for replication-inhibitory compounds against SARS-CoV derived from a full length Nsp.

The Nsp10/Nsp16 interaction was initially identified by a yeast-two-hybrid screening and verified in detail by X-ray crystallography of the SARS-CoV complex including SAH [112; 114]. In contrast to Nsp16, another interaction partner of Nsp10, which is Nsp14, owns both guanine-*N*7-methyltransferase (for RNA capping) and exoribonuclease (ExoN) activity. The feature of both activities in close proximity within one Nsp is unique to coronaviruses [115]. The DEDD-motif of Nsp14 is highly conserved among the ExoN superfamily. Nsp14 is capable of hydrolyzing ssRNA and dsRNA to produce fragments of 10 nt or less and was therefore suggested to participate in minus-strand discontinuous transcription and recombination. It was further demonstrated that SARS-CoV Nsp14 ExoN activity is enhanced with Mg²⁺ as a cofactor, which is preferred compared to Mn²⁺ or Zn²⁺ [116].

Recently, a homology model of Nsp14 was published combined with site-directed mutagenesis of key residues of Nsp14; inter alia a conserved DXGXPXA motif (amino acid 331-338) to bind the co-factor SAM was revealed [115], whereas high-resolution data is not available. Again, in the presence of Nsp10 the ExoN activity was shown to be increased (20-fold at equimolar ratio), while it did not affect the transferase activity of Nsp14. The Nsp10 mutants G69A, H80A and D82A displayed a significantly reduced affinity towards Nsp14 and resulted in a strong decrease of Nsp14 ExoN activity compared to the wild type complex [117].

A mutagenesis screening for essentially conserved residues within SARS-CoV Nsp10 [118] revealed a stable core that is resistant to mutations. However, different mutations in the periphery of both Zn-binding sites resulted in non viable phenotypes. Overall, an essential function of Nsp10 in

RNA synthesis was observed *in vivo*. Nsp10 is suggested to be a common mediator, e.g. associated with RNA mismatch excision, while an enzymatic function is still unknown. By regulating Nsp16, Nsp10 is also indicated to be involved in host cell immune evasion processes [119]. In addition to the interaction with Nsp16 and Nsp14, some regulatory host cell interaction partners of SARS-CoV Nsp10 have been suggested as well, including the transcription factors BTF3, ATF5, NADH4L and cytochrome oxidase II [120].

1.7. Advanced methods in structural biology:

Structure, function and oligomerization

The application of X-ray crystallography, which is well established since many years to obtain high-resolution structures from biomolecules, critically requires the growth of a crystal of sufficient size and quality. Within the recent years, significant improvements have been made in the field of X-ray crystallography, particularly by applying highly brilliant X-rays from 3rd generation synchrotrons or from the recently commissioned pulsed X-ray free-electron lasers (FEL). In 2012, a high-resolution protein structure was solved for the first time by exposing small crystals (10 μm^3 average volume) spontaneously grown within living *Sf9* insect cells with FEL pulses. A suspension of hundreds of thousands of *TbCatB* crystals that have so far been considered useless for crystallographic analysis due to their small size was sprayed into the laser beam using a liquid jet and successfully produced high-resolution diffraction data that can be combined into a three-dimensional dataset [121; 122]. Thus, the extreme energy within the ultrashort femtosecond pulses of an FEL enables the diffraction data collection from protein crystals down to a nanometre size, without any indication for radiation damage [123]. This new approach is named serial femtosecond crystallography (SFX) [124]. Crystal growth *in vivo* has the advantage that large amounts of tiny crystals, as required for FELs, can be easily produced by enlarging the cell number. Therefore, *in vivo* crystal growth suits very well for FEL-based SFX. In addition, microfocus beamlines at 3rd generation synchrotrons are already able to produce diffraction data from crystals in the low μm size range.

Beside the crystallographic techniques also solution scattering methods (e.g. small-angle X-ray scattering, SAXS) have been benefited from third-generation synchrotron sources, e.g. PETRA III (DESY, Hamburg). The possibility to study a protein, including its shape, flexibility and variability (e.g. depending on pH, temperature or a ligand) in solution in a wide range of masses (1 to 100 nm particle diameter) represents the general potential of SAXS. Usually, the dependence of the intensity of scattered hard X-rays from the scattering angle is recorded. Technical improvements including the high brilliance of PETRA III and the beam focussing allow obtaining improved structural details of a molecule's shape and reduce the experimental time and sample consume significantly. SAXS techniques are additionally applicable for the characterization of advanced materials like

nanocomposites. In this context, metal-coating on the surface of large tobacco mosaic virus (TMV) particles that may serve as a template for nanowire formation was successfully analyzed by SAXS [125]. Moreover, SAXS provided insights into the 480 kDa dodecameric DnaB/C “helicase loader” complex of *E. coli* [126] as well as in sugar-induced dynamic conformational changes in the tertiary structure of the botulinum toxin complex [127].

A complementary method to study the oligomeric state of biomolecules represents native mass spectrometry that recently gained more attention in protein biochemistry. Mass spectrometry in general was always an essential technique, particularly in organic chemistry. The discovery of electrospray ionization was honoured with the chemistry Nobel Prize in 2002 (J. Fenn & K. Tanaka); the Nobel lecture was entitled “Electrospray Wings for Molecular Elephants”. At low speed, ESI actually represents a structure-conserving method for ionization and transfer into the gas phase that allowed the determination of the oligomeric composition of a considerably huge TMV particle at high mass accuracy. Remarkably, the ionization did not alter the infectivity of these particles [128]. To highlight the impact of mass spectrometry for biochemistry it should be mentioned that it is indispensable to analyze labelling of proteins (e.g. H-D-exchange), small tightly bound ligands, and simple cross-linking reactions to obtain structural information. Various chemical reagents are available to covalently link primary amino groups (e.g. by forming a Schiff base or an amide bond) as well as thiol groups or carboxy groups, allowing to analyze the oligomeric state and eventually identify the reaction site within a biomolecule. Recently, *ion-mobility mass spectrometry* established a new dimension of ion separation in mass spectrometry. Determining the drift-speed of an ion in a gas-filled chamber of the spectrometer ions with similar m/z -ratio in complex protein/peptide mixtures can be separated. The drift time strongly depends on the shape and the collision cross section Ω of a molecule (for review see [129]). Recently, the tertiary structure of the tetradecameric GroEL (around 800 kDa) was studied by a combination of surface induced dissociation (SID) – alternatively to collision-induced dissociation (CID) – and ion-mobility mass spectrometry, revealing heptameric building blocks [130].

Within this thesis, the advanced and optimized techniques for structural biology, e.g. SAXS and crystallographic methods, were combined with well-established conventional methods, e.g. saturation transfer difference NMR spectroscopy, surface plasmon resonance (SPR) spectroscopy, and mass spectrometry, to provide broad structural and functional insights into the putative coronaviral RTC, focussing on Nsp7-10 and its precursor pp710.

1.8. Aims of this project

Different coronaviruses are still emerging or expected to potentially emerge in the future. As reported for SARS-CoV and MERS-CoV, the highly pathogenic ones provide or will provide a major

1. Introduction

world-wide health thread particularly to mammals including humans. Although coronaviral main proteases have been successfully exploited as first drug targets, these compounds did not reach the market so far. Thus, the current lack of therapies to treat coronaviral infections promotes the need to provide new perspectives and additional potential drug targets. Different mature non-structural proteins (Nsps) as well as their pre-processed polyprotein precursors are considered to constitute the essential viral RTC complex. However, to target the assembly of the RTC for therapeutic purpose strongly requires a basic functional and structural knowledge about the involved proteins, their oligomerization states, and their specific interactions.

In this study, polyprotein Nsp7-10 (pp710) is selected for functional and structural investigation, since its Nsp domains are suggested to be involved in the RTC formation. Investigating the pp710 processing by SARS-CoV main protease in terms of protease specificity, oligomerization of the intermediate products, and interaction of the partly and fully processed proteins will result in new insights about the polyprotein maturation. This finally promotes understanding of the architecture of the large hypothetical coronaviral RTC. For instance it was speculated about the impact of Nsp10 on Nsp7, Nsp8 and Nsp9 and the authentic oligomeric state of Nsp10, e.g. [131]. Nsp7-10 comprises the C-terminal part of pp1a which is assumed to be cleaved off from the large transmembrane Nsp6 by Nsp5 with a high turn over rate and is therefore considered physiologically relevant. Deleting either Nsp7 or Nsp8 or Nsp9 or Nsp10 in MHV-A59 has been revealed to be lethal [107]. This project largely addresses a comparative investigation of *Alpha*- (HCoV-229E) and a highly pathogenic *Betacoronavirus* (SARS-CoV). Purification strategies that were developed to purify mature Nsps and polyproteins are assumed to be at least partly applicable to other coronaviruses that may emerge in the future. Along with the structural aspects, potential functions of pre-processed polyproteins during proteolytic Nsp maturation are poorly understood or even unknown. Remarkably, different oligomeric complexes of Nsp7 interacting with Nsp8 were revealed by investigating different coronaviruses. Those crystallographic high-resolution structures are worth to be further investigated in solution applying e. g. SAXS and mass spectrometry. In the course of this thesis, the oligomeric states of mature Nsps remain to be compared to the respective precursor of different coronaviruses to understand the composition of RTC building blocks. Basic knowledge about structural details of mature Nsps compared to their precursors, particularly involving coronaviruses that are only distantly related to SARS-CoV, is expected to provide perspectives to interfere with the function/oligomerization of an Nsp by anti-viral compounds in the future. For instance Nsp10 is essentially involved in different protein interactions and might be a valuable target aimed to be characterized. Furthermore, starting points to inhibit polyprotein processing and a detailed understanding of the structural basis of inhibition by substrate-like inhibitors and its consequences are highly appreciated. Particularly, the understanding of the monomer-dimer equilibrium of the main protease (Nsp5) that might depend on the presence of a substrate remains incomplete [56; 132].

2. Material and methods

2.1. Material and devices

2.1.1. Devices

Table 1. Selected devices

centrifuge	centrifuge 5804R/5810R/5415R/5424 (Eppendorf, Germany) centrifuge Minispin [®] Plus (Eppendorf, Germany) Optima TL ultracentrifuge (Beckman Coulter, USA)
thermomixer	Thermomixer comfort (Eppendorf, Germany)
spectrophotometer	GeneQuant 1300 (GE Healthcare, UK) Nanodrop 2000c (Thermo Scientific, Peqlab, Germany) GENios microplate reader (Tecan, Schweiz)
incubator	37 °C incubator Kelvitron [®] T (Thermo scientific, USA) 4 °C incubator (Rubarth, Germany) 20 °C incubator (Rubarth, Germany)
(orbital) shaker	IRC-1-U (Adolf Kühner AG, Switzerland) Innova [®] 43/43R (New Brunswick Scientific, USA) Innova [®] 4330 (New Brunswick Scientific, USA) GFL 3017 (GFL, Germany)
sonifier	Branson Sonifier 250/450 (Emerson Electric Co, USA)
balance	TE3102S (Sartorius AG, Germany) LP224S-0CE (Sartorius AG, Germany)
pH meter	SevenEASY (Mettler Toledo, USA)
FPLC	ÄKTA Purifier P-901 (GE Healthcare, UK) with fraction collector and UV detector
micropipette	Micropipette Research (Eppendorf, Germany)
microwave	microwave MR-6450 (Hitachi, Japan)
freezer (-20 °C)	Liebherr premium (Liebherr, Germany)
SPR spectroscopy	Biacore T-1000 (Biacore, Germany)
thermocycler	Mastercycler [®] gradient, Mastercycler [®] personal (Eppendorf, Germany)
hot-plate magnetic stirrer	VMS-A (VWR, USA) MR 3001 (Heidolph, Germany)
CD spectrometer	J-815 (Jasco, UK)
electrophoresis power supply	EV 231 (Peqlab, Germany) Power PAC 200 (Bio-Rad, Germany)
UV-light source	CrystalLIGHT 100 (Nabitec, Germany)
pipetting robot	Honeybee 961 (Genomic Solutions, USA) Lissy (Zinsser, Germany)
DLS instrument	SpectroSIZE 300 (Xtal-Concepts) Spectro Light 500 (Xtal-Concepts)
acryl amide gel chamber	SE275 (Hoefer)
western blot transfer chamber	TE 22 Mini Tank Transfer Unit (GE-Healthcare)
agarose gel chamber	SE260 Mighty Small II Deluxe Mini electrophoresis unit (Hoefer)
crystal plate incubator	RUMED 3001 (Rubarth, Germany) incubators
crystal imaging	CrystalScore (Diversified Scientific Inc., USA) microscope SZX12 with camera DP10 (both Olympus, Japan)

2.1.2. Plastic consumables

Reaction tubes and pipette tips were obtained from Sarstedt (Germany) and Eppendorf (Germany)

2. Material and methods

2.1.3. Plasmids and oligonucleotides

Table 2 and 3-6 list the plasmids and oligonucleotides used for cloning in the course of this project. The commercial primers T7/T7 terminator, IBA fw/rev, pFastBac fw/rev were usually used for sequencing as described later. Table 7 contains the desoxyoligonucleotides to obtain the corresponding ssRNA by *in vitro* T7 transcription. Strains of *Escherichia coli* used to maintain plasmids and to synthesize the proteins of interest are described in table 8 and were cultivated in either of the two media from table 9.

Table 2. Plasmid table

plasmid	size	features	supplier
pRSETA	2897 bp	amp ^R , enterokinase cleavage site, N-term. (his) ₆ -tag	Invitrogen
pGEX-6p-1	4984 bp	amp ^R , Prescission TM protease site, N-term. GST-tag	GE Healthcare
pET-11d	5677 bp	amp ^R	Agilent
pASK-IBA 3plus	3247 bp	amp ^R , C-term. Strep-tag	IBA
pASK-IBA 33plus	3250 bp	amp ^R , C-term. (his) ₆ -tag	IBA
pASK-IBA 5plus	3260 bp	amp ^R , N-term. Strep-tag	IBA
pASK-IBA 35plus	3263 bp	amp ^R , N-term. (his) ₆ -tag	IBA
pASK-IBA 16	3335 bp	amp ^R , TEV protease cleavage site, ompA signal peptide, N-term. strep tag,	IBA
pFastBac1	4775 bp	amp ^R	Invitrogen
pFastBacHTb	4856 bp	amp ^R , (his) ₆ -tag	Invitrogen

Table 3. Primer table I (HCoV-229E), cloning into pRSETA

name	primer (5' → 3')	length
SF_Ensp7_2_fw	GCGGATCCTCCATCTTGCACTCTAAATTGACTGATCTTAAGTG	43
SF_Ensp7_rev	GCGAATTCGCATACCAACAAAAGAAGATG	29
SF_Ensp8_fw	GCGGATCCCTTATTTTGAAGCACTCCATT	31
SF_Ensp8_re	GCGAATTCATCTTGCCCGCATTATTTTCATT	31
SF_Ensp9_fw	GCGGATCCTTTGGCTCTGATTTTGACTTGT	31
SF_Ensp9_re	GCGAATTCCTAAACAGCAAAAAGAACAATGTGT	32
SF_Ensp10_fw	GCGGATCCAGGAGAGGTGCTGTTTTGGGTAC	32
SF_Ensp10_rev	GCGAATTCGATCTTAATGGTGATGATGGTGATG	33
SFpK16_N8C_E_fw	ATGGTAGGTCTCAGCGCCGACACTATCCTTAATATGGCACGT	42
SFpK16_N8C_E_re	ATGGTAGGTCTCATATCACTGCAATTTAACGACACGTTTACACA	42

Table 4. Primer table II (SARS-CoV), cloning into pASK IBA vectors

name	primer (5' → 3')	length
3+Nsp7 F	ATGGTAGGTCTCAAATGTCTAAAATGTCTGACGTAAAGTGCA	42
3+Nsp8 F	ATGGTAGGTCTCAAATGGCTATTGCTTCAGAATTTAGTTCTTT	43
3+Nsp8 R	ATGGTAGGTCTCAGCGCTCTGTAGTTAACAGCTGAGTTGGC	42
3+Nsp9 R	ATGGTAGGTCTCAGCGCTCTGAAGACGTACTGTAGCAGCTA	41

3+Nsp10 R	ATGGTAGGTCTCAGCGCTCTGCATCAAGGGTTCGCGGAG	39
5+Nsp7 F	ATGGTAGGTCTCAGCGCCTCTAAAATGTCTGACGTAAAGTGCA	43
5+Nsp8 F	ATGGTAGGTCTCAGCGCCGCTATTGCTTCAGAATTTAGTTCTTT	44
5+Nsp8 R	ATGGTAGGTCTCATATCACTGTAGTTAACAGCTGAGTTGGC	42
5+Nsp9 R	ATGGTAGGTCTCATATCACTGAAGACGTACTGTAGCAGCTA	41
5+Nsp10 R	ATGGTAGGTCTCATATCACTGCATCAAGGGTTCGCGGAG	39

Table 5. Primer table III, cloning into pFastBac HTB/ pFastBac1

name	primer (5' → 3')	length
SFpB1_N8E_fw	gcggatccATGAGTGTTCATCTTCTTTTGGTGGTA	36
SFpB1_N9E_re	gcgaattcTTATTTGCCAGCTTGCAATCTCACAGT	35
SFpB1_N10E_re	gcggatccTTGGTCACAACCTACAGCCATAACC	32
SFpB1_N7SA_fw_2	gcggatccATGGTACAGTCTAAAATGTCTGACGTAAAG	38
SFpB1_N10SA_re	gcggatccTTGGTCACAACCTACAGCCATAACC	32

Table 6. Mutagenesis primer for either an HCoV-229E (blue) or a SARS-CoV polyprotein gene template (green)

name	primer (5' → 3')	length	mutation*
78_E_QN_f	C G A C T C C A T T T T G A A T A G T G T T G C A T C T T C T T T T G T T G G T A T G C C	45	Q83N
78_E_QN_r	G G C A T A C C A A C A A A A G A A G A T G C A A C A C T A T T C A A A A T G G A G T C G	45	
910_E_QN_f	G T G C C A C T G T G A G A T T G A A T G C T G G C A A A C A G A C T G A G	38	Q387N
910_E_QN_r	C T C A G T C T G T T T G C C A G C A T T C A A T C T C A C A G T G G C A C	38	
78_E_QA_f	G A G A A C G A C T C C A T T T T G G C A A G T G T G C A T C T T C T T T T G T T G G	44	Q83A
78_E_QA_r	C C A A C A A A A G A A G A T G C A A C A C T T G C C A A A A T G G A G T C G T T C T C	44	
89_E_QA_f	G T G A A C G T G T C G T T A A A T T G G C G A A C A A T G A A A T A A T G C C G G G C	44	Q278A
89_E_QA_r	G C C C G G C A T T A T T T C A T T G T T C G C C A A T T T A A C G A C A C G T T C A C	44	
910_E_QA_f	G G T G C C A C T G T G A G A T T G G C A G C T G G C A A C A G A C T G A G	39	Q387A
910_E_QA_r	C T C A G T C T G T T T G C C A G C T G C C A A T C T C A C A G T G G C A C C	39	
10His_E_QA_f	G T G A C C G G A C T G C T A T C G C A C A T C A C C A T C A T C A C C	36	Q525A
10His_E_QA_r	G G T G A T G A T G G T G A T G T G C G A T A G C A G T C C G G T C A C	36	
710_E_C347S_f	C A G T T G A A T T G G A A C C A C C T A G C A G A T T T G T T A T A G A C A C	40	C347S
710_E_C347S_r	G T G T C T A T A A C A A A T C T G C T A G G T G G T T C C A A T T C A A C T G	40	
710_E_G378E_f	G A G A G G T G C T G T T T T G G A A T A C A T T G G T G C C A C T G T G	37	G378E
710_E_G378E_r	C A C A G T G G C A C C A A T G T A T T C C A A A A C A G C A C C T C T C	37	
8_E_T123C_f	C T T T C C G T T A T C C C T G C T T G T T C T G C A G C C A G G C T C G T C	39	T206C
8_E_T123C_r	G A C G A G C C T G G C T G C A G A A C A A G C A G G G A T A A C G G A A A G	39	
710_S_G381E_f	C T T A A A C A A C C T A A A T A G A G A A A T G G T G C T G G G C A G T T T A G C	42	G381E
710_S_G381E_r	G C T A A A C T G C C C A G C A C C A T T T C T C T A T T T A G G T T G T T A A G	42	
78_S_QN_f	C T C G A T A A C C G T G C T A C T C T T A A T G C T A T T G C T T C A G A A T T T A G	44	Q84N
78_S_QN_r	C T A A A T T C T G A A G C A A T A G C A T T A A G A G T A G C A C G G T T A T C G A G	44	
V11E L14R_f	C G T A A A G T G C A C A T C T G A G G T A C T G C G C T C G G T T C T T C A A C A A C	44	V11E/

2. Material and methods

V11E L14R_r	GTTGTTGAAGAACCGAGCGCAGTACCTCAGATGTGCACCTTTACG	44	L14R
5_S_R298A_f	CACCATTGTGATGTTGTTGCACAATGCTCTGGTGTAC	37	R298A
5_S_R298A_r	GTAACACCAGAGCATTGTGCAACAACATCAAATGGTG	37	(Nsp5)

*except for the SARS-CoV M^{pro} mutation R298A, the amino acid position corresponds to the sequence of native pp710.

Table 7. Oligonucleotides of the SARS-CoV 3'-UTR applied to T7-transcription

name	sequence (5' → 3')	length
3'-UTR ΔHVR	TTTTTTTTTTTTTGTTCATTCTCCTAAGAAGCTATTTAAATCACATGGCTATGTGAGATTAAAGT TAACTAAACCTACTTGTGCTGTTTAGTTACGAGAATTCATTCCCTATAGTGAGTCGTATTACAT	127
3'-UTR ΔHVR-SH	AATTTTAAATACATTAATGCGGGTAAAAATATCTTAGATGTCTGTTTCGACTATGGTCTTAAAGC TGCATTCATTAATCTCTTTTCTAGGATAGTTCTTTAATACCTATAGTGAGTCGTATTACAT	127

2.1.4. *E. coli* culture and common solutions

Table 8. Strains of *E. coli*

strain	genotype	resistance/ selection marker	supplier
BL21 Star TM (DE3)	F ⁻ ompT hsdS _B (r _B ⁻ , m _B ⁻) gal dcm rne 131 (DE3)	-	Invitrogen
BL21 AI	F ⁻ ompT gal dcm Ion hsdS _B (r _B ⁻ m _B ⁻) araB::T7RNAP-tetA	-	Invitrogen
Origami TM (DE3)	Δ(<i>ara-leu</i>)7697 Δ <i>lacX74</i> Δ <i>phoA</i> PvuII <i>phoR</i> <i>araD139</i> <i>ahpC</i> <i>galE</i> <i>galK</i> <i>rpsLF</i> ⁺ [<i>lac</i> ⁺ <i>lacI</i> ^q <i>pro</i>] (DE3) <i>gor522::Tn10</i> <i>trxB</i> (Kan ^R , Str ^R , Tet ^R)	kanamycine, tetracycline, streptomycine	Novagen
DH5α TM	F ⁻ end A1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d <i>lacZ</i> ΔM15 Δ(<i>lacZYA</i> -argF)U169, hsdR17(r _K ⁻ m _K ⁺), λ-	nalidixic acid	Invitrogen
XL10-Gold	endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(<i>mcrA</i>) 183 Δ(<i>mcrCB</i> - <i>hsdSMR</i> - <i>mrr</i>)173 tet ^R F ⁺ [<i>proAB</i> <i>lacI</i> ^q ΔM15 Tn10(Tet ^R Amy Cm ^R)	nalidixic acid, tetracycline, chloramphenicol	Stratagene
XL1-Blue	endA1 gyrA96(<i>nal</i> ^R) thi-1 recA1 relA1 lac glnV44 F ⁺ [<i>proAB</i> ⁺ <i>lacI</i> ^q Δ(<i>lacZ</i>)M15] hsdR17(r _K ⁻ m _K ⁺)	nalidixic acid, tetracycline,	Stratagene

Table 9. Media for cultivation of *E. coli*

medium	composition
LB medium [133]	10 g tryptone, 10 g NaCl, 5 g yeast extract (for solid LB medium 1.5% (w/v) agar agar was added)
2YT Medium	16 g l ⁻¹ tryptone, 5 g l ⁻¹ NaCl, 10 g l ⁻¹ yeast extract

Table 10. Common solutions, electrophoresis and affinity chromatography

solution	composition
Tris-acetate-EDTA (1×TAE) (Agarose gel electrode buffer)	80 mM Tris 40 mM acetic acid 2 mM EDTA
Tris-borate-EDTA (1×TBE) (native PAGE, agarose gel electrode buffer)	89 mM Tris base 89 mM boric acid 2 mM EDTA
coomassie staining solution	0.25% (w/v) coomassie brilliant blue G-250 25% (v/v) 2-propanol 10% (v/v) acetic acid
coomassie destaining solution	20% (v/v) acetic acid
SDS-PAGE electrode buffer	250 mM Tris 1.9 M glycine 1% (w/v) SDS
SDS-PAGE sample buffer (2×)	3.55 ml H ₂ O 1.25 ml 0.5 M Tris-HCl 2.5 ml glycerol 2.0 ml 10% (w/v) SDS 0.2 ml 0.5% (w/v) bromophenol blue 20 mM DTT
stacking gel buffer	0.5 M Tris-HCl, pH 6.5
separating gel buffer	1.5 M Tris-HCl, pH 8.5
lysis buffer*	50 mM NaH ₂ PO ₄ , pH 7.9 / 50 mM Tris-HCl pH 7.9 300 mM NaCl (15 mM imidazole for his-tag-based purification)
Strep-Tactin resin elution buffer	100 mM Tris-HCl, pH 8.0 260 mM NaCl 1 mM EDTA 2.5 mM desthiobiotin
Ni-NTA resin elution buffer*	lysis buffer supplemented with imidazole
GST resin elution buffer*	PBS-lysis buffer containing 200 mM NaCl in total, supplemented with reduced glutathione

* common buffer composition; changes of pH, ionic strength and additives are stated elsewhere below.

Staining of PAGE gels with AgNO₃, providing a higher sensitivity was carried out according to [134]. Aqueous solutions in contact with RNA were treated with 1% (v/v) DEPC prior to use.

2.2. Methods

2.2.1. Agarose gel electrophoresis

After boiling, 0.8-2% agarose in TAE buffer was poured into a chamber (Peqlab) and allowed to cool down. The gel was overlaid with TAE buffer. A Power Pac 300 power supply (Biorad) was connected to the chamber and typically adjusted to 95 V (approx. 70 mA) for nucleic acid separation (loading dye obtained from Thermo Scientific; method described by Adkins and Burmeister [135] and subsequent detection by either ethidium bromide (Carl Roth) or SYBR Gold (Invitrogen).

2.2.2. Polymerase chain reaction (PCR)

The components listed below were typically mixed in a reaction tube for a PCR [136] and placed in a thermocycler.

DNA template	15 - 100 ng
polymerase buffer (10×) (Thermo Scientific)	5 μ l
dNTPs, 2 mM each	4 μ l
primer forward, 10 pmol μ l ⁻¹	1 μ l
primer revers, 10 pmol μ l ⁻¹	1 μ l
<i>pfu</i> / <i>taq</i> polymerase (2.5 U μ l ⁻¹ / 1 U μ l ⁻¹) (Thermo Scientific)	1 μ l
MgCl ₂ (2.5 mM; for Mg ²⁺ -free <i>taq</i> polymerase buffer)	2 μ l
dH ₂ O	ad 50 μ l

Basic PCR program:

I	95 °C	90 s	(initial denaturation)
II	95 °C	40 s	(denaturation)
III	55 °C	50 s	(annealing)
IV	72 °C (68 °C)	60 s per 900 nt	(elongation)

(steps II-IV were repeated 30 times followed by 72 °C for 6 min. and storage at 4 °C)

To clone a gene of interest into a vector, the PCR product was purified by agarose gel electrophoresis and cleaved with the desired enzymes (e.g. *Bam*HI and *Eco*RI) in a suitable buffer as recommended. Restriction enzymes were obtained from Thermo Scientific and NEB. A 1.5-fold excess of purified digestion product over plasmid was incubated with T4 ligase according to the instructions of Thermo scientific. Up to 5 μ l of a 20 μ l reaction containing 20-50 ng vector and 1 U of T4 ligase were used for transformation of *E. coli*. After analytical PCR/digestion, the sequence of the gene of interest obtained from a single clone was verified by a sequencing reaction according to Sanger [137] by GATC Biotech AG (Konstanz).

2.2.3. Mutagenesis PCR

Forward and reverse primer designed for mutagenesis "whole-plasmid" PCRs are pair-wise reverse and complement to each other. The reaction and PCR program (20 cycles, primer annealing at 55 °C) were set up according to the instructions of the Quick Change II – *Site-directed mutagenesis kit* (Invitrogen) that was used. Afterwards, the reaction mixture was supplemented with 10 U of *DpnI* and incubated for another 1 h at 37 °C to digest the non-mutated methylated template. Up to 5 µl of PCR product were used for a transformation of chemically competent *E. coli* XL1 blue cells. The sequence/mutation of the corresponding gene was verified by a sequencing reaction with isolated plasmid (GATC Biotech AG).

2.2.4. Transformation of *E. coli*

Chemically competent cells of various *E. coli* strains, prepared according to [138; 139] and stored at -80 °C, were thawed on ice. Plasmid DNA (40-80 ng) was added to the cell suspension and incubated on ice for 20 minutes. Subsequently, the cells were exposed to 42 °C for 45 s and cooled down on ice again, before they were spread on LB agar (1.8% (w/v) agar) with the suitable antibiotic and allowed to incubate over night at 37 °C.

2.2.5. Plasmid isolation

Plasmids were isolated from *E. coli* cells by using the peqGOLD Plasmid Miniprep Kit 1 (PEQLAB) according to the manufacturer's instructions.

2.2.6. Heterologous gene overexpression in *E. coli*

Frequently, the initial pre-culture of *E. coli* was inoculated with cells in the logarithmic growth phase that had been stored at -80 °C in the presence of 30% (v/v) glycerol. A pre-culture of *E. coli* cells of a strain optimized for gene overexpression was grown over night in either LB- or 2YT medium in the presence of the required antibiotic(s), usually 100 mg l⁻¹ ampicillin. Fresh 2YT medium (1.2 l per 5 l Erlenmeyer flask) was inoculated with the pre-culture. The cells were grown at 37 °C until a specific optical density determined by a spectrophotometer was obtained. For promoter induction, either L(+)-arabinose (20% stock solution in dH₂O), IPTG (1 M in dH₂O) or AHT (2 g l⁻¹ in DMF) were added. After gene expression for several hours, the cells were finally harvested by centrifugation and stored at -20 °C after removing the supernatant (table 30 is a summary of the specific optimized conditions, see appendix).

The SARS main protease (M^{pro}, Nsp5) gene cloned into pGEX-6p-1 was kindly provided by Prof. R. Hilgenfeld (University of Lübeck). Gene expression and purification of the SARS main protease was performed as previously described by [78].

2.2.7. Gene expression and *in vivo* crystallization experiments in *Sf9* insect cells

Full-length pp710 of SARS-CoV was cloned into pFastBac1 using the *Bam*H1 and *Eco*RI restriction sites. *E. coli* DH10 bac cells were transformed with the desired plasmid after verification of the sequence. The cells were additionally allowed to gently shake at 37 °C for 4 h for double homologous recombination. Cells harvested from an overnight culture of a single clone in the presence of gentamycin, tetracycline and kanamycin were lysed by the method of alkaline lysis (Birnboim and Doly, 1979) to finally purify the bacmid by subsequent precipitation using ethanol. The dried DNA was resuspended in dH₂O and short-term stored at 4 °C. Recombination in *E. coli* DH10 bac cells was verified by PCR using the isolated bacmid as a template and pUC/M13 primer binding to the gene flanking regions. For transfection, 1 µg of bacmid in 100 µl Ex-Cell Medium was gently mixed with 2 µl ESCORT IV reagent and incubated at room temperature for 20 min. This mixture was drop wise added to 2 ml Ex-CellMedium containing $1.6 \cdot 10^6$ *Sf9* insect cells in a 6-well plate (greiner bio-one) followed by incubation for 5 h at 27 °C. The medium was replaced by fresh medium containing penicillin and streptomycin. After additional incubation for 4 days at 27 °C the medium was aspired yielding the P1 virus stock. The P2 and P3 virus stocks were created by supplementing the same amount of cells with 400 µl of the P1 stock, followed by incubation at 27 °C. After 4 days the P2 virus stock was harvested and used to access the P3 virus stock in the same way. Finally, the P3 virus stock was used to screen the infected cells for *in vivo* grown crystals. To obtain this, $1.6 \cdot 10^6$ *Sf9* insect cells were supplemented with different volumes of the P3 virus stock (2 µl, 10 µl, 20 µl, 100 µl, 200 µl) in a 6-well plate (1 well comprises the negative control) and incubated at room temperature and at 27 °C for up to 10 days with daily analyzes by light microscopy.

2.2.8. Cell lysis and chromatography

2.2.8.1. Affinity chromatography

For the purification of **oligo-histidine-tagged proteins** (HCoV-229E pp710 and genes cloned into pRSETA or pASK IBA 33plus/35plus), the harvested cells were lysed in an appropriate buffer by sonication ([140]; Branson Sonifier 250). Subsequently, the lysate was centrifuged for 45 min (10000-15000×g) at 4 °C. The supernatant was treated with benzonase (SIGMA) (ideally in a lysis buffer with < 300 mM NaCl) at room temperature for 1 h, if an increase in viscosity or protein-bound nucleic acids were expected. The supernatant was applied to a column containing Ni-NTA resin and incubated for 5 min, while slowly shaking. The resin was washed with additional 8-10 CVs of wash buffer before eluting the target protein from the resin by stepwise increasing the imidazole concentration to up to 400 mM within 12 CVs. All purification steps were verified by SDS-PAGE. For benzonase treatment after purification by affinity chromatography, 1 mM MgCl was added in a dialysis tube.

Proteins with N-or C-terminal **strep-tag** were purified using the high affinity to Strep-Tactin resin (IBA). An additional SA linker (*SAWSHPQFEK*) is provided by the vector pASK IBA 3plus (C-terminal tag) and *MASWSHPQFEKGA* comprises the N-terminal tag encoded by pASK IBA 5plus. Cells were disrupted by sonication as described above. The soluble protein fraction was allowed to pass very slowly through a column containing equilibrated resin. The resin was subsequently washed with ten CVs of lysis buffer. Bound protein molecules were finally eluted with five CVs of Strep-Tactin resin elution buffer (2 ml fractions each) containing 2.5 mM D-desthiobiotin.

Pp78 of HCoV-229E was recombinantly produced fused to an N-terminal **GST-tag**. After sonication and centrifugation the soluble cell lysate was incubated with GST resin (QIAGEN) carefully shaking in a column for 20 min at 6 °C. Subsequently, the resin was washed with eight CVs. The protein was eluted by a stepwise increase of reduced L-glutathione (γ -glutamyl-cysteinyl-glycine; GSH) to up to 30 mM. For lysis and affinity chromatography elution of pp78-GST fusionprotein, PBS [141] with a NaCl concentration increased to 200 mM was used.

All kinds of affinity chromatography resin were maintained according to the instructions of the manufacturer. The GST-tag of HCoV-229E pp78 and the his₆-tag of SARS-CoV M^{pro} were removed particularly for crystallization by prescission protease (GE-Healthcare) at a ratio of 20:1 in the presence of 2 mM DTT either for 30 h at 6 °C or over night at 20 °C.

2.2.8.2. Size-exclusion chromatography (SEC)

An ÄKTA FPLC purification system (ÄKTA Purifier P-901; GE Healthcare, UK) connected to a UV detector and a fraction collector was provided and maintained at 16 °C. The device including the column was previously rinsed with water and buffer and separation parameters were optimized controlled by the accompanying UNICORN software, which also supported the chromatogram evaluation. Applied columns obtained from GE Healthcare are listed below.

- Hi Load 16/60 Superdex 75
- Hi Load 16/60 Superdex 200
- Superose 12 10/300 GL

Lysozyme (hen egg, 14 kDa; SIGMA), thaumatin (*Thaumatococcus daniellii*; 22 kDa; SIGMA), ovalbumin (hen egg; 44 kDa), conalbumin (chicken egg white; 75 kDa), aldolase (rabbit muscle; 158 kDa), ferritin (440 kDa; horse spleen), thyroglobulin (bovine thyroid; 669 kDa) and Blue Dextran 2000 (all GE Healthcare) were used for calibration of SEC columns. To change the buffer composition of a protein solution in between of purification/experimental steps the protein was either dialysed to

2. Material and methods

introduce/remove compounds via a dialysis membrane with a sufficient MWCO or applied to an ultra filtration membrane (AMICON; 3 kDa/10 kDa/30 kDa/50 kDa MWCO) used in a centrifuge according to the company's instructions.

2.2.9. Protein quantification

The protein concentration in solution was usually determined by measuring the specific absorbance of ultra violet light at a wavelength of 280 nm according to the law of Lambert and Beer – see equation 1.

$$A_{280\text{ nm}} = \epsilon d c \quad (1)$$

The theoretical molar extinction coefficient ϵ [$\text{M}^{-1} \text{cm}^{-1}$] was calculated by ProtParam ([142]; ExPASy, SIB); d represents the path length passed by the UV-light (cuvette diameter) and c the protein concentration of interest.

Table 11. List of theoretical molar extinction coefficients, assuming all thiol groups are reduced. All extinction coefficients applied in the course of this project were calculated by ProtParam ([142]; SIB)

HCoV-229E	ϵ [$\text{M}^{-1} \text{cm}^{-1}$]	SARS-CoV	ϵ [$\text{M}^{-1} \text{cm}^{-1}$]
pp710 (C-terminal his ₆ -tag)	58330	pp710 (strep-tag)	58330
pp78 (N-terminal his ₆ -tag, long linker)	32430	pp710 (his-tag)	52830
pp78 (without GST-tag)	25440	pp79 (strep-tag)	45380
Nsp7	8480	pp78 (strep-tag)	32430
Nsp8	16960	pp89 (N-term. strep-tag)	39880
Nsp9	15930	Nsp9	12950
Nsp10	16960	M ^{pro} (C-terminal his ₆ -tag)	32890

2.2.10. SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is a simple method to determine size and purity of a protein sample under denaturing conditions. It is based on the electrophoretic mobility of proteins that are negatively charged by SDS (Laemmli, 1970). Discontinuous PAGE gels consist of a narrow stacking gel (4% acryl amide, Tris pH 6.8) to focus the sample molecules and a separation gel Tris-buffered at pH 8.8. Protein samples were mixed with 2× concentrated sample buffer and incubated at 95 °C for 5 min for denaturation. After pouring and polymerization of the gel due to addition of APS and TEMED, the gel was vertically placed in a Mighty Small II gel chamber (Hoefer Inc, USA) and connected to an EV 231 power supply (Peqlab, Germany) to adjust the electric field. Electrophoresis was terminated as soon as bromophenol blue reached the low end of the gel. Typically the gel was stained for at least one hour in coomassie staining solution and subsequently destained in 20% (v/v) acetic acid until a sufficient contrast was obtained. The gels were documented by a colour scanner (Epson). Alternatively, proteins

were either stained by AgNO₃ [134] or prepared for western blotting. To approximate the mass of a protein the molecular weight markers #26610, #26628, #26632, #26634 (all Thermo Scientific) were applied.

2.2.11. Native gel electrophoresis

2.2.11.1. Clear/blue native PAGE

Pre-cast polyacrylamide gels (3-12% and 4-16% gradient gels) for blue or clear native PAGE were obtained from SERVA (Heidelberg) and used according to the instructions in the manual applying a mighty small II PAGE chamber (Hoefer, USA). Furthermore, a standard protein marker to approximate the molecular weight, sample buffer as well as anode and cathode buffer was purchased from SERVA.

2.2.11.2. Isoelectric focussing/non-equilibrium pH-gradient gel electrophoresis (NEPHGE)

Pre-cast ampholyt polyacrylamide gels obtained from SERVA (pH range 3.5-9.5) were used for isoelectric focussing (or NEPHGE) according to the instructions of SERVA using SERVA anode/cathode buffer, 2× IEF sample buffer, SERVA Violet 17 for staining and SERVA IEF Marker 3-10 Liquid Mix.

2.2.11.3. Zone-interference gel electrophoresis (ZIGE)

The principle to investigate protein-ligand interactions by ZIGE was described by Abrahams [143] and performed accordingly. A 1% (w/v) agarose gel was prepared with Tris-borate buffer supplemented with 100 mM NaCl and 400 μM MgCl₂. Electrophoresis of pp78 with the horizontal gel was performed in a Peqlab 40-1410 chamber strictly on ice. Electrophoretic separation was terminated after approximately 2 h at 200 V. To visualize RNA after electrophoresis, the gel was fixed in 3.5% (w/v) α-sulfosalicylic acid and 10% (w/v) trichloroacetic acid. The protein was stained by Coomassie Brilliant Blue G-250.

2.2.12. Western blot

Proteins were separated by SDS-PAGE prior to electrophoretic transfer to a PVDF membrane that was activated with methanol [144]. Membrane and acrylamide gel on two layers of Whatman paper covered with sponges were put into a wet-blot chamber filled with CAPS-buffer (10 mM CAPS, pH 11) for the transfer for 3 h at 6 °C and 50 V voltage. Afterwards, the membrane was washed twice with TBS (750 mM NaCl, 1 M Tris-HCl, pH 7.5). Unspecific binding to the membrane was blocked by incubation with blocking buffer (0.5% (w/v) BSA in TBS) for 30 min before the membrane was washed again with TBS. The membrane was incubated with mouse anti-tetra-histidine IgG1 (QIAGEN) in blocking buffer for 2 h at room temperature. After washing with TBST (5× TBS buffer,

2. Material and methods

0.05% (v/v) Tween 20) and TBS the membrane was incubated with the secondary goat anti-mouse antibody linked with APC for 45 min. The membrane was washed again with TBS and TBST and transferred to reaction buffer supplemented with 50 g l⁻¹ NBT (dissolved in 70% DMF) and 20 g l⁻¹ BCIP. At a sufficient level of staining the membrane was rinsed with deionized water to stop the reaction and finally documented by a colour scanner.

2.2.13. Quantification of free thiol groups

Determination of free thiol groups in proteins by Elman's reagent (5, 5'-dithiobis-(2-nitrobenzoic acid, DTNB) was carried out in transparent flat bottom 96-well microtiter plates (Greiner). Proteins were adjusted to a concentration of 40 μM in 20 mM Bis-Tris-propane supplemented with 300 mM NaCl and adjusted to pH 7.5. 50 μl of the protein solution were incubated with an equal volume of 1.5 mM DTNB. The absorbance of the solution at 405 nm was monitored by a GENios spectrophotometer/ELISA reader (TECAN) and XFLUOR4 software. An extinction coefficient of 14150 M⁻¹ cm⁻¹ was used for data evaluation. The assay was additionally calibrated using reduced L-glutathione.

2.2.14. Quantification of enzymatic activity of M^{Pro}

The proteolytic activity of SARS-CoV M^{Pro} was determined in flat-bottom 96-well microtiter plates (Greiner) using a total volume of 100 μl per well with a final concentration of 400 μM TSAVLQ-pNA (purity 98.5%; GenScript, NJ, USA) and 2-20 μM M^{Pro} purified by Ni-NTA affinity chromatography. Again, the absorbance of the solution was monitored at 405 nm using a GENios spectrophotometer (TECAN). Catalysis was promoted by adding 1 mM TCEP.

2.2.15. Protein processing by M^{Pro}

Prior to polyprotein/tag cleavage by M^{Pro}, imidazole (if present) was removed by dialysis since it inhibits the proteolytic activity. If not stated otherwise the substrate in its specific buffer was mixed with SARS-CoV M^{Pro} at a molar ratio of 10:1 in the presence of 4 mM DTT at pH 7.5-8. Proteolytic cleavage was performed at either 20 °C or 6 °C at least over night. The processing reaction was verified by SDS-PAGE. Afterwards the protease was typically removed by incubating the solution with equilibrated Ni-NTA resin slowly shaking for 20 min. The resin was separated either by filtration or centrifugation at 400×g. The resin was washed with buffer twice to optimize the yield of cleaved protein.

2.2.16. T7 transcription, labelling of RNA and filter retention RNA-binding assay

A conserved 105 nt fragment of the SARS-CoV 3'-UTR (see also [108]) lacking a hypervariable region (3'-UTR Δ HVR) and a shuffled RNA fragment comprising the same nucleotides, but a completely different sequence (3'-UTR Δ HVR-SH) was produced by *in vitro* T7 transcription. Initially, a T7 promoter primer was hybridized with the complementary part of the ssDNA template for both fragments. The hybridization mixture (5 μ M ssDNA template, 10 μ M T7 primer in 1 \times hybridization buffer [0.2 M Tris-HCl, pH 8.0, 0.5 M NaCl, 10 mM EDTA], dH₂O ad 20 μ l) was heated to 95 °C for 2 min and slowly cooled down to RT. The T7 transcription was conducted for 2 h at 37 °C. The reaction setup is described below and the products were verified by denaturing PAGE (10% (w/v) acrylamide/bisacrylamide (19:1) in 1 \times TBE containing 8 M urea). Therefore samples were mixed with an equal amount of RNA sample buffer (see below). The desired product was eluted from a denaturing PAGE gel. RNA in the gel was detected by *UV shadowing* on a TLC plate containing fluorescent manganese activated zinc silicate. Elution of the RNA from the gel by diffusion was performed by incubation of the excised gel piece in 0.3 M sodium acetate, pH 5.4 two times for 1 h at 37 °C. RNA in the supernatant was precipitated using 2-propanol.

T7 transcription

	<i>wt</i> (3'-UTR Δ HVR)	<i>shuffled</i> (3'-UTR Δ HVR-SH)
hybridization reaction	2.5 μ l	5 μ l
3 \times transcription buffer	16.6 μ l	16.6 μ l
100 μ M MgCl ₂	15 μ l	20 μ l
NTP-Mix [25 mM]	5 μ l	5 μ l
T7-RNA polymerase [150 U/ μ L]	2.5 μ l	2.5 μ l
dH ₂ O	ad 50 μ l	ad 50 μ l

RNA sample buffer (2 \times)

80% (v/v) 8 M urea

10% (v/v) glycerin

10% (v/v) 0.5 M EDTA, pH 8.0

For radioactive labelling, purified RNA molecules were dephosphorylated by alkaline phosphatase (FastAP, Thermo Scientific) for 30 min at 37 °C according to the instructions of the supplier in a total reaction volume of 40 μ l. Subsequently, the reaction was terminated at 75 °C for 5 min. The RNA was phosphorylated again by T4 polynucleotide kinase (T4-PNK) and γ -[³²P]-ATP (10 μ Ci μ l⁻¹) as a substrate. Two units T4-PNK were incubated with 0.4 μ M RNA and 3 μ l γ -[³²P]-ATP in 1 \times PNK buffer A in a total reaction volume of 50 μ l.

2. Material and methods

Various concentrations of pp710 were incubated with a constant amount of both RNA fragments. The solution was drawn through a nitro cellulose membrane (Carl Roth), while the protein binds to the membrane. Unbound nucleic acid should flow through, whereas the RNA molecules that bind to the protein remain on the filter detected using a phosphor imager (BioRad, München)

2.2.17. RNA-dependent RNA polymerase (RdRp) activity assay

The assay addressing RdRp activity was performed in cooperation with *Y. Xiao* (University of Lübeck, Germany). For a single reaction 9 μ l of the mixture described below was mixed with 1 μ l of protein solution. The solution was incubated at 30 °C for 1 h, terminated by the addition of EDTA. If the α -phosphate group (of one kind of nucleotide) carries the radioactive label, reaction products of various sizes are presumed to carry the radioactive phosphor; free diphosphate is a side-product. Radioactively labeled reaction products were analyzed and separated by denaturing PAGE (17.5% acrylamide and 7 M urea) in TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.0). The gel was screened and scanned with a phosphor imager (Typhoon; GE Healthcare).

Tris, pH 8.0	5 μ l	(25 mM f.c.)
KCl	2 μ l	(2 M stock solution)
MgCl ₂	50 μ l	(10 mM stock solution)
ATP/CTP/UTP	1 μ l each	
α - ³² P-GTP	1 μ l	
RNAsin® (Promega)	2 μ l	
RNA template*	10 μ l (2 μ M)	
dH ₂ O	ad 110 μ l	

*FCoV [strain FIPV WSU-79/1146] 3'-UTR, stemloop1; 67 nucleotides:

5'-GGCAACCCGAUGUCUAAAACUUGUCUUUCCGAGGAAUACGGGUCAUCGCGCUGCC
UACUCUUGUAC-3' (supplied by Riboxx)

2.2.18. STD-NMR spectroscopy

Saturation transfer difference (STD) NMR spectroscopy was performed in cooperation with Dr. *Miriam Kötzler* (University of Hamburg), who performed the data processing in addition to the experiments, in the research group of Prof. *Bernd Meyer*. The development and optimization of the method widely refers to B. Meyer and Co-workers. Applying a Bruker Avance 700 MHz NMR spectrometer (Bruker BioSpin, USA) equipped with a TXI triple resonance cryoprobe head incorporating gradients in the z-axis, saturation transfer difference (STD) NMR spectra were recorded at 300 K. Each 160 μ L samples in a 3 mm sample tubes contained 20 μ M protein and 2.0 mM ligand if not indicated otherwise. Deuterium oxide buffered with 30 mM phosphate and containing 2 mM

magnesium chloride, 0.01% sodium azide and 200 mM NaCl adjusted to pH 7.9 or 160 mM sodium chloride at pH 7.3 for polyproteins and for SARS-CoV main protease, respectively, was used as solvent. Polyproteins under investigation are SARS-CoV pp710 (N-terminal strep-tag) and SARS-CoV pp78 (N-terminal his₆-tag) as well as HCoV-229E pp710 (C-terminal his₆-tag) and HCoV-229E pp78, after cleaving the N-terminal GST-tag off the fusion protein.

Data was acquired using a Bruker standard sequence (stddiff.3) incorporating T_{1ρ} relaxation filter sequence with a spin-lock duration of 10 ms to suppress residual protein resonances. On resonance irradiation was applied at 0 ppm for all ligands and off resonance irradiation at 40 ppm. Saturation was achieved by a cascade of Gaussian pulses with duration of 50 ms (field strength 100 Hz) to give a total saturation time of 2 s. STD NMR spectra and reference spectra were acquired with 32,768 data points and a total of 256 transients. FIDs were multiplied with an exponential function (line broadening 1 Hz) before Fourier transformation. All ligands were assigned using 1D proton and carbon and 2D correlation NMR spectroscopy as necessary. To determine the size of STD effects, resonances in the STD spectrum were integrated with respect to the reference spectrum. Therefore, an absolute STD effect of 100% results if the resonances in both spectra show the same intensity. As different relaxation time constants may bias the epitope, all longitudinal relaxation times were determined by means of an inversion recovery experiment and absolute STD effects were divided by the T₁ of the respective proton. These corrected absolute STD effects are then normalized to the largest STD effect in the molecule to yield the relative STD effects in order to map the ligand epitope. All ligands were also measured in the absence of protein in order to ensure that the observed saturation is not biased by other effects than saturation transfer from the protein.

2.2.19. Surface plasmon resonance (SPR) spectroscopy

SPR spectroscopy was performed with a Biacore T1000 device in the laboratory of Prof. *I. Kursula* (Helmholtz Centre for Infection Research, Hamburg). A CM5 sensor chip (carboxy-methyl surface, Biacore) was activated by an NHS-EDC solution (molar ratio 1:1) at a flow rate of 5 μl min⁻¹. At the same speed a target protein was immobilized after diluting it with 50 mM sodium acetate buffer (pH 4.0) to 1 μM final concentration. Remaining free binding sites were saturated with ethanolamine for 7 min at 5 μl min⁻¹. The ligand of interest was typically passing the immobilized protein at a speed of 20 μl min⁻¹. The maximum resonance was determined for various ligand concentrations and plotted according to the one-site-binding model. The K_D-value depends on association and dissociation rate as follows:

$$K_D = \frac{k_{off}}{k_{on}} \quad (2)$$

2.2.20. Circular dichroism (CD) spectroscopy

CD spectroscopy is a common method to obtain insights into a protein's secondary structure content and to a lower extent into the tertiary structure. It allows structural comparison of protein variants or proteins obtained from different expression systems. CD spectroscopy is very useful for characterizing small chiral molecules and macromolecules like DNA or RNA as well. The CD spectrometer (J-815, Jasco, UK) equipped with a peltier element was calibrated according to the supplier's instructions. The peltier element allowed precise investigations concerning the thermal stability of a certain protein fold. The *ellipticity* of the sample was typically measured in a 1 mm quartz cuvette in a wavelength interval ranging from 185-260 nm (*far-UV* spectrum) or in a 1 cm cuvette for wavelength values from 260-320 nm (*near-UV* spectrum). The baseline recorded for the corresponding buffer was subtracted. The ellipticity θ is defined as the difference in absorbance of clockwise and counter clockwise circular polarized light.

$$\theta = \frac{180 \cdot \ln 10}{4 \cdot \pi} \cdot (A_L - A_R) \quad (3)$$

The ellipticity that was determined was converted to MRE (equation (4); mean residue ellipticity, [deg cm² dmol⁻¹], according to [145]) according to the following equation (MRW: mean amino acid residue weight, c: protein concentration, d: path length through the sample): For the entire recorded spectrum the total absorbance of the sample including all salts and additives should not exceed 3 AU. Standard curves determined by [146] served as analytical references, e. g. landmarks of the ellipticity minima indicating α -helical structure.

$$MRE = \frac{MRW \cdot \theta}{10 \cdot cd} \quad (4)$$

To investigate the folding stability of a protein the melting temperature (T_m) was determined three times individually. A Boltzmann function was fit to the experimentally determined dependency of the ellipticity on the temperature. The T_m value of each protein is the global maximum of $d\theta/dT$, i.e. the temperature of the highest loss of ellipticity according to the Boltzmann function. If not stated otherwise T_m was determined at a wavelength of 220 nm.

2.2.21. Dynamic light scattering

To investigate the dispersity and the numerical hydrodynamic radius (and its time-dependent change) of macromolecules in solution a spectroLIGHT 300/500 DLS instrument (nabitec/Xtal concepts) was applied. This allows to use either at least 12 μ l in a quartz cuvette or at least 2 μ l of sample in situ, e.g. in a terazaki 96 well plate overlaid with paraffin oil to reduce evaporation. The sample was irradiated by a red light class 3b laser ($\lambda = 690$ nm; laser power 10-50 mW) to detect

isotropic scattering at an angle of 90 degree. In preparation, the samples were centrifuged for 30 min. at $16,100 \times g$, if not stated otherwise, and the desired volume was drawn from slightly below the surface. The R_h is experimentally determined by its dependence on viscosity η and the diffusion coefficient D_0 (Stokes-Einstein equation).

$$R_h = \frac{KT}{6\pi\eta D_0} \quad (5)$$

A theoretical hydrodynamic radius of *nearly globular* particles may be estimated as follows. [ρ : Density of the protein; N_A : Avogadro constant; h : Hydration volume ($0.35 \text{ g H}_2\text{O (g protein)}^{-1}$); V_S : Specific particle volume (commonly set to $0.73 \text{ cm}^3 \text{ g}^{-1}$)]

$$R_h^{theo} = \left[\frac{3M(V_S + h)}{4\pi \cdot N_A} \right]^{\frac{1}{3}} \quad (6)$$

Isolating the molecular mass M the equation is rearranged as follows.

$$M_W^{theo} = N_A \cdot \rho \cdot R_h^{2.3} \cdot \frac{4}{3} \pi \quad (7)$$

2.2.22. Small-angle X-ray scattering (SAXS)

A monodisperse solution containing purified protein was applied to SAXS at the EMBL BioSAXS beamlines X33 of the DORIS storage ring (Hamburg, Germany) and P12 of the PETRA III 3rd generation synchrotron source. SAXS aimed to obtain low-resolution structural data of randomly oriented particles in solution. The scattering data was recorded by a 2D Photon counting Pilatus 1M-W pixel X-ray detector (beamline X33) or a 2D photon counting Pilatus 2M pixel X-ray detector (beamline P12) respectively. The sample was positioned 2.7 m away from the detector at DORIS and 3.1 m at PETRA III. Typically 20-100 μl sample solution each (with stepwise increasing protein concentration) was exposed to the beam with a wavelength of 1.54 \AA (beamline X33) or 0.124 \AA (beamline P12) at $10 \text{ }^\circ\text{C}$. The scattering profiles recorded in the momentum transfer range $0.01 < s \leq 6 \text{ nm}^{-1}$ were initially plotted and evaluated using Primus [147] (e.g. containing AutoRg, SASPLOT). To determine the radius of gyration (R_g) the Guinier approximation was used, presuming that the scattering intensity at very small angles ($s < 1.3 R_g^{-1}$) is given by $I(s) = I(0) \exp[-(sR_g)^2/3]$ [148], derived from

2. Material and methods

$$\ln I(s) = 1 - \frac{R_g^2 s^2}{3} \quad \text{and} \quad s = \frac{4\pi \sin \theta}{\lambda} \quad (8/9)$$

Linear Guinier plots indicated monodispersity. The pair-distribution function was calculated by GNOM [149]. *Ab initio* models were calculated and superimposed using DAMMIN/DAMMIF [150] and the automated mode of DAMAVER [151]. CRY SOL [152] allows comparison/superposition of the scattering data to known high-resolution structures. Crysol is aware of the hydration shell and calculates a spherically averaged scattering pattern using multipole expansion of the scattering amplitudes. Varying the background, hydration shell contrast and relative amount of displaced solvent, the χ -value is minimized. This value comprises a quality measure of the deviation of the fitted function and the scattering data. SASREF/GASBOR [153; 154] and MIXTURE [147] were applied both for rigid-body modelling in addition.

Bovine serum albumin (66 kDa; in 50 mM HEPES, pH 7.5) and maltose binding protein served as a molecular weight reference.

The following equation that represents a suitable approximation for nearly globular particles was used to compare the radius of gyration (R_g) and the hydrodynamic radius (R_h) as determined by DLS.

$$R_h = R_g \sqrt{\frac{5}{3}} \quad (10)$$

2.2.23. Crystallization, data collection and model building

In the beginning, to adjust the protein concentration that is considered to be favourable for crystallization, a pre-crystallization test (Hampton) was performed according to the instructions. Initial screenings to obtain promising crystallization conditions was carried out applying the pipetting robot Honeybee 961 (Genomic solutions, UK) combined with either a 2-well MRC plate (NUNC) or a Nextal Qia1 plate (Qiagen, Germany). Usually, 300 nl of protein solution were mixed with 300 nl of precipitant solution and equilibrated against 50 μ l reservoir solution (sitting drop, vapour diffusion) at a specific constant temperature (at first 20 °C). Sets of crystallization solutions that were screened comprise JCSG+, Classic suite, Cryos and ComPAS (all Qiagen) as well as PACT premier, Morpheus and Stura footprint screen (Molecular Dimensions). Promising conditions were refined by either hanging/sitting drop vapour diffusion experiments, counter diffusion in a GCB or crystallization under oil in a 96-well terazaki plate. For tiny crystals microseeding was applied. Differentiation between protein and salt crystals was performed by analysing protein solution drops with the CrystalScore device (Diversified Scientific Inc, USA) according to the manufacturer's specifications to detect specific fluorescence of the protein crystals.

For HCoV-229E Nsp10 a single crystal was cryoprotected by a transfer into the reservoir solution containing 1% MPD and 16% glycerol in a nylon loop and exposed to an X-ray beam (wavelength 0.81 Å) at the consortium beamline X13 of the DORIS storage ring (DESY, Hamburg). Diffraction data was collected to up to 1.8 Å applying a MARCCD detector. MOSFLM [155] and SCALA [156] were used for indexing and data reduction. Initial phases were obtained by molecular replacement using PHASER [157]. SARS-CoV Nsp10 (pdb code 2FYG) was used as a search model. Refinement of the electron density towards the diffraction data after phasing was performed using REFMAC V5.2.0019 [158]. Model building was further conducted by consecutive rounds of manual optimization in COOT [159; 160] and restraint refinement using REFMAC. The data processing software was similarly applied to process SARS-CoV M^{pro} diffraction data.

Synchrotron radiation sources and crystallography beamlines

P14 Beamline Petra III, DESY (Hamburg)

Synchrotron source: PETRA III storage ring

Wavelength: 1.69 Å

Focal Spot: 10x10 μm²

Detector: PILATUS 6M

X13 Consortium Beamline, HASYLAB/DESY (Hamburg)

Synchrotron source: DORIS III storage ring

Wavelength: 0.81 Å

Focal spot: 2 mm x 0.4 mm

Detector: MARCCD 165 mm

2.2.24. Mass spectrometry (MS)

2.2.24.1. MS-based protein identification

Digestion by trypsin and mass spectrometry data collection was performed in cooperation with the research group of Prof. *H. Schlüter* (UKE, Hamburg). Coomassie-stained protein samples were excised from an SDS-PAGE gel and destained. If not stated otherwise, disulfides were reduced by adding DTT dissolved in NaHCO₃ and modified by iodoacetamide. Digestion by trypsin was typically conducted over night by adding trypsin in trypsin-resuspension buffer (sequencing grade; Promega) to the sample. Peptides were extracted from the gel by adding 50% (v/v) ACN/0.1% TFA. Peptide samples were desalted by reversed phase chromatography. In addition to an ESI-Q-TOF instrument (see also 2.2.24.2) an ESI ion trap (LC/MSD Trap XCT Ultra II) instrument was used to detect peptides that were identified and listed using the mascot search engine (Matrix Science).

2. Material and methods

2.2.24.2. Analytical covalent cross-linking

In-gel digestion of protein samples by trypsin to investigate cross-linking reactions and mass spectrometry data collection was done in cooperation with the research group of Prof. *H. Schlüter* (UKE, Hamburg), particularly *Marcel Kwiatkowski*. Peptides were analyzed by a nanoUPLC device coupled to an ESI-Q-TOF system (QTOF Premier, Micromass/Waters, UK). Peptides eluted from the precolumn (nanoACQUITY UPLC column, C18, 180 μm \times 200 mm, Waters, UK; solvent A: 0.1% FA; solvent B: 99.9% ACN, 0.1% FA; flow rate: 5 $\mu\text{l min}^{-1}$) towards a separation column (nanoACQUITY UPLC column, C18, 75 μm \times 150 mm, Waters, UK; flow rate: 300 nl min^{-1}). Full scan spectra were recorded and analyzed in DDA mode (m/z range: 300-1,800; scan rate: 0.6 s; inter-scan-delay 0.05 s) Two most intense peaks of the full-scan spectrum were selected for fragmentation (fragmentation m/z range: 100-1,800; scan rate: 3.95 s; inter-scan-delay 0.05 s). Selected m/z values were blocked for 0.5 s after fragmentation by the DDA algorithm. Cross-linking reagents are listed in table 12 below. In case of BS2G (bis[sulfosuccinimidyl] glutarate) a mixture of unmodified BS2G- d_0 and a heavy isotope (BS2G- d_4 ; two deuterium labels at carbon atom 2 and 4 each) carrying four deuterium labels was used in one reaction as a “mass-label” to detect cross-linking sites. Detection of both isotopes improves the credibility of a putative cross-linking site. BS2G is an *N*-hydroxy-succinimide ester forming an amide bond with primary amino groups, glutaraldehyde links amino groups as well, while EDC (1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride) requires one carboxy group and a slightly acidic buffer, MES pH 6.5 was used, to complete the two-step reaction. A 10 μM SARS-CoV pp710 solution at pH 7.9 was mixed with an equal volume of EDC (freshly dissolved in 500 mM MES, pH 6.5; final concentration 1 g l^{-1}). The reaction was quenched by adding DTT to a final concentration of 5 mM. The cross-linking reaction with BS2G was performed as specified by the protocol of the supplier Thermo scientific. Reaction products were separated by SDS-PAGE. Some products were cut from an SDS-PAGE gel for trypsin digestion followed by mass spectrometry. Additionally, a negative control without cross-linking reagent was included in mass spectrometry experiments. For cross-linking by EDC, the obtained spectra were analyzed by XQuest [161], BS2G linked peptides were identified by the XQuest/XProphet software package [161; 162] in the iontag mode. For XQuest/XProphet data evaluation, peptides of 3-50 amino acids in length with up to 4 missed trypsin cleavage sites were taken into account, presuming a minimum charge of 2 for cross-linked peptides. Fixed mass modifications are 57.0215 Da and 15.9949 Da. All mass tolerance values were set to default parameters.

Table 12. Cross-linking compounds

cross-linking reagent	cross-linking site	spacer arm length	mass / cross-link mass difference [g mol ⁻¹]
glutaraldehyde (AppliChem)	primary amino groups	5 Å	100.12 / 64.08
EDC-HCl* (AppliChem)	links pr. amino gr. with carboxy group	0 Å	191.7 / -18.02
BS2G-d ₀ /d ₄ (Thermo Scientific)	primary amino groups	7.7 Å	530.35 (534.37) / 96.02 (100.04) [monolink: 114.03 (118.05)]

* 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride

¹ bis[sulfosuccinimidyl] glutarate

2.2.24.3. Native mass spectrometry

Native mass spectrometry experiments were performed by Dr. *Charlotte Utrecht* (European XFEL GmbH, Hamburg; Uppsala University, Sweden). Aiming to analyze protein-protein interactions by native mass spectrometry including collision-induced dissociation (CID), the proteins of interest were applied in a buffer containing 350 mM ammonium acetate adjusted to pH 7.9. Ammonium acetate provides a valuable buffer capacity in a wide pH range. Samples were supplemented with up to 1 mM DTT to reduce intermolecular disulfide bonds and stabilize the activity of SARS-CoV M^{pro}, if a cleavage reaction is investigated. M^{pro} wt and M^{pro} R298A were applied in a buffer containing 250 mM ammonium acetate adjusted to pH 7.4. Gold-coated pointed glass capillaries were prepared by a micropipette puller (Sutter Instruments) and filled with the sample typically containing 5-10 µM protein for subsequent low-speed electrospray ionization at an ESI-TOF LCT (Waters, UK) or ESI-Q-TOF instrument (Q-TOF II; Micromass/Waters) [163]. Q-TOF experiments (in MS/MS mode) allowed analysing the composition of protein complexes by collision-induced dissociation in a xenon gas collision chamber, essentially as previously described [164]. Xenon at a pressure of 1-2 · 10⁻² mbar was used as collision gas. Caesium iodide allowed calibration of the spectra. The capillary voltage was in the range 1,300-1,500 V, the sample cone ranged from 120 to 200 V. Spectra were processed by Massign [165] starting with linearization of the m/z-axis and including simulation of component spectra and mass assignment.

2.2.25. *In silico* docking/modelling

2.2.25.1. Online server

The I-Tasser homology modelling server was used to calculate/obtain homology models (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>) [166]. The FTmap server provided grid-based small-molecule docking (protein “hot spot” mapping) applying a fast Fourier transform correlation algorithm (<http://ftmap.bu.edu/>) [167]. The electrostatic surface potential of HCoV-229E Nsp10 was calculated by the PBEQ-Solver (<http://www.charmm-gui.org/input/pbeqsolver>) [168]. Additionally, SITEHOUND-web [169] was used for ligand binding site prediction (<http://sitehound.sanchezlab.org>).

2. Material and methods

2.2.25.2. *In silico* modelling of the interaction of Nsp10 with Nsp16

SARS-CoV Nsp10 interacting with SARS-CoV Nsp16 (pdb code 2XYQ) was used as a template for *in silico* modelling of the interface of the homologue proteins by Sybyl-X 1.2 (SYBYL, Tripos International, St. Louis, USA). SARS-CoV Nsp10 was superposed (“fit monomer” option) with HCoV-229E Nsp10 and subsequently replaced, while SARS-CoV Nsp16 was similarly superposed by a homology model [I-Tasser modelling server 166, RMSD 2.7 ± 2.0 Å; C-score: 1.73, TM-score: 0.96; 57% sequence identity] to obtain the homologue complex of HCoV-229E by the command “merge”. All four combinations including the hypothetical chimeric complexes involving a SARS-CoV and an HCoV-229E protein were modeled and comparatively analyzed for putative non-covalent interactions at the interface. Evaluation was supported by the PISA server (EMBL-EBI). A quantitative surface charge distribution was calculated using the charge model according to Gasteiger/Hückel. Images were created by PyMOL Molecular Graphics System (Schrödinger).

2.2.25.3. Docking

The subprogram *Flexidock*TM implemented in Sybyl-X 1.2 was used for docking of modified peptide substrates to the surface of SARS-CoV M^{pro}. Water was removed; charges were calculated according to Gasteiger-Hückel. The maximum number of iterations was 200,000. The peptides were prepositioned at the active site taking the approximate position of the glutamine residue of a ligand in an M^{pro} crystal structure [78] into consideration.

3. Results and discussion

The aim of this study is to obtain new insights into the structure and the function of pp710, i.e. Nsp7-10 of two different distantly related coronaviruses and its proteolytic processing into mature Nsps. SARS-CoV emerging in the beginning of 2003 is a highly pathogenic *Betacoronavirus* like the currently emerging MERS-CoV, while the *Alphacoronavirus* HCoV-229E that is usually not lethal upon infection was chosen to be the second source of pp710. As a pre-requirement to perform this study, the pp710 genes as well as smaller processing fragments were successfully cloned and overexpressed to allow protein purification. The large number of completely sequenced coronavirus genomes already available in databases provides a basic understanding of the genome structure as described in the introduction section. A fundamental biochemical and biophysical characterization of pp710 of HCoV-229E as well as SARS-CoV, both comprising around 58 kDa of the C-terminus of the associated polyprotein pp1a, but lacking the Nsp11 peptide, has been performed. Due to a high degree of structural similarity, pp710 of HCoV-229E also served as a template/substrate for SARS-CoV main protease Nsp5 to study the proteolytic processing reaction and to further purify mature Nsps, which allowed for instance the structural characterization of HCoV-229E Nsp10 by SAXS and X-ray crystallography.

3.1. Sequence analysis and cloning

Comparing pp710 of HCoV-229E and SARS-CoV the amino acid sequence identity is 47% (fig. 7; table 13). The percentage of identical residues within the individual Nsp domains ranges between 52.6% for Nsp10 and 38.6% for Nsp7. HCoV-229E Nsp8, Nsp9 and Nsp10 are slightly shorter than their homologues in SARS-CoV. Interestingly, SARS-CoV Nsp10 (*Betacoronavirus*) shares a higher sequence identity with Nsp10 of the *Gammacoronavirus* IBV than with that of other *Betacoronaviruses*, e.g. MHV and BCoV, as well as with that of the distantly related HCoV-229E (*Alphacoronavirus*). It highlights that the virus phylogeny strongly depends on the applied marker sequence. Table 13 lists characteristics of HCoV-229E and SARS-CoV pp710 as well as processing products and mature Nsps that represent targets proteins of this study. Sequences were obtained from the NCBI nucleotide database and the Uniprot database.

Referring to the aims of the project, the gene of polyprotein Nsp7-10 (pp710) of HCoV-229E cloned into pET11d and the cDNA of the SARS-CoV replicon including all 16 non-structural proteins (both kindly provided by Prof. R. Hilgenfeld, University of Lübeck) were used as a template for PCR to clone the pp710 gene of both viruses as well as fragments of it (tab. 13). into plasmids suitable for heterologous overexpression in bacteria. The gene fragments were amplified, cleaved with the appropriate restriction enzymes to generate “sticky ends” and ligated into the cleaved vector as

3. Results and discussion

described in the methods section. HCoV-229E pp78 (polyprotein Nsp7-8) cloned into pGEX 6p-1 offering an N-terminally GST-tagged fusion protein was kindly provided by *Y. Xiao* (University of Lübeck). The vector pGEX-6p-1 carrying the SARS-CoV M^{pro} gene (Nsp5) was provided by *Prof. Hilgenfeld* as well. It contains an N-terminal GST-tag that is autoproteolytically processed after translation to generate the authentic N-terminus and a C-terminal his₆-tag cleavable by precision protease (rhinovirus 3CL protease; GE Healthcare). Bacteria culture conditions are listed in table 30 in the appendix.

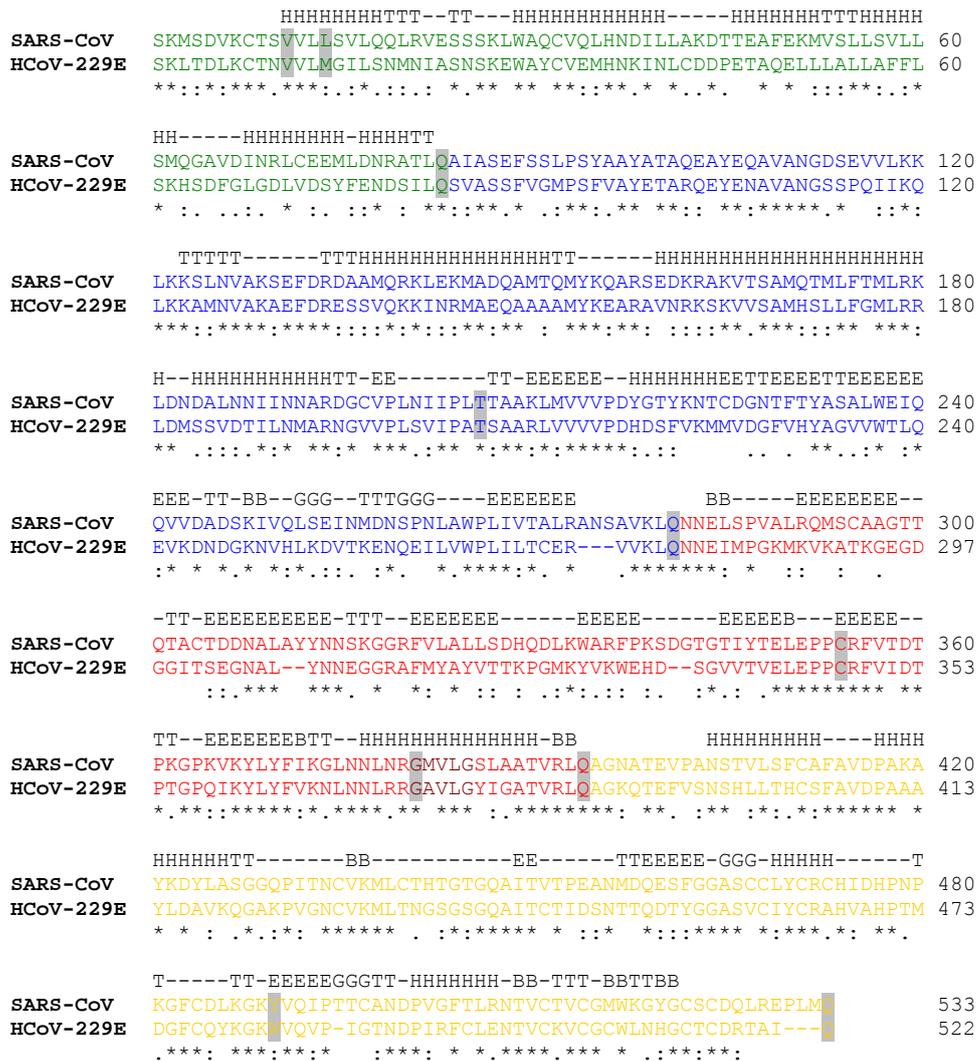


Figure 7. Alignment of polyprotein Nsp7-10 of SARS-CoV and HCoV-229E. Secondary structure information (*black*) is based on SARS-CoV Nsp7 (*green*; pdb code 2KYS), Nsp8 (*blue*; pdb code 2AHM), Nsp9 (*red*; pdb code 2J97) and Nsp10 (*yellow*; pdb code 2FYG). H: α -helix; T: turn; B: β -bridge; E: β -sheet; G: 3_{10} helix. Amino acid positions targeted by mutagenesis for structural investigation starting in chapter 3.5 are highlighted in *grey*. A reported dimerization motif of mature Nsp9 is coloured *dark red* [102]. Polyprotein Nsp7-10 cloned into pET11d contains a C-terminal his₆-tag without additional artificial amino acids. Pp710 of SARS-CoV contains either a strep- or his₆-tag provided by the plasmids pASK-IBA3plus, pASK-IBA 5plus, pASK-IBA 33plus and pASK-IBA 35plus.

Table 13. Overview of target proteins of HCoV-229E (blue) and SARS-CoV (black) obtained by gene cloning and recombinant bacterial expression

protein	length [amino acids]	molecular mass [Da]	theor. pI*	sequence identity with SARS-CoV [identical residues/total length (% of identity)]
pp1a (HCoV-229E; Nsp1-11)	4085	n. d.	n. d.	n. d.
pp710 (HCoV-229E)	522	57310.9	8.03	246/522 (47.1%)
pp78 (HCoV-229E)	278	30923.7	n.d.	125/278 (45.0%)
Nsp7 (HCoV-229E)	83	9300.6	4.38	32/83 (38.6%)
Nsp8 (HCoV-229E)	195	21623.1	9.32	93/195 (47.7%)
Nsp9 (HCoV-229E)	109	12045.8	9.24	50/109 (45.9%)
Nsp10 (HCoV-229E)	135	14395.3	7.66	71/135 (52.6%)
Nsp11 (HCoV-229E) [SFDNSYLNESGALVPLD]	17	1840.9	3.49	n.d.
pp1a (SARS-CoV)	4382	n. d.	n. d.	-
pp710 (SARS-CoV)	533	58324.9	7.33	-
pp79 (SARS-CoV)	394	43584.0	7.99	-
pp78 (SARS-CoV)	281	31182.9	n.d.	-
Nsp7 (SARS-CoV)	83	9267.8	5.18	-
Nsp8 (SARS-CoV)	198	21915.1	6.58	-
Nsp9 (SARS-CoV)	113	12401.1	9.1	-
Nsp10 (SARS-CoV)	139	14843.9	6.3	-

*calculated by ProtParam ([142]; SIB)

3.2. HCoV-229E and SARS-CoV pp710

3.2.1. Gene overexpression in *E. coli* and polyprotein purification

C-terminally his₆-tagged HCoV-229E pp710 was successfully purified to more than 90% purity after Ni-NTA affinity chromatography as judged by SDS-PAGE (fig. 8A). Compared to a one-step elution with 400 mM imidazole the purity was improved by applying a stepwise increase of imidazole (15-400 mM) for protein elution. The identity of the protein was verified by ESI-MS/MS after trypsin digestion, applying soluble protein directly from an affinity chromatography elution fraction and additionally a single coomassie-stained band that was cut out of an SDS-PAGE gel (data not shown).

To further increase the purity and to analyze the aggregation state of HCoV-229E pp710 in solution, a Hi-Load 16/60 Superdex 200 size-exclusion chromatography column was applied (fig. 8B). Depending on the buffer conditions a maximum of four peaks corresponding to different oligomeric states were resolved. The peak identity of one SEC experiment (fig. 8B, *dark blue* chromatogram) was exemplarily verified by reducing SDS-PAGE (fig. 8C). SEC indicated the formation of dimeric and

3. Results and discussion

higher oligomeric states in solution. The dimeric state is indicated by approx. the mol weight characterizing peak III at $V_E = 66.2$ ml ($k_{av} = 0.30$) obtained under non-reducing conditions. Analyzing protein from the dimer peak by SDS-PAGE showed a predominant protein band corresponding to the MW of a pp710 monomer, indicating the major presence of a non-covalent dimer in solution. However, if the SEC is performed with reducing conditions (*yellow*), peak III is slightly reduced, confirming also the presence of disulfide-linked dimers. The comparison of SEC chromatograms obtained by applying different protein concentrations (fig. 8B, *dark blue* and *red* chromatogram) and buffer compositions (fig. 8B, *green* and *red* chromatogram) revealed that the proportion of oligomeric states is obviously concentration- and buffer-dependent. A peak corresponding to monomeric protein elutes at $V_E = 77.7$ ml ($k_{av} = 0.45$), which increased in absorption under reducing conditions, by high ionic strength and at low protein concentration (fig. 8). Isoelectric focussing of HCoV-229E pp710 revealed a band at pH 6.0 (data not shown), lower than pH 7.4 as calculated by ProtParam for pp710 with his₆-tag ([142], SIB).

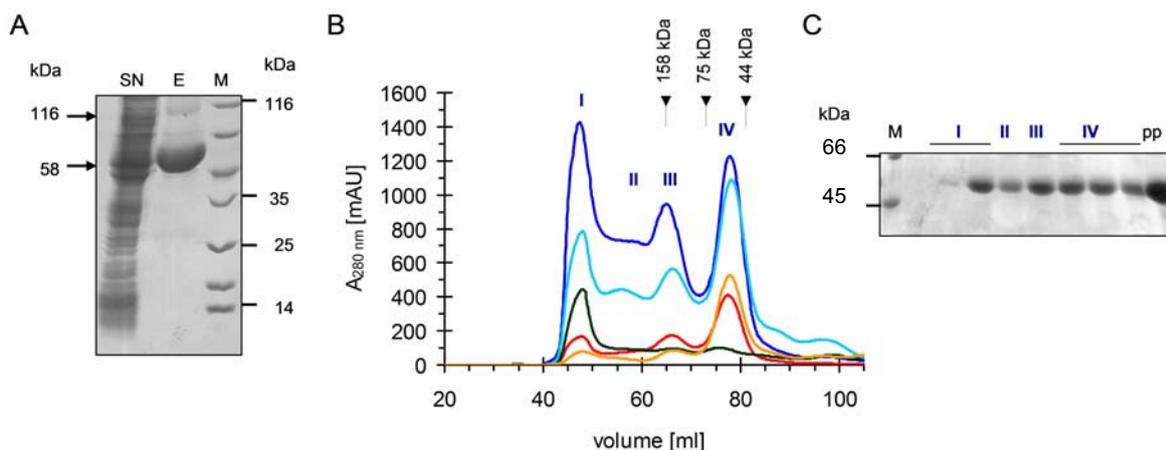


Figure 8. Purification of **HCoV-229E pp710**. **(A)** The success of the purification procedure is visualized by non-reducing SDS-PAGE and coomassie staining [SN: supernatant after disrupting the harvested *E. coli* cells and centrifugation; E: protein eluted from the Ni-NTA affinity resin; M: molecular weight marker]. A small amount of covalent dimer was identified (116 kDa) next to the monomeric pp710 (58 kDa). **(B)** The aggregation state of pp710 depending on the buffer conditions was analyzed by SEC (HiLoad 16/60 Superdex 200). The molecular weight calibration of the column by standard proteins is shown. For each experiment, 5 ml of protein solution were loaded onto the column. *Dark blue*: 300 mM NaCl, 30 mM NaH₂PO₄, pH 7.9; *cyan*: 300 mM NaCl 30 mM Tris pH 7.9; *dark green*: 120 mM NaCl, 20 mM NaH₂PO₄ pH 7.9; *yellow*: 300 mM NaCl, 30 mM NaH₂PO₄ pH 7.9 3 mM DTT; *red*: 400 mM NaCl, 5% (v/v) glycerol, 40 mM NaH₂PO₄ pH 7.9. **(C)** The fractions corresponding to the absorption peaks I-IV of one experiment (*dark blue* chromatogram) were analyzed by reducing SDS-PAGE [pp: pp710 SEC column input; I-IV: SEC peaks].

Applying a similar one-step purification strategy, four variants of SARS-CoV pp710 were successfully purified by affinity chromatography, i.e. an N-terminal and a C-terminal his₆-tag variant using a Ni-NTA resin column (SDS-PAGE gel, fig. 9A) as well as an N-terminal and a C-terminal strep-tag variant using Strep-Tactin resin. Solubility and purity of the differently tagged SARS-CoV proteins turned out to be highly similar to HCoV-229E pp710. Calibrated SEC revealed the existence of different oligomeric states in solution, including a monomeric and a dimeric state (fig. 9B), as previously observed for the homologue HCoV-229E pp710 (fig. 8). However, the SEC analysis of pp710 from both coronaviruses (fig. 8B, 9B) consistently showed that the pp710 monomer is the predominant state in solution, compared to other oligomers.

Interestingly, supplementing the *E. coli* expression culture of SARS-CoV pp710 with 500 μ M Zn²⁺ resulted in a significantly reduced amount of oligomers larger than a dimer (fig. 9B, *dark blue* chromatogram). The effect of Zn²⁺ may originate from two highly conserved Zn²⁺-binding sites within Nsp10. Investigating the isolated Nsp10 domain of MHV, it was previously shown that Zn²⁺ significantly altered the oligomerization behaviour by partly de-aggregation, as determined by dynamic light scattering techniques [170].

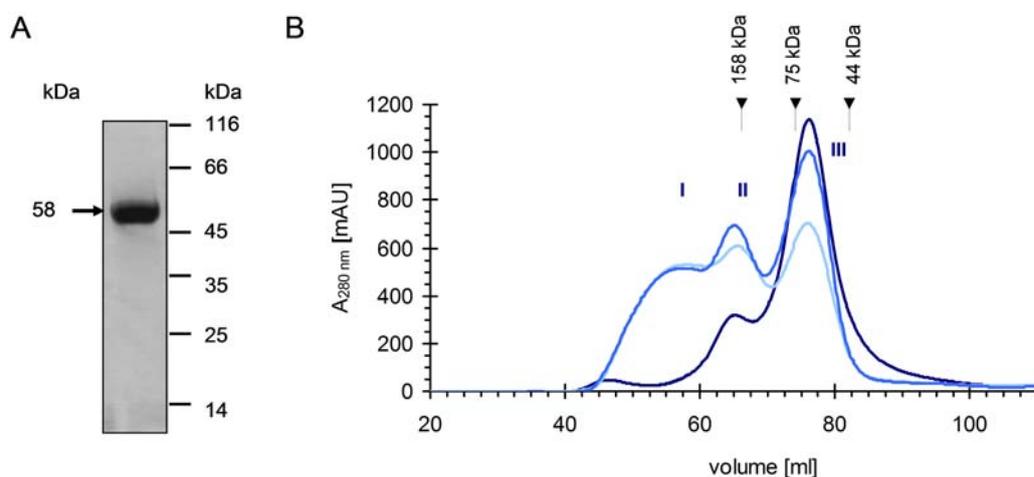


Figure 9. (A) Purified SARS-CoV pp710 visualized by reducing SDS-PAGE and coomassie staining. (B) Calibrated SEC (Hi-Load 16/60 Superdex 200) reveals two well defined peaks (II and III) conserved in all chromatograms, corresponding to a dimeric and monomeric state in solution, respectively. The *light blue* chromatogram was obtained in 280 mM NaCl and 30 mM NaH₂PO₄, pH 7.9. Exchanging sodium phosphate by 30 mM Tris resulted in the *blue* chromatogram. A significantly reduced amount of higher oligomers was observed after supplementing the *E. coli* expression culture with 500 μ M Zn²⁺ and adding 2 mM DTT to the purified protein (*dark blue*).

3. Results and discussion

In summary, SEC revealed a co-existence of different oligomeric states of both purified pp710 homologues in solution which is subsequently investigated in more detail using native PAGE, DLS and SAXS techniques.

3.2.2. Variable oligomerization of pp710

3.2.2.1. Blue native PAGE & Elman's reagent

Addressing to investigate the number, the approximate size and the quantity of oligomeric states of pp710 in solution, blue native PAGE was performed. During this electrophoresis technique the proteins are additionally negatively charged on the surface by interaction with a trace amount of coomassie stain. Nevertheless, the native state of a protein remains widely conserved [171], allowing studying protein-protein interactions with low sample consumption. Analysing both homologues, a comparable number and estimated quantity of four oligomeric states up to a putative octamer were consistently identified for HCoV-229E and SARS-CoV pp710 by native blue PAGE under non-reducing and even under mildly reducing conditions (fig. 10A/C). The monomeric state of both homologue proteins appears to be the most abundant one.

However, for HCoV-229E pp710, a significant amount of oligomers with a molecular weight above 400 kDa was observed after incubation of the sample purified by affinity chromatography for 7 days (fig. 10A). Probing the time-dependent oxidation of cysteine side chains, i.e. the formation of disulfide bonds, by Elman's reagent (5,5'-dithiobis -2-nitrobenzoic acid, DTNB) under non-reducing conditions revealed a continuous oxidation down to around two reduced thiol groups per molecule after 18 days (fig. 10B). Consequently, the high molecular weight oligomers of HCoV-229E are mainly a result of covalent disulfide bonds formation due to aerobic incubation of pp710 in solution. This is further supported by the observation that reducing conditions partly disassemble these oligomers, resulting in a significant increase of the amount of monomeric and dimeric states in solution (fig. 10A). The remaining high molecular weight fraction is stable in reducing conditions, indicating non-covalent interactions. Blue native PAGE of large HCoV-229E pp710 oligomers separated by SEC (fig. 8B, peak I, *dark blue* chromatogram) showed a disassembly into smaller states up to a monomer after 7 days incubation (not shown), attributed to the non-covalently linked oligomers. Remarkably, two thiol groups of HCoV-229E remained reduced for at least 18 days at 6 °C under oxidizing conditions (fig. 10B). One of them might be C281 within Nsp8 that is buried according to the high resolution structure (pdb code 2AHM) [98]. The second cysteine residue could not be attributed so far.

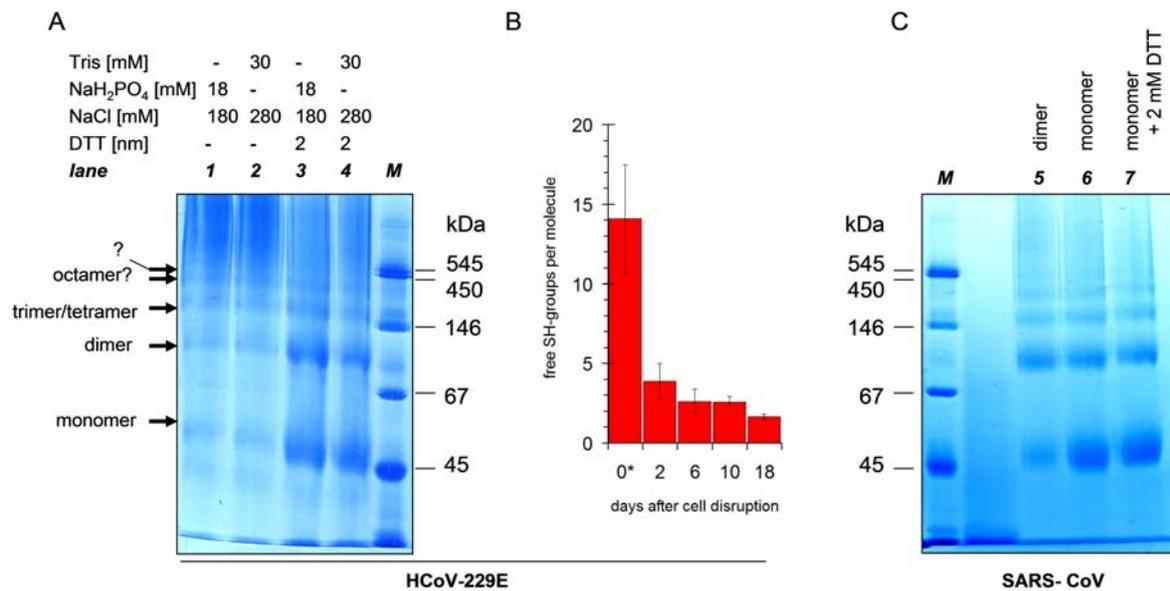


Figure 10. Investigation of the pp710 oligomerization. (A) Blue native PAGE of HCoV-229E pp710: Lane 1-4: Pp710 purified by affinity chromatography and incubated for 7 days at 6 °C; the buffer composition is displayed above the gel. The pH was consistently adjusted to pH 7.9. (B) Stepwise oxidation of thiol groups of HCoV-229E pp710 photometrically determined by DTNB up to 18 days after cell disruption. *approximately 3 h after cell lysis in elution buffer. (C) Blue native PAGE of SARS-CoV pp710: Lane 5: Dimer after incubation at 6 °C for 7 days; lane 6: Monomer, after 7 days; lane 7: Monomer, after 7 days supplemented with 2 mM DTT. The protein buffer contains 280 mM NaCl and 18 mM sodium phosphate, pH 7.9.

For freshly prepared SARS-CoV pp710, the monomeric and the dimeric state were isolated by SEC (fig. 9B, light blue chromatogram, peak III and II). The homogeneity of both samples was verified by dynamic light scattering (table 14, monomer and dimer sample). Incubating monomeric SARS-CoV pp710 for 7 days, oligomeric states appear on the gel, including a dimer, tetramer and octamer (fig. 10C, lane 6). Since these oligomers are stable at mild reducing conditions (fig. 10C, lane 7), a non-covalent interaction is suggested. The purified pp710 dimer also tends to form these oligomers after one week, even a slight disassembling into the monomeric state has been detected (fig. 10C, lane 5). Interestingly, high molecular weight oligomers as observed for HCoV-229E pp710 did not form to the same extent during incubation for 7 days (fig 10C, lane 5 and 6), which could be attributed to an altered distribution of cysteine residues in SARS-CoV pp710 (fig. 7).

Overall, the native PAGE experiments indicated an equilibrium of monomer and non-covalent dimer, as well as to a lesser extent higher oligomers for both pp710 homologues in solution. Moreover, disulfide bridged high molecular weight oligomers seem to represent artefacts that are formed by oxidation during storage under aerobic conditions.

3. Results and discussion

3.2.2.2. Dynamic light scattering (DLS)

To further characterize the stability of the isolated SARS-CoV pp710 monomer and dimer in solution and to analyze the influence of buffer conditions on the aggregation behaviour of monomeric pp710 and its high molecular weight oligomers, dynamic light scattering techniques were applied. This method is based on the scattering of a red-light laser on protein molecules that is focussed on the sample. The fluctuation of the scattering intensity due to Brownian motion of the molecules in solution is correlated with the hydrodynamic radius (R_h) of a particle, allowing the molecular weight estimation in homogenous solutions. Pure monomer and dimer fractions separated by SEC were investigated in comparison to a mixture of oligomeric states obtained directly after affinity chromatography.

Monomeric SARS-CoV pp710 isolated by SEC (fig. 9B, peak III) is characterized by a hydrodynamic radius of $R_h = 3.8 \pm 0.1$ nm, while an $R_h = 6.5 \pm 1.1$ nm was determined for pure dimeric pp710 (table 14). For the mixture of oligomeric states a hydrodynamic radius of 11.9 ± 3.8 nm was determined. To access the stability of monomeric SARS-CoV pp710, the solution was monitored by DLS for up to 7 days. After 12 hours the hydrodynamic radius was stable ($R_h = 3.8 \pm 0.3$ nm). Even after two days of incubation, only a marginally increased radius of $R_h = 4.1 \pm 0.1$ nm was detected, indicating the formation of a small amount of pp710 dimers. Within one week, the hydrodynamic radius of monomeric pp710 increased up to 6.9 ± 1.1 nm, highly similar to that determined for a pure dimeric state, confirming the formation of a slow equilibrium as observed by blue native PAGE (fig. 10C). However, the latter method additionally showed protein bands corresponding to higher oligomeric states after 7 days of incubation.

The presence of 1 mM DTT during incubation of the monomeric pp710 fraction slightly decreased the hydrodynamic radius after 7 days, supporting the coexistence of non-covalent and disulphide linked dimers. In contrast, large concentrations of reducing agents (10 mM β -ME, 30 mM TCEP) significantly increased the aggregation of SARS-CoV pp710. On the other hand, EDTA, $MgSO_4$, $CaCl_2$ or $ZnCl_2$ as well as 10% glycerol, 0.5% β -OG or 2 mM CHAPS did not affect the dimerization of pp710 over 7 days. Only 1% DDM slightly reduced the hydrodynamic radius to $R_h = 5.1 \pm 0.1$ nm, attributed to the partial disassembling of non-covalent dimers. Interestingly, the presence of 1 mM $MnCl_2$ induced the formation of large oligomers, suggesting specific manganese binding sites in pp710 of SARS-CoV. This is consistent with the report of a manganese-dependency of the RNA-dependant RNA polymerase activity of mature SARS-CoV Nsp8 [94].

Table 14. Summary of SARS-CoV pp710 (N-terminal strep tag) DLS experiments. The basic protein buffer included 280 mM NaCl and 30 mM NaH₂PO₄ at pH 7.9.

sample	additive	radius			count rate
		peak 1	peak 2	peak 3	
monomer	untreated	3.8 ± 0.1 nm	-	-	57.2 ± 1.3 kHz
dimer	untreated	6.5 ± 1.1 nm	-	-	45.3 ± 13.3 kHz
oligomer*	untreated	11.9 ± 3.8 nm	-	-	400 ± 22.4 kHz
monomer (after one week)	untreated	6.9 ± 1.1 nm	-	-	213.0 ± 10.5 kHz
	1 mM DTT	4.7 ± 0.2 nm	-	-	57.9 ± 1.2 kHz
	10 mM β-ME	16.2 ± 1.1 nm	-	-	1110.5 ± 34.6 kHz
	30 mM TCEP	17.1 ± 0.7 nm	87.1 ± 8.1 nm	-	1477.7 ± 85.8 kHz
	1% DDM	5.1 ± 0.1 nm	-	-	398.8 ± 22.2 kHz
	0.5% β-OG	7.2 ± 0.7 nm	-	-	175.2 ± 7.8 kHz
	10% glycerol	6.8 ± 0.2 nm	-	-	108.7 ± 2.8 kHz
	2 mM CHAPS	6.4 ± 0.7 nm	-	-	133.6 ± 9.1 kHz
	1 mM ZnCl ₂	6.0 ± 0.2 nm	-	-	113.3 ± 2.4 kHz
	1 mM MnCl ₂	11.4 ± 1.0 nm	78.7 ± 8.7 nm	-	652.1 ± 39.9 kHz
	1 mM MgSO ₄	7.0 ± 1.4 nm	-	-	160.3 ± 8.9 kHz
	1 mM CaCl ₂	7.0 ± 0.3 nm	76.8 ± 12.2 nm	-	148.6 ± 16.6 kHz
	1 mM EDTA	7.0 ± 0.3 nm	-	-	146.1 ± 5.2 kHz
	mixture of oligomeric states (after 30 min)	500 mM NaCl	7.8 ± 1.2 nm	17.3 ± 2.2 nm	-
1 mM DTT		9.8 ± 2.0 nm	21.5 ± 4.8 nm	-	659.4 ± 27.8 kHz
10 mM β-ME		11.1 ± 1.0 nm	20.2 ± 3.6 nm	-	935.4 ± 37.6 kHz
30 mM TCEP		100.9 ± 13.4 nm	-	-	3506.9 ± 358.7 kHz
0.5% β-OG		21.0 ± 1.9 nm	-	-	645.6 ± 27.8 kHz
2 mM CHAPS		18.3 ± 1.3 nm	-	-	783.5 ± 28.7 kHz
1 mM MgSO ₄		19.2 ± 3.7 nm	-	-	932.7 ± 46.5 kHz
1 mM CaCl ₂		35.1 ± 6.7 nm	-	-	1872.9 ± 102.4 kHz
1 mM EDTA		16.4 ± 4.5 nm	-	-	864.2 ± 34.1 kHz
mixture of oligomeric states (after one week)	500 mM NaCl	17.6 ± 3.0 nm	-	-	826.1 ± 28.1 kHz
	1 mM DTT	16.3 ± 2.6 nm	43.1 ± 7.2 nm	-	2317.9 ± 102.5 kHz
	10 mM β-ME	20.7 ± 2.7 nm	37.3 ± 10.7 nm	-	4300.1 ± 152.2 kHz
	30 mM TCEP	23.6 ± 3.2 nm	82.6 ± 19.8 nm	-	1892.7 ± 116.3 kHz
	0.5% DDM	7.5 ± 0.5 nm	20.7 ± 4.6 nm	143.8 ± 38.1 nm	725.4 ± 27.0 kHz
	0.5% β-OG	17.6 ± 3.1 nm	33.4 ± 7.2 nm	-	3038.8 ± 113.4 kHz
	10% glycerol	27.2 ± 4.6 nm	-	-	1131.2 ± 53.5 kHz
	2 mM CHAPS	38.1 ± 4.4 nm	-	-	2652.7 ± 112.6 kHz
	1 mM ZnCl ₂	71.3 ± 13.7 nm	-	-	1826.9 ± 104.7 kHz
	1 mM MnCl ₂	38.7 ± 7.4 nm	-	-	2983.6 ± 130.5 kHz
	1 mM MgSO ₄	18.8 ± 4.7 nm	-	-	2203.7 ± 81.7 kHz
	1 mM CaCl ₂	30.6 ± 4.1 nm	-	-	2799.8 ± 114.3 kHz
	1 mM EDTA	25.9 ± 3.7 nm	-	-	2402.9 ± 89.3 kHz

*mixture of oligomeric states of SARS-CoV pp710 after affinity chromatography purification

3.2.2.3. Native ESI-Q-TOF mass spectrometry

The previous investigations indicated the presence of a continuum of variable oligomeric states for pp710 in solution with a relatively high abundance of the monomeric and the dimeric state. A semi-quantitative method providing a very high mass accuracy is native mass spectrometry, which is applied to exactly determine the size and ratio of the different oligomeric states of SARS-CoV pp710 in ammonium acetate buffer at pH 7.9. Ammonium acetate is volatile in a vacuum and therefore a suitable salt for mass spectrometry. An ESI-Q-TOF instrument uses electrospray ionization (ESI) of the proteins that pass a quadrupole (Q) component on the way to the time-of-flight (TOF) mass detector. Mass spectrometry experiments were performed in cooperation with Dr. C. Uetrecht at the University of Utrecht.

3. Results and discussion

Native ESI-Q-TOF applying 5 μ M SARS-CoV pp710, purified by affinity chromatography, detected the coexistence of a series of different oligomers in solution, up to a state in the range of a decamer characterized by approximately 630 kDa (fig. 11A/B). In agreement with SEC experiments and blue native PAGE the monomeric state with a calculated mass of 59.695 kDa for pp710 with N-terminal strep tag is most abundant. The relative content of the putative monomer (\approx 60 kDa), dimer (\approx 120 kDa) and trimer (\approx 180 kDa) is determined to 22.9%, 19.8% and 10.7% of the total intensity at the TOF detector, respectively (fig. 11A). Therefore, with an increase in molecular weight up to a tetramer, the relative fraction of the respective state is indicated to decrease. Comparing panel A and B of figure 11, a small mass variation of monomeric pp710 probably due to a ligand or truncation of the termini is indicated at an increased collision voltage of 100 V. Two further masses assigned to the spectrum in panel B are $119,537 \pm 23$ Da and $179,273 \pm 30$ Da well fitting dimeric and trimeric SARS-CoV pp710, respectively.

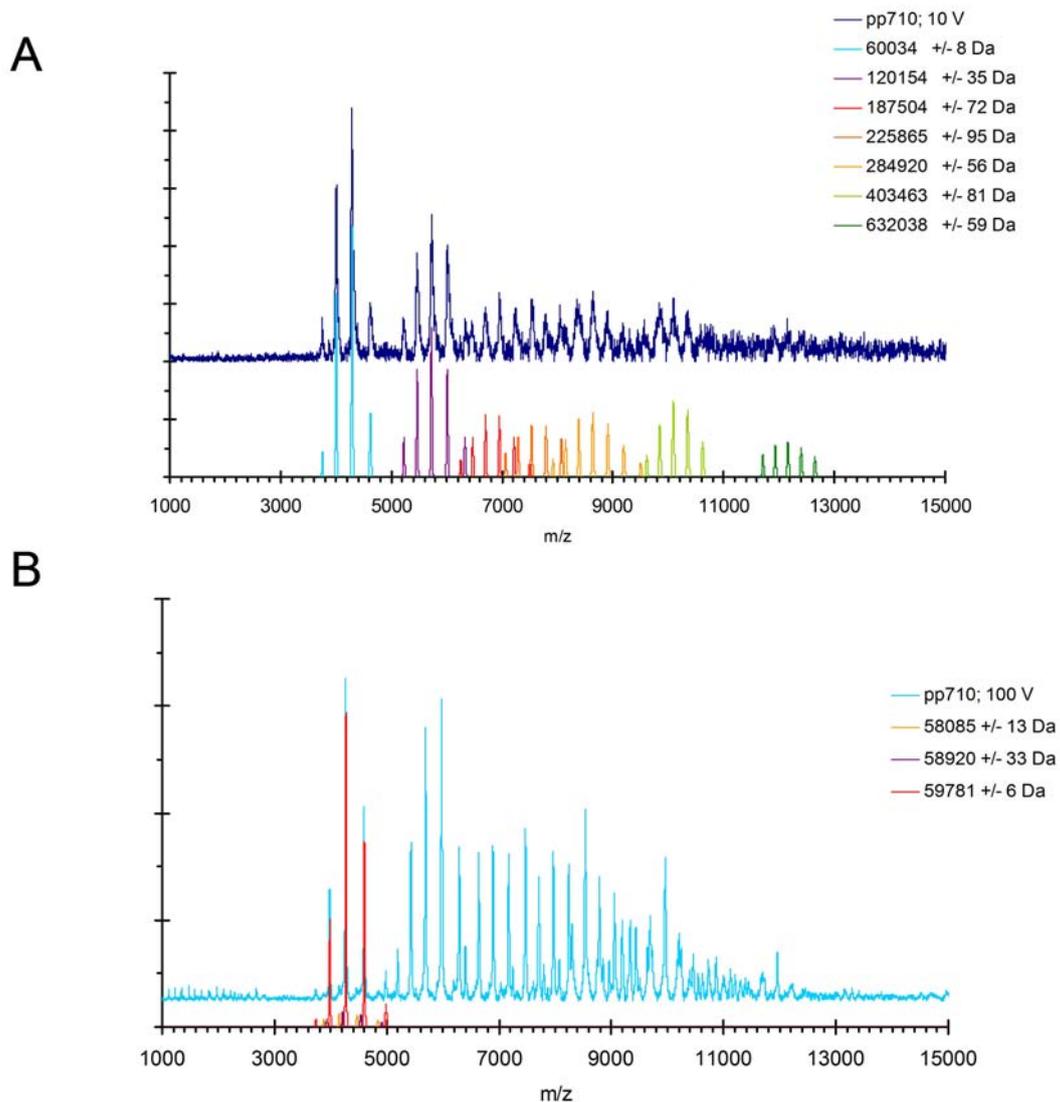


Figure 11. Oligomerization of SARS-CoV pp710 investigated by native ESI-Q-TOF MS. **(A)** Pp710 mass spectrum with different ion series after linearization, smoothing and background subtraction with the raw data (*dark blue*). The calculated mass of pp710 with N-terminal strep tag is 59.695 kDa. The expected oligomeric species were simulated by Massign and presented in the same spectrum (*different colours*). Different oligomeric states were detected in solution, i.e. a putative monomer (60,034 Da), dimer (120,154 Da), trimer (187,504 Da) and other states up to a putative decamer (632,038 Da). **(B)** The collision voltage is increased to 100 V for further fragmentation of oligomers/protein complexes (*light blue*). The persisting oligomers could be covalently linked. Simulated individual mass species that approximately fit the monomeric state are coloured and superposed with original data.

The results obtained for both pp710 homologues by SEC, blue native PAGE, DLS and native mass spectrometry essentially agree that the pp710 monomer, representing the most abundant state, is in equilibrium with a prominent dimer and with larger oligomeric states. The oligomerization process, triggered partly by non-covalent interactions, but also partly by disulphide formation, can be artificial or of relevance for the native function of pp710 *in vivo*. Larger oligomers detected by either method

3. Results and discussion

are slightly different in quantity and their maximum size, which might be related to individual limitations of detection and quantification by the respective method, e.g. incomplete ionization in mass spectrometry or different charge states of a protein in blue native PAGE.

However, oligomers of pp710 might be essential for their function. Extended oligomers of HCoV-229E Nsp9 were already suggested to be involved in RNA binding [101], probably involving a similar mechanism as known for single stranded DNA binding protein (SSB) of *E. coli* [172]. Other proteins that bind nucleic acids, e.g. the Trp RNA-binding attenuation protein (TRAP) and DNA replication associated proteins, like the processivity factor PCNA [173], form large ring structures directly associated with their function [174]. Orb2, a member of the RNA-binding CPEB family (cytoplasmic polyadenylation element-binding proteins), is even capable of forming amyloid fibers that are involved in long-term memory [175]. However, these proteins are suggested to be functionally only distantly related to pp710. Poliovirus RNA-dependent RNA polymerase (Mahoney type 1 poliovirus polymerase), which seems to be closely related to the putative function of Nsp8, forms a small oligomer as observed by X-ray crystallography that is reported to be essential for efficient RNA-binding and formation of the active centre [176]. Although high molecular weight oligomers have been detected for both pp710 homologues in this study, the applied techniques did not provide sufficient indications for the formation of ordered fibril-shaped or ring-shaped aggregates under *in vitro* conditions that may correlate with a coordinative “structural” function.

Different hetero-oligomeric states have been reported for complexes of mature Nsp7 and Nsp8 by high-resolution crystal structures. For SARS-CoV Nsp7 and Nsp8, a putative hexadecameric state was described that is assembled of eight subunits of Nsp7 and Nsp8 each [98], whereas FCoV Nsp7 and Nsp8 form a 2:1 heterotrimeric complex [96]. Both structures involve a specific Nsp7 dimer interface. Although HCoV-229E and SARS-CoV pp710 unlikely form a significant hexadecamer-like state in solution, as previously shown, mentioned complexes reveal that specific non-covalent Nsp-Nsp interactions exist, which could be also involved in the assembly of pp710 oligomers. Moreover, these examples additionally highlight that a comparison of different Nsps/polyproteins of distantly related coronaviruses is highly important. Mature Nsp9 of HCoV-229E forms a disulphide-linked dimeric state in solution that is reported to have a higher affinity to nucleic acid [101], indicating that at least the detected covalently linked dimers of both pp710 homologues could be of functional relevance or of regulatory impact. In this context, it was for example observed that DNA-binding specificity for the tumour repressor protein p53 is dependent on the reduced state of cysteines [177]. Despite the unfavourable redox potential in the cytosol, disulfide formation is also critical for the function of NF- κ B, PaX, Fos, OxyR and eukaryotic transcription factors (for review see [178]). Additionally, non-covalent oligomerization beyond a dimeric state was reported for mature HCoV-229E Nsp9, triggered by the presence of sulfate ions [101]. A comparable specific oligomerization effect was not detected

for SARS-CoV and HCoV-229E pp710 up to a concentration of 1 mM MgSO₄. This could be due to sterically hindered sulphate binding sites within the polyprotein.

The oligomerization of pp710 that is under investigation in this study may result from a combination of different (independent) homo-dimerization or homo-oligomerization modes, e.g. via Nsp-Nsp homointeractions which would favour higher oligomers in the polyprotein state that are probably irrelevant and decreased in size after cleavage/maturation, which is subsequently investigated in more detail. However, intermolecular interactions of the individual Nsp domains within the polyprotein oligomers appear at least weakened under the chosen experimental conditions compared to that of mature Nsps. For the complex of SARS-CoV Nsp7 and Nsp8 [98] or for HCoV-229E Nsp9 [101], the monomeric state is not the most abundant one under a wide range of conditions, in contrast to both pp710 homologues.

3.2.3. Structural characterization of pp710

3.2.3.1. Secondary structure content

To compare the secondary structure composition of *Alphacoronavirus* pp710 (HCoV-229E) and *Betacoronavirus* pp710 (SARS-CoV), circular dichroism (CD) spectroscopy was applied. Far-UV CD-spectra, i.e. the ellipticity of the sample in the wavelength range from 190 nm to 260 nm, were recorded (fig. 12A). The applied buffer contained 250 mM NaF and 20 mM NaH₂PO₄ at pH 7.9. Sodium chloride was substituted by sodium fluoride to reduce the absorbance of circular polarized light by the buffer during the experiment. Comparing HCoV-229E and SARS-CoV pp710, almost superimposable ellipticity curves (fig. 12A) indicate a highly similar secondary structure composition for both pp710 homologues.

Two distinct minima at a wavelength of 207 nm and 222 nm are characteristic for a predominant α -helical secondary structure content. This reflects the secondary structure of mature Nsp7 and Nsp8, as revealed by known high-resolution structures [96; 98]. Small fractions of β -sheet and random coil structure have also been deduced as well from the spectra in figure 12A, which is in agreement with a high β -sheet content of HCoV-229E Nsp9 [101] and SARS-CoV Nsp9 [102]. The far-UV spectrum of HCoV-229E is not affected by sodium acetate buffer (pH 5.0), by replacement of NaF by NaCl, or by a four-fold molar excess of ZnCl₂ (not shown).

For an initial comparison of the tertiary structure, near-UV spectra ranging from a wavelength of 250 to 320 nm were recorded (fig. 12B), representing a structural “fingerprint” of the protein. The increased absolute ellipticity of HCoV-229E pp710 around 260-270 nm may correspond to additional disulfide bonds. The peak positions in this area are sensitive towards the local orientation of phenylalanine residues (255-260 nm) and tyrosine (260-275 nm). In addition, an ellipticity maximum

3. Results and discussion

at approximately 290 nm corresponding to the absorbance of tryptophan is highly conserved for all analyzed pp710 samples of HCoV-229E and SARS-CoV pp710, including a purified monomer and a mixture of oligomeric states respectively. This consistency indicates consistent characteristics of the tertiary structure, even for different oligomeric states of pp710. Beside differences in the tertiary structure of both pp710 homologues indicated by the near-UV spectrum, the secondary structure is highly similar. Moreover, a mixture of different secondary structure elements with a dominating fraction of α -helical structure initially well reflects the sum of the secondary structure composition of the four mature Nsp domains.

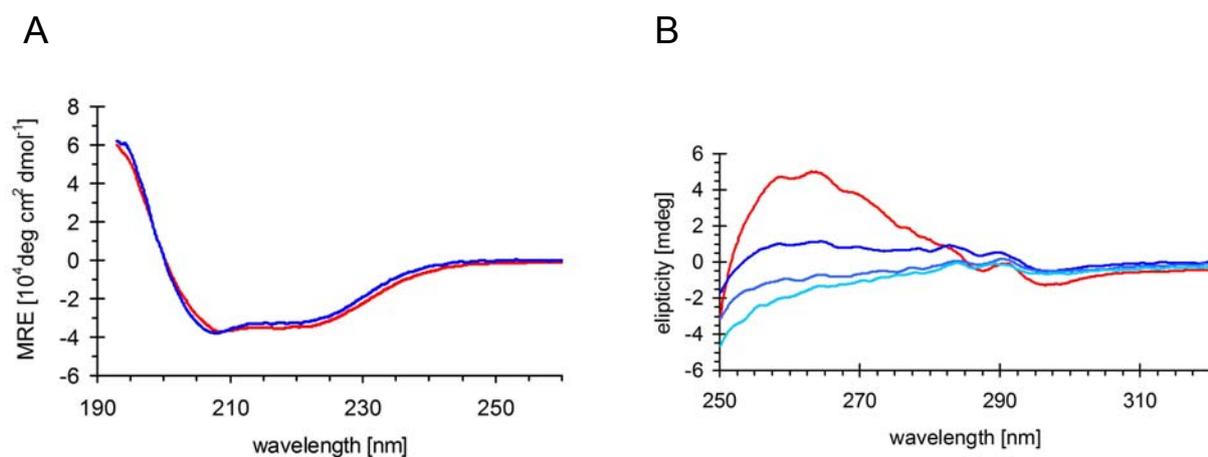


Figure 12. (A) Far-UV CD spectra of SARS-CoV pp710 (*blue*) and HCoV-229E pp710 (*red*). (B) Near-UV "fingerprint" spectra (right panel): HCoV-229E pp710 (*red*) purified by affinity chromatography and treated with benzonase in 280 mM NaCl 30 mM NaH_2PO_4 pH7.9. Monomeric SARS-CoV pp710 purified by SEC (*dark blue*), a mixture of oligomeric states purified by affinity chromatography (*mid blue*) and the same mixture supplemented with 1 mM DTT (*cyan*) are shown for comparison.

3.2.3.2. Tertiary and quaternary structure of pp710

High resolution X-ray crystallography provides essential insights into protein molecules; however it cannot be ruled out that there are structural artefacts due to flexibility, diverse oligomerization, a high protein concentration or non-physiological compounds in the solutions during crystallization. Therefore, a combination with "in-solution methods" like small-angle X-ray scattering (SAXS) is highly beneficial and indispensable in order to analyze particularly partly flexible multi-domain proteins. In this project SAXS aims at analyzing and comparing the tertiary structure of HCoV-229E and SARS-CoV pp710, which should be interpreted by known high-resolution protein structures of individual domains. Monomeric and dimeric states of both pp710 homologues will be compared in more structural detail, having in mind that their secondary structure is widely conserved (fig. 12). In preparation for SAXS experiments, the monodispersity of the samples was verified by

DLS (see also chapter 3.2.2, table 15). The SAXS sample with the lowest concentration of both proteins was directly taken from SEC without modifications of the buffer or adjustment of the concentration; the concentration of the others was adjusted by ultra filtration. SARS-CoV pp710 scattering data was collected at DORIS beamline X33. The scattering pattern of monomeric and dimeric HCoV-229E separated by SEC were recorded at two different synchrotron radiation sources, since the advanced 3rd generation synchrotron source PETRAIII beamline P12 (DESY, Germany) became recently available in the course of this study. The scattering intensity was investigated up to a scattering vector s of 6 nm^{-1} .

Initially, it was observed that purification and SAXS data collection in the presence of 2 mM DTT reduced the determined radius of gyration (R_g) as a measure of size (and shape) of HCoV-229E pp710 by 5-10% (fig. 13A). This effect again demonstrates the existence of a minor fraction of covalently linked oligomers under non-reducing conditions, as previously observed by SEC and blue native PAGE (see section 3.2.1/3.2.2). In polydisperse solutions the determined R_g refers to an average value of all present particles. Moreover, it was observed that the radius of gyration R_g and the particle volume determined for SARS-CoV and HCoV-229E pp710 monomer and dimer samples are clearly concentration-dependent under non-reducing conditions and in the presence of 2 mM DTT (fig. 13A). To get rid of this property of pp710 during data evaluation, the scattering data was extrapolated to infinite dilution. It is concluded that the data of HCoV-229E recorded at beamline P12 in an early stage of operation and beamline X33 is highly similar (fig. 13B). While SAXS data is frequently visualized by the plot of the scattering intensity I versus scattering vector s (see figure 14 for comparison), figure 13B displays the alternative Kratky plot. In this type of data presentation, the scattering intensity I multiplied by the square of the scattering vector s , which directly depends on the scattering angle (see methods section), is plotted against the scattering vector s . For highly flexible particles or unfolded proteins the data points do not converge to the scattering vector axis at higher scattering angles, instead the product Is^2 is continuously increasing. Therefore, the plot allows a flexibility prediction, where in this case the bending of the graph deduced from the raw data is indicative for a multi-domain protein with a relatively low overall flexibility and a rather compact/rigid particle, particularly for the dimer. It is conceivable that high local flexibility within the monomer of pp710 of both virus species is additionally stabilized within the dimer, which could theoretically be a physiological relevance *in vivo*.

The approximated molecular weight after extrapolation of the scattering data to infinite dilution based on the porod volume well matches the value expected for a monomer and dimer of HCoV-229E and SARS-CoV, respectively. SAXS results are summarized in table 15.

3. Results and discussion

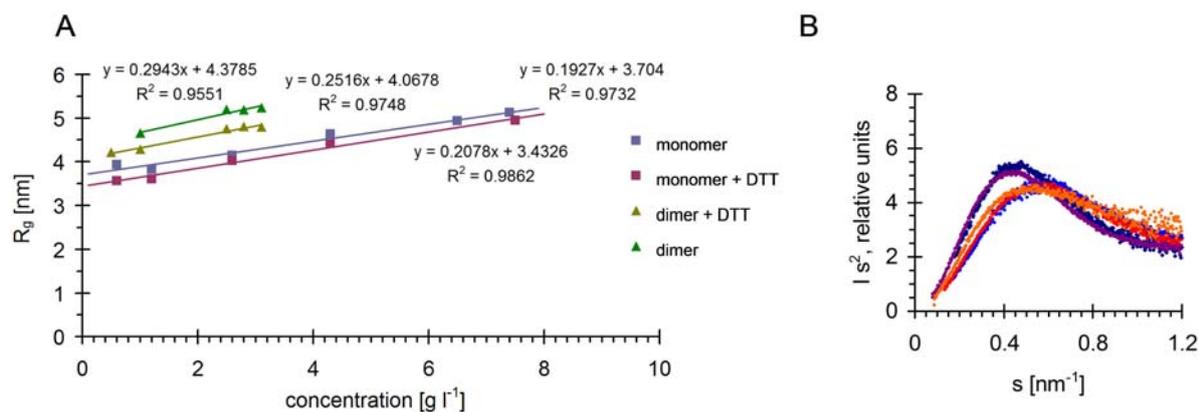


Figure 13. SAXS investigation of HCoV-229E and SARS-CoV pp710. **(A)** Concentration-dependent equilibrium of oligomeric states of HCoV-229E pp710 as investigated by SAXS. Monomeric and dimeric pp710 isolated by SEC were analyzed in the presence and absence of 2 mM DTT, as shown in the legend. **(B)**: Kratky plot ($I s^2$ versus s) comparing dimer (*dark blue*) and monomer (*blue*) samples of HCoV-229E investigated at PETRA III P12 and dimer (*purple*) and monomer (*red*) samples applied to beamline DORIS X33. SARS-CoV pp710 monomer (DORIS X33; *orange*) is shown for comparison, displaying slight intensity differences between 0.2 and 0.4 nm^{-1} that affect the radius of gyration R_g .

Table 15. Summary of size properties of pp710 of HCoV-229E and SARS-CoV under reducing conditions

virus	protein	R_h [nm] ¹	R_g [nm] ²	R_h from R_g [nm] ³	D_{max} [nm]	NSD ⁴
SARS-CoV ⁵	pp710 monomer	4.6 ± 0.2	3.99 ± 0.02^7	5.2	13.1	0.74 ± 0.05
	pp710 dimer	6.5 ± 1.1	5.42 ± 0.01	7.0	18.0	1.29 ± 0.08
HCoV-229E ⁶	pp710 monomer	4.8 ± 0.1	3.43 ± 0.10	4.4	10.5	0.72 ± 0.10
	pp710 dimer	6.5 ± 1.1	4.07 ± 0.08	5.3	14.3	0.74 ± 0.10

¹ hydrodynamic radius determined by DLS

² radius of gyration taken from the distance-distribution function (P(R)-function) of the molecule determined by GNOM and verified by AutoRg implemented in PRIMUS; commonly the raw scattering data was extrapolated to infinite dilution, if a concentration-dependency of the R_g was observed.

$$^3 R_h = R_g \sqrt{\frac{5}{3}} \quad (\text{approximation of } R_h, \text{ see method section})$$

⁴ normalized spatial discrepancy, calculated by the “automated mode” of DAMAVER involving at least 10 *ab initio* models calculated by DAMMIF. NSD is a quantitative measure of similarity of particles including all its components. For identical models the NSD is equal to zero [179].

⁵ 280 mM NaCl, 30 mM NaH_2PO_4 , pH 7.9, 2 mM DTT, proteins contains an N-terminal strep-tag

⁶ 300 mM NaCl, 30 mM NaH_2PO_4 , pH 7.9, 2 mM DTT, protein contains a C-terminal his-tag

⁷ reduced to 3.66 nm for protein originating from an expression culture supplemented with 500 μM Zn^{2+} and in the presence of 3 mM DTT instead of 2 mM DTT.

Characteristic parameters of the size and shape of both pp710 homologues in comparison to other related proteins and to SEC experiments are listed and discussed later (chapter 3.10, table 24 and 25). However, an increased radius of gyration and maximum diameter of monomeric and dimeric SARS-CoV pp710 compared to its homologue obviously indicates a slightly different geometry, i.e. a more elongated shape (tab. 15). Nevertheless, the approximated molecular weight of monomeric SARS-CoV and HCoV-229E pp710 is in the expected range, as further verified by an approach reported recently [180]. To apply this approach, as an example, the scattering pattern of SARS-CoV pp710 monomer was extrapolated to $I(0)$, i.e. obtaining the scattering intensity at an angle of $s = 0$ by the fit graph of the modelling software DAMMIF. This intensity was used to calculate the variable V_c according to [180], resulting in 483.9 \AA^2 and $Q_R = 6397.6 \text{ \AA}^3$. In a plot of I multiplied by s versus s , the area that is enclosed by the s -axis and the graph was determined by integration. For integration, the data was cut to a maximum scattering vector of 0.6 \AA^{-1} . This integral depends on the molecular weight of pp710 via a suggested power-law-relationship for protein samples. Using additional constants experimentally determined by Rambo & Tainer [180], the calculated mass for putative monomeric SARS-CoV pp710 was determined to 52 kDa (theoretical $MW_{pp710} \approx 59 \text{ kDa}$). This method is applicable for flexible/unfolded proteins as well, since in contrast to a Kratky plot (I multiplied by s^2 plotted against s), the applied plot I versus s is converging at higher angles anyway, as required for the calculation of the porod invariant Q . However, in this case the approach sufficiently proofed the existence of a monomeric state of SARS-CoV pp710 in solution, providing a valuable basis to model the shape of pp710.

3. Results and discussion

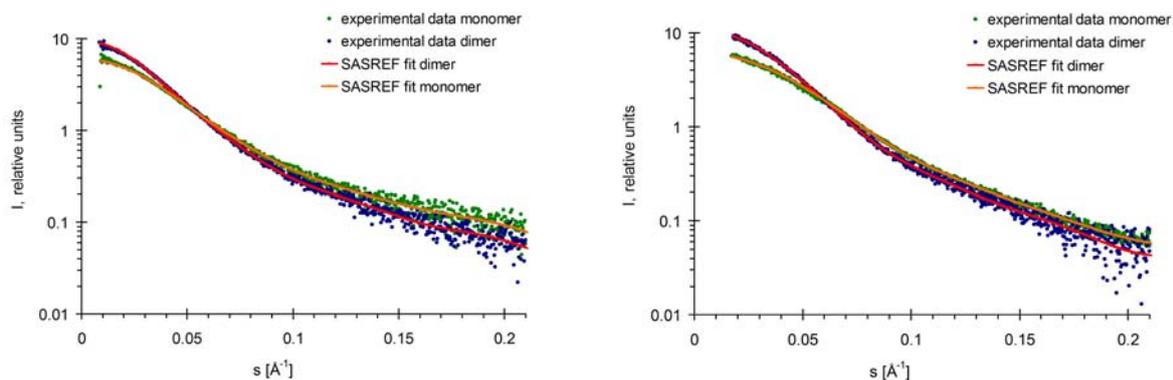


Figure 14. Small-angle X-ray scattering data of pp710 used for rigid-body modelling (*left panel*: SARS-CoV; *right panel*: HCoV-229E). The optimized fit graph of the model generated by SASREF based on high-resolution structures with modelling restrains described in the text is shown for comparison [pdb codes 2AHM (SARS-CoV Nsp7, Nsp8), 2J97 (HCoV-229E Nsp9)/2UW7 (SARS-CoV Nsp9), 2FYG (SARS-CoV Nsp10)]. These SASREF graphs demonstrate that the scattering data is well interpreted by the high-resolution structures even with the restraint of a covalent connection of the Nsp subunits. It is therefore assumed that the overall fold and shape of the mature Nsp is already preformed within the pp710 molecules. The χ -value calculated by SASREF is a quality measure of the similarity of scattering data and generated model. A low χ -value indicates a good data fit. For HCoV-229E pp710 monomer the χ -value is 1.38, for SARS-CoV pp710 monomer the χ -value is 0.97. [The value I in this figure is a measure of the relative scattering intensity; s is the scattering vector depending on the X-ray scattering angle 2θ and the constant wavelength.]

To calculate a low-resolution surface structure of both states the angle-dependent scattering intensity was extrapolated to a theoretical concentration of zero. The *ab initio* model of HCoV-229E pp710 monomer and dimer was created by DAMMIF, resulting in a set of well superimposable models. This is demonstrated by a low normalized spatial discrepancy (NSD; tab. 15), in other words a highly similar relative positions of the “dummy spheres” in different models. An exemplary *ab initio* model is displayed in figure 15 (grey spheres). A requirement for *ab initio* modelling is the distance distribution function (P(R)-function) obtained by indirect Fourier transform calculated by GNOM. The distance distribution function represents the abundance of distances between two random points in the molecule and is therefore characteristic for the shape of the molecule. The molecular weight of HCoV-229E pp710 dimer based on the calculated scattering intensity for $s = 0$ was approximated to 118 kDa (monomer expected: 59 kDa). Assuming P2 symmetry, the *ab initio* model for this HCoV-229E pp710 dimer is elongated with a thickened globular core and moderately curved lengthening (fig. 15). The maximum diameter is more than twice as high as the minimum diameter. Independent rigid-body modelling well reflects that shape (fig. 15). A rigid-body model aims to interpret scattering data by known structures of domains/fragments of the protein, for instance to additionally propose inter-domain contacts and the dimer interface of HCoV-229E pp710. The software SASREF optimizes the orientation of subunits in cycles of translation/rotation to optimally fit the SAXS scattering pattern (fit function of monomer/dimer, fig. 14), indicated by a low χ -value in the desired range of angles. For

rigid-body modelling (fig. 15) of the dimeric state using the subunits Nsp7, Nsp8, Nsp9 and Nsp10, the dimer and monomer scattering data were equally taken into account by SASREF to additionally verify the model. It is considered that dimerization does not significantly influence the tertiary structure of both protomers. Therefore, one protomer of the dimer model fits the scattering data of the monomer. As a restraint for rigid-body modelling by SASREF the sequential covalent linkage of the Nsps in the polyprotein was used, i.e. the distance of C-terminus to N-terminus linkage was set to 5 Å. It has to be carefully remarked that rigid high-resolution structures were used for modelling [pdb codes 2AHM (SARS-CoV Nsp7, Nsp8), 2J97 (HCoV-229E Nsp9)/2UW7 (SARS-CoV Nsp9), 2FYG (SARS-CoV Nsp10)]. Moreover, while high-resolution structures of all four Nsps from SARS-CoV are available, only the Nsp9 domain of HCoV-229E pp710 was reported at high resolution so far [101].

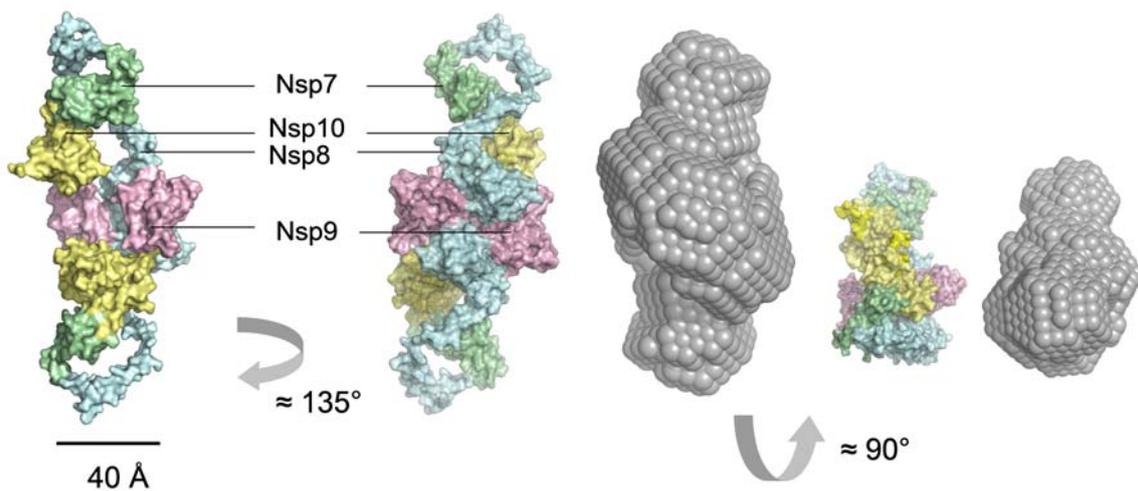


Figure 15. Rigid-body and *ab initio* model (grey spheres) of HCoV-229E pp710 dimer owing a C-terminal histag. The experimental scattering pattern recorded in 30 mM NaH_2PO_4 pH 7.9 supplemented with 300 mM NaCl and 3 mM DTT and the optimized theoretical scattering data calculated by SASREF is shown in figure 14. For rigid-body modelling, the high-resolution structures of SARS-CoV Nsp7 (*green*), SARS-CoV Nsp8 (*blue*), HCoV-229E Nsp9 (*pink*) and SARS-CoV Nsp10 (*yellow*) were used.

In the rigid-body model of dimeric HCoV-229E pp710, the Nsp8 "domain" of both protomers is located at the dimerization interface, whereas the C-terminal thickened globular domain of the Nsp8 domains are oriented towards each other. Nsp9 is located close to the interface in the dimer model as well, although the model suggests either a minor or no contribution to the binding affinity due to sterically impaired dimerization. The verification or disproof of an Nsp9-Nsp9 interface within the pp710 dimer is of specific interest, since mature SARS-CoV and HCoV-229E Nsp9 is known to consistently form a tight dimer via a GXXXG motif [101; 102]. Nsp10 is predicted to be "loosely" attached to the quaternary structure with either a minor or no contribution to inter-domain interactions at all. Since rigid structures were used for modelling, some additional flexibility for Nsp8 has to be

3. Results and discussion

considered, because different isoforms are already known [98]. Moreover, 42 residues of chain E and 54 residues of chain F at the N-terminus of Nsp8 that are not defined by the high-resolution structure of the bended Nsp8 chains within the Nsp7-Nsp8-hexadecamer used for modelling have to be taken into account. Consequently, Nsp7 is assumed to be able to be "bended down" to occupy a cavity in the rigid HCoV-229E pp710 model close to the C-terminal domain of Nsp8. This unblocks a suitable volume for the additional N-terminal residues of Nsp8 that are missing in the high-resolution structure. This bending of the N-terminus would additionally enable Nsp7 to occupy the same binding site reported for mature SARS-CoV Nsp7 and Nsp8 within the Nsp(7+8) high-resolution structure mentioned before [98]. Although the SARS-CoV pp710 *ab initio* model is more elongated than that of HCoV-229E, the bending is similarly conceivable. The difference of the models is actually supposed to originate from the structure of the Nsp8 N-terminus. It has to be considered that rigid high-resolution structures of mature Nsps incompletely reflect the potentially different tertiary structure within a polyprotein precursor due to the spatial restraint of the covalent linkage. However, according to the structure of the other Nsp8 chains in the high resolution structure of SARS-CoV, the Nsp8 N-terminus can at least adopt either a straight helical or a partly flexible conformation. More details are further discussed in this study in the context of the investigation of polyprotein Nsp7-8 (pp78, chapter 3.11).

As already indicated by SEC, SARS-CoV pp710 behaves similar in terms of oligomerization compared to HCoV-229E pp710. The obtained scattering data of SARS-CoV pp710 was also used as a template for *ab initio* modelling by DAMMIF/DAMAVAR (fig. 14) and rigid-body modelling by SASREF (fig. 15), respectively.

The *ab initio* model of the SARS-CoV pp710 dimer is similarly elongated with a slightly higher maximum diameter, reflecting the small difference in the hydrodynamic radius of both homologue polyproteins determined by DLS. A comparative rigid-body modelling approach was performed by GASBOR that is capable of indicating cavities and internal voids in a molecule by using "dummy C α -atoms", thus this approach might be favourable for multi-domain-proteins. However, it resulted in similar models compared to DAMMIF models (not shown). Fig. 16 displays the result of the rigid-body modelling, suggesting a very similar orientation of the four Nsp domains compared to HCoV-229E pp710, particularly for the dimer interface. Full-length SARS-CoV Nsp8 with stretched N-terminus well interpreted the slightly different shape of the "elongated termini" of the SARS-CoV pp710 *ab initio* model. However, some flexibility of Nsp8 to transfer covalently linked Nsp7 into an alternative position than calculated in the rigid model, close to the globular C-terminus of Nsp8 is suggested for SARS-CoV pp710 as well. The hypothesis of an Nsp7-Nsp8 interaction within pp710 is based on the well-characterized tight interaction of Nsp7 and Nsp8 in the mature state, after the polyprotein was processed by a viral protease [96; 98]. Furthermore, it is not excluded that not all

Nsp7 domains in their respective pp710 molecule are interacting with one putative Nsp8 interface, which could result in minor shape heterogeneity of Nsp8 within pp710.

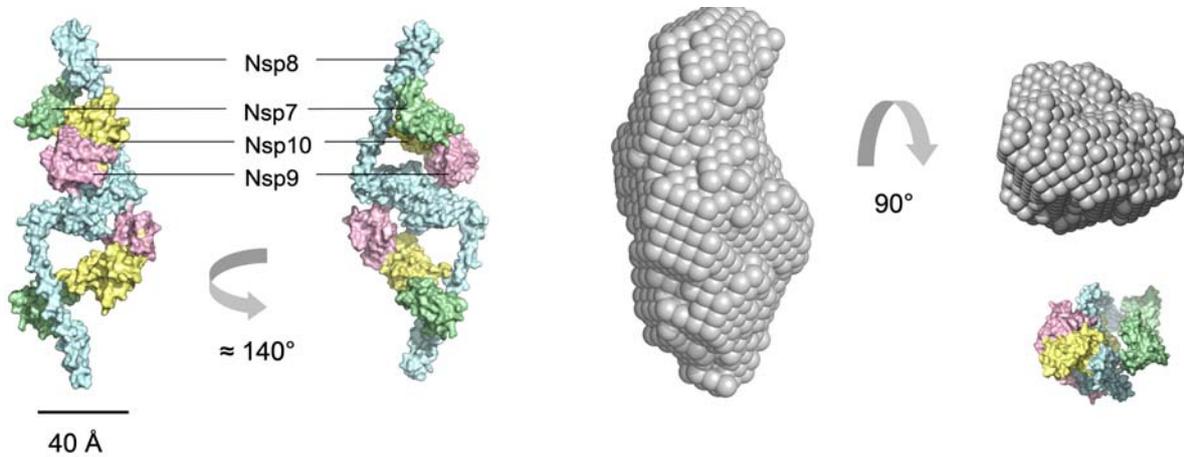


Figure 16. Rigid-body and *ab initio* model (grey spheres) of SARS-CoV pp710 dimer (N-terminal strep-tag). The SAXS pattern was recorded in 30 mM NaH₂PO₄, pH 7.9, supplemented with 280 mM NaCl and 3 mM DTT investigated by SAXS. Referring to the scattering data in figure 14, rigid-body model and *ab initio* model (grey spheres) are depicted in different orientations. For rigid-body modelling, the high-resolution structures of SARS-CoV Nsp7 (green), SARS-CoV Nsp8 (blue), SARS-CoV Nsp9 (pink) and SARS-CoV Nsp10 (yellow) were used.

3.2.3.3. Cross-linking and mass spectrometry for tertiary structure mapping of pp710

To investigate the oligomerization and particularly inter-domain and inter-molecular binding-epitopes of pp710 in solution in more detail, covalent cross-linking was combined with mass spectrometry. This approach allows the identification of surface areas in close contact at a low concentration under almost physiological conditions. This data is required to further verify the previously proposed models of pp710 based on SAXS measurements. Interaction sites are covalently linked by a reagent with typically two reactive groups targeting a specific functional group on the protein surface. After this reaction, tryptic peptides of the products are analyzed by mass spectrometry to detect additional masses arising from the cross-linking reaction. Beside the reactive group the reagents significantly differ in size that is critical for the distance of protein side chains that can be linked. EDC (1-ethyl-3-[3-dimethylaminopropyl]carbodiimide) and BS2G (bis[sulfosuccinimidyl] glutarate) were selected in this study. EDC is a zero-length cross-linker, therefore, a successful link of two amino acids requires a distance close to 0 Å, while BS2G contains a “spacer arm” length of 7.7 Å (see table 12). Cross-linking with BS2G was performed with an equimolar mixture of non-deuterated and four times deuterated BS2G (BS2G-d₄) to eventually verify the results by detecting reaction products containing both isotopes with the expected mass difference. For this cross-linking approach,

3. Results and discussion

SARS-CoV pp710 was chosen, since the individual domains were characterized in more detail so far, compared to HCoV-229E pp710. Moreover, the relatively high sequence identity and the similar low-resolution models (see section 3.2.3.2) suggest comparable domain interactions within both pp710 homologues.

The EDC reaction products of SARS-CoV pp710 were separated by SDS-PAGE (not shown). Clearly detected monomeric, dimeric, and a larger oligomeric state, were digested by trypsin. In addition, cross-linked pp710 was incubated with SARS-CoV main protease to remove the Nsp10 domain. The pp79 cleavage product was additionally cut from the gel and digested by trypsin, enabling the analysis of the Nsp7, Nsp8 and Nsp9 domains interactions with a minimum number of peptides. All peptides were applied to ESI-MS/MS analysis. The obtained MS data was analyzed for cross-linking reaction products that are absent in a control sample without EDC using the XQuest online server [161]. Identified putative links are listed in table 16.

Table 16. List of putative peptides of SARS-CoV pp710 cross-linked by EDC as analyzed by XQuest [161]

number	sample	peptide1	peptide 2	m/z
1	monomer	DAAMQRK (Nsp8)	LKKSLNVAK (Nsp8)	901.03
2		GFCDLKGK (Nsp10)	SEDKRAK (Nsp8)	841.44
3		GKYVQIPTTXANDPVGFTLR (Nsp10)	KLKK (Nsp8)	912.19
4	dimer	AYKDYLASGGQPITNXVK (Nsp10)	GPKVK (Nsp9)	832.10
5		AKVTSAMQTMLFTMLR (Nsp8)	EPLMQ (Nsp10)	809.76
6		LCEEMLDNR (Nsp7)	LKKSLNVAK (Nsp8)	1052.58
7	large oligomer	FVTDTPKGPK (Nsp9)	SDGTGTIYTELEPPCR (Nsp9)	703.09
8		MSDVKCTSVVLLSVLQQLR (Nsp7)*	SEDKR (Nsp8)	912.17
9		GFCDLKGK (Nsp10)	SEDKRAK (Nsp8)	841.44
10	pp79 (digested)	GPKVK (Nsp9)	MVLLSVLLSMQGAVDINR (Nsp7)	852.47
11	pp710)	SEFDR (Nsp8)	VKYLYFIKGLNNLNR (Nsp9)	830.44

* lysine in this peptide of Nsp7 was identified as cross-linking site linked to the neighboring tryptic peptide of SEDVK of Nsp8 using BS2G for cross-linking.

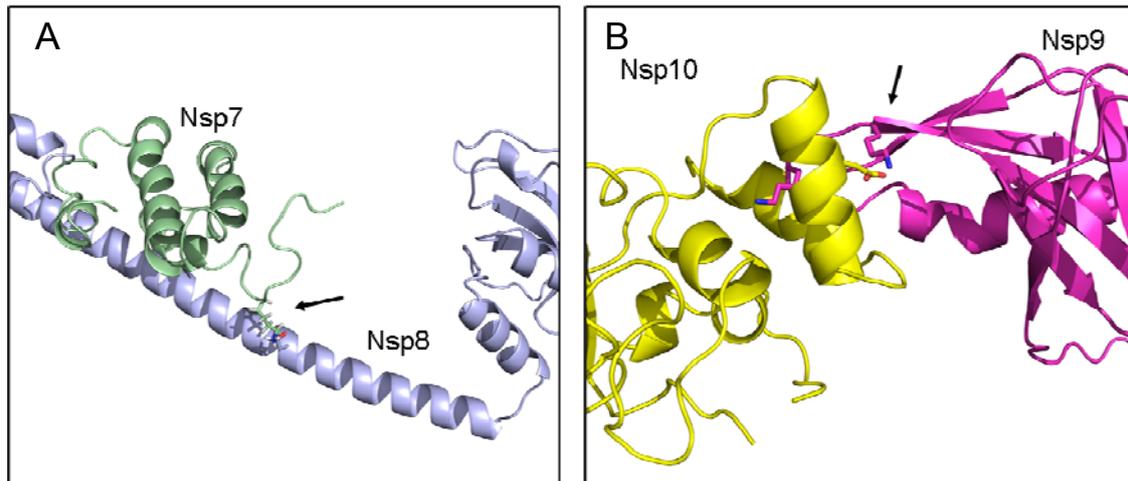


Figure 17. Two putative cross-linking sites within SARS-CoV pp710 (cross-link number 8 and 4, according to tab. 16; the acidic amino acid and the lysine of the linked peptides are highlighted as sticks). **(A)** Nsp7 (*green*) and Nsp8 (*blue*, pdb code 2AHM chain H; Nsp7 was manually positioned considering the covalent connection). This Nsp7-Nsp8 cross-link can reasonably be related to an epitope revealed at high-resolution for mature Nsp7 that binds close to the C-terminus of Nsp8 (pdb code 2AHM), considering that the putative flexible N-terminus of Nsp7 is oriented towards the stretched α -helical N-terminus of Nsp8 (based on SAXS data discussed in chapter 3.11). **(B)** SARS-CoV Nsp9 (*magenta*; pdb code 1UW7) and Nsp10 (*yellow*; pdb code 2FYG).

Similarly, N-terminally strep-tagged SARS-CoV pp710 was cross-linked by BS2G (with and without deuterium “label”) as displayed in fig. 18. SARS-CoV polyprotein Nsp7-8 (pp78; see section 3.9) was included and cross-linked as well to compare its oligomerization to SARS-CoV pp710. While non-linked “monomeric” pp78 visualized by coomassie-stained SDS-PAGE almost completely disappeared during the reaction in the presence of 25 mM BS2, pp710 formed a significantly lower fraction of covalent oligomers in this reaction. Based on this observation, a different oligomerization of pp78 and pp710 is hypothesized, i.e. different dissociation constants. However, the presence of the “domains” Nsp9 and Nsp10 may influence the formation of a comparably small fraction of dimer and a larger fraction of putative trimer of pp710 compared to pp78. The indicated specific trimeric state – as at least a small amount is detectable by mass spectrometry – could be mediated by trimerization of Nsp10, as observed within a controversially discussed dodecameric high-resolution structure ([131; 118; 181]). Nevertheless, this is highly speculative since a similar oligomeric state of Nsp10 has never been observed in solution and Nsp7, Nsp8 and Nsp9 could contribute as well to form higher oligomeric states, including intermolecular disulfide bonds.

3. Results and discussion

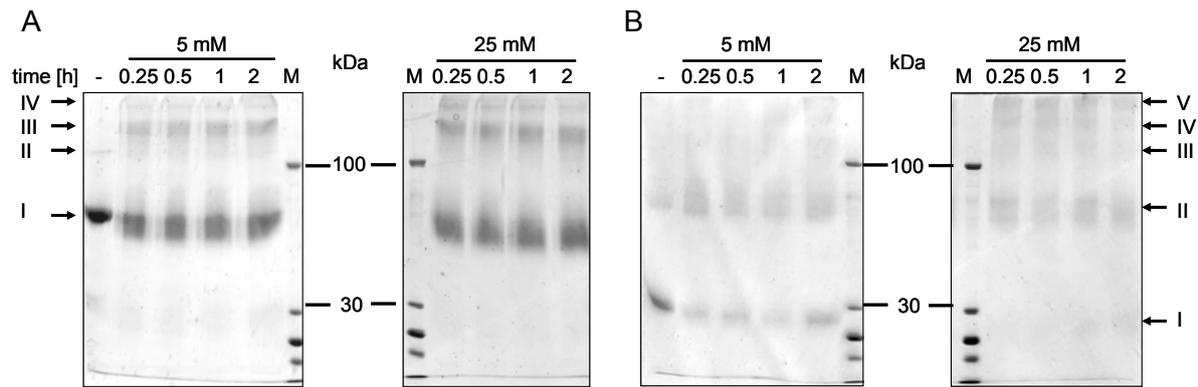


Figure 18. Two different total concentrations of an equimolar mixture of **BS2G-d₀/d₄** were incubated with 5 μ M SARS-CoV pp710 (**A**) and 5 μ M SARS-CoV pp78 (**B**) for different time periods up to 2 h. Both proteins harbor an N-terminal strep tag. Reaction products were visualized by SDS-PAGE and coomassie staining. Oligomeric states are continuously labeled with numbers I-V, where I is the monomer.

All oligomeric states identified by SDS-PAGE were cut from the gel and applied to tryptic in-gel digestion. Trypsin cleaves N-terminally of lysine and arginine residues. The tryptic peptides were separated according to their mass to charge ratio by an ESI-Q-TOF instrument with significantly advanced separation capacity. The XQuest/XProphet software package [161; 162] was applied to search MS/MS spectra of suitable masses for fragment ions of peptides cross-linked by BS2G-d₀ and/or deuterated BS2G-d₄. Therefore, the identification of cross-linking sites is significantly improved compared to the EDC-based cross-linking investigated by an ESI-ion trap instrument (LC/MSD Trap XCT Ultra II). Oxidation and carbamido methyl modifications of amino acid side chains were taken into consideration for all spectra. MS/MS spectra of putative BS2G cross-link events are summarized in the appendix. The basic method including the beneficial use of isotope labels for verification of the results has proven to be powerful in modelling particularly of large multi-domain proteins [182; 183], optionally combined with homology modelling and different “low-resolution” methods summarized by Venselaar *et al.* [184] and native mass spectrometry [185].

Initially, to establish the method two putative intra-domain cross-linking sites within SARS-CoV pp710 indicated by XQuest/XProphet are described (fig. 19). Subsequently, 5 additional putative interdomain cross-linking sites identified in the sample of monomeric pp710 and 1 intra- or interdomain cross-linking site of Nsp8 in the sample of presumably trimeric pp710 (oligomeric state III in fig. 18A), as cut from an SDS-PAGE, are presented in the following paragraphs. These cross-links were selected according to the highest fragment ion coverage in the MS/MS spectra. Spectra of all putative BS2G cross-linking sites are displayed in the appendix in figure 87A and figure 87B. These figures include two additional cross-linking sites not presented in the main text in the sample of larger oligomers (oligomeric state IV in fig. 18A), which are hardly interpretable, since different perhaps unspecific oligomerization epitopes are expected in such oligomers that are not related to the

SAXS models presented in chapter 3.2.3.2. A control sample of SARS-CoV pp710, free of BS2G was included as a negative control.

At first, an intramolecular covalent link of K84 and K86 of SARS-CoV Nsp9 was identified (fig. 19, upper panel). Both isotopes of the cross-linking product were identified by high fragment-ion coverage of the sequence. Additionally, this result is strongly supported by the close proximity of these residues in the corresponding high-resolution structure of mature Nsp9 (fig. 19, upper panel). A putative intramolecular link of K72 and K79 of Nsp8 by BS2G-d₀ can be explained by the predominantly helical structure of the N-terminal half of mature Nsp8, which enables the residues to localize in close contact – remarkably on the same side of the helix. Nevertheless, loop segments within the Nsp8 N-terminus might have a flexible random structure that allows the helical N-terminus to adopt a “bended” conformation as revealed in Nsp8 high-resolution structures [96; 98]. The alternative conformation of Nsp8 in the same high-resolution structures is a prolonged stretched α -helical N-terminus as observed for SARS-CoV [98] and feline CoV (FCoV) [96]. According to the geometry of the BS2G molecule and the orientation of K72 and K79 in the “stretched” conformation of the α -helix (see fig. 19, *light blue*), a cross-link might be sterically impaired. However, the cross-linking reaction between K72 and K79 could have been allowed by the bended conformation, similar to the Nsp8 structure depicted in cyan (fig. 19), even though the number of detected fragment ions is rather low compared to the intrapeptide Nsp9 cross-link described before. However, both examples further indicate that the method in principle is sufficient to obtain structural information.

3. Results and discussion

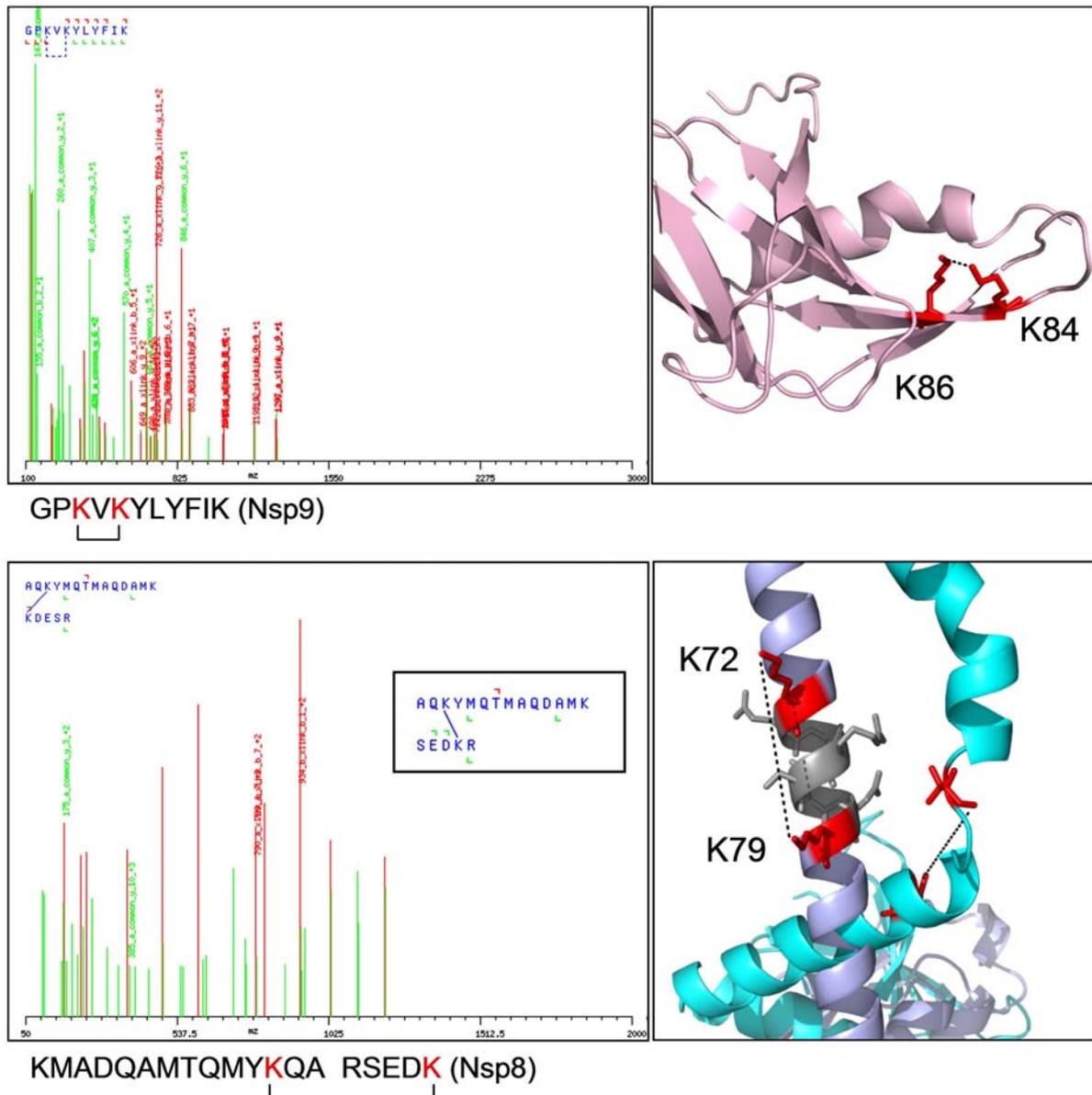


Figure 19. Intra-domain cross-linking sites of SARS-CoV pp710. The left panel displays the MS/MS spectra. Fragment ions that were identified in a respective MS/MS spectrum with the default mass tolerance are labelled either *green* (common mass) or *red* (mass containing the cross-linker). The lysine residues labelled *red* below the spectra were linked by BS2G. The covalent link of K84 and K86 (upper panel) was identified by high fragmentation coverage of the sequence; both isotopes BS2G- d_0 and - d_4 of the peptide were identified, the BS2G d_0 -link is shown (precursor mass: 726.4146 Da, charge 2+, isotope mass difference: 4.0237 Da). This link is strongly supported by the high-resolution structure of SARS-CoV Nsp9 [102] (upper right panel) and at least further reveals that the method is capable of identifying surface areas in close proximity. Similarly, the link between K72 and K79 within Nsp8 by BS2G- d_0 in the monomer sample can be well explained. Assuming a prolonged stretched α -helical N-terminus of Nsp8 the flexible lysine residues are facing in the same direction (SARS-CoV Nsp8, pdb code 2AHM, chain H in *light blue*), separated by two helix windings. The “bended” N-terminal conformation of Nsp8 suggests a more probable close position of both lysines separated by a short turn motif (FCoV Nsp8, pdb code 3UB0, chain A in *cyan*). Peaks fitting the putative cross-link site are labelled. A similar cross-linked peptide of Nsp8 created by alternative trypsin cleavage was identified by XQuest/XProphet and is depicted in the enframed box.

Further cross-linking sites suggested by mass spectrometry indicate the close proximity of the elongated helical subdomain of Nsp8 with that of Nsp9, i.e. Nsp8 K46 linked with Nsp9 K77 and Nsp8 K39 connected to Nsp9 K89. Due to the structure of BS2G, a distance of around 8 Å between those sites is expected, whereas EDC presented earlier requires a much lower distance. Assuming that one suggested link could be “false positive” in the spectra, the other Nsp8-Nsp9-link is a valuable validation. An additional cross-verification is presented by the detected peptide SEFDR (Nsp8) that is assumed to be linked by EDC with VKYLYFIKGLNNLNR (Nsp9). In parallel, a BS2G-mediated link was identified for KLKSLNVAKSEFDRDAAMQ and YLYFIKGLNNLNR, notably both peptides overlap with the previously mentioned EDC linked Nsp8 and Nsp9 peptides. Consequently, according to the rigid-body model of both homologues of pp710, Nsp9 seemed to be further translated towards the cross-linking lysine residues of Nsp8 (*red* arrow, fig. 20), as it is well visualized for HCoV-229E pp710 as well. Therefore, Nsp9 might even displace Nsp10 that is considered to be “loosely” attached to Nsp9, particularly as no specific interaction of mature Nsp10 with the other three mature Nsps is indicated, as investigated and discussed in chapter 3.4.1. A translation of Nsp9 is furthermore supported by the shape of the *ab initio* model, providing sufficient space for Nsp9, and the N-terminus would still be in sufficiently close proximity to the C-terminus of Nsp8 (fig. 20, compare left and right panel). In parallel, the N-terminal helical domain of Nsp8 is assumed to bend in the opposite direction (*green* arrow, fig 20). For comparison to the rigid-body model of HCoV-229E pp710 the *ab initio* models of SARS-CoV pp710 and HCoV-229E pp710 are shown (fig. 20, left panel). Nsp8-Nsp9 cross-link events together with three cross-links of Nsp8 with Nsp7 or the strep tag fused to the N-terminus of Nsp7 are summarized in figure 20, respectively. All five cross-linking sites were identified in the sample of monomeric pp710. Consequently, these links have to be intramolecular.

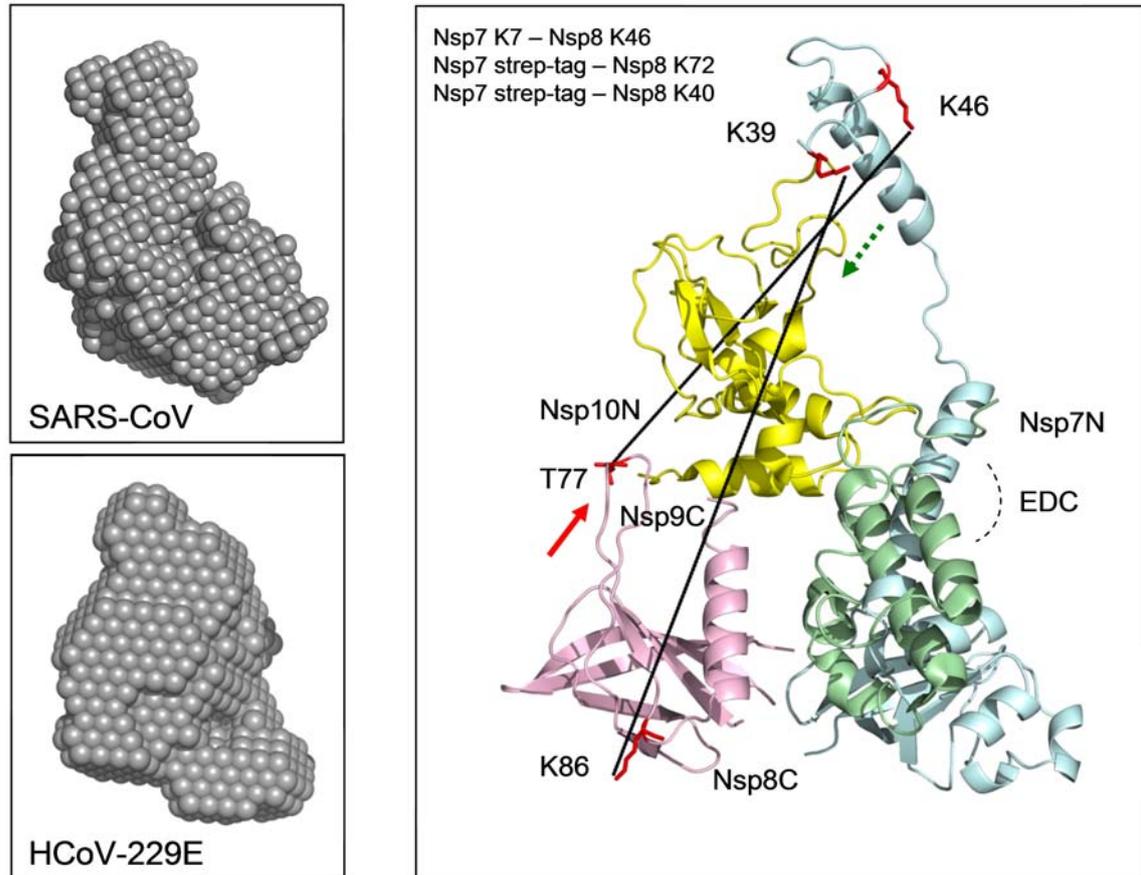


Figure 20. Based on the high sequence identity of 47.1% and similar *ab initio* models shown in the left panel, the structure of SARS-CoV and HCoV-229E pp710 is interpreted taking SAXS and mass spectrometry data into consideration. Cross-linking sites of SARS-CoV pp710 connecting Nsp8 (blue) and Nsp9 (pink) are visualized in the rigid-body model of monomeric HCoV-229E pp710 (right panel, solid black lines, involved residues are coloured red). The arrows indicate the suggested direction of subdomain translation based on the Nsp8-Nsp9 cross-link sites identified by MS. Some termini of Nsp domains are labelled for clarity. Three cross-linking sites involving the N-terminus or the N-terminal strep-tag of Nsp7 (Nsp7N; Nsp7 is coloured green) and the elongated helical domain of Nsp8 are listed above the rigid-body model and are visualized in fig. 21. The dashed line represents an EDC cross-linking site connecting Nsp7 and Nsp8, as previously depicted in figure 17A, in good agreement with the putative BS2G cross-linking sites. Nsp10 is coloured yellow.

The mass spectrometry-based cross-linking analysis of pp710 indicated that the highly elongated N-terminus of Nsp8 represents an uncertain aspect within the rigid-body models of both pp710 homologues. According to the *ab initio* models of pp710 (fig. 20, left panel), the structure of the N-terminus of Nsp8 appears to be slightly different, comparing both coronavirus species. As previously mentioned, a high-resolution structure contained one isoform of Nsp8 with stretched N-terminal α -helical domain and another one with a specific degree of bending of the N-terminus, allowed by short flexible loop structures in between of helical structure [96; 98]. Crystal contacts identified by CryCo 5.1 analysis [186] are suggested to stabilize the elongated N-terminus in the crystal. Considering this

putative flexibility, the identified Nsp8-Nsp8-link in the pp710 dimer sample (KLKKSLNVAKSEFD linked to KLDNDALNNIINNAR) would be well explained by a relative close proximity of both Nsp8 termini, due to tight bending. Nsp7 is expected to bind to Nsp8 close to the Nsp8 C-terminus, which indeed results in a close proximity of both Nsp8 termini within pp710. Consequently, the Nsp8-Nsp8-link is suggested to be an intra-molecular link, rather than associated with the dimer interface involving two different Nsp8 molecules. Considering that the Nsp8 N-terminus is at least partly bending towards Nsp9 and Nsp7 (fig. 20, right panel, rigid-body model), Nsp8 could most likely not reach the elongated termini of both pp710 *ab initio* models (fig. 20), and therefore not span the entire molecule according to the maximum diameter determined by SAXS. Thus, mass spectrometry data questions the position of Nsp10 relative to Nsp8, since a minor translation of Nsp10 would allow the required bending of the Nsp8 N-terminus and the maximum diameter of the *ab initio* model would be increased by Nsp10 at the same time, corresponding to the *ab initio* model (fig. 21). To enable the hypothesized “bending” of the Nsp8 N-terminus, SARS-CoV Nsp8 (chain H, extracted from pdb code 2AHM) was cut into 5 fragments (amino acids 1-34 (fragment 1), 35-73 (fragment 2), 74-113 (fragment 3), 114-124 (fragment 4), 125-195 (fragment 5, C-terminus)) to design the final concluding structural model (fig. 21). These Nsp8 fragments were designed according to reported putative sites of flexibility within the molecule [96; 98], and personal communication with *Y. Xiao* (University of Lübeck). Additionally, a loop comprising residues 99-101 could be flexible as well to potentially increase maximum diameter of the proposed model, which is slightly lower than determined by SAXS.

As observed for the complex of mature Nsp7 and Nsp8, Nsp7 is expected to be located close to the globular Nsp8 C-terminus, involving hydrophobic residues in the N-terminal helix 1 and helix 3 of Nsp7 [98; 96]. The cross-linking sites connecting Nsp8 and Nsp7 or the strep tag N-terminally linked to Nsp7 are in good agreement with this model: Nsp8 K46 linked to Nsp7 K7 and a lysine within the strep-tag linked to K72 and K40 of Nsp8, respectively. The two different cross-links of the strep-tag could be explained by its flexibility.

There is a valuable agreement comparing the results obtained by either cross-linking reagent with each other. For instance, using EDC XQuest suggested a link connecting MSDVKCTSVVLLSVLQQLR (Nsp7) and SEDKR (Nsp8). A cross-link suggested by BS2G is KMSDVKCTSVVLLSVLQQL linked to MADQAMTQMYKQAR which is located next to the N-terminus of SEDKR in Nsp8, further securizing the orientation of the Nsp7 N-terminus towards a defined fragment of the corresponding Nsp8 helix. This is in good agreement with the interaction of the N-terminal helix 1 of Nsp7 with a central part of the Nsp8 chain in the complex of the two mature Nsp8s, allowing to suggest a very similar orientation of Nsp7 helix 1 within pp710. Therefore, a conserved orientation of helix 3 is hypothesized as well, which also contributes to the interaction of mature Nsp7 and Nsp8 [98]. Finally, a summarizing rigid-body model of SARS-CoV pp710 involving

3. Results and discussion

known high-resolution structures of SARS-CoV and five individual fragments of Nsp8 considered all putative BS2G cross-linking events of Nsp7 with Nsp8 (*orange* spheres in fig. 21) as well as of Nsp8 with Nsp9 (*red* spheres) of SARS-CoV, as displayed in Fig. 21. The residues that were linked according to mass spectrometry are located in close proximity, mainly by the rearrangement of the structure of Nsp8. All four Nsp domains of SARS-CoV pp710 are presumed to have flexible termini according to the high-resolution structures. The flexibility allows an adjustment of the position of the cleavage sites within pp710 to guarantee that the main protease cleavage site to process pp710 is well surface-exposed, as already suggested by the model in fig. 21.

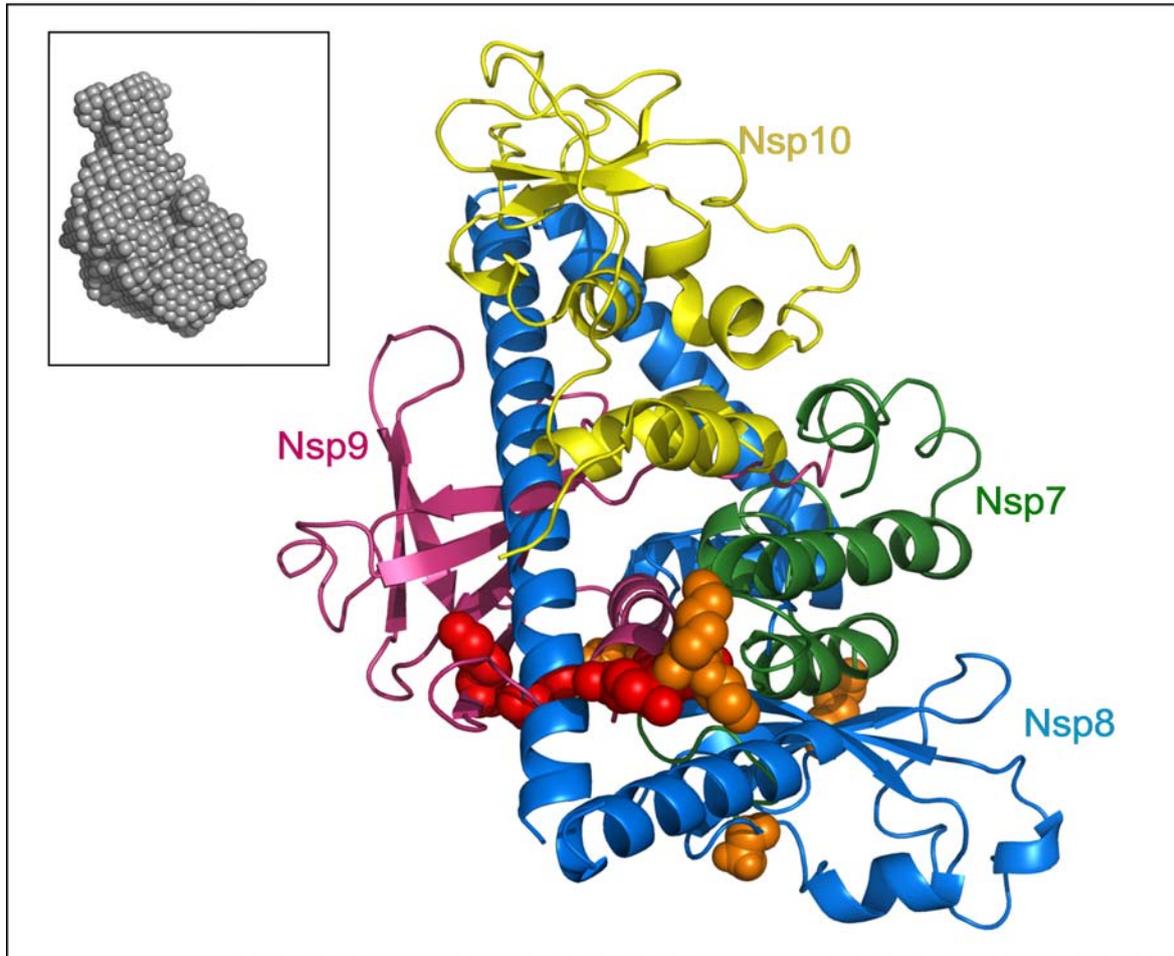


Figure 21. Optimized structural model of monomeric SARS-CoV pp710, taking the identified cross-linking sites of SARS-CoV pp710 into consideration. Lysine residues involved in two cross-links connecting Nsp9 and Nsp8 are labelled by *red* spheres (Nsp8 K46 – Nsp9 K77; Nsp8 K39 – Nsp9 K86), three sites linking Nsp7 and Nsp8 (see fig. 20) are labelled in *orange* (Nsp7 K7 – Nsp8 K46; Nsp7-tag – Nsp8 K72; Nsp7-tag – Nsp8 K40). Nsp7, Nsp9, Nsp10 and five fragments of SARS-CoV Nsp8 (comprising chain H, extracted from pdb code 2AHM) were used for modelling. The fragments of Nsp8 individually used to model the bended conformation of the full-length Nsp8 domain within pp710 are: amino acids 1-34 (fragment 1), 35-73 (fragment 2) 74-113 (fragment 3), 114-124 (fragment 4), 125-195 (fragment 5, C-terminus). It may additionally be assumed that a loop comprising residues 99-101 is flexible as well to potentially increase the maximum diameter of the model, which is slightly lower than determined by SAXS. Fragments of Nsp8 were manually positioned according to the *ab initio* model. The *ab initio* model of SARS-CoV is shown again for comparison. All covalent Nsp-Nsp contacts within pp710 are sufficiently surface-exposed to allow cleavage by the coronavirus main protease.

This concluding model of SARS-CoV pp710 that is drawn here allows a valuable prediction concerning the overall structure of pp710 with slight differences between pp710 of an *Alphacoronavirus* compared to a *Betacoronavirus* (fig. 20). It also highlights the hypothetical previously unknown interface of mature Nsp8 and Nsp9, although there are contradictions in the literature concerning the strength of this interaction [101; 102]. Moreover, it is strongly suggested that Nsp7 in the polyprotein state is already capable of occupying a binding site with Nsp8 close to its C-terminus, as reported for the complex of mature Nsp7 and Nsp8 [96; 98]. Finally it was observed that pp710 of SARS-CoV has a significantly higher radius of gyration and a higher maximum diameter than the HCoV-229E homologue. However, except these differences, the low-resolution surface structures of HCoV-229E and SARS-CoV clearly share significant similarities (fig. 20).

3.2.4. Trials to crystallize pp710

Knowing that pp710 displayed variable oligomerization in a slow equilibrium and that it further represents a multi-domain protein that is assumed to be locally flexible, it was not considered to be easy to crystallize the full-length pp710 proteins. Both homologues were applied to a conventional robot-assisted crystallization condition screening. Applying a solution with predominantly monomeric or dimeric pp710 of either SARS-CoV or HCoV-229E to approximately 2000 different crystallization conditions each resulted in a few promising observations (like phase separation) that were worth to be refined. However, refinement failed to result in crystals suitable for a diffraction analysis up to now. One exemplary crystallization condition is displayed in figure 22. Refinement approaches included microseeding of very tiny crystals and “crystalline precipitation” as well as counter diffusion experiments. Moreover, smaller fragments of pp710, e.g. pp79 (polyprotein Nsp7-9) and pp78 (polyprotein Nsp7-8) (described in chapter 3.9) as well as and some of the mutated variants of pp710 (described in chapter 3.5) were included in crystallization trials. Limited proteolysis using a serial dilution of trypsin, chymotrypsin, papain and subtilisin did not reveal a stable domain of more than approximately 10-12 kDa in size and was therefore not considered to be beneficial for crystallization of a rigid pp710 fragment. Similarly, covalent cross-linking trials did not result in homogeneous solutions that are favourable for crystallization trials.

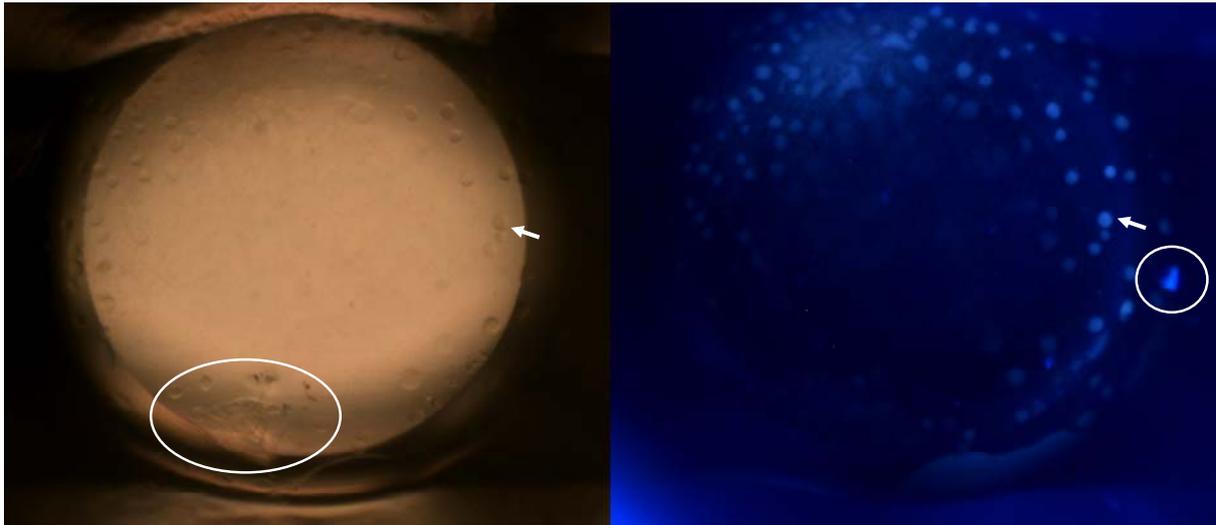


Figure 22. Crystallization of HCoV-229E pp710. A 300 nl drop of 40 μM HCoV-229E pp710 in Tris buffer containing 260 mM NaCl was mixed with an equal amount of 0.3 M magnesium formate, 0.1 M Bis-Tris pH 5.5 and sealed in a 96-well Nextal plate (Qiagen). Crystallization success was investigated and documented after three weeks under visible light (left image) and excited by UV-light (right image). The drop contained monomeric protein purified by SEC supplemented with 2 mM DTT. For orientation, the white arrow points at the same spheroid “drop” in both images. Thin two-dimensional crystals are encircled in the left panel. On the right image, a putative crystal with dimensions in the low μm range that is fluorescent is highlighted.

It was recently observed and reported by members of our research group that heterogeneous gene overexpression in baculovirus-infected *Sf9* insect cells can yield small crystals ($10 \mu\text{m}^3$ average volume) grown *in vivo* within the living cells. As a first proof-of-principle, glycosylated cathepsin B of the parasite *T. brucei* crystallized within the late ER was isolated and successfully applied to the novel technique of serial femtosecond X-ray crystallography at a free-electron laser to solve the protein structure [122]. Based on this promising approach, *in vivo* crystallization of SARS-CoV pp710 was investigated following the established protocol, representing the very first trial for directed crystallization of a polyprotein within insect cells. The corresponding full-length gene was integrated by double homologous recombination into a bacmid that is controlled by a polyhedron promoter to express the pp710 gene and release the protein in the cytosol of *Sf9* insect cells. The insect cells showed the typical morphology of infected cells using a light microscope, crystalline structures within the cells incubated at either 27 °C or room temperature were not observed. Although various different approaches to obtain crystals of pp710 were applied, high resolution structural data could not be obtained within this study.

3.2.5. Insights into the function of pp710

3.2.5.1. Native gel electrophoresis: Interaction of pp710 with nucleic acid

Native PAGE is useful to initially detect a protein's ability to interact with nucleic acid. However, this method does not allow a reliable K_D -value calculation. Particularly for very weak protein-nucleic acid interactions an electrophoretic mobility shift is hardly quantitative. Nevertheless, an interaction of HCoV-229E pp710 with random ssDNA (fig. 23A), random dsDNA (fig. 23B, upper panel) as well as tRNA (fig. 23 B, lower panel) was qualitatively verified for the full-length pp710 of HCoV-229E in a concentration-dependent manner. Even though RNA is the physiologically relevant ligand of coronaviral Nsps, those three different types of nucleic acid provide structural diversity and the opportunity to compare the results with previous studies about the interaction of Nsps with different kinds of nucleic acid. In the presence of protein, a putative protein-nucleic acid complex that is reduced in electrophoretic mobility became visible (fig. 23A) and more abundant (fig. 23B) with increasing protein concentrations. Since the stain used for the detection in the assay is specific for nucleic acid, the electrophoretic mobility shift is due to the interaction of the different nucleic acids with pp710.

These results indicated that pp710, the precursor of four mature Nsps, is already capable of binding nucleic acid, supporting that pp710 could have an impact on the replication of the virus. Interactions of mature Nsp domains with nucleic acid in the μM range were previously observed *in vitro*, for instance for mature SARS-CoV Nsp10 with dsDNA, dsRNA and ssRNA [111], Nsp9 with ssDNA and ssRNA [101; 187] and HCoV-229E Nsp8 with tRNA and other RNA fragments [103]. In particular, an interaction of processed MHV Nsp10 with tRNA was demonstrated by using an electrophoresis technique with horizontal agarose gels, called zone-interference gel electrophoresis [170] and revealed a low μM dissociation constant.

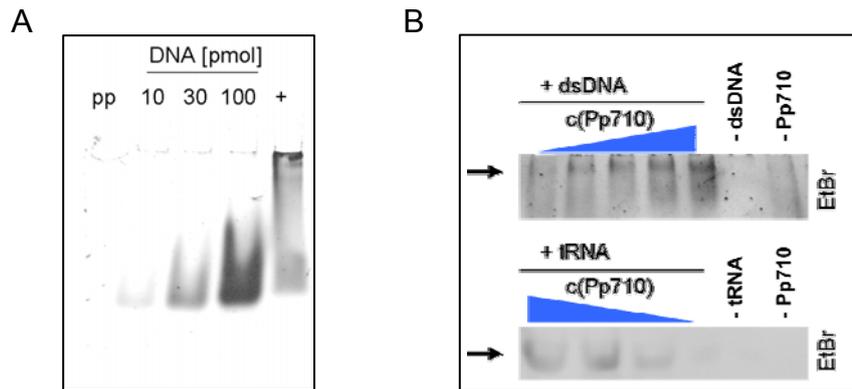


Figure 23. Native PAGE under non-reducing conditions demonstrating an interaction of HCoV-229E pp710 with a random ssDNA, dsDNA and yeast tRNA (Roche). Pp710 was initially treated with benzonase and applied to the native PAGE in sodium phosphate buffer at pH 7.9 supplemented with 100 mM NaCl and 2 mM MgCl₂. (A) Interaction of pp710 with ssDNA. “pp”, 300 pmol polyprotein; “+”, 300 pmol HCoV-229E pp710 mixed with 50 pmol ssDNA (5'-GCGAATTCCTAAACAGCAAAAGAACAATGTGT-3'). Other lanes contain 10 to 100 pmol pure ssDNA. (B) Interaction of pp710 with 60 pmol dsDNA (5'-GCATTTGTCGACCAT-3') and 20 pmol tRNA. Double stranded DNA was heated to 80 °C for three minutes and cooled down at room temperature. The amount of pp710 ranges between 60 and 300 pmol in both gels. A nucleic acid-free and a protein-free control was included as well (“-dsDNA/-tRNA” and “-Pp710”). To detect dsDNA, gels were stained with EtBr. Detection of ssDNA and tRNA was accomplished by incubating the gels with SybrGOLD diluted in electrophoresis buffer.

3.2.5.2. Pp710 interacts with RNA derived from the 3'-UTR

Considering that SARS-CoV Nsp8 is known to specifically bind genomic RNA in an early stage of replication [108], this investigation aims to analyze the protein-RNA interaction of the precursor SARS-CoV pp710. It is assumed that a filter retention assay, usually applied to reveal protein-nucleic acid interactions in the range of μ M dissociation constants, is suitable to detect pp710-RNA interactions, obtaining further insights into a putative function of pp710.

A structurally conserved fragment of the SARS-CoV genomic 3'-untranslated region (3'-UTRdeltaHVR, 105 nt, see methods for reference) and a shuffled sequence with identical nucleotide composition (3'-UTRdeltaHVR-SH) were produced by T7 polymerase *in vitro* transcription. RNA transcription products were separated on a denaturing acrylamide gel (fig. 24A/B). These fragments represented the ligands applied to the assay. The secondary structure of the 3'-UTR fragment, lacking a hyper variable region (HVR), is shown in figure 24C, as predicted by the software *mfold* [188]. The RNA fragments were terminally labelled with ³²P-phosphate to enable a sensitive and specific detection. Various concentrations of SARS-CoV pp710 were incubated with a constant amount of both RNA fragments. The solution was drawn through a membrane, while only the protein bound to the

3. Results and discussion

membrane. Unbound nucleic acid were not retained, whereas labeled RNA molecules that interacted with the protein remained on the filter and were subsequently detected and quantified using a phosphor imager. A positive control, i.e. an RNA aptamer binding to streptavidin in the nM range, was included for comparison (fig. 24C).

In spite of a relatively high background due to unspecific binding of RNA on the membrane, a concentration-dependent interaction is strongly suggested for both fragments (fig. 24D) by plotting the detected radioactivity against the protein concentration (not shown). The detected RNA-interaction of pp710 additionally indicated that the polyprotein precursor represents a functional component of the RTC. However, due to the interaction with the shuffled sequence it appears that pp710 does not only specifically recognize the sequence and secondary structure of the putative Nsp8 binding site within the 3'-UTR of the SARS-CoV genome. This is in good agreement with the observation that even mature Nsp8 does not seem to be highly sequence-specific with a long recognition site [103], but clearly prefers 5'-(G/U)CC-3' in a template to initiate the synthesis of a short complementary fragment [94]. Furthermore, mature coronaviral Nsp9 and Nsp10 were not reported to harbour a sequence-dependency for nucleic acid interaction [111; 187] and likely also contribute to the interaction of pp710 with nucleic acid. However, an interaction of Nsp9 with hairpin SL2 of the 3'-UTR was demonstrated [104]. This interaction as well as the FCoV Nsp8 RNA polymerase activity significantly depends on the environmental pH, preferring approximately pH 7-7.5 in both interactions [96]. This is reflected by the binding buffer at pH 7.6 used in this experiment. Facing the problem that in a filter retention assay a small amount of RNA is unspecifically binding to the membrane and that Nsp9 and Nsp10 may contribute to sequence independent interaction of pp710 with nucleic acid, it was decided to continue by investigating a specific catalytic activity, i.e. polymerase activity, of pp710 and its processing products. However it is summarized, that pp710 harbours at least one accessible binding site to interact with DNA and RNA. Moreover, SARS-CoV and HCoV-229E pp710 appear to have either no or a weak overall sequence specificity in binding to nucleic acid.

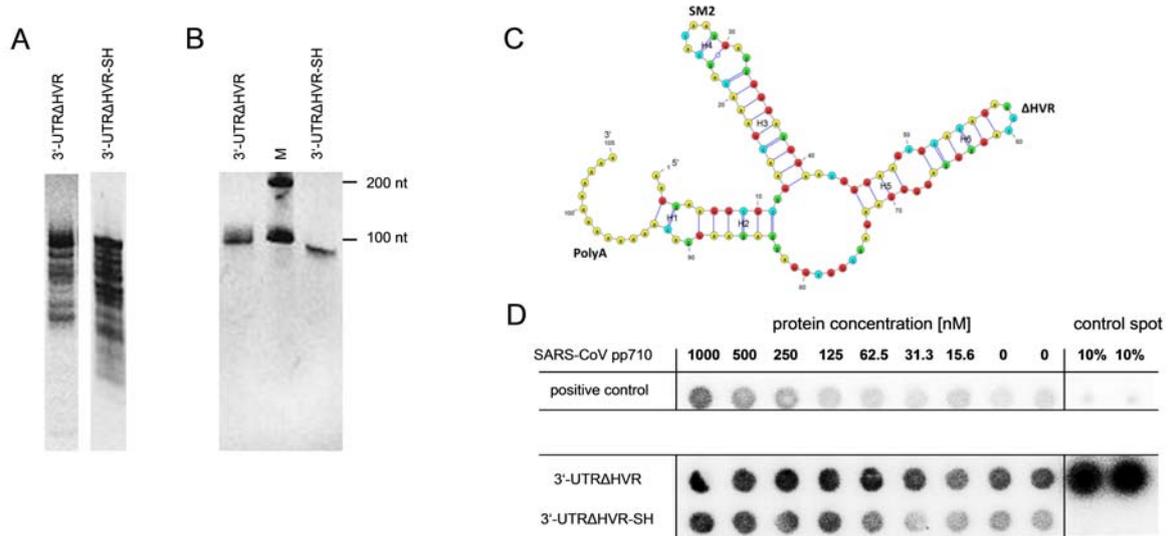


Figure 24. Interaction of SARS-CoV pp710 with RNA. (A) *In vitro* transcription products of two different DNA templates derived from the SARS-CoV 3'UTR were separated on a denaturing urea acrylamide gel. (B) The desired RNA fragments, i.e. the authentic RNA fragment and a shuffled variant, isolated by gel extraction showed high homogeneity. (C) Prediction of the secondary structure of the 3'-UTR lacking a HVR using *mfold* [188]. (D) Phosphorimager screen of the filter-retention assay involving both RNA fragments and N-terminally strep-tagged SARS-CoV pp710. 10% of the total amount of RNA per sample in the respective lane is spotted directly on the membrane for control. An RNA aptamer provided by *S. Kruspe* (University Hamburg) was applied as positive controls.

3.3. SARS-CoV M^{pro} (Nsp5): Structural properties and monomer-dimer equilibrium

During replication in a host cell, coronaviral pp710 is cleaved by the viral main protease to release mature Nsps. Pp710 itself is considered to be an intermediate product during processing of a large fragment of pp1a (Nsp4-Nsp16). To analyze the time-course of the processing pathway of pp710 in the presence of the SARS-CoV main protease (M^{pro}) and eventually purify cleavage products of pp710 for structural characterization, the protease had to be purified and initially characterized.

SARS-CoV M^{pro} was heterologously expressed as previously described [78] and purified by Ni-NTA affinity chromatography and size-exclusion chromatography (fig. 25A). The presence of the C-terminal his-tag was verified by Western blot. Isoelectric focusing (IEF) determined the pI to 6.6 (fig. 25A). The N-terminal GST-tag to optimize the protein solubility was cleaved off *in vivo* by N-terminal autocleavage of the protease, comprising a first opportunity of activity verification by analysing the protein size by SDS-PAGE. The gene of the protease mutant R298A was generated by mutagenesis PCR, the protein was purified applying the same protocol. Amino acid R298 is known to be critical for

3. Results and discussion

the dimerization of M^{pro} [56]. Thus, the mutant is interesting to analyze the monomer-dimer equilibrium and protein interactions with the main protease in more detail. The mutant is still capable of GST-tag autocleavage in *E. coli* cells, even though the dimeric main protease is known to be the catalytically active state [48; 49]. In agreement with an expected predominantly monomeric state of the mutant, the hydrodynamic radius determined by DLS peaks at $R_h = 3.5 \pm 0.3$ nm, whereas an $R_h = 4.2 \pm 0.1$ nm was determined for the dimeric wild type of SARS M^{pro} applying 90 μ M protein each. The melting temperatures of M^{pro} wild type and the mutant R298A determined by CD spectroscopy are significantly different in 150 mM NaCl, 15 mM NaH₂PO₄, pH 7.6. For the wild type enzyme the most significant change in absolute ellipticity monitored at 220 nm was detected at $T_m = 55.8 \pm 1.5$ °C. For the mutant a $T_m = 53.1 \pm 0.7$ °C was detected. These values correspond to the turning point of an optimized Boltzmann function calculated by Origin 8.5, fitting the raw data. The increased thermal stability of the wild type is expected to be a result of the dimerization, perhaps with a minor contribution of a local structural rearrangement due to the mutation [56]. In analytical size-exclusion chromatography, 90 μ M M^{pro} R298A were applied to the Superose 12 10/300 GL SEC column in the absence of a substrate. In spite of the mutation, a minor fraction of the mutated protease was observed to be still dimeric. In contrast M^{pro} wild type displayed no detectable amount of monomer and a small peak corresponding to a large oligomer beside the dimeric state (see also fig. 44). SAXS measurements further supported the predominant monomeric state of M^{pro} R298A, in contrast to the protease wild type. The altered pair-distribution function (P(R)-function (for explanation see chapter 3.2.5) indicated a different geometry and size for both molecules (fig. 25B). The three domains of the high-resolution structure of the monomeric mutant are well superimposable with the *ab initio* model, indicating that the correct overall fold is obtained (fig. 25C; DAMMIF), based on the P(R)-function of M^{pro} R298A (fig. 25 B). The low χ -value of the high-resolution structure ($R_g = 23.34$) of 0.98 (determined by CRY SOL) compared to the SAXS scattering pattern additionally demonstrates that both structures are well superimposable. Mass spectrometry that ensures a high mass resolution and accuracy did not detect a specific large oligomeric state which was previously reported by [189]. Nevertheless, detection of predominantly dimeric M^{pro} wt and predominantly monomeric mutant (fig. 25D) comprised a valuable test system for the method, qualifying it for analysing cleavage reactions of SARS-CoV M^{pro} (chapter 3.4). The activity of M^{pro} R289A is reduced by 92% compared to the wild type enzyme, as determined by monitoring the cleavage of the chromogenic substrate TSAVLQ-pNA at a wavelength of 405 nm. The loss of activity was already reported [56]. However, there are indications that certain protease substrates significantly alter the monomer-dimer equilibrium of M^{pro} [56; 132] demonstrated by AUC. An octameric state of the wild type main protease according to [189] has not been observed by the applied methods. A minor amount of octamer created by domain-swapping could artificially arise from either the termini of the protein or the experimental conditions. However, the octamer may theoretically be considered as a specific stencil, perhaps underrepresented in a host cell and promoting efficient polyprotein cleavage.

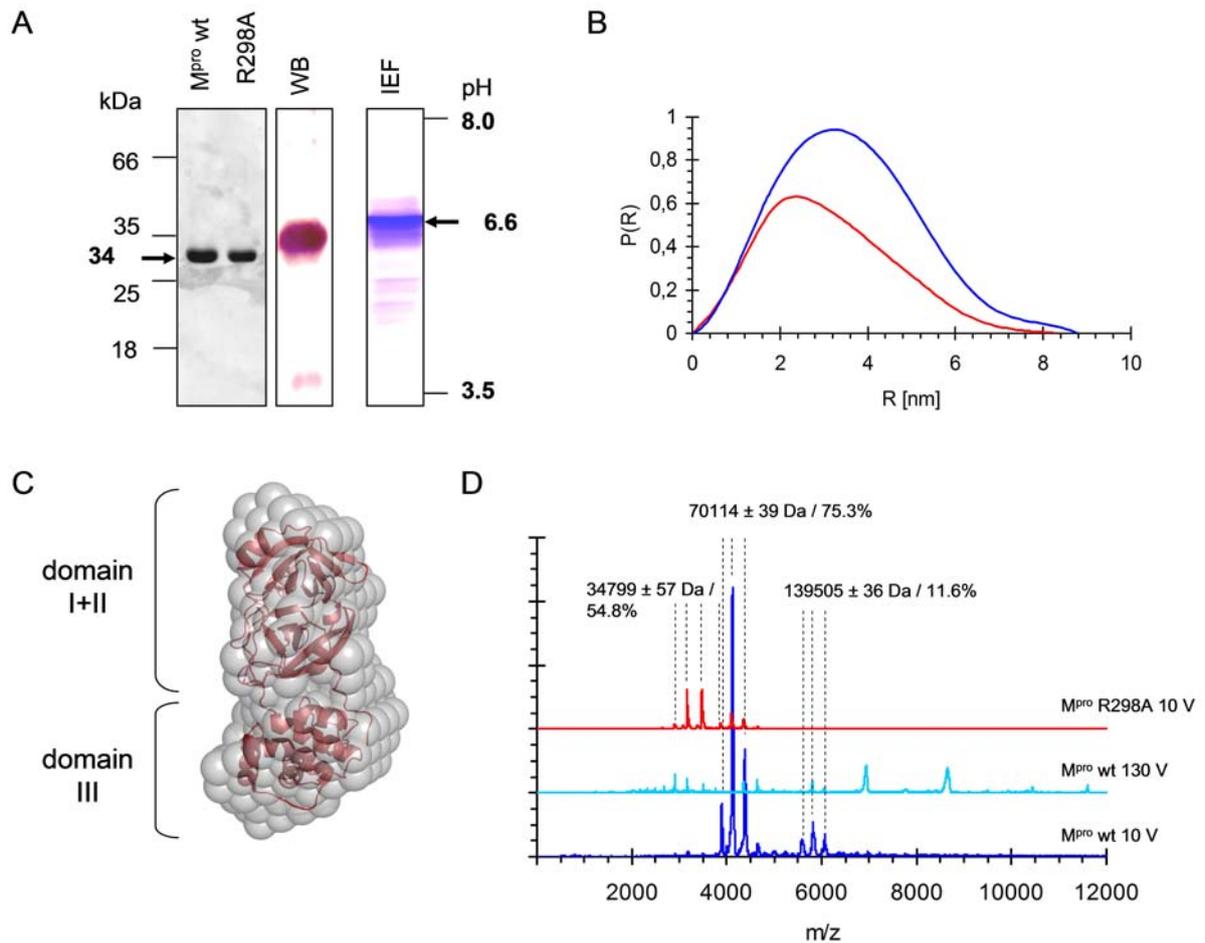


Figure 25. (A) SDS-PAGE analysis of purified SARS-CoV M^{pro} (Nsp5) and its mutant R298A after Ni-NTA affinity chromatography and size-exclusion chromatography. Western blot (WB) analysis of M^{pro} wild type verifies the presence of the his₆-tag by incubating the blot membrane with a primary goat anti-tetra-his antibody (QIAGEN) and a secondary anti-goat antibody conjugated with alkaline phosphatase. Isoelectric focusing (IEF) reveals a pI of 6.6. IEF Marker 3-10 (“liquid mix”, SERVA) was used as a standard. The gel was stained with SERVA violet. (B) The difference between the dimeric state of M^{pro} wild type (*blue*) and purified monomeric R298A mutant (*red*) is obvious from the P(R)-functions calculated from recorded SAXS datasets by GNOM. The maximum of the P(R)-function is shifted to a lower radius R and the maximum diameter for the mutant is slightly reduced. (C) The *ab initio* model of the R298A mutant fits the superposed high-resolution structure of the M^{pro} monomer (pdb code 3M3T, [56]). (D) Native ESI-Q-TOF mass spectrometry investigating the mass and oligomeric states of M^{pro} in solution. The wild type is predominantly dimeric (MW = 69.8 kDa) and the mutant is predominantly monomeric (MW = 34.8 kDa) in solution. Collision-induced dissociation of the dimer at 130 V collision voltage in the mass spectrometer resulted in the detection of the monomer subunit. The verified dimeric state of the wild type main protease is detectable by native ESI-Q-TOF MS.

3.4. Proteolytic processing of polyprotein Nsp7-10 by SARS-CoV M^{pro}

Purified SARS-CoV M^{pro} was incubated with the substrate pp710 of SARS-CoV and also of HCoV-229E, which is presumable cleavable by SARS-CoV M^{pro} with the same specificity, due to the highly conserved protease structure. *In vivo*, it was recently shown that a chimeric exchange of main proteases within different *Betacoronavirus* species resulted in only marginally decreased fitness of the virus [190]. The aim of this investigation is to semi-quantitatively analyze the specificity of the cleavage reaction with the full-length substrate pp710 comprising four Nsp domains. A generalized scheme of cleavage sites is presented in fig. 26A. The following protease cleavage experiments depicted in figure 26B, 27A and 28 were performed in sodium phosphate buffer pH 7.9 supplemented with 3 mM DTT and 280 mM NaCl. Using a simple SDS-PAGE approach, the overall order and time course of proteolytic cleavage of HCoV-229E and SARS-CoV pp710 by SARS-CoV M^{pro} was determined to be similar at a protease to substrate ratio of 1:10. An SDS-PAGE gel of the cleavage reaction involving SARS-CoV pp710 monitored for up to 44 h is displayed in figure 26B, while the first 20 min of the reaction with HCoV-229E pp710 as a substrate is depicted in figure 27A.

There was no indication that the ratio of M^{pro} and pp710 substrate alters the specificity or the cleavage rate of either pp710 homologue. Several intermediate processing products of pp710 were detected to be remarkably stable at the applied experimental conditions. Finally mature Nsp7, Nsp8, Nsp9 and Nsp10 were released (fig. 26, fig. 27).

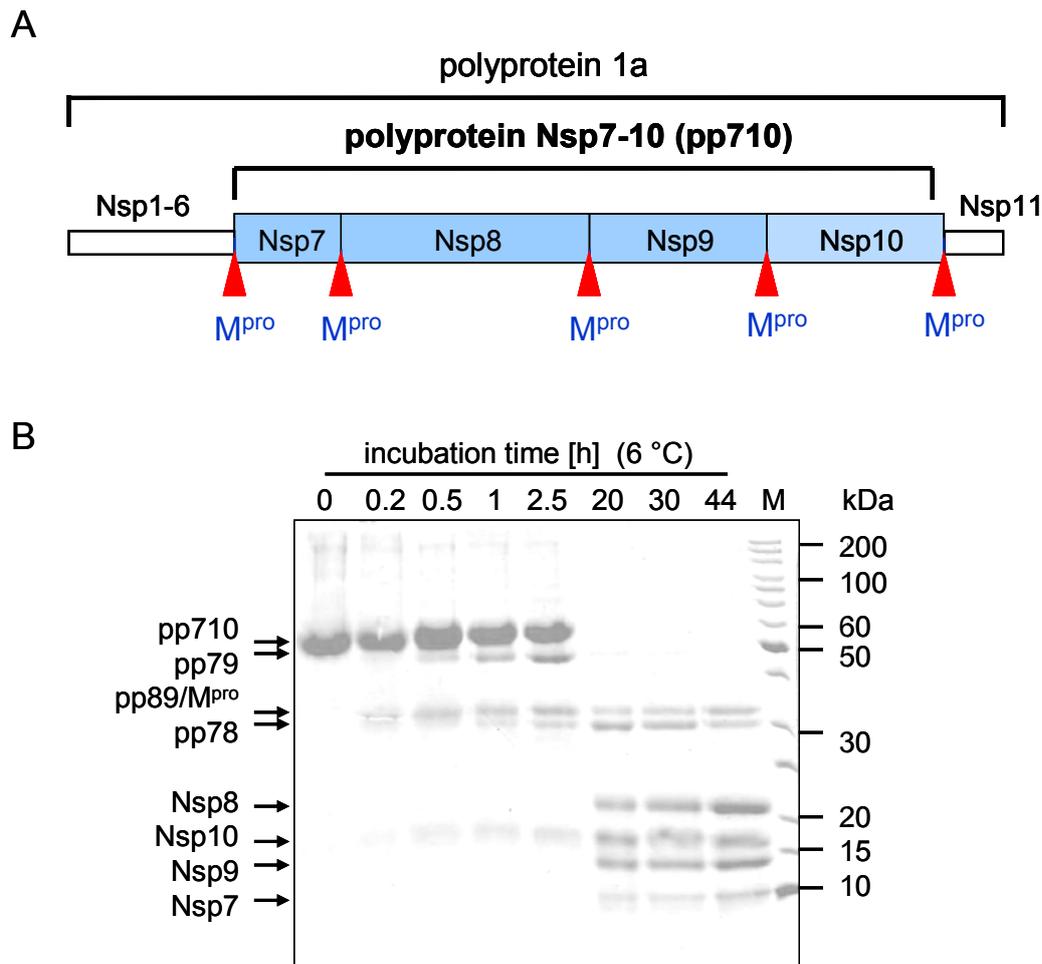


Figure 26. Proteolytic processing of SARS-CoV pp710. **(A)** Schematic overview of pp710, the main protease substrate under investigation. **(B)** Time course of SARS-CoV pp710 cleavage. A 10-fold molar excess of **SARS-CoV pp710** was incubated with M^{pro} and analyzed by SDS-PAGE at different time points up to a total incubation time of 44 h. For proteolytic processing, monomeric pp710 was purified by SEC. For the processing intermediates pp78 and pp89, an increased stability was indicated.

Noteworthy, SARS-CoV pp78 (Nsp7-Nsp8; connected by the cleavage site $\text{RATLQ}\downarrow\text{AIASE}$) as well as its homologue from HCoV-229E appears to be substantially more stable than SARS-CoV pp910 (Nsp9-Nsp10; $\text{TVRLQ}\downarrow\text{AGNAT}$) towards proteolytic cleavage (fig. 26 B). Pp78 and pp89 were still detected after 44 h of incubation. Therefore, the turnover-rate of the third cleavage site in between of Nsp8 and Nsp9 ($\text{AVKLQ}\downarrow\text{NNELS}$) appears to be low as well. Particularly, HCoV-229E pp910 (polyprotein Nsp9-10, ≈ 26 kDa) and pp78 that are released upon cleavage of the site in between Nsp8 and Nsp9 within HCoV-229E pp710 are hardly detectable.

3. Results and discussion

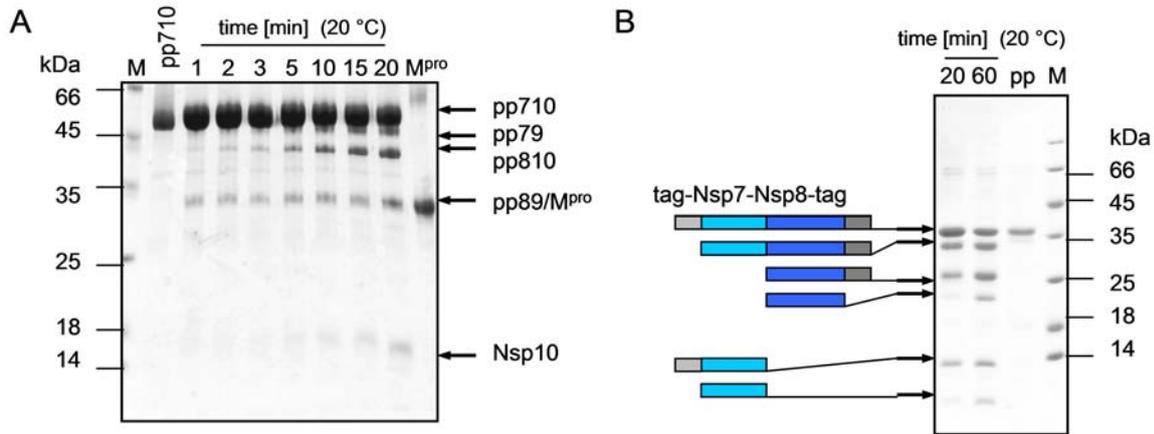


Figure 27. Time course of HCoV-229E pp710 processing by SARS-CoV pp710 M^{pro}. (A) Pp710 was incubated with SARS-CoV M^{pro} at a molar ratio of 10:1 at 20 °C. Samples were analyzed after different time points up to 20 min by SDS-PAGE. The cleavage products that were assigned to the detected protein bands are displayed. (B) Proteolytic processing of a 20-fold molar amount of HCoV-229E pp78 prepared with N- and C-terminal tag by SARS-CoV M^{pro} for up to 60 min. The cleavage products that were assigned to the detected protein bands are schematically displayed.

To investigate the cleavage sites flanking Nsp7 and Nsp8, including the Nsp8-Nsp9 cleavage site, pp78 of HCoV-229E was cloned with an N-terminal his₆-tag connected by a long linker (\approx 4 kDa) that contains the native M^{pro} cleavage site located between the membrane protein Nsp6 and Nsp7. The amino acids SILQ (P4-P1) of the cleavage site were introduced by the forward primer. The C-terminus of pp78 contained 9 N-terminal amino acids of Nsp9 and 24 C-terminal amino acids encoded by the vector pRSETA up to its “TAG” stop codon. Analysing the cleavage of pp78 by M^{pro} by SDS-PAGE (fig. 27B) in the presence of 2 mM DTT revealed that the amount of pp78 cleaved off from N-terminal tag and C-terminal tag, i.e. N-terminus of Nsp9 is very low. The specific activity of M^{pro} cleaving the site which links Nsp7 and Nsp8 seems to be remarkably higher. It appears that the cleavage site linking Nsp7 and Nsp8 as well as the site linking Nsp7 and the N-terminal tag are cleaved first. Consequently, after cleaving the covalent bond linking Nsp7 and Nsp8, the majority of Nsp7 and Nsp8 could still be covalently linked to the “membrane anchor” Nsp6 *in vivo* or more probable to Nsp9 during processing of the polyprotein. However, this is in contradiction to the Nsp6-Nsp7 cleavage site sequence (HCoV-229E) that is remarkably close to the consensus sequence of SARS-CoV M^{pro} as calculated by NetCorona 1.0 [40] (table 17). The relatively high stability of the cleavage site between SARS-CoV Nsp7 and Nsp8 (fig. 26B) compared to the less stable site in HCoV-229E should be further analyzed by (semi-)quantitative methods applying different oligomeric states.

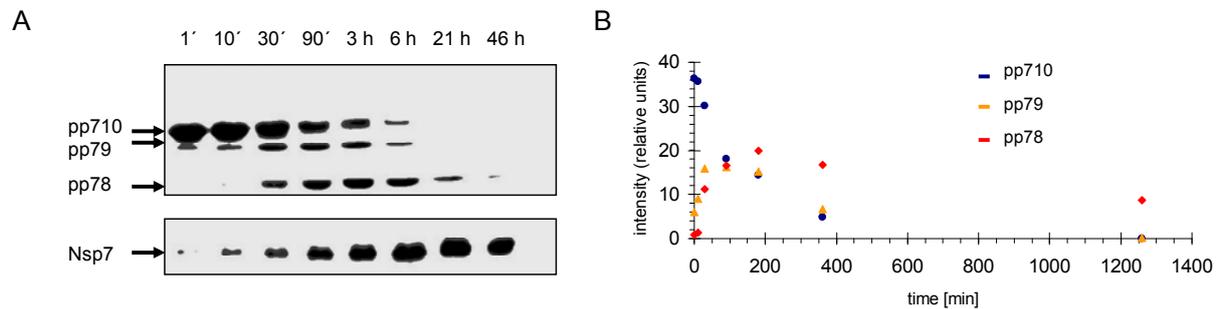


Figure 28. (A) Western blot analysis of SARS-CoV pp710 processing. (A) A ten-fold excess of **SARS-CoV pp710** was incubated with SARS-CoV M^{pro} at 20 °C for up to 46 h. *Western blot* after different time periods revealed a defined specificity of proteolytic cleavage. The primary antibody detected the N-terminal strep-tag of pp710. NBT and BCIP were used for detection. (B) The intensity of the signals on the membrane was quantified using *ImageJ*; the result is displayed for an incubation time of up to 21 h: pp710 in *blue*, pp79 in *yellow*, pp78 in *red*. The cleavage site connecting Nsp9 and Nsp10 is clearly preferred compared to the site that links Nsp8 and Nsp9. After 6 h the remaining amount of pp710 and pp79 was highly similar, while both species were completely processed after 21 h. Pp78 showed maximum concentration after 3 h of incubation, while most intensive signals for pp79 were detected after approximately 1.5 h.

For an additional semiquantitative comparison of cleavage products a western blot with SARS-CoV pp710 after affinity chromatography was performed, assuming to achieve increased quantification accuracy. A tenfold excess of SARS-CoV pp710 with N-terminal strep-tag was mixed with SARS-CoV M^{pro} and incubated at 20 °C with 3 mM DTT to further analyze the order and the time-course of the reaction (fig. 28). Western blot analysis revealed a defined specificity of cleavage. After 46 h of incubation, all intermediate cleavage products have been processed. The cleavage site connecting Nsp9 and Nsp10 is clearly preferred for proteolytic processing compared to the site that links Nsp8 and Nsp9 as well as Nsp7 and Nsp8 (fig. 28), in agreement with previous SDS-PAGE analysis (fig. 26B, 27A). After 6 h the remaining amounts of pp710 and pp79 are very similar after 21 h both species are gone. Pp78 peaks later than pp79 after approximately 3 h. Consequently the cleavage at the C-terminus of Nsp8 is again not preferred (see fig. 27A, for comparison). The reduced cleavage preference of this site can be explained by the large asparagine residue in P1' position (table 17). This effect was previously observed using short peptide substrates [65] and is supported by the rapid release of Nsp10 compared to free pp910 (fig. 27A). According to [191], a residue with a volume smaller than approximately 50 Å³ in P1' position is strongly preferred by M^{pro}. Similar to the semi-quantitative analysis by SDS-PAGE (fig. 26B and 27A) the online-server NetCorona 1.0 [40] (<http://www.cbs.dtu.dk/services/NetCorona/>) calculated that the cleavage site connecting SARS-CoV Nsp8 and Nsp9 is not perfectly matching the consensus sequence – indicated by a low score – and considered to be cleaved slowly, even though the score is still higher than for the cleavage site in

3. Results and discussion

between of Nsp7 and Nsp8. However, there are partly huge differences of the NetCorona 1.0 “cleavage score”, particularly of the Nsp6-Nsp7 “cleavage-score”, comparing different coronaviruses, e.g. HCoV-229E and SARS-CoV (tab. 17). This observation may reflect that the NetCorona “cleavage score” can hardly predict the actual turnover rate (tab. 17).

Table 17. Alignment of M^{pro} recognition/cleavage sites*: Cleavage sites between Nsp7-10 are highly similar comparing HCoV-229E and SARS-CoV. An unusual NNE motif in P1'-P3' position (blue), an alternative cleavage site (green) and some species-specific variations in P2' position are labelled (orange).

cleavage site	HCoV-229E	SARS-CoV	MERS-CoV	MHV ¹
Nsp4 - Nsp5	GSTLQ↓AGLRK	SAVLQ↓SGFRK	SGVLQ↓SGLVK	TSFLQ↓SGIVK
Nsp5 - Nsp6	GVNLQ↓SGKTT	GVTFQ↓GKFKK	GVVMQ↓SGVRK	GVKLQ↓SKRTR
Nsp6 - Nsp7	VSTVQ↓SKLTD(0.905)	VATVQ↓SKMSD(0.783)	VAAMQ↓SKLTD	VSQIQ↓SRLTD
Nsp7 - Nsp8	DSLQ↓SVASS(0.897)	RATLQ↓AIASE(0.838)	PSVLQ↓ATLSE	LOALQ↓SEFVN
Nsp8 - Nsp9	VVKLQ↓NNEIM(0.849)	AVKLQ↓NNEIS(0.860)	AVKLQ↓NNEIK	TVVLQ↓NNEIM
Nsp9 - Nsp10	TVRLQ↓AGKQT(0.880)	TVRLQ↓AGNAT(0.904)	TVRLQ↓AGSNT	TVRLQ↓AGTAT
Nsp10 - Nsp11	RTAIQ↓SFDNS(0.801)	EPLMQ↓SADAS(0.650)	DSLRLQ↓AALPQ	GSQFQ↓SKDTN
Nsp12 - Nsp13	STVLQ↓AAGLC	HTVLQ↓AVGAC	PTTLQ↓AVGSC	SAVMQ↓SVGAC
Nsp13 - Nsp14	MTDLQ↓SESSC	VATLQ↓AENVT	SSPLQ↓SYACW	NPRLQ↓CTTNL
Nsp14 - Nsp15	EVNLQ↓GLENI	FTRLQ↓SLENV	FTKVQ↓GLENI	FTRLQ↓SLENV
Nsp15 - Nsp16	YPQLQ↓SAEWK	YPKLQ↓ASQAW	YPRLQ↓ASADW	YPRLQ↓AAADW

* some cleavage sites were analyzed by NetCorona 1.0 [40]; “cleavage-score” [0;1] is specified in brackets

¹ murine hepatitis virus

Moreover, the western blot indicates that Nsp10 was cleaved off faster than Nsp7. Although the release of Nsp7 is already detectable in the beginning of the reaction, the amount quantified by using *ImageJ* (<http://rsbweb.nih.gov/ij/>) is at least slightly reduced compared to pp79, which is released after cleaving off Nsp10 from pp710 (fig. 28). Notably, the maximum intensity detected for pp78 was higher than for pp79, which is another indication for the stability of the Nsp7-Nsp8 cleavage site of SARS-CoV, even though the number of data points in the time-course of cleavage was limited in this experiment.

In SARS-CoV the Nsp7-Nsp8 cleavage site contains a threonine in P3 position instead of an isoleucine in HCoV-229E, but this position was generally observed to have a minor influence on k_{cat}/K_m [192]. Nonetheless, this putative effect on the cleavage reaction is probably too small to see it by SDS-PAGE or Western Blot for a full-length polyprotein. Interestingly, the k_{cat}/K_m -values listed by Grum-Tokars *et al.* [193] for 11-mer peptides spanning the M^{pro} cleavage sites (P6-P5') derived from SARS-CoV pp1ab indicate only a slightly lower catalytic efficiency at the cleavage site connecting Nsp7 and Nsp8 compared to the slowly cleaved site linking Nsp8 and Nsp9. A difference of 3% of the absolute k_{cat}/K_m -value of the rapidly cleaved N-terminal autocleavage site, i.e. the site connecting

Nsp4 and the protease, of the SARS-CoV M^{Pro} was determined. This report also aimed at standardizing conditions for an M^{Pro} activity assay, since the use of substrates of different length affects the kinetics of the protease [193].

A potential effect of the oligomeric state of pp710 on cleavage by M^{Pro} remains enigmatic. Even though the dimeric state of pp710 was enriched in the Western Blot experiment after SEC and larger oligomers were cleavable similarly well with a comparably high stability of the cleavage site linking Nsp8 and Nsp9. Using HCoV-229E pp710 obtained from the SEC peak containing larger oligomers the ratio of pp79 and pp810 seems to be slightly altered during cleavage compared to HCoV-229E pp710 after affinity chromatography (SDS-PAGE; figure 45). Even though these isolated oligomers tend to readjust an equilibrium in solution including smaller oligomers as observed by blue native PAGE (data not shown). Applying an authentic mixture of oligomeric states of pp710 to SAXS results in the calculation of an elongated average *ab initio* model, suggesting that some higher oligomers might be even more elongated than the dimer (data not shown). Therefore, the ratio of surface to volume would be increased to favour a rapid cleavage with a similar specificity compared to monomeric pp710. A cleavage site with an alternative recognition site in close proximity like confirmed for pp78 cleavage in MHV [107; 194] (table 17; *green*) is irrelevant for either HCoV-229E or SARS-CoV according to the amino acid sequence.

Since none of the SDS-PAGE gels to investigate the reaction of either HCoV-229E or SARS-CoV pp710 displayed a significant amount of pp910, due to rapid cleavage of the respective site (fig. 27A), the following model is promoted. Nsp10 needs to be released quickly from other proteins of pp1a to allow interactions, e. g. with either Nsp14 or Nsp16 for a time-efficient RNA-capping after infection. On the other hand, Nsp10 needs to be part of pp1a to enable a higher protein concentration, compared to Nsps that are not part of pp1a. As mentioned above the transcription level of pp1a exceeds the level of pp1ab by a factor of three to five, therefore a high level of Nsp10 could efficiently interact with Nsp14 and Nsp16 at the same time as speculated by [117], considering that their interfaces on the surface of Nsp10 are overlapping. In conclusion, the remarkable stability of the cleavage site connecting Nsp8 and Nsp9 as well as the persistence of pp78 for some hours at the applied experimental *in vitro* conditions promote the hypothesis that intermediate partly processed polyproteins may act as essential building blocks of the coronaviral RTC.

3.4.1. Interactions of mature Nsps

Next, the cleavage products obtained after pp710 processing have been investigated to identify non-covalent interactions among each other. Insights into Nsp-Nsp interactions will additionally verify possibilities to purify individual Nsps out of this cleavage mixture. Processed SARS-CoV pp710 obtained after incubation for 60 h with SARS-CoV M^{Pro} (fig. 26 B; SDS-PAGE) was applied to Strep-Tactin resin to obtain indications for protein-protein interactions by a "pull-down" approach (fig. 30C).

3. Results and discussion

Only N-terminally strep-tagged Nsp7 could bind specifically and efficiently to the resin. Thus, the detection of approximately half of the loaded untagged Nsp8 molecules in the elution fraction should be due to the interaction with bound Nsp7. A minor amount of Nsp9 co-eluted with Nsp7 and Nsp8 as well, while the amount of eluted Nsp10 is not significant.

The interaction of mature Nsps was further verified for HCoV-229E pp710 by size-exclusion chromatography analysis (figure 29A). As shown by SDS-PAGE analysis of peak C of the SEC chromatogram (fig. 29B), Nsp8 also co-localized with Nsp7 and a minor amount of Nsp9 of HCoV-229E. An interaction of Nsp8 with Nsp7 is suggested (fig. 29B), as already reported for SARS-CoV [98] and FCoV [96] in high resolution structures. The dimeric his₆-tagged SARS-CoV main protease (Nsp5; \approx 68 kDa) was also part of this peak, but not assumed to interact with mature Nsps in this mixture, identifying the protease a valuable additional mass standard. Nsp10 and a major amount of Nsp9 are co-localizing in peak D. Peaks comprising the *orange* chromatogram (fig. 29A; SEC buffer: 250 mM NaCl, 30 mM NaH₂PO₄, pH 7.9) were similarly observed, if the chromatography was performed at the same concentration of the protein mixture, but with a lower ionic strength (140 mM NaCl, 30 mM NaH₂PO₄ pH 7.5) that is considered to be closer to physiological conditions (data not shown). These results suggests that, under the applied conditions, there is either no or only a very weak interaction of Nsp10 with Nsp7 and Nsp8, consistently for SARS-CoV and HCoV-229E. After incubation of the processed Nsps of HCoV-229E with Ni-NTA resin prior to SEC to remove his-tagged M^{Pro}, Nsp10 interacted with the Ni-NTA resin, allowing the purification of HCoV-229E Nsp9 by SEC (fig. 29A, peak H). Comparing the *blue* and *orange* chromatogram (with and without further incubation with Ni-NTA resin after cleavage) peak H containing Nsp9 is remarkably shifted compared to peak D, which contains co-localizing Nsp9 and Nsp10. The position of SEC peak H at lower protein concentration ($V_E = 81$ ml) was indicative for a shift to the position characteristic for the monomeric state of Nsp9. Beside the well-known dimer, the monomer might appear at relatively low concentrations. The *blue* chromatogram was obtained with a buffer at pH 7.5 containing 50 mM Tris and 140 mM NaCl.

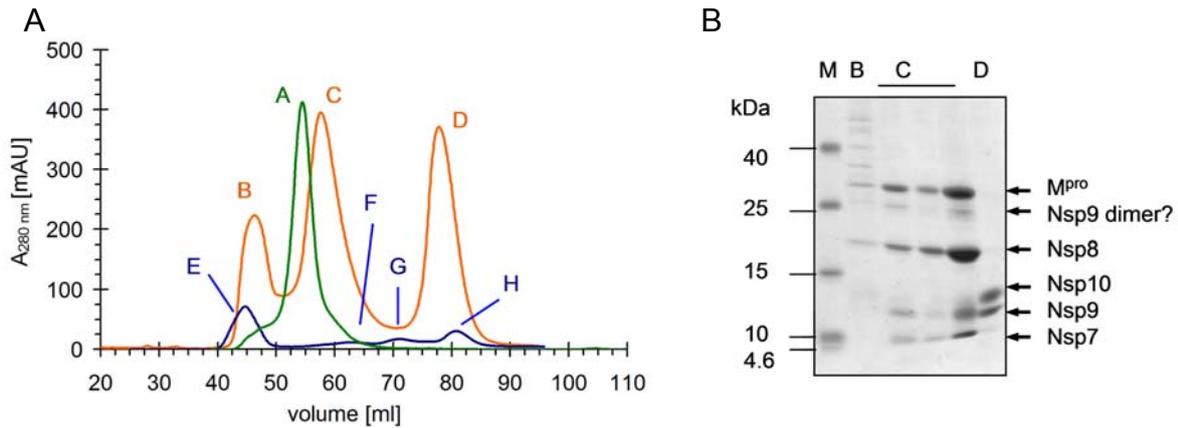


Figure 29. Interaction of cleavage products of HCoV-229E pp710. **(A)** Co-localization of HCoV-229E Nsp7, Nsp8, and Nsp9 indicated by SEC (*HiLoad 16/60, Superdex 75*). Purified monomeric HCoV-229E pp710 (58 kDa) is shown as an additional mass calibration (*green* chromatogram, peak A). With SARS-CoV M^{pro} completely processed HCoV-229E pp710 was initially incubated with Ni-NTA resin to remove the his-tagged protease and subsequently analyzed by SEC (*blue*), or directly analyzed by SEC in the presence of M^{pro} (*orange*). **(B)** Non-reducing SDS-PAGE gel investigating the protein composition of the SEC elution peaks B, C and D. Peak H was assigned to Nsp9, peak F was assigned to Nsp7 and Nsp8.

For comparison, completely cleaved SARS-CoV pp710 was applied to SEC as well (fig. 30B). According to the mass calibration, monomeric Nsp10 elutes at a volume comparable to HCoV-229E Nsp10, also co-eluting with Nsp9. The elution volume of SARS-CoV Nsp8 is comparable as well, reflecting a molecular weight of roughly 60 kDa. Altogether, different common structural characteristics of the corresponding part of the RTC are indicated.

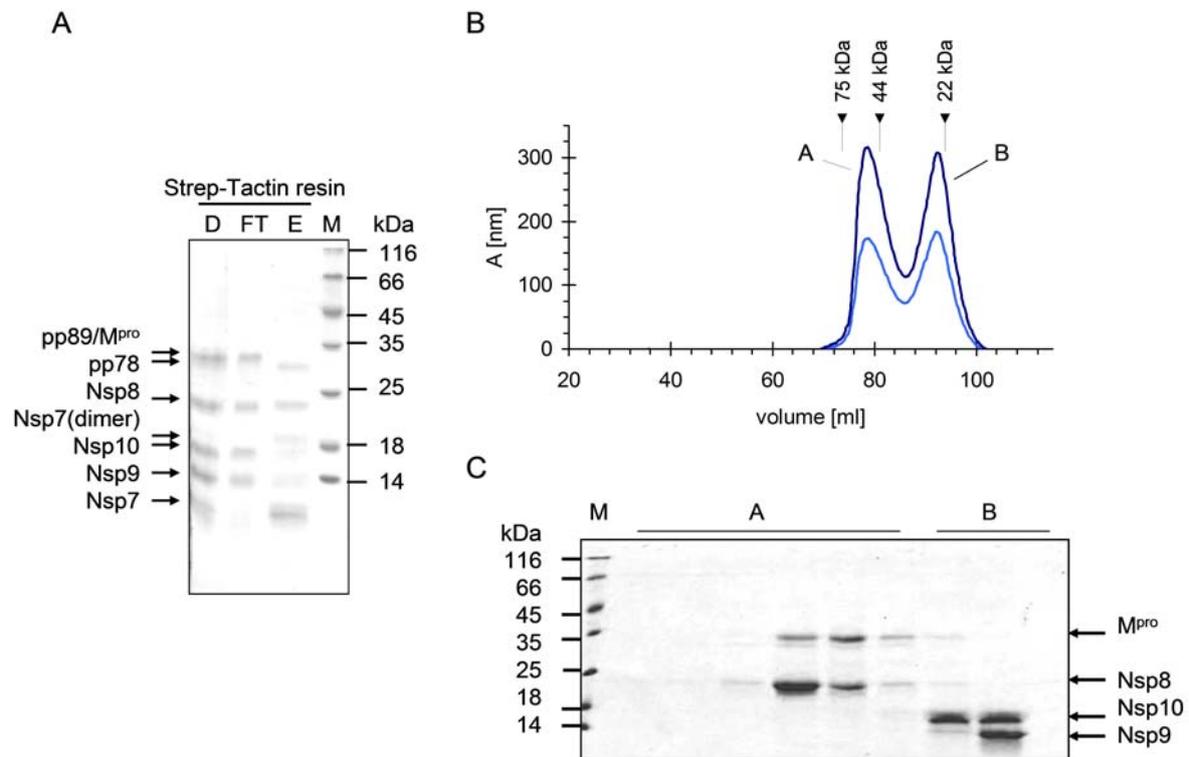


Figure 30. Interaction of cleavage products of SARS-CoV pp710. (A) The mixture of final cleavage products was applied to Strep-Tactin resin. Since Nsp7 contained an N-terminal strep tag, bound Nsp7 could "capture" interaction partners. D: SARS-CoV pp710 cleaved by M^{pro} after 60 h incubation; FT: "Flow-through", not interacting with the resin; E: Elution by 2 mM desthiobiotin; M: Molecular weight marker. (B) Cleavage products of SARS-CoV pp710 after incubation with SARS-CoV M^{pro} separated by SEC (*HiLoad 16/60 Superdex 200*). Two peaks were detected with an average elution volume of 78.7 ml and 92.3 ml, respectively, both corresponding to masses below 100 kDa. The absorbance was monitored at 280 nm (*dark blue*) and 260 nm (*light blue*). The protein was applied in 40 mM sodium phosphate buffer (pH 7.9) supplemented with 250 mM NaCl. (C) The composition of peaks A and B was revealed by reducing SDS-PAGE.

3.4.2. Investigation of pp710 processing and cleavage product interactions by native mass spectrometry

After analyzing the time course of pp710 maturation and the interactions of the cleavage intermediates and products by SDS-PAGE, western blot, and SEC techniques, the reaction was investigated by native mass spectrometry. This approach provides a higher mass accuracy, mass species resolution and allows mass determination at relatively low protein concentrations. After mixing 4 μ M SARS-CoV pp710 with 0.4 μ M SARS-CoV M^{pro}, masses corresponding to free monomeric Nsp10 (14.98 kDa) and, to a lower amount, remarkably monomeric Nsp9 (12.40 kDa) were detected, already after 80 s of incubation (fig. 31). Therefore, Nsp9 and Nsp10 do not seem to tightly interact with either Nsp7 or Nsp8 in the frame of experimental conditions. Nsp7 and Nsp8

which are known to interact with each other as well as a main protease dimer are likely overlapping with masses roughly around 60 kDa, which remains to be verified.

The mass of monomeric Nsp10 was slightly higher than expected, due to additional cation binding that occupied, for example, the Zn²⁺-binding sites. An intensity quantification of the ion peaks indicated that the amount of monomeric Nsp10 is significantly higher than the combined amount of monomeric and dimeric Nsp9 after 80 (appendix, fig. 86). During the course of the reaction, the relative intensity increase of monomeric Nsp9 is significantly lower than that of monomeric Nsp10 over more than 3 h. However, this could correlate with another homo- or heterooligomeric state of Nsp9, e.g. an interaction with pp78 or pp89 that would result in a continuous mass shift during the reaction. Furthermore, a 24 kDa mass species that fits an Nsp9 dimer increased in later stages of the reaction (appendix, fig. 86), particularly significant after 18 days. Monomeric Nsp9 is not necessarily in contradiction to an essential dimerization of Nsp9 *in vivo* [104]. In fact, it represents a further indication for a concentration-dependent monomeric state that is predominant at relatively low protein concentrations [102] in an early stage of the processing reaction. Applying isolated HCoV-229E Nsp9 to mass spectrometry in the same buffer, a monomeric state was detected as well (see chapter 3.12). Moreover, in contrast to the previous DLS experiments (chapter 3.3), highly sensitive mass spectrometry detected also small amounts of monomeric SARS-CoV M^{pro} at a concentration of 0.4 μM (34.87 kDa; fig. 31).

Interestingly, in addition to a mass of 120 kDa at the beginning of the processing reaction (fig. 31) considered to represent a non-covalent dimeric state of SARS-CoV pp710 (1 mM DTT was added to the reaction mixture), no other oligomers with a mass above 120 kDa was detected during the reaction. Most particles co-localize at around 60 kDa. This mass presumably comprises a huge variety of different interaction states involving Nsp8 and Nsp7. Free monomeric Nsp7 was not detected. The shape of peaks putatively corresponding to monomeric M^{pro} altered during the reaction, suggesting an overlay with other mass states at 3,000-3,600 m/z-value, perhaps around 40 kDa (fig. 31).

Considering that the concentration of Nsps at ER-derived membranes during virus replication is relatively high in an advanced infection stage [47], the impact of concentration-dependent oligomerization of Nsps is not reflected by the mass spectrometry experiments. However, it is worth to remember that the detection of main protease activity in HCoV-229E by a specific antiserum indicated a very low abundance [195]. To judge the regulatory impact of concentration-dependent oligomerization of Nsps *in vivo* in different infection stages, including the phenomenon of domain-swapping, it would be highly supporting to understand the absolute local protein concentration profile around the “replicase membrane structures” in more detail in the future.

3. Results and discussion

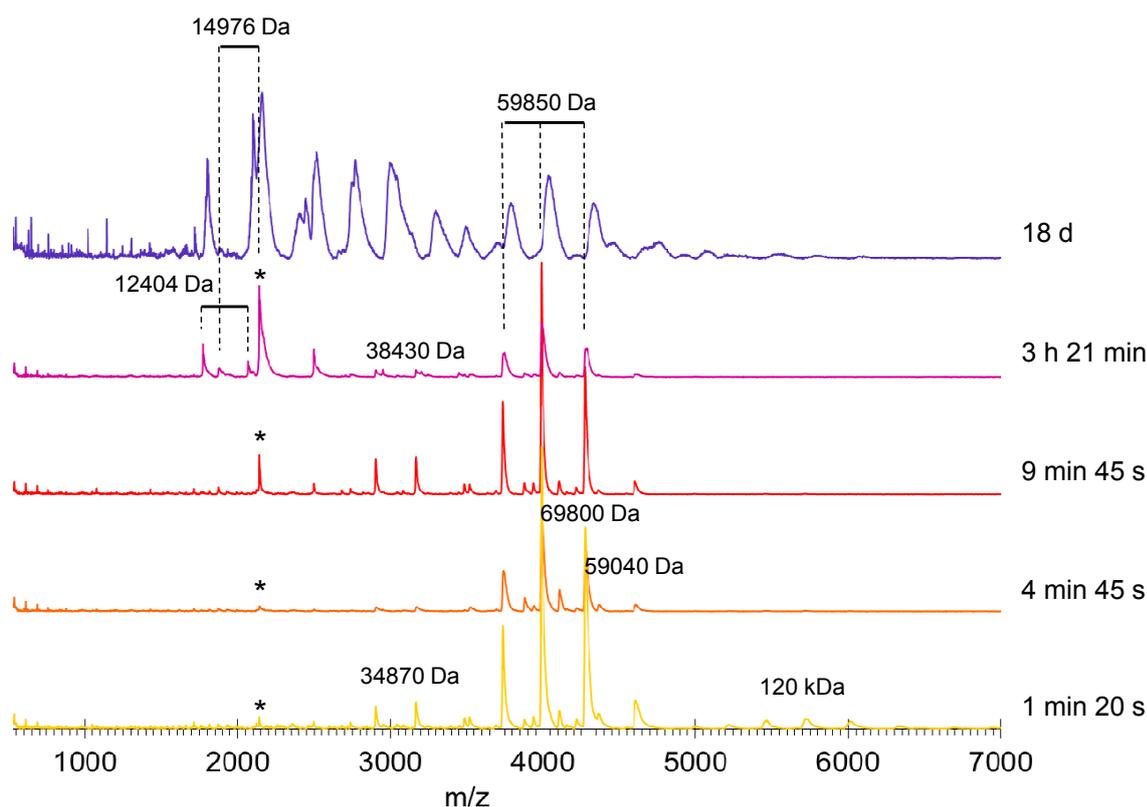


Figure 31. *Native ESI-TOF MS* analysis of proteolytic processing of **SARS-CoV pp710**. The time-course of mass changes during the cleavage reaction is revealed in a specific range of m/z -values. Pp710 in 350 mM ammonium acetate at pH 7.9 and SARS-CoV M^{pro} in 250 mM ammonium acetate adjusted to pH 7.4 were mixed (4 μ M:0.4 μ M final concentration) and analyzed by an LCT2 instrument (Waters) for relevant masses related to intermediates or mature cleavage products of pp710. The activity of the protease at these conditions was previously confirmed by SDS-PAGE (not shown). Monomeric (58 kDa) and dimeric pp710 (120 kDa) were present at the start of the reaction. An increasing amount of monomeric Nsp10 (asterisk, 14,976 Da, calculated by ProtParam, SIB [142]: 14,400 Da¹) and a small amount of monomeric Nsp9 (12.4 kDa, calculated: 12,400 Da) were detected. See appendix for the development of the relative fraction of mass species (fig. 86).

In a similar mass spectrometry experiment the cleavage of HCoV-229E pp710 by SARS-CoV M^{pro} was monitored by ESI-TOF, consistently revealing the release of mature monomeric Nsp9 and Nsp10 (fig. 32). A mass around 31 kDa is indicative for monomeric pp78 or a homodimer of Nsp7 and Nsp8 significantly increasing during the reaction, although a marginal amount of free Nsp7, below 1% of the total intensity of detected masses, was also observed. Remaining pp710 (58598 Da) and monomeric/dimeric main protease were detected and assigned in figure 32 as well. A mass of 33.9 kDa may be indicative for monomeric pp89, which is reduced in a later stage of the reaction. For one ion series, the mass of 42.9 kDa (42.8 kDa after 10 min 15 s) was assigned, increasing to around 8% of the total intensity (m/z 500-11,000) after 5 min 20 s in accordance with a heterotrimeric interaction of Nsp7, Nsp8 and Nsp9, characterized by an expected molecular weight of 42,969 Da. Alternatively, this mass species could be assigned to monomeric pp79 or an interaction of mature Nsp7 with pp89,

although a corresponding mass was absent in the processing reaction of SARS-CoV pp710 (fig. 31). An interaction of Nsp7 and Nsp8 [96] as well as Nsp8 and Nsp9 [102] was previously detected by other *in vitro* methods, which could indicate a complex formation of all three Nsps. Furthermore, Bost *et al.* [196] suggested an interaction of Nsp8 and Nsp9 *in vivo*. Some additional intermediate products that have not been well resolved by mass spectrometry were most likely overlapping with peaks of the SARS-CoV main protease at around 60-70 kDa, suggested by the broad shape of these peaks (figure 31, reaction after 1 min 20 s; fig. 32).

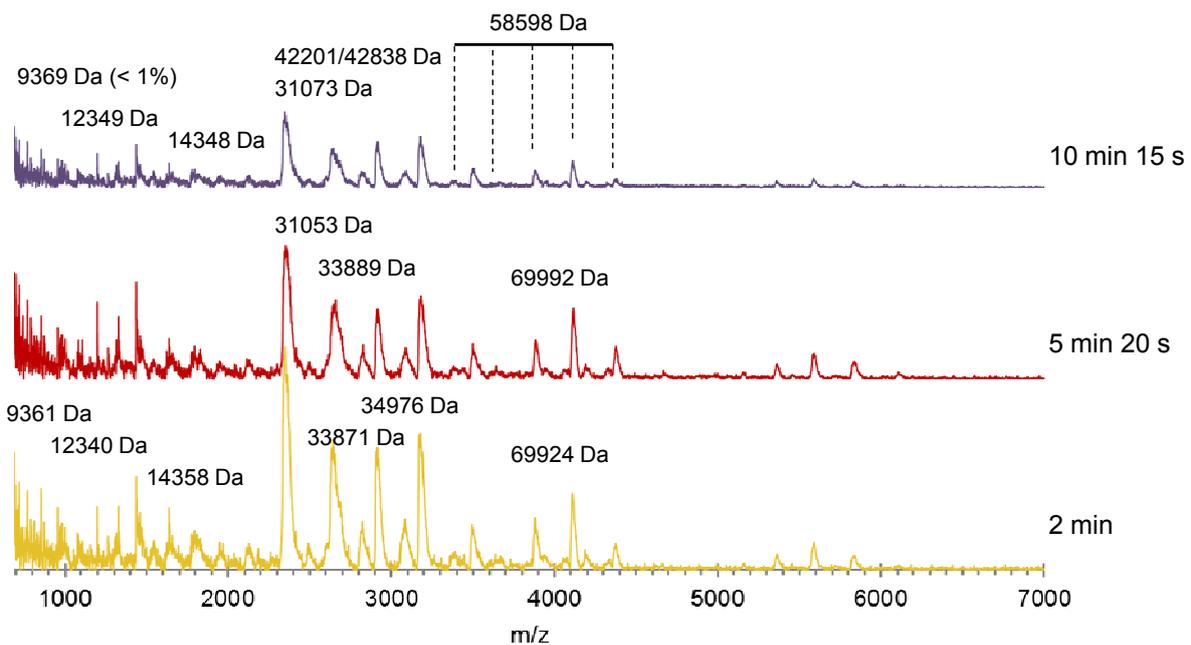


Figure 32. *Native ESI-TOF MS* analysis of proteolytic processing of **HCoV-229E pp710**. The time-course of mass changes during the cleavage reaction is revealed in a specific range of m/z -values within 10 min. The ratio of pp710 and protease is 10:1, incubated at 20 °C in 350 mM ammonium acetate buffer, pH 7.9. Similar to the reaction with SARS-CoV pp710 the masses expected for monomeric Nsp9 and monomeric Nsp10 as well as a large number of masses in the range of 40 to 70 kDa were observed.

In summary, several homo- and heterooligomeric states were assigned in the course of the pp710 processing pathway by mass spectrometry, assumed to be at least temporary components of the coronaviral RTC “core”. The oligomerization of mature Nsps after cleavage of pp710 appears to be highly conserved among *Alpha- and Betacoronaviruses*, as indicated by SEC and mass species like monomeric Nsp10 observed by mass spectrometry during cleavage of both pp710 homologues. Furthermore, taking a mixture of pp710 cleavage products in a certain stage of processing, it appears to be difficult to purify either Nsp7 or Nsp8 or a short temporary existing polyproteins derived from pp710 for further characterization. In the following chapter an alternatively developed strategy is presented.

3.5. Interaction of M^{pro} with substrates lacking glutamine in P1 position

3.5.1. Full-length polyproteins

As an additional strategy to purify individual polyproteins (e.g. polyprotein Nsp7-9 (pp79)) and mature Nsps originating from purified pp710 of HCoV-229E it was aimed to selectively block a protease cleavage site by site-directed mutagenesis. The glutamine residue in P1 position at selected main protease cleavage sites was exchanged by either alanine or asparagine which completely abolished the cleavage activity as proven by SDS-PAGE (fig. 33 C; column input).

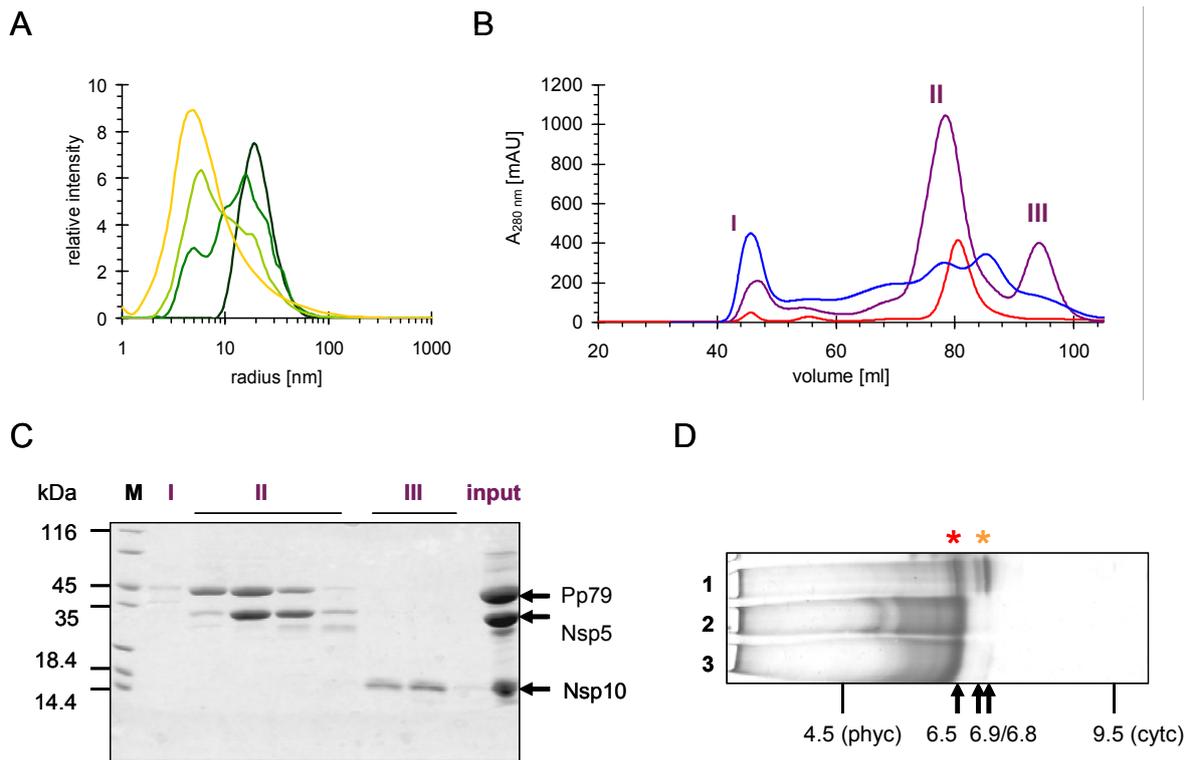


Figure 33. Investigation of the interaction of pp79 Q83A/Q278A with SARS-CoV main protease (A) Specific de-aggregation of HCoV-229E pp79 Q83A/Q278A by addition of SARS-CoV M^{pro} was detected by DLS measurements. Pp79, monitored before adding M^{pro} ($R_h = 19.4 \pm 1.2$ nm/ 18.8 ± 2.2 nm; *dark green*), 15 min after adding an equimolar ratio (1:1) of protease ($R_h = 11.9 \pm 2.4$ nm/ $4.5 + 0.5$ nm; *green*), as well as the same equimolar sample after 18 h ($R_h = 6.6 \pm 1.0$ nm; *light green*) and after 24 h of incubation ($R_h = 5.1 \pm 0.28$ nm; *yellow*). (B) Size distribution profile of processed pp710 Q83A/Q278A obtained from SEC (Hi Load 16/60 Superdex 200) after incubation with an equimolar amount of M^{pro} (*purple*) and the supernatant after an additional incubation with Ni-NTA resin binding the his₆-tagged protease (*blue*). Purified dimeric M^{pro} is shown for comparison (*red*). (C) SDS-PAGE investigating the composition of the elution fractions referring to the *purple* chromatogram in panel B. Notably, a minor amount of pp79 is aggregating (panel C, peak I). (D) NEPHGE; lane1: Freshly prepared processed pp710 Q83A/A278A (pp79 + M^{pro} + Nsp10), lane 2: Pp79/M^{pro}, left edge of the SEC peak II (panel B), lane 3: Pp79/M^{pro} right edge of the SEC peak II (panel B); NEPHGE marker (SERVA), phyc: Phytochrome C, cytc: Cytochrome C. All three lanes show one protein band at pI = 6.5 assumed to contain pp79 Q83A/Q278A and main protease.

Exchanging either Q83 and/or Q278 in HCoV-229E pp710 by alanine resulted in increased aggregation of the mutated polyprotein as evident by dynamic light scattering (DLS) and size-exclusion chromatography, which likely results from an enhanced local hydrophobicity ("hydrophobic patch", fig. 33A and table 24). The two mutations Q83A and Q278A target the P1 position of the cleavage sites connecting Nsp7 and Nsp8 as well as Nsp8 and Nsp9, respectively, which are consequently not cleaved anymore. In spite of large aggregates, the double mutant containing both glutamine/alanine substitutions (Q83A/Q278A) was completely cleaved by a catalytic amount of SARS-CoV M^{pro}, resulting in the release of mature Nsp10 and intact pp79 (fig. 33C). Applying size-exclusion chromatography (SEC) to the protein solution after cleavage, pp79 and Nsp10 were successfully separated from each other (fig. 33B, *purple* chromatogram). This result further indicated that Nsp10 did not interact with Nsp7-9 in the polyprotein state. For the mutated pp79 Q83A/Q278A isolated from peak I a hydrodynamic radius of $R_h = 18.8 \pm 2.2$ nm was determined by DLS (fig. 33A, *dark green* graph), indicating large aggregation. Adding an equal molar amount of SARS-CoV M^{pro} to the pp79 aggregates, the hydrodynamic radius stepwise decreased down to a monodisperse solution ($R_h = 5.1 \pm 0.3$ nm) after 24 h of incubation at 6 °C (fig. 33A, *yellow* graph). The observation of aggregate dissociation is highly indicative for a direct interaction of both proteins, most probably shielding the increased hydrophobicity at the mutated cleavage site of the polyprotein, although the protease is not capable of cleaving anymore. Thus, it was subsequently investigated, if new insights into the putative interaction of pp79 and M^{pro} could be obtained allowing insights into a state of M^{pro} binding to the cleavage site.

Indeed, SEC revealed a co-localization of both proteins (fig. 33B *purple* chromatogram; fig. 33C, peak II). To further probe the interaction, pp79 and his₆-tagged protease obtained from SEC peak II were incubated with Ni-NTA resin. After separating the resin by centrifugation, the supernatant contained almost no main protease and additionally only a reduced amount pp79; while an increased amount of aggregated pp79 was detected by SEC analysis (fig. 33B, *blue* chromatogram). This agrees with the postulated interaction of pp79 and M^{pro}, since free pp79 is presumed to be unable to bind to the Ni²⁺ ions on the resin. Nevertheless, SEC revealed only a marginal amount of larger heterooligomeric states of the putative interaction (fig. 33 B/C) which could be dimeric pp79 binding two protease dimers at one mutated site, where a protease dimer could theoretically bind another substrate. Such interactions might be sterically impaired or, regarding the dimeric protease, result from the observation of an asymmetric fold, i.e. only one active centre within the dimer is able to catalyse a reaction at the same time [52].

In SEC, the pp79-M^{pro}-complex is characterized by a retention volume of 78.4 ml (16/60 HiLoad Superdex 200; 41.0 ml void volume), main protease peaks at 80.5 ml (fig. 33B *purple* chromatogram).

3. Results and discussion

Purified pp79 with main protease as well as a mixture containing SARS-CoV M^{PrO} and HCoV-229E pp710 (Q83A/Q278A; molar ratio 1:1) were further applied to NEPHGE. Similar to isoelectric focussing (IEF), this native pH-gradient electrophoresis technique is particularly suitable for moderate pI values in the range between 6 and 11). Similar to the IEF result (chapter 3.13, fig. 77A) two bands (pH 6.8 and 6.9) arise from Nsp10 (fig. 33D, *yellow asterisk*), whereas main protease and pp79 Q83A/Q278A again seem to co-localize at around pH 6.5 (fig. 33 D, *red asterisk*), as previously observed.

To quantify the binding affinity of SARS-CoV M^{PrO} to the double mutant pp79 Q83A/Q278A, *surface plasmon resonance spectroscopy* was performed, allowing the K_D -value determination. HCoV-229E pp79 (49 fmol) was immobilized by EDC/NHS coupling to the dextran layer functionalized CM5-chip. M^{PrO} was injected at concentrations ranging from 0-400 μM at a flow rate of $20 \mu\text{l min}^{-1}$ and the association time of the protein was limited to 240 s. The binding equilibrium was reached within this time even at low protease concentrations (fig. 34A). The maximum RU value was plotted against the ligand concentration and the fit of the data points resulted in a dissociation constant of $K_D = 11 \pm 0.8 \mu\text{M}$, according to the *one-site-binding model*.

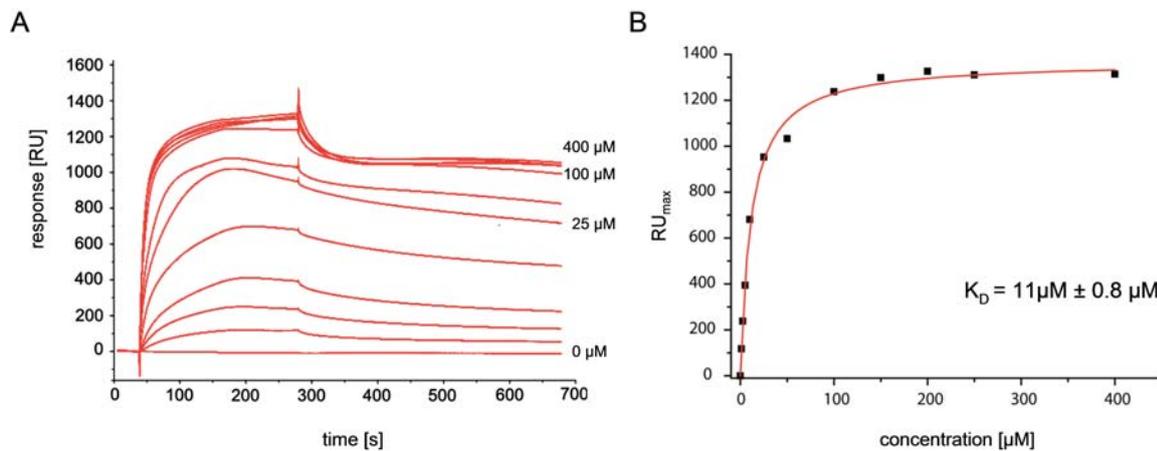


Figure 34. Interaction of HCoV-229E pp79 Q83A/Q278A and SARS-CoV M^{PrO} analyzed by surface plasmon resonance spectroscopy. (A) Sensorgram of the injected M^{PrO} and (B) the dependence of the maximum resonance on the concentration of SARS-CoV M^{PrO} as determined by *SPR spectroscopy*. The fit resulted in a dissociation constant in the low μM -range according to the *one-site-binding model*, proving the suggested interaction.

Insights into size and geometry of the pp79/M^{PrO} interaction were obtained by SAXS analysis. The complex isolated by SEC after mixing mutated HCoV-229E pp710 with SARS-CoV M^{PrO} (molar ratio 1:1) was analyzed in the presence of 400 mM and 100 mM NaCl, respectively, combined with ultracentrifugation at either $280,000 \times g$ or $541,000 \times g$ for 30 min. Results are listed in table 18. Apparently, the smallest R_g value was determined at low ionic strength after centrifugation at $541,000 \times g$. Shape and volume of the particle fit the expected parameters of monomeric pp79 (molecular mass 43 kDa) (table 18). Considering an average protein density of 1.22 g cm^{-3} [197] a

particle volume of 65 nm³ results in a calculated molecular mass of 48 kDa. Since pp79 contained two putative protease binding sites at the mutated polyprotein cleavage sites connecting Nsp7 and Nsp8 as well as Nsp8 and Nsp9 and since either monomeric or dimeric protease could bind to the two cleavage sites, rigid body modelling based on the scattering data was applied to determine the preferred binding mode. It was specified that only one of the cleavage sites is occupied by protease. Consequently, four different complexes were modelled (table 19). A low χ -value of the rigid-body fit function calculated by SASREF and a low penalty indicate high similarity to the raw scattering data. Scattering vectors up to $s \leq 0.24 \text{ \AA}^{-1}$ were taken into consideration for three independent rounds of SASREF modelling with default penalty score specifications. The high-resolution structures included for modelling as well as the obtained quality measures are specified in table 19. According to rigid-body modelling, it is statistically preferred that one of the mutated cleavage sites of monomeric pp79 is occupied by monomeric M^{pro} (tab. 19) – for both cleavage sites the χ -value is in a similar range still considering that there might be mixtures of oligomeric interaction states.

Table 18. Pp79 and M^{pro}: SAXS result summary of R_g, D_{max} and particle volume depending on sample preparation conditions.

Condition	concentration [g l ⁻¹]*	R _g [nm]	D _{max} [nm]	V [nm ³]
280,000 × g, 400 mM NaCl	2.3	5.1	15.4	212
280,000 × g, 100 mM NaCl	0.5; 1.0; 1.9	4.1	14.2	175
541,000 × g, 400 mM NaCl	0.6; 1.2; 2.3	3.5	12.3	97
541,000 × g, 100 mM NaCl	0.5; 1.0; 1.9	2.9	10.1	65

* scattering data obtained for different protein concentrations was merged, lacking a concentration-dependent aggregation or particle repulsion.

Table 19. Pp79 and M^{pro}: Samples centrifuged at 541,000 × g supplemented with 400 mM NaCl further analyzed by SASREF [input high-resolution structures: pdb code 2AHM (SARS-CoV Nsp7/Nsp8), 2J97 (HCoV-229E Nsp9), 2C3S (SARS-CoV M^{pro})]; all putative complexes were modelled considering P1 symmetry.]

M ^{pro}	monomeric M ^{pro} *		dimeric M ^{pro}	
binding site	Nsp7/8	Nsp8/9	Nsp7/8	Nsp8/9
χ -value	0.82 ± 0.03	0.83 ± 0.02	1.08 ± 0.12	1.04 ± 0.12
penalty	0.06 ± 0.09	0.20 ± 0.12	0.27 ± 0.16	0.17 ± 0.08

* one protomer of the dimer

The interaction mode of polyprotein and protease appeared to be worth for further investigation to enlighten the process of processing of RTC building blocks in more detail. The single mutant HCoV-229E pp710 Q278A containing a C-terminal his₆-tag was generated, purified and cleaved by incubation with an equimolar amount of SARS-CoV M^{pro}. Since the mutation is located at the

3. Results and discussion

cleavage site connecting Nsp8 and Nsp9, an interaction of protease and pp89 Q278A is expected, along with the cleavage products of mature Nsp7 and Nsp10. Applying HCoV-229E pp710 Q278A in the presence of an equimolar amount of M^{pro} to SEC analysis in a revealed the expected co-elution of M^{pro} and pp89QA, which have a similar molecular weight around 34 kDa. The identity of both proteins in the SDS-PAGE gel (fig. 35B) was verified by ESI-Q-TOF MS after trypsin digestion. Interestingly, mature Nsp7 (9.3 kDa) also co-eluted and therefore interacted with pp89QA (fig. 35B, peak B), presuming that Nsp7 did not interact with M^{pro}, which is rather unlikely. Furthermore, Nsp10 was separated by SEC (fig. 35B, peak C), as previously observed after cleaving HCoV-229E pp79 Q83A/Q278A (fig. 33).

Since the processing rate of pp89 by SARS-CoV M^{pro} was previously demonstrated to be very slow (chapter 3.4), the putative interaction of pp89 and Nsp7 could represent a major intermediate building block of the RTC, presumably even *in vivo*. Thus, Nsp7 is suggested to be capable of interacting not only with mature Nsp8 [96; 98], but also with larger pre-processed polyproteins (e.g. pp89 and pp810), even though a direct interaction of Nsp7 and Nsp9 was not excluded so far. The abundance of pp810 (Nsp8-10) is considered to be rather low due to the rapid cleavage of the Nsp9-Nsp10 cleavage site.

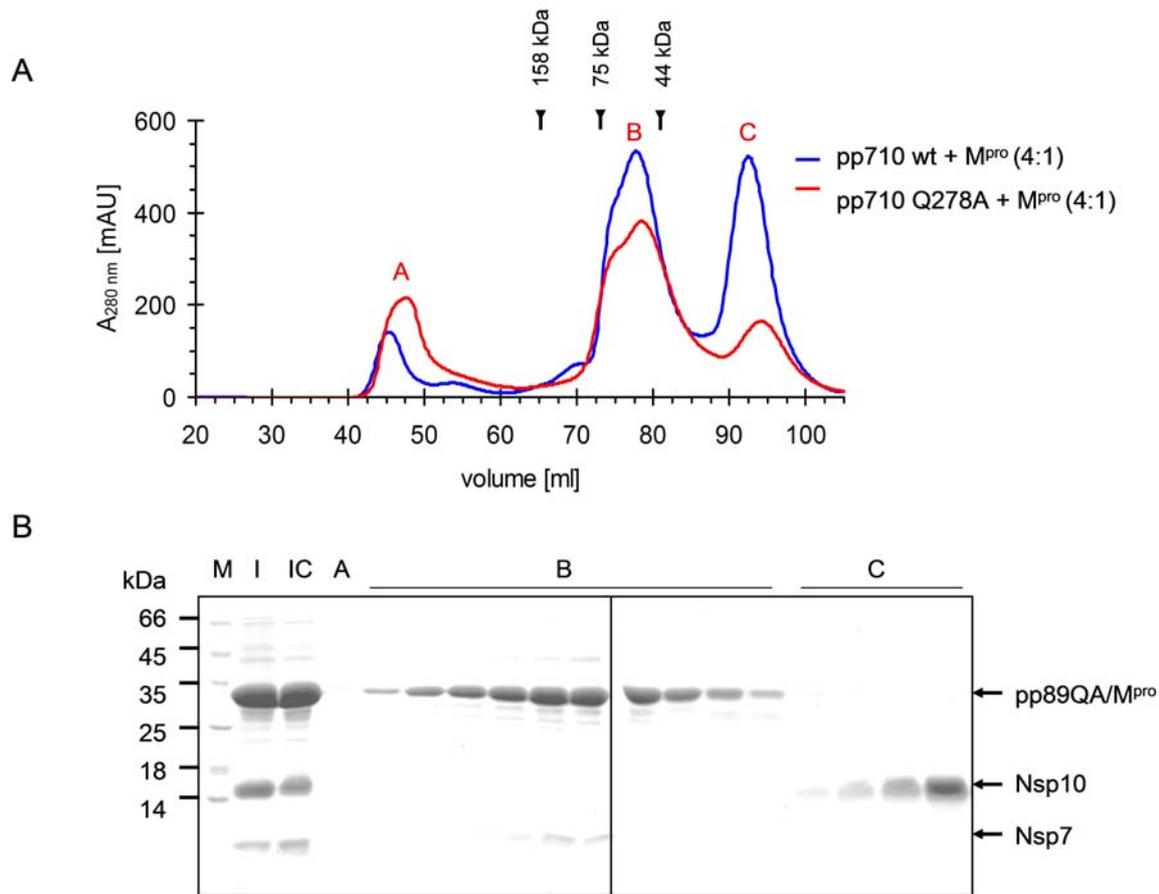


Figure 35. HCoV-229E pp710 Q278A processed by SARS-CoV M^{pro}. (A) As evident in a SEC experiment (HiLoad 16/60 Superdex 200), pp89 is co-localizing with M^{pro} and Nsp7 (peak B, red chromatogram) assuming that the Nsp7 is directly interacting with Nsp8. The approximated total MW is below 100 kDa (red chromatogram, peak B), while the proteins might be capable of additional oligomerization. The positions of the major peaks are highly similar compared to the “mature” state after completely cleaving pp710 WT (blue chromatogram). (B) SDS-PAGE analysis revealing the composition of peaks A-C (red chromatogram, see panel A). M: protein standard; I: column input, cleaved HCoV-229E pp710 Q278A and protease; IC: column input after ultracentrifugation at 280,000 × g for 30 min.

Low-resolution structural data of the complex consisting of Nsp7, pp89QA, and M^{pro} co-eluting from SEC column should be obtained applying SAXS techniques. The P(R)-function depicted in fig. 36B is indicative for an elongated particle. The strongly elongated *ab initio* model of the complex surface well matched a straight elongated helical N-terminus of Nsp8 and an interface of Nsp7 close to a hinge loop, approximately involving residues 100-120 of Nsp8 (fig. 36). Therefore the fold of Nsp8 in this polyprotein complex is assumed to be more similar to the isoform with stretched N-terminus reported within the SARS-CoV Nsp(7+8) complex high-resolution structure (pdb 2AHM, chain H) than to the alternative bended conformation that was observed in the same complex. As a requirement for rigid-body modelling by SASREF, Nsp9 was covalently linked to the C-terminus of Nsp8. An interface of close contact between Nsp7 and Nsp9 is not indicated by the *ab initio* model, in agreement

3. Results and discussion

with a SEC analysis of completely processed wild type HCoV-229E pp710, which showed that a strong interaction between Nsp7 and Nsp9 is absent (chapter 3.4.1/3.4.2). Therefore, a previous yeast-2-hybrid approach that displayed an initial indication of a direct Nsp7-Nsp9 interaction in one of two conceivable fusion protein orientations [87] is not supported at the applied experimental conditions.

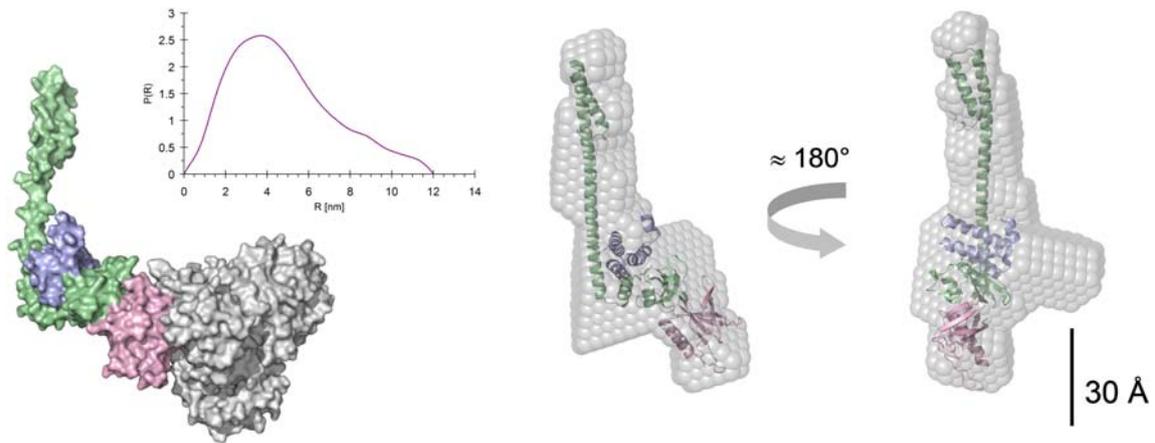


Figure 36. SAXS investigation of pp89 interaction with M^{pro} and Nsp7. (A) SASREF rigid-body model of the single mutant pp89 Q278A in complex with Nsp7 and main protease, consisting of SARS-CoV Nsp7 (blue; extracted from pdb code 2AHM) and Nsp8 (green; chain H, pdb code 2AHM), HCoV-229E Nsp9 (pink; pdb code 2J97), and dimeric main protease (grey, pdb code 1UJ1). Modelling restraints were defined that Nsp8 and Nsp9 are covalently linked by C- to N-terminus (distance 5 Å) and that main protease interacts with the protease recognition site (P5-P5') connecting Nsp8 and Nsp9 within the polyprotein. (B) The maximum diameter of the complex is 12 nm, according to the distance distribution function (P(R)-function) (C) Ten *ab initio* models generated by DAMMIF are characterized by a mean normalized spatial discrepancy value of 0.91 ± 0.04 , calculated by DAMAVER. The average model is shown. The calculated particle volume of 135 nm^3 corresponded to an approximated molecular weight of 82 kDa [$MW_{\text{Nsp7+8+9}} = 43 \text{ kDa}$, $MW_{\text{protease}} = 34 \text{ kDa}$]. The *ab initio* and the rigid-body model without M^{pro} are well superimposable. The remaining volume of the *ab initio* model can be explained by a certain fraction of main protease molecules at the expected position.

Moreover, an average rigid-body model of pp89-Nsp7- M^{pro} complex was calculated that fits well to the *ab initio* model previously described (fig. 36C). Additional particle volume in the *ab initio* model that is not occupied by Nsp7, Nsp8 or Nsp9 in the rigid-body model is interpretable by the main protease, partly filling the putative binding site at the mutation spot. The volume-based molecular weight determination approximately reflected an equimolar ratio of all molecules within the complex. However, a mixture of free cleavage site, binding sites occupied by a dimeric protease molecule or heterodimers containing monomeric main protease cannot be excluded. This might be explained by the slow catalytic efficiency of the protease, cleaving at a site with a large asparagine residue in P1 position [192] which may result in a higher dissociation constant of both interaction partners, as investigated in chapter 3.5.3. Beside the monomeric state, the Nsp7 and Nsp8 interface within the *ab initio* model of the pp89-Nsp7- M^{pro} complex is in good agreement with that observed in a mature

Nsp7-Nsp8 complex of SARS-CoV [98], although the individual Nsp structures were independently fitted into the SAXS model. Furthermore, the SAXS model provided an indication for an at least weak interaction of Nsp8 and Nsp9 in the pre-processed state, as discussed for the mature Nsps [101] (chapter 3.12.1).

The proteolytic processing of the HCoV-229E pp710 Q278A mutant with SARS-CoV M^{pro} wt was additionally investigated applying native ESI-TOF MS to verify suggested protein-protein interactions as well as the masses of the corresponding complexes (fig. 37A). Around 40% of the total intensity at the detector after 1 min of incubation resulted from a mass of around 70 kDa ($69,866 \pm 47$ Da) that increased during the next 3 h and would fit the mass of dimeric protease. At the same time, a 4% fraction (with reference to the total intensity in the displayed m/z range) of an approximately 600 Da lower mass at 69,313 Da appeared with a relatively high standard deviation and in close proximity to other m/z species that may correspond to a 1 to 1 interaction of main protease and mutated pp89. A mass of 34,974 Da for monomeric M^{pro} was determined by native ESI-Q-TOF MS/MS spectra and 33,969 Da is the expected mass of pp89. In addition, two masses above 100 kDa were detected. Particularly the mass of 139.8 kDa which is too high for representing a pp710 dimer, well fits a dimer of pp89 interacting with a dimer of protease, where the mass of 105 kDa could represent a monomer of pp89 interacting with dimeric protease. The relative fraction of 34,915 \pm 6 Da corresponding to monomeric M^{pro} significantly decreased during incubation (fig. 37A), probably due to a concentration-dependent interaction with a polyprotein substrate in an advanced stage of cleavage or a substrate-induced alteration of the monomer-dimer equilibrium of the main protease. For comparison, the proteolytic processing of pp710 Q278A was also performed with the predominantly monomeric main protease mutant R298A to probe the requirement of the dimeric protease wild type for the formation of a specific interaction state. In fact, the two mass species above 100 kDa observed in the presence of wt M^{pro} were not detected in the reaction with the R298A mutant (fig 37B, see also chapter 3.3). Thereby, it is indicated that these mass species (105 kDa, 139.8 kDa) involve and require the native dimeric state of the wild type. In agreement with SEC (chapter 3.4.1), only a small amount of free mature Nsp7 at a mass of 9,337 Da was detected by mass spectrometry after a reaction time of more than 3 h in the presence of the wild type protease (fig. 37A), indicating that mature Nsp7 partly interacted with pp89. The determined mass of $43,013 \pm 78$ Da after 3 h 24 min (fig. 37A) agrees with the expected mass of pp89 interacting with Nsp7. Even though, the whole putative complex of monomeric Nsp7 interacting with monomeric pp89 binding to main protease was not observed among the assigned masses, it was possible to identify masses indicative for fragments of the expected complex with two interaction partners, i.e. pp89 and Nsp7 as well as pp89 and protease.

3. Results and discussion

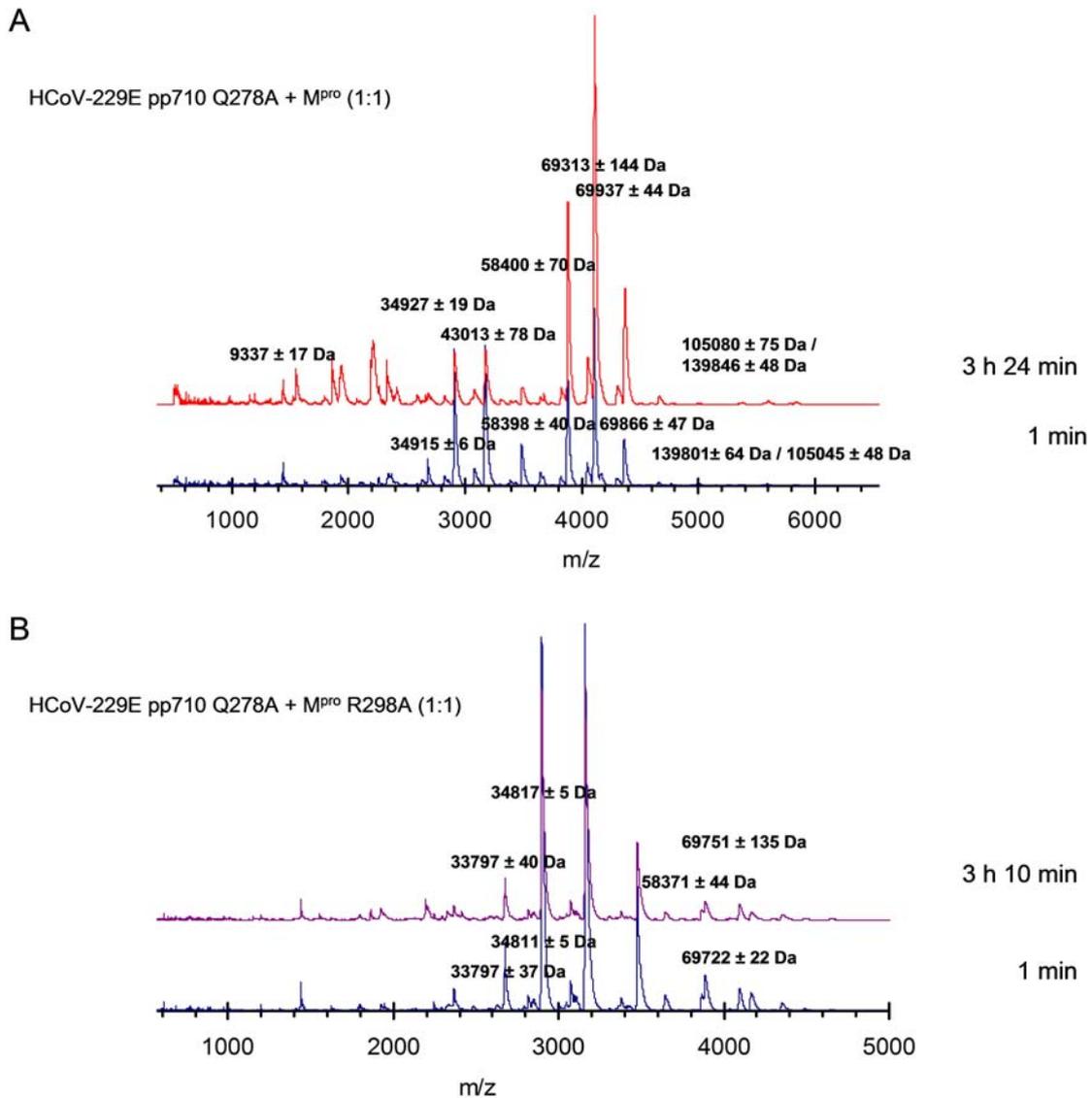


Figure 37. Native ESI-TOF MS investigating the interaction of M^{PRO} wt or M^{PRO} R298A with HCoV-229E pp710 Q278A. (A) MS spectra of pp710 Q278A incubated with an equimolar ratio of M^{PRO} for 1 min and for 204 min at 20 °C, respectively. (B) Pp710 Q278A was mixed with an equal molar amount of mutant M^{PRO} R298A and investigated by native ESI-TOF MS at similar time points. The displayed masses were assigned according to a simulation performed with Massign.

Trying to understand the composition of detected mass species (fig. 37 A/B) it has to be considered that nucleic acids present in the protein preparation from *E. coli* cell lysis could bind particularly to pp89. Applying SARS-CoV pp89 containing a C-terminal strep-tag, separately purified by Strep-Tactin affinity chromatography and in the absence of Nsp7, to SEC, the absorbance ratio at 260 nm/280 nm of a putative monomeric state is strongly increased and higher than in a fraction of larger oligomers (data not shown). This observation suggests that nucleic acid is binding and de-

99

aggregating pp89. However, prior to the mass spectrometry experiments described in this chapter a probably high content of nucleic acid of the samples was strongly reduced by incubation of the protein with benzonase prior to the affinity chromatography.

The formation of specific protease-polyprotein substrate complexes after mutation of the protease recognition site was revealed by DLS and additionally verified by SEC and MS experiments.

3.5.2. Full-length polyproteins with Q/N substitution in P1 position

For comparison, a different amino acid exchange in P1 position within a different polyprotein substrate on the way to form the RTC is the focus of the forthcoming section. Assuming an increased authenticity and affinity of the protease-polyprotein interaction, an exchange of glutamine in P1 position of the respective cleavage site by asparagine was favoured in forthcoming experiments to further enlighten and understand this interaction described in the beginning of this chapter. Moreover, an exchange by asparagine indeed reduced the additional aggregation of the respective polyprotein compared to the exchange by alanine presented earlier.

In contrast to the pp79 double mutant that was described at the beginning of this chapter (fig. 33), pp78 Q83N contains only one mutated cleavage site for stable binding of M^{pro}. Using SDS-PAGE it was verified that the Q→N mutation likewise completely inhibited the proteolytic processing at the cleavage site connecting Nsp7 and Nsp8 (fig. 38). SEC analysis of HCoV-229E pp710 Q83N processed by SARS-CoV M^{pro} displayed a significant co-localization of mutated pp78 and M^{pro} within a symmetric peak (fig. 38A, *light blue* chromatogram) which is again indicative for a major fraction of a relatively small heterooligomeric interaction state of pp78 and M^{pro}. The composition of the peaks A, D and E (fig. 38A, *blue* chromatogram) was investigated by SDS-PAGE (fig. 38B), as for the peaks A, B, C and E of the *green* chromatogram (fig. 38A) investigated by SDS-PAGE, shown in figure 38B (lower panel). The *green* chromatogram (fig. 38A) of incompletely processed pp710 Q83N reveals co-localization of the main protease with pp78 Q83N and to a lower extent with pp79 Q83N; polyproteins are present in excess. However, it is remarkable that pp79 and pp78 peak separately, indicating that the formation of a heterodimer of pp78 and pp79 is not preferred. Nsp9 and Nsp10 peak separately (fig. 38, peak E in all chromatograms) and are therefore not interacting with either pp78 or M^{pro} for the chosen experimental conditions.

3. Results and discussion

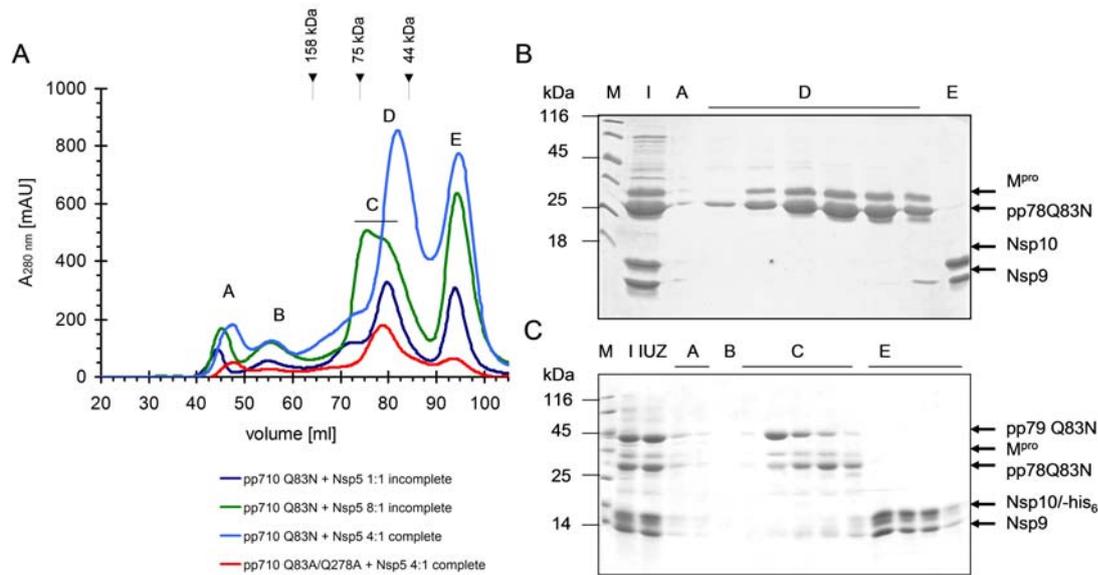


Figure 38. Proteolytic processing of HCoV-229E pp710 Q83N by SARS-CoV M^{pro} . (A) SEC analysis (HiLoad 16/60 Superdex 200) of either completely or incompletely processed pp710 Q83N. (B) SDS-PAGE analysis of the *light blue* chromatogram. Further cleavage of the pp78 product is forestalled due to the Q83N mutation as evident from SDS-PAGE analysis. (C) Reducing SDS-PAGE analysis of the *green* chromatogram. In addition to the co-localization of mutated pp710 fragments and M^{pro} during cleavage, the separate elution of pp79 Q83N and pp78 Q83N may be noted (peak C, *green* chromatogram). For the chosen experimental conditions – including a phosphate buffer with 250 mM NaCl – this implicates that there is no tendency to form a heterodimer, considering that in parallel both polyproteins are at least partly interacting with M^{pro} . [peaks A-E are labelled, M: protein molecular weight standard, I: SEC column input, IUZ: column input after ultracentrifugation]

In terms of the generally accepted fact that the dimerization of the SARS main protease is required for catalytic activity [49-51], small peaks additionally observed by SEC might contain dimeric protease interacting with a polyprotein. Nevertheless, as shown by the experiments described above, binding to a native substrate does not seem to necessarily require the dimeric state of M^{pro} , but there could be a significant influence on the monomer-dimer equilibrium of the protease in the presence of a full-length substrate to trigger catalysis in trans. In this context, a substrate-induced dimerization of M^{pro} was observed using the short cleavable peptide substrate TSAVLQ-pNA [48; 132]. Both studies applied analytical ultra-centrifugation to investigate a putative re-dimerization of mutated monomeric SARS-CoV main protease in solution. Moreover, it was reported that NaCl influences the monomer-dimer equilibrium of M^{pro} , reflecting that the majority of the binding energy results from hydrophilic interaction [198], even though it required more than 1 M of NaCl to release a significant amount of monomer from dimeric protease. This effect is therefore most likely not relevant for the experimental conditions applied in this study.

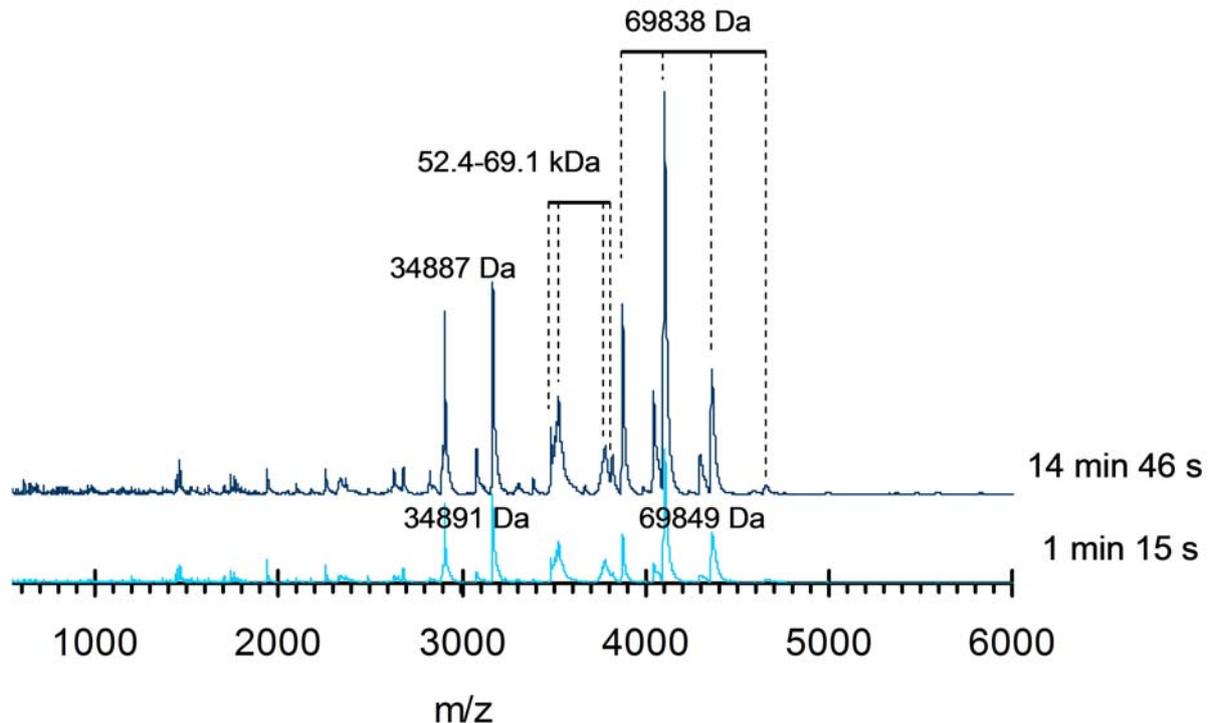


Figure 39. Native ESI-TOF MS investigating the interaction of SARS-CoV M^{pro} with HCoV-229E pp710 Q83N incubated at 20 °C and at equimolar ratio for 75 s and for 14 min 46 s, respectively. Peaks corresponding to 52.4 – 69.1 kDa that are not well resolved indicate the interaction of the SARS-CoV M^{pro} with pp78 Q83N (MW_{expected} = 65.7 kDa for a heterodimer). Other depicted masses fit a monomeric and dimeric state of SARS-CoV M^{pro}.

Aiming a high mass accuracy to investigate protein-protein interactions ESI-TOF was applied again to investigate the proteolytic processing of the herein described HCoV-229E pp710 Q83N mutant (fig. 39). Beside monomeric and dimeric protease (34,891 Da/69,849 Da), different mass species in close proximity to each other on the m/z-axis corresponding to masses ranging from 52.4 to 69.1 kDa are observed during the reaction that would be in good agreement with a heterodimer of pp78 Q83N (expected MW = 30.9 kDa) with authentic ends and SARS-CoV main protease (expected MW_{monomer} = 34.9 kDa). However, for this polyprotein the additional peaks could at least partly contain dimeric pp78 (expected MW = 61.8 kDa), even though, considering the absence of the protease-polyprotein complex, particularly a monomeric state of pp78 would be expected to be more abundant (see chapter 3.9.2 for comparison).

To confirm the complex formation of SARS-CoV M^{pro} with the pp78 Q83N and to investigate the ability of the main protease to dimerize during interaction, native PAGE was applied. This experiment addressed if pre-dominantly monomeric M^{pro}, like the mutant R298A, is able to interact with a mutated cleavage site in a monomeric state or if partly dimerization induced by the substrate occurs. N-terminally GST-tagged pp78 Q83N was purified and the tag was removed by precision protease as described in the method section (fig. 40 A). Pp78 Q83N was pre-incubated with main

3. Results and discussion

protease for 2 h at 22 °C, while the protease cleaved off six amino acids of the C-terminus of Nsp6 together with four additional residues that remain after removing the N-terminal GST-tag. A native polyacrylamide gel was loaded with SARS-CoV M^{pro} wt, the R298A mutant and HCoV-229E pp78 Q83N as well as mixtures of the protease and pp78 Q83N (fig. 40B). Since the pI of the protease was determined to be around 6.6 (fig. 25), both variants of the M^{pro} migrated slowly towards the anode at pH 6.5 and stopped at the increased pH value of 8.0 in the lower part of the gel. The electrophoretic mobility of pp78 Q83N was significantly altered in the presence of main protease. Interestingly, the extend of the shift was even more significant in the presence of the M^{pro} mutant R298A (fig. 40A). A small piece (approx. 15 mm²) of the lowest band in the lane containing a 1:1 molar ratio of pp78 Q83N and M^{pro} R298A was carefully excised. After *in-gel* digestion with trypsin peptides were applied to an iontrap ESI-MS/MS instrument. Pp78Q83N as well as M^{pro} were both identified with adequate sequence coverage. Consequently, both the gel shift and the co-localization within the same protein band clearly demonstrated a significant interaction of pp78 Q83N with M^{pro}, even more pronounced with the monomeric state of the protease.

The complex of M^{pro} wt and pp78 Q83N at a 1:1 molar ratio was applied to SAXS; an elongated particle ($R_g = 2.96 \pm 0.01$ nm) that is well interpretable by a 1:1 heterodimeric interaction was revealed (fig. 40C). Nevertheless, it is not ruled out that there is some remaining homodimer of either protein inside that fits the approximated molecular weight similarly well. The maximum diameter of 9.6 nm is close to the respective value of monomeric pp78 ($D_{max} = 8.7$ nm) indicating that at low concentrations and after extrapolation to infinite dilution pp78 is monomeric in this mixture. Nonetheless, pp78 tends to form an increased amount of dimer at higher concentrations as indicated by an increasing radius of gyration. The sample concentrations ranged from 0.8 to 6.5 g l⁻¹ (not shown).

Mass spectrometry spectra displayed in figure 39 (see also fig. 37) basically agree with the existence of a putative 1:1 interaction of mutated pp78/pp89 and protease, perhaps in addition to a minor amount of additional heterooligomers, corresponding to peaks indicated in the respective figures.

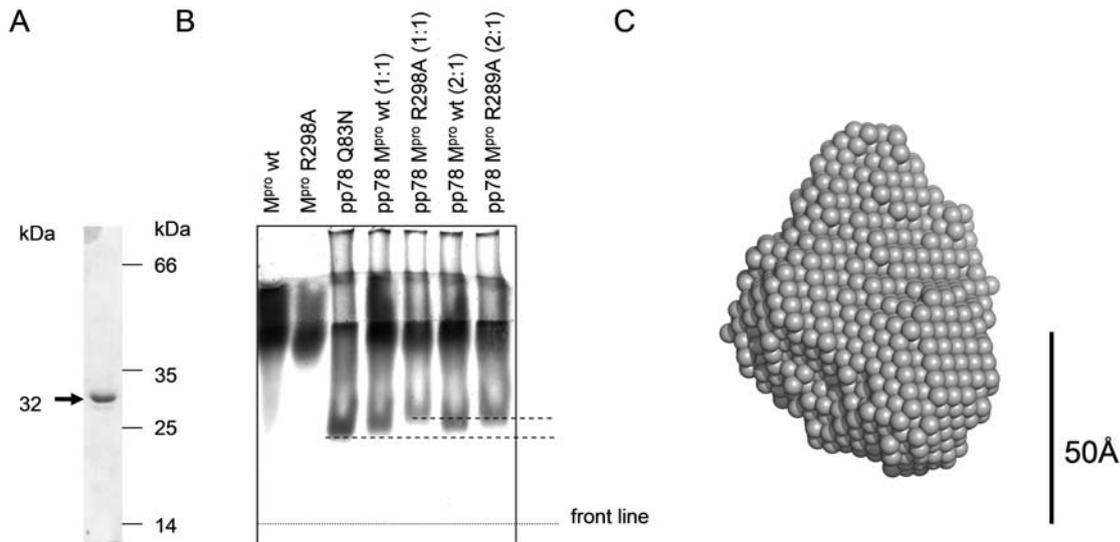


Figure 40. Investigation of the interaction of HCoV-229E pp78 Q83N with SARS-CoV M^{pro} and its mutant R298A. (A) SDS-PAGE gel showing pure pp78 Q83N (B) Native electrophoretic mobility shift assay: The upper gel layer contains 4% polyacrylamide (PAA) at pH 6.5; middle layer: 10% PAA pH 6.5; low separating layer: 12% PAA, pH 8.0. Electrophoresis was performed in TBE buffer supplemented with 100 mM NaCl. The electrophoretic mobility shift upon the addition of M^{pro} R298A is highlighted by dashed lines. The molar ratio of protease and pp78 Q83N is depicted above. Pp78 Q83N was pre-incubated with main protease for 2 h at 22 °C before electrophoresis (C) Average *ab initio* model of a particle in solution representing a 1:1 pp78/M^{pro}-mixture considerably able to interact at equimolar ratio obtained by SAXS. The average NSD-value of 10 models is 1.68 ± 0.06 , with a radius of gyration of $R_g = 2.96 \text{ nm} \pm 0.01$ and a maximum diameter of 9.6 nm. All samples were supplemented with 3 mM DTT to avoid radiation damage, which did not alter the results compared to non-reducing conditions.

3.5.3. Inhibition of M^{pro} by peptides derived from polyprotein recognition sites

After obtaining fundamental insights into the interaction of “uncleavable” polyproteins with M^{pro}, short peptides derived from the main protease cleavage sites were analyzed in terms of interaction and inhibition of M^{pro}. *In silico* analysis using FlexidockTM as well as *in vitro* characterization using SPR spectroscopy and native mass spectrometry was performed to quantify the binding affinity of the peptides. Initially, pentapeptides containing either Q, N or A in P1 position of the protease cleavage site were analyzed *in silico* concerning the impact of this residue (fig. 41) on the affinity, compared to peptides carrying a serine or asparagine residue in P1' position, respectively (see table 20 for overview).

3. Results and discussion

3.5.3.1. *In silico* peptide docking

Pentapeptides representing native and modified cleavage-recognition sites of SARS-CoV M^{pro} (pdb code 3SND) were used for *in silico* docking to its surface as described in the method section. The calculated binding energy for all ligands is listed in table 20. Substitution of the essential glutamine in P1 position of the peptide by asparagine resulted in the formation of a stabilizing hydrogen bond with G143 (2.43 Å) of the enzyme, while a relatively weak hydrogen bond with N142 of the enzyme (3.39 Å) was observed for a native glutamine in P1 position, according to the docking approach (fig 42). In contrast, alanine substitution in P1 position did not result in an alternative hydrogen bond to strengthen the peptide-enzyme interaction. Therefore, according to reduced hydrophobicity and the increased binding affinity, peptides containing asparagine in P1 position, considerably close to a native substrate, were favoured to investigate the inhibition of the protease. It was predicted that the peptide KLQNN is also capable of interacting with HCoV-229E M^{pro} in a similar way (fig. 42, HCoV-229E M^{pro} in *green*).

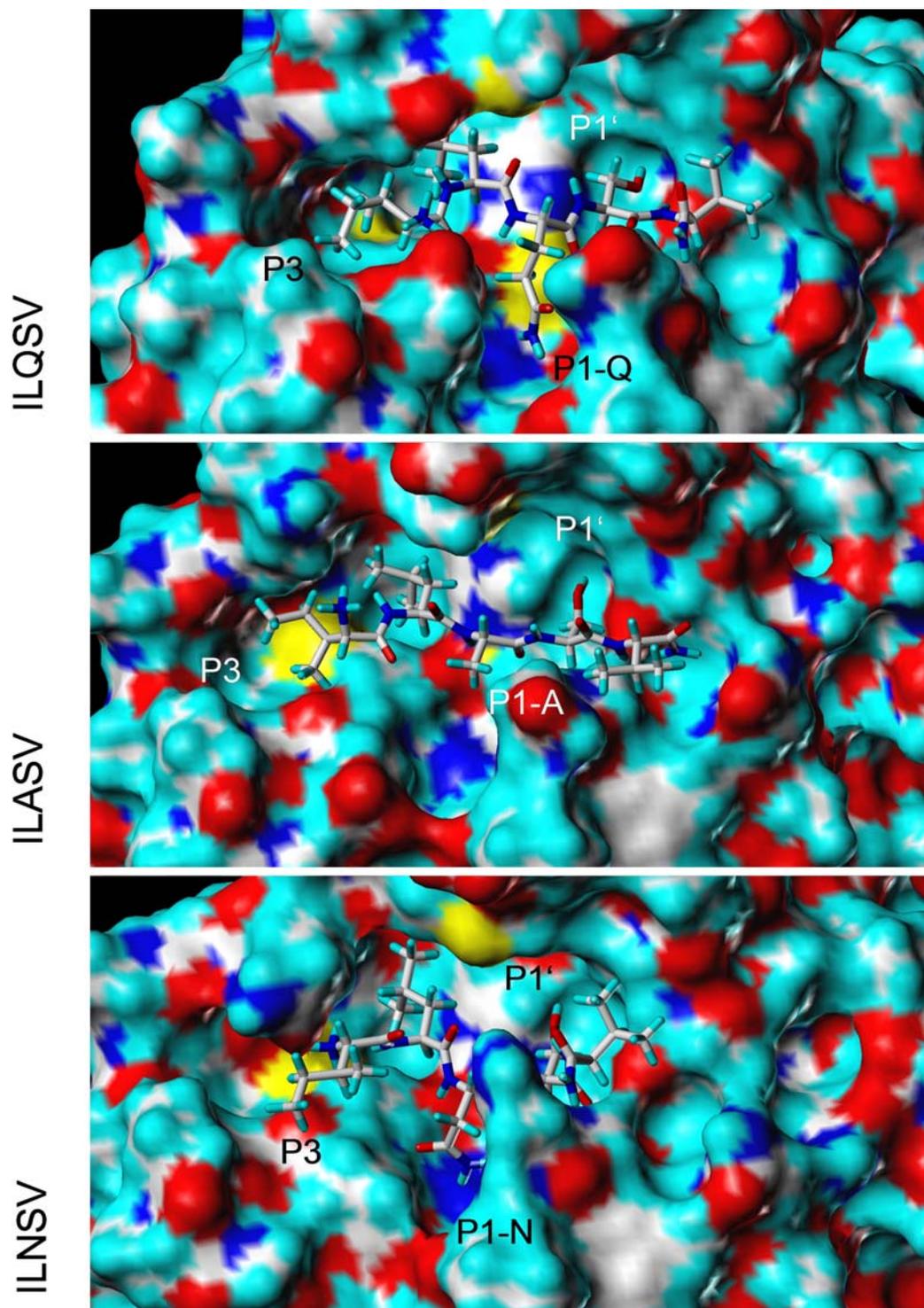


Figure 41. Pentapeptides representing native and modified cleavage-recognition sites of SARS-CoV M^{pro} (pdb code 3SND) were used for docking using FlexidockTM. The calculated binding energies were listed in table 20. Substitution of the essential glutamine in P1 position by asparagine results in a putative stabilizing hydrogen bond with G143 (2.43 Å) of the enzyme, which did not form in case of an alanine substitution.

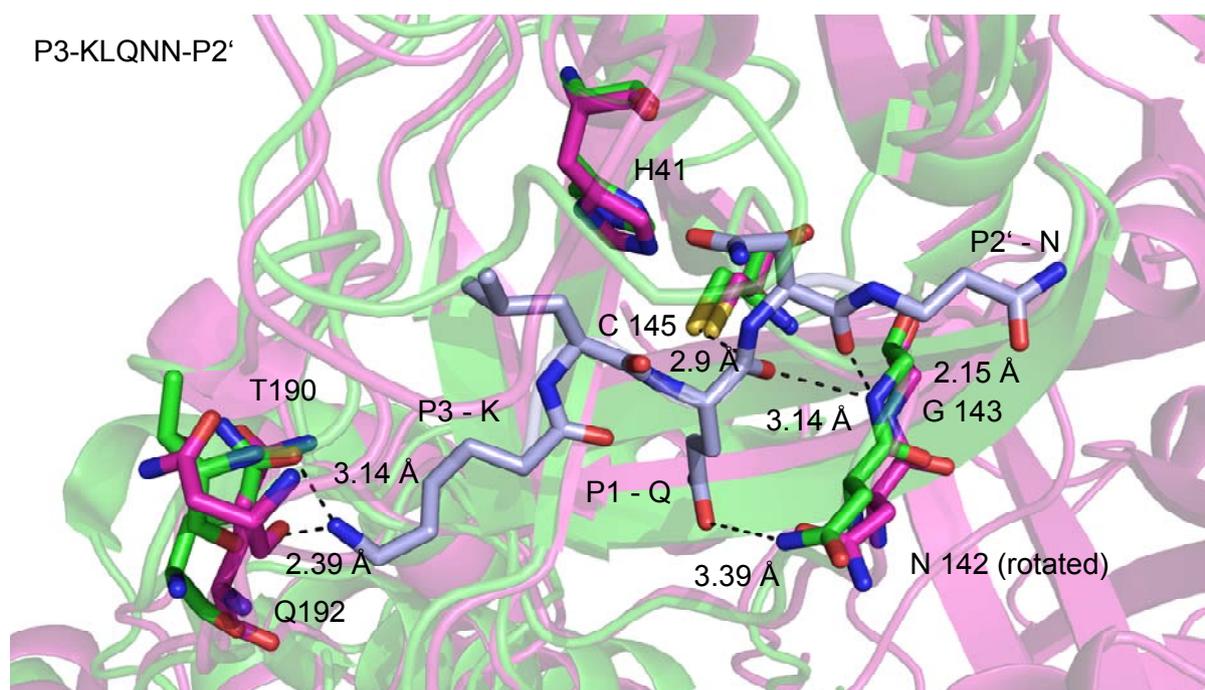


Figure 42. The interaction of P3-KLQNN-P2' and related peptides with M^{pro} is mainly stabilized by hydrogen bonds. The image displays hydrogen bonds of SARS-CoV M^{pro} (pdb code 3SND; *green*) and the peptide superposed with HCoV-229E M^{pro} (pdb code 2ZU2, *pink*). [N142 of SARS-CoV M^{pro} was rotated around the C α -C β and the C β -C γ bond to reduce the distance of the hydrogen bond with the P1-glutamine]

Table 20. Overview of pentapeptides docked to the active site of SARS-CoV M^{pro} and binding energies calculated by *in silico* docking.

peptide (location in the pp710 sequence)	binding energy
P3- KLQNN -P2' (Nsp8-Nsp9)	-1222.9 kcal mol ⁻¹
P3- KLANN -P2' (Nsp8-Nsp9)	-1032.7 kcal mol ⁻¹
P3- ILQSV -P2' (Nsp7-Nsp8)	-1088.6 kcal mol ⁻¹
P3- ILASV -P2' (Nsp7-Nsp8)	-964.3 kcal mol ⁻¹
P3- ILNSV -P2' (Nsp7-Nsp8)	-996.6 kcal mol ⁻¹

The calculated binding energy of -1223 kcal mol⁻¹ for the ligand P3-KLQNN-P2' is primarily based on hydrogen bonds between the ligand's lysine and M^{pro} T190 and Q192 as well as ligand's serine/asparagine with M^{pro} G143 (fig. 42). Replacing glutamine by alanine destroys a hydrogen bond involving SARS-CoV M^{pro} N142 that specifically positioned the substrate into the active centre. This is reflected by a loss in binding energy of almost 200 kcal mol⁻¹. Substitution of glutamine by asparagine established a comparable hydrogen bond with G143, therefore limiting the loss in binding energy (table 20). The substrate stabilization due to P1-glutamine is supported by the high-resolution structure of SARS-CoV M^{pro} lacking the catalytic histidine in complex with a cleavable substrate (pdb code 2Q6G; [75]). In contrast to a small serine of the ligand in this protease high-resolution structure,

the asparagine in position P1' of P3- KLQNN -P2' is facing out of the pocket with its amide group. This agrees with the observation that, apart substitutions with proline and acidic amino acids that are disallowed due to electrostatic repulsion, all other amino acids in P1 position allowed a certain degree of cleavage [40; 65]. However, in the absence of suitable reactive groups, potential inhibitors of this type do not yet allow to target the catalytic cysteine by a covalent bond, e.g. by a thiohemiacetal involved in inhibition by aldehydes. The performed docking approaches provided insights into the potential to target coronaviral main proteases by nearly native substrates involving the S1' and S2' site, in addition to the other subsites already targeted by the comparable peptide aldehyde inhibitors [78].

The proposed binding affinity and the inhibitory effect of the peptides remained to be verified *in vitro*, which may additionally contribute to the understanding of the specificity of the polyprotein maturation process. Therefore, a number of peptides were synthesized, according to table 21. Peptide V comprised the N-terminal autocleavage site of the main protease, which was included as a control peptide that is known to be cleaved with a high catalytic efficiency [192].

Table 21. Synthetic peptides under investigation, derived from pp1a of HCoV-229E. Peptides were acetylated at the N-terminus and unmodified at the C-terminus (purity > 95%, obtained from Genescript).

	peptide	molecular weight [Da]
I	TVRLNAGKQ (Nsp9-Nsp10)	1027.2
II	RTAINSFND (Nsp10-Nsp11)	1078.1
III	DSILNSVAS (Nsp7-Nsp8)	946.0
IV	VVKLNNEI (Nsp8-Nsp9)	1083.2
V	SAVLQSGFRK (Nsp4-Nsp5)	1134.2

The secondary structure of each peptide was investigated by CD spectroscopy in the far-UV wavelength range (192-260 nm) in buffer, water and in the presence of 2,2,2-trifluoroethanol (TFE) (fig. 43A), since the structure of a cleavage site loop could potentially influence polyprotein cleavage. The spectra of all peptides corresponding to the different cleavage sites displayed a random coil “fold” characterized by a strong ellipticity minimum below 200 nm and no positive ellipticity in the entire spectrum. Only the CD spectrum of peptide IV displayed a significantly different shape and intensity of the minimum, probably corresponding to a β -turn. This peptide displayed a low binding affinity to the main protease in SPR spectroscopy (fig. 44A) and the respective cleavage site between Nsp8 and Nsp9 is cleaved relatively slow (chapter 3.4), which could be influenced by the local secondary structure. Moreover, even 18% (v/v) TFE failed to induce another secondary structure of the peptides I and III. Therefore, although a limited number of cleavage-site-derived peptides were tested, the potential inhibition by the peptides and the cleavage of the corresponding native sites may not be influenced by the secondary structure, at least for peptides I-III. A significant effect of α -helix and β -

3. Results and discussion

sheet propensities of a SARS-CoV main protease substrate was reported previously, at some amino acid positions β -sheet propensities increase the turnover rate [191; 192; 199]. However, it has to be mentioned that the structural properties within a full-length polyprotein, including additional non-covalent interactions, might influence the structure of the cleavage site loop in a different way compared to the isolated peptides.

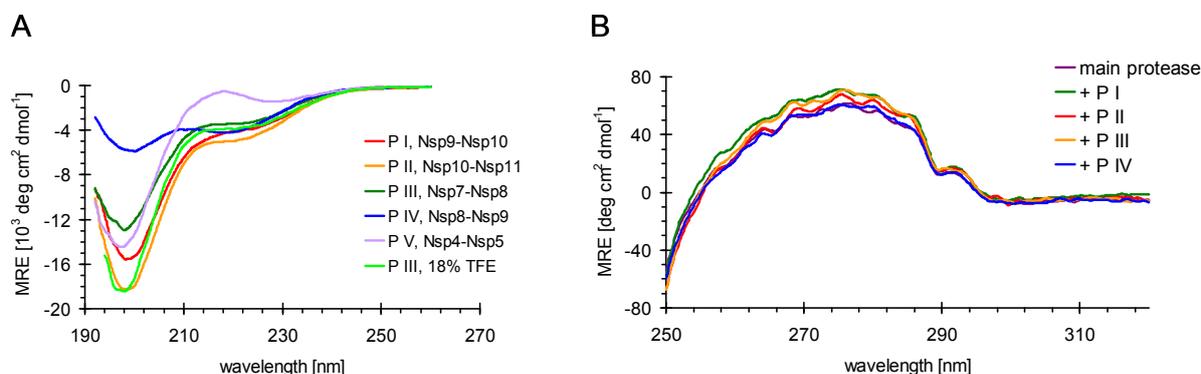


Figure 43. (A) Far-UV CD spectra of synthetic peptides derived from M^{pro} cleavage sites of HCoV-229E pp1a, as listed in table 21. Measurements have been performed in 50 mM NaCl and 8 mM sodium phosphate buffer, pH 7.6. Peptides I-III and V displayed a comparable spectrum characteristic for a random coil structure, while peptide IV was significantly different at lower wavelengths, (B) Near-UV spectra of M^{pro} incubated with a 10-fold molar excess of the respective peptide confirmed that the (tertiary) structure of M^{pro}, particularly the aromatic amino acids, was not affected upon peptide binding.

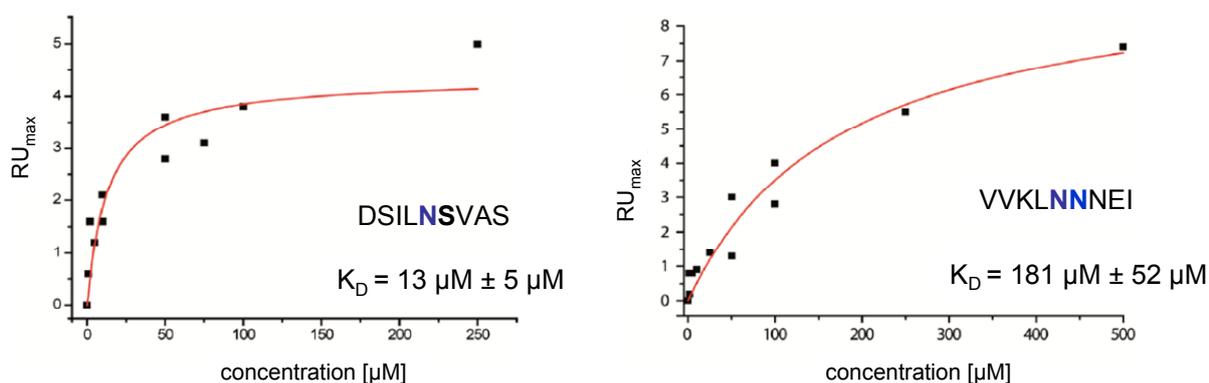


Figure 44. SPR spectroscopy experiment investigating the interaction of SARS-CoV M^{pro} with two inhibitory peptides. The dependence of the maximum resonance on the concentration of the respective peptide is plotted.

To verify that binding of the peptides did not trigger a significant effect on the tertiary structure of the protease, near-UV CD-spectra of 150 μM SARS-CoV M^{pro} were additionally recorded in the presence of a 7-fold molar excess of each peptide (fig. 43B). Beside the intensity of the maximum ellipticity caused by the aromatic residues, the peaks were not significantly shifted (e. g. by

quenching), indicating a conserved tertiary structure upon binding. This observation was basically confirmed by SAXS measurements. The R_g of the M^{pro} was only marginally reduced in the presence of the peptides (data not shown). The C-terminal his-tag was cleaved off prior to the experiment resulting in authentic termini of the protease. Moreover, it was not possible to significantly alter either the monomer-dimer equilibrium or the pI of M^{pro} in the presence of a 10-fold molar excess of peptide (data not shown).

Inspired by the observation of Wu *et al.* [132] that the presence of an excess of the artificial substrate TSAVLQ-pNA can trigger the dimerization of the monomeric M^{pro} mutant R298A the peptide VVKLNNEI (IV, table 21), representing the cleavage site connecting HCoV-229E Nsp8 and Nsp9, was applied to analytical size-exclusion chromatography. According to SEC (fig. 45), the amount of dimeric mutated main protease was slightly increased in the presence of a 5-fold molar excess of peptide, similar to the effect of a 2-fold molar excess of TSAVLQ-pNA. However, M^{pro} R298A appeared to be still partly monomeric – for the chosen experimental conditions.

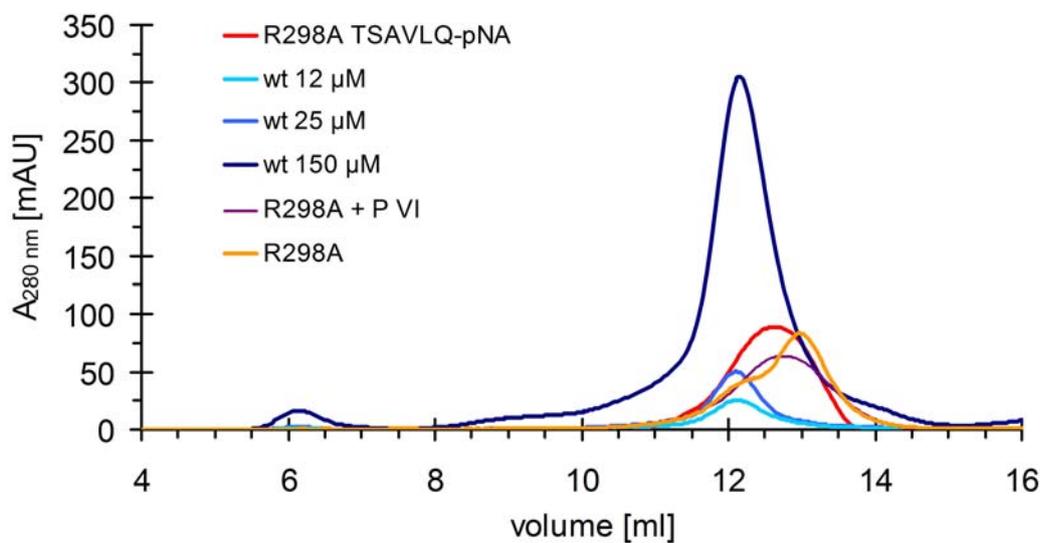


Figure 45. Dimerization of SARS-CoV M^{pro} . M^{pro} wt applied to a Superose 12 10/300 GL SEC column consistently displayed a dimeric state in a concentrations range from 12 to 150 μM (different shades of *blue*; 100 μl each). In contrast, M^{pro} R298A displayed a predominantly monomeric state (*yellow*). Incubating the mutant with a 6-fold molar excess of peptide VVKLNNEI for 1 h at 4 $^{\circ}\text{C}$ resulted in the *purple* chromatogram, while a comparable incubation of M^{pro} with a two-fold molar excess of the synthetic cleavable peptide TSAVLQ-pNA yielded a similar chromatogram (*red*). [Chromatography buffer: 40 mM NaH_2PO_4 , 250 mM NaCl, pH 7.8.]

3. Results and discussion

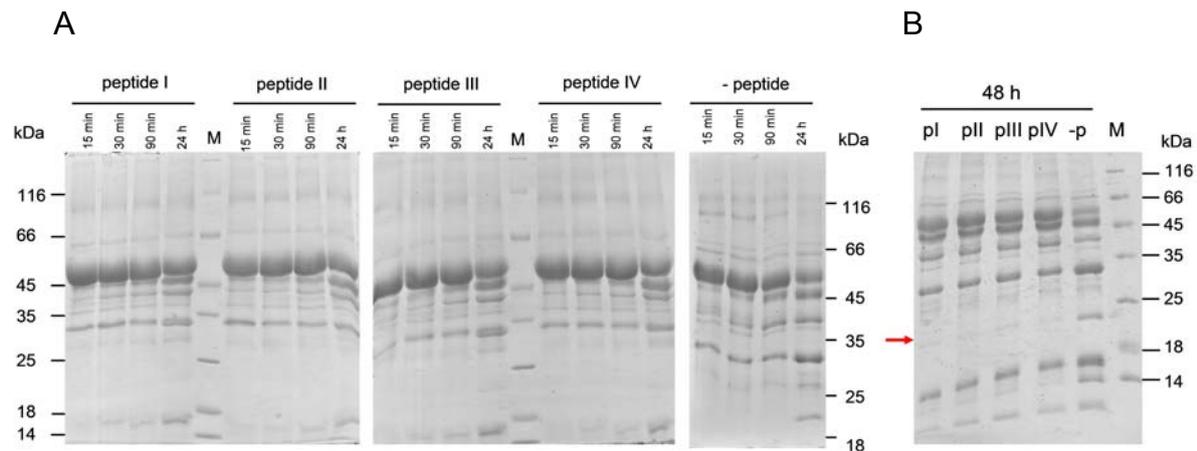


Figure 46. (A) Processing of native HCoV-229E pp710 by SARS-CoV M^{pro} was inhibited by four different peptides as demonstrated by SDS-PAGE. Pp710 was cleaved by M^{pro} in the presence of a 9-fold molar excess of peptide I to IV and 1 mM DTT up to 24 h. As a control, the same reaction was performed in the absence of peptides (-peptide). (B) The amount of all four mature Nsps was reduced after 48 h incubation at 6 °C (red arrow points at Nsp8). The lane “-p” represents an inhibitor-free control. The inhibitory effect of each peptide is obvious at this ratio, e.g. by the strongly reduced amount of mature Nsp8 (21 kDa, red arrow)

The impact of peptide I-IV on the proteolytic processing pathway of HCoV-229E pp710 was analyzed by SDS-PAGE (fig. 46), consistently revealing a remarkable inhibition of SARS-CoV M^{pro}, indicated by the reduced release of mature Nsps. Particularly the release of Nsp8 (21 kDa) and Nsp9 (12 kDa) was obviously reduced (fig. 46, red arrow). The extend of inhibition at a 9-fold excess is in a similar range for all peptides, comprising a starting point for further optimized protease inhibitors. The affinity of the peptide inhibitors described in this study is still in the range of a native substrate. Applying TSAVLQ-pNA as an artificial substrate for a colorimetric quantification of the M^{pro} activity in the presence of the peptides, no significant inhibitory effect was observed. This could be attributed to the fact that the expected dissociation constants of the substrate-like peptides were not increased compared to that of the substrate in the assay, which was reported to be in the low μ M range for mature SARS-CoV M^{pro} [200].

Furthermore, no inhibition was observed supplementing the protease with either a similar excess of *Onchocerca volvulus* onchocystatin (OV7; Dr. R. Eberle, University Hamburg) or with the peptide pEQW that was identified as a broad spectrum protease inhibitor in snake venom (Dr. A. Munawar, University Hamburg). These conclusions were drawn based on SDS-PAGE analysis of the cleavage reaction.

The specific binding affinity of the peptides I, III and IV towards SARS-CoV M^{pro} was determined applying surface plasmon resonance spectroscopy. Remarkably, an approximately 10-fold lower binding affinity was revealed for peptide IV containing an asparagine in P1' position (fig. 44B) compared to peptide III (DSILNSVAS (Nsp7-Nsp8) (fig 44A). The determined K_D -value of peptide I

(TVRLNAGKQ; $K_D = 16 \mu\text{M}$; $\text{RU}_{\text{max}} = 7 \text{RU}$; not shown) was in the same range as that of peptide III. This observation differed from the binding energy calculated for shorter peptides (5 amino acids), indicating that the docking of ILQSV may not well reflect the native conditions, resulting in an underestimated binding energy, even though binding energy does not necessarily correlate with catalytic activity in general. A significant increase in binding energy by flanking residues of ILQSV is rather improbable. A weak specificity in position P4 [193] occupied by serine and valine, respectively, is not expected to have a significant impact on the binding energy in this case, comparing the docking and the SPR spectroscopy approach. However, the reduced affinity of peptide IV correlated well with the observation that the cleavage site connecting Nsp8 and Nsp9 is poorly cleaved (see fig. 26B, 27A). Nevertheless, all K_D -values were in the μM range and SDS-PAGE displayed a similar inhibitory effect of all four peptides (fig. 46).

For comparison and further validation, the binding affinity of peptides I-IV to the SARS-CoV M^{pro} was additionally analyzed by native mass spectrometry in cooperation with Dr. C. *Utrecht* at the University of Utrecht (fig. 47). The affinity of peptide IV was almost too low to be detectable. The displayed spectra allowed a relative comparison of the peptide affinity, largely reflecting the order of K_D -values obtained by SPR spectroscopy. Peptide DSILNSVAS displayed the lowest K_D -value, significantly lower than the peptide IV with asparagine in P1' position [VVKLNNNEI (Nsp8-Nsp9)]. Differing from SPR spectroscopy, the peptide TVRLNAGKQ (Nsp9-Nsp10) seemed to bind significantly weaker than peptide III [DSILNSVAS (Nsp7-Nsp8)] in mass spectrometry experiments, instead of exhibiting a comparable affinity (fig.46, SPR spectroscopy). This could be related to the ESI process in mass spectrometry. Remarkably, an interaction of monomeric and dimeric protease with peptide molecules is indicated.

Interestingly, the affinity of RTAINSFDN (Nsp10-Nsp11) was significantly higher than the affinity of peptide IV. Considering that this affinity correlates with a moderately high turn-over rate at the Nsp10-Nsp11 cleavage site (see also table 16), it might be important for the viral replication to cleave off Nsp11. Nevertheless, the short Nsp11 peptide is not conserved in different coronavirus genomes and a high-resolution structure of SARS-CoV Nsp10 attached to Nsp11 did not reveal any ordered structure of Nsp11 [131]. Furthermore, there is no indication yet that Nsp11 could interfere with an interaction of Nsp10 and another protein.

3. Results and discussion

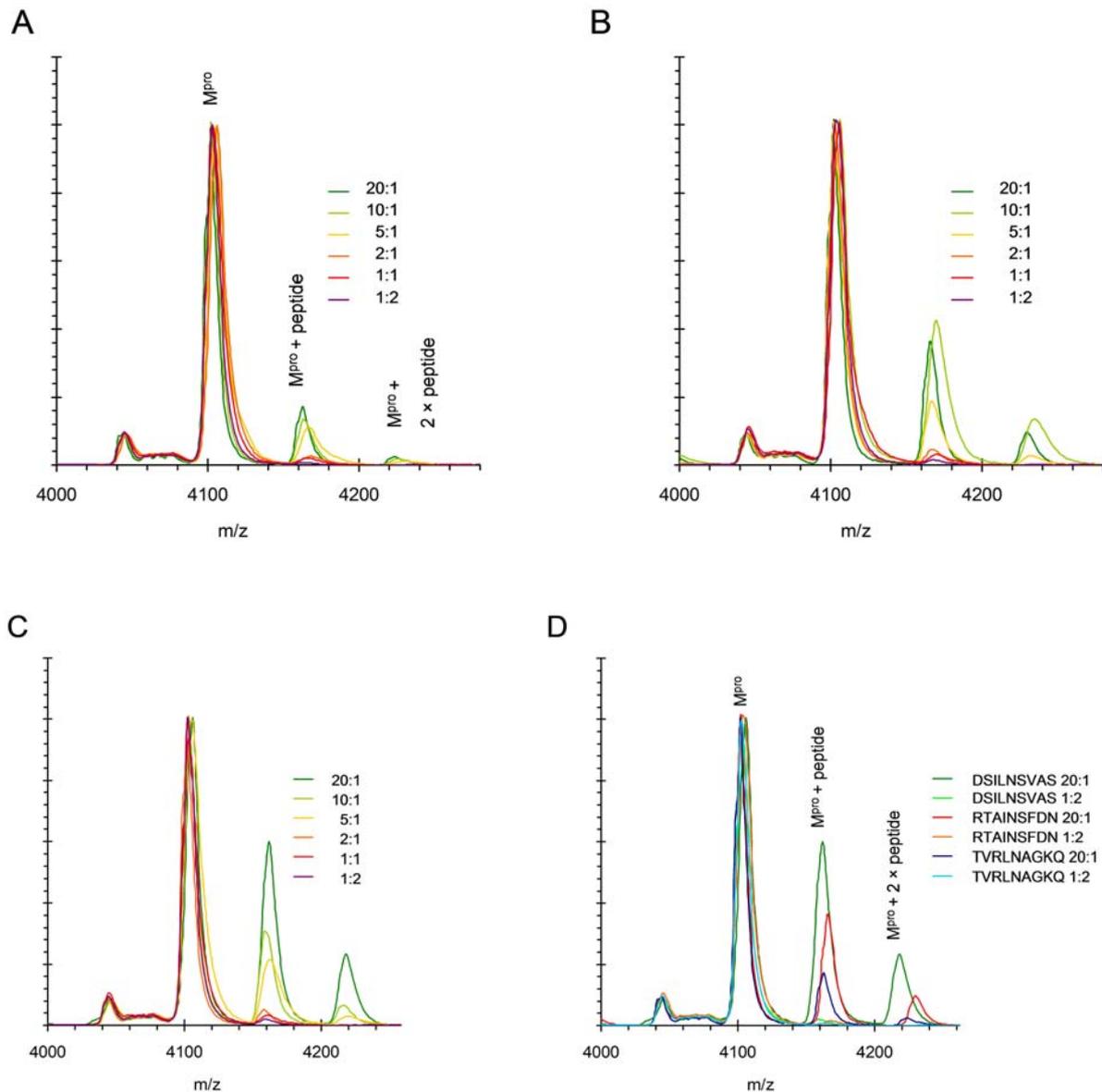


Figure 47. Native ESI-TOF MS experiment (LCT device, Waters) to quantify the affinity of substrate-like inhibitory peptides to SARS-CoV M^{pro}: Peptide and enzyme were incubated for 30 min at 20 °C in the presence of 1 mM DTT and 250 mM ammonium acetate. A peak corresponding to dimeric main protease without peptide ("M^{pro}") is shown for every peptide and concentration ratio, together with a shifted peak originating from a main protease dimer-peptide complex ("M^{pro} + peptide"). The intensity of the latter peak relative to the other M^{pro} peak allows a relative comparison of the affinity of the different peptides. Increased intensity of the "M^{pro} + peptide" peak indicates an increased affinity. Notably, the binding mode is not suggested to be cooperative and, presuming that the signal at around 4220 m/z as part of the respective ion series is not an artefact, both active sites within a M^{pro} dimer can be occupied at the same time ("M^{pro} + 2 x peptide peak"). For all peptides, the relative differences of the binding affinities were verified by other m/z-states, including the ion series corresponding to monomeric main protease binding to one peptide. (A) peptide I (TVRLNAGKQ); (B) peptide II (RTAINSF DN); (C) peptide III (DSILNSVAS); (D) comparison of peptides I-III.

A BLAST (NCBI) database search for similar inhibitory or cleavable peptide sequences identified fragments of human transmembrane receptors. Considering that binding of M^{pro} to a host-cell protein might take place during the replication cycle, the best hits are listed in tab. 22, although the abundance of main protease recognition sites in proteins of non-viral species is considerably low.

Table 22. Selected examples of protein fragments identified by a BLAST search with the highest similarity to the autocleavage site of SARS-CoV M^{pro} and to the inhibitory peptide DSILNSVAS (owing an asparagine in P1 position), respectively.

inhibitory peptide/ cleavage site	peptide match	protein [organism]	identity
DSILNSVAS	SILNSVAS	olfactory receptor [<i>Homo sapiens</i>]	8/8
	SILNSVAS	seven transmembrane helix receptor [<i>Homo sapiens</i>]	8/8
	DSILGSVAS	GK14735 [<i>Drosophila willistoni</i>]	8/9
SAVLQSGFRK	AVLQSGYRK NetCorona score 0.822*	β -galactosidase [<i>Aspergillus kawachii</i>]	8/9
-	LANLQSTYDK ¹ NetCorona score 0.539*	β -1,4-galactanase from <i>Aspergillus aculeatus</i> (pdb code 1FHL)	-

* (the threshold for a sequence considered to be cleavable is > 0.5)

¹ forms an α -helix in the high-resolution structure, in contrast to the inhibitory peptides investigated in this study, which could affect cleavage [191]

In summary, inhibition of proteolytic processing of a native full-length polyprotein by four peptides designed in this study was demonstrated. The dissociation constants were in the moderately low μ M range, comprising a novel aspect and perspective for coronavirus main protease inhibitor development. A considerably different inhibitor of the SARS-CoV main protease, the plant flavonoid epigallocatechin gallate is investigated in the following chapter, which was applied to co-crystallization trials along with the peptides described in chapter 3.7.

3.6. Interaction of M^{pro} with flavonoid derivatives

The inhibition of SARS-CoV M^{pro} by various plant flavonoids was already reported several times [68; 201] (see also introduction, part 1.3). Nevertheless, besides *in silico* predictions of the binding epitopes, there is still a lack of experimental structural data. Therefore, STD-NMR spectroscopy in parallel to co-crystallization trials was performed to characterize the corresponding binding epitope. STD-NMR spectroscopy allows determining the epitope of a small ligand binding to a comparably large protein. Upon an electromagnetic radio-frequency excitation pulse to the protein the saturation of the nucleus spins will spread to a bound ligand, indicating a close proximity of the respective part of the ligand to the protein.

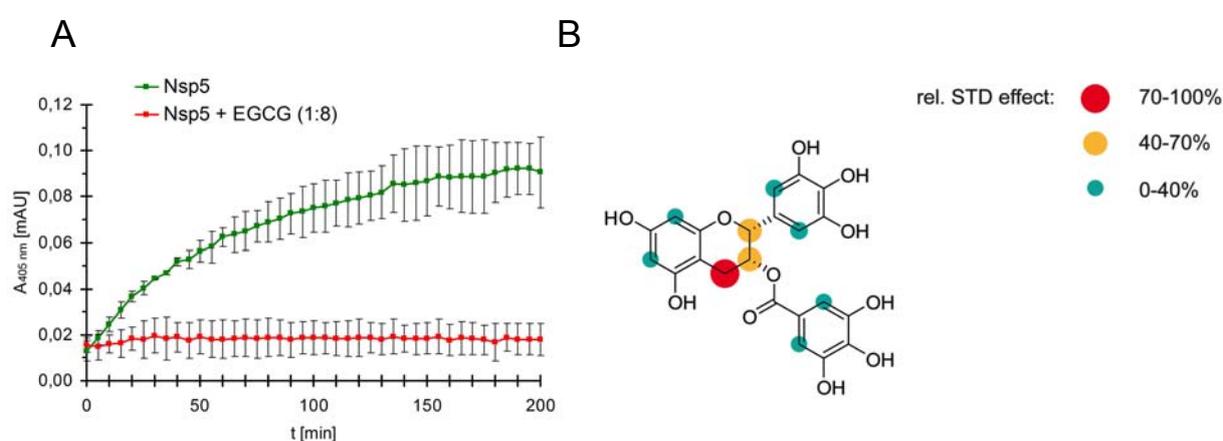


Figure 48. (A) Inhibition of SARS-CoV M^{pro} by epigallocatechin gallate was verified photometrically using the chromogenic substrate TSAVLQ-pNA (C-terminal p-nitroanilide, MW = 755.8 Da). Triplicate measurements have been performed with main protease (*green*) and main protease mixed with EGCG at a molar ratio of 1:8 (*red*). (B) The interaction of M^{pro} with epigallocatechin gallate was investigated by STD-NMR spectroscopy. The protein was freshly prepared and supplemented with a 100-fold molar excess of ligand. The ligand epitope was investigated immediately to prevent a significant amount of keto-enol tautomerism. The NMR sample was supplemented with 2 mM deuterated DTT. The relative STD effect is indicated by a 3-level colour code, where a high relative STD effect indicated close proximity to the protein surface. The highest STD effect was set to 100%.

Initially, the inhibition of SARS-CoV M^{pro} by epigallocatechin gallate (EGCG) was verified photometrically in three independent experiments. An 8-fold molar amount of EGCG almost completely abolished the catalytic activity of M^{pro} (fig. 48A, red data points). The ligand epitope of EGCG was investigated by STD NMR spectroscopy, indicating that the oxane ring is located very close (and oriented parallel) to the protease surface. Therefore, the geometry of the benzoxane moiety is considered to be critical for the level of interaction. Aiming at correlating structure and inhibitory effect of flavonoids, the binding mode of the diverging quercetin-3- β -galactoside was predicted *in*

silico [202]. Despite the different sterical properties of the galactoside moiety compared to the aromatic ring, the orientation of the ligand appears to be similar compared to the binding epitope of EGCG. Moreover, even though the area around the poly-hydroxylated six-rings is not perfectly covered by STD-NMR spectroscopy-relevant hydrogen atoms, the experimentally determined epitope essentially agrees with the epitope of the related gallocatechin gallate as predicted *in silico* [67]

Since other phenolic plant compounds were proven to harbour potential to inhibit SARS-CoV M^{pro} [203; 204], it is worth to further investigate more and particularly larger compounds of the diverse secondary metabolism of higher plants. A docking approach with amentoflavone which is a biflavonoid indicated a significantly different epitope, occupying more space in the active centre cleft. The IC₅₀-value of this compound is already as low as 8 μM [201]. Another flavonoid, tannic acid that has a huge inhibitory potential (IC₅₀ = 3 μM) [68] consists of five galloylester moieties surrounding a glucose ring. The structure of tannic acid suggests that specific binding to more than one active centre might be simultaneously possible, which could result in a chain-like oligomerization of M^{pro}. At least different epitopes within this tannic acid molecule are conceivable. However, investigations by DLS did not reveal a specific effect of tannic acid on the oligomerization of M^{pro}, but indicated that tannic acid induced a continuous unspecific oligomerization of the protease already at equimolar amounts, preventing further crystallization trials. Native ESI-TOF mass spectrometry did not reveal an altered oligomerization of the protease.

3.7. Co-crystallization trials with SARS-CoV M^{pro}

To analyze the binding epitope of epigallocatechin gallate (EGCG) in addition to the characterization by STD-NMR spectroscopy in more detail, co-crystallization trials were set up. SARS-CoV main protease mixed with a 2-fold molar amount of EGCG formed a dimer in a monodisperse solution, as verified by DLS (data not shown). Even though crystallization of SARS-CoV main protease was successful (fig. 49, particularly panel I), co-crystallization trials with EGCG were not successful, i.e. screening 300 conditions containing various PEGs as well as applying published conditions [78] and the counter diffusion technique (fig. 49I/J). Even a low excess of EGCG significantly impaired the formation of protein crystals. An optimization of the co-crystallization conditions by applying a continuous gradient of PEG 6000 precipitant solution in a Granada Crystallization Box[®] (GCB) did not result in the formation of diffraction quality crystals (fig. 49J). Soaking of main protease crystals to a final concentration of 1 mM EGCG in the drop worked fine and did not significantly affect the diffraction quality of the M^{pro} crystals. Additionally, SARS-CoV M^{pro} crystals were soaked with the substrate-like peptides listed in table 21 to obtain high-resolution structures of the respective complex (see fig. 49 for a summary of experimental conditions).

Nevertheless, all diffraction data sets that were collected and processed so far did not reveal

3. Results and discussion

sufficient electron density of the respective ligand in one of the active centres of the M^{pro} dimer. The highest resolution obtained for an M^{pro} crystal diffraction dataset was 1.6 Å. Datasets were successfully indexed revealing either a P1, C121 or P2₁2₁2 space group.

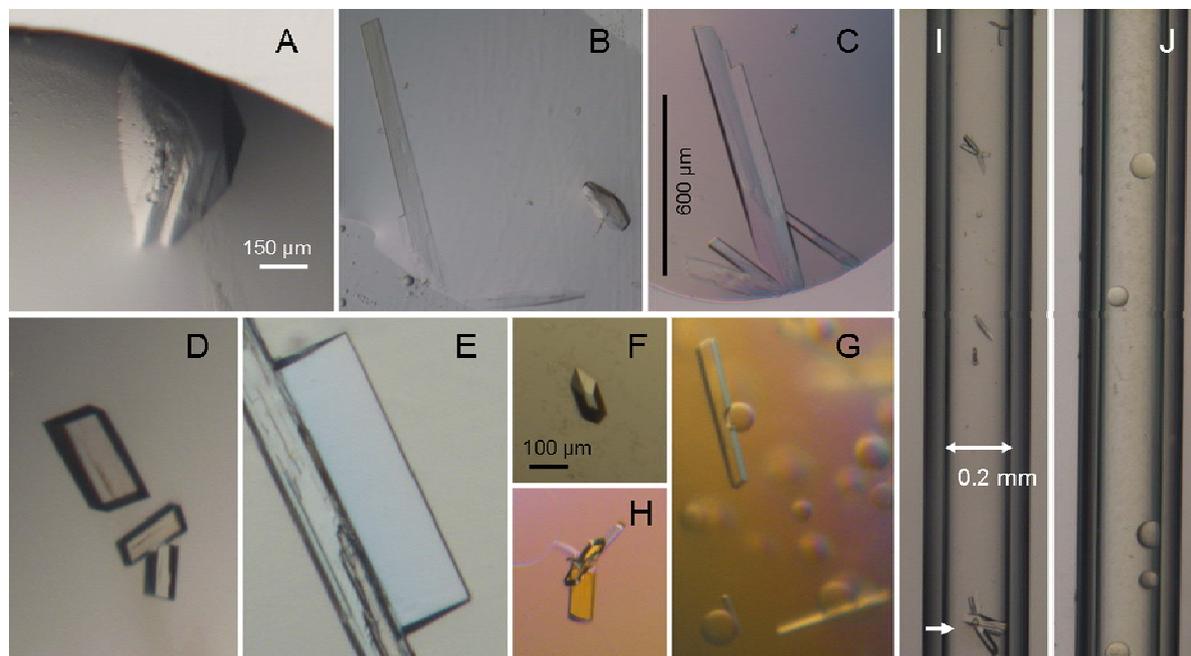


Figure 49. SARS-CoV M^{pro} co-crystallization trials. A-G: Vapour-diffusion experiment in a linbro 24-well plate; if not stated otherwise 10 g l⁻¹ M^{pro} in 1 mM DTT, 20 mM NaCl and 30 mM Tris pH 7.5 were mixed with an equal volume of reservoir solution, 2 μ l drops were equilibrated against 1ml reservoir precipitant solution each. (A) Reservoir: 8% PEG 6,000, 0.1 M MES pH 6.0, 3% DMSO, 1% MPD, 1 mM EDTA. Cube-shaped crystal soaked with 1 mM EGCG over night, diffracted at beamline X13 (DORIS) to a resolution of up to 2.2 Å. (B) 8 g l⁻¹ M^{pro} mixed with an 8-fold amount of peptide III; reservoir: 5.5% PEG 6000, 0.1 M MES pH 6.0, 2% MPD; two different crystal morphologies. (C) 8 g l⁻¹ M^{pro} mixed with an 7-fold amount of peptide II; reservoir: 5% PEG 6000, 0.1 M MES pH 6.0, 2% MPD, 1 mM EDTA, 2 mM DTT. (D) Main protease crystal soaked with a final concentration of 500 μ M EGCG for 5 h, diffracted at beamline X13 (DORIS) up to a resolution of 2.8 Å; reservoir: 6% PEG 6,000, 0.1 M MES pH 6.0, 3% DMSO. (E) Brick-shaped crystal (approx. 350 μ m in length) diffracted to a resolution of up to 1.6 Å at beamline P14 (PETRA III). (F) 6 g l⁻¹ M^{pro} 200 mM NaCl in the protein buffer; reservoir 0.8 M sodium malonate, 0.1 M MES pH 6.0, 2% 2-propanol. (G) 7 g l⁻¹ M^{pro} ; 200 mM NaCl in the protein buffer; reservoir 0.8 M sodium malonate, 0.1 M MES pH 6.0, 0.5% 2-propanol mixed with a 10-fold amount of peptide III. (H) 6 g l⁻¹ M^{pro} ; 100 mM NaCl in the protein buffer; reservoir 0.8 M sodium malonate, 0.1 M MES pH 6.0, 1% 2-propanol; mixed with a 10-fold amount of peptide II. (I) 180 μ M purified monodisperse M^{pro} (180 μ M) was drawn into a capillary that was punched into a Granada Crystallization Box[®] (GCB) sealed with agarose. The displayed GCB was filled with 10% PEG 6000, 0.1 M MES pH 6.0, 2% MPD, 2% DMSO and 1 mM DTT and incubated to allow crystallization by counter-diffusion. (J) Aiming co-crystallization a two-fold molar excess of EGCG was mixed with 180 μ M purified monodisperse M^{pro} and drawn into a another capillary that was set up in a GCB essentially as described for panel I.

3.8. Purification and oligomerization of SARS-CoV pp79

3.8.1. Dimerization of SARS-CoV pp79

As previously presented, HCoV-229E pp79 Q83A/Q278A formed a highly aggregated state in the absence of main protease that prevented any structural characterization; presumably due to locally increased hydrophobicity at the mutation sites (section 3.5). Therefore, the wild type SARS-CoV pp79 gene was separately cloned into pASK-IBA plasmids by *BsaI* restriction and recombinantly expressed to enable a structural comparison of pp79 with its precursor SARS-CoV pp710 and its maturation product pp78 that is formed by cleaving of Nsp9. SARS-CoV pp79 fused with an N-terminal strep-tag was purified by affinity chromatography (fig. 50A). The expected secondary structure content of the protein was verified by CD spectroscopy (data not shown). Along with significant amounts of high molecular weight oligomers, a monomeric and a dimeric state of SARS-CoV pp79 were detected in solution and separated by SEC (fig. 50B). Therefore, SARS-CoV pp79 and SARS-CoV pp710 consistently shared these states, as previously observed (fig. 9B). The abundance of the pp79 monomer is only slightly increased compared to the dimer in 280 mM NaCl, 30 mM NaH₂PO₄, pH 7.9 (fig. 50B). The monodispersity of the monomeric state was additionally verified by DLS. A long-term DLS experiment (> 140 h) demonstrated a remarkably good stability of the monomeric state ($R_h = 4.5 \pm 0.1$ nm) that was incubated for 5 d at 6 °C prior to the DLS experiment (fig. 50C). After 40 h, the hydrodynamic radius started to slowly increase, indicating the formation of dimers and subsequently higher oligomers that continuously increased in size. Therefore, a reformation of the dimer was observed, although this was most likely attributed to an unspecific effect of storage than to the establishment of an equilibrium, as the hydrodynamic radius continuously increased. An explanation for the significantly increased stability of the pp79 monomer compared to the monomeric state of SARS-CoV pp710 (see chapters 3.2.2.2 and 3.2.2.1) could be the absence of 13 cysteine residues located in SARS-CoV Nsp10, which might be responsible to form aggregates via intermolecular disulfide bonds, as observed for purified HCoV-229E Nsp10 (see chapter 3.13).

To obtain a low-resolution structural model of pp79, particularly for comparison of the monomeric and the dimeric state of pp79 with that of pp710, SAXS measurements were performed. Monodisperse monomeric and dimeric pp79 containing an N-terminal strep-tag isolated by SEC were exposed to synchrotron radiation at DORIS storage ring beamline X33. Based on the scattering data recorded for different concentrations (1 - 4.5 g l⁻¹) and after extrapolation to infinite dilution, *ab initio* low-resolution models were calculated (fig. 51) to predict the dimerization mode. Significant parameters of monomeric and dimeric pp79 as determined by SAXS, including a quality estimate of the *ab initio* models are presented in table 25.

3. Results and discussion

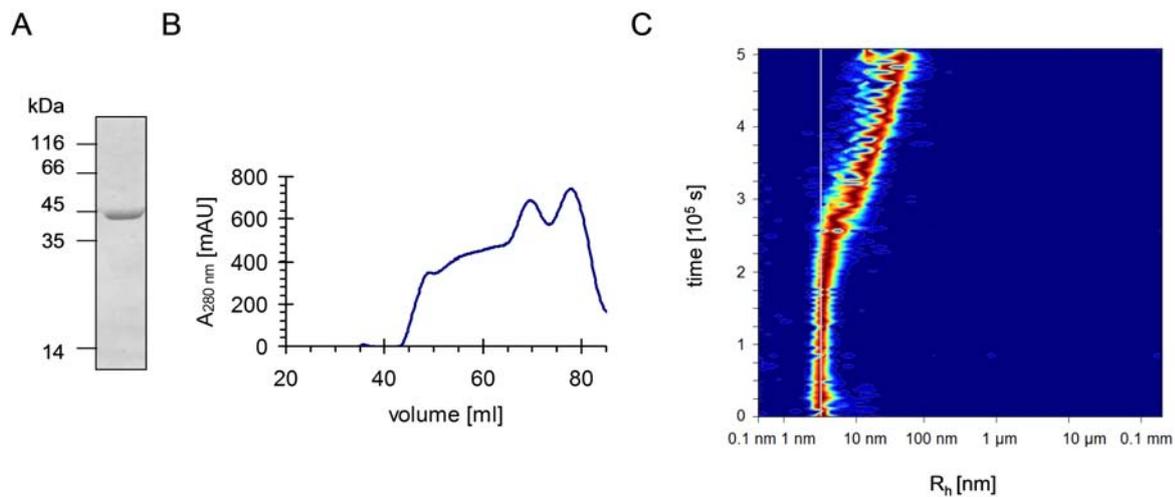


Figure 50. (A) N-terminally strep-tagged SARS-CoV pp79 purified by Strep-Tactin affinity chromatography is visualized by reducing SDS-PAGE. (B) Size-exclusion chromatography (HiLoad 16/60 Superdex 200) of the purified protein revealed a co-existence of different oligomeric states with the monomer that eluted at a volume of 78 ml in solution, corresponding to 54 kDa (expected $MW_{pp79 \text{ monomer}} = 44 \text{ kDa}$). (C) The monomeric state of pp79 isolated by SEC was ($R_h = 4.5 \pm 0.1 \text{ nm}$, grey line) was investigated by long-term DLS measurements. A 2 μl drop (60 μM protein in 280 mM NaCl, 40 mM NaH_2PO_4 , pH 7.9) in a 96-well Terazaki plate (NUNC) covered by paraffin oil was applied to record the DLS data and store the protein. After around 40 h of incubation at 20 $^\circ\text{C}$, the detected hydrodynamic radius indicated the formation of a significant amount of dimer, followed by subsequent strong aggregation. The particle abundance is colour-coded in the displayed heat map, ranging from *blue* (low) to *dark red* (high) abundance.

The obtained *ab initio* models of both states reasonably allowed predicting the dimerization site of two rigid monomers forming an elongated “V”-shaped dimer as displayed in fig. 51. As previously observed for pp710, it is well-conceivable that the dimerization of pp79 was mediated by Nsp8, orienting its C-terminal towards the dimerization interface (rigid-body model not shown). Moreover, assuming that Nsp10 does not provide additional non-covalent interactions with other domains within either SARS-CoV or HCoV-229E pp710, as investigated and discussed in chapter 3.4.1, it was reasonable to suggest a conserved dimer interface in pp79 and pp710, if Nsp10 does not sterically mask the interface in pp710. However, this appeared to be unlikely, based on the previously evaluated model of pp710 presented in chapter 3.2.3.

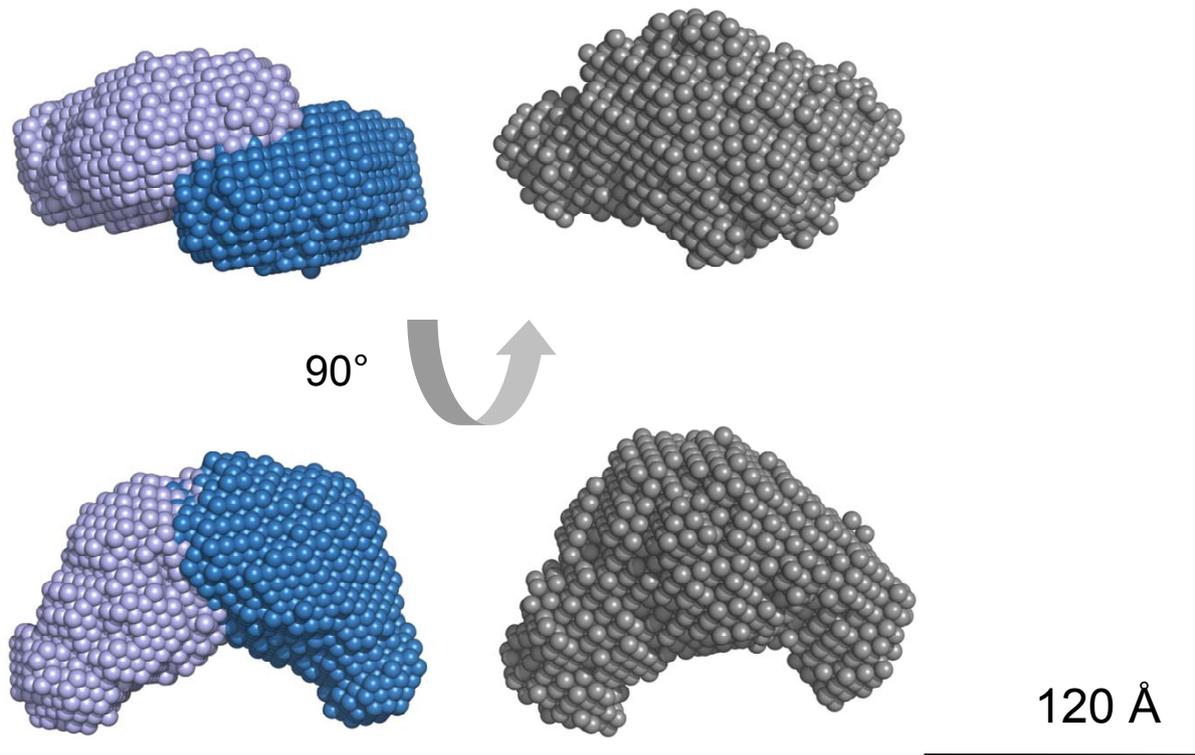


Figure 51. Prediction of the dimerization of SARS-CoV pp79 investigated by SAXS technique. Two copies of the average *ab initio* model of monomeric SARS-CoV pp79 are shown in blue and blue-grey, respectively. The predicted rigid assembly of the two monomers well interpreted the experimental *ab initio* shape of the pp79 dimer (grey). The dimerization involved the thickened end of the longitudinal model axis of the monomer to form a smooth “V”-like dimer shape.

3.8.2. Mutation of a GXXXG motif does not alter dimerization of SARS-CoV pp79 and pp710

SARS-CoV Nsp9 was reported to form a dimer via a ¹⁰⁰GXXXG¹⁰⁴ loop motif by a rigid “helix-groove” packing allowed by the small glycine residues located at the C-terminal helix of Nsp9 [102; 104]. The GX₃GX₂(G/A/S) motif is highly conserved in coronaviral Nsp9 and transmembrane helices of various proteins “packing” with similar motifs [205], e.g. the amyloid beta protein precursor [206]. Exchanging one of the two glycines by glutamate resulted in the formation of monomeric Nsp9 according to multi-angle light scattering tests [104]. The impact of this loop structure on the dimerization of SARS-CoV pp79 and pp710 was investigated by replacing G100 of the Nsp9 domain in both proteins by glutamate. The mutants were purified by Strep-Tactin affinity chromatography. The chromatogram obtained from SEC experiments demonstrated that the freshly prepared SARS-CoV pp710 G381E and pp79 G381E still formed a similar amount of dimer compared to the wild type protein (fig. 50A). Dimeric pp79 G381E isolated by SEC was applied to SAXS analysis for comparison with the wild type protein (fig. 50B). The superimposable scattering data at comparable protein concentrations strongly indicated that the intact GXXXG loop, which is crucial for

3. Results and discussion

dimerization of mature Nsp9 [104], is not relevant for dimer formation of SARS-CoV pp79, as for the SARS-CoV pp710 dimer formation as investigated by SAXS (chapter 3.2.3.2) as well. Furthermore, this result implied that also a mature Nsp9 monomer can not bind to pp79 via the GXXXG interface. Searching the sequence of coronaviral replicase polyproteins for another conserved GXXXG motif, which could be indicative for another interaction partner, had no result. However, such a motif was identified in the spike protein transmembrane domain, unique to SARS-CoV and stabilizing the trimer formation of the spike protein [207].

In addition to the GXXXG-dependent dimerization of Nsp9, β -sheet 5 (residues 63-69) of SARS-CoV Nsp9 was also reported to trigger dimerization, as indicated by a high-resolution crystal structure [101]. Nevertheless, mutation of the GXXXG motif introducing either glutamate or aspartate instead of glycine resulted in severe growth defects of the virus and was reported to be critical even for the formation of the RTC complex of SARS-CoV and IBV [104; 208], highlighting the importance of this motive *in vivo*. Therefore, it has to be assumed that this helix-helix dimerization is either much weaker in pp79 than another dimerization that was observed by SAXS or that the helix-helix dimer mediated by GXXXG is entirely formed by uncovering this interface after cleaving of Nsp9 by the main protease. The latter explanation is suggested to be more likely, since the mutation introduced into the GXXXG motif did not have any effect on the quaternary structure of pp79. SARS-CoV pp79 as well as SARS-CoV pp710 is not significantly impaired in forming higher oligomers. Therefore, larger oligomers appear to be formed by other protein-protein interfaces as well. Consequently, proteolytic processing between Nsp8 and Nsp9 in the precursor state is assumed to be essential for viral replication to allow the “helix-helix” dimerization of mature Nsp9. This suggestion is in agreement with the *in vivo* observation by Deming *et al.* [107] that disruption of the Nsp8-Nsp9 cleavage site by replacing P1-glutamine by alanine in MHV is lethal, because the Nsp9 dimerization is essential *in vivo* [104]. This phenotype was similarly observed for the mutation in P1 position in between of Nsp7 and Nsp8 – taking both alternative cleavage sites of MHV at this position (see table 17) into consideration – no syncytium formation or subgenomic mRNA was detected after implanting the mutated RNA by electroporation.

Interestingly, a comparable exchange of the essential P1 glutamine by alanine between either Nsp7 and Nsp8 or Nsp8 and Nsp9 of HCoV-229E pp710 and pp79 consistently induced aggregation as revealed in this study, well explained by locally increased hydrophobicity. Therefore, it is highly suggested to substitute the P1-glutamine by asparagine in future *in vivo* experiments to keep the local charge close to native, as it was investigated in chapter 3.5.

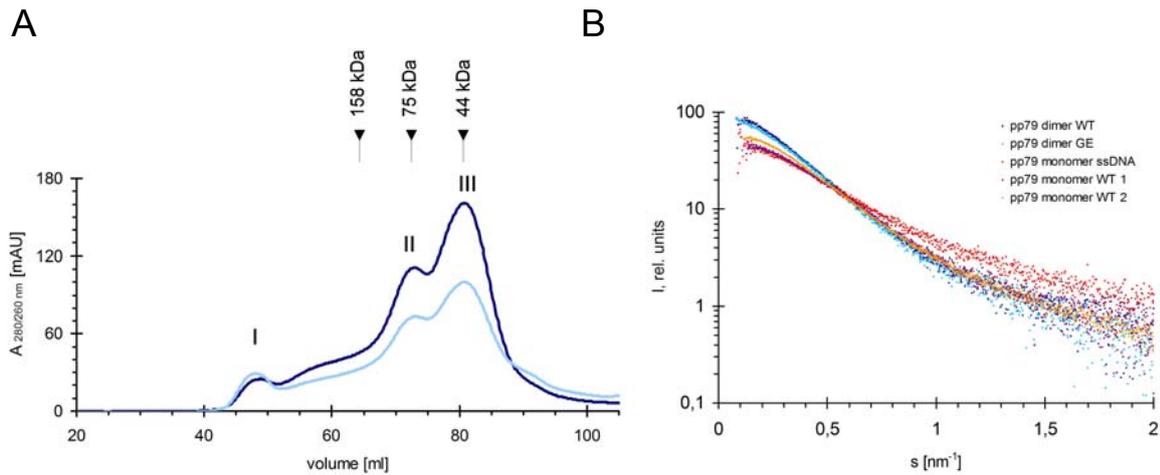


Figure 52. (A) Particle size distribution of pp79 G381E obtained from SEC. Absorbance was monitored at 280 nm (*dark blue*) and 260 nm (*light blue*). Freshly prepared pp79 G381E was supplemented with 1 mM TCEP to disassemble disulphide-linked covalent dimers. The SEC column was equilibrated with 200 mM NaCl, 30 mM sodium phosphate, pH 7.9. Standard protein masses are indicated. (B) Monomeric and dimeric state of pp79 G381E were investigated by SAXS. The scattering pattern $I(s)$ versus s of dimeric pp79 wt (*dark blue*) and pp79 G381E mutant (*cyan*) are well superimposable, indicating no change in size or shape of the molecules.

At similar buffer conditions, the addition of an equimolar amount of ssDNA (5'-GATTAGTTACCTGAC-3') and a final $MgCl_2$ concentration of 1 mM to SARS-CoV pp79 resulted in a slightly altered SAXS scattering curve compared to that of the pp79 monomer in solution (fig. 52B, *red* and *purple* pattern). While the progression at low angles was highly similar to that of monomeric Nsp9, increased scattering intensity has been detected at higher angles, indicating a slightly changed tertiary structure of pp79 in the presence of ssDNA, although the overall size corresponded to a monomeric molecule. However, the binding affinity of ssDNA could be significantly weakened by the relatively high ionic strength of the applied buffer conditions (200 mM NaCl). The alteration of the tertiary structure in the presence of nucleic acid is in agreement with conformational changes of Nsp8 upon RNA-binding already suggested by preliminary FRET results [96].

3.9. Polyprotein Nsp7-8 (pp78)

3.9.1. Oligomerization of HCoV-229E and SARS-CoV pp78

For structural comparison with pp79 and pp710, both homologues of pp78 were separately cloned and successfully purified from an *E. coli* cell lysate by affinity chromatography, exploiting either the interaction of an N- or C-terminal strep-tag fused to SARS-CoV pp78 with Strep-Tactin resin or of an N-terminal GST-tag fused to HCoV-229E pp78 with glutathione resin (fig. 53A, lane 1/2). The GST-tag of HCoV-229E pp78 was removed by *prescission* protease cleavage. Remaining impurities and the free GST-tag were separated by incubating the solution with glutathione resin again (fig. 53C). Purification of GST-tagged HCoV-229E pp78 from *E. coli* BL21 AI cells commonly resulted in a minor double-band of the target protein as verified by mass spectrometry. This effect was attributed to the artificial cleavage of some GST molecules from the fusion protein already in the *E. coli* cells.

Analytical SEC (HiLoad 16/60 Superdex 200) performed to initially identify the oligomeric states of both pp78 homologues in solution consistently revealed a predominant mixture of monomer and dimer of pp78 for both virus species (fig. 53A/B). For HCoV-229E, the monomeric state of pp78 was clearly preferred compared to the dimeric state at two different buffer conditions (fig. 53B, peak II), contrary to the ratio of SARS-CoV pp710 as previously revealed by SEC (fig. 9). Both states were additionally verified by native mass spectrometry (data not shown). High molecular weight oligomers have not been detected. A slightly smaller fragment of HCoV-229E pp78 observed by SDS-PAGE could falsify the ratio of monomer and dimer observed by SEC. In contrast, SARS-CoV pp78 with N-terminal strep-tag showed a prevalence for the dimer in solution, and significant amounts of large oligomers have additionally been detected (fig. 53B, peak II). Moreover, the oligomerization of SARS-CoV pp78 was significantly affected by the position of the purification tag, since pp78 with C-terminal strep-tag predominantly formed large aggregates, together with only small and not well separated fractions of dimeric and monomeric protein. The co-existence of a monomeric and a dimeric state of SARS-CoV pp78 with N-terminal strep-tag was verified by native blue PAGE as well (fig. 53C). The ratio of absorbance at wavelengths of 280 nm and 260 nm was constant for all peaks of both homologue proteins (fig. 53A/B), indicating that there is no significant enrichment of nucleic acid carried over from *E. coli* cells.

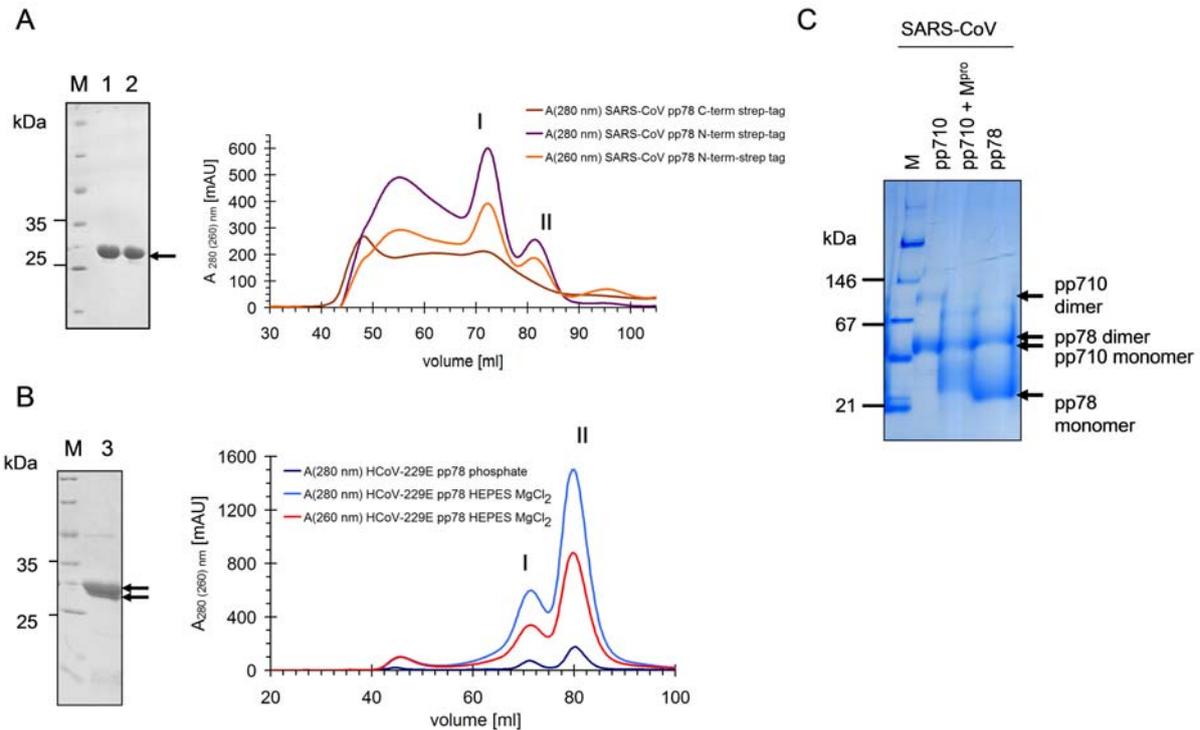


Figure 53. Purification and characterization of pp78 homologues in solution. **(A)** SDS-PAGE analysis (left) of SARS-CoV pp78 purification by affinity chromatography, exploiting either an N-terminal (lane 1) or C-terminal strep-tag (lane 2). The corresponding SEC chromatograms display a putative monomeric and dimeric state along with larger oligomers that are pre-dominant for C-terminally tagged protein. **(B)** The purity of HCoV-229E pp78 was verified by SDS-PAGE applied to SEC (HiLoad 16/60 Superdex 200) revealing two major oligomeric states (peak I and II), assumed to be dimer and monomer. **(C)** N-terminally strep-tagged SARS-CoV pp78 was investigated by blue native PAGE (right panel). SARS-CoV pp710 and a ten-fold amount of pp710 incubated with M^{Pro} for 30 min at 20 °C are shown for comparison. Monomeric pp710 roughly co-localizes with dimeric pp78. The existence of monomeric and dimeric states of pp78 was confirmed.

For comparison of folding stability, the melting temperature of HCoV-229E and SARS-CoV pp78 with N-terminal strep-tag was examined by CD spectroscopy in a temperature interval up to 80 °C, estimating the conservation of the overall sum of non-covalent interactions stabilizing the pp78-fold of an *Alpha*- and a *Betacoronavirus*. The thermal stability which is characterized by the turning point of an optimised Boltzmann fit (T_m , “melting point”), determined based on the ellipticity monitored at a wavelength of 220 nm, was identical within the error range for both homologue proteins determined at identical buffer conditions. For HCoV-229E pp78, a T_m of 61.8 ± 1.4 °C was obtained, while the T_m of SARS-CoV pp78 accounted to 61.3 ± 0.7 °C. For an initial comparison of the low-resolution structure, monomeric and dimeric state of SARS-CoV and HCoV-229E pp78 observed at higher concentration were isolated by SEC and applied to SAXS experiments. Protein concentrations ranged from 1 up to 4.5 g l^{-1} . SAXS results for the monomeric and the dimeric state of both pp78 homologues are presented in table 23 and 25, verifying the expected molecular weight of

3. Results and discussion

monomer and dimer. Low resolution structural models of the proteins calculated based on the scattering data using DAMMIF are discussed in detail in chapter 3.10.

Table 23. Summary of structural properties of pp78 obtained by SAXS and DLS

virus	protein	R _h [nm]	R _g [nm]	D _{max} [nm]
SARS-CoV	pp78 monomer	3.3 ± 0.3	3.20 ± 0.01	10.0
	pp78 dimer	5.3 ± 0.1	3.88 ± 0.01	13.0
HCoV-229E	pp78 monomer (GST-tag cleaved)	3.4 ± 0.4	2.65 ± 0.11	8.7
	pp78 dimer (GST-tag cleaved)	4.5 ± 0.2	3.37 ± 0.11	11.4

Exploiting the high mass accuracy of native mass spectrometry, minor amounts of tri- and tetrameric SARS-CoV pp78 were detected in ammonium acetate buffer at pH 7.9, co-existing with predominant mono- and dimeric states (fig. 54), as basically observed by SEC analysis (fig. 53 A). Larger oligomers were almost absent in the m/z spectra. The oligomeric states were specifically fragmented by CID, resulting in the detection of a mass of 32,354 Da. This species is in good agreement with the monomer mass detected at lower collision voltages (fig. 54, approx. 3,000-3,300 m/z) and the calculated mass of pp78 of 31.16 kDa + 1.48 kDa accounting to the strep-tag provided by the pASK-IBA5plus vector.

In conclusion, highly abundant mono- and dimeric states were identified for HCoV-229E and SARS-CoV pp78, even though the ratio is altered comparing the two different homologues of pp78. For SARS-CoV pp78 the aggregation is altered by the tag position. Moreover, R_g and maximum diameter of HCoV-229E pp78 are significantly lower compared to the homologue protein, as observed similarly for HCoV-229E and SARS-CoV pp710 (chapter 3.2.3.2), strongly indicating an altered tertiary structure. Most probably the effect is related to the tertiary structure of Nsp8, since structural variation of Nsp8 due to flexibility was observed in the three high-resolution structures of a mature coronavirus Nsp(7+8) complex [96; 98; 99].

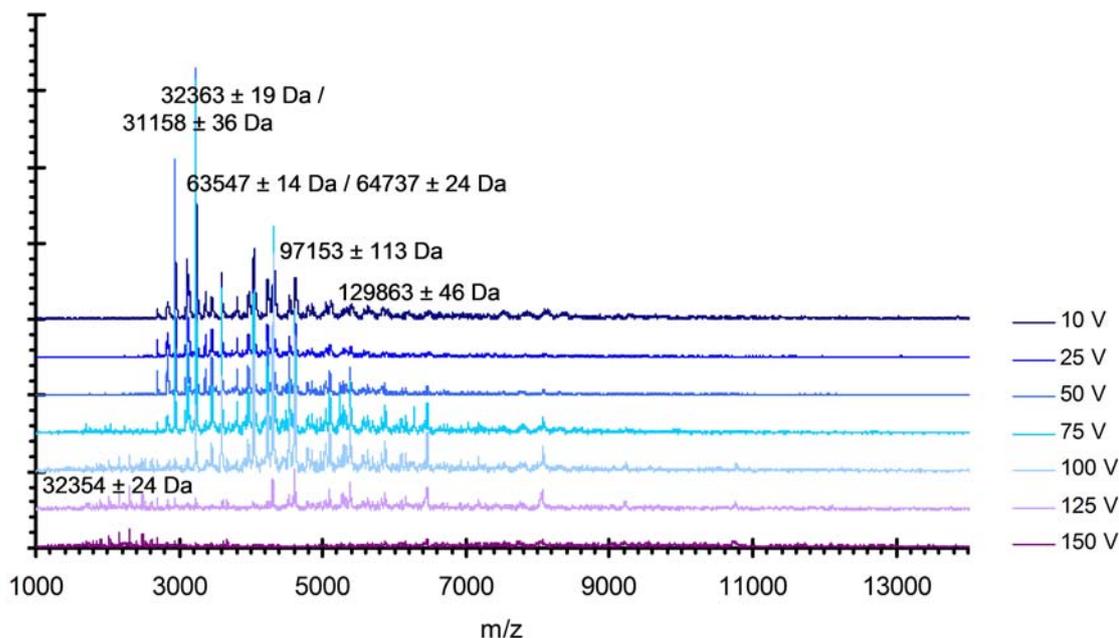


Figure 54. N-terminally strep-tagged pp78 of SARS-CoV investigated by native ESI-Q-TOF. A co-existence of the monomer with a defined dimeric, trimeric and tetrameric state was revealed in solution and additionally fragmented by CID. The abundance of the specific states significantly decreased with enhanced oligomerization. Moreover, increasing the collision voltage resulted in a reduced abundance of pp78 oligomers. Finally, at 150 V collision voltage, only monomeric pp78 remained (32,292 Da/32,382 Da), corresponding to peaks below 3,000 m/z-value that are already visible at lower voltage (32,354 Da at 125 V).

3.9.2. SARS-CoV pp78 binds tRNA with a low μM affinity

Using conventional native PAGE techniques to verify the interaction of HCoV-229E pp710 with nucleic acids (chapter 3.2.7), a relatively low amount of “shifted” protein-nucleic acid complexes was observed, indicating rather weak interactions. However, such weak interactions with nucleic acids can be detected and quantified applying the method of zone-interference gel electrophoresis (ZIGE). According to [143], ZIGE is particularly valuable for low affinity interactions in the μM range, because the protein is in contact with the ligand for a longer time during electrophoresis. Considering that mature Nsp8 is reported to comprise RNA-dependent RNA polymerase activity [94; 97], tRNA was used as an easily accessible random RNA pool for interaction tests. As described in the same report, the K_D -value is determined to represent the negative reciprocal of the slope of the linear function that is displayed in fig. 55A, depending on the migration distance of the protein in the presence of various ligand concentrations. This approximation revealed an interaction of SARS-CoV pp78 with tRNA in the low micromolar range in the presence of low quantities of Mg^{2+} . Similar K_D -values were previously reported for mature Nsp8 of HCoV-229E and SARS-CoV [103]. This implies that the affinity of Nsp8 towards tRNA, as well as probably for different other random RNA fragments, remains highly conserved in the pre-processed pp78 state.

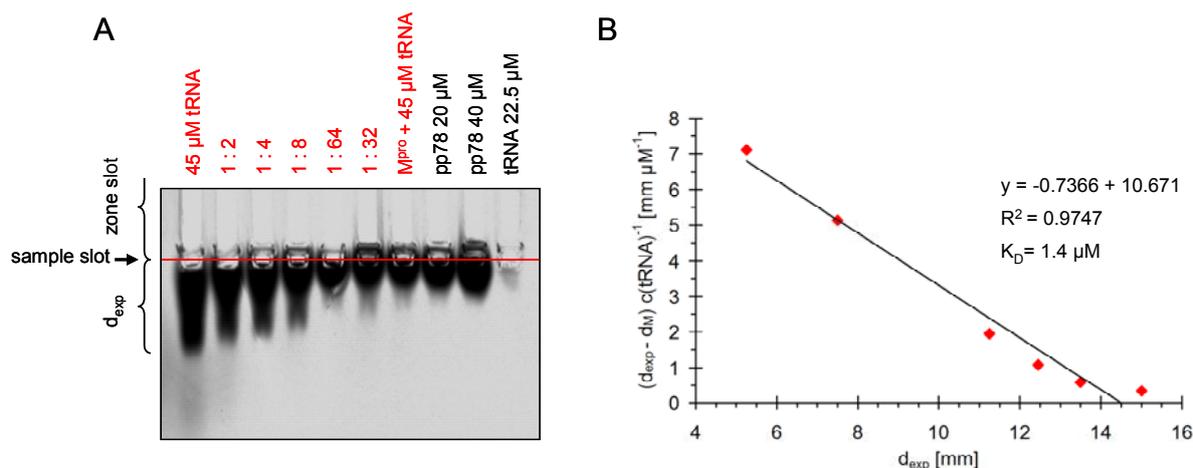


Figure 55. Zone-interference gel electrophoresis. (A) 10 μl of 20 μM SARS pp78 (N-terminal his₆-tag) were applied to the sample slots of an agarose gel (0.8% (w/v) agarose in TB buffer supplemented with 100 μM MgCl₂ and 100 mM NaCl). A serial dilution of yeast tRNA (Roche) in dH₂O (treated with 1% (v/v) DEPC) was applied to the zone slots (100 μl each; concentration in *red*). The voltage of the electric field was adjusted to 100 V for 90 min, the chamber was properly cooled on ice and the gel was stained with coomassie, detecting proteins afterwards. (B) The K_D -value was calculated according to [143]: The migration distance of the protein with tRNA (d_{exp}) was plotted against $d_{exp} - d_M$ (d_M : migration distance of the protein in the absence of the ligand) multiplied by the concentration of tRNA, whereas the slope of the displayed function represents $-K_D^{-1}$. SARS-CoV M^{pro} was used as a negative control and did not display an electrophoretic mobility shift at the applied experimental conditions.

3.10. Structural/functional comparison of pre-processed polyproteins of SARS-CoV and HCoV-229E

Summarizing the oligomerization characteristics of pp710, pp79 and pp78 presented within this thesis, a mixture of predominantly monomeric and dimeric states of the polyproteins was consistently identified in solution, along with additional less abundant higher oligomeric states. The associated molecular weights of the different protein species were approximated based on the retention volumes on the calibrated HiLoad 16/60 Superdex 200 SEC column (table 24). It was indicated that pp710, pp79 and pp78 consistently formed a non-covalent dimeric state. The corresponding structural parameters of the investigated polyprotein monomers and dimers revealed by DLS and SAXS experiments are summarized in table 25. Overestimation of the molecular weight by SEC and DLS for some protein species was most likely caused by the elongated shapes of the molecules, as demonstrate by SAXS data. Noteworthy, the maximum diameter of monomeric pp78 is relatively close to the value determined for monomeric pp710, promoting the suggestion that Nsp8 is the decisive factor for the elongated geometry of the polyproteins.

Interestingly, results obtained from SEC analysis (fig. 8/9), blue native PAGE (fig. 10), and cross-linking experiments with BS2G (fig. 18) consistently strengthened the suggestion that the equilibrium position of SARS-CoV pp710 is close to the monomeric state in solution, while C-terminal truncation by proteolytic processing that results in the formation of pp79 and subsequently of pp78 continuously shifts the equilibrium position to the dimeric state. Nevertheless, all three pre-processed polyprotein proteins additionally formed moderately higher oligomers, up to a tetrameric state for SARS-CoV pp78, as detected by mass spectrometry.

Table 24. Molecular weight approximation of polyprotein species based on calibrated SEC

virus	absorbance peak	retention volume [ml]	calculated mass [kDa] ¹	oligomeric state
SARS-CoV	pp710 peak I	65.1	159	dimer
	pp710 peak II	76.2	63	monomer
	pp710 peak I (30 mM HEPES buffer, expression culture : + 0.5 mM Zn ²⁺)	65	160	dimer
	pp710 peak II (30 mM HEPES buffer, expression culture: + 0.5 mM Zn ²⁺)	76.1	63	monomer
	pp710 G381E peak I	65.5	154	dimer
	pp710 G381E peak II	74.7	71	monomer
	pp79 peak I	69.6	109	dimer
	pp79 peak II	78	54	monomer
	pp78 peak I	72.2	88	dimer
	pp78 peak II	81.5	40	monomer
HCoV-229E	pp710 “aggregate” *	47.9	> 600	-
	pp710 peak I	66	148	dimer
	pp710 peak II	77.5	56	monomer
	pp710 Q83A peak I*	47.1	> 600	-
	pp78 peak I (GST-tag cleaved)	71.4	94	dimer
	pp78 peak II (GST-tag cleaved)	80.2	45	monomer

*not considered to be a defined oligomeric state

¹Thaumatococcus (*Thaumatococcus daniellii*; 22 kDa) (Sigma), ovalbumin (hen egg; 44 kDa), conalbumin (chicken egg white; 75 kDa), aldolase (rabbit muscle; 158 kDa), thyroglobulin (bovine thyroid; 669 kDa) (all GE Healthcare) were used for column calibration.

Table 25. Summary of structural properties obtained by SAXS and DLS

virus	protein	R_h [nm] ¹	R_g [nm] ²	R_h from R_g [nm] ³	D_{max} [nm]	NSD ⁴
SARS-CoV ⁵	pp78 monomer	3.3 ± 0.3	3.20 ± 0.01	4.1	10.0	0.98 ± 0.08
	pp78 dimer	5.3 ± 0.1	3.88 ± 0.01	5.0	13.0	0.80 ± 0.05
	pp79 monomer	4.5 ± 0.1	3.82 ± 0.02	4.9	13.1	0.65 ± 0.04
	pp79 dimer	5.8 ± 0.1	4.76 ± 0.02	6.1	17.9	0.58 ± 0.03
	pp710 monomer	4.6 ± 0.2	3.99 ± 0.02 ⁷	5.2	13.1	0.74 ± 0.05
	pp710 dimer	6.5 ± 1.1	5.42 ± 0.01	7.0	18.0	1.29 ± 0.08
HCoV-229E ⁶	pp78 monomer (GST-tag cleaved)	3.4 ± 0.4	2.65 ± 0.11	3.4	8.7	1.43 ± 0.06
	pp78 dimer (GST-tag cleaved)	4.5 ± 0.2	3.37 ± 0.11	4.3	11.4	1.45 ± 0.06
	pp710 monomer (C-terminal his ₆ -tag)	4.8 ± 0.1	3.43 ± 0.10	4.4	10.5	0.72 ± 0.10
	pp710 dimer (C-terminal his ₆ -tag)	6.5 ± 1.1	4.07 ± 0.08	5.3	14.3	0.74 ± 0.10

¹ hydrodynamic radius, determined by DLS

² radius of gyration, taken from the P(R)-function determined by GNOM and verified by AutoRg implemented in PRIMUS; commonly the raw scattering data was extrapolated to infinite dilution, if a concentration-dependency of the R_g was observed.

$$^3 \quad R_h = R_g \sqrt{\frac{5}{3}} \quad (\text{approximation of } R_h, \text{ see method section})$$

⁴ normalized spatial discrepancy, calculated by the “automated mode” of DAMAVER involving at least 10 *ab initio* models calculated by DAMMIF. NSD is a quantitative measure of similarity of particles including all its components. For identical models the NSD is equal to zero [179].

⁵ 280 mM NaCl, 30 mM NaH₂PO₄, pH 7.9, 2 mM DTT, all proteins owned an N-terminal strep-tag

⁶ 300 mM NaCl, 30 mM NaH₂PO₄, pH 7.9, 2 mM DTT

⁷ reduced to 3.66 nm for protein originating from an expression culture supplemented with 500 μM Zn²⁺ and in the presence of 3 mM DTT.

3.10.1. Conservation of the Nsps’ secondary structure during maturation

During proteolytic pp710 maturation, the intermediate products pp78 and pp79 shared similar oligomerization characteristics, although the equilibrium position appeared to be different. Thus, the structural conservation of the fold of the Nsp subdomains within the polyproteins during maturation remained to be investigated. High-resolution structures of all mature Nsps assembling pp710 have already been reported for both coronaviruses [101; 102], except for Nsp7, Nsp8 and Nsp10 of HCoV-

229E, representing a valuable comparison source for the recorded CD data. For Nsp9, the high resolution structures of HCoV-229E (pdb code 2J97) and SARS-CoV (pdb code 1QZ8) are highly similar, as determined by an average RMSD value of 1.483 Å (PyMOL Molecular Graphics System, Schrödinger) for the backbone atoms and expected due to the high sequence identity of 45.9%. Since the latter value is in a similar range also for Nsp7, Nsp8 and Nsp10 of both viruses (see table 13), at least basically homologue structures were assumed for the other three Nsps. In this context, CD-spectra of pp710 revealed indeed a comparable secondary structure composition of the SARS-CoV and HCoV-229E homologues in this study (chapter 3.2.3.1).

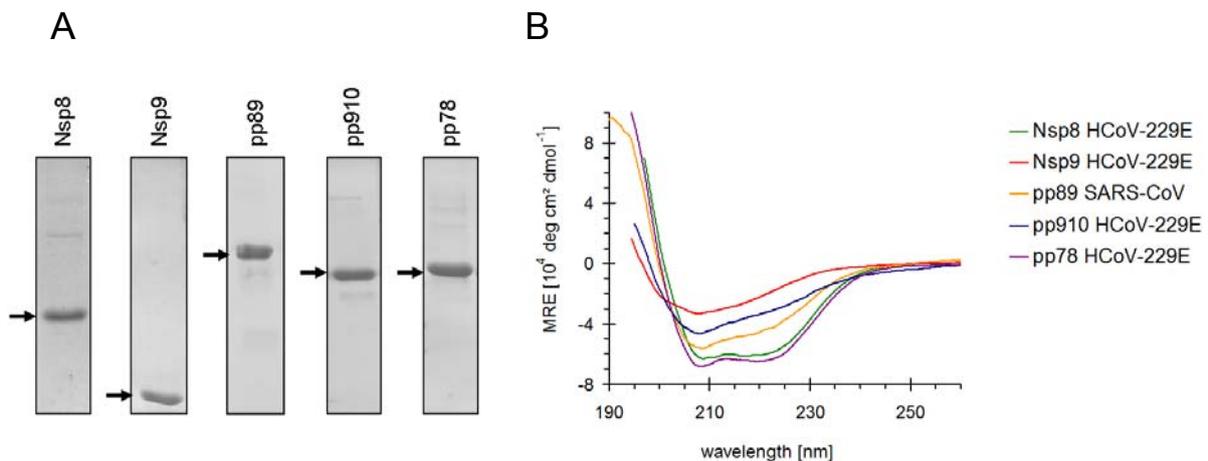


Figure 56. Investigation of the secondary structure content of selected polyproteins in comparison to individual mature Nsps. (A) Purification of the proteins of interest was verified by reducing SDS-PAGE. To purify HCoV-229E Nsp8, the phosphate buffer used during cells lysis was supplemented with 10% glycerol and 0.01% β -OG, while it was supplemented with 5% glycerol for purification of N-terminally strep-tagged pp89 of SARS-CoV. Authentic termini of HCoV-229E Nsp8 (21.6 kDa) and Nsp9 (12.0 kDa) as well as pp910 (26.4 kDa, the cleavage site was mutated in P1-position Q \rightarrow N) were generated by M^{pro} cleavage. HCoV-229E pp78 contained six amino acids of Nsp6 at the N-terminus in addition to four amino acids upstream related to the cloning process and the *prescission* protease cleavage site. (B) Far-UV CD spectra of the samples described in (A), measured in sodium phosphate buffer supplemented with 200 mM NaF.

The CD spectrum of SARS-CoV pp89 indicated a mixed α -helical and β -sheet secondary structure that approximately fits the merged spectra obtained for individual mature Nsp8 and Nsp9 (both HCoV-229E). Considering that HCoV-229E Nsp10 (see chapter 3.13) contains a slightly increased content of α -helix than Nsp9, which is β -sheet-enriched, these secondary structures have also been well preformed in pp910, since its α -helix-content was higher than that of Nsp9, with a similar percentage of β -sheet structure, according to the CD spectrum. Consequently, no indications for significant differences in the secondary structure composition have been obtained for the pre- and

3. Results and discussion

post-cleavage states of the proteins investigated in this study, while the SAXS results already indicate at least common characteristics of the tertiary structure of pp710 at low resolution (chapter 3.2.3.2/3.2.3.3).

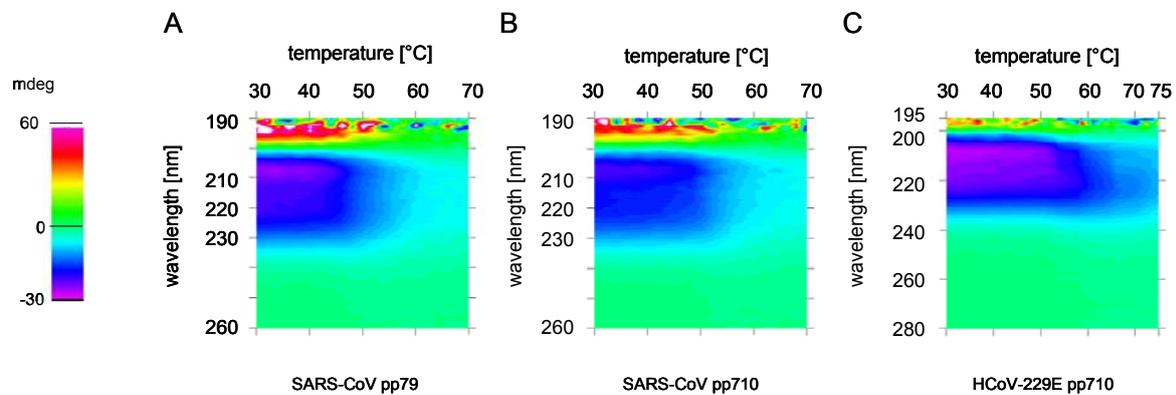


Figure 57. In comparative melting experiments the ellipticity of three proteins was continuously recorded by a CD spectrometer in phosphate buffer supplemented with 300 mM NaF, pH 7.9. The ellipticity is colour-coded, light green-blue comprises an ellipticity of zero, according to the displayed scale on the left. At a wavelength of 220 nm the melting temperature of (A) SARS-CoV pp79 is 50.5 °C, (B) for SARS-CoV pp710 $T_m = 52.5$ °C, and (C) $T_m = 61.0$ °C for HCoV-229E pp710. Beside this difference the ellipticity spectrum at room temperature is highly similar for all three proteins, indicating a similarly high α -rich fold.

Comparing the thermal stability of pp79 and pp710 by CD spectroscopy, results in the colour-coded interval CD spectra with an additional temperature axis displayed in figure 57. Complete far-UV CD spectra were recorded in temperature intervals of 1 °C at a continuous scanning speed of 100 nm min⁻¹ without time delay. The determined thermal stability of SARS-CoV pp79 ($T_m = 50.5$ °C) was slightly decreased compared to that of corresponding pp710 ($T_m = 52.5$ °C), reflected by the temperature corresponding to the most significant loss of ellipticity at a wavelength of 220 nm. The difference might be explained by a decreased number of unspecific intermolecular disulfide bonds formed involving cysteine residues of SARS-CoV Nsp10 that is absent in pp79. For HCoV-229E pp710, a significantly increased stability ($T_m = 61.0$ °C) compared to its homologue was detected. A slightly altered monomer-dimer equilibrium could have an effect on the stability, together with the different amount of cysteine residues in pp710 of SARS-CoV (21 cysteines) and HCoV-229E (17 cysteines) that could form intra- or intermolecular disulfide bonds. Remarkably, in all three spectra the ellipticity crossed the value of zero ellipticity (x-axis) at almost the same wavelength at 20 °C, indicating a similar ratio of α -helix and β -sheet. The cooperative continuous unfolding reaction represents another interesting consistent characteristic. There was no obvious step-wise domain unfolding, suggesting that the stability of one Nsp domain affects the others. Since β -sheet-enriched Nsp9, which was observed to be significantly stabilized (see fig. 77C), showed a comparably low mean residue ellipticity that was partly overlaid by the denaturation of α -helix structure during the

thermal unfolding experiment of pp710, almost preventing to follow the impact of the Nsp9 structure on the stability of other Nsp domains. In contrast to the cooperative continuous unfolding mechanism observed for coronaviral polyproteins in this study, immunoglobulin G, for example, that also represents a large multi-domain protein is reported to follow a typical unfolding reaction in two defined transition stages [209]. Despite recent progress in analyzing near UV spectra, in establishing protocols to analyze thermodynamic values by CD spectroscopy [210] and in deconvoluting secondary structure fractions, e.g. by using the convex constraint algorithm [211], the calculation and separation of the relative secondary structure fractions of different domains of proteins still lacks some accuracy.

3.10.2. RNA-dependent RNA polymerase (RdRp) activity assay

To probe a putative polymerase activity of different pp710 intermediate cleavage products, polyproteins that are supposed to have a functional impact on the RTC, a sensitive RdRp assay was carried out with strong support by *Yibei Xiao* (University of Lübeck). Using ATP that was radioactively labelled with γ - ^{32}P -phosphate hydrolytic activity of SARS-CoV pp79 (fig. 58A, lane 4), HCoV-229E pp78 (lane 7) and digested HCoV-229E pp710 (lane 9) was demonstrated. This enzyme activity was indicated by the detection of radioactively labelled diphosphate released from the labelled nucleotide. A negative control without protein was included (lane 6). To reveal polymerase activity, the analyzed proteins were incubated with α - ^{32}P -labeled GTP, resulting in the incorporation of the sensitive label into products of the polymerase reaction. The overall efficiency of catalysis and synthesis of a certain fragment length probably depends on the secondary structure of the RNA template, therefore an authentic 67 nt stem loop structure was selected with a sequence derived from FCoV. Polymerase activity was successfully demonstrated for SARS-CoV pp79 (fig. 58B, lane 4). Moreover, his₆-tagged HCoV-229E pp78 was capable of synthesizing polymerase products of various size (fig. 58B, lane 7), whereas HCoV-229E pp710 displayed only weak RdRp activity with a preference for producing short RNA fragments (fig. 58B, lane 8). The concentration of some polymerase products might be below the detection limit. Interestingly, HCoV-229E pp78 (with cleaved GST-tag, but GPLG left at the N-terminus) displayed polymerase activity without significant product size specificity, while SARS-CoV pp78 (N-terminal his₆-tag) synthesized predominantly small RNA fragments (4-6 nt). The RdRp activity of HCoV-229E pp710 before and after proteolytic processing by SARS-CoV M^{pro} was weak, but detectable in comparison to the negative control (lane 6 in each panel).

Overall, considering the product fragment length, it additionally has to be considered that RdRp activity might be influenced by the storage conditions, the ionic strength of the buffer, as well as by the tag length and position. The potential impact of the protein's oligomeric state is most interesting to identify a regulatory effect *in vivo/in vitro* of the concentration-dependent oligomerization observed for the two homologues of pp710 (fig. 8/9). Having in mind that pp710 was predominantly monomeric

3. Results and discussion

under various conditions (fig. 8/9), this state could be correlated with the synthesis of small RNA fragments, in agreement with the primase activity postulated for HCoV-229E pp710 [96]. *De novo* RNA synthesis of different fragment sizes by SARS-CoV pp79 was revealed for the first time in this study. The pp79 preparation purified by SEC was assumed to be enriched in monomeric and dimeric pp79 as verified by DLS prior to the experiment (not shown). However, the conditions of the polymerase assay may further affect the oligomeric state of the protein. Based on the obtained results, pp79 is considered to represent a functional building block of the RTC. Consequently, it is indicated that neither a hetero-hexadecameric or similar oligomeric state nor an N- or C-terminal cleavage of Nsp7 or Nsp9, respectively, is required for RdRp activity of Nsp8.

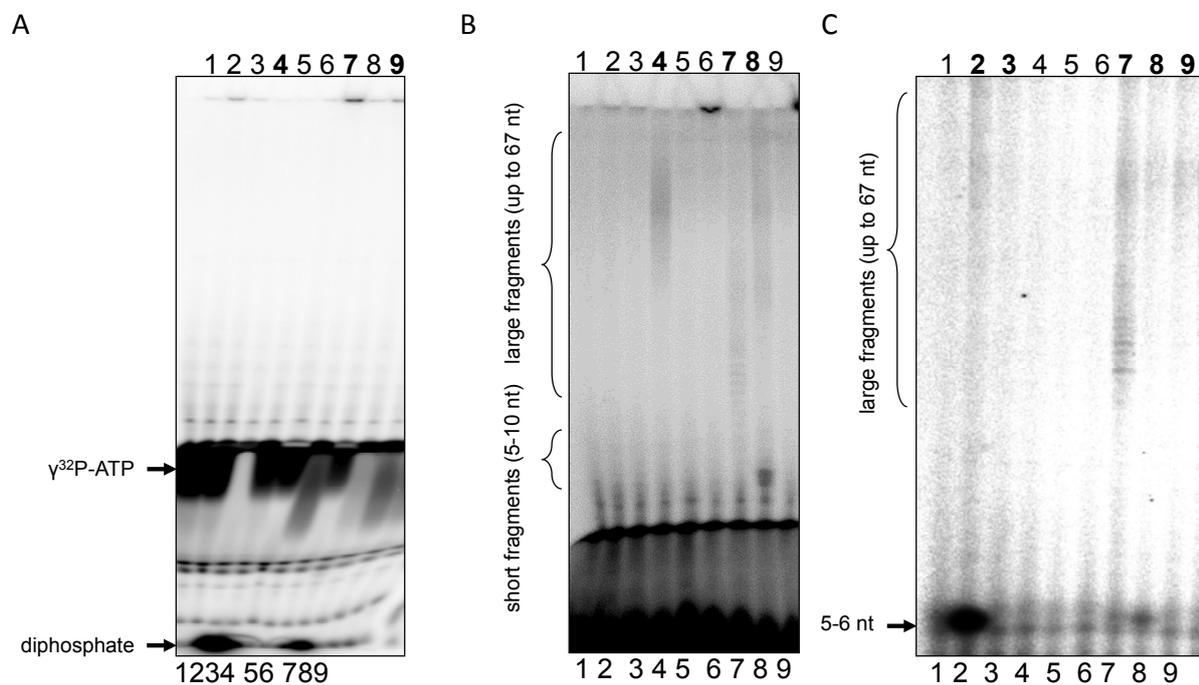


Figure 58. Products of the RdRp assay, performed as described in the method section, were separated on a denaturing PAGE gel. Separation of the RdRp activity assay products using γ - 32 P-labeled ATP (panel A) and α - 32 P-labeled GTP (panel B and C) is depicted. SARS-CoV pp79 (panel B, lane 4) owing an N-terminal strep-tag displays polymerase activity synthesizing even larger fragments up to almost template size (panel A and B, lane 4). SARS-CoV pp78 owing an N-terminal his₆-tag displays polymerase activity as well (panel C, lane 2) even though for the chosen experimental conditions the synthesis of a short oligoribonucleotide is specifically preferred. Lane 6 in every panel comprises the negative control without protein.

3.10.3. Identification of binding epitopes of nucleoside derivatives by saturation-transfer-difference (STD) NMR spectroscopy

STD-NMR spectroscopy was performed in cooperation with Dr. *Miriam Kötzler* (University of Hamburg), who further processed the experimental data. STD-NMR spectroscopy essentially allows determining the epitope of a small ligand binding to a comparably large protein. A high level of saturation-transfer-difference at a hydrogen atom of the bound ligand upon a resonance-saturating pulse selectively targeting the protein, reflects a close proximity of the respective part of the ligand to the protein surface. On the other hand, a low relative STD-effect indicates a larger distance of the hydrogen atom to the protein surface, facing away from the protein. In the last chapter, RNA-dependent RNA polymerase activity (EC 2.7.7.48) was successfully demonstrated for different polyproteins (fig. 58), as previously shown for mature Nsp8 of SARS-CoV [94; 97]. To identify binding epitopes and binding modes of nucleoside derivatives at the surface of polyproteins, the interaction with ATP, AMP, uridine, dGDP and NAD was investigated applying STD-NMR spectroscopy. For all ligands displayed herein (fig. 59/60), a well-defined specific ligand epitope, highlighting the importance of specific functional groups, is presented.

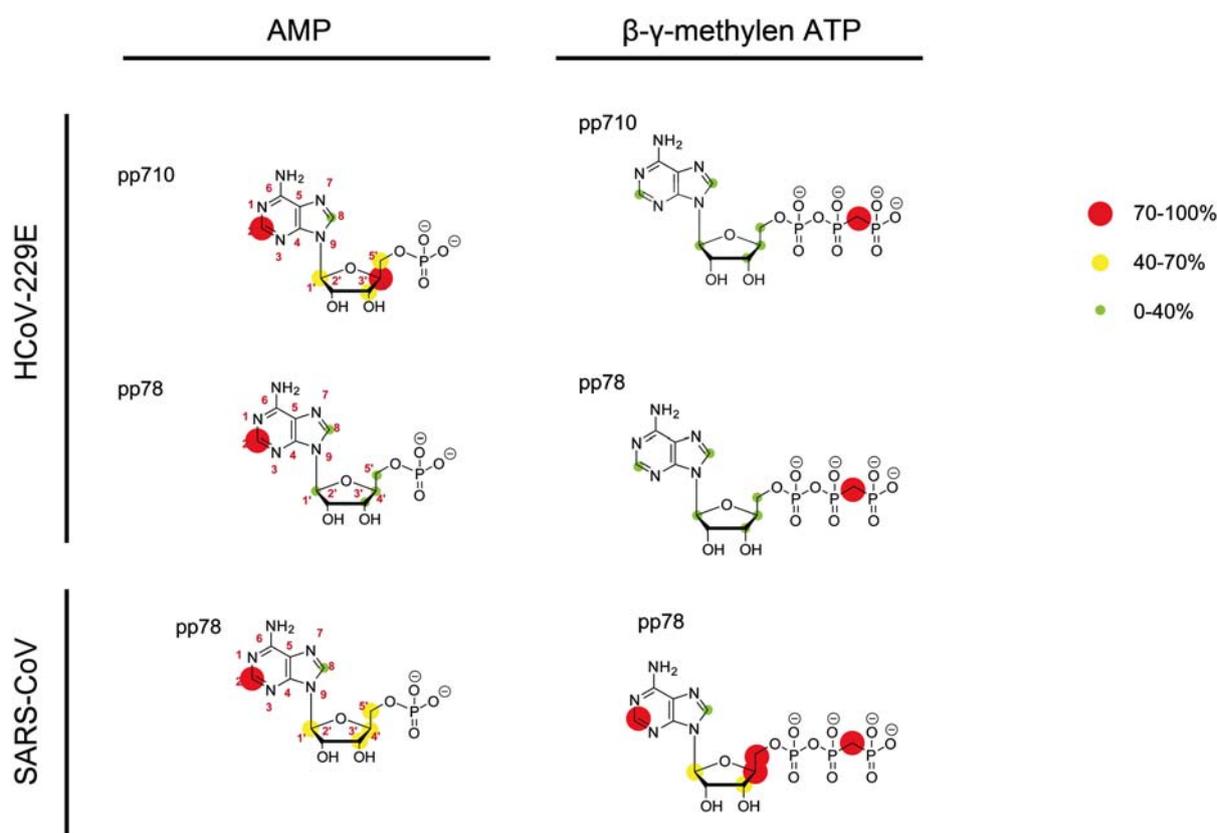


Figure 59. Identification of binding epitopes of AMP (left) and β,γ -methylen-ATP (right) for coronaviral polyproteins by STD-NMR spectroscopy. The epitopes are visualized by colour-coded relative STD effects at three different levels at the respective hydrogen atom that is not exchanging with deuterium from the solvent. The highest absolute saturation transfer difference in a molecule is set to 100%. The similarities of the β,γ -methylen-ATP binding mode with either HCoV-229E pp78 or pp710 suggest that there is no influence by either Nsp9 or Nsp10.

Belonging to the enzyme class of transferases, RNA-dependent RNA polymerases are presumed to hydrolyse the β,γ -phosphorester bond of a nucleoside triphosphate like ATP. Diphosphate is released and the RNA is elongated by AMP forming another phosphorester bond involving the 3'-hydroxy group of the terminal ribose moiety. Investigating the ligand epitope of AMP binding to pp78, the highest saturation transfer difference is observed at the C₆-ring of the purine moiety. This observation is consistent for pp78 of HCoV-229E and of SARS-CoV. As soon as a triphosphate ester is attached to the nucleoside (e.g. β,γ -methylen-ATP), forming a nucleotide substrate, a remarkably different epitope including the phosphate moiety that turned into close proximity to the protein surface is indicated. This effect was quantified by the two additional hydrogen atoms at the methylen group of β,γ -methylen-ATP and equals 100% relative STD effect for all three proteins. A specific recognition and additional interaction around the sugar moiety is hardly visible for pp78 and pp710 HCoV-229E. For SARS-CoV pp78 the STD effect around the ribose moiety is increased, particularly at the 4' and 5' carbon atom (72% and 79%, respectively). This ligand orientation would overall well match the specific recognition of the triphosphate moiety of an RNA polymerase [212] or at least ATP-

hydrolysis activity for HCoV-229E and SARS-CoV pp78 as well as for HCoV-229E pp710. Notably, the epitope determined in the presence of pp78 is not significantly different in the presence of pp710, demonstrating that Nsp9 and Nsp10 did not alter the epitope by changing the structure of the pp78 binding site or by providing an alternative binding site, e.g. recognizing the nucleobase moiety. Investigating uridine that lacks a phosphate ester, the epitope of the interaction with SARS-CoV pp78 was different compared to adenosine with one phosphate group (AMP) due to a high saturation around the ribose – particularly at the 4' and 5' carbon atom. In protein samples mixed with uridine a varying degree of its hydrolysis was observed referring to the depicted reaction (fig. 60). Therefore, it was not possible to perform a reliable epitope-determination of uridine with either purified HCoV-229E pp710 or HCoV-229E pp78. It is unknown whether this could be an intrinsic activity of the polyproteins under certain experimental conditions.

Nicotinamide adenine dinucleotide (NAD) which is a common co-factor in redox reactions, is binding to pp710 and pp78 in a specific orientation as well (fig. 60). Similar to the binding epitope determined for AMP a high STD effect is detected at the adenine base. Even though, in agreement with the expected function, there was no significant interaction distributed throughout the whole molecule and no expectation of a redox activity, the adenosine group is sterically able to bind to the polyproteins. In contrast, the nicotinic acid 6-ring which is similarly connected to the other ribose moiety by an *N*-glycosidic bond clearly seems to be specifically averted from the protein surface. Nevertheless, there is a minor increase of saturation transfer difference at the ribose moieties of NAD binding to HCoV-229E pp710 in comparison to HCoV-229E pp78. Applying desoxyguanosindiphosphate (dGDP), which is not expected to be a suitable substrate for an RNA polymerase, again at a molar ratio of 100:1 with HCoV-229E pp78, no significant STD was displayed. Therefore a specific binding of dGDP was not indicated.

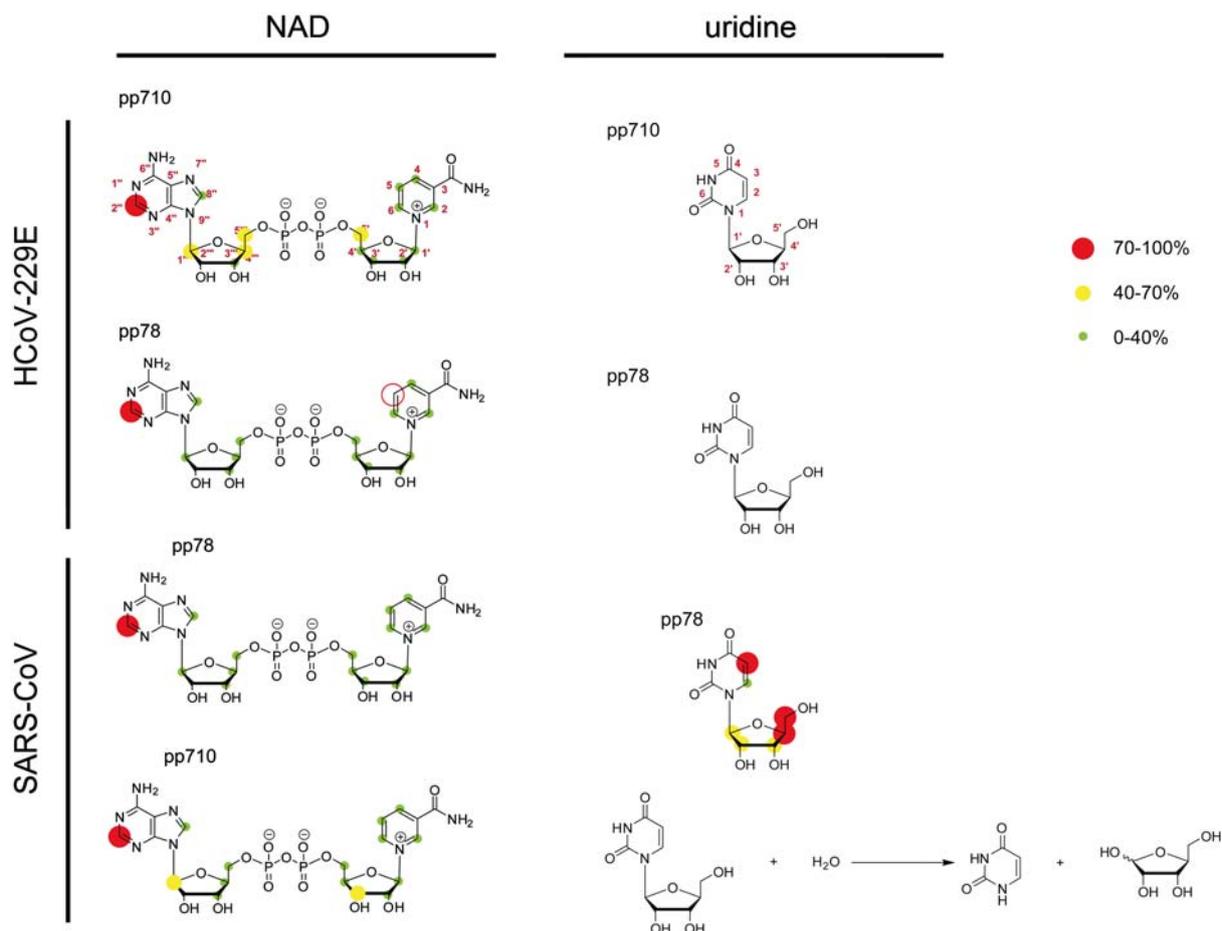


Figure 60. Identification of binding epitopes of NAD (left) and uridine (right) for coronaviral polyproteins by STD-NMR spectroscopy. The relative saturation transfer difference was assigned by the colour-coding as introduced in fig. 59. The highest absolute saturation transfer difference in a molecule is set to 100%. * Due to hydrolysis of the *N*-glycosidic bond in uridine, as evident from the NMR spectra, it was not possible to perform a reliable epitope-determination of uridine with either purified HCoV-229E pp710 or HCoV-229E pp78; the reaction is depicted.

For comparison, particularly to the epitope determined for β,γ -methylene-ATP, the binding site of ATP and UTP in two RNA polymerase high-resolution structures, respectively, is shown in figure 61. Although in detail both binding sites seem to disagree with the ligand epitopes determined by NMR spectroscopy, particularly the ATP epitope of SARS-CoV pp78 shares some aspects with the UTP ligand epitope in panel A and B, e.g. the close proximity of the β,γ -phosphorester groups to the protein surface. At the binding site of the HCV RNA polymerase. Specifically the γ -phosphate of UTP is stabilized by a hydrogen bond with F224, which fits the observed high relative STD-effect at the methylene group of β,γ -methylene-ATP well. The high relative STD effect at this position is conserved for both coronavirus species as well as the low STD-effect around the nucleobase in the presence of a β - and γ -phosphate group.

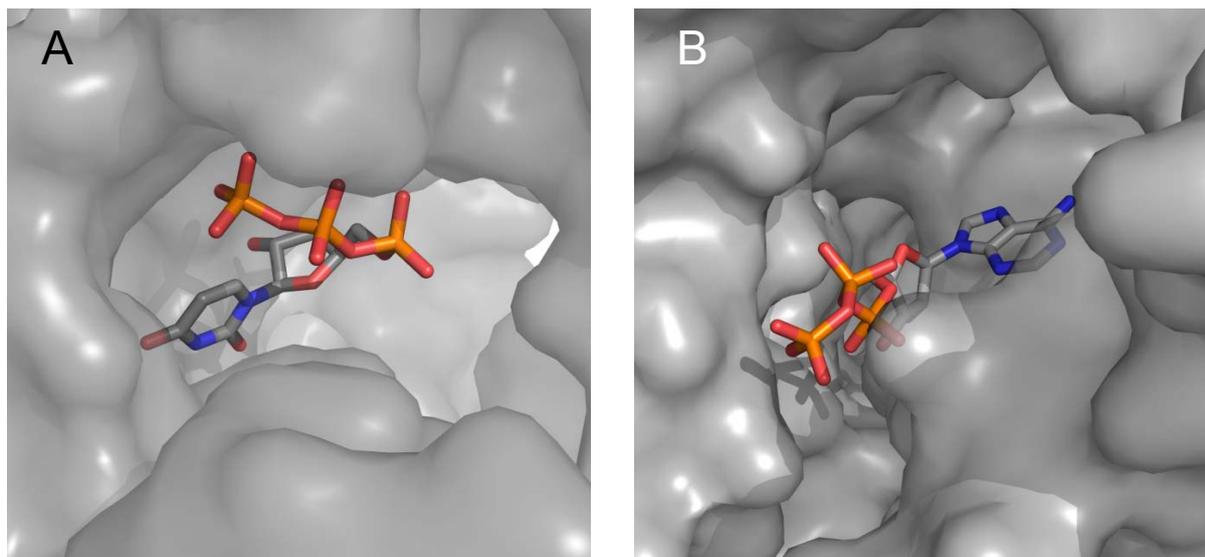


Figure 61. Two RNA polymerase high resolution structures revealing the respective nucleotide binding site and ligand orientation. **(A)** Panel A displays the HCV RNA polymerase (pdb code 1GX6) [212] in complex with UTP forming four hydrogen bonds. Additionally four Mn^{2+} ions surround the ligand in the high resolution structure. **(B)** The ATP epitope defined by three hydrogen bonds of the RNA-dependent RNA polymerase of the dsRNA bacteriophage phi6 is shown in panel B (pdb code 1HI1) [213]. [Carbon atoms are coloured *grey* and phosphorus atoms *orange*.]

In summary, there are strong similarities of the epitopes comparing pp710 and pp78 of either virus indicating that the epitope is widely determined by the Nsp7 and the Nsp8 domain. Furthermore it is suggested that adenine and –if present – a triphosphate moiety is specifically recognized and buried in the binding pocket of the respective protein, where this specificity is slightly weaker for the SARS-CoV polyproteins. Overall, the tendency to recognize the γ -phosphate moiety of a nucleotide is indicative for a conserved polymerase activity mechanism.

3.11. Interaction of Nsp7 and Nsp8

3.11.1. Oligomerization of SARS-CoV Nsp7 and Nsp8

The interaction of mature coronaviral Nsp7 and Nsp8 (Nsp(7+8)) is already well investigated for two coronaviruses. Crystallographic data of an 8:8 hexadecamer of SARS-CoV and a heterotrimer of FCoV involving one Nsp8 and two Nsp7 subunits is available. A fundamental biochemical and biophysical characterization of HCoV-229E Nsp(7+8) and SARS-CoV Nsp(7+8) in comparison to their precursor pp78 is performed in this chapter.

3. Results and discussion

The putative complex of mature SARS-CoV and HCoV-229E Nsp7 (10 kDa, tab. 13) and Nsp8 (21 kDa) was formed by cleaving pp78 with SARS-CoV main protease. In analytical calibrated SEC (HiLoad 16/60 Superdex 200) SARS-CoV Nsp7 and Nsp8 co-localized and peaked at a retention volume corresponding to a molecular weight of 62 kDa (fig. 62A). This corresponds to an interaction ratio of 2:2 at the applied experimental conditions. At a retention volume of around 85 ml, the absorbance ratio $A_{260\text{ nm}}/A_{280\text{ nm}}$ significantly increased, likely originating from remaining nucleic acid carried over from *E. coli* cells. For further experiments, SARS-CoV M^{pro} was removed from the cleavage mixture by incubating the solution with Ni-NTA resin. Figure 62B shows the purified Nsp(7+8) complex of SARS-CoV (lane II) and HCoV-229E (lane III) verified by SDS-PAGE. Analytical SEC (Superose 12 10/300 GL) revealed that the applied buffer had a slight influence on the oligomerization of SARS-CoV Nsp(7+8) (fig. 62C; see fig. 62A for comparison). However, a state indicative for a 2:2-interaction of Nsp7 and Nsp8 corresponding to roughly 60-70 kDa was most abundant, together with higher oligomeric states in the range of 100-180 kDa, as demonstrated by clear native PAGE (fig. 62D). The protein was carefully excised from the gel and further analyzed by mass spectrometry (ESI ion trap device) after trypsin digestion. Nsp7 and Nsp8 were identified with a sequence coverage of 84% and 57%, respectively. A coomassie-stained putative complex with a similar molecular weight was obtained by blue native PAGE as well (not shown). These results strongly indicate the existence of a highly abundant 2:2 interaction state of Nsp7 and Nsp8 in solution, clearly smaller in size than the hexadecameric state previously observed in a high-resolution crystal structure [98] which might exist in addition. The proposed 2:2 interaction of Nsp7 and Nsp8 is in agreement with the oligomerization of the homologue proteins of the *Alphacoronavirus* HCoV-229E as previously indicated by SEC (fig. 29).

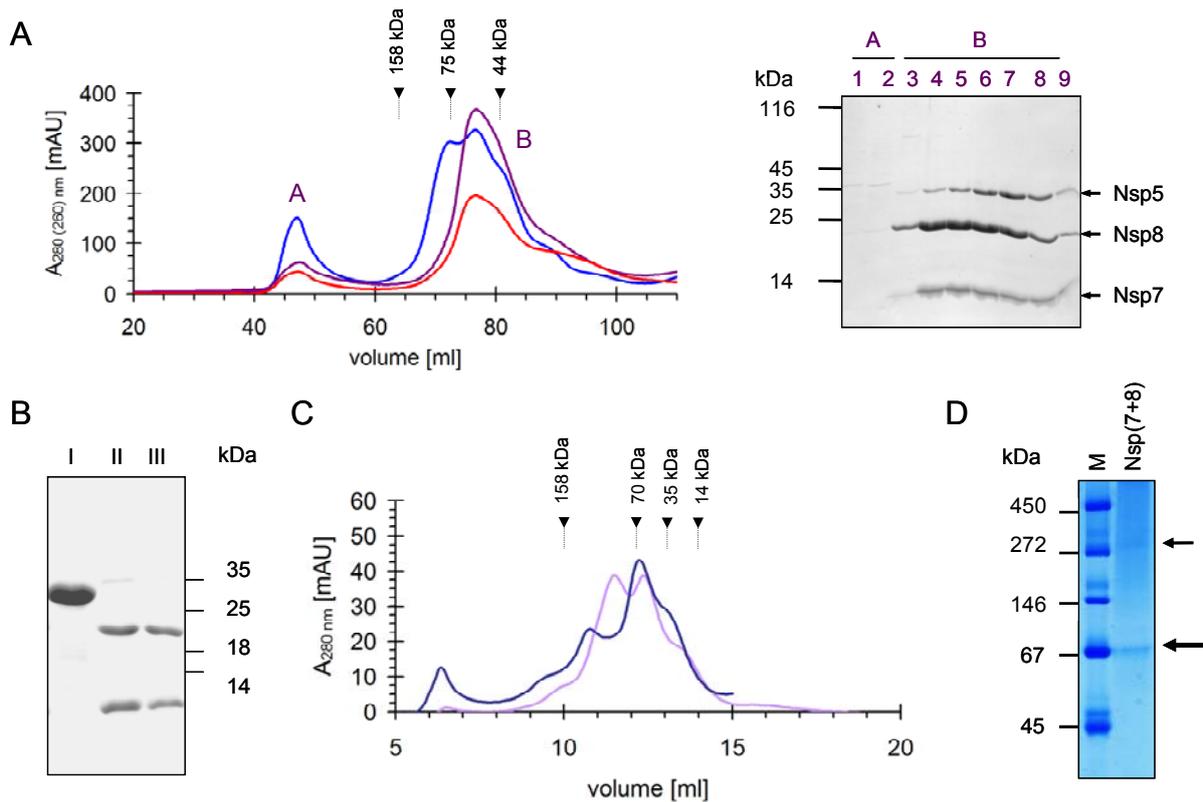


Figure 62. Interaction of mature Nsp(7+8). **(A)** SARS-CoV pp78 containing an N-terminal strep-tag was processed by SARS-CoV M^{pro} at a molar ratio of 4:1 and subsequently applied to SEC (Hi-Load 16/60 Superdex 200). Nsp7 and Nsp8 co-eluted and peaked slightly earlier than M^{pro} (purple chromatogram, A_{280} nm), as evident from SDS-PAGE analysis. The red chromatogram displays the absorbance of the same sample at 220 nm. Processing and analysing pp78 containing a C-terminal strep-tag fused to Nsp8 by SEC resulted in the blue chromatogram, while buffer and sample preparation were identical. **(B)** M^{pro} was removed by subsequent incubation with equilibrated Ni-NTA resin and centrifugation to obtain the pure Nsp(7+8) complex [lane I: SARS-CoV pp78, lane II: SARS-CoV Nsp(7+8), lane III: HCoV-229E Nsp(7+8)]. **(C)** The Nsp(7+8)-complex containing N-terminally strep-tagged Nsp7 was analyzed by SEC (Superose 12 10/300 GL) in HEPES buffer (dark blue) and in 350 mM ammonium acetate adjusted to pH 7.9, respectively (light pink). Besides a small protein fraction directly eluting from the column (MW > 300 kDa, “exclusion volume“), the estimated MW of all additional states is predicted to be smaller than 200 kDa. **(D)** A defined oligomeric state of Nsp(7+8) characterized by a molecular weight of 60 to 70 kDa is revealed by clear-native PAGE, corresponding to a heterotetramer.

To verify interaction states of SARS-CoV Nsp(7+8) at a high mass accuracy, native mass spectrometry was applied. The spectra displayed in fig. 63 support the previous observations, demonstrating a high abundance of a mass in agreement with a heterotetrameric 2:2-interaction (65,011 Da). This mass species was fragmented by collision-induced dissociation, yielding monomeric Nsp7 and Nsp8 (fig. 63C). However, smaller fractions of a heterodimer (32,547 Da) and a putative

3. Results and discussion

4:4-interaction (heterooctamer; 131,802 Da) requiring a second dimerization of the 2:2-heterotetrameric state were also detected.

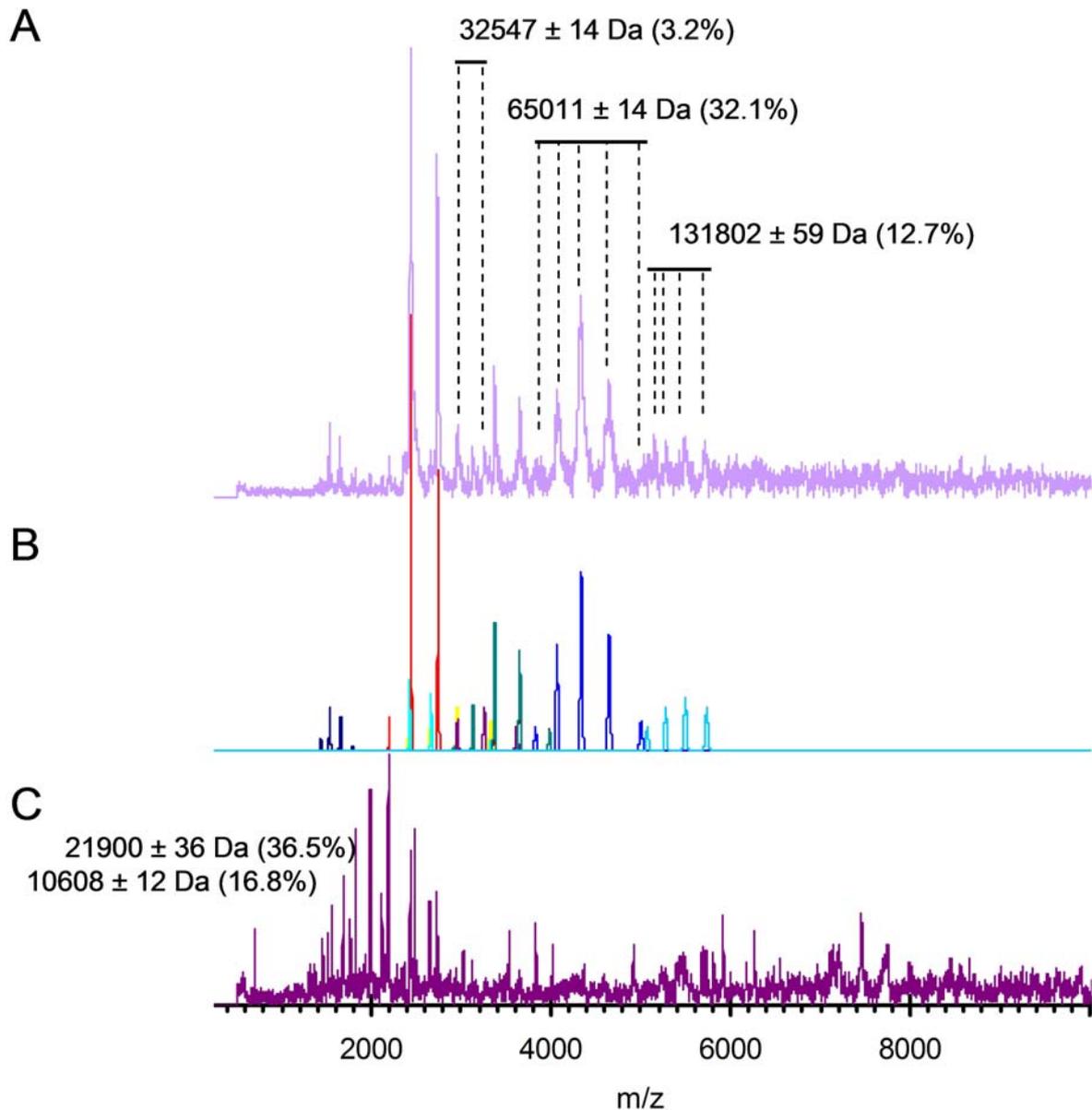


Figure 63. An equimolar mixture of Nsp7 and Nsp8 was obtained by completely cleaving SARS-CoV pp78 with N-terminal strep-tag by M^{pro} . The protease was removed prior to the experiment by Ni-NTA resin. (A) The m/z-spectrum at 10 V collision voltage. (B) Referring to panel A, individual mass species of the spectrum above are shown that were simulated by Massign. Nsp7 and Nsp8 were analyzed by an ESI-Q-TOF instrument revealing different oligomeric states; more than 30% of the total intensity in the displayed m/z range (around 65 kDa) fits a 2:2-interaction of both proteins along with a lower amount of a 4:4-interaction state indicative for a further dimerization mode. This observation nicely agrees with SEC experiments (fig. 62). The broadness of the peaks indicates a minor mass variation, perhaps due to small ligands carried over from *E. coli* or ions from the buffer. Another oligomeric state fits an Nsp8 dimer (green in the panel B), most probably overlaying with a marginal amount of a 2:1-(Nsp7:Nsp8)-complex fragment (purple). (C) Stepwise CID released a dimeric Nsp8 “core” and

free Nsp8 ($21,900 \pm 36$ Da), in parallel free Nsp7 ($10,608 \pm 12$ Da) is enriching by increasing the voltage (lower panel). The depicted purple spectrum was obtained at 120V collision voltage.

Due to the high sequence identity of Nsp7 and Nsp8 of distantly related coronaviruses (see also tab. 13), a 2:2-interaction state as proposed for the SARS-CoV Nsp7 and Nsp8 complex might be conceivable for other closely-related *Betacoronaviruses* as well, e.g. MERS-CoV, along with other oligomeric states. As shown above, mass spectrometry demonstrated the formation of small amounts of these additional oligomeric states, suggesting a model of a condition-dependent transition and balance of oligomers similar to the previous observations for polyprotein pp710 (e.g. chapter 3.2). This could also include specific oligomeric states with a molecular mass of more than 200 kDa that were not clearly detected by the methods applied in this study, but have been previously reported in a crystalline environment [98]. For mass spectrometry ammonium acetate buffer had to be applied, that altered the SEC chromatogram of the Nsp(7+8) complex compared to HEPES buffer (fig. 62C). A second dominating peak appeared at lower elution volumes, corresponding to a state larger than a 2:2 interaction. Thus, slight variations in the oligomerization pathway are assumed to depend on the buffer conditions required for the applied technique.

However, an impact of the purification tag position on the formation of oligomeric states could not be excluded. In chemical cross-linking experiments performed by te Velthuis *et al.* [97], C-terminal his-tagged Nsp8 formed an octameric state, while an N-terminal his-tag hindered the formation of higher order oligomers. Thus, the formation of a hexadecameric state of Nsp(7+8) could depend on correct N-terminal processing of Nsp8. Nevertheless, even the variation of the strep-tag position in pp78 did not result in any indication for a significant amount of a hexadecameric interaction state in this study. It has to be mentioned that chemical cross-linking as performed by te Velthuis *et al.* might have the disadvantage that it is difficult to distinguish between continuums of different oligomeric states, e.g. a mixture of dimer, tetramer and octamer and/or even larger (specific) oligomers. Additionally, cross-linking depends on the location of suitable amino acid residues at the interaction interface. Up to now, it has not been sufficiently proven by applying different methods at conditions close to physiological that octameric Nsp8 or hexadecameric Nsp(7+8) represent exclusive aggregation states of these proteins in solution, even though comparative SEC at a relatively low mass resolution was performed previously [97] indicating a hexadecameric interaction of SARS-CoV Nsp7 and Nsp8.

According to the present results it is suggested that a mixture of different oligomeric states is formed in solution, with at least a preference for a 2:2 interaction for the respective complex of HCoV-229E and SARS-CoV. This interaction state was independently observed in SEC experiments for SARS-CoV Nsp7 and Nsp8 after cleavage of pp78 containing different tags as well as authentic termini (chapter 3.4.1). This hypothesis agrees well with a significant amount of dimeric pp78 detected in solution (see chapter 3.9) and allows the conclusion that even distantly related

coronaviruses share the capability to form different small (hetero-)oligomeric states of mature Nsp7 and Nsp8, also displaying a similar oligomerization compared to the precursor pp78. To further investigate whether the domain-interactions within monomeric and dimeric states of pp78 are similar to that within heterotetrameric states of mature Nsp7 and Nsp8 in solution, SAXS technique was subsequently applied.

3.11.2. Nsp(7+8) in comparison to its precursor pp78 by SAXS

To address the question of the interaction of Nsp7 and Nsp8 in the precursor state pp78, SAXS experiments were performed as described. SAXS particularly aimed at enlightening the geometry of the N-terminus of Nsp8 within pp78, which was shown to fold into a long flexible α -helix in a high-resolution crystal structure of SARS-CoV Nsp(7+8) [98]. The fold of the N-terminal domain of Nsp8 in pp78 might be influenced by non-covalent interactions with Nsp7 at interfaces different to the covalent link of both proteins. Consequently, the aim was to predict the binding epitope(s) of SARS-CoV Nsp7 and Nsp8 in the consecutive precursor state. Monomeric and dimeric pp78 of SARS-CoV and HCoV-229E were purified by SEC as described in fig. 53 (chapter 3.9.1). The monodispersity of the sample was verified by DLS, revealing a hydrodynamic radius of 3.3 ± 0.3 nm for monomeric SARS-CoV pp78 and 5.3 ± 0.1 nm for dimeric pp78 of SARS-CoV. In comparison, the hydrodynamic radius of monomeric HCoV-229E pp78 is 3.4 ± 0.4 nm and 4.5 ± 0.2 nm for dimeric HCoV-229E pp78. Results of SAXS measurements at protein concentrations of 1 - 4.5 g l⁻¹ resulted in R_g and D_{max} values specified in table 25. A comparison of the distance-distribution functions of SARS-CoV pp78 monomer and dimer is shown in figure 64A, demonstrating the increased maximum diameter of the dimer (*yellow* curve, $D_{max} = 13$ nm) compared to that of the monomer (*red* curve, $D_{max} = 10$ nm). Furthermore, the incubation of monomeric SARS-CoV pp78 with SARS M^{pro} (ratio 10:1) demonstrated that the distance-distribution function continuously changed during proteolytic processing of the pp78 precursor. After 5 min (*dark blue* curve) the maximum diameter of the molecules increased by 2 nm, while an additional increase up to 13 nm was observed after 15 min of incubation (*light blue*). *Ab initio* models of dimeric pp78 of SARS-CoV and HCoV-229E (see also chapter 3.10) were calculated based on the distance-distribution functions (fig. 64B/C). Both dimeric pp78 molecules shared an elongated shape and a similar overall geometry, even though radius of gyration and maximum diameter of HCoV-229E pp78 are slightly smaller ($D_{max} = 11.4$ nm, $R_g = 3.37$ nm) compared to SARS-CoV pp78 ($D_{max} = 13$ nm, $R_g = 3.88$ nm).

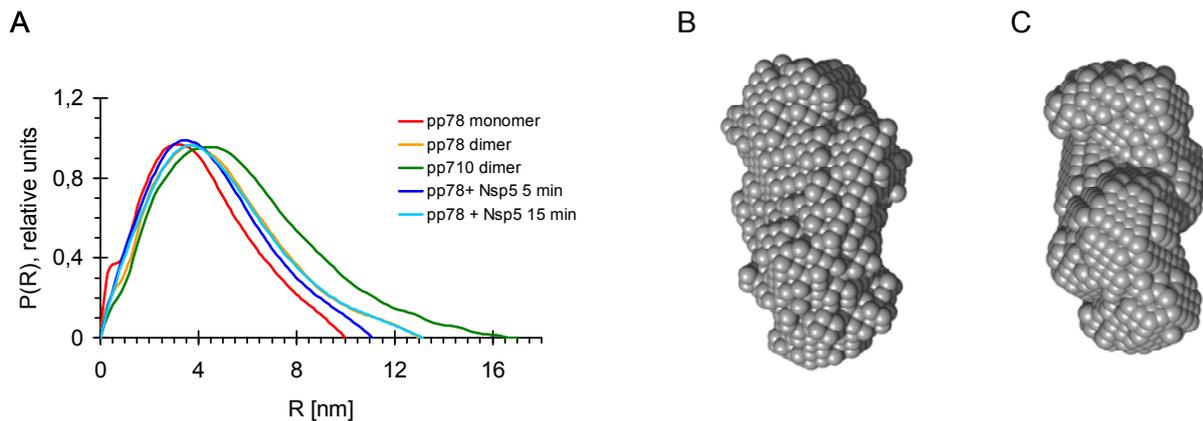


Figure 64. SAXS analysis of pp78. (A) P(R)-functions of SARS-CoV pp78 before and during incubation with SARS-CoV M^{pro}. The maximum diameter corresponds to the point of intersection with the radius axis. (B) *Ab initio* model of dimeric SARS-CoV pp78 with a maximum diameter of 13 nm; (C) *Ab initio* model of HCoV-229E pp78 shown for comparison ($D_{\max} = 11.4$ nm).

Evaluating the scattering pattern of monomeric HCoV-229E pp78 resulted in a valuable rigid-body model using the high-resolution structures of Nsp7 (*green*, pdb code 1YSY) and Nsp8 (*blue*, chain A and H, extracted from pdb code 2AHM) (fig. 65A). 68 cycles of refinement have been performed (penalty 0.4442, best chi value of the scattering fit: 1.0656, considering scattering vector values $0.01 < s \leq 0.28 \text{ \AA}^{-1}$). The restraint that Nsp7 and Nsp8 are linked C-terminus to N-terminus was set for rigid-body modelling and the rigid “stretched” isoform of SARS-CoV Nsp8 was used as a template, together with Nsp7. However, this rigid Nsp8 structure seemed to be insufficient to interpret the *ab initio* model (calculated by DAMMIF) of monomeric pp78 correctly (fig. 65C), as already discussed concerning full-length HCoV-229E pp710 (chapter 3.2.3.2/3.2.3.3). Since Nsp7 is reported to occupy a known interaction epitope at Nsp8 in the complex of mature Nsp7 and Nsp8 [98], a comparable interaction could be expected for pp78. The epitope shared by SARS-CoV Nsp7 and Nsp8 (chain A and E) involves four hydrogen bonds (Nsp7 Q19/Nsp8 T84, Nsp7 S61/Nsp8 P116, Nsp7 S57/Nsp8 I120, Nsp7 S54/Nsp8 I120). Assuming a specific degree of flexibility in the N-terminal domain of Nsp8, this epitope is also conceivable for HCoV-229E pp78. To take the flexibility of Nsp8 into account, Nsp8 was divided into four individual fragments that were used as templates for improved rigid-body modelling by SASREF (fig. 65B). The fragments of Nsp8 were designed according to reported sites of flexibility based on high-resolution structures [96; 98] and personal communication with Y. Xiao (University of Lübeck).

The resulting rigid-body model fitted the corresponding *ab initio* model even better, notably with a significant bending of two fragments of the Nsp8 N-terminus (fig. 65C). A manual repositioning of Nsp7 according to free volume within the *ab initio* model of the “scattering envelope” of SARS-CoV pp78 not defined by Nsp8 was performed. An *ab initio* model of monomeric pp78 of HCoV-229E is displayed for comparison (fig. 65D). For SARS-CoV pp78, the position of Nsp7 obtained by rigid-

3. Results and discussion

body modelling (fig. 65C, *green*) was changed according to the known interface of mature Nsp7 and Nsp8, resulting in the red Nsp7 molecule. This optimized rigid-body model fitted the *ab initio* model equally well (fig. 65C). To still allow the covalent bond between Nsp7 and Nsp8, a higher degree of bending of the N-terminal domain of Nsp8 is suggested. The resulting *ab initio* model of SARS-CoV pp78 shares common characteristics in terms of shape and tertiary structure with HCoV-229E pp78, for which the *ab initio* model was additionally well interpreted by an independent homology model that was calculated using the I-Tasser modelling server based of sequence homology (fig. 65E).

Fig. 65F displays the rigid body and the *ab initio* model of dimeric HCoV-229E pp78 based on the scattering data in solution. The “bended” SARS-CoV Nsp8-chain (*blue*) taken from the Nsp(7+8) complex (pdb code 2AHM) is truncated at the N-terminus, because this part of Nsp8 was not defined by the electron density [98]; therefore the native Nsp8 N-terminus is assumed to allow Nsp7 to be located even closer to the C-terminus of Nsp8 as observed for the complex of mature SARS-CoV Nsp7 and Nsp8. The interface of the HCoV-229E pp78 dimer is predicted to involve the C-terminal domain of Nsp8, as previously observed for the pp710 dimer of HCoV-229E (chapter 3.2.3.2/3.2.3.3). According to the rigid-body model the dimerization of HCoV-229E pp78, based on the proposed symmetry and shape, as calculated by SASREF, is not expected to involve intermolecular Nsp7-Nsp7 interactions at the dimerization interface, although one pp78 molecule could rotate around the longitudinal axis to get the Nsp7 domains in the pp78 dimer in contact with each other. Moreover, this Nsp7-Nsp7 dimerization at the dimer interface of pp78 would provide an additional stabilization of the pp78 dimer. Furthermore, dimerization of mature SARS-CoV Nsp7 was previously indicated in solution (chapter 3.11.3) and agrees with a dimerization of mature FCoV Nsp7 in solution detected by covalent cross-linking [96].

The dimerization of mature HCoV-229E Nsp(7+8) appeared to be mediated by the C-terminus of Nsp8 as well, forming a largely elongated molecule as evident from the *ab initio* model (fig. 65G). Comparison of HCoV-229E pp78 and the non-covalent complex of mature Nsp7 and Nsp8 (fig. 65, F/G) revealed that the complex is further elongated upon processing by main protease, well explaining the increase of the maximum diameter observed in a SAXS experiment that followed the cleavage reaction (fig. 65A). However, in summary, the C-terminal globular domain of Nsp8 still appeared to be essential for dimerization even after complete maturation of Nsp7 and Nsp8.

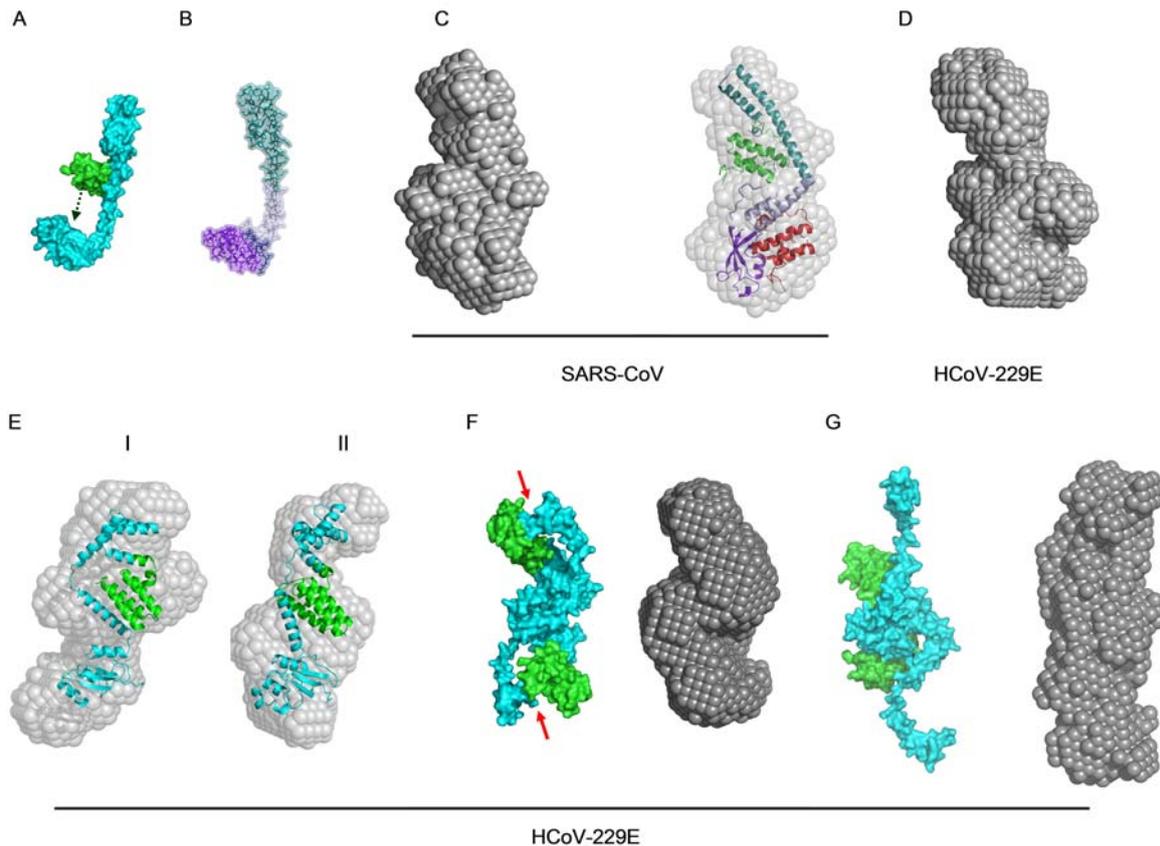


Figure 65. Comparison of DAMMIF *ab initio* models of SARS-CoV and HCoV-229E pp78 as well as Nsp(7+8) of HCoV-229E obtained by SAXS and interpretation by high-resolution structures. **(A)** Rigid-body model based on scattering data of monomeric HCoV-229E pp78 using the rigid full-length SARS-CoV Nsp8 structure taken from the Nsp(7+8) complex (pdb code 2AHM, chain H) and Nsp7 (pdb code 1YSY) with the restrain of a covalent connection. These prerequisites did not allow a sufficient interpretation of the data that suggested a translation of Nsp7 (*black* arrow). **(B)** Nsp8 was segmented into four individual fragments to simulate the sites of flexibility: amino acids 1-73 (fragment 1, *blue-green*), 74-113 (fragment 2, *grey*), 114-124 (fragment 3, *blue*), 125-195 (fragment 4; *purple*). **(C)/(D)** Low resolution *ab initio* structural models of the pp78 homologues of SARS-CoV and HCoV-229E are presented for comparison. The rigid-body model of SARS-CoV pp78 calculated using four Nsp8 fragments that were independently positioned by SASREF is fitted into the corresponding *ab initio* model in panel C. The Nsp7 molecule (*green*) was manually positioned close to the Nsp8 C-terminus, where the calculated low-resolution envelope provided sufficient space (*red*). **(E)** Two possibilities to interpret the *ab initio* model of monomeric HCoV-229E pp78 by a homology model obtained from the I-Tasser modelling server (RMSD: 13.2 ± 4.1 Å, C-score: -3.01, TM-score: 0.37 ± 0.13). Superposition II was favoured because Nsp7 (*green*) can occupy a part of the *ab initio* model close to the Nsp8 (*blue*) C-terminus. **(F)** Rigid-body model of dimeric HCoV-229E pp78. Truncated and bended Nsp8 was used as a template (*blue*, chain E, pdb code 2AHM), together with Nsp 7 (*green*, pdb code 1YSY). Cleavage sites connecting Nsp7 and Nsp8 are labelled by *red* arrows. **(G)** Rigid-body model of the mature Nsp(7+8) dimer of HCoV-229E obtained by SASREF (88 cycles of refinement, penalty 0.0001, best χ -value of the scattering fit: 0.90532, considering scattering vector values $0.01 < s \leq 0.24$ Å⁻¹). The averaged low resolution *ab initio* model of mature HCoV-229E Nsp(7+8) presented for comparison displayed an increased maximum diameter ($D_{\max} = 14.5$ nm) compared to dimeric HCoV-229E pp78 ($D_{\max} = 11.4$ nm). The mean NSD of 10 DAMMIF models is 1.124 ± 0.044 .

3. Results and discussion

To compare the stability of the fold of HCoV-229E pp78 with the mature complex consisting of Nsp7 and Nsp8 CD-spectroscopy was applied, having in mind that previously significantly different low-resolution structures were presented (fig. 65). At 30 °C, a far-UV CD spectrum with two minima characteristic for a high α -helix content is observed, reflecting the expected fold of Nsp7 and Nsp8 [96; 98]. However, the similar temperature stability of HCoV-229E pp78 in the pre-cleavage state compared to Nsp(7+8) in the post-cleavage state (fig. 66) indicated that the overall secondary structure and the sum of inter-domain interactions are not significantly altered. The slight difference could originate from differences in the aggregation behaviour, the different purification protocols or theoretically a minor difference of the Nsp7-Nsp8 interaction site in the polyprotein state.

Assuming that pp78, pp79 and pp710 have different equilibrium positions of oligomeric states, this could have impact on the melting temperature as well as an altered structure of the protomers. For example, as cross-linking with BS2G suggested, SARS-CoV pp710 is monomeric at a higher percentage compared to SARS-CoV pp78. This would correlate with the higher thermal stability of SARS-CoV pp78. Applying purified monomeric and dimeric SARS-CoV pp710 to CD spectroscopy there was no significant difference, even though higher oligomers appeared to be marginally more stable (data not shown). This observation might already be close to the resolution limit of relatively similar melting temperatures. Nonetheless, for instance for the L-tryptophan biosynthesis regulator (TRAP) [174] tight specific oligomerization and ligand-binding was correlated with significantly higher melting temperatures determined by CD spectroscopy. Additionally, this was observed for the reduced thermal stability of M^{pro} R298A ($\Delta T_m = 2.7$ °C; $\Delta G_{\text{dissociation}} = 4.2$ kcal mol⁻¹ (PISA,EMBL-EBI, calculated for pdb code 1Q2W)) compared to the wild type known to form a relatively tight dimer [53; 60].

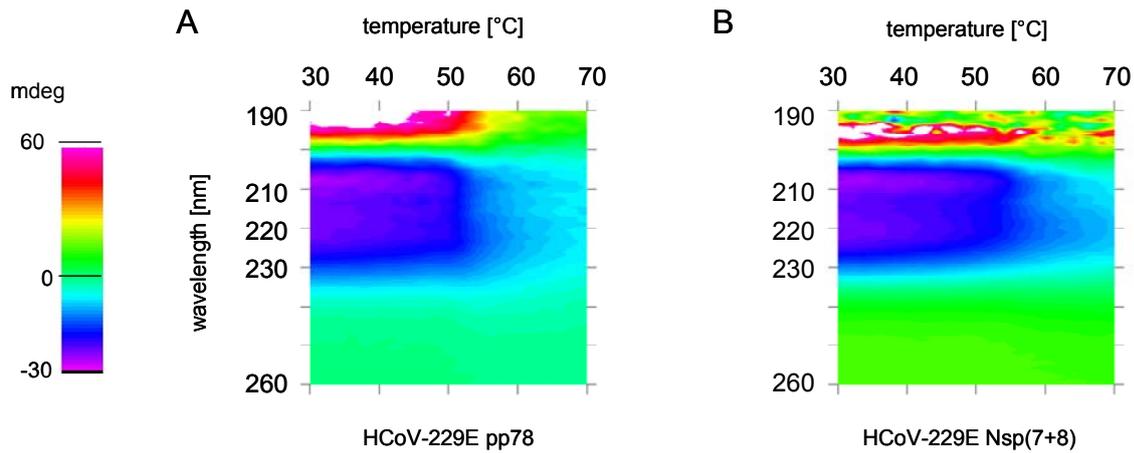


Figure 66. Investigation of the thermal stability of HCoV-229E pp78 (A) and Nsp7+8 (B) applying CD spectroscopy. Temperature interval scanning experiments were performed at temperatures ranging from 30 °C to 70 °C. The typical shape of the spectrum at 30 °C involving two minima characteristic for high α -helix content is conserved. The highest loss of absolute ellipticity of the minima is observed in a similar range close to 60 °C, even though the stability of the mature Nsp(7+8) complex determined at 220 nm is a little bit higher. The original two minima observed at around room temperature tend to fuse to form one resulting flat minimum above 65 °C, after the collapse of most of the helix structure.

To obtain indication for putative ligand binding sites at the C-terminus of SARS-CoV Nsp8, small-molecule docking was applied (fig. 67). The FT map server [167] identifies putative ligand binding sites on the surface of a protein by screening small organic molecules of a default library in terms of ligand binding energy. Applying this algorithm to SARS-CoV Nsp8 (pdb code 2AHM, chain H), FTmap identified one “hot-spot” for ligand binding at a helix in the almost globular C-terminal domain of the protein (*red circle*). This observation could not only be indicative for RNA/nucleotide binding, but also for an interaction with small regulatory ligands *in vivo*. This C-terminal domain was previously considered to be a favourable protein binding site and probably part of the Nsp8 dimerization interface (fig. 65F/G). Interestingly, the C-terminal domain of Nsp8 exhibits structural similarities with hnRNP (pdb code 2UP1), i.e. a human single strand nucleic acid-binding protein [94]. Therefore, the C-terminal domain could be involved in catalysis or at least RNA binding, which has been reported for the hexadameric complex of mature SARS Nsp(7+8). This is supported by the suggestion that an N-terminally truncated isoform of Nsp8 is able to preserve nucleic acid binding and exhibits a regulatory effect on the hexadameric Nsp(7+8) complex. However, as already mentioned, such an oligomeric state was not detected in the course of this project.

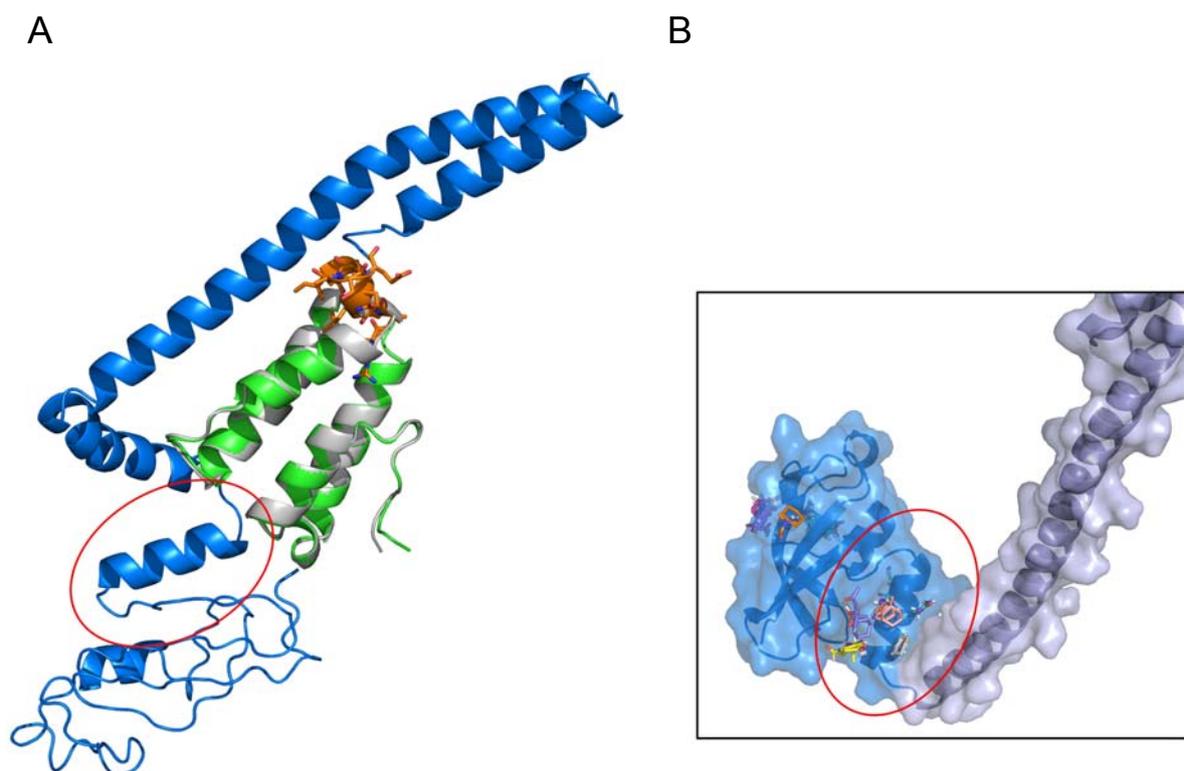


Figure 67. (A) Homology model of SARS-CoV pp78 (Nsp7: *green*, Nsp8: *blue*; calculated by the I-Tasser modelling server) superposed with SARS-CoV Nsp7 (pdb code 2AHM; *grey*). Remarkably, Nsp7 is located close to the epitope reported for the SARS-CoV Nsp(7+8) complex. A sharp turn of the N-terminal Nsp8 α -helix enabled this interaction. Notably, this model is well superimposable with the *ab initio* model of monomeric HCoV-229E pp78 (fig. 65E). Side chains of residues P5-P5' of the main protease cleavage site connecting both Nsp8s are displayed in *orange* stick representation. (B) Grid-based small-molecule docking applying the FTmap algorithm. The C-terminal domain of Nsp8 (*marine blue*; pdb code 2AHM, chain H). Different docking events are located at the highlighted helix of Nsp8 (red circle) perhaps correlating to a binding site of a small organic ligand *in vivo*.

In conclusion, the Nsp7 epitope appears to be highly similar within the mature Nsp(7+8) complex and the precursor pp78. However, upon cleavage of pp78 by the main protease the Nsp8 N-terminus becomes more elongated and stretched, presumably with some flexibility, resulting in a significantly increased D_{\max} of the mature HCoV-229E Nsp(7+8) complex compared to pp78. Within pp78, the N-terminal domain of Nsp8 is assumed to bridge the distance to the putative Nsp7 binding site. The resulting bended conformation of Nsp8 well fits the corresponding part of the *ab initio* model of SARS-CoV and HCoV-229E pp78 and explains its slightly reduced R_G compared to the stretched geometry of mature Nsp(7+8) (fig. 65G; see also pp89 in fig. 36). It is also concluded that the oligomerization of pp78 is widely unaffected by cleavage by the main protease, a SARS-CoV pp78 dimer and a heterotetrameric 2:2 interaction of SARS-CoV Nsp7 and Nsp8 are most abundant.

3.11.3. SARS-CoV Nsp7 forms a dimer in solution

To investigate the impact of a hydrophobic patch located within the N-terminal α -helix of Nsp7 on the interaction with Nsp8 the double mutant Nsp7 V11E/L14R of SARS-CoV pp78 was designed. The side chains of these two residues point in different orientations (pdb code 2AHM) [98; 100]. Parts of the sequence ¹¹VVLL¹⁴ of SARS-CoV Nsp7 are surface exposed, while intramolecular hydrophobic interactions are also formed within the Nsp(7+8) high-resolution structure [98]. Putative intermolecular hydrophobic interaction of the helices of two Nsp7 molecules could also influence the oligomerization of pp78. Introducing large polar residues may alter the peripheral hydrophobicity of the potential binding epitope, resulting in an altered hydrophobic interaction involving the N-terminal helix of Nsp7. Furthermore, V11 is known to hydrophobically interact with S20 of another Nsp7 molecule in the SARS-CoV Nsp(7+8) complex, stabilizing an Nsp7 dimer, while each Nsp7 molecule is additionally interacting with one Nsp8 molecule to form a building block of the Nsp(7+8) hexadecamer [98].

The obtained double mutant V11E/L14R of SARS-CoV pp78 was purified by affinity chromatography to homogeneity (fig. 68B). Interestingly, the protein yield of the double mutant obtained from *E. coli* cells was significantly increased compared to the wild type protein, which could correlate with improved protein solubility due to locally reduced surface hydrophobicity. After cleavage of the precursor pp78 by SARS-CoV M^{pro}, calibrated SEC showed two main peaks (fig. 68, dark *green* chromatogram). As shown by non-reducing SDS-PAGE analysis mature Nsp7 V11E/L14R was successfully separated from mature Nsp8 (fig. 68B). The approximated molecular weight of mutated mature Nsp7 optimally fits a dimeric state ($V_E = 91.5 \text{ ml} \rightarrow 17.2 \text{ kDa}$). A minor amount of remaining pp78 co-localized with mature Nsp8 at a retention volume between 75-80 ml, indicative for an interaction of pp78 and Nsp8 (fig. 68A, see *red* chromatogram compared to peak A). Applying fractions of peak A to SAXS, elongated particles at a molecular weight below 70 kDa were reconstructed (data not shown). The mutated precursor pp78 still dimerized in solution (experimentally calculated $MW_{pp78} = 62 \text{ kDa}$) in a concentration-dependent manner (fig. 68A). Interestingly, additional larger oligomers previously observed by SEC with pp78 wild type (fig. 53A) were absent. The observation that a major amount of SARS-CoV Nsp7 V11E/L14R was separated from Nsp8 by SEC is an indication that the corresponding conserved hydrophobic patch within α -helix 1 has a minor contribution to the interaction. In spite of the mutation, CD spectroscopy displayed a conserved α -helix-rich fold of Nsp7 (fig. 71), in good agreement with known high-resolution structures [98; 100].

3. Results and discussion

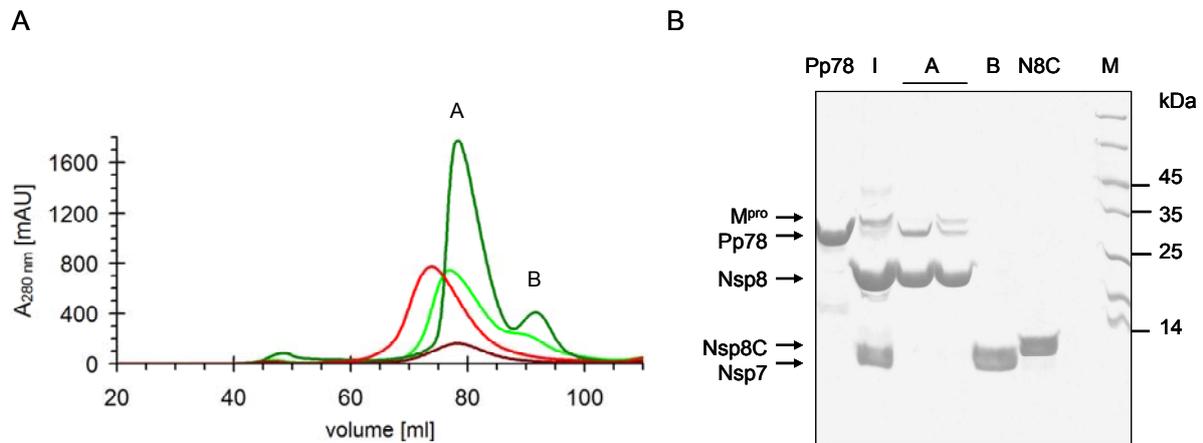


Figure 68. Proteolytic cleavage of mutated SARS-CoV pp78 by SARS M^{pro} . **(A)** Different stages of the processing reaction of pp78 V11E/L14R were applied to SEC (HiLoad 16/60 Superdex 200). The excess of pp78 towards M^{pro} was 7-fold. *Red* chromatogram: 6 mg pp78 V11E/L14R, *brown*: 1 mg pp78 V11E/L14R *light green*: 60% of pp78 V11E/L14R cleaved by M^{pro} , *dark green*: 90% cleaved, as judged by SDS-PAGE. All samples were buffered in 30 mM NaH_2PO_4 , pH 7.6, supplemented with 250 mM NaCl. **(B)** Individual fractions referring to the *dark green* chromatogram were analyzed by SDS-PAGE. Pp78: Purified pp78 V11E/L14R after affinity chromatography; I: Column input (pp78 DM after M^{pro} cleavage); A, B: Fractions of respective SEC peaks; Nsp8C: C-terminal domain of Nsp8 as a molecular weight standard for 12.8 kDa.

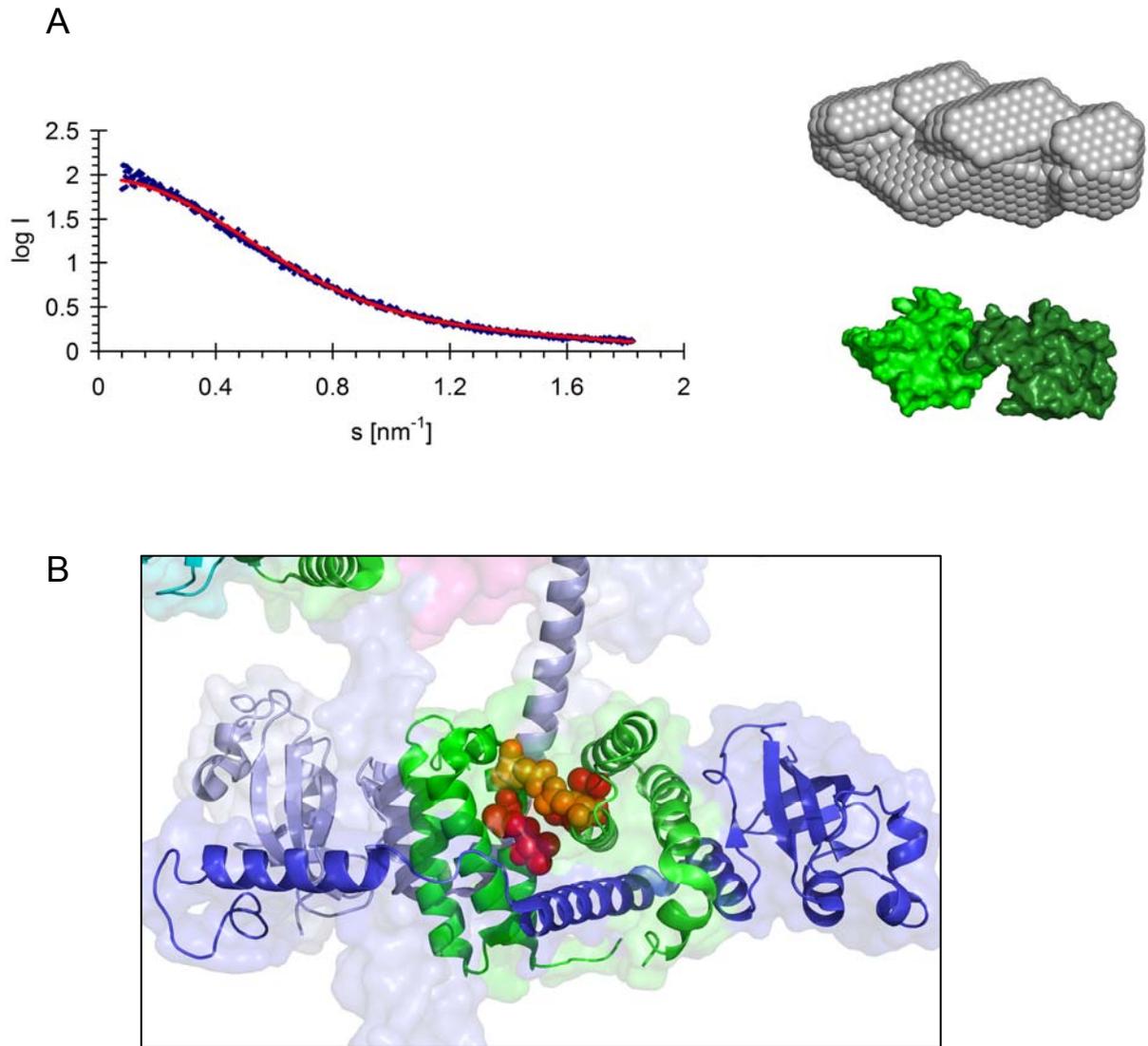


Figure 69. **(A)** SAXS investigation revealed an elongated dimer of Nsp7 V11E/L14R: Scattering data at protein concentrations of up to 2.8 g l^{-1} did not indicate a concentration-dependency of the particle size. Optimized fit function of the rigid-body model (*red*) created by SASREF considering P2 symmetry, i.e. a dimer, well fits the experimental scattering data (*blue*). The *rigid-body* model of Nsp7 is depicted on the right (template: FCoV Nsp7 extracted from pdb code 3UB0). The model is characterized by a chi-value of 1.1972 after 58 cycles (penalty 0.0016; default penalty scores and weight) using a maximum scattering vector s of 0.1826 nm^{-1} (1707 data points). **(B)** The Nsp7 dimer (*green*) is stabilized by hydrogen bonds of Q18 (*orange spheres*). The Nsp7-Nsp7 and Nsp7-Nsp8 (*blue*) interaction within the Nsp(7+8) hexadecameric complex (pdb code 2AHM) is visualized. Nsp7 Q18 is labeled by *orange spheres*; V11 and L14 are displayed as *red spheres*.

3.11.4. C-terminal domain of HCoV-229E Nsp8

To investigate a putative involvement of the Nsp8 C-terminus in dimerization of pre-processed polyproteins as suggested by SAXS (chapter 3.11.2), Nsp8 D104-Q195 was cloned into pASK IBA 16 providing an ompA-tag for exporting the protein to the periplasmic space. The N-terminus elongated by a strep-tag and a six amino acid linker (SGGGGG) may mimic a part of the flexible partly helical Nsp8 N-terminus. The protein was purified by Strep Tactin affinity chromatography. The ompA-tag was cleaved during secretion to the periplasmic space while the strep-tag was completely cleaved off by tobacco etch virus protease (Sigma-Aldrich) (fig. 70, SDS-PAGE lane 1/2). In the absence of the tag, chemical cross-linking with glutaraldehyde, linking lysine residues, reveals a dimeric state under reducing conditions along with monomeric protein (fig. 70, SDS-PAGE, lane 3/4), indicating that the C-terminus of Nsp8 could mediate the dimerization of native Nsp8 or even a larger polyprotein.

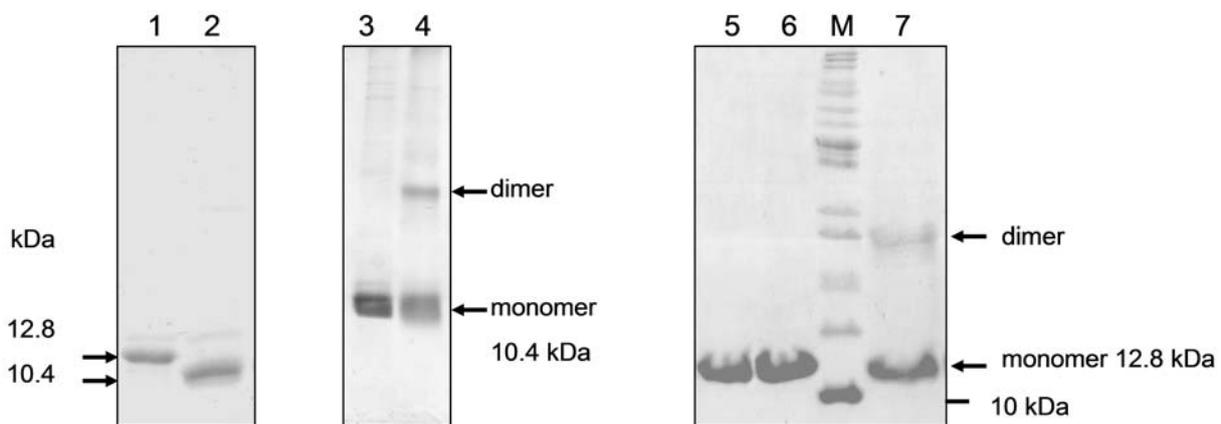


Figure 70. SDS-PAGE gel of purified **C-terminal domain of HCoV-229E Nsp8 (D104-Q195)**. Lane 1: Uncleaved, after Strep-Tactin affinity column purification, reducing conditions; lane 2: Purified after cleavage of the strep-tag with TEV protease (50:1, at 4 °C for 24 h), reducing; lane 3: Control prior to cross-linking, reducing; lane 4: After incubation with 0.05% glutaraldehyde for 30 min at 20 °C. The reaction was terminated by adding Tris to a final concentration of 200 mM. Beside the monomeric state, a specifically cross-linked dimer is revealed, reducing; lane 5: Purified T123C mutant, reducing conditions. The Nsp8 C-terminus was covalently cross-linked by glutaraldehyde revealing a specific dimerization; lane 6: Wild type, non-reducing conditions; lane 7: Nsp8 C-terminus T123C, non-reducing, stored for 2 days after purification. M: Marker.

Recently, domain swapping of the Nsp8 C-terminal domain was suggested to cause dimerization within a complex of mature Nsp7 and mature Nsp8 of HCoV-229E by a preliminary high resolution crystal structure (*Yibei Xiao*, University of Lübeck, personal communication). To investigate if particularly a hinge loop (¹¹⁵VPLSVIPATSAARLVV¹³⁰) of the globular C-terminus of Nsp8 is involved in dimerization of isolated Nsp8 or if domain swapping of Nsp8 occurred at the dimer interface in solution, the mutation T123C was introduced to the hinge loop of HCoV-229E Nsp8 C-terminal domain. If the introduced surface-exposed cysteine is located at the dimer interface, it is supposed to come in close proximity to the introduced cysteine of another Nsp8 C-terminus molecule

during dimerization, resulting in the formation of a disulfide bond under non-reducing conditions. Wild type Nsp8 D104-Q195 contains only one native cysteine residue that is not exposed to the surface. Since secretion into the periplasmic space should favour the formation of disulfide bonds within the *E. coli* cells, mutated protein Nsp8 was cloned as an ompA-tag fusion protein. The protein was purified by Strep-Tactin affinity chromatography (fig. 70B, lane 5). To verify that the truncated and mutated variant of Nsp8 is well folded, CD spectroscopy was performed in comparison to other Nsps (fig. 71). The spectra demonstrated that the C-terminal domain of Nsp8 was rich in β -sheet and random coil structure, while full-length Nsp8 was largely composed of α -helical structure due to its N-terminus. A fraction of random coil is in agreement high-resolution structures of the Nsp(7+8) complex [96; 98], even though a higher content of β -sheet would be expected.

The α -helix-rich fold of HCoV-229E Nsp8 was indicated by two predominant minima of the ellipticity around 208 nm and 222 nm that were absent in the spectra of Nsp8 C-terminus wt and Nsp8 C-terminus T123C.

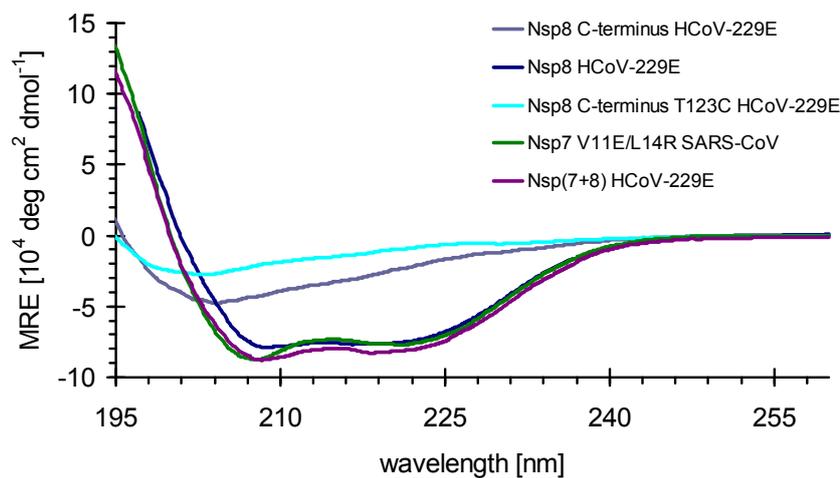


Figure 71. Far-UV CD-spectra ranging from 260 to 195 nm wavelength, collected in sodium phosphate buffer. The spectrum of Nsp7 V11E/L14R (*green*) is characteristic for an α -helix-rich protein, confirming its proper native fold. The complex of full-length mature Nsp(7+8) of HCoV-229E (*wine red*) and full-length Nsp8 of HCoV-229E (*dark blue*) showed a comparable α -helix-rich fold. Wild type Nsp8 C-terminal domain (*grey-blue*) and its T123C mutant (*cyan*) showed a relatively low absolute ellipticity and a single minimum around 204 nm and 201 nm, respectively, indicating a fold dominated by random coil and β -sheet fold, but a very low α -helix content, in agreement with reported crystal structures [98], even though folding artefacts due to the truncation or mutation could not be not excluded.

To probe the close proximity of the introduced cysteine residues within a dimer of the Nsp8 C-terminal domain, non-reducing SDS-PAGE of the purified T123C mutant of Nsp8 D104-Q195 was performed immediately after purification as well as after two days incubation under oxidising conditions (fig. 70B, lane 6 and 7). In contrast to the wild type protein, the mutant formed a covalent dimer during incubation, representing a first indication that the additionally introduced cysteine

3. Results and discussion

residue formed a specific link with the cysteine of another molecule of the Nsp8 C-terminal domain. Additional covalent oligomers of mutant or wild type were absent in non-reducing SDS-PAGE.

The indicated covalent dimerization was further analyzed by analytical size-exclusion chromatography (HiLoad 16/60, Superdex 200, GE Healthcare). A single protein peak was observed for pure Strep-tagged Nsp8 C-terminus wild type at an elution volume of 90.9 ml, corresponding to a molecular weight of 29.4 kDa that represents a dimeric state of the protein (fig. 72A). Under identical non-reducing conditions, a comparable sample of the corresponding T123C mutant eluted slightly earlier than the wild type at a volume of 88.7 ml from the SEC column, corresponding to a molecular weight of 33.6 kDa. This indicated the presence of trimeric protein, consisting of a dimer covalently linked at the mutated site that interacted non-covalently with further molecules. After removing the strep-tag under reducing conditions and further incubating the mutant for 24 h at 6 °C, SEC displayed a similar dimeric state ($V_E = 92.5$ ml). The elution is slightly shifted due to the removal of the tag compared to the strep tagged wild type. Significant amounts of higher oligomeric states have not been detected for all samples (fig. 72A).

Mutated, strep-tagged C-terminal Nsp8 (fig. 72A *light blue* chromatogram) was additionally investigated by SAXS technique, resulting in $R_g = 2.57 \pm 0.01$ and $D_{max} = 9$ nm. Based on the scattering data in solution, a low-resolution *ab initio* model was calculated considering P2 symmetry for a dimeric state, which is depicted in fig. 72B. As already indicated by SEC, the approximated molecular weight of the Nsp8 C-terminus T123C mutant of 30 kDa, based on the *ab initio* model calculation by DAMMIF, appeared to be slightly increased compared to that expected for a dimer ($MW_{dimer} = 26$ kDa). In conclusion, this truncated Nsp8 "model system" to investigate dimerization of Nsp8 could be either altered in oligomerization due to the truncation at the N-terminus or display an artificial folding in the absence of the N-terminus in *E. coli*, as indicated by CD spectroscopy.

The oligomerization might also be a matter of the concentration, considering the hypothesized involvement of domain-swapping. Moreover, as observed by non-reducing SDS-PAGE the relative fraction of covalently linked dimers of the Nsp8 C-terminus T123C which were presumed to be indicative for the potential dimer interface, is rather low (fig. 70 lane 7). Finally, it still has to be considered that the introduced surface-exposed cysteine residue could result in an unspecifically formed dimer. However, the dimeric state of mutated Nsp8 C-terminus appeared to be predominant in solution. Moreover, the calculated *ab initio* model is well superimposable with the corresponding fragment of the *ab initio* model of full-length pp78 (fig. 72B, see also fig. 65F for comparison), indicating a relevance of this putative dimeric Nsp8 C-terminus molecule in the dimerization of pp78.

To verify the identity of the disulfide bond observed by non-reducing SDS-PAGE of the Nsp8 C-terminus T123C mutant, the covalent dimer (fig. 70, lane 7) was cut from the gel and digested by trypsin for ESI-Q-TOF mass spectrometry (fig. 73). At sufficient fragment ion coverage, mass

spectrometry detected the expected peptide containing a disulfide bond formed by C123 of two different protomers within the dimer. For comparison, dimeric pp710 with the same cysteine mutation (HCoV-229E pp710 T206C) was purified, stored under aerobic conditions, and applied to non-reducing SDS-PAGE one week after cell lysis. A corresponding dimer band was detected and subsequently applied to mass spectrometry, revealing the existence of the C206 intermolecular disulfide bond that linked two identical peptides of different pp710 molecules as well (fig. 73).

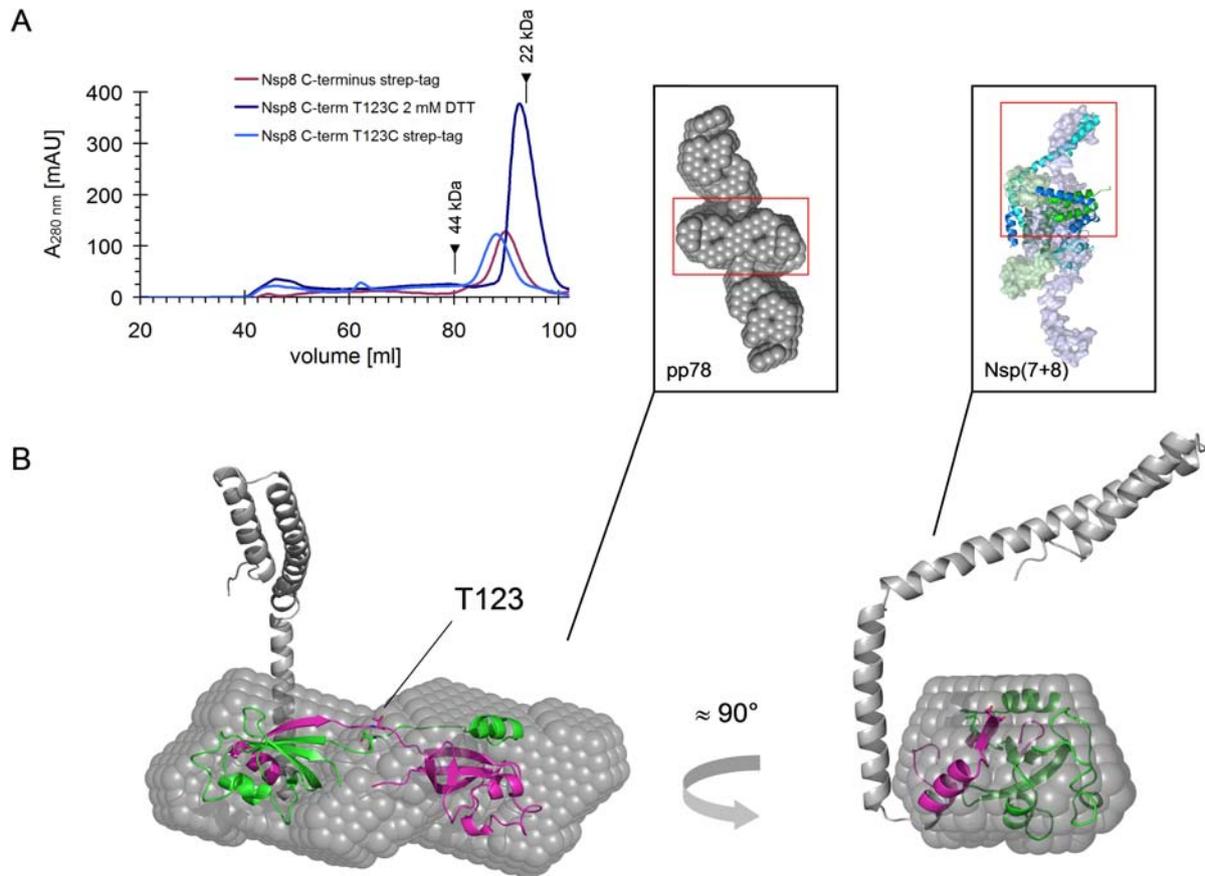


Figure 72. Dimerization of HCoV-229E Nsp8. (A) Analytical SEC (HiLoad 16/60 Superdex 200) of Nsp8 C-terminus wt and T123C mutant in 200 mM NaCl and 40 mM Na₂PO₄, pH 7.6, indicative for a dimeric state of the T123C mutant under reducing conditions. Moreover, the strep-tagged C-terminal domain of Nsp8 wt forms a dimer in solution. The putative dimeric state is not affected by the N-terminal strep-tag of the protein. (B) An *ab initio* model calculated based on the SAXS data obtained for HCoV-229E Nsp8 C-terminus T123C mutant was superposed with a dimer of Nsp8 fragments (*green/purple*) obtained from a preliminary high resolution structure of HCoV-229E Nsp(7+8) heterotetramer (*Yibei Xiao*, University of Lübeck, personal communication, unpublished data). Both Nsp7 subunits and the N-terminus of one Nsp8 protomer were removed for clarity. The T123 mutation site is highlighted. The Nsp8 C-terminus T123C mutant is characterized in solution by $R_g = 2.27 \pm 0.04$ nm and $D_{max} = 9$ nm. The mean NSD of 10 *ab initio* models was 1.143 ± 0.266 . A single *ab initio* model of pp78 V11E/L14R is displayed for comparison (left inset). The *ab initio* model of the mutated Nsp8 C-terminus fits well to a domain highlighted in *red*. Including the stretched N-terminus of Nsp8 and Nsp7 of the preliminary heterotetrameric high-resolution structure, this structure is well superimposable with the *ab initio* model of HCoV-229E Nsp(7+8) (enframed image on the right; Nsp8 in *blue*, Nsp7 is coloured *green*).

3. Results and discussion

In summary, the obtained results suggested that C-terminal Nsp8 T123C as well as HCoV-229E pp710 T206C (T123 in Nsp8 and T206 in pp710 represent identical residues) covalently dimerized in a comparable way via the introduced cysteine residue. Consequently, the dimerization of Nsp8 appeared to be conserved during maturation of HCoV-229E pp710. Moreover, considering the *ab initio* models of pp710 this dimerization site was suggested to be an essential intermolecular interaction within the investigated dimer of HCoV-229E pp710 (fig. 15).

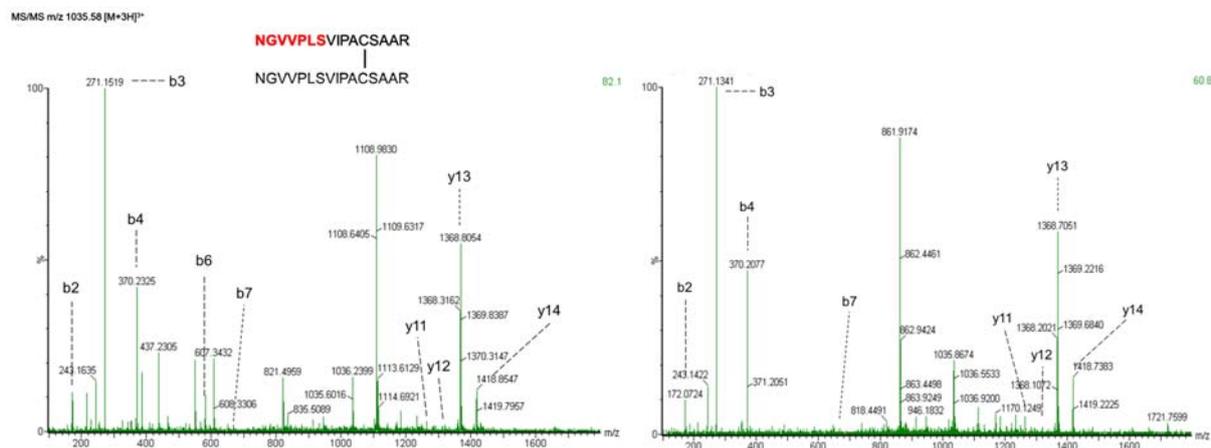


Figure 73. The dimeric state of HCoV-229E Nsp8 C-terminal domain T123C (left) and HCoV-229E pp710 T206C as purified by SDS-PAGE, digested by trypsin and applied to ESI mass spectrometry. As evident from the MS/MS spectra depicted above both peptide samples contain a dimer of the mutated peptide containing a cysteine. Therefore, both mutated proteins are capable of dimerizing via a disulfide involving the introduced cysteine of both protomers. Both spectra show the same fragment ion coverage of the b- and y- ion series, i.e. fragment ions corresponding to the sequence NGVVPLS are labelled in both spectra. The coverage and the expected disulfide-bonded product in both spectra are depicted above the left spectrum.

The previous chapters contain fundamental insights into the interaction states of Nsp7 and Nsp8 in comparison to the precursor pp78. Overall, assuming that the dimerization interface is conserved for the pre-processed pp78, Nsp7-9 and Nsp7-10, the putative dimerization via intermolecular Nsp8-Nsp8 interactions could be a common regulatory mechanism to mask a ligand-binding site upon dimerization. In this context, an increasing concentration of polyproteins translated during replication that results in a concentration-dependent increase of the dimeric state might influence a signalling pathway. This process could further induce termination or down-regulation of RNA polymerase activity. Concentration-dependent oligomerization, as it was observed in SAXS experiments with pp78, is a common characteristic among coronaviral replicase proteins [102; 51]. For instance, after a few translation events in an early stage of replication pp710 is predominantly monomeric, while the dimeric state is subsequently continuously increasing which may display a significant regulatory potential to influence the RTC composition. Particularly, having in mind that the Nsp8-C-terminus that was suggested to be involved in dimerization is indicated to bind small organic ligands (fig. 66B),

where the respective binding sites might be masked or altered in the dimeric state of Nsp8.

Additionally, a monomer-dimer equilibrium was suggested for mature Nsp9 [102], where the monomeric state is favoured at concentrations below approximately 1.5 g l^{-1} , as further verified by native mass spectrometry data in this study. In contrast to those examples, the monomer-dimer equilibrium of SARS-CoV M^{Pro}, which is required in an early stage of replication to release different mature replicase Nsps, shifts to the active dimeric state at comparably low concentrations, supported by SEC results presented previously (fig. 44) and by reports demonstrating a tight dimerization [60]. The observation that all coronaviral main proteases (3CL proteases) require dimerization to be catalytically active [199; 214; 215] is in contrast to the related picornavirus 3C protease that does not require dimerization. However, the molecular basis for the different activity is not understood so far.

In terms of catalytic activity the surface region around D50 and D52 of Nsp8 was predicted to be critical for RdRp activity, associated with RNA binding [97]. However, this region was not identified as a binding site for small molecules of the default library of FTmap in the docking approach (fig. 66B), which may allow to deduce a nucleotide binding site. This could be explained by the suggested requirement of a conformational change of Nsp8 to become catalytically active [96]. However, SAXS data did not indicate an obvious tertiary structure modification within Nsp8 upon dimerization. Parts of the helical N-terminus of Nsp8 are predicted to be flexible or disordered, as indicated by the online server *fold index*. However, this is particularly hard to interpret by the SAXS technique due to the limited resolution. Therefore high-resolution data is required or alternatively a method to probe changes in the tertiary structure upon nucleic acid binding, e.g. fluorescence resonance energy transfer measurements. Beyond that, the structural variability, i.e. the fold and the oligomerization, of mature Nsp8 in solution is not well understood. Assuming that the α -helix-rich N-terminus retains its secondary structure as supported by CD-spectroscopy, different tertiary structures of the elongated helical N-terminus of Nsp8 have been considered, including a stretched variant [98] visualized in figure 65B and a variant with a bended N-terminus by approximately 90° [98]. Moreover, a tightly bended conformation that enables the covalently linked Nsp7 to interact with Nsp8 relatively close to its globular C-terminus was proposed in this study. The low resolution shape of the mature Nsp(7+8) of HCoV-229E as well as of pp89 indicated a preferably straight and stretched geometry of approximately the first 100 N-terminal amino acids of Nsp8. Thus, the geometry of the Nsp8 N-terminus proposed for pp78 requires the already reported flexibility between segments of the extended α -helix [96; 98] and perhaps some flexibility of more than 40 amino acids at the N-terminus that are not defined in the structure of the “bended” Nsp8 chains of the hexadecameric SARS-CoV Nsp(7+8) complex (pdb code 2AHM). Nsp7 can keep the binding epitope known for the mature Nsp7-Nsp8 interaction. After proteolytic processing, the maximum diameter of the mature Nsp(7+8)-complex increased, indicative for a small change in the tertiary structure of the Nsp8 N-terminus into the more stretched variant. Major structural rearrangements during processing of pp78 were not predicted.

3. Results and discussion

This model of Nsp(7+8) interaction was independently supported by SAXS and mass spectrometry data of pp710 that resulted in the proposal of a comparable Nsp7-Nsp8 interface also in pp710. Consequently, the Nsp(7+8) interface appears to be well conserved in various polyprotein precursor proteins. The pre-formation of the Nsp(7+8) epitope might be required in an early stage of the nascent polyprotein to stabilize both proteins and gain efficient enzymatic activity already at the polyprotein stage. This suggestion could explain the observation that exchanging the order of the Nsp7 and Nsp8 gene fragments of ORF1a resulted in a non-viable phenotype of MHV [107].

The slightly increased values of R_h and R_g determined in chapter 3.2.3 for SARS-CoV pp710 compared to its homologue of HCoV-229E (see table 25) can also be well explained by an altered, probably taxon-related conformation of the Nsp8 N-terminus. However, it is not excluded that during cleavage of pp78 an essential interaction epitope is uncovered that does not significantly effect the binding of Nsp7 to the Nsp8 C-terminus. A better understanding of the regulatory impact of the Nsp8 geometry requires a fundamental knowledge about Nsp8 interaction partners and the corresponding binding epitopes that are unknown so far. The dimerization of Nsp8 via its C-terminal domain appears to be equally possible for the Nsp8 domain in the state of pp78 during proteolytic maturation.

The folding of the strongly N-terminally truncated and dimeric variant of Nsp8 might be influenced by the absence of Nsp7, since all high-resolution crystal structures that are available so far for comparison of the Nsp8 structure in solution contain Nsp7 [96; 98; 99]. Even though a slightly altered folding of Nsp8 due to the truncation is not excluded, the dimerization of the Nsp8 C-terminal domain in solution was demonstrated by SEC and covalent cross-linking. Moreover, the formation of a disulfide bond of C123 with C123 of another Nsp8 molecule was verified further supporting the hypothesized interface. It was further observed that a mutant of SARS-CoV Nsp7 is dimeric and additionally Nsp7 is in close contact to the dimer interface of dimeric SARS-CoV and HCoV-229E pp78 according to the SAXS data. Therefore, dimerization of Nsp7 even within the polyprotein state is conceivable, but remains to be verified by a high-resolution structure of pp78.

3.12. HCoV-229E Nsp9

The structures of HCoV-229E Nsp9 and its C69S mutant, as well as of SARS-CoV Nsp9 were previously solved by X-ray crystallography [101; 102]. The C69S mutant was investigated to verify the impact of an intermolecular disulfide-bond within the wild type molecule on the oligomerization of Nsp9.

To analyze the dimerization of HCoV-229E Nsp9 in solution and to investigate its additional oligomerization, SAXS and native MS experiments have been performed. HCoV-229E Nsp9 wild type and the Nsp9 C69S mutant were obtained by proteolytic processing of pure HCoV-229E pp710 wild

type and pp710 C347S mutant, respectively, with SARS-CoV M^{Pro}. The pp710 C347S mutant was purified followed the established protocol for wild type pp710, yielding similar amounts of protein. After SEC chromatography of the processed protein mixture (HiLoad 16/60 Superdex 200), mature Nsp10 co-localized with mature Nsp9 wt or with the Nsp9 mutant, respectively, in the elution fractions (fig. 74A; see also chapter 3.4.1). To isolate Nsp9 wt/mutant, mature Nsp10 was precipitated by directly adding 10% of the total volume of 10 mM ZnCl₂ to 100 μM protein in solution. A final concentration of 20 mM CoCl₂ had a similar effect. After centrifugation, a pure and monodisperse solution of Nsp9 was obtained (fig. 74A). HCoV-229E Nsp9 C69S obtained by this method formed crystals in a hanging drop vapour diffusion experiment with a reservoir solution containing 2 M ammonium sulphate and either 0.1 M Bis-Tris pH 5.5 or 0.2 M NaCl and 0.1 M sodium cacodylate pH 6.5 which is close to conditions that were already reported [101].

Alternatively, to purify Nsp9, Nsp10 was removed by incubating the mixture of mature Nsp9 and Nsp10 with Ni-NTA resin that weakly binds Nsp10, although it did not contain a his₆-tag. The HCoV-229E Nsp9 wt gene was additionally cloned into pRSETA vector and separately expressed in *E. coli* cells, resulting in approximately 10-20 mg pure protein from one litre of *E. coli* culture.

The secondary structure composition of HCoV-229E Nsp9 wt determined by CD spectroscopy well reflected that of the high-resolution structure, characterized by a high random coil and β-sheet content (data not shown). Investigating Nsp9 C69S by SAXS (fig. 74B) demonstrated that this mutant formed a dimer in solution in a buffer containing 250 mM NaCl and 20 mM NaH₂PO₄ pH 7.9, as discussed before, while higher oligomers were absent. Since the only native cysteine of wild type Nsp9 was substituted by serine in this mutant, the dimer was formed by non-covalent interactions. In agreement with SAXS data and with the previous investigation of polyprotein cleavage by native mass spectrometry (fig. 31, 32), native ESI-Q-TOF spectra of isolated Nsp9 wt detected dimeric Nsp9 (expected mass: 24,092 Da, observed mass: 24,113 Da/24,164 Da). In addition, mass spectrometry at a protein concentration of 5 μM revealed a similar amount of monomeric Nsp9 (expected mass: 12,046 Da, observed mass: 12,069 ± 17 Da) (fig. 75). Two slightly different masses of the dimer (ΔMW ≈ 50 Da) were suggested to result from binding of additional ions to Nsp9, e.g. one Ca²⁺ or two Mg²⁺ ions per Nsp9 molecule.

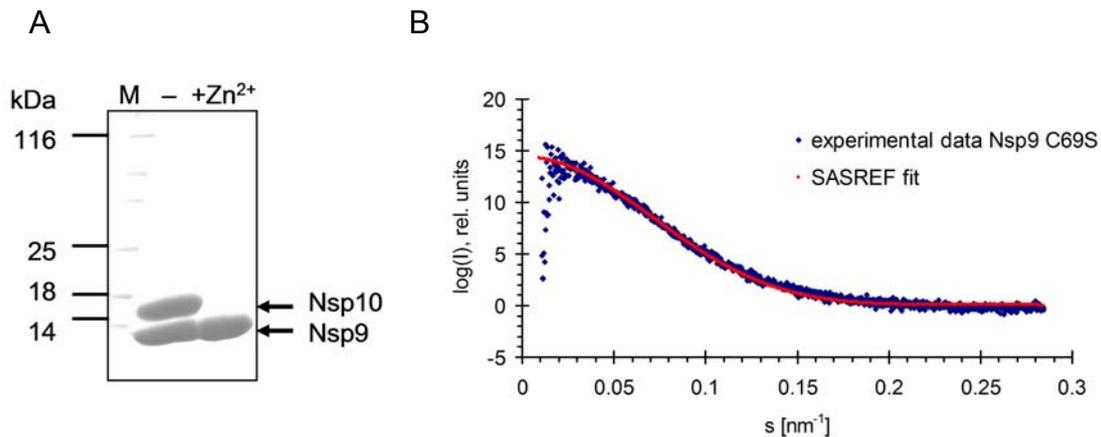


Figure 74. (A) Reducing SDS-PAGE of Nsp10 and Nsp9 co-localizing in an elution fraction after SEC of the processing mixture of pp710. After adding ZnCl_2 , Nsp10 precipitated, resulting in pure Nsp9 in solution. The Nsp9 C69S mutant was isolated following the same protocol. (B) SAXS data of Nsp9 C69S was collected at a concentration of $120 \mu\text{M}$ protein in 250 mM NaCl , $20 \text{ mM NaH}_2\text{PO}_4$ pH 7.9. The scattering plot was interpreted by rigid-body modelling using P2 symmetry (dimer) with SASREF for a momentum transfer of up to 0.2 nm^{-1} . The rigid high-resolution structure input was HCoV-229E Nsp9 wt (pdb code 2J97). The approximated molecular weight of 20 kDa well fitted the expected dimer as analyzed by SASREF (100 cycles; χ -value: 1.12678; penalty: 0.2315).

In the following section, a putative interaction of purified HCoV-229E Nsp9 C69S, lacking the only cysteine residue, with the complex of Nsp7 and Nsp8 is investigated by DLS.

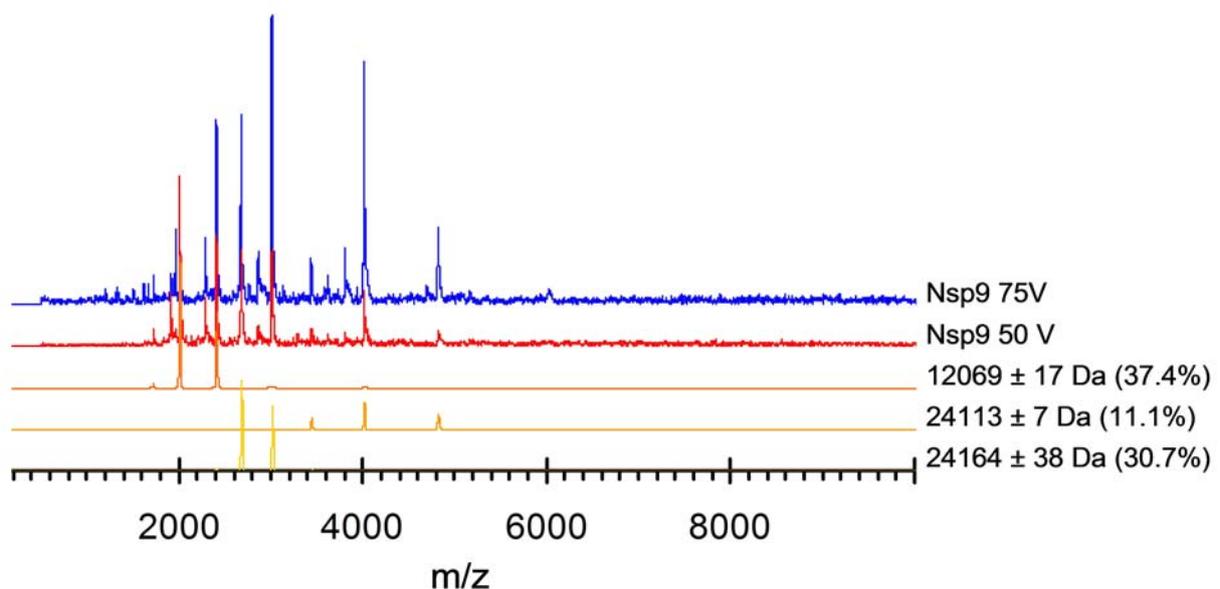


Figure 75. Native ESI-Q-TOF spectra of HCoV-229E Nsp9 wt at two different collision conditions (blue/red). Three components of the *red* Nsp9 spectrum, representing a monomeric and a dimeric state as well as a dimer that binds additional ions, have been simulated. $5 \mu\text{M}$ protein dissolved in $350 \text{ mM ammonium acetate}$, pH 7.9 was injected.

3.12.1. DLS and SEC do not suggest a tight interaction of Nsp9 with Nsp(7+8)

In vivo Nsp9 is most abundant in the perinuclear region of MCoV-A59-infected DBT cells and co-localizing with other components of the RTC including Nsp8, referred to as p1a-22 by the authors [196]. Aiming to investigate whether HCoV-229E Nsp9 influences the oligomeric state of the complex of HCoV-229E Nsp7 and Nsp8, DLS was applied. Nsp9 C69S and Nsp(7+8) were applied to DLS separately as well as an equimolar mixture of all three Nsps assuming that Nsp7 and Nsp8 interact with an equimolar complex stoichiometry as previously suggested by SEC (fig. 29B). However SEC indicated only a partly co-localization of Nsp9 with Nsp(7+8), indicating a rather weak interaction in the frame of experimental conditions. For DLS measurements a 2 μ l drop in a Terazaki 96-well plate covered with paraffin oil was prepared for every sample.

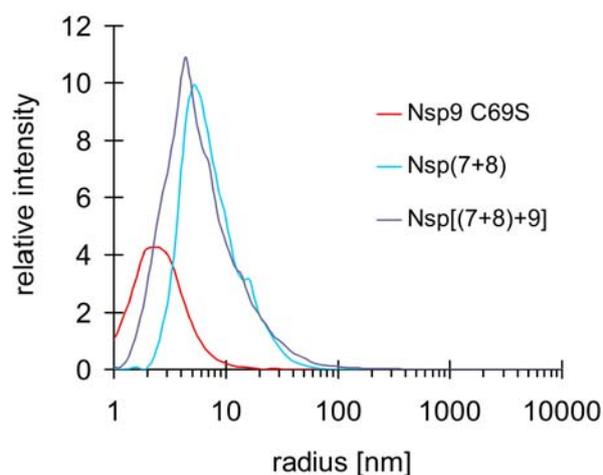


Table 26. Hydrodynamic radii determined by DLS

Nsp9 C69S	Nsp7+8	Nsp[(7+8)+9]
1.9 ± 0.2 nm	5.9 ± 0.9 nm	4.5 ± 0.3 nm

Figure 76. Half-logarithmic size-distribution plot obtained by DLS experiments to investigate the effect of HCoV-229E Nsp9 on the complex formed by mature Nsp7 and Nsp8 (Nsp(7+8)). For comparison, a hydrodynamic radius of 2.3 ± 0.2 nm was determined for isolated HCoV-229E Nsp9 wt (3 g l^{-1}) in a regular quartz cuvette with 15 μ l sample volume under identical conditions. Due to the substitution of Nsp9 C69 by serine a disulfide bond with either Nsp7 or Nsp8 was excluded. The protein buffer consisted of 250 mM NaCl, 30 mM NaH_2PO_4 , pH 7.6. Samples were pre-incubated for 1 h at 20 $^\circ\text{C}$.

The determined hydrodynamic radii (table 26) were not indicative for an interaction of HCoV-229E Nsp9 with the Nsp(7+8) complex. There was a minor decrease in hydrodynamic radius upon addition of Nsp9, likely representing an average signal of the components (fig. 76). Therefore, it was not possible to detect larger hetero-oligomers consisting of all three Nsps. This widely agrees with different SEC experiments in this study (fig. 29, 30, 38) showing that not more than a marginal amount of Nsp9 co-localizes with Nsp8. Previously, a reduction of the hydrodynamic radius of HCoV-229E Nsp(7+8) in the presence of HCoV-229E Nsp9 was observed by Dr. R. Ponnusamy using DLS

(University of Lübeck, unpublished). Moreover, the interaction of Nsp8 and Nsp9 of SARS-CoV was qualitatively previously indicated by sedimentation equilibrium experiments [102], promoting the DLS experiment in this study. However, a complex of all three Nsp8, Nsp9 and Nsp7 might require additional proteins or ligands *in vivo*. High-resolution structures of such a large complex would be highly beneficial, even though it might not be easy to obtain a crystal to collect diffraction data, if dissociation constants are rather low. An indirect large interaction state of Nsp7, Nsp8 and Nsp9 might functionally require Nsp9 to coordinate the binding of (long) RNA fragments to allow most efficient RNA polymerase activity of Nsp(7+8). A model deduced from variable oligomeric states of Nsp9 is a complex of an elongated Nsp9 oligomer wrapping around ssRNA, most probably mediated by the previously described OB fold [101].

3.13. HCoV-229E Nsp10

3.13.1. Purification, folding and oligomeric state

Nsp10 is a regulatory protein for which several interaction partners of the host and virus proteome were identified, some of them have been well characterized [112; 114; 115; 120]. The crystal structure of Nsp10 of the *Betacoronavirus* SARS-CoV was already reported [111], but there is still no atomic scale structural information of Nsp10 of an *Alphacoronavirus*, e.g. HCoV-229E, available for comparison.

For isolation of Nsp10 with authentic termini, HCoV-229E pp710 Q278A (chapter 3.5.1) was processed using M^{Pro}. After separation of Nsp10 by SEC the purity was verified by SDS-PAGE (fig. 77A) and the identity of the target protein was additionally verified by ESI-MS/MS after digestion by trypsin. This strategy is particularly valuable since an approach to purify Nsp10 after cloning the HCoV-229E Nsp10 gene into the expression vector pRSETA resulted in a relatively poor protein yield due to significantly reduced expression levels compared to the entire pp710.

Isoelectric focussing of pure Nsp10 revealed two major bands around pI = 7 (fig. 77B) under non-reducing conditions, which could be explained by the presence of different oligomeric states due to unspecific intermolecular disulfide bonds or two alternative folding states with a different surface charge. The homogeneity under reducing conditions (2 mM DTT) was confirmed by dynamic light scattering (DLS) ($R_H = 2.1 \text{ nm} \pm 0.2 \text{ nm}$) and analytical SEC (fig. 77C), indicating a monomeric state in solution, according to a calculated molecular weight of 13.9 kDa in SEC. However, under non-reducing conditions additional oligomerization was observed, likely correlating with unspecific oxidation of cysteine thiol groups, since HCoV-229E Nsp10 contains 12 cysteine residues. The hydrodynamic radius observed by DLS after storing the protein for 1 week under aerobic conditions significantly increased to $3.13 \pm 0.41 \text{ nm}$, corresponding to an approximately tetrameric state assuming a roughly globular shape of the protein. To verify the fold of monomeric Nsp10 CD

spectroscopy was applied, revealing a mixture of α -helix and β -sheet secondary structures, together with a random coil fraction. At least a part of the Nsp10 fold showed a remarkable stability up to a temperature of 70 °C (fig. 77D). Moreover, a tolerance to EDTA treatment demonstrated that the secondary structure stability does not significantly depend on the occupancy of two conserved Zn^{2+} -binding sites that have been identified by sequence comparison.

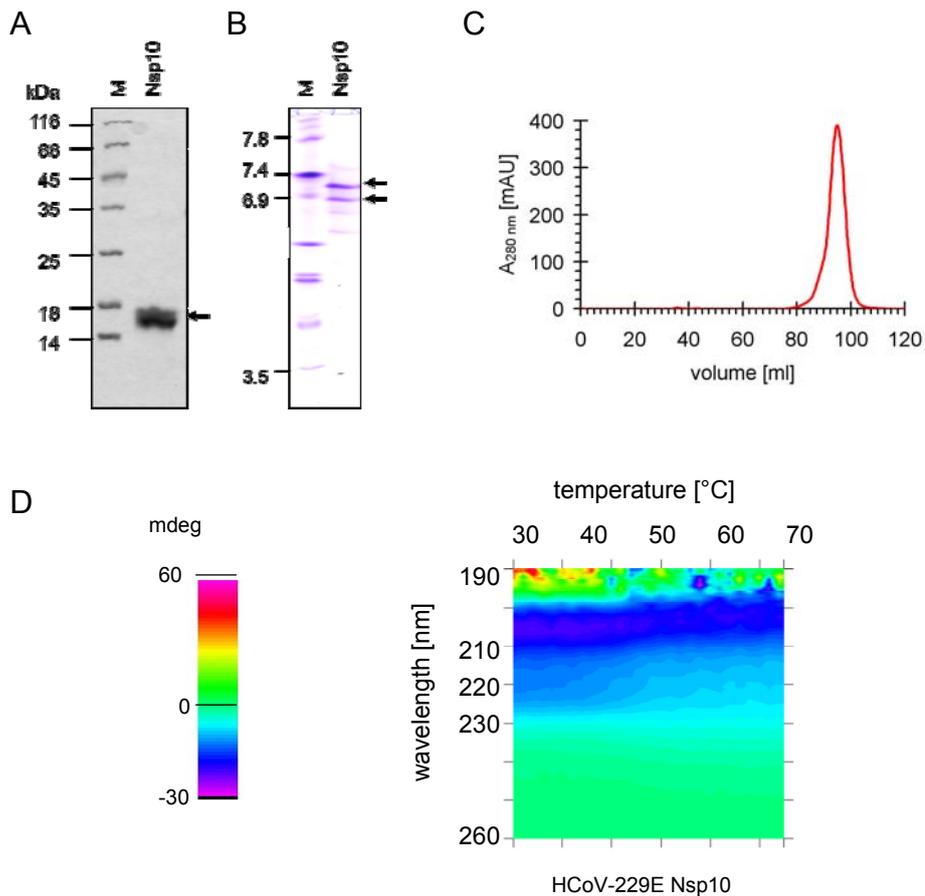


Figure 77. (A) Purification and characterization of HCoV-229E Nsp10. (A) Reducing SDS-PAGE analysis revealed pure Nsp10 after pp710 cleavage and isolation by SEC (theor. $MW_{\text{monomer}} = 14.4$ kDa). (B) Isoelectric focussing of pure Nsp10 under non-reducing conditions revealed two major charge states in a pH-range between 6.7 and 7.2. (C) SEC (HiLoad 16/60 Superdex 200) applying Nsp10: The protein eluted in one single peak characterized by $V_E = 94.1$ ml ($k_{av} = 0.66$) corresponding to a calculated molecular weight of 13.9 kDa. (D) Thermal stability of Nsp10 analyzed by CD spectroscopy. Nsp10 is characterized by a mixture of α -helix and β -sheet structure. A significant β -sheet percentage of Nsp10 is stable at 70 °C, while most α -helix structure is lost at temperatures above 50 °C.

It is well established that SARS-CoV Nsp10 unspecifically binds to ss/dsDNA and ss/dsRNA [111]. Thus, SEC (Superose 12 10/300 GL column) was applied to probe an interaction of HCoV-229E with nucleic acids. As depicted in fig. 78, the retention volume of Nsp10 shifted upon the

3. Results and discussion

addition of ssDNA to lower values, indicating an increase in particle size and therefore an interaction. As expected based on the sequence identity of 52.6% to SARS-CoV Nsp10 (fig. 7/79), there was no indication for altered functional properties of Nsp10 of an *Alphacoronavirus*, which is further supported by the structural data presented in the subsequent paragraphs.

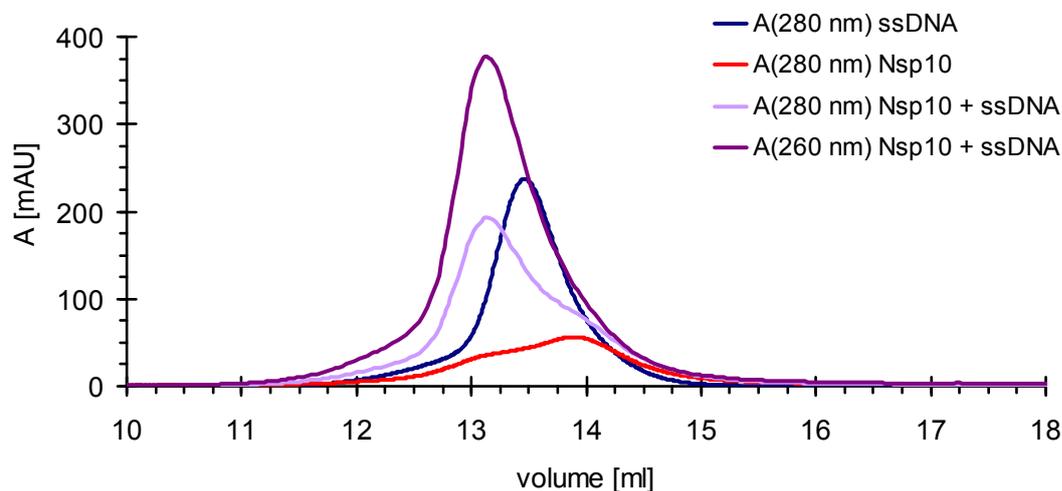


Figure 78. SEC analysis of nucleic acid interaction with HCoV-229E Nsp10 under non-reducing conditions. Supplementing Nsp10 with a 10-fold molar amount of a random 15-mer of ssDNA (5'-GATTAGTTACCTGAC-3') resulted in a significantly lower retention volume of both interaction partners monitored at a wavelength of 280 nm and 260 nm (*purple*), compared to the individual components (*red/blue*). This clearly indicates an interaction between Nsp10 and nucleic acids.

Aligning the sequences of Nsp10 from different coronavirus taxa (fig. 79) revealed a high degree of conservation, for instance highlighting the importance of the two Zn^{2+} -binding sites revealed for SARS-CoV Nsp10 [111]. However, the conservation of the C-terminal part was detected to be particularly low (fig. 79). Among the amino acids that were identified in a mutagenesis approach to influence the interaction of SARS-CoV Nsp10 with Nsp14 [117], amino acid position 82 displays the highest variability among the *Coronaviridae* with a preference for an acidic amino acid. Whereas residues Cys42, Gly70 and His80 that were crucial for the Nsp14 (ExoN) interaction of SARS-CoV Nsp10 are conserved, HCoV-229E Nsp10 harbours an alanine residue at position 82 instead of an aspartate. Nevertheless, the lack of structural data of Nsp14 and of the Nsp10/Nsp14 complex prevented a prediction whether an altered Nsp14 structure of HCoV-229E might compensate the amino acid exchange in terms of binding energy.

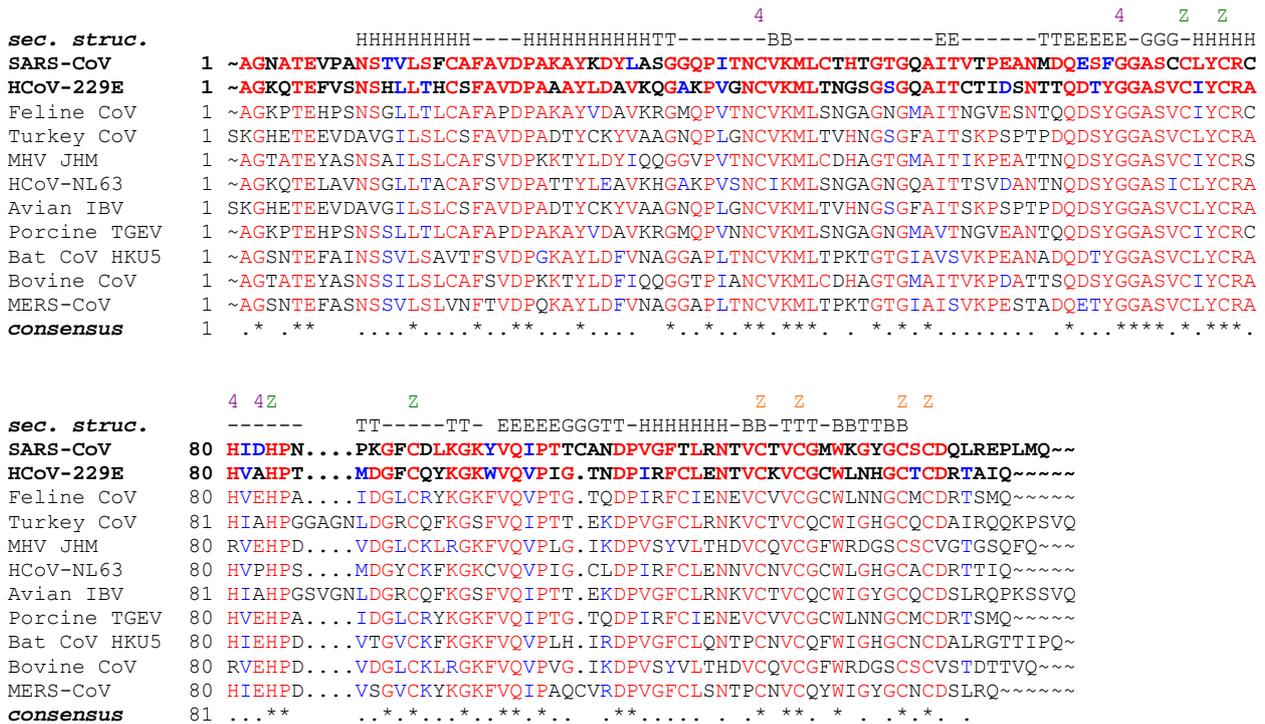


Figure 79. Amino acid sequence alignment of *Alpha*-, *Beta*- and *Gammacoronavirus* Nsp10 created by Clustal Omega (EMBL-EBI) and visualized by Boxshade V 3.2 (SIB). SARS-CoV and HCoV-229E Nsp10 share a sequence identity of 52.6%. An insertion of four predominantly small amino acids is found in turkey CoV and avian IBV. 4: residue involved in the interaction with Nsp14 [117]; Z: residue involved in zinc binding; H: α -helix; T: turn; B: β -bridge; E: β -sheet; G: 3_{10} helix; secondary structure according to SARS-CoV Nsp10 (pdb code 2FYG).

3.13.2. Crystallization of HCoV-229E Nsp10

Aiming at obtaining high-resolution structural data of HCoV-229E Nsp10, the concentration of pure Nsp10 was increased to 520 μM using an AMICON ultra filtration unit. Initial robot-assisted crystallization trials including approximately 500 different crystallization conditions were performed using 96-well MRC2 crystallization plates (Molecular Dimensions) and the sitting drop vapour diffusion method. A promising crystallization condition obtained within the first screening consisted of a precipitant composition of 1.26 M $(\text{NH}_4)_2\text{SO}_4$, 0.2 M Li_4SO_4 , 0.1 M Tris pH 8.5. After further optimization of this initial crystallization condition, final crystals were obtained by mixing 1.5 μl of protein solution with an equal volume of precipitant solution containing 0.9 M $(\text{NH}_4)_2\text{SO}_4$, 0.2 M Li_2SO_4 , 2 mM DTT, 0.1 M Tris pH 7.5. Crystals of polyhedral-bipyramidal shape grew using a hanging drop setup with 1000 μl of reservoir solution at 293 K to a final size of approximately $150 \times 150 \times 150 \mu\text{m}^3$ after three days. Using different ratios of $(\text{NH}_4)_2\text{SO}_4$ and Li_2SO_4 in the reservoir solution, additional crystals of similar quality and size were obtained. The successfully grown Nsp10 crystals are depicted in figure 80A-D. Without supplementing the SEC buffer (280 mM NaCl, 2 mM

3. Results and discussion

DTT, 25 mM HEPES pH 7.6) for the final purification of Nsp10 with 500 μ M ZnCl₂ it was not possible to obtain crystals of similar size and quality, although around 500 different crystallization conditions were initially screened again.

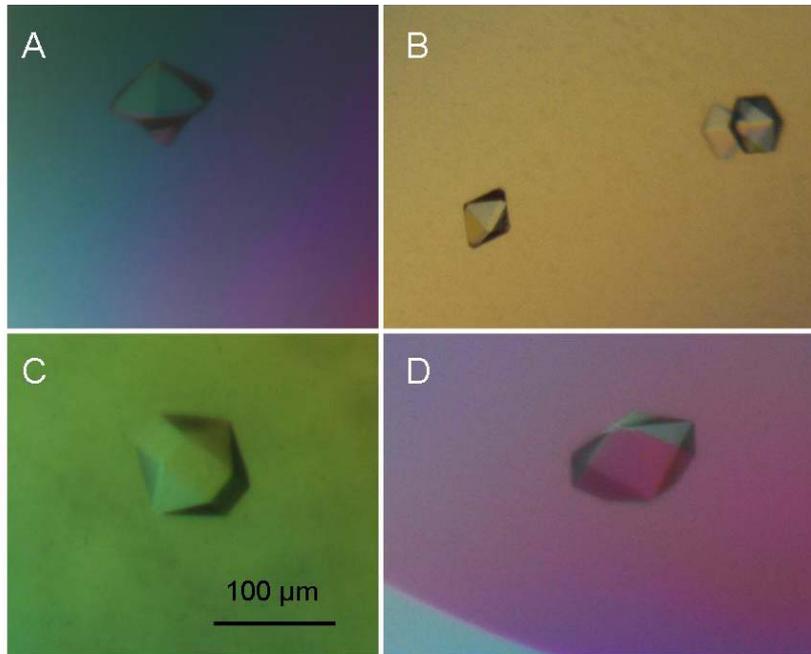


Figure 80. Polyhedral-bipyramidal shaped crystals of HCoV-229E Nsp10 obtained from hanging drop vapor diffusion experiments applying the following reservoir buffer conditions, diffracted to 1.85 Å at DORIS beamline X13: (A) 1.6 M (NH₄)₂SO₄, 0.1 M Tris pH 8.0; (B) 1.2 M (NH₄)₂SO₄, 0.1 M Tris pH 7.5, 2 mM DTT; (C) 1.25 M (NH₄)₂SO₄, 0.2 M LiSO₄ 0.1 M Tris pH 7.5; diffracted to 1.77 Å at PETRA III beamline P14; (D) 0.5 M (NH₄)₂SO₄, 1 M LiSO₄, 0.1 M sodium citrate.

Table 27. X-ray data collection and refinement statistics for HCoV-229E Nsp10. (Numbers in parentheses refer to the highest resolution shell.)

space group	P6 ₁
a (Å)	72.98
b (Å)	72.98
c (Å)	50.67
V _M (Å ³ /Da)	2.6
solvent content (%)	52.66
completeness of data (%)	100 (100)
no. of total reflections	74,272 (10796)
no. of unique reflections	9,778 (1421)
average I/σ intensity	16.7 (5.8)
resolution (Å)	26.83-2.05 (2.15-2.05)
redundancy	7.6 (7.6)
R _{merge} (%)	10.6 (34.7)
no. of reflections used in refinement	9,310
no. of reflections used in R _{free}	455
R _{crystal} (%)	17.5
R _{free} (%)	22.1
average B-factor (Å ²)	
main-chain atoms	15.227
side chain atoms	19.188
root mean square deviation	
bonds (Å)	0.0185
bond angles (°)	1.6534
residues in regions of the Ramachandran plot (%)	
most favoured	95 (114 aa)
allowed	5 (6 aa)
disallowed	0
generally allowed	0

Optimized Nsp10 crystals displayed a polyhedral bi-pyramidal-like shape (fig. 80). Diffraction data was collected to up to 1.8 Å at DORIS storage ring (consortium beamline X33, DESY, Hamburg) Diffraction data was indexed considering space group P6₁ and successfully processed. The crystal contains one molecule per asymmetric unit and a symmetry-related hexamer. Phasing was performed by molecular replacement using the structure of SARS-CoV Nsp10 (pdb code 2FYG) as a search model. Diffraction data up to 2.05 Å resolution was used for modelling and refinement.

The refined HCoV-229E Nsp10 structure shares the overall fold of SARS-CoV Nsp10 (pdb code 2FYG), as revealed by an RMSD of 0.97 Å for the backbone atoms (fig. 81A/B). Two predominant N-terminal α-helices up to residue 33 are connected by a short loop structure. Another α-helix of similar size as in SARS-CoV Nsp10 is located at the C-terminus (residues 106-114), flanked by the two highly conserved Zn-binding sites. A small “core” of three anti-parallel β-sheets is conserved as well. The third strand is connected by a loop structure comprising 20 residues. The Zn²⁺-binding sites involve identical residues in the high-resolution structures of HCoV-229E (fig. 81) and SARS-CoV

3. Results and discussion

Nsp10 (pdb code 2FYG) [111]. Both binding sites are occupied by a zinc ion in the HCoV-229E Nsp10 structure and well-defined in the electron density map. The Cys₄ binding site is reported to be grouped as a distant *gap-knuckle* type binding site, while the motif including C74, C77, C90, and H83 was previously identified as unique novel type [111]. The latter type of binding site is still unique to coronaviruses, since the DALI online server (Holm and Sander, 1998) is unable to identify a similar structure (Z -score > 2.0). It is not clear yet, whether or why particularly coronaviruses contain a large number of unique folds, as also revealed for Nsp7 [98], for the SARS-CoV nucleocapsid protein [216] and others.

Binding of a glycerol molecule as observed in the high resolution structure of SARS-CoV Nsp10 [111] was not observed. However, electron density was identified that was well interpreted by a calcium ion coordinated by water molecules in close proximity to residue A71 and to a substantial surface area of predominantly positive charge. There is no indication that this ion is involved in a regulatory mechanism, even though bivalent cations potentially alter the fold of RNAs and/or improve the affinity of RNA towards proteins [94; 217].

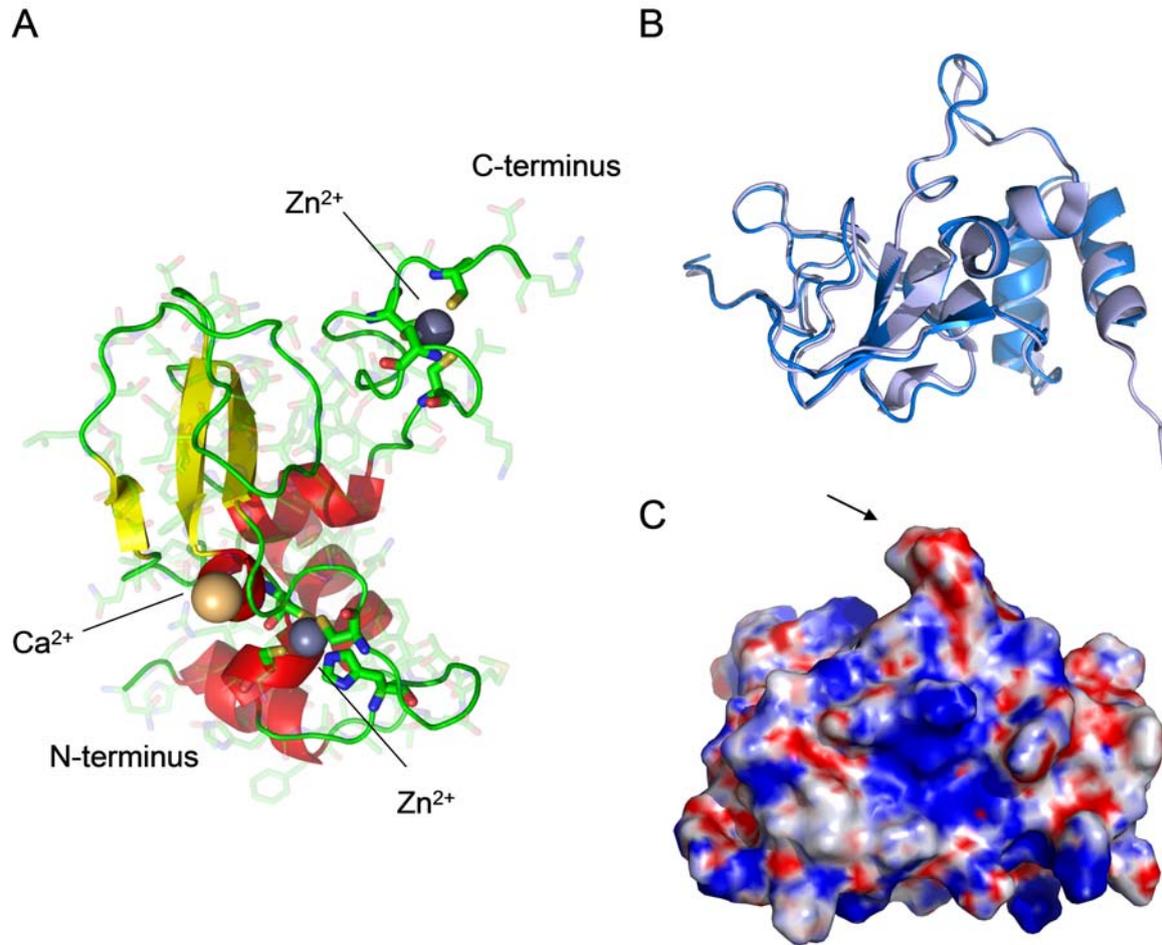


Figure 81. (A) Cartoon and stick visualization of the high-resolution structure of HCoV-229E Nsp10. Two conserved Zn²⁺-binding sites were occupied, the involved amino acids are highlighted. An additional Ca²⁺-binding site was identified. (B) Superposed structures of SARS-CoV Nsp10 (pdb code 2FYG; *light grey blue*) and HCoV-229E Nsp10 (*marine blue*). An RMSD value for the backbone atoms of 0.97 Å was determined by NOC vers. 3; <http://noch.sourceforge.net/>. (C) Surface charge of HCoV-229E Nsp10 as analyzed by PBEQ-Solver [168] displays a rim of predominantly positive charge (*blue*), as similarly reported for SARS-CoV Nsp10 [111]. Negative charge is coloured in red. The loop containing D87 sticking out of the protein results in additional negative charge (*black arrow*) in contrast to positive charge provided by K87 in the homologue loop of SARS-CoV Nsp10.

Analyzing the electrostatic surface potential of HCoV-229E Nsp10, a belt of positive charge involving K93 and R78 reported for SARS-CoV Nsp10 [111] is similarly present (fig. 81C), supporting the hypothesis that the homologue proteins share the same functional characteristics. For SARS-CoV Nsp10 residues forming this area remain conserved among most, but not all coronaviruses. Avian IBV and turkey CoV, for example, share a hydrophobic valine in position 11 (fig. 79). The charged residues K93 and R78 are conserved, with the exception that a few coronaviruses have an arginine in position 93. The positively charged belt is supposed to be associated with the RNA-binding function of SARS-CoV Nsp10 [111]. It was reported that the mutation Q65E

3. Results and discussion

disrupts the patch of positive charge [111]. The Nsp10 Q65E mutant in MHV resulted in the observation of a temperature-sensitive phenotype and the conclusion that Nsp10 even seems to be a critical factor in the regulation of main protease activity, implicating the regulation of polyprotein maturation [118].

In contrast, a surface loop containing residue D87 that sticks out of the protein provides additional negative charge that is not conserved (fig. 81C). The homologue loop of SARS-CoV Nsp10 contains a lysine residue at this position. Different interaction partners of Nsp10 have been suggested, without structural knowledge of the putative interface. Charge differences at the protein surface could have an impact on the interaction partner, depending on the coronavirus species. For example, the interaction with the cellular oxido-reductase machine was investigated by transfecting human embryo lung fibroblasts by the Nsp10 gene, resulting in a depolarized inner mitochondrial membrane [120]. However, a preferred binding site for small organic molecules is not indicated in close proximity to the described loop, as investigated by small molecule docking using the FTmap server [167].

To map and analyze hypothetical small molecule binding sites on the surface HCoV-229E Nsp10 and to identify differences compared to SARS-CoV Nsp10, the FTmap and also the SiteHound online server were applied. These algorithms map cavities and surface substructures that could potentially be occupied by a ligand. Comparing the optimised shape and distribution of all ligands modelled by SiteHound, significantly different results have indeed been observed for the two Nsp10 homologues (fig. 82, left panel). However, having a closer look at the few specific binding sites with defined molecules identified by small-molecule-docking using FTmap indicates strong similarities. One “cross cluster” including small molecules (acetone, benzene, isobutanol, benzaldehyde and others) that are predicted by FTmap to bind at a specific HCoV-229E Nsp10 site (fig. 82, *orange* circle) involves two hydrogen bonds of G94 and K95 for binding, in close proximity to the Ca²⁺-binding site observed in the high-resolution structure (fig. 81). Remarkably, the glycerol binding site identified by Joseph *et al.* [111] co-localizes with one docking cluster of small molecules on the surface of SARS-CoV Nsp10 (*pink* carbon atoms), highlighting the plausibility of this binding site. At the corresponding position on the surface of HCoV-229E Nsp10 another comparable docking cluster was identified (*yellow* carbon atoms), pointing to a similar binding cavity. The enframed “backside” of the protein harbours a similarly large binding cavity in both proteins, containing more than one docking cluster, although the geometry (a more elongated deep cleft is observed in HCoV-229E, *green circle*) and composition of the clusters is slightly different. In conclusion, several candidate binding sites of small molecules have been predicted on the surface of HCoV-229E Nsp10, most of them remain conserved for SARS-CoV Nsp10, promoting further experiments to develop larger high affinity ligands. Such compounds could, for instance, interfere with the reported interaction of Nsp10 with Nsp14 or Nsp16. This is of specific interest in terms of compound development to treat coronaviral infections.

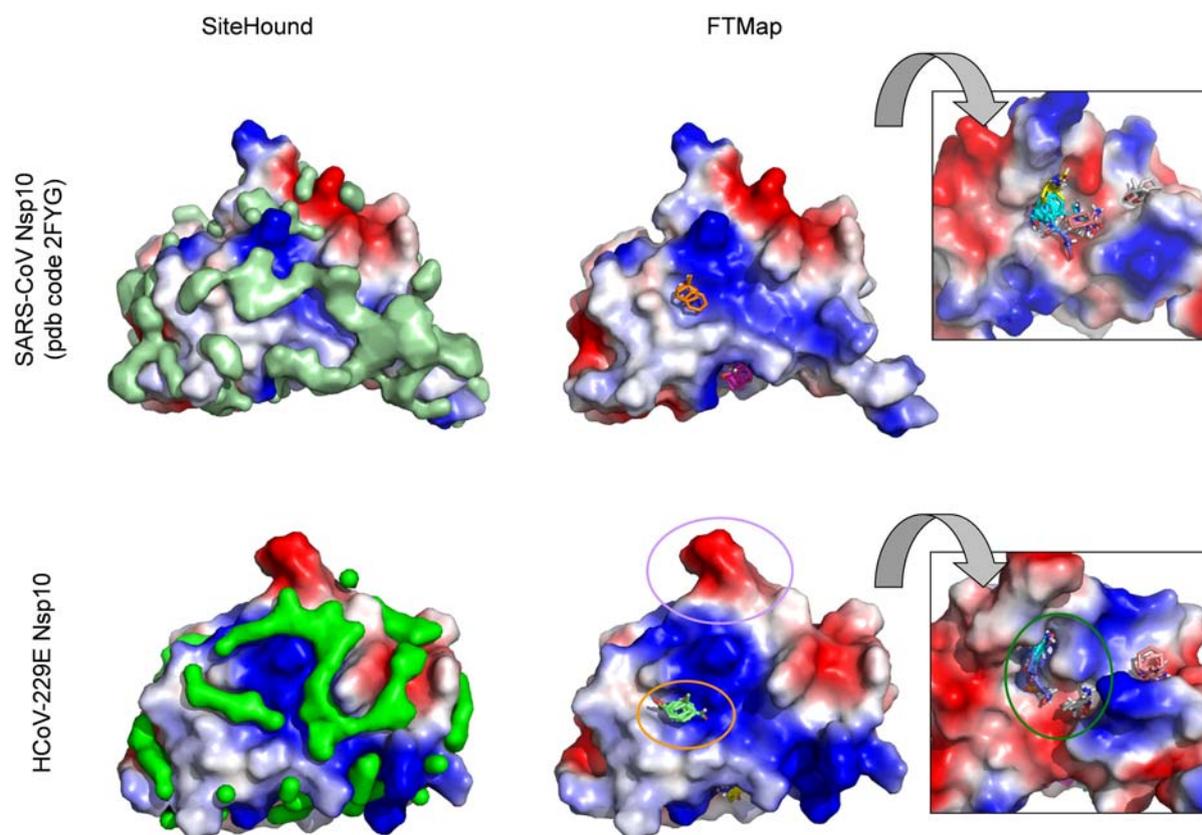


Figure 82. Ligand binding site prediction based on the surface topology of Nsp10 using *SiteHound* (left panel) and grid-based docking of small molecules performed by FTmap (right panel). The *SiteHound* server was used with the implemented carbon probe to identify potential binding sites of drug-like molecules. The predicted geometry of ligands resulting from all “docking clusters” is depicted with solid surface in *green*. FTmap was applied using the entire default library of ligands. An area of altered surface charge is *encircled purple*. The electrostatic potential was calculated by PyMOL Molecular Graphics System. The insets on the right show the surface on the “backside” of the protein (180° rotation).

3.13.3. Structural characterization in solution

To verify the tertiary structure of HCoV-229E Nsp10 and its monomeric state in solution independent of the crystal environment and to investigate the impact of Zn^{2+} on the Nsp10 structure, SAXS experiments were performed (fig. 83).

3. Results and discussion

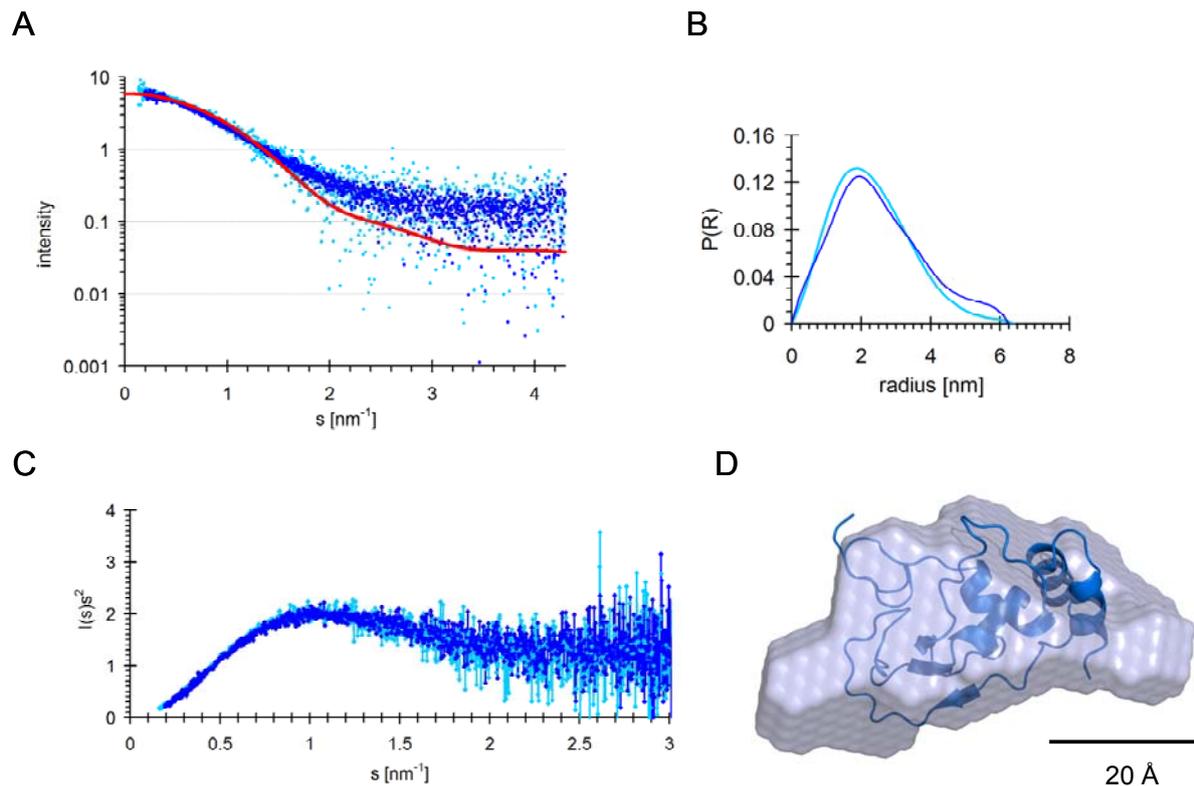


Figure 83. Structure of HCoV-229E Nsp10 in solution determined by small-angle X-ray scattering. **(A)** Scattering pattern $I(s)$ plotted against s [nm^{-1}] of HCoV-229E Nsp10 in 400 mM NaCl, 1 mM EDTA, 5% glycerol, 40 mM NaH_2PO_4 , pH 7.9 (*dark blue*), and 280 mM NaCl, 2 mM DTT, 25 mM HEPES pH 7.6 supplemented with 500 μM ZnCl_2 (*light blue*), respectively. The calculated CRY SOL graph (*red*) corresponds to the theoretical scattering of the high-resolution structure of SARS-CoV Nsp10 (pdb code 2FYG). **(B)** Pair-distribution function of Nsp10 resulting from a Fourier transform by GNOM in the absence (*dark blue*) and presence (*light blue*) of Zn^{2+} . **(C)** Kratky plot ($I(s)s^2$ plotted against s) of HCoV-229E Nsp10 with the same colour code. **(D)** Superposition of monomeric HCoV-229E Nsp10 crystal structure and a single *ab initio* model of the zinc-containing samples (DAMMIF). Superposition and comparison of ten individual DAMMIF models was done using DAMMAVER resulting in a mean NSD of 0.810 (NSD variation: 0.039).

The scattering pattern at small angles recorded for Nsp10 in the presence and absence of zinc ions appeared to be highly similar. A slightly elongated shape characterized by an R_g of 1.81 ± 0.02 nm and a maximum diameter of $D_{\text{max}} = 6.4$ nm is indicated in the presence of Zn^{2+} , while an R_g of 1.95 ± 0.08 nm and a $D_{\text{max}} = 6.3$ nm were obtained in the absence of zinc ions. Therefore, although a slight difference in the $P(R)$ -function was detected, the occupied state of the Zn-binding sites did not quantitatively trigger oligomerization under reducing conditions. This represents another strong indication for a functional authentic monomeric state of the protein in the range of concentrations applied in this experiment. However, a high-resolution crystal structure of SARS-CoV Nsp10 (pdb code 2G9T) displayed a dodecameric state characterized by an envelope diameter and R_g that is approximately three times higher than obtained for the monomers (table 28). Even though a

dodecameric state was sterically excluded for the complex consisting of Nsp10 and Nsp16, a physiological relevance of a “free” Nsp10 dodecamer was not excluded [181]. Nonetheless, considering that pp910 cleavage is dispensable for the viability of murine hepatitis virus [107], dodecamer-formation would consequently need to allow the formation of a pp910 oligomer of similar size at least for MHV.

To compare the shape of HCoV-229E Nsp10 in solution with the high-resolution structures of SARS-CoV and HCoV-229E Nsp10, respectively, Crysol [152] was used to determine the discrepancy. A chi value of ≈ 1 for the comparison with both monomeric structures (SARS-CoV, pdb code 2FYG; HCoV-229E, this study) consistently indicates high structural similarity with the solution state (table 28). A slightly increased experimentally determined R_g value might result from additional flexibility in solution. For example, the exposed loop region containing the residues 85 to 90 ($^{85}\text{TMDGFC}^{90}$) of HCoV-229E Nsp10 displays significantly increased B-values and high local RMSD values ($> 3\text{\AA}$) compared to the SARS-CoV homologue and was therefore considered to be flexible. The solvent accessibility of this loop is significantly altered in the crystal environment, according to the online server CryCo 5.2 [186]. Moreover, particularly close to the N- and C-terminus of the Nsp10 structure, according to the superposition with a DAMMIF model, some additional volume was calculated by DAMMIF, additionally indicating increased flexibility in solution. This was also indicated for larger regions throughout the molecule by the characteristic uncommon Kratky plot of the scattering pattern [$I(s)s^2$] (fig. 83C) that is not clearly converging towards the radius axis at higher angles. According to CryCo 5.2 the N-terminal helix is largely kept in place by crystal contacts, which is similarly observed in the crystal structure of SARS-CoV Nsp10 [111]. This might, for example, explain a slight increase in flexibility in solution observed by SAXS. A certain degree of structural variability to adapt to different interaction partners is consequently not excluded.

Table 28. Comparison of SAXS data and different rigid Nsp10 high-resolution structures by an optimized CRY SOL fit

	Nsp10 (HCoV-229E; SAXS)	Nsp10 (HCoV-229E)	2FYG (SARS-CoV)	2G9T (SARS-CoV)*
R_g [nm]	1.81	1.67	1.71	5.22
solution scattering fit: χ -value	-	0.999	0.969	4.59
envelope diameter [nm]	6.4 (= D_{\max})	5.1	5.8	1.7
R_h	2.1 nm \pm 0.2 nm (+ 2 mM DTT); calc. estimation ¹ : 1.8 nm	-	-	-
oligomeric state	monomer	monomer	monomer	dodecamer

* the protein contains the Nsp11 peptide at the C-terminus

¹ according to [218]

3.13.4. Impact of Nsp10 on the interaction with HCoV-229E Nsp16

For activation of the 2'-O-methyltransferase Nsp16 in SARS-CoV the interaction with Nsp10 was shown to be required. The respective complex of the SARS-CoV Nsps was solved at atomic resolution [112]. In order to investigate a homologue interaction between both Nsps for HCoV-229E, the corresponding interface was examined *in silico*.

The structure of free SARS-CoV Nsp10 (pdb entry 2FYG) with a solvent-exposed Nsp16 interface does not significantly differ from that within the Nsp10/Nsp16 complex of SARS-CoV (pdb code 2XYQ). Since Nsp16 binding does not induce significant structural changes in Nsp10, an *in silico* investigation of the same interaction in HCoV-229E using the solved crystal structure of free HCoV-229E Nsp10 is validated. The high-resolution-structure of SARS-CoV Nsp10 interacting with SARS-CoV Nsp16 (pdb code 2XYQ) was used as a template to model the equivalent interaction for the HCoV-229E Nsps. Using Sybyl-X 1.2, SARS-CoV Nsp10 was replaced by HCoV-229E Nsp10 and SARS-CoV Nsp16 by a homology model built by I-TASSER. The predicted complexes were characterized by the PISA server (EMBL-EBI; table 29). A quantification of the surface charge of one Nsp10 homologue interacting with either SARS-CoV Nsp16 or the HCoV-229E Nsp16 homology model is displayed in the appendix (fig. 89), revealing a high conservation of large charge patches that have been described for SARS-CoV [111; 112].

Table 29. Comparative summary of the Nsp10/Nsp16 interfaces calculated by the PISA server (EMBL-EBI)

complex		solvent accessible area [Å ²]	interface area [Å ²]	solvation energy [kcal mol ⁻¹]		
				isolated structure	gain on complex formation	p-value
I	HCoV-229E Nsp10	7029.1	892 (12.7%)	-94.8	-8.0 (8.4%)	0.174
	HCoV-229E Nsp16	14738.0	910 (6.2%)	-278.8	-5.8 (2.1%)	0.315
II	HCoV-229E Nsp10	7020.8	896.3 (12.8%)	-95.0	-7.7 (7.8%)	0.232
	SARS-CoV Nsp16	13915.4	840.9 (6.0%)	-288.4	-6.2 (2.1%)	0.142
III	SARS-CoV Nsp10	7908.0	895.7 (11.3%)	-99.7	-7.5 (7.5%)	0.219
	HCoV-229E Nsp16	14752.9	904.5 (6.1%)	-278.5	-5.7 (2.1%)	0.322
IV	SARS-CoV Nsp10	7215.1	944 (13.1%)	-101.7	-8.0 (7.9%)	0.164
	SARS-CoV Nsp16	13739.2	891 (6.5%)	-290.5	-6.2 (2.1%)	0.173

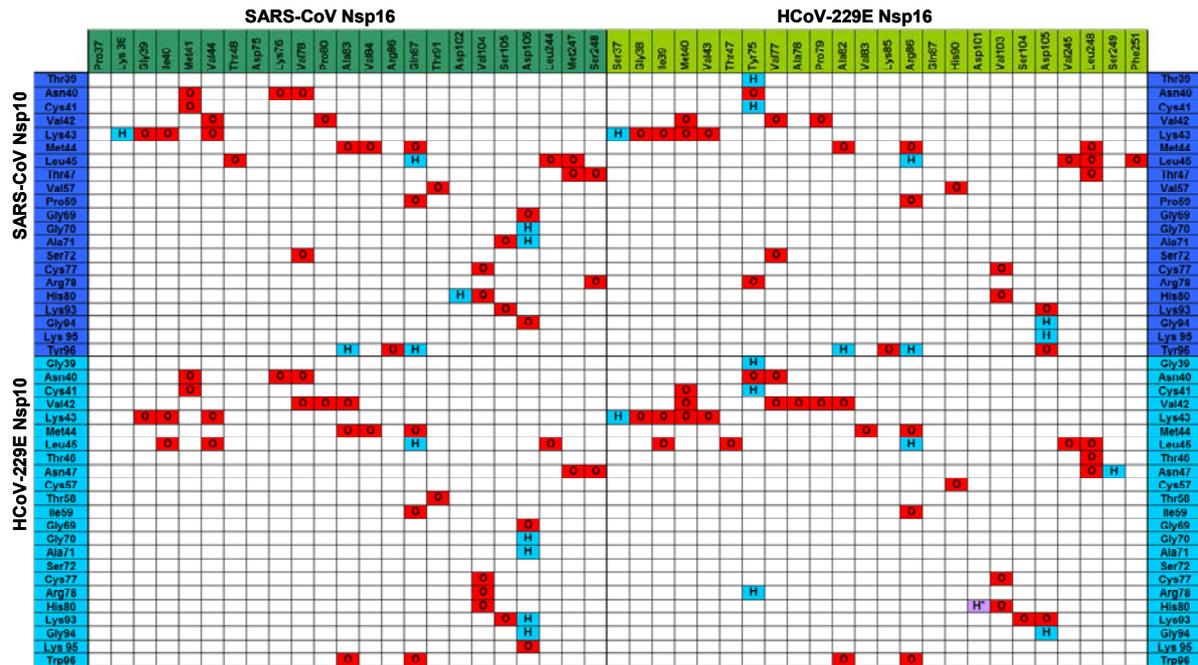


Figure 84. Schematic overview of non-covalent interactions of SARS-CoV/HCov-229E Nsp10/Nsp16 including the two chimeric complexes. A hydrogen bond between two residues is labelled with “H”; “O” indicates hydrophobic interaction. Upon the exchange of either Nsp10 or Nsp16 of SARS-CoV by the homologue of HCov-229E, the pattern of hydrogen bonds and hydrophobic interactions significantly changed. Particularly S37, Y75 and S249 of HCov-229E Nsp16 (homology model based on SARS-CoV Nsp16) provide additional hydrogen bonds, interacting with HCov-229E Nsp10, compared to the complex of SARS-CoV. On the other hand, the exchange of SARS-CoV Nsp10 Y96 by W96 in HCov-229E reduces the number of hydrogen bonds at the interface by two.

* by rotation of Asp101 C_α-C_β (hydrogen bond labelled *purple*)

A schematic summary of hydrogen bonds and hydrophobic interactions of the Nsp(10+16) complexes of SARS-CoV and HCov-229E, including the chimeric complexes consisting of proteins of two different coronaviruses used for validation, is displayed in figure 84. The predicted interface of Nsp10 and Nsp16 from HCov-229E displays additional hydrogen bonds. Nsp16 Y75 that is not conserved in SARS-CoV or FCoV (S75 in FCoV; fig. 84/85/88, appendix) establishes additional hydrogen bonds, independent of which Nsp10 homologue forms the complex. One interaction partner is Nsp10 C41, which is conserved in the SARS-CoV homologue, but not involved in hydrogen bonding with SARS-CoV Nsp16. A major difference of the interface is related to residue 96 of Nsp10, which is harboured by a tyrosine in SARS-CoV and a tryptophan in HCov-229E, respectively. It was already reported that Nsp10 Y96 represents a critical “hot spot” for the interaction with Nsp16, as identified by alanine-scanning mutagenesis combined with the determination of residual Nsp16 activity in SARS-CoV; V42, M44, A71, K93 and G94 were additionally identified in this approach [219]. Nevertheless, the significant impact of Y96A is suggested to depend on the individual amino acid environment of SARS-CoV Nsp10. In HCov-229E (Nsp10+16), W96 of Nsp10 contributes to the

3. Results and discussion

binding energy by hydrophobic interactions to A82 and N86, instead of two hydrogen bonds of Y96 in the SARS-CoV complex with A83 and N87. It is predicted that W96 of HCoV-229E is unable to participate in a hydrogen bond, while in compensation HCoV-229E Nsp16 Tyr75 participates in additional hydrogen bonds.

Similar to SARS-CoV Nsp16 D102, the corresponding D101 of the HCoV-229E homologue is able to participate in a hydrogen bond with Nsp10 H80 assuming free rotation of the aspartate residue. Both residues are highly conserved, while Nsp10 H80 overlaps with the binding interface of Nsp14. According to this model of HCoV-229E Nsp(10+16) one further loop-stabilizing hydrogen bond of S249 with N47 is possible (fig. 84/85), even though the main chain orientation of the involved loop structures is similar. Remarkably, feline and turkey coronavirus contain a proline and an aliphatic alanine residue in position 249 in Nsp16. The flexible loop region 100-108 of the SARS-CoV Nsp16 which is essentially involved in binding the cofactor SAM, is significantly stabilized by Nsp10 [181]. This is well reflected in the model of the HCoV-229E complex as the conserved D105 of HCoV-229E Nsp16 (D105 in FCoV Nsp16, D106 in SARS-CoV) forms a hydrogen bond.

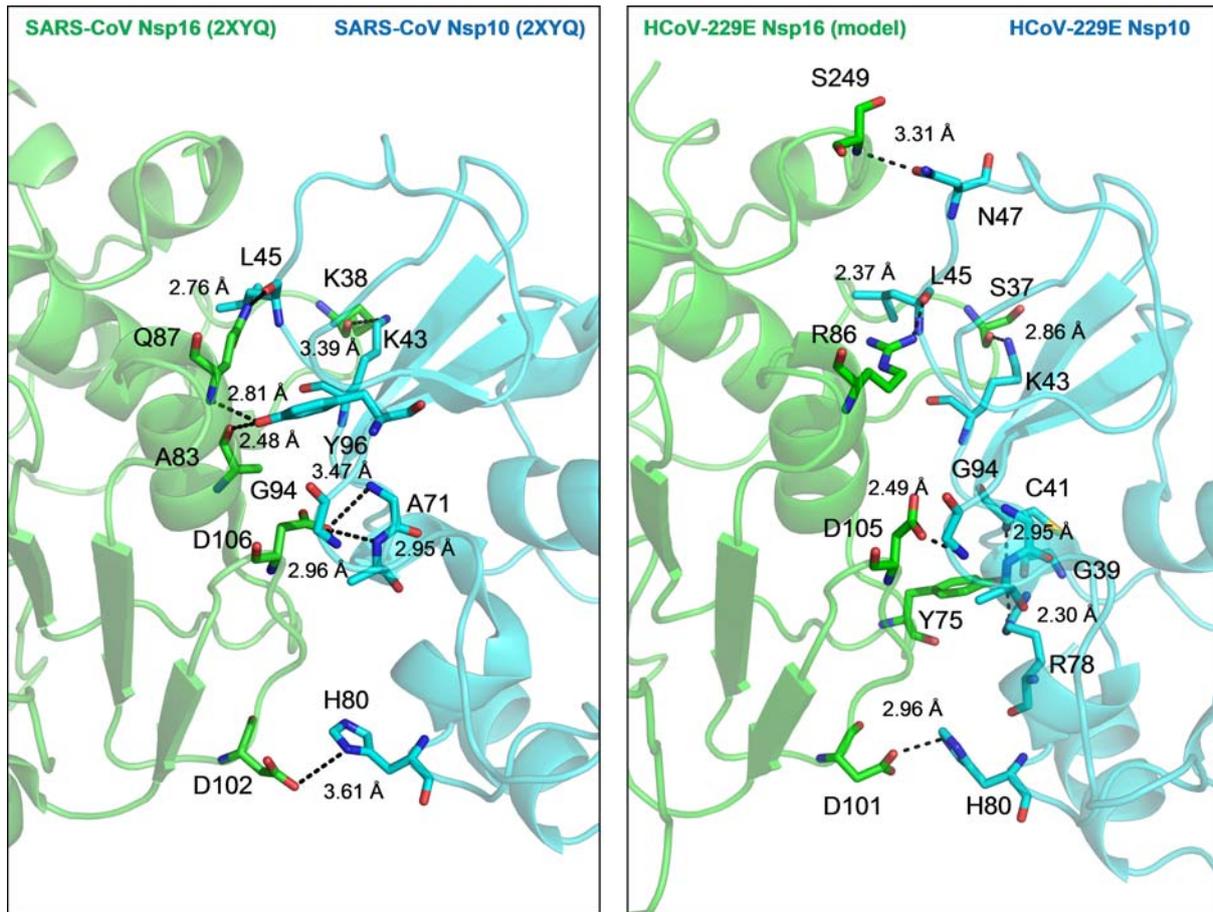


Figure 85. Close-up view of hydrogen bonds (dashed lines) that are part of the modelled interface of HCoV-229E Nsp10 and HCoV-229E Nsp16 (right) in comparison to the homologue complex of SARS-CoV (left). Nsp16 is coloured *green* and Nsp10 is coloured *blue* in both panels. Residues involved in H-bonds are displayed in stick representation.

In summary, the solvent accessible area and the buried interface area of all putative Nsp(10+16) complexes are similar within a fluctuation range of 10% (table 29). The interface area of SARS-CoV Nsp10 is remarkably decreased when Nsp16 of SARS-CoV is substituted by its homologue. Nevertheless, the majority of hydrogen bonds is rearranged, to an even higher extend while the Nsp16 homologue is “exchanged” in the model (fig. 84) which is also reflected in the final model of HCoV-229E Nsp(10+16) “exchanging” both interaction partners of the SARS-CoV complex. Today, SARS-CoV Nsp16 is still the only 2'-O-methyltransferase demonstrated to require the stimulation by another protein to induce its catalytic activity, in contrast to FCoV Nsp16 that is catalytically active in the absence of Nsp10 [220]. Even though it cannot be ruled out that this activation is not necessary in case of HCoV-229E, a valuable model is provided, indicating that a similarly tight interaction could occur between HCoV-229E Nsp10 and Nsp16, but involving remarkably different amino acids.

4. Conclusion

Addressing the question whether coronaviral pre-processed polyproteins, e.g. pp710, pp79 and pp78, already act as functional (performed) building blocks of the RTC, it was shown in this study that polymerase activity, secondary structure and oligomerization characteristics detected for the precursor states remain largely comparable to that of the mature Nsps. The observation that a mutation within the GXXXG-motif, important for dimerization of mature Nsp9 [102; 104], is dispensable for the oligomerization of pp79 and pp710 indicates that this dimerization site is uncovered not before Nsp9 is N-terminally processed. Considering that the dimerization of Nsp9 is essential for viral replication [104], this observation could explain why processing of pp89 is not dispensable *in vivo*, as observed by Deming *et al.* [107]. Nevertheless, cleavage of the respective site was determined to be comparably slow in agreement with the relatively high K_D -value of the inhibitory peptide derived from the Nsp8-Nsp9 cleavage site and with previous studies [65; 191]. In contrast, cleavage of pp910 is dispensable for viability of MHV *in vivo* [107], corresponding to the observation that Nsp9 and Nsp10 do not share a joint binding interface that could, for instance, impair dimerization of Nsp9. Dimerization of SARS-CoV pp910 was observed by Z. J. Miknis (dissertation, State University of New York, unpublished) in a SAXS experiment. However, in light of the rapid processing of pp910 by the main protease and the missing interaction of both mature Nsps, the physiological impact of the precursor pp910 is most likely rather low or absent.

As observed for other enzymatic activities of mature Nsps so far, the detected polymerase activity, i.e. at least the nucleotide ligand epitopes determined by STD-NMR spectroscopy, appears to be highly conserved for polyproteins of SARS-CoV and of the distantly related Alphacoronavirus HCoV-229E. A high level of conservation of the SAM and RNA binding site of HCoV-229E Nsp16 and Nsp10 and its complex is also indicated, compared to the reported SARS-CoV Nsp(10+16) complex [112; 114]. Despite the known degree of sequence variability at the interface, conserved interaction epitopes are expected for the building blocks of the RTC of different coronaviruses, as widely suggested in this study for Nsp10 and Nsp16. Since the overall fold of mammalian and viral methyltransferases, e.g. Nsp16 is highly conserved, it seems to be more promising in terms of developing antiviral compounds to focus on Nsp10 and its binding interfaces, instead to interfere with the activity of Nsp16 itself, as suggested by [181]. Along with other putative regulatory functions of Nsp10 [117; 118] this statement highlights an important role of Nsp10 in coronavirus research.

In terms of oligomerization it was not possible to verify the formation of a specific hexadecameric 8:8-Nsp(7+8) complex, which was reported by [98] in the context of a high-resolution crystal structure. However, smaller hetero-oligomers exist in solution, perhaps in addition to a hexadecamer. Therefore, the significant different oligomerization states identified for homologue

complexes strongly depend on the applied method for investigation and the associated (artificial) conditions. A dodecameric state of Nsp10 composed of trimers observed within a protein crystal structure [131], for example, is in contradiction to all results obtained by analysing isolated Nsp10 in solution. In a wide frame of experimental conditions Nsp10 was demonstrated to be monomeric, while a comparably large oligomer of the Nsp(10+16) complex is sterically improbable. Different high molecular weight oligomers observed for pp710 were highly heterogeneous and not yet correlated to a physiological function, beside the possibility to provide an increased stability towards stress in a host cell, e.g. induced by proteases or temperature. The detected equilibrium of the monomeric and dimeric state of pp710 appears to be physiologically more relevant. The presented low-resolution model of pp710 that is strongly supported by mass spectrometry data indicates that the dimerization via the C-terminal domain of Nsp8 appears to be conserved for pp710 and pp78 during maturation of Nsps as well as for *Alpha-* and *Betacoronaviruses*. In fact, a significant flexibility of the N-terminus of the Nsp8 domain within pp710 and pp78 is strongly suggested.

Finally, specific interactions of SARS-CoV main protease with pp710 cleavage sites lacking the crucial glutamine in P1 position were observed. However, even though both proteins can form a homodimer there is no significant formation of specific high-order oligomers of the protease-polyprotein complex, probably influenced by a dependence on the protein concentration, by an asymmetric fold of the protomers within a dimer [52] or sterical limitations. Based on reports concerning the monomer-dimer equilibrium of SARS-CoV M^{pro} [48; 132] it is conceivable that the interaction is at least partly a 1:1 interaction involving different native full-length substrates – like in a set of mass spectrometry experiments.

The major aim of future studies will still be the elucidation of high-resolution structures of precursor states of coronaviral Nsps. Probably involving novel crystallographic techniques, both pp710 and partly processed fragments as well as polyprotein-M^{pro}-complexes mentioned above represent important targets to confirm the low-resolution models and the structural and functional suggestions presented in this study. A clarification of further Nsp-Nsp-interactions and intra-polyprotein-domain-interactions will broaden the basis of structural knowledge to understand the architecture of the coronaviral RTC complex and to develop perspectives and strategies to specifically interfere with its formation. In addition, next to the well-established attempts to target the coronaviral main protease Nsp5 for drug development, also Nsp10 appears to be an interesting target for modulation of the methyltransferase activity of its interaction partner Nsp16. In this context, additional efforts might be promising to target oligomeric states of Nsps or precursor proteins by dimerization inhibitors and to investigate the effect on virus replication. A mechanism to inhibit the essential dimerization of the SARS-CoV main protease by a peptide was previously described [221].

5. Summary

Coronaviruses (CoVs) comprise a group of enveloped (-)ssRNA viruses. SARS-CoV and the currently emerging MERS-CoV characterized by a fatality rate of around 50% are considered to be a major world-wide health issue with high economic impact. Although most coronaviral infections are comparably harmless for humans (e.g. for HCoV-229E), SARS-CoV and MERS-CoV clearly demonstrated that severe respiratory diseases are likely to be spread by novel highly pathogenic coronavirus species in the future, infecting various animals and humans. The coronaviral main protease has been identified as a suitable target to treat infections, but such compounds did not reach the market yet. A large hypothetical transcriptase/replicase complex (RTC) of coronaviruses is assumed to be another promising drug target, even though its formation remains poorly understood so far. It is established that the RTC essentially involves several mature non-structural-proteins (Nsps) released from the primary polyprotein translation products of the viral RNA by viral proteases, while the impact of their sequentially cleaved polyprotein-precursors is unknown.

Addressing important questions of structure, oligomeric composition and function of polyprotein precursors that may act as putative RTC building blocks, the C-terminus of SARS-CoV polyprotein 1a comprising Nsp7-10 (pp710) was recombinantly produced in *E. coli* and purified. The homologue protein of the distantly related *Alphacoronavirus* HCoV-229E that causes common cold was purified for comparison as well. In addition, specific fragments of these large pp710 molecules that were observed along the pathway of processing by a viral main protease *in vitro* were purified to allow a structural comparison with incompletely cleaved polyprotein intermediates and mature non-structural proteins. For the first time, a complex cleavage reaction of a viral polyprotein was monitored at high mass accuracy by mass spectrometry, revealing for example a 2:2-interaction state of mature SARS-CoV Nsp7 and Nsp8. It was shown that RNA-dependent RNA polymerase activity of Nsp8 does not require a hexadecameric interaction with Nsp7, which was previously reported. Nsp8 is capable of *de novo* synthesis of RNA fragments (> 10 nt), even in the precursor state covalently linked to Nsp7 and Nsp9. The binding epitope of nucleotides resulting in catalytic activity was characterized for two Nsp8 precursors by saturation transfer difference NMR spectroscopy. The observation of catalytic activity and interaction with nucleic acids promotes the idea of functionally relevant polyproteins acting as RTC building blocks. Furthermore, size exclusion-chromatography and small-angle X-ray scattering (SAXS) demonstrate a concentration-dependant oligomerization of SARS-CoV and HCoV-229E pp710, which is hypothesized to represent a regulatory mechanism during coronavirus infection. A low-resolution structural model for the domain interactions within the monomer of pp710 as well as for the intermolecular dimer interactions is proposed, verified by SAXS data of truncated variants and mass spectrometry of cross-linking experiments. Based on SAXS data it is further suggested that the interaction of Nsp7 and Nsp8 is preformed in the “polyprotein state“, without significant

rearrangements during proteolytic maturation, exploiting a putative flexibility of the N-terminus of Nsp8 to interact with the C-terminus of Nsp7.

For proteolytic processing of a polyprotein, coronaviral main proteases require a glutamine in P1-position of the recognition site. Exchanging this essential residue by asparagine inhibits the cleavage completely, but interestingly still allows a strong interaction with the active centre of the protease. This state of interaction, likely sharing structural aspects with the native processing pathway *in vivo*, was investigated involving full-length polyprotein substrates. Peptides derived from main protease recognition sites carrying a P1-position mutation actually displayed inhibition of the protease activity; the lowest K_D -value determined by SPR spectroscopy is in the low μM range. The interaction was verified by native mass spectrometry. Consequently, these peptides comprise a starting point for novel main protease inhibitors.

The high-resolution structure of HCoV-229E Nsp10 was solved by X-ray crystallography, comprising the first Nsp10 structure of an *Alphacoronavirus*. As demonstrated, Nsp10 containing two highly conserved and occupied Zn^{2+} -binding sites is monomeric in solution. Interaction with Nsp7, Nsp8 or Nsp9 was not observed in a wide frame of experimental conditions. *In silico* modelling predicts an interaction with HCoV-229E Nsp16 and therefore the modulation of the Nsp16 2'-O-methyltransferase activity highly similar to that reported for SARS-CoV Nsp10 and Nsp16. Sequence variations within the interface are compensated by establishing alternative interactions.

In summary, pp710 as well as its partly processed and full processed maturation products that are assumed to constitute building blocks of the essential coronaviral RTC have been characterized in terms of function and structure applying state-of-the-art methods in structural biology. The direct comparison of equivalent proteins from *Alphacoronavirus* HCoV-229E and *Betacoronavirus* SARS-CoV overall revealed a high structural and functional conservation, despite the distant relationship within the *Coronaviridae* which enlightens conserved key-events in the process of RTC-formation of *Alpha*- and *Betacoronaviruses*. Significant differences were only detected in the tertiary structure of pp710 and pp78, according to SAXS.

6. Zusammenfassung

Coronaviren (CoV) sind membranumhüllte (-)ssRNA Viren, die Menschen bzw. verschiedene Tiere infizieren und in der Regel eher milde Erkältungskrankheiten auslösen. Das SARS-CoV und das sich aktuell ausbreitende MERS-CoV, welches bislang eine Sterblichkeitsrate von etwa 50% aufweist, repräsentieren jedoch Beispiele, die ein ernstes weltumspannendes Gesundheitsproblem darstellen. Neben den vielen bekannten nicht-pathogenen Coronaviren, wie z.B. HCoV-229E, müssen in Zukunft auch weitere Ausbrüche von schweren Atemwegserkrankungen befürchtet werden, die durch neuartige hoch-pathogene Coronaviren, die den Wirt wechseln, ausgelöst werden. Die coronavirale Haupt-Protease wurde bereits als effektives Ziel zur Behandlung von Infektionen identifiziert, bisher haben derartige Medikamente allerdings noch keine Marktreife erlangt. Eine weitere perspektivische Zielstruktur könnte der essentielle Transkriptase/Replikase-Komplex (RTC) darstellen, dessen Existenz und Aufbau unter anderem aus viralen Nicht-Struktur-Proteinen (Nsp) postuliert wird. Die genaue Zusammensetzung des RTC und vor allem die Bedeutung der primären Translationsprodukte des Virus, also der Polyproteine, aus denen durch virale Proteasen zunächst teilweise prozessierte Intermediate und final die reifen Nsps sequentiell freigesetzt werden, sind derzeit jedoch nahezu vollständig unbekannt.

Um wichtige Fragestellungen bezüglich der Struktur, der Stöchiometrie von oligomeren Zuständen sowie der Funktion von Polyprotein-Precursoren, die als mögliche Bausteine des RTC fungieren können, wurde ein großer Teil des C-Terminus des SARS-CoV Polyproteins 1a, bestehend aus den vier Domänen Nsp7-Nsp10 (pp710), rekombinant in *E. coli* produziert und gereinigt. Das homologe Protein des entfernt verwandten Erkältungserregers HCoV-229E wurde für vergleichende Untersuchungen ebenfalls hergestellt und gereinigt. Zusätzlich wurden spezifische Fragmente isoliert, die während der proteolytischen Prozessierung von pp710 durch die virale Haupt-Protease nachgewiesen wurden, um einen direkten Vergleich mit unvollständig prozessierten Polyprotein-Intermediaten und reifen Nsps zu ermöglichen. Meines Wissens wurde zum ersten Mal eine solche komplexe Prozessierungsreaktion eines viralen Polyproteins mit hoher Massengenauigkeit durch native Massenspektrometrie verfolgt, wodurch unter anderem ein 2:2-Interaktionszustand von SARS-CoV Nsp7 und der RNA-abhängigen RNA-Polymerase Nsp8 in Lösung identifiziert werden konnte. Weiterhin wurde nachgewiesen, dass die Polymerase-Aktivität nicht zwingend einen hexadecameren Zustand von SARS-CoV Nsp7 und Nsp8 erfordert, der zuvor postuliert wurde. Bereits im Polyprotein-Zustand vor der N- und C-terminalen Prozessierung ist die Nsp8-Domäne in der Lage, größere RNA-Fragmente (> 10 nt) zu synthetisieren. Das Bindungsepitop von Nukleotiden an zwei Nsp8-Vorläuferproteine wurde mittels Sättigungs-Transfer-Differenz NMR Spektroskopie charakterisiert. Der Nachweis einer Interaktion mit verschiedenen Nukleinsäuren und der katalytischen Aktivität von Polyproteinen bekräftigt die Hypothese, dass unvollständig prozessierte Polyproteine bereits funktionelle Bausteine des RTC mit „vorgeformten“ Struktureigenschaften sein können. Weiterhin

wurde durch Größen-Ausschluss-Chromatographie (SEC) und Röntgen-Kleinwinkel-Beugung (SAXS) eine konzentrationsabhängige Oligomerisierung von SARS-CoV und HCoV-229E pp710 nachgewiesen, was möglicherweise einen Regulationsmechanismus während der Coronavirus-Infektion darstellen könnte. Ein strukturelles Modell der Inter-Domänen-Interaktionen im pp710-Monomer sowie der intermolekularen Dimer-Interaktionen wurde mit niedriger Auflösung erstellt und basierend auf SAXS- und Massenspektrometrie-Daten verifiziert. Dieses Modell sowie das SAXS-Modell von pp78 (Polyprotein Nsp7-Nsp8) legen nahe, dass die Interaktion von „reifem“ Nsp7 und Nsp8 im Polyprotein-Stadium bereits vorgeformt ist, ohne dass während der Prozessierung signifikante Umlagerungen von Nsp7 erforderlich wären. Bei konserviertem Bindungssepitop kann die im Polyprotein-Stadium erforderlich kovalente Verknüpfung von Nsp7 und Nsp8 durch die postulierte Flexibilität des N-Terminus von Nsp8 ermöglicht werden.

Die proteolytische Prozessierung des Polyproteins durch die coronavirale Haupt-Protease erfordert ein Glutamin in P1-Position der Erkennungsstelle. Ein Austausch dieses essentiellen Restes gegen Asparagin inhibiert die Katalyse vollständig, erlaubt aber interessanterweise noch immer eine starke Interaktion von Protease und Substrat. Dieser Interaktionszustand, der wahrscheinlich strukturelle Aspekte des nativen Prozessierungsablaufs *in vivo* widerspiegelt, wurde mit vollständigen Polyprotein-Substraten mittels SAXS-Technik charakterisiert. Peptide mit einer Mutation in P1-Position, die von den nativen Protease-Schnittstellen abgeleitet wurden, inhibierten die Protease. Der niedrigste durch SPR-Spektroskopie bestimmte K_D -Wert liegt im niedrigen μM -Bereich. Die Interaktion wurde durch Massenspektrometrie bestätigt. Folglich repräsentieren diese Peptide einen weiteren Ansatzpunkt zur Entwicklung neuartiger Haupt-Protease-Inhibitoren.

Des Weiteren wird mit der röntgenkristallographischen Aufklärung der Struktur des HCoV-229E Nsp10, einem vielseitigen regulatorischen coronaviralen Protein, die erste Nsp10-Struktur eines *Alphacoronavirus* präsentiert. Nsp10 beinhaltet zwei hochkonservierte und besetzte Zn^{2+} -Bindungsstellen und weist in Lösung eindeutig einen monomeren Zustand auf. Unter verschiedensten experimentellen Bedingungen konnte keine Interaktion mit reifem Nsp7, Nsp8 oder Nsp9 nachgewiesen werden. Durch *in silico* Modellierung wurde eine Interaktion von Nsp10 mit HCoV-229E Nsp16 und somit eine Regulation der 2'-O-Methyltransferase-Aktivität von Nsp16 entsprechend der für SARS-CoV berichteten vorhergesagt. Sequenzvariationen im Bereich der Interaktionsflächen werden durch Etablierung alternativer Wechselwirkungen kompensiert.

Insgesamt wurde das coronavirale Polyprotein pp710 sowie teilweise und auch vollständig prozessierte Spaltprodukte, die als unterschiedliche „Bausteine“ des essentiellen RTC postuliert werden, bezüglich ihrer funktionellen und strukturellen Eigenschaften charakterisiert. Der direkte Vergleich entsprechender Proteine des *Alphacoronavirus* HCoV-229E und des *Betacoronavirus*

SARS-CoV zeigte generell ein hohes Ausmaß an funktioneller und struktureller Konservierung, trotz der Entfernung beider Gruppen innerhalb der Coronavirus-Familie. Signifikante Unterschiede beschränken sich auf die Tertiärstruktur von pp710 und pp78, basierend auf SAXS-Daten.

7. References

- [1] Fauquet, CM.; Fargette, D. International Committee on Taxonomy of Viruses and the 3,142 unassigned species. *Virology journal*. 2005. 2. 64.
- [2] den Boon, J A; Snijder, EJ.; Chirnside, ED.; de Vries, A A; Horzinek, MC.; Spaan, WJ. Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *Journal of virology*. 1991. 65. 6. 2910–2920.
- [3] Snijder, EJ.; Bredenbeek, PJ.; Dobbe, JC.; Thiel, V; Ziebuhr, J; Poon, Leo L M; Guan, Y; Rozanov, M; Spaan, Willy J M; Gorbalenya, AE. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *Journal of molecular biology*. 2003. 331. 5. 991–1004.
- [4] Tyrrell, DA.; Bynoe, ML. Cultivation Of A Novel Type Of Common-Cold Virus In Organ Cultures. *British medical journal*. 1965. 1. 5448. 1467–1470.
- [5] Peiris, J S M; Lai, ST.; Poon, L L M; Guan, Y; Yam, L Y C; Lim, W; Nicholls, J; Yee, W K S; Yan, WW.; Cheung, MT.; Cheng, V C C; Chan, KH.; Tsang, D N C; Yung, R W H; Ng, TK.; Yuen, KY. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*. 2003. 361. 9366. 1319–1325.
- [6] Donnelly, CA.; Ghani, AC.; Leung, GM.; Hedley, AJ.; Fraser, C; Riley, S; Abu-Raddad, LJ.; Ho, L; Thach, T; Chau, P; Chan, K; Lam, T; Tse, L; Tsang, T; Liu, S; Kong, James H B; Lau, Edith M C; Ferguson, NM.; Anderson, RM. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet*. 2003. 361. 9371. 1761–1766.
- [7] Guan, Y; Zheng, BJ.; He, YQ.; Liu, XL.; Zhuang, ZX.; Cheung, CL.; Luo, SW.; Li, PH.; Zhang, LJ.; Guan, YJ.; Butt, KM.; Wong, KL.; Chan, KW.; Lim, W; Shortridge, KF.; Yuen, KY.; Peiris, J S M; Poon, L L M. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science (New York, N.Y.)*. 2003. 302. 5643. 276–278.
- [8] Lim, PLian; Lee, THong; Rowe, EK. Middle East Respiratory Syndrome coronavirus (MERS CoV): Update 2013. *Current infectious disease reports*. 2013.
- [9] van der Hoek, Lia; Pyrc, K; Jebbink, MF.; Vermeulen-Oost, W; Berkhout, Ron J M; Wolthers, KC.; Wertheim-van Dillen, Pauline M E; Kaandorp, J; Spaargaren, J; Berkhout, B. Identification of a new human coronavirus. *Nature medicine*. 2004. 10. 4. 368–373.
- [10] Rottier, Peter J M; Nakamura, K; Schellen, P; Volders, H; Haijema, BJan. Acquisition of macrophage tropism during the pathogenesis of feline infectious peritonitis is determined by mutations in the feline coronavirus spike protein. *Journal of virology*. 2005. 79. 22. 14122–14130.
- [11] Stadler, K; Massignani, V; Eickmann, M; Becker, S; Abrignani, S; Klenk, H; Rappuoli, R. SARS--beginning to understand a new virus. *Nature reviews. Microbiology*. 2003. 1. 3. 209–218.
- [12] Larson, HE.; Reed, SE.; TYRRELL, DA. Isolation of rhinoviruses and coronaviruses from 38 colds in adults. *Journal of medical virology*. 1980. 5. 3. 221–229.
- [13] Pene, F; Merlat, A; Vabret, A; Rozenberg, F; Buzyn, A; Dreyfus, F; Cariou, A; Freymuth, F; Lebon, P. Coronavirus 229E-related pneumonia in immunocompromised patients. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2003. 37. 7. 929–932.
- [14] Marra, MA.; Jones, Steven J M; Astell, CR.; Holt, RA.; Brooks-Wilson, A; Butterfield, Yaron S N; Khattra, J; Asano, JK.; Barber, SA.; Chan, SY.; Cloutier, A; Coughlin, SM.; Freeman, D; Girn, N; Griffith, OL.; Leach, SR.; Mayo, M; McDonald, H; Montgomery, SB.; Pandoh, PK.;

- Petrescu, AS.; Robertson, AGordon; Schein, JE.; Siddiqui, A; Smailus, DE.; Stott, JM.; Yang, GS.; Plummer, F; Andonov, A; Artsob, H; Bastien, N; Bernard, K; Booth, TF.; Bowness, D; Czub, M; Drebot, M; Fernando, L; Flick, R; Garbutt, M; Gray, M; Grolla, A; Jones, S; Feldmann, H; Meyers, A; Kabani, A; Li, Y; Normand, S; Stroher, U; Tipples, GA.; Tyler, S; Vogrig, R; Ward, D; Watson, B; Brunham, RC.; Krajdén, M; Petric, M; Skowronski, DM.; Upton, C; Roper, RL. The Genome sequence of the SARS-associated coronavirus. *Science (New York, N.Y.)*. 2003. 300. 5624. 1399–1404.
- [15] Raman, S; Bouma, P; Williams, GD.; Brian, DA. Stem-loop III in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. *Journal of virology*. 2003. 77. 12. 6720–6730.
- [16] Goebel, SJ.; Miller, TB.; Bennett, CJ.; Bernard, KA.; Masters, PS. A hypervariable region within the 3' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis. *Journal of virology*. 2007. 81. 3. 1274–1287.
- [17] Spagnolo, JF.; Hogue, BG. Host protein interactions with the 3' end of bovine coronavirus RNA and the requirement of the poly(A) tail for coronavirus defective genome replication. *Journal of virology*. 2000. 74. 11. 5053–5065.
- [18] Sawicki, SG.; Sawicki, DL.; Siddell, SG. A Contemporary View of Coronavirus Transcription. *Journal of Virology*. 2006. 81. 1. 20–29.
- [19] Enjuanes, L; Almazán, F; Sola, I; Zuñiga, S. Biochemical aspects of coronavirus replication and virus-host interaction. *Annual review of microbiology*. 2006. 60. 211–230.
- [20] Thiel, V; Ivanov, KA.; Putics, A; Hertzog, T; Schelle, B; Bayer, S; Weissbrich, B; Snijder, EJ.; Rabenau, H; Doerr, HWilhelm; Gorbalenya, AE.; Ziebuhr, J. Mechanisms and enzymes involved in SARS coronavirus genome expression. *The Journal of general virology*. 2003. 84. Pt 9. 2305–2315.
- [21] Morales, L; Mateos-Gomez, PA.; Capiscol, C; del Palacio, L; Enjuanes, L; Sola, I. Transmissible gastroenteritis coronavirus genome packaging signal is located at the 5' end of the genome and promotes viral RNA incorporation into virions in a replication independent process. *Journal of Virology*. 2013.
- [22] Escors, D; Ortego, J; Laude, H; Enjuanes, L. The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. *Journal of virology*. 2001. 75. 3. 1312–1324.
- [23] Lu, G; Hu, Y; Wang, Q; Qi, J; Gao, F; Li, Y; Zhang, Y; Zhang, W; Yuan, Y; Bao, J; Zhang, B; Shi, Y; Yan, J; Gao, GF. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature*. 2013. 500. 7461. 227–231.
- [24] Bosch, BJan; de Haan, Cornelis A M; Rottier, Peter J M. Coronavirus spike glycoprotein, extended at the carboxy terminus with green fluorescent protein, is assembly competent. *Journal of virology*. 2004. 78. 14. 7369–7378.
- [25] Wang, S; Guo, F; Liu, K; Wang, H; Rao, S; Yang, P; Jiang, C. Endocytosis of the receptor-binding domain of SARS-CoV spike protein together with virus receptor ACE2. *Virus research*. 2008. 136. 1-2. 8–15.
- [26] Pewe, L; Zhou, H; Netland, J; Tangudu, C; Olivares, H; Shi, L; Look, D; Gallagher, T; Perlman, S. A severe acute respiratory syndrome-associated coronavirus-specific protein enhances virulence of an attenuated murine coronavirus. *Journal of virology*. 2005. 79. 17. 11335–11342.
- [27] Tangudu, C; Olivares, H; Netland, J; Perlman, S; Gallagher, T. Severe acute respiratory syndrome coronavirus protein 6 accelerates murine coronavirus infections. *Journal of virology*. 2007. 81. 3. 1220–1229.
- [28] Law, P. T. W. The 3a protein of severe acute respiratory syndrome-associated coronavirus induces apoptosis in Vero E6 cells. *Journal of General Virology*. 2005. 86. 7. 1921–1930.

- [29] Yuan, X; Wu, J; Shan, Y; Yao, Z; Dong, B; Chen, B; Zhao, Z; Wang, S; Chen, J; Cong, Y. SARS coronavirus 7a protein blocks cell cycle progression at G0/G1 phase via the cyclin D3/pRb pathway. *Virology*. 2006. 346. 1. 74–85.
- [30] van Hemert, Martijn J; de Wilde, Adriaan H; Gorbalenya, AE.; Snijder, EJ. The in vitro RNA synthesizing activity of the isolated arterivirus replication/transcription complex is dependent on a host factor. *The Journal of biological chemistry*. 2008. 283. 24. 16525–16536.
- [31] Su, M; Chang, C; Chu, C; Tsai, C; Chang, K. An atypical RNA pseudoknot stimulator and an upstream attenuation signal for -1 ribosomal frameshifting of SARS coronavirus. *Nucleic acids research*. 2005. 33. 13. 4265–4275.
- [32] Garcia (Ed.) 2012 – Viral Genomes - Molecular Structure, Diversity, Gene Expression Mechanisms and Host-Virus Interactions, InTech, ISBN: 978-953-51-0098-0
- [33] Knoops, K; Kikkert, M; Worm, Sjoerd H E van den; Zevenhoven-Dobbe, JC.; van der Meer, Yvonne; Koster, AJ.; Mommaas, AMieke; Snijder, EJ. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS biology*. 2008. 6. 9. e226.
- [34] Angelini, MM.; Akhlaghpour, M; Neuman, BW.; Buchmeier, MJ. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio*. 2013. 4. 4.
- [35] Snijder, EJ.; van Tol, H; Roos, N; Pedersen, KW. Non-structural proteins 2 and 3 interact to modify host cell membranes during the formation of the arterivirus replication complex. *The Journal of general virology*. 2001. 82. Pt 5. 985–994.
- [36] Clementz, MA.; Kanjanahaluethai, A; O'Brien, TE.; Baker, SC. Mutation in murine coronavirus replication protein nsp4 alters assembly of double membrane vesicles. *Virology*. 2008. 375. 1. 118–129.
- [37] Cho, JKeun; Curtis-Long, MJ.; Lee, KHo; Kim, DWook; Ryu, HWon; Yuk, HJoo; Park, KHun. Geranylated flavonoids displaying SARS-CoV papain-like protease inhibition from the fruits of *Paulownia tomentosa*. *Bioorganic & medicinal chemistry*. 2013. 21. 11. 3051–3057.
- [38] Tan, J; Kusov, Y; Mutschall, D; Tech, S; Nagarajan, K; Hilgenfeld, R; Schmidt, CL. The "SARS-unique domain" (SUD) of SARS coronavirus is an oligo(G)-binding protein. *Biochemical and biophysical research communications*. 2007. 364. 4. 877–882.
- [39] Tan, J; Vonnrhein, C; Smart, OS.; Bricogne, G; Bollati, M; Kusov, Y; Hansen, G; Mesters, JR.; Schmidt, CL.; Hilgenfeld, R. The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *PLoS pathogens*. 2009. 5. 5. e1000428.
- [40] Kiemer, L; Lund, O; Brunak, S; Blom, N. Coronavirus 3CLpro proteinase cleavage sites: possible relevance to SARS virus pathology. *BMC bioinformatics*. 2004. 5. 72.
- [41] Denison, MR.; Zoltick, PW.; Hughes, SA.; Giangreco, B; Olson, AL.; Perlman, S; Leibowitz, JL.; Weiss, SR. Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. *Virology*. 1992. 189. 1. 274–284.
- [42] Denison, MR.; Hughes, SA.; Weiss, SR. Identification and characterization of a 65-kDa protein processed from the gene 1 polyprotein of the murine coronavirus MHV-A59. *Virology*. 1995. 207. 1. 316–320.
- [43] Lu, XT.; Sims, AC.; Denison, MR. Mouse hepatitis virus 3C-like protease cleaves a 22-kilodalton protein from the open reading frame 1a polyprotein in virus-infected cells and in vitro. *Journal of virology*. 1998. 72. 3. 2265–2271.
- [44] Schiller, JJ.; Kanjanahaluethai, A; Baker, SC. Processing of the coronavirus MHV-JHM polymerase polyprotein: identification of precursors and proteolytic products spanning 400 kilodaltons of ORF1a. *Virology*. 1998. 242. 2. 288–302.

- [45] Schiller, JJ.; Baker, SC. Maturation of the polymerase polyprotein of the coronavirus MHV strain JHM involves a cascade of proteolytic processing events. *Advances in experimental medicine and biology*. 1998. 440. 135–139.
- [46] Heusipp, G; Grötzinger, C; Herold, J; Siddell, SG.; Ziebuhr, J. Identification and subcellular localization of a 41 kDa, polyprotein 1ab processing product in human coronavirus 229E-infected cells. *The Journal of general virology*. 1997. 78 (Pt 11). 2789–2794.
- [47] Hagemeyer, MC.; Rottier, Peter J M; de Haan, Cornelis A M. Biogenesis and dynamics of the coronavirus replicative structures. *Viruses*. 2012. 4. 11. 3245–3269.
- [48] Chen, S; Jonas, F; Shen, C; Hilgenfeld, R; Higenfeld, R. Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode. *Protein & cell*. 2010. 1. 1. 59–74.
- [49] Barrila, J; Bacha, U; Freire, E. Long-range cooperative interactions modulate dimerization in SARS 3CLpro. *Biochemistry*. 2006. 45. 50. 14908–14916.
- [50] Chen, S; Hu, T; Zhang, J; Chen, J; Chen, K; Ding, J; Jiang, H; Shen, X. Mutation of Gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3C-like protease: crystal structure with molecular dynamics simulations. *The Journal of biological chemistry*. 2008. 283. 1. 554–564.
- [51] Chen, S; Jonas, F; Shen, C; Hilgenfeld, R; Higenfeld, R. Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode. *Protein & cell*. 2010. 1. 1. 59–74.
- [52] Chen, H; Wei, P; Huang, C; Tan, L; Liu, Y; Lai, L. Only one protomer is active in the dimer of SARS 3C-like proteinase. *The Journal of biological chemistry*. 2006. 281. 20. 13894–13898.
- [53] Yang, H; Yang, M; Ding, Y; Liu, Y; Lou, Z; Zhou, Z; Sun, L; Mo, L; Ye, S; Pang, H; Gao, GF.; Anand, K; Bartlam, M; Hilgenfeld, R; Rao, Z. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proceedings of the National Academy of Sciences of the United States of America*. 2003. 100. 23. 13190–13195.
- [54] Hilgenfeld, R; Anand, K; Mesters, JR.; Rao, Z; Shen, X; Jiang, H; Tan, J; Verschueren, Koen H G. Structure and dynamics of SARS coronavirus main proteinase (Mpro). *Advances in experimental medicine and biology*. 2006. 581. 585–591.
- [55] Steuber, H; Hilgenfeld, R. Recent advances in targeting viral proteases for the discovery of novel antivirals. *Current topics in medicinal chemistry*. 2010. 10. 3. 323–345.
- [56] Shi, J; Sivaraman, J; Song, J. Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. *Journal of virology*. 2008. 82. 9. 4620–4629.
- [57] Zhang, S; Zhong, N; Xue, F; Kang, X; Ren, X; Chen, J; Jin, C; Lou, Z; Xia, B. Three-dimensional domain swapping as a mechanism to lock the active conformation in a super-active octamer of SARS-CoV main protease. *Protein & cell*. 2010. 1. 4. 371–383.
- [58] Chou, C; Chang, H; Hsu, W; Lin, T; Lin, C; Chang, G. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry*. 2004. 43. 47. 14958–14970.
- [59] Chen, S; Chen, L; Tan, J; Chen, J; Du, L; Sun, T; Shen, J; Chen, K; Jiang, H; Shen, X. Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. *The Journal of biological chemistry*. 2005. 280. 1. 164–173.
- [60] Graziano, V; McGrath, WJ.; Yang, L; Mangel, WF. SARS CoV main proteinase: The monomer-dimer equilibrium dissociation constant. *Biochemistry*. 2006. 45. 49. 14632–14641.

- [61] Huang, C; Wei, P; Fan, K; Liu, Y; Lai, L. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry*. 2004. 43. 15. 4568–4574.
- [62] Solowiej, J; Thomson, JA.; Ryan, K; Luo, C; He, M; Lou, J; Murray, BW. Steady-State and Pre-Steady-State Kinetic Evaluation of Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL pro Cysteine Protease: Development of an Ion-Pair Model for Catalysis. *Biochemistry*. 2008. 47. 8. 2617–2630.
- [63] Anand, K; Ziebuhr, J; Wadhwani, P; Mesters, JR.; Hilgenfeld, R. Coronavirus main proteinase (3CL^{pro}) structure: basis for design of anti-SARS drugs. *Science (New York, N.Y.)*. 2003. 300. 5626. 1763–1767.
- [64] Goetz, DH.; Choe, Y; Hansell, E; Chen, YT.; McDowell, M; Jonsson, CB.; Roush, WR.; McKerrow, J; Craik, CS. Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS coronavirus. *Biochemistry*. 2007. 46. 30. 8744–8752.
- [65] Chu, LMatthew; Choy, W; Tsai, S; Rao, Z; Ngai, S. Rapid peptide-based screening on the substrate specificity of severe acute respiratory syndrome (SARS) coronavirus 3C-like protease by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Protein science : a publication of the Protein Society*. 2006. 15. 4. 699–709.
- [66] Mandadapu, SRao; Weerawarna, PM.; Prior, AM.; Uy, Roxanne Adeline Z; Aravapalli, S; Alliston, KR.; Lushington, GH.; Kim, Y; Hua, DH.; Chang, K; Groutas, WC. Macrocyclic inhibitors of 3C and 3C-like proteases of picornavirus, norovirus, and coronavirus. *Bioorganic & medicinal chemistry letters*. 2013. 23. 13. 3709–3712.
- [67] Nguyen, Thi Thanh Hanh; Woo, H; Kang, H; van Nguyen, D; Kim, Y; Kim, D; Ahn, S; Xia, Y; Kim, D. Flavonoid-mediated inhibition of SARS coronavirus 3C-like protease expressed in *Pichia pastoris*. *Biotechnology Letters*. 2012. 34. 5. 831–838.
- [68] Chen, C; Lin, Coney P C; Huang, K; Chen, W; Hsieh, H; Liang, P; Hsu, JT-A. Inhibition of SARS-CoV 3C-like Protease Activity by Theaflavin-3,3'-digallate (TF3). *Evidence-based complementary and alternative medicine : eCAM*. 2005. 2. 2. 209–215.
- [69] Liu, X; Kim, J; Li, Y; Li, J; Liu, F; Chen, X. Tannic acid stimulates glucose transport and inhibits adipocyte differentiation in 3T3-L1 cells. *The Journal of nutrition*. 2005. 135. 2. 165–171.
- [70] Tadera, K; Minami, Y; Takamatsu, K; Matsuoka, T. Inhibition of alpha-glucosidase and alpha-amylase by flavonoids. *Journal of nutritional science and vitaminology*. 2006. 52. 2. 149–153.
- [71] Groot, H de; Rauen, U. Tissue injury by reactive oxygen species and the protective effects of flavonoids. *Fundamental & clinical pharmacology*. 1998. 12. 3. 249–255.
- [72] Thanigaimalai, P; Konno, S; Yamamoto, T; Koiwai, Y; Taguchi, A; Takayama, K; Yakushiji, F; Akaji, K; Chen, S; Naser-Tavakolian, A; Schön, A; Freire, E; Hayashi, Y. Development of potent dipeptide-type SARS-CoV 3CL protease inhibitors with novel P3 scaffolds: Design, synthesis, biological evaluation, and docking studies. *European journal of medicinal chemistry*. 2013. 68C. 372–384.
- [73] Thanigaimalai, P; Konno, S; Yamamoto, T; Koiwai, Y; Taguchi, A; Takayama, K; Yakushiji, F; Akaji, K; Kiso, Y; Kawasaki, Y; Chen, S; Naser-Tavakolian, A; Schön, A; Freire, E; Hayashi, Y. Design, synthesis, and biological evaluation of novel dipeptide-type SARS-CoV 3CL protease inhibitors: structure-activity relationship study. *European journal of medicinal chemistry*. 2013. 65. 436–447.
- [74] Ren, Z; Yan, L; Zhang, N; Guo, Y; Yang, C; Lou, Z; Rao, Z. The newly emerged SARS-Like coronavirus HCoV-EMC also has an “Achilles’ heel”: current effective inhibitor targeting a 3C-like protease. *Protein & cell*. 2013. 4. 4. 248–250.

- [75] Xue, X; Yu, H; Yang, H; Xue, F; Wu, Z; Shen, W; Li, J; Zhou, Z; Ding, Y; Zhao, Q; Zhang, XC.; Liao, M; Bartlam, M; Rao, Z. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *Journal of virology*. 2008. 82. 5. 2515–2527.
- [76] Shie, J; Fang, J; Kuo, T; Kuo, C; Liang, P; Huang, H; Wu, Y; Jan, J; Cheng, YE.; Wong, C. Inhibition of the severe acute respiratory syndrome 3CL protease by peptidomimetic alpha,beta-unsaturated esters. *Bioorganic & medicinal chemistry*. 2005. 13. 17. 5240–5252.
- [77] Lee, C; Kuo, C; Ko, T; Hsu, M; Tsui, Y; Chang, S; Yang, S; Chen, S; Chen, H; Hsu, M; Shih, S; Liang, P; Wang, AH-J. Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds. *The Journal of biological chemistry*. 2009. 284. 12. 7646–7655.
- [78] Zhu, L; George, S; Schmidt, MF.; Al-Gharabli, SI.; Rademann, J; Hilgenfeld, R. Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. *Antiviral research*. 2011. 92. 2. 204–212.
- [79] Lee, T; Cherney, MM.; Liu, J; James, KELLIS; Powers, JC.; Eltis, LD.; James, Michael N G. Crystal structures reveal an induced-fit binding of a substrate-like Aza-peptide epoxide to SARS coronavirus main peptidase. *Journal of molecular biology*. 2007. 366. 3. 916–932.
- [80] Chuck, C; Chen, C; Ke, Z; Chi-Cheong Wan, D; Chow, H; Wong, K. Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases. *European Journal of Medicinal Chemistry*. 2013. 59. 1–6.
- [81] Wu, C; King, K; Kuo, C; Fang, J; Wu, Y; Ho, M; Liao, C; Shie, J; Liang, P; Wong, C. Stable benzotriazole esters as mechanism-based inactivators of the severe acute respiratory syndrome 3CL protease. *Chemistry & biology*. 2006. 13. 3. 261–268.
- [82] Jansson, AM. Structure of Alphacoronavirus Transmissible Gastroenteritis Virus nsp1 Has Implications for Coronavirus nsp1 Function and Evolution. *Journal of Virology*. 2013. 87. 5. 2949–2955.
- [83] Lokugamage, KG.; Narayanan, K; Huang, C; Makino, S. Severe Acute Respiratory Syndrome Coronavirus Protein nsp1 Is a Novel Eukaryotic Translation Inhibitor That Represses Multiple Steps of Translation Initiation. *Journal of Virology*. 2012. 86. 24. 13598–13608.
- [84] Yu, K; Ming, Z; Li, Y; Chen, C; Bao, Z; Ren, Z; Liu, B; Tao, W; Rao, Z; Lou, Z. Purification, crystallization and preliminary X-ray analysis of nonstructural protein 2 (nsp2) from avian infectious bronchitis virus. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*. 2012. 68. 6. 716–719.
- [85] Wojdyla, JA.; Manolaridis, I; Snijder, EJ.; Gorbalenya, AE.; Coutard, B; Piotrowski, Y; Hilgenfeld, R; Tucker, PA. Structure of the X (ADRP) domain of nsp3 from feline coronavirus. *Acta crystallographica. Section D, Biological crystallography*. 2009. 65. Pt 12. 1292–1300.
- [86] Yu, M; Lee, J; Lee, JMoo; Kim, Y; Chin, Y; Jee, J; Keum, Y; Jeong, Y. Identification of myricetin and scutellarein as novel chemical inhibitors of the SARS coronavirus helicase, nsP13. *Bioorganic & medicinal chemistry letters*. 2012. 22. 12. 4049–4054.
- [87] Brunn, A von; Teepe, C; Simpson, JC.; Pepperkok, R; Friedel, CC.; Zimmer, R; Roberts, R; Baric, R; Haas, J. Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFome. *PloS one*. 2007. 2. 5. e459.
- [88] Pan, J; Peng, X; Gao, Y; Li, Z; Lu, X; Chen, Y; Ishaq, M; Liu, D; Dediego, ML.; Enjuanes, L; Guo, D. Genome-wide analysis of protein-protein interactions and involvement of viral proteins in SARS-CoV replication. *PloS one*. 2008. 3. 10. e3299.
- [89] Ito, N; Mossel, EC.; Narayanan, K; Popov, VL.; Huang, C; Inoue, T; Peters, CJ.; Makino, S. Severe acute respiratory syndrome coronavirus 3a protein is a viral structural protein. *Journal of virology*. 2005. 79. 5. 3182–3186.

- [90] Shen, S; Lin, P; Chao, Y; Zhang, A; Yang, X; Lim, S; Gee, Hong, W; Tan, Y. The severe acute respiratory syndrome coronavirus 3a is a novel structural protein. *Biochemical and biophysical research communications*. 2005. 330. 1. 286–292.
- [91] Palaninathan, S.; Bhardwaj, K.; Alcantara, J.M.O.; Guarino, L.; Yi, L.L.; Kao, C.C.; Sacchettini, J. Crystal structure of Nsp15-H234A mutant- Hexamer in asymmetric unit. *Journal of Biological Chemistry*. 2008. 283. 3655 – 3664.
- [92] Bhardwaj, K; Liu, P; Leibowitz, J.L.; Kao, C.C. The Coronavirus Endoribonuclease Nsp15 Interacts with Retinoblastoma Tumor Suppressor Protein. *Journal of Virology*. 2012. 86. 8. 4294–4304.
- [93] Ahn, D; Choi, J; Taylor, D.R.; Oh, J. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. *Archives of Virology*. 2012. 157. 11. 2095–2104.
- [94] Imbert, I; Guillemot, J; Bourhis, J; Bussetta, C; Coutard, B; Egloff, M; Ferron, F; Gorbalenya, A.E.; Canard, B. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *The EMBO journal*. 2006. 25. 20. 4933–4942.
- [95] Hansen, J.L.; Long, A.M.; Schultz, S.C. Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure*. 1997. 5. 8. 1109–1122.
- [96] Xiao, Y; Ma, Q; Restle, T; Shang, W; Svergun, D.I.; Ponnusamy, R; Sczakiel, G; Hilgenfeld, R. Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. *Journal of virology*. 2012. 86. 8. 4444–4454.
- [97] te Velthuis, Aartjan J W; van den Worm, Sjoerd H E; Snijder, E.J. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic acids research*. 2012. 40. 4. 1737–1747.
- [98] Zhai, Y; Sun, F; Li, X; Pang, H; Xu, X; Bartlam, M; Rao, Z. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. *Nature structural & molecular biology*. 2005. 12. 11. 980–986.
- [99] Li, S; Zhao, Q; Zhang, Y; Zhang, Y; Bartlam, M; Li, X; Rao, Z. New nsp8 isoform suggests mechanism for tuning viral RNA synthesis. *Protein & cell*. 2010. 1. 2. 198–204.
- [100] Johnson, M.A.; Jaudzems, K; Wüthrich, K. NMR Structure of the SARS-CoV Nonstructural Protein 7 in Solution at pH 6.5. *Journal of molecular biology*. 2010. 402. 4. 619–628.
- [101] Ponnusamy, R; Moll, R; Weimar, T; Mesters, J.R.; Hilgenfeld, R. Variable oligomerization modes in coronavirus non-structural protein 9. *Journal of molecular biology*. 2008. 383. 5. 1081–1096.
- [102] Sutton, G; Fry, E; Carter, L; Sainsbury, S; Walter, T; Nettleship, J; Berrow, N; Owens, R; Gilbert, R; Davidson, A; Siddell, S; Poon, Leo L M; Diprose, J; Alderton, D; Walsh, M; Grimes, J.M.; Stuart, D.I. The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure (London, England : 1993)*. 2004. 12. 2. 341–353.
- [103] Ponnusamy, R; Mesters, J.R.; Ziebuhr, J; Moll, R; Hilgenfeld, R. Non structural proteins 8 and 9 of human coronavirus 229E. *Advances in experimental medicine and biology*. 2006. 581. 49–54.
- [104] Miknis, Z.J.; Donaldson, E.F.; Umland, T.C.; Rimmer, R.A.; Baric, R.S.; Schultz, L.Wayne. Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *Journal of virology*. 2009. 83. 7. 3007–3018.
- [105] Webster, G; Genschel, J; Curth, U; Urbanke, C; Kang, C; Hilgenfeld, R. A common core for binding single-stranded DNA: structural comparison of the single-stranded DNA-binding proteins (SSB) from *E. coli* and human mitochondria. *FEBS letters*. 1997. 411. 2-3. 313–316.

- [106] Mapelli, M; Panjkar, S; Tucker, PA. The crystal structure of the herpes simplex virus 1 ssDNA-binding protein suggests the structural basis for flexible, cooperative single-stranded DNA binding. *The Journal of biological chemistry*. 2005. 280. 4. 2990–2997.
- [107] Deming, DJ.; Graham, RL.; Denison, MR.; Baric, RS. Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. *Journal of virology*. 2007. 81. 19. 10280–10291.
- [108] Züst, R; Miller, TB.; Goebel, SJ.; Thiel, V; Masters, PS. Genetic interactions between an essential 3' cis-acting RNA pseudoknot, replicase gene products, and the extreme 3' end of the mouse coronavirus genome. *Journal of virology*. 2008. 82. 3. 1214–1228.
- [109] Geng, H; Liu, Y; Chan, W; Lo, AWing-Ip; Au, DMun-Yee; Waye, MMiu-Yee; Ho, Y. The putative protein 6 of the severe acute respiratory syndrome-associated coronavirus: expression and functional characterization. *FEBS letters*. 2005. 579. 30. 6763–6768.
- [110] Lai, MM.; Stohlman, SA. Genome structure of mouse hepatitis virus: comparative analysis by oligonucleotide mapping. *Advances in experimental medicine and biology*. 1981. 142. 69–82.
- [111] Joseph, JS.; Saikatendu, KSingh; Subramanian, V; Neuman, BW.; Brooun, A; Griffith, M; Moy, K; Yadav, MK.; Velasquez, J; Buchmeier, MJ.; Stevens, RC.; Kuhn, P. Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs. *Journal of virology*. 2006. 80. 16. 7894–7901.
- [112] 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. Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS pathogens*. 2011. 7. 5. e1002059.
- [113] Ke, M; Chen, Y; Wu, A; Sun, Y; Su, C; Wu, H; Jin, X; Tao, J; Wang, Y; Ma, X; Pan, J; Guo, D. Short peptides derived from the interaction domain of SARS coronavirus nonstructural protein nsp10 can suppress the 2'-O-methyltransferase activity of nsp10/nsp16 complex. *Virus research*. 2012. 167. 2. 322–328.
- [114] Debarnot, C; Imbert, I; Ferron, F; Gluais, L; Varlet, I; Papageorgiou, N; Bouvet, M; Lescar, J; Decroly, E; Canard, B. Crystallization and diffraction analysis of the SARS coronavirus nsp10-nsp16 complex. *Acta crystallographica. Section F, Structural biology and crystallization communications*. 2011. 67. Pt 3. 404–408.
- [115] Chen, Y; Tao, J; Sun, Y; Wu, A; Su, C; Gao, G; Cai, H; Qiu, S; Wu, Y; Ahola, T; Guo, D. Structure-function analysis of severe acute respiratory syndrome coronavirus RNA cap guanine-N7-methyltransferase. *Journal of virology*. 2013. 87. 11. 6296–6305.
- [116] Chen, P; Jiang, M; Hu, T; Liu, Q; Chen, XS.; Guo, D. Biochemical characterization of exoribonuclease encoded by SARS coronavirus. *Journal of biochemistry and molecular biology*. 2007. 40. 5. 649–655.
- [117] Bouvet, M; Imbert, I; Subissi, L; Gluais, L; Canard, B; Decroly, E. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proceedings of the National Academy of Sciences of the United States of America*. 2012. 109. 24. 9372–9377.
- [118] Donaldson, EF.; Sims, AC.; Graham, RL.; Denison, MR.; Baric, RS. Murine hepatitis virus replicase protein nsp10 is a critical regulator of viral RNA synthesis. *Journal of virology*. 2007. 81. 12. 6356–6368.
- [119] Daffis, S; Szretter, KJ.; Schriewer, J; Li, J; Youn, S; Errett, J; Lin, T; Schneller, S; Züst, R; Dong, H; Thiel, V; Sen, GC.; Fensterl, V; Klimstra, WB.; Pierson, TC.; Buller, RMark; Gale, M; Shi, P; Diamond, MS. 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature*. 2010. 468. 7322. 452–456.
- [120] Li, Q; Wang, L; Dong, C; Che, Y; Jiang, L; Liu, L; Zhao, H; Liao, Y; Sheng, Y; Dong, S; Ma, S. The interaction of the SARS coronavirus non-structural protein 10 with the cellular oxido-

- reductase system causes an extensive cytopathic effect. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2005. 34. 2. 133–139.
- [121] Koopmann, R; Cupelli, K; Redecke, L; Nass, K; Deponte, DP.; White, TA.; Stellato, F; Rehders, D; Liang, M; Andreasson, J; Aquila, A; Bajt, S; Barthelmess, M; Barty, A; Bogan, MJ.; Bostedt, C; Boutet, S; Bozek, JD.; Caleman, C; Coppola, N; Davidsson, J; Doak, RBruce; Ekeberg, T; Epp, SW.; Erk, B; Fleckenstein, H; Foucar, L; Graafsma, H; Gumprecht, L; Hajdu, J; Hampton, CY.; Hartmann, A; Hartmann, R; Hauser, G; Hirsemann, H; Holl, P; Hunter, MS.; Kassemeyer, S; Kirian, RA.; Lomb, L; Maia, Filipe R N C; Kimmel, N; Martin, AV.; Messerschmidt, M; Reich, C; Rolles, D; Rudek, B; Rudenko, A; Schlichting, I; Schulz, J; Seibert, MMarvin; Shoeman, RL.; Sierra, RG.; Soltau, H; Stern, S; Strüder, L; Timneanu, N; Ullrich, J; Wang, X; Weidenspointner, G; Weierstall, U; Williams, GJ.; Wunderer, CB.; Fromme, P; Spence, John C H; Stehle, T; Chapman, HN.; Betzel, C; Duszenko, M. In vivo protein crystallization opens new routes in structural biology. *Nature methods*. 2012. 9. 3. 259–262.
- [122] Redecke, L; Nass, K; Deponte, DP.; White, TA.; Rehders, D; Barty, A; Stellato, F; Liang, M; Barends, Thomas R M; Boutet, S; Williams, GJ.; Messerschmidt, M; Seibert, MMarvin; Aquila, A; Arnlund, D; Bajt, S; Barth, T; Bogan, MJ.; Caleman, C; Chao, T; Doak, RBruce; Fleckenstein, H; Frank, M; Fromme, R; Galli, L; Grotjohann, I; Hunter, MS.; Johansson, LC.; Kassemeyer, S; Katona, G; Kirian, RA.; Koopmann, R; Kupitz, C; Lomb, L; Martin, AV.; Mogk, S; Neutze, R; Shoeman, RL.; Steinbrener, J; Timneanu, N; Wang, D; Weierstall, U; Zatsepin, NA.; Spence, John C H; Fromme, P; Schlichting, I; Duszenko, M; Betzel, C; Chapman, HN. Natively inhibited Trypanosoma brucei cathepsin B structure determined by using an X-ray laser. *Science (New York, N.Y.)*. 2013. 339. 6116. 227–230.
- [123] Boutet, S; Lomb, L; Williams, GJ.; Barends, Thomas R M; Aquila, A; Doak, RBruce; Weierstall, U; Deponte, DP.; Steinbrener, J; Shoeman, RL.; Messerschmidt, M; Barty, A; White, TA.; Kassemeyer, S; Kirian, RA.; Seibert, MMarvin; Montanez, PA.; Kenney, C; Herbst, R; Hart, P; Pines, J; Haller, G; Gruner, SM.; Philipp, HT.; Tate, MW.; Hromalik, M; Koerner, LJ.; van Bakel, N; Morse, J; Ghonsalves, W; Arnlund, D; Bogan, MJ.; Caleman, C; Fromme, R; Hampton, CY.; Hunter, MS.; Johansson, LC.; Katona, G; Kupitz, C; Liang, M; Martin, AV.; Nass, K; Redecke, L; Stellato, F; Timneanu, N; Wang, D; Zatsepin, NA.; Schafer, D; Defever, J; Neutze, R; Fromme, P; Spence, John C H; Chapman, HN.; Schlichting, I. High-resolution protein structure determination by serial femtosecond crystallography. *Science (New York, N.Y.)*. 2012. 337. 6092. 362–364.
- [124] Chapman, HN.; Fromme, P; Barty, A; White, TA.; Kirian, RA.; Aquila, A; Hunter, MS.; Schulz, J; Deponte, DP.; Weierstall, U; Doak, RBruce; Maia, Filipe R N C; Martin, AV.; Schlichting, I; Lomb, L; Coppola, N; Shoeman, RL.; Epp, SW.; Hartmann, R; Rolles, D; Rudenko, A; Foucar, L; Kimmel, N; Weidenspointner, G; Holl, P; Liang, M; Barthelmess, M; Caleman, C; Boutet, S; Bogan, MJ.; Krzywinski, J; Bostedt, C; Bajt, S; Gumprecht, L; Rudek, B; Erk, B; Schmidt, C; Hömke, A; Reich, C; Pietschner, D; Strüder, L; Hauser, G; Gorke, H; Ullrich, J; Herrmann, S; Schaller, G; Schopper, F; Soltau, H; Kühnel, K; Messerschmidt, M; Bozek, JD.; Hau-Riege, SP.; Frank, M; Hampton, CY.; Sierra, RG.; Starodub, D; Williams, GJ.; Hajdu, J; Timneanu, N; Seibert, MMarvin; Andreasson, J; Rocker, A; Jönsson, O; Svenda, M; Stern, S; Nass, K; Andritschke, R; Schröter, C; Krasniqi, F; Bott, M; Schmidt, KE.; Wang, X; Grotjohann, I; Holton, JM.; Barends, Thomas R M; Neutze, R; Marchesini, S; Fromme, R; Schorb, S; Rupp, D; Adolph, M; Gorkhover, T; Andersson, I; Hirsemann, H; Potdevin, G; Graafsma, H; Nilsson, B; Spence, John C H. Femtosecond X-ray protein nanocrystallography. *Nature*. 2011. 470. 7332. 73–77.
- [125] Freer, AS.; Guarnaccio, L; Wafford, K; Smith, J; Steilberg, J; Culver, JN.; Harris, MT. SAXS characterization of genetically engineered tobacco mosaic virus nanorods coated with palladium in the absence of external reducing agents. *Journal of colloid and interface science*. 2013. 392. 213–218.
- [126] Arias-Palomo, E; O'Shea, VL.; Hood, IV.; Berger, JM. The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell*. 2013. 153. 2. 438–448.

- [127] Sagane, Y; Hayashi, S; Matsumoto, T; Miyashita, S; Inui, K; Miyata, K; Yajima, S; Suzuki, T; Hasegawa, K; Yamano, A; Nishikawa, A; Ohyama, T; Watanabe, T; Niwa, K. Sugar-induced conformational change found in the HA-33/HA-17 trimer of the botulinum toxin complex. *Biochemical and biophysical research communications*. 2013. 438. 3. 483–487.
- [128] Siuzdak, G; Bothner, B; Yeager, M; Brugidou, C; Fauquet, CM.; Hoey, K; Chang, CM. Mass spectrometry and viral analysis. *Chemistry & biology*. 1996. 3. 1. 45–48.
- [129] Uetrecht, C; Heck, Albert J R. Modern biomolecular mass spectrometry and its role in studying virus structure, dynamics, and assembly. *Angewandte Chemie (International ed. in English)*. 2011. 50. 36. 8248–8262.
- [130] Zhou, M; Jones, CM.; Wysocki, VH. Dissecting the Large Noncovalent Protein Complex GroEL with Surface-Induced Dissociation and Ion Mobility-Mass Spectrometry. *Analytical chemistry*. 2013. 85. 17. 8262–8267.
- [131] Su, D; Lou, Z; Sun, F; Zhai, Y; Yang, H; Zhang, R; Joachimiak, A; Zhang, XC.; Bartlam, M; Rao, Z. Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10. *Journal of virology*. 2006. 80. 16. 7902–7908.
- [132] Wu, CGuo; Cheng, SChun; Chen, SChuan; Li, JYan; Fang, YHsuan; Chen, YHung; Chou, CYuan. Mechanism for controlling the monomer-dimer conversion of SARS coronavirus main protease. *Acta crystallographica. Section D, Biological crystallography*. 2013. 69. Pt 5. 747–755.
- [133] Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of bacteriology*. 1951. 62. 3. 293–300.
- [134] Chevallet, M; Luche, S; Rabilloud, T. Silver staining of proteins in polyacrylamide gels. *Nature protocols*. 2006. 1. 4. 1852–1858.
- [135] Adkins, S; Burmeister, M. Visualization of DNA in agarose gels as migrating colored bands: applications for preparative gels and educational demonstrations. *Analytical biochemistry*. 1996. 240. 1. 17–23.
- [136] Mullis, KB. The unusual origin of the polymerase chain reaction. *Scientific American*. 1990. 262. 4. 56-61, 64-5.
- [137] Sanger, F; Coulson, AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*. 1975. 94. 3. 441–448.
- [138] Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology*. 1983. 166. 4. 557–580.
- [139] Inoue, H; Nojima, H; Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene*. 1990. 96. 1. 23–28.
- [140] Clarke, PR.; Hill, CR. Physical and chemical aspects of ultrasonic disruption of cells. *The Journal of the Acoustical Society of America*. 1970. 47. 2. 649–653.
- [141] Green, Sambrook 2012 – Molecular cloning, 4th edition, Cold Spring Harbor Laboratory Press, ISBN: 1936113422
- [142] Gasteiger, E; Gattiker, A; Hoogland, C; Ivanyi, I; Appel, RD.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*. 2003. 31. 13. 3784–3788.
- [143] Abrahams, JP.; Kraal, B; Bosch, L. Zone-interference gel electrophoresis: a new method for studying weak protein-nucleic acid complexes under native equilibrium conditions. *Nucleic acids research*. 1988. 16. 21. 10099–10108.
- [144] Renart, J; Reiser, J; Stark, GR. Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure. *Proceedings of the National Academy of Sciences of the United States of America*. 1979. 76. 7. 3116–3120.

- [145] Kelly, SM.; Jess, TJ.; Price, NC. How to study proteins by circular dichroism. *Biochimica et biophysica acta*. 2005. 1751. 2. 119–139.
- [146] Yang, JT.; Wu, CS.; Martinez, HM. Calculation of protein conformation from circular dichroism. *Methods in enzymology*. 1986. 130. 208–269.
- [147] Konarev, PV.; Volkov, VV.; Sokolova, AV.; Koch, MH. J.; Svergun, DI. PRIMUS. *Journal of Applied Crystallography*. 2003. 36. 5. 1277–1282.
- [148] Guinier, A. La Diffraction des rayons X aux très petits angles - application à l'étude de phénomènes ultramicroscopiques.... 1939
- [149] Svergun, DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*. 1992. 25. 4. 495–503.
- [150] Franke, D; Svergun, DI. DAMMIF , a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*. 2009. 42. 2. 342–346.
- [151] Volkov, VV.; Svergun, DI. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography*. 2003. 36. 3. 860–864.
- [152] Svergun, D; Barberato, C; Koch, MH. J. CRY SOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*. 1995. 28. 6. 768–773.
- [153] Petoukhov, MV.; Svergun, DI. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical journal*. 2005. 89. 2. 1237–1250.
- [154] Svergun, DI.; Petoukhov, MV.; Koch, MH. Determination of domain structure of proteins from X-ray solution scattering. *Biophysical journal*. 2001. 80. 6. 2946–2953.
- [155] Powell, HR.; Johnson, O; Leslie, Andrew G W. Autoindexing diffraction images with iMosflm. *Acta crystallographica. Section D, Biological crystallography*. 2013. 69. Pt 7. 1195–1203.
- [156] Kabsch, W. Automatic indexing of rotation diffraction patterns. *Journal of Applied Crystallography*. 1988. 21. 1. 67–72.
- [157] McCoy, AJ.; Grosse-Kunstleve, RW.; Adams, PD.; Winn, MD.; Storoni, LC.; Read, RJ. Phaser crystallographic software. *Journal of Applied Crystallography*. 2007. 40. Pt 4. 658–674.
- [158] Murshudov, GN.; Grebenko, AI.; Barynin, V; Dauter, Z; Wilson, KS.; Vainshtein, BK.; Melik-Adamyanyan, W; Bravo, J; Ferrán, JM.; Ferrer, JC.; Switala, J; Loewen, PC.; Fita, I. Structure of the heme d of *Penicillium vitale* and *Escherichia coli* catalases. *The Journal of biological chemistry*. 1996. 271. 15. 8863–8868.
- [159] Emsley, P; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography*. 2004. 60. Pt 12 Pt 1. 2126–2132.
- [160] Emsley, P; Lohkamp, B; Scott, WG.; Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography*. 2010. 66. Pt 4. 486–501.
- [161] Rinner, O; Seebacher, J; Walzthoeni, T; Mueller, LN.; Beck, M; Schmidt, A; Mueller, M; Aebersold, R. Identification of cross-linked peptides from large sequence databases. *Nature methods*. 2008. 5. 4. 315–318.
- [162] Walzthoeni, T; Claassen, M; Leitner, A; Herzog, F; Bohn, S; Forster, F; Beck, M; Aebersold, R. False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nature methods*. 2012. 9. 9. 901–903.
- [163] van den Heuvel, Robert H H; van Duijn, E; Mazon, H; Synowsky, SA.; Lorenzen, K; Versluis, C; Brouns, Stan J J; Langridge, D; van der Oost, John; Hoyes, J; Heck, Albert J R. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Analytical chemistry*. 2006. 78. 21. 7473–7483.

- [164] Synowsky, SA.; Heck, AJ.R. The yeast Ski complex is a hetero-tetramer. *Protein Science*. 2007. 17. 1. 119–125.
- [165] Morgner, N; Robinson, CV. Massign: an assignment strategy for maximizing information from the mass spectra of heterogeneous protein assemblies. *Analytical chemistry*. 2012. 84. 6. 2939–2948.
- [166] Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*. 2008. 9. 40.
- [167] Brenke, R; Kozakov, D; Chuang, G; Beglov, D; Hall, D; Landon, MR.; Mattos, C; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics*. 2009. 25. 5. 621–627.
- [168] Jo, S; Vargyas, M; Vasko-Szedlar, J; Roux, B; Im, W. PBEQ-Solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Research*. 2008. 36. Web Server. W270.
- [169] Hernandez, M; Ghersi, D; Sanchez, R. SITEHOUND-web: a server for ligand binding site identification in protein structures. *Nucleic Acids Research*. 2009. 37. Web Server. W413.
- [170] Matthes, N; Mesters, JR.; Coutard, B; Canard, B; Snijder, EJ.; Moll, R; Hilgenfeld, R. The non-structural protein Nsp10 of mouse hepatitis virus binds zinc ions and nucleic acids. *FEBS letters*. 2006. 580. 17. 4143–4149.
- [171] Schagger, H; Jagow, G von. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Analytical biochemistry*. 1991. 199. 2. 223–231.
- [172] Raghunathan, S; Kozlov, AG.; Lohman, TM.; Waksman, G. Structure of the DNA binding domain of *E. coli* SSB bound to ssDNA. *Nature structural biology*. 2000. 7. 8. 648–652.
- [173] Krishna, TS.; Kong, XP.; Gary, S; Burgers, PM.; Kuriyan, J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell*. 1994. 79. 7. 1233–1243.
- [174] Bayfield, OW.; Chen, C; Patterson, AR.; Luan, W; Smits, C; Gollnick, P; Antson, AA. Trp RNA-binding attenuation protein: modifying symmetry and stability of a circular oligomer. *PloS one*. 2012. 7. 9. e44309.
- [175] Majumdar, A; Cesario, WColón; White-Grindley, E; Jiang, H; Ren, F; Khan, MRepon; Li, L; Choi, EMan-Lik; Kannan, K; Guo, F; Unruh, J; Slaughter, B; Si, K. Critical role of amyloid-like oligomers of *Drosophila* Orb2 in the persistence of memory. *Cell*. 2012. 148. 3. 515–529.
- [176] Hobson, SD.; Rosenblum, ES.; Richards, OC.; Richmond, K; Kirkegaard, K; Schultz, SC. Oligomeric structures of poliovirus polymerase are important for function. *The EMBO journal*. 2001. 20. 5. 1153–1163.
- [177] Parks, D; Bolinger, R; Mann, K. Redox state regulates binding of p53 to sequence-specific DNA, but not to non-specific or mismatched DNA. *Nucleic acids research*. 1997. 25. 6. 1289–1295.
- [178] Bauer, CE.; Elsen, S; Bird, TH. Mechanisms for redox control of gene expression. *Annual review of microbiology*. 1999. 53. 495–523.
- [179] Kozin, MB.; Svergun, DI. Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography*. 2001. 34. 1. 33–41.
- [180] Rambo, RP.; Tainer, JA. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*. 2013. 496. 7446. 477–481.
- [181] 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. Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS pathogens*. 2011. 7. 10. e1002294.

- [182] Muller, DR.; Schindler, P; Towbin, H; Wirth, U; Voshol, H; Hoving, S; Steinmetz, MO. Isotope-tagged cross-linking reagents. A new tool in mass spectrometric protein interaction analysis. *Analytical chemistry*. 2001. 73. 9. 1927–1934.
- [183] Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of structural biology*. 2011. 173. 3. 530–540.
- [184] Venselaar, H; Joosten, RP.; Vroling, B; Baakman, Coos A B; Hekkelman, ML.; Krieger, E; Vriend, G. Homology modelling and spectroscopy, a never-ending love story. *European biophysics journal : EBJ*. 2010. 39. 4. 551–563.
- [185] Fu, C; Uetrecht, C; Kang, S; Morais, MC.; Heck, A. J. R.; Walter, MR.; Prevelige, PE. A Docking Model Based on Mass Spectrometric and Biochemical Data Describes Phage Packaging Motor Incorporation. *Molecular & Cellular Proteomics*. 2010. 9. 8. 1764–1773.
- [186] Eyal, E; Gerzon, S; Potapov, V; Edelman, M; Sobolev, V. The limit of accuracy of protein modeling: influence of crystal packing on protein structure. *Journal of molecular biology*. 2005. 351. 2. 431–442.
- [187] Egloff, M; Ferron, F; Campanacci, V; Longhi, S; Rancurel, C; Dutartre, H; Snijder, EJ.; Gorbalenya, AE.; Cambillau, C; Canard, B. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proceedings of the National Academy of Sciences of the United States of America*. 2004. 101. 11. 3792–3796.
- [188] Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*. 2003. 31. 13. 3406–3415.
- [189] Zhang, S; Zhong, N; Xue, F; Kang, X; Ren, X; Chen, J; Jin, C; Lou, Z; Xia, B. Three-dimensional domain swapping as a mechanism to lock the active conformation in a super-active octamer of SARS-CoV main protease. *Protein & cell*. 2010. 1. 4. 371–383.
- [190] Stobart, CC.; Sexton, NR.; Munjal, H; Lu, X; Molland, KL.; Tomar, S; Mesecar, AD.; Denison, MR. Chimeric exchange of coronavirus nsp5 proteases (3CL^{pro}) identifies common and divergent regulatory determinants of protease activity. *Journal of Virology*. 2013.
- [191] Chuck, C; Chong, L; Chen, C; Chow, H; Wan, DChi-Cheong; Wong, K. Profiling of substrate specificity of SARS-CoV 3CL. *PloS one*. 2010. 5. 10. e13197.
- [192] Fan, K; Ma, L; Han, X; Liang, H; Wei, P; Liu, Y; Lai, L. The substrate specificity of SARS coronavirus 3C-like proteinase. *Biochemical and Biophysical Research Communications*. 2005. 329. 3. 934–940.
- [193] Grum-Tokars, V; Ratia, K; Begaye, A; Baker, SC.; Mesecar, AD. Evaluating the 3C-like protease activity of SARS-Coronavirus: recommendations for standardized assays for drug discovery. *Virus research*. 2008. 133. 1. 63–73.
- [194] Deming, DJ.; Graham, RL.; Denison, MR.; Baric, RS. MHV-A59 ORF1a replicase protein nsp7-nsp10 processing in replication. *Advances in experimental medicine and biology*. 2006. 581. 101–104.
- [195] Ziebuhr, J; Herold, J; Siddell, SG. Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *Journal of virology*. 1995. 69. 7. 4331–4338.
- [196] Bost, AG.; Carnahan, RH.; Lu, XT.; Denison, MR. Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. *Journal of virology*. 2000. 74. 7. 3379–3387.
- [197] Quillin, ML.; Matthews, BW. Accurate calculation of the density of proteins. *Acta crystallographica. Section D, Biological crystallography*. 2000. 56. Pt 7. 791–794.

- [198] Chou, C; Chang, H; Hsu, W; Lin, T; Lin, C; Chang, G. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry*. 2004. 43. 47. 14958–14970.
- [199] Fan, K; Wei, P; Feng, Q; Chen, S; Huang, C; Ma, L; Lai, B; Pei, J; Liu, Y; Chen, J; Lai, L. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *The Journal of biological chemistry*. 2004. 279. 3. 1637–1642.
- [200] Li, C; Qi, Y; Teng, X; Yang, Z; Wei, P; Zhang, C; Tan, L; Zhou, L; Liu, Y; Lai, L. Maturation mechanism of severe acute respiratory syndrome (SARS) coronavirus 3C-like proteinase. *The Journal of biological chemistry*. 2010. 285. 36. 28134–28140.
- [201] Ryu, YBae; Jeong, HJae; Kim, JHoon; Kim, YMin; Park, J; Kim, D; Nguyen, Thi Thanh Hanh; Park, S; Chang, JSun; Park, KHun; Rho, M; Lee, WSong. Biflavonoids from *Torreya nucifera* displaying SARS-CoV 3CL(pro) inhibition. *Bioorganic & medicinal chemistry*. 2010. 18. 22. 7940–7947.
- [202] Chen, L; Li, J; Luo, C; Liu, H; Xu, W; Chen, G; Liew, OWah; Zhu, W; Puah, CMok; Shen, X; Jiang, H. Binding interaction of quercetin-3- β -galactoside and its synthetic derivatives with SARS-CoV 3CLpro: Structure–activity relationship studies reveal salient pharmacophore features. *Bioorganic & Medicinal Chemistry*. 2006. 14. 24. 8295–8306.
- [203] Kim, DWook; Seo, KHye; Curtis-Long, MJ.; Oh, KYeol; Oh, J; Cho, JKeun; Lee, KHoo; Park, KHun. Phenolic phytochemical displaying SARS-CoV papain-like protease inhibition from the seeds of *Psoralea corylifolia*. *Journal of enzyme inhibition and medicinal chemistry*. 2013.
- [204] Lin, C; Tsai, F; Tsai, C; Lai, C; Wan, L; Ho, T; Hsieh, C; Chao, PLee. Anti-SARS coronavirus 3C-like protease effects of *Isatis indigotica* root and plant-derived phenolic compounds. *Antiviral research*. 2005. 68. 1. 36–42.
- [205] Senes, A; Gerstein, M; Engelman, DM. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *Journal of molecular biology*. 2000. 296. 3. 921–936.
- [206] Mao, G; Tan, J; Cui, M; Chui, D; Xu, X. The GxxxG motif in the transmembrane domain of AbetaPP plays an essential role in the interaction of CTF beta with the gamma-secretase complex and the formation of amyloid-beta. *Journal of Alzheimer's disease : JAD*. 2009. 18. 1. 167–176.
- [207] Arbely, E; Granot, Z; Kass, I; Orly, J; Arkin, IT. A trimerizing GxxxG motif is uniquely inserted in the severe acute respiratory syndrome (SARS) coronavirus spike protein transmembrane domain. *Biochemistry*. 2006. 45. 38. 11349–11356.
- [208] Chen, B; Fang, S; Tam, JP.; Liu, DXiang. 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*. 2009. 383. 2. 328–337.
- [209] Vermeer, AW.; Norde, W. The thermal stability of immunoglobulin: unfolding and aggregation of a multi-domain protein. *Biophysical journal*. 2000. 78. 1. 394–404.
- [210] Greenfield, NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature protocols*. 2006. 1. 6. 2527–2535.
- [211] Perczel, A; Park, K; Fasman, GD. Deconvolution of the circular dichroism spectra of proteins: the circular dichroism spectra of the antiparallel beta-sheet in proteins. *Proteins*. 1992. 13. 1. 57–69.
- [212] Bressanelli, S; Tomei, L; Rey, FA.; Francesco, R de. Structural analysis of the hepatitis C virus RNA polymerase in complex with ribonucleotides. *Journal of virology*. 2002. 76. 7. 3482–3492.

- [213] Butcher, SJ.; Grimes, JM.; Makeyev, EV.; Bamford, DH.; Stuart, DI. A mechanism for initiating RNA-dependent RNA polymerization. *Nature*. 2001. 410. 6825. 235–240.
- [214] Shi, J; Wei, Z; Song, J. Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors. *The Journal of biological chemistry*. 2004. 279. 23. 24765–24773.
- [215] Shi, J; Song, J. The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *The FEBS journal*. 2006. 273. 5. 1035–1045.
- [216] Huang, Q; Yu, L; Petros, AM.; Gunasekera, A; Liu, Z; Xu, N; Hajduk, P; Mack, J; Fesik, SW.; Olejniczak, ET. Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry*. 2004. 43. 20. 6059–6063.
- [217] Petrov, AS.; Bernier, CR.; Hsiao, C; Okafor, CDenise; Tannenbaum, E; Stern, J; Gaucher, E; Schneider, D; Hud, NV.; Harvey, SC.; Dean Williams, L. RNA–Magnesium–Protein Interactions in Large Ribosomal Subunit. *The Journal of Physical Chemistry B*. 2012. 116. 28. 8113–8120.
- [218] Georgieva, D; Koker, M; Redecke, L; Perbandt, M; Clos, J; Bredehorst, R; Genov, N; Betzel, C. Oligomerization of the proteolytic products is an intrinsic property of prion proteins. *Biochemical and biophysical research communications*. 2004. 323. 4. 1278–1286.
- [219] Lugari, A; Betzi, S; Decroly, E; Bonnaud, E; Hermant, A; Guillemot, J; Debarnot, C; Borg, J; Bouvet, M; Canard, B; Morelli, X; Lécine, P. Molecular mapping of the RNA Cap 2'-O-methyltransferase activation interface between severe acute respiratory syndrome coronavirus nsp10 and nsp16. *The Journal of biological chemistry*. 2010. 285. 43. 33230–33241.
- [220] Decroly, E; Imbert, I; Coutard, B; Bouvet, M; Selisko, B; Alvarez, K; Gorbalenya, AE.; Snijder, EJ.; Canard, B. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. *Journal of virology*. 2008. 82. 16. 8071–8084.
- [221] Wei, P; Fan, K; Chen, H; Ma, L; Huang, C; Tan, L; Xi, D; Li, C; Liu, Y; Cao, A; Lai, L. The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochemical and biophysical research communications*. 2006. 339. 3. 865–872.

Recent review on SARS-CoV/MERS-CoV emergence and antiviral research:

- [222] Hilgenfeld, R; Peiris, M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Research*. 2013. 100. 1. 286–295.

8. Appendix

Table 30. Summary of optimized parameters for gene over-expression in *E. coli*

protein	vector	strain*	inductor	time/temperature
pp710 HCoV-229E	pET11d	BL21 (DE3) star	0.8 mM IPTG	4 h / 37 °C
pp910 HCoV-229E	pRSETA	BL21 AI	0.2% L-arabinose	6 h / 25 °C
pp78 HCoV-229E	pGEX-6p-1	BL21 AI (BL 21 (DE3) star)	0.2% L-arabinose (1 mM IPTG)	6 h / 25 °C (overnight / 22 °C)
pp89 HCoV-229E	pRSETA	BL21 AI (BL21 (DE3) star)	0.2% L-arabinose (1 mM IPTG)	4 h / 30 °C
pp78 HCoV-229E	pRSETA	BL21 AI (BL21 (DE3) star)	0.2% L-arabinose (1 mM IPTG)	6 h / 25-30 °C
Nsp7 HCoV-229E	pRSETA	BL21 AI (BL21 (DE3) star)	1 mM IPTG	5 h / 25 °C
Nsp8 HCoV-229E	pRSETA	BL21 (DE3) star	1 mM IPTG	4 h / 25 °C
Nsp9 HCoV-229E	pRSETA	BL 21 (DE3) star (Origami)	1 mM IPTG	4 h / 37 °C
Nsp8 HCoV-229E (D104-Q195), C-term.	pASK 16	BL21 (DE3) star	0.2 mg L ⁻¹ AHT	5 h / 25 °C
pp710 SARS-CoV	pASK 3+/5+/33+/55+	BL21 (DE3) star	0.2 mg L ⁻¹ AHT	3.5 h / 37 °C
pp79 SARS-CoV	pASK 3+/5+/33+/55+	BL21 (DE3) star	0.2 mg L ⁻¹ AHT	3.5 h / 37 °C
pp78 SARS-CoV	pASK 3+/5+/33+/55+	BL21 (DE3) star	0.2 mg L ⁻¹ AHT	3.5 h / 37 °C
pp89 SARS-CoV	pASK 3+/5+/33+/55+	BL21 (DE3) star	0.2 mg L ⁻¹ AHT	3.5 h / 37 °C

* strains in brackets resulted in a similar protein yield

Amino acid sequence of recombinantly produced SARS-CoV main protease (N-term → C-term)

SGFRKMAFSPGKVEGCMVQVTCGTTTLNGLWDDTVYCPRHVICTAEDMLNPNYEDLLIRKSNHSFLV
 QAGNVQLRVIGHSMQNCLLRKVDTSNPKTPKYKFVRIQPGQTFVSLACYNGSPSGVYQCAMRPNHTI
 KGSFLNGSCGSVGFNIDYDCVSFCYMHMELPTGVHAGTDLEGKFGPFVDRQTAQAAGTDTITLNV
 LAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMCAALKEL
 LQNGMNGRTILGSTILEDEFTPFVDVVRQCSGVTFQGPHHHHHH

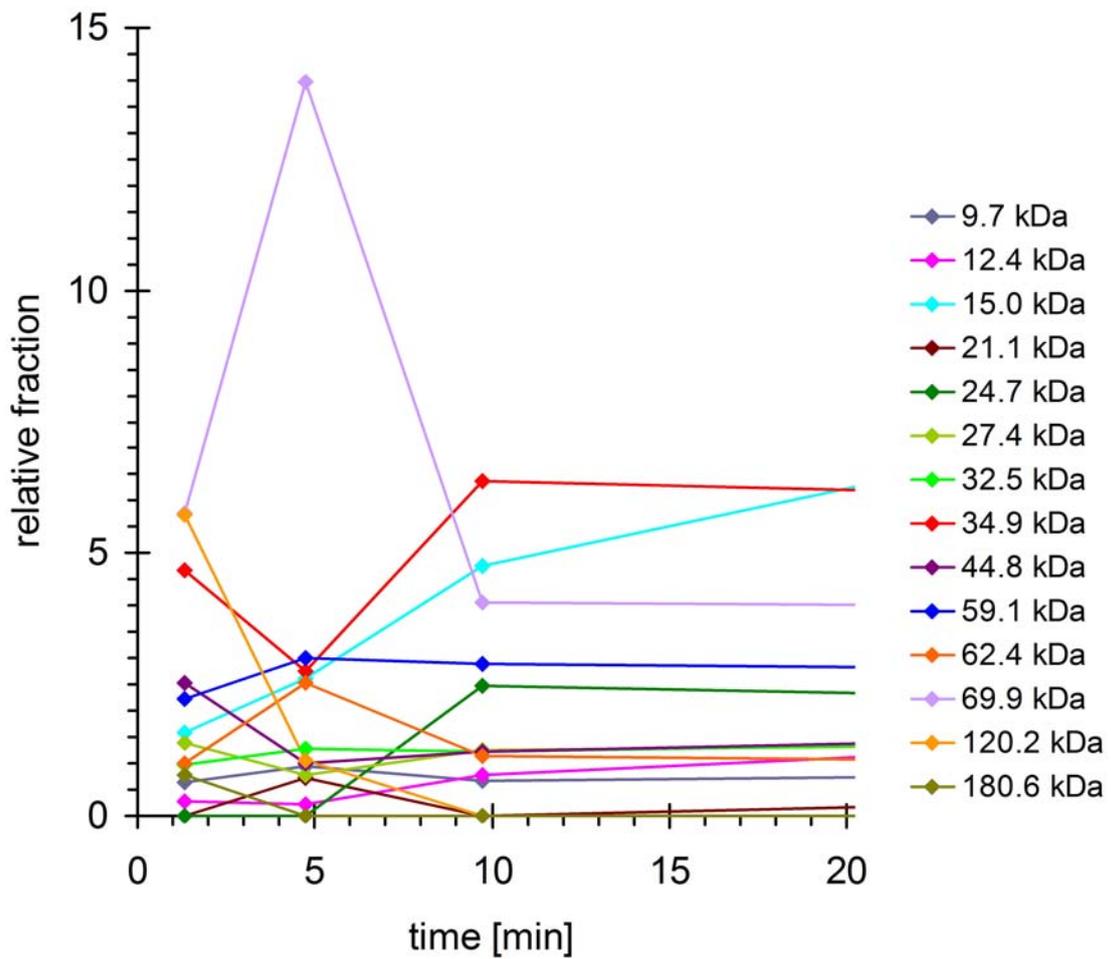


Figure 86. Semi-quantitative analysis of the relative content of mass species during processing of SARS-CoV pp710 by SARS-CoV M^{pro}. Polyprotein and main protease were incubated at a molar ratio of 10:1. Data was obtained by ESI-TOF MS, the relative fraction refers to the percentage of the total intensity at the TOF detector (see chapter 3.4.4).

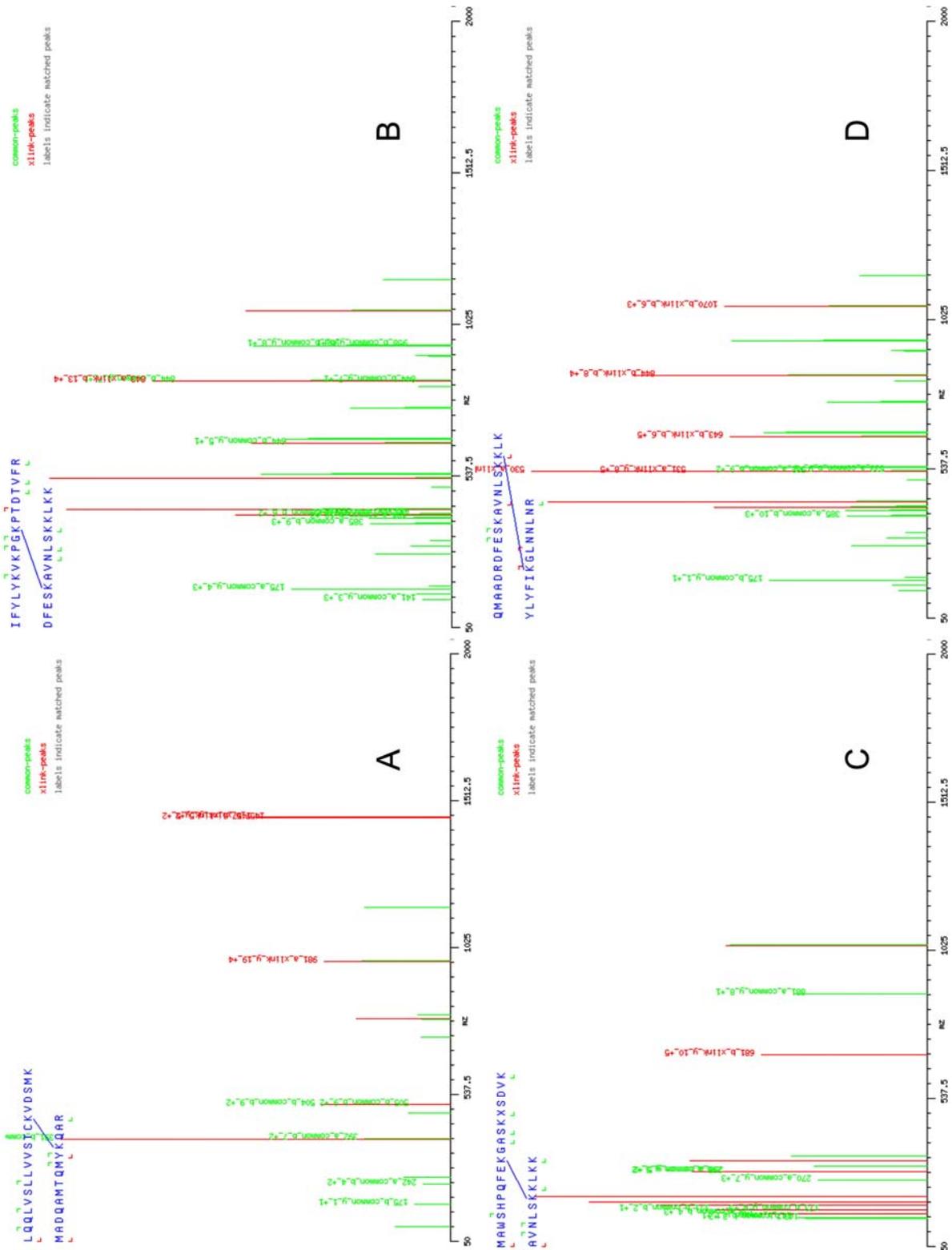


Figure 87A. MS/MS spectra of putative BS2G cross-linking sites within SARS-CoV pp710. All the depicted MS/MS spectra were obtained from trypsin digested monomeric pp710 cut from an SDS-PAGE gel. In spectra **A**, **B** and **D** the detected masses refer to a cross-link with the heavy deuterated isotope BS2G-d₄. Spectrum **C** shows a cross-link with the non-deuterated BS2G-d₀. Ion masses are labelled either *green* (common mass) or *red* (mass containing the cross-linker). The detected masses are also labelled in the peptide sequence, either red or green to see the sequence coverage. All labelled peaks match an expected fragment mass.

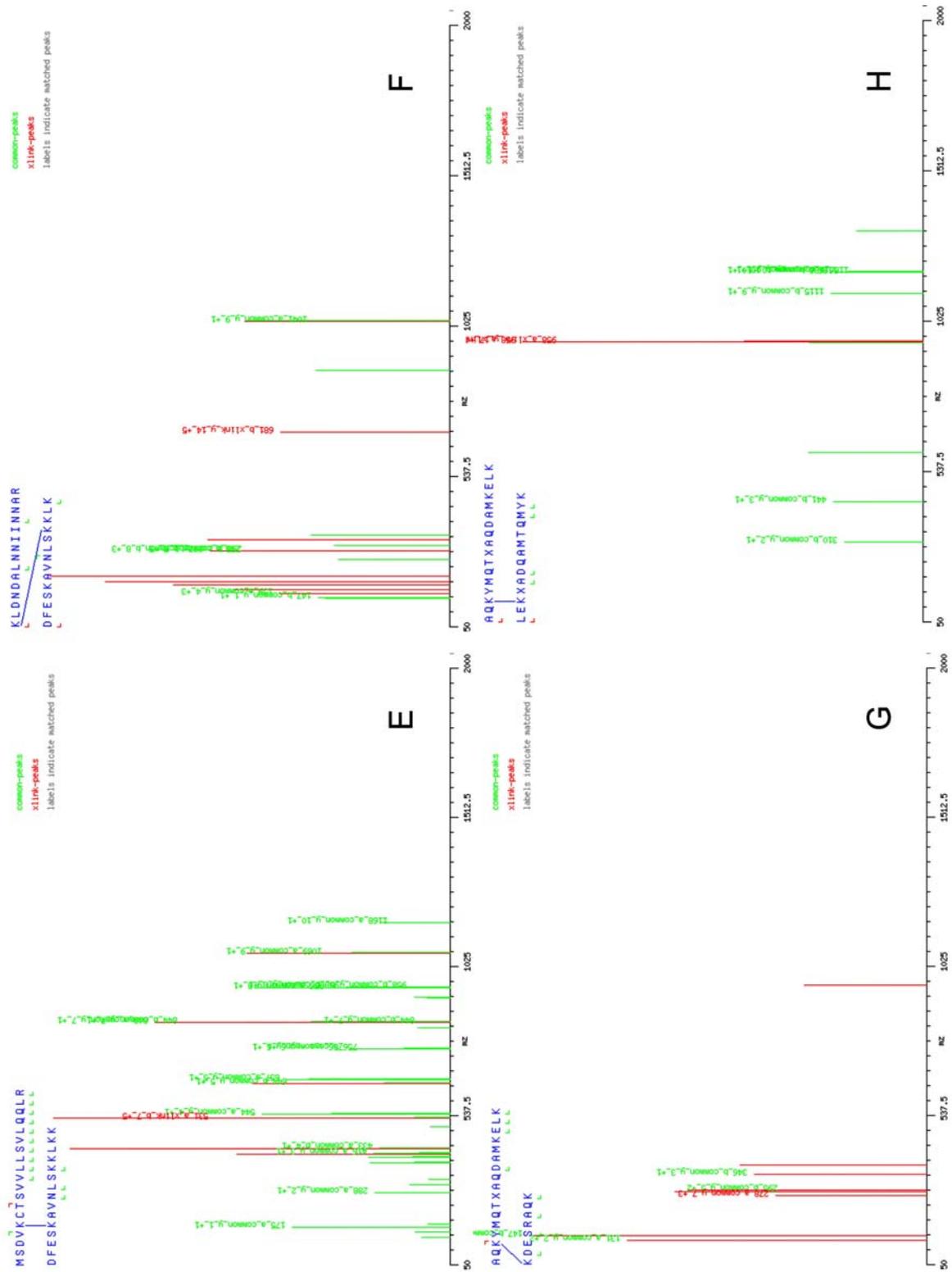


Figure 87B. MS/MS spectra of putative BS2G cross-linking sites within SARS-CoV pp710. Spectrum **E** indicates cross-linking of monomeric pp710 with BS2G-d₀ cut from an SDS-PAGE gel. Spectrum **F** shows a cross-link site within trimeric pp710 with BS2G-d₄. Spectrum **G** represents a putative cross-link with BS2G-d₄ in a sample of larger oligomers. Spectrum **H** shows a cross-linked peptide containing the light isotope BS2G-d₀ in an oligomer fraction.

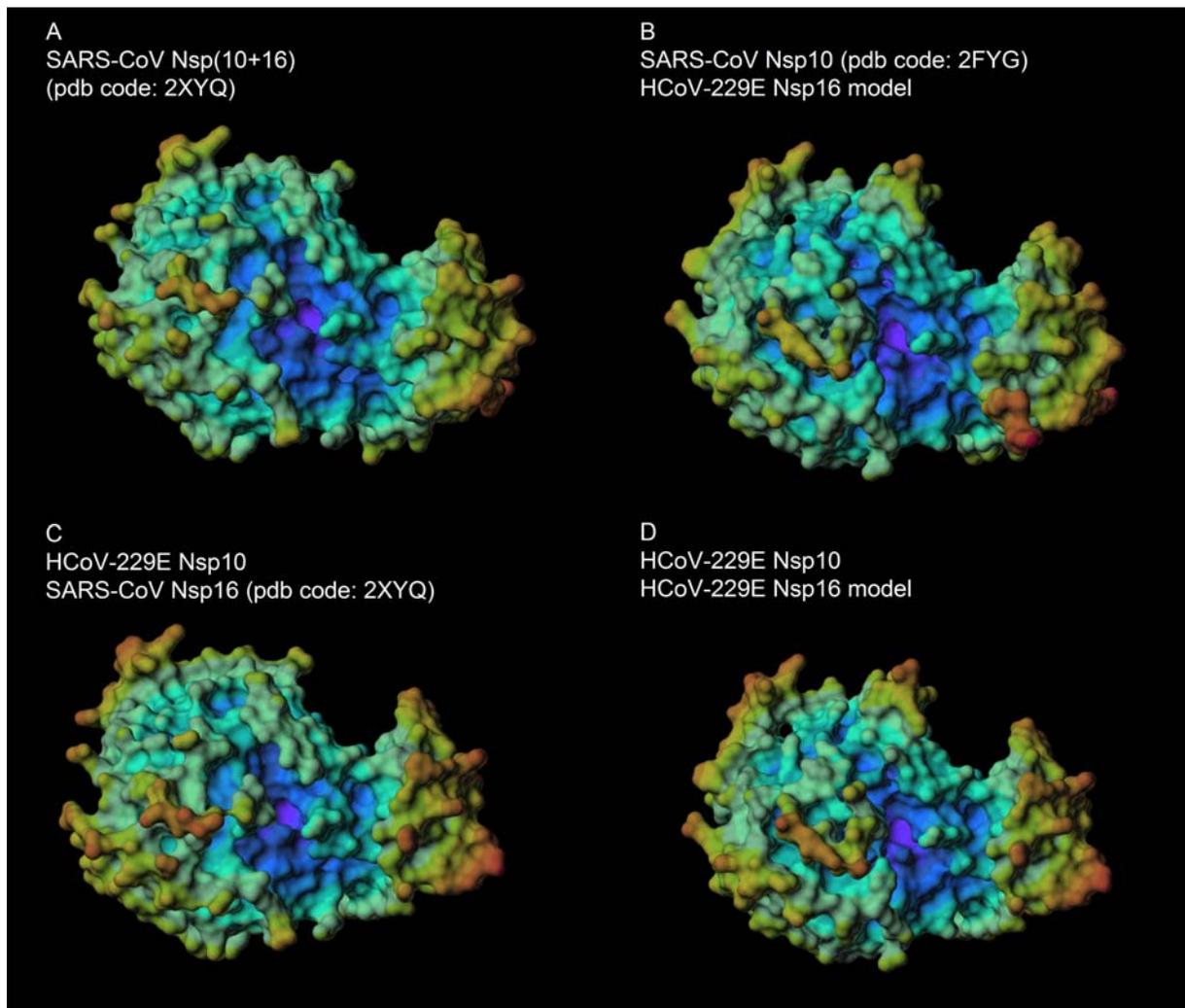


Figure 89. Electrostatic surface potential calculated by Sybyl-X 1.2 comparing the interaction of two different homologues of Nsp10 and Nsp16, particularly the resulting conserved patch of positive charge (*blue/purple*).

9. Risk and safety statements

9.1. Chemicals (GHS classification)

Table 31. Chemicals (GHS classification)

compound	CAS-No.	supplier	GHS hazard	hazard statements	precautionary statements
acetic acid	64-19-7	Chem-solute	GHS02 GHS05	H226, H314	P280, P305+351+338, P310
acrylamide 30%	79-06-1	Carl Roth	GHS06 GHS08	H301, H312, H316, H317, H319, H332, H340, H350, H361f, H372	P201, P280, P301+310, P305+351+338, P308+313
ATP	34369-07-8	Sigma	-	-	-
agarose	9012-36-6	Serva	-	-	-
ammonium acetate	631-61-8	Sigma	-	-	-
(NH ₄) ₂ SO ₄	7283-20-2	Carl Roth	-	-	-
ampicillin	69-52-3	Carl Roth	GHS08	H334, H317	P280, P261, P302+P352, P342+P311
APS	7727-54-0	Carl Roth	GHS03 GHS07 GHS08	H272, H302, H315, H317, H319, H334; H335	P280, P305+351+338, P302+352, P304+341, P342+311
bis-Tris	6976-37-0	Sigma	GHS07	H315, H319, H335	P261, P305 + P351 + P338
bromophenol blue	115-39-9	Applichem	-	-	-
boric acid	10043-35-3	Carl Roth		H360	P201, P202, P281, P303+313, P405, P501
CaCl ₂	10043-52-4	Merck	GHS07	H319	P305+351+338
citric acid	77-92-9	Sigma	GHS05	H318	P305+351+338, P311
coomassie brilliant blue R250	6104-59-2	Serva	-	-	-
EGCG	989-51-5	Sigma	-	-	-

compound	CAS-No.	supplier	GHS hazard	hazard statements	precautionary statements
DMSO	67-68-5	Sigma	GHS08	H315, H319	P280, P302+352, P305 +351+338, P332+313, P362
DTT	578517	Applichem	GHS07	H302, H315, H319, H335	P302+352, P305+351+338
EDC	25952-53-8	Applichem	GHS05, GHS07	H315, H318, H335	P261, P280, P305 + P351 + P338
EDTA	60-00-4	Sigma	GHS07	H319	P305+351+338
ethanol	64-17-5	Carl Roth	GHS02	H225	P210
ethidium bromide	1239-45-8	Sigma	GHS06, GHS08	H302, H330, H341	P260, P281, P284, P310
glutaraldehyde	111-30-8	Applichem	GHS05, GHS07, GHS08, GHS09	H302 + H332, H314, H317, H334, H335, H400	P261, P273, P280, P305 + P351 + P338, P310
glycerol	56-81-5	Sigma	-	-	-
GSH	70-18-8	Carl Roth	-	-	-
guanidin hydrochlorid	50-01-1	Applichem	GHS07	H302, H315, H319	P305+P351+P388, P302+P352
HEPES	7365-45-9	Sigma	-	-	-
hydrochloric acid >25%	7647-01-0	Merck	GHS05, GHS07	H314, H335	P261, P280, P310, P305+351+338
imidazole	288-32-4	Carl Roth	GHS05, GHS06, GHS08	H301; H314; H361	P260, P281, P303+P361+P353, P301+P330+P331, P305+P351+P338, P308+P313
IPTG	367-93-1	Carl Roth	-	-	-
isopropanol	67-63-0	Carl Roth	GHS02, GHS07	H225, H319, H336.	P210, P233, P305+351+338
KCl	7447-40-7	Carl Roth	-	-	-
K₂HPO₄	7758-11-4	Carl Roth	-	-	-
KH₂PO₄	7778-77-0	Merck	-	-	-

compound	CAS-No.	supplier	GHS hazard	hazard statements	precautionary statements
Li₂SO₄	10102-25-7	Merck	GHS07	H302	-
MgCl₂	7786-30-3	Carl Roth	-	-	-
MgSO₄	7487-88-9	Merck	-	-	-
MPD	107-41-5	Carl Roth	GHS07	H315, H319	-
2-mercapto-ethanol	60-24-2	Fisher Scientific	GHS06, GHS09	H302, H411, H315, H335, H311, H319	P280, P312, P302+P350, P261, P273, P301+P312, P305+351+338
NaOAc	127-09-3	Applichem	-	-	-
Na₂HPO₄	7558-79-4	Carl Roth	-	-	-
NaH₂PO₄	10049-21-5	Applichem	-	-	-
NaF	7681-49-4	Merck	-	H301, H315, H319	P301 + P310, P305 +P351 + P338
NaCl	7647-14-5	Carl Roth	-	-	-
NaH₂PO₄	10049-21-5	Applichem	-	-	-
NaOH	1310-73-2	Merck	GHS05	H314	P280, P310, P305+351+338
Na₃-citrate	6132-04-3	Sigma	-	-	-
NiSO₄	10101-97-0	Applichem	GHS08, GHS09, GHS07	H332, H315, H334, H317, H341, H350i, H360D, H372 H410	P280, P273, P201, P342+P311, P308+P313, P302+P352
paraffin	8002-74-2	Applichem	-	-	-
PEG 400	25322-68-3	Sigma	-	-	-
PEG 6000	25322-68-3	Merck	-	-	-
PEG 8000	25322-68-3	Sigma	-	-	-
PMSF	329-98-6	Applichem	GHS06, GHS05	H301, H314	P280, P305+P351+P338, P310
SDS	151-21-3	Sigma	GHS02 GHS06	H228, H302, H311, H315, H319, H335	P210, P261, P280, P312, P305+351+338
silver nitrate	7761-88-8	Carl Roth	GHS09	H272, H314, H410	P273, P280, P301+310+331, P305+351+338, P309+310

compound	CAS-No.	supplier	GHS hazard	hazard statements	precautionary statements
tannic acid	1401-55-4	Sigma	-	-	-
TCA	76-03-9	Sigma	GHS05 GHS09	H314, H410	P273, P280, P301+330+331, P305+351+338, P309+P310
TEMED	110-18-9	Merck	GHS02 GHS05 GHS07	H225, H302, H314, H332	P261, P280, P305+351+338
Tris	1185-53-1	Fluka	GHS07	H315, H319, H335	P261, P305+351+338
yeast extract	8013-01-2	Serva	-	-	-
ZnCl₂	7646-85-7	Sigma	GHS05 , GHS07 , GHS09	H302-H314-H410	P273, P280, P305 + P351 + P338, P310, P501
β-OG	29836-26-8	Carl Roth	-	-	-

9.2. Commercial crystallization solutions

Table 32. Commercial crystallization solutions and kits

name	supplier	risk label	risk phrases	safety phrases
Floppy Choppy	Jena Bio Science	C, Xn, Xi	R35, R41, R42, R36/37/38	S22, S26, S45, S24/25, S36/37/39:
PCT	Hampton	-	-	-
Classic Suite	Qiagen	T, N	R10, R45, R46, R60, R61, R23/25, R36/37/38, R48/20/22, R51/53	S20, S26, S45, S53, S36/37/39.
PACT premier	Molecular Dimensions	T	R23/25, R52/53	S20, S36, S45, S61
Morpheus	Molecular Dimensions	T, N	R10, R45, R46, R60, R61, R63, R23/25, R36/37/38, R48/20/22, R51/53	S20, S26, S45, S53, S61, S36/37/39
ComPAS Suite	Qiagen	T	R10, R45, R23/24/25, R36/38, R39/23/24/25, R51/53	S13, S26, S45, S53, S61, S36/37/39.
Stura/Footprint	Molecular Dimensions	T, N	R10, R45, R46, R60, R61, R25, R36/37/38, R48/20/22, R51/53	S20, S26, S45, S53, S61, S36/37/39
JCSG+ Suite	Qiagen	T, N	R10, R21, R41, R45, R23/25, R37/38, R51/53	S13, S20, S26, S45, S53, S36/37/39
Ammoniumsulfate Suite	Qiagen	T+, N	R10, R25, R26, R45, R46, R60, R61, R48/23/25, R51/53	S45, S53, S61, S36/37.
Cryos Suite	Qiagen	T, N	R10, R45, R46, R60, R61, R23/25, R36/37/38, R48/20/22, R51/53	S20, S26, S45, S53, S61, S36/37/39.
peqGOLD Plasmid Mini Kit	peqlab	-	-	-
NucleoSpin® Gel and PCR Clean-up	Macherey & Nagel	Xn	R 20/21/22, R32, 52/53	S13, S61

9.3. GHS and risk symbols and information about hazard-, risk-, safety- and precaution-statements

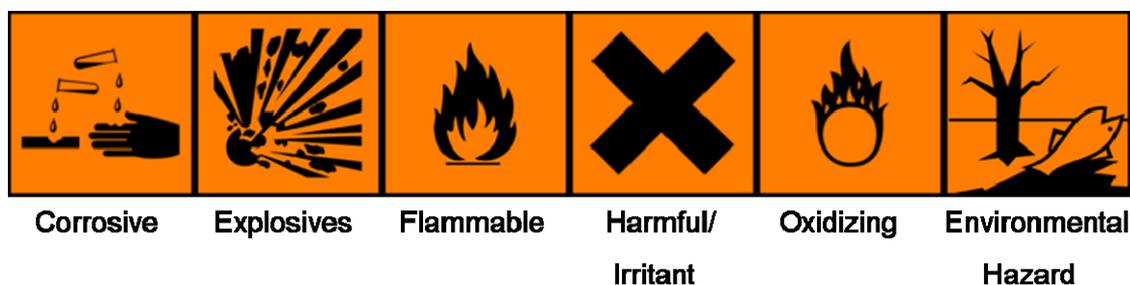


Figure 90. Hazard symbols (according to <http://www.hse.gov.uk/>) for formulations and respective risk labels



Figure 91. GHS pictograms (source: <http://www.unece.org/trans/danger/publi/ghs/pictograms.html>).

Table 33. GHS hazard statements

GHS hazard statements	
H 225	Highly flammable liquid and vapour
H 226	Flammable liquid and vapour
H 228	Flammable solid
H 272	May intensify fire; oxidizer
H 301	Toxic if swallowed
H 302	Harmful if swallowed
H 311	Toxic in contact with skin
H 312	Harmful in contact with skin
H 314	Causes severe skin burns and eye damage
H 315	Causes skin irritation
H 316	Causes mild skin irritation

H 317	May cause an allergic skin reaction
H 318	Causes serious eye damage
H 319	Causes serious eye irritation
H 330	Fatal if inhaled
H 331	Toxic if inhaled
H 332	Harmful if inhaled
H 334	May cause allergy or asthma symptoms or breathing difficulties if inhaled
H 335	May cause respiratory irritation
H 336	May cause drowsiness or dizziness
H 340	May cause genetic defects
H 341	Suspected of causing genetic defects
H 350	May cause cancer
H 350i	May cause cancer by inhalation
H 360	May damage fertility or the unborn child
H 360D	May damage the unborn child
H 361	Suspected of damaging fertility or the unborn child
H361f	Suspected of damaging fertility
H 370	Causes damage to organs
H 372	Causes damage to organs through prolonged or repeated exposure
H 410	Very toxic to aquatic life with long lasting effects
H 411	Toxic to aquatic life with long lasting effects

Table 34. GHS precautionary statements

GHS precautionary statements	
P201	Obtain special instructions before use
P210	Keep away from heat/sparks/open flames/hot surfaces – No smoking
P233	Keep container tightly closed
P260	Do not breathe dust/fume/gas/mist/vapors/spray
P261	Avoid breathing dust/fume/gas/mist/vapors/spray
P264	Wash ... thoroughly after handling
P273	Avoid release to the environment
P281	Use personal protective equipment as required
P280	Wear protective gloves/protective clothing/eye protection/face protection
P284	Wear respiratory protection
P309	IF exposed or you feel unwell: ...
P310	Immediately call a POISON CENTER or doctor/physician
P311	Call a POISON CENTER or doctor/physician
P312	Call a POISON CENTER or doctor/physician if you feel unwell
P321	Specific treatment (see respective MSDS)
P362	Take off contaminated clothing and wash before reuse
P501	Dispose of contents/container to ...
P301+310	IF SWALLOWED: Immediately call a POISON CENTER or doctor/physician
P301+P312	IF SWALLOWED: Call a POISON CENTER or doctor/physician if you feel unwell
P301+P330+P331	IF SWALLOWED: Rinse mouth. Do NOT induce vomiting
P302+P352	IF ON SKIN: Wash with soap and water
P303+P361+P353	IF ON SKIN (or hair): Remove/Take off immediately all contaminated clothing. Rinse skin with water/shower

P304+341	IF INHALED: If breathing is difficult, remove victim to fresh air and keep at rest in a position comfortable for breathing
P305+351+338	IF IN EYES: Rinse cautiously with water for several minutes. Remove contact lenses if present and easy to do – continue rinsing
P308+313	IF exposed or concerned: Get medical advice/attention
P332+313	If skin irritation occurs: Get medical advice/attention
P342+311	Call a POISON CENTER or doctor/physician
P403+233	Store in a well-ventilated place. Keep container tightly closed

Table 35. Risk statements

risk statements	
R8	Contact with combustible material
R10	May cause fire
R20	Flammable
R21	Harmful by inhalation
R22	Harmful in contact with skin
R25	Harmful if swallowed
R35	Toxic if swallowed
R36	Causes severe burns
R38	Irritating to eyesIrritating to skin
R41	Risk of serious damage to eyes
R42	May cause sensitization by inhalation
R43	May cause sensitization by skin contact
R45	May cause cancer
R46	May cause heritable genetic damage
R60	May impair fertility

R61	May cause harm to the unborn child
R39/23/24/25	Toxic: danger of very serious irreversible effects through inhalation, in contact with skin and if swallowed
R36/37/38	Irritating to eyes, respiratory system and skin
R23/24/25	Toxic by inhalation, in contact with skin and if swallowed
R20/21/22	Harmful by inhalation, in contact with skin and if swallowed
R48/20/22	Harmful: danger of serious damage to health by prolonged exposure through inhalation and if swallowed
R23/25	Toxic by inhalation and if swallowed
R36/38	Irritating to eyes and skin
R51/53	Toxic to aquatic organisms, may cause long-term adverse effects in the aquatic environment
R37/38	Irritating to respiratory system and skin

safety statements	
S20	When using do not eat or drink. Do not breathe dust
S22	In case of contact with eyes, rinse immediately with plenty of water and seek medical advice
S26	In case of accident or if you feel unwell seek medical advice immediately (show the label where possible).
S28	After contact with skin, wash immediately with plenty of ... <i>(to be specified by the manufacturer)</i>
S45	If swallowed, seek medical advice immediately and show this container or label
S46	Avoid exposure - obtain special instructions before use
S53	Avoid release to the environment
S61	Refer to special instructions/safety data sheet
S24/25	Avoid contact with skin and eyes
S36/37	Wear suitable protective clothing and gloves
S36/37/39	Wear suitable protective clothing, gloves and eye/face protection

10. Poster presentations and publications

Parts of this thesis were presented as a poster at the following conferences.

International Conference on Antivirals for Neglected and Emerging Viruses (ICAV 9), Lübeck, 10th-13th October, 2010

S. Falke, A. Kikhney, D. Svergun, R. Hilgenfeld, L. Redecke

“First Structural insights into Polyprotein Nsp7-10 from Human Coronavirus 229E”

23rd Annual Meeting of the Society for Virology, Kiel, 6th-9th March, 2013

S. Falke, R. Schubert, C. Uetrecht, J. Schulz, A. Kikhney, D. Svergun, R. Hilgenfeld, L. Redecke

“Structural Insights into Coronaviral Polyprotein Nsp7-10”

A talk was held by Sven Falke at the “5th International PhD Symposium” at the Helmholtz Zentrum für Infektionsforschung (HZI), Braunschweig, 15th December, 2011

“Structural Characterization of Non-Structural Proteins and Preprocessed Polyproteins within the Coronaviral Transcriptase/Replicase Complex”

Publications

S. Falke, M. Klinge, D. Rehders, A. Kikhney, D. Svergun, R. Hilgenfeld, L. Redecke

“First structural insights into Alphacoronavirus Nsp10 and its impact on the Interaction with Nsp16”, to be submitted

S. Falke, R. Schubert, C. Uetrecht, A. Heck, M. Kötzler, B. Meyer, A. Kikhney, M.

Kwiatkowski, H. Schlüter, D. Svergun, R. Hilgenfeld, L. Redecke

“Elucidating low-resolution structure and processing of coronaviral polyprotein Nsp7-10 in vitro”, to be submitted

11. Acknowledgement

In the end of more than three years of this project I am so glad for meeting a large number of skilled people who strongly supported and positively influenced my work. At first I would like to express an outstanding “thank you” to Dr. *Lars Redecke* heading the junior-group SIAS for giving me the opportunity to move on in science. His balanced mood and continuous support with reliable teamwork helped me a lot. This has also included a few consequent and strict but helpful advices at the right time and discussions beyond biochemistry. A special gratitude is addressed to Prof. *Rolf Hilgenfeld* who provided the truly interesting topic and was always available to share his huge experience, enthusiasm and knowledge in many fruitful discussions over the years. Furthermore, I thank Prof. *Christian Betzel* for really valuable discussions and so many small aspects of support and career guidance as well as the opportunity to work in his labs and use his infrastructure. I thank Prof. *Reinhard Bredehorst* for being the second reviewer of the thesis and Prof. *Bernd Meyer* for his commitment to be disputation referee.

From the very beginning of my project Dr. *Dirk Rehders* supported me with SPR spectroscopy and *in silico* modelling and set a good “lab atmosphere”; therefore he deserves a huge acknowledgement. Without Dr. *Charlotte Uetrecht* the native mass spectrometry experiments would not have been possible. Thank you, it was a good cooperation – not forgetting about Prof. *Albert Heck's* group in Utrecht. One important partner in terms of speculation about coronavirus proteins is *Yibei Xiao*. Thank you for the cooperation on analysing polymerase activity and for being always available by E-mail. I also thank Dr. *Alexey Kikhney* for a strong support in learning how to analyze SAXS data and Prof. *Dmitri Svergun* and other people from EMBL for support and coordination during the synchrotron experiments.

Ich mache die zweite Hälfte mal auf Deutsch und hoffe, dass ich niemanden vergesse. Weit oben auf der Liste stehen auch *Robin Schubert* und *Marco Klinge*. Ein großes Lob und Dank gilt euch für viel Kreativität, großes Engagement und Fingerspitzengefühl in der Weiterentwicklung des Projekts. Danke für viel Spaß. Ebenso gilt mein Dank weiteren Leuten die das Projekt mit vollem Einsatz und Freude an unterschiedlichen Stellen praktisch unterstützt haben: *Timur Yorgan*, *Henrike Friedenberger*, *Bianca Schlee*, *Raphael Klein* – und natürlich *Birgitta Fried*, der immer positive gute Geist des Labors. Der Gruppe von Prof. *Hartmut Schlüter* dank ich für massenspektrometrische Untersuchungen und Beratung, insbesondere *Marcel Kwiatkowski* war eine große Unterstützung. Dr. *Miriam Koetzler* danke ich sehr für ihre Bereitschaft NMR Experimente durchzuführen und mir die Technik etwas näher zu bringen. Darüber hinaus bedanke ich mich herzlich bei allen Doktoranden der Graduiertenschule SDI. Natürlich bin ich noch allen Leuten zu Dank verpflichtet, die über die Jahre im Labor waren oder immer noch dort sind bzw. den Alltag an der Uni etwas schöner, interessanter, spannender gemacht haben. An alle Gäste und Kollegen von SIAS, den Arbeitskreisen Betzel und Hahn und an der Uni insgesamt: Danke euch allen!

Ganz besonders danke ich meiner Familie, die mir diesen Schritt zum großen Teil erst ermöglicht hat und mich immer zu 100% unterstützt hat. Danke an alle Freunde, ich kann mich glücklich schätzen.

H. F. V. S.

12. Curriculum vitae

- entfällt aus datenschutzrechtlichen Gründen -

Selbstständigkeitserklärung

Hiermit erkläre ich an Eides statt, dass ich die an der Universität Hamburg zur Promotion eingereichte Dissertation mit dem Titel „Coronaviral Polyprotein *Nsp7-10*: Proteolytic Processing and Dynamic Interactions within the Transcriptase/Replicase Complex“ im Institut für Biochemie des Fachbereichs Chemie der Universität Hamburg ohne sonstige nicht angeführte Hilfe durchgeführt und bei der Abfassung der Dissertation keine anderen als die dort aufgeführten Hilfsmittel benutzt habe.

Ferner versichere ich, dass ich bisher an keiner in- oder ausländischen Universität ein Gesuch um Zulassung zur Promotion eingereicht und daher weder diese noch eine andere Arbeit als Dissertation vorgelegt habe.

Sven Falke

