Aus dem Institut für Biochemie
der Universität zu Lübeck
Direktor: Prof. Dr. rer. nat. Dr. h.c. Rolf Hilgenfeld

Crystallographic and Biochemical Investigations on Coronavirus

Replication Proteins – Non-Structural Proteins 8 and 9

Inauguraldissertation

zur

Erlangung der Doktorwürde

der Universität zu Lübeck

- Aus der Technisch-Naturwissenschaftlichen Fakultät -

vorgelegt von

Rajesh Ponnusamy

aus Thirukkovilur, Indien

Lübeck, 2009

1. Berichterstatter/Berichterstatterin: Prof. Dr. rer. nat. Dr. h.c. Rolf Hilgenfeld

2. Berichterstatter/Berichterstatterin: Prof. Dr. rer. nat. Thomas Peters

Tag der mündlichen Prüfung: 08.06.2010

Zum Druck genehmigt. Lübeck, den 09.06.2010

Acknowledgements

TABLE OF CONTENTS

ABBREVIATIONS

1,8-ANS	1-Anilinonaphthalene-8-sulfonic acid
1D-NMR	one dimensional-nuclear magnetic resonance
Amp	ampicillin
CD	circular dichroism
Cm	chloramphenicol
CoV	coronavirus
DLS	dynamic light-scattering
DMSO	dimethylsulfoxid
DNA	deoxyribonucleic acid
ds	double-stranded
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EF TU	elongation factor thermo unstable
eIF4F	eukaryotic translation initiation factor 4F
gRNA	genomic RNA
H_2O_2	hydrogen peroxide
HCoV-229E	human coronavirus 229E
hnRNPA1	heterogeneous nuclear ribonucleoprotein A1
HVR	hypervariable region
ICP8	infected-cell protein 8
IPTG	isopropyl-β-D-thiogalactopyranosid
K_D	dissociation constant
kDa	kiloDalton
M^{pro}	main proteinase
min	minute
µl	microlitre
µM	micromolar
mM	millimolar

MPD	2-methyl-2,4-pentanediol
mRNA	messenger RNA
nm	nanometer
Nsp	non-structural protein
NTA	nitrilotriacetic acid
OB	oligonucleotide/oligosaccharide-binding
ORF	open reading frame
PDB	Protein Data Bank
pp1a	polyprotein 1a
RdRp	RNA-dependent RNA polymerase
r.m.s	root-mean-square
RNA	ribonucleic acid
rpm	rounds per minute
RTC	replicase/transcriptase complex
RU	resonance units
SA chip	streptavidin-coated sensor chip
SARS-CoV	severe acute respiratory syndrome coronavirus
SDS	sodium dodecyl sulfate
sgmRNA	subgenomic RNA
SPR	surface plasmon resonance
ss	single-stranded
SSB	single-stranded DNA-binding protein
tRNA	transfer RNA
TRS	transcription-regulating sequence
UTR	untranslated region
v/v	volume/volume
wt	wild type
ZIGE	zone-interference gel electrophoresis

1. Introduction

1.1 Nidoviruses

The name *Nidovirales* (from the Latin *nidus*, nest) refers to the 3´-coterminal nested set of subgenomic (sg) viral mRNAs that is produced during infection. *Nidovirales* is an order comprising three families, *Coronaviridae* (consisting of the genera *Coronavirus* and *Torovirus*), *Arteriviridae* (consisting of the genus *Arterivirus*) and *Roniviridae* (consisting of the genus *Okavirus*) (Fauquet *et al.,* 2005). *Arteriviridae* and *Roniviridae* are only known to infect animals, whereas member of the *Coronaviridae* family infects both animals and humans. The common features of nidoviruses are the genome organization and replication strategy. These genomes share the structure of eukaryotic mRNA and involves host-cell protein(s) for gene expression. The order *Nidovirales* contains the coronavirus (MHV-A59) that has the largest known non-segmented positive-stranded RNA genome. The virus morphology and the size of the genome are very different among nidoviruses.

1.2 Coronaviruses

The first isolated coronaviruses were Infectious Bronchitis Virus (IBV) in the 1930s, and Mouse Hepatitis Virus (MHV) and Transmissible Gastroenteritis virus (TGEV) in the 1940s. The name coronavirus is derived from the Greek κορώνα, meaning crown, as the virus envelope appears to have a crown-like shape under the electron microscope. The first two human coronaviruses identified are Human Coronavirus 229E (HCoV-229E) and Human Coronavirus OC43 (HCoV-OC43) in the 1960s. Since then, several other new coronaviruses have been identified, which infect a wide range of species. These coronaviruses are classified into three different groups based on the serological and antigenic-cross reactivity. Group I and II coronaviruses mainly infect mammals, whereas group III exclusively infect birds, but also very recently identified in the blue whale (Cavanagh, 2005; Mihindukulasuriya *et al.,* 2008; Table 1.1). The coronaviruses have been studied extensively from 1965 to the mid-1980s and believed to be relatively harmless respiratory pathogens causing common cold. This conviction ended with the outbreak of Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) in Southern

China during late 2002 (Drosten *et al.,* 2003; Ksiazek *et al.,* 2003; Kuiken *et al.,* 2003; Peiris *et al.,* 2003). SARS-CoV is classified as an outliner of group II coronaviruses (Snijder *et al.,* 2003). The SARS outbreak in humans was due to zoonotic transition, most likely from bats (Groneberg *et al.,* 2005; Roberts *et al.,* 2005) and was transmitted via infected civet cats. Unlike other coronaviruses, SARS-CoV can infect a wide range of mammals, including humans, nonhuman primates, Himalayan palm civets, raccoon dogs, cats, and dogs (Lau *et al.,* 2005; Li *et al.,* 2005). After the SARS-CoV outbreak, research

Table 1.1: Coronaviruses, host ranges, and diseases

Group	Virus	Host	Respiratory infection	Enteric Infection	Hepatitis	Neurologic infection	Other[&]
I	HCoV-229E	Human	X			?	
	HCoV-NL63	Human	X				
	TGEV, PRCoV	Pig	X	X			X
	CCoV	Dog		X			
	FECoV	Cat		X			
	FIPV	Cat	X	X	X	X	X
	RbCoV	Rabbit		X			X
II	HCoV-OC43	Human	X	?		?	
	HCoV-HKU1	Human	X				
	SARS-CoV	Bat	X	X			X
	MHV	Mouse	X	X	X	X	
	SDAV	Rat					X
	HEV	Pig	X	X		X	
	BCoV	Cow	X	X			
	Bat SARS-like CoV	Bat**					
III	IBV	Chicken	X		X		X
	TCoV	Turkey	X	X			
	SW1	Whale	X				

HCoV-229E, -NL63, -OC43, and -HKU1, human respiratory coronaviruses; SARS-CoV, severe acute respiratory syndrome coronavirus; TGEV, porcine transmissible gastroenteritis virus; PRCoV, porcine respiratory coronavirus; CCoV, canine coronavirus; FECoV, feline enteric coronavirus; FIPV, feline infectious peritonitis virus; RbCoV, rabbit coronavirus; MHV, murine hepatitis virus; SDAV, sialodacryoadenitis virus; HEV, porcine hemagglutinating encephalomyelitis virus; BCoV, bovine coronavirus; IBV, avian infectious bronchitis virus; TCoV, turkey coronavirus. [&]Other diseases caused by coronaviruses include infectious peritonitis, immunologic disorders, runting, nephritis, pancreatitis, parotitis, myocarditis, and sialodacryoadenitis. **It is not known if bat SARS-like CoV infection causes illness. Table adopted from the review article by Lai *et al.* (2006).

interest in coronaviruses increased dramatically. Several new coronaviruses have been identified and reported from humans, bats and whale (van der Hoek *et al.,* 2004; Woo *et al.,* 2005; Mihindukulasuriya *et al.,* 2008) and additional coronaviruses are likely to be identified in the near future.

1.3 Structure of the coronavirion and its components

The coronavirions are spherical enveloped particles about 100 to 160 nm in diameter (Fig. 1.1 a, b). The virion consists of a single-stranded, positive-sense genomic RNA 27 to 32 kb in size (Lai & Cavanagh, 1997) wrapped around the nucleocapsid protein (N). The nucleocapsid-RNA complex is anchored inside the virion with the help of the membrane (M) protein. The M protein is one of the three canonical membrane proteins, which are anchored in a lipid envelope. The two other membrane proteins are the envelope protein (E) and the spike protein (S). Commonly, group II coronaviruses have an additional fourth membrane protein, the hemagglutinin-esterase protein (HE) (de Haan & Rottier, 2005). Recently, Nsp3 (multi-domain protein), Nsp5 (cysteine protease), Nsp2, ORF9b, and ORF3a were also identified to be incorporated inside the virion (Neumann *et al.,* 2008). But their encapsulation within the virus is not clear and yet to be elucidated in detail.
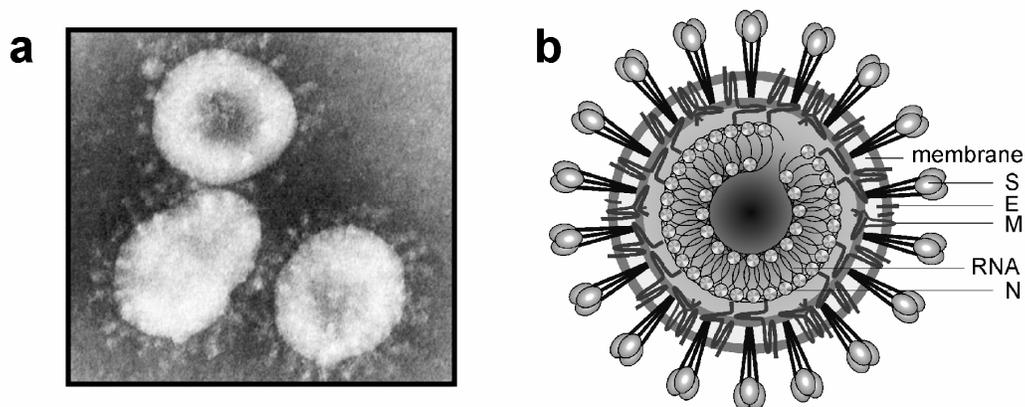
Fig. 1.1 Electron micrograph and systematic representation of Coronavirus

(a) Electron micrograph picture of HCoV-229E with approximately 60 x magnification (F.A. Murphy, School of Veterinary Medicine, University of California, Davis (http://www.vetmed.ucdavis.edu/viruses/download.html)). (b) Artistic view of coronavirus virion structure. Viral membrane proteins present in the virion envelope are depicted. Adopted from Siddell *et al.* (2005).

1.4 Virus attachments and disassembly

The first step of virus entry is the attachment of spike protein to the receptors. The coronavirus spike protein is a membrane-anchored glycoprotein that is found on the virion surface in a trimeric form (Neuman *et al.,* 2006). The spike protein consists of a receptor-binding domain in the N-terminal half (S1), and a membrane-anchored domain in the C-terminal half (S2). The linker is responsible for the fusogenic activity using two heptad repeats, HR1 and HR2. Group I coronaviruses mainly use CD13 (aminopeptidase N) as a receptor (Tusell *et al.,* 2007). However, the recently identified group-I HCoV-NL63 uses a zinc peptidase, ACE2 (Angiotensin Converting Enzyme 2), as a receptor, similar to the SARS-CoV (Hofmann *et al.,* 2005; Li *et al.,* 2003). MHV from group II uses carcinoembryonic antigen-related cell adhesion molecules as a cellular receptor (Williams *et al.,* 1991). Receptors of group III coronaviruses have not yet been identified. Virus entry into the target cell is facilitated either by direct fusion with the plasma membrane or by endocytosis. The key step in direct fusion with the plasma membrane is a conformational change of the spike protein, which is responsible for bringing the viral and the host-cell membranes into close proximity (Bosch *et al.,* 2004). Recently, Wang *et al.* (2008) discovered that SARS-CoV uses receptor-dependent endocytosis. This process occurs at acidic pH and uses lipid rafts (sphingolipid- and cholesterol-rich domains of plasma membrane that contain a variety of signaling and transport proteins) for entry. The exact mechanisms of entry of coronaviruses from other serological groups are so far unclear.

1.5 Molecular biology of coronaviruses

1.5.1 Coronavirus genome organization

Coronavirus genome length varies from 27,317 nt for HCoV-229E to 31,357 nt for MHV-A59. The genomic RNA is capped and polyadenylated. The organization is very similar among the coronaviruses and within the order, nidovirus. The 5′ region starts with the leader sequence varying between 65 and 98 nt in length; this leader sequence is present in all the genomic and the subgenomic messenger RNAs (sgmRNAs). The sgmRNAs is responsible for the synthesis of the structural and accessory proteins. The leader sequence is followed by the 5′-Un-Translated Region (UTR) of 209 to 528 nt in

length. Another UTR is found at the 3´ end of the genome, with the length varying from 288 to 506 nt, followed by a poly(A) tail of varying length. These 5´ and 3´ UTRs are important for RNA replication and transcription (see below). The replicase and transcriptase complex of the virus is encoded within two large overlapping open reading frames (ORF), ORF 1a and ORF 1b, present in the 5´ proximal region of the genome. ORF 1a and 1b are about 20 to 22 kb in length and cover almost two third of the entire genome. The remaining one third of the genome consists of 7 to 14 ORFs in the 3´ proximal region. This 3´ region includes the structural proteins and the group-specific accessory proteins. The order of the genes encoding the ORF 1a and 1b (also called *Rep* for replicase) and the structural protein is: 5´-*Rep*-S-E-M-N-3´, taking into account that in between these genes are ORFs, encoding nonstructural proteins, which are specific for the different coronavirus groups (Fig. 1.2).

1.5.2 Higher-order structures in UTR regions

Higher-order structures in UTR regions are very important components of the coronavirus life cycle. Often these higher order UTR regions are the anchoring point for the formation of the replicase/transcriptase complex, which includes several host and viral proteins, and for the protection of nascent RNA from degradation (Andino *et al.,* 1990, 1993; Galán *et al.,* 2009). Both 5´ and 3´ UTR regions are essential for positive-strand RNA synthesis. However, only the last 55 nucleotides and the poly(A) tail present in the 3´-UTR are decisive for minus-strand RNA synthesis (Lin *et al.,* 1994). The 5´-UTR region varies in length from 209 to 528 nt among the coronaviruses. At least five stem-loop regions have been mapped in the 5´-UTR region among all three groups of coronaviruses (Raman *et al.,* 2003). Within the 5´-UTR, the conserved transcription-regulating sequence (TRS, also known as intergenic sequence) "UCUAAAC" is present, and to this sequence nucleocapsid (N) protein binds with high affinity (Nelson *et al.,* 2000). The polypyrimidine tract binding protein also binds near the leader adjacent TRS sequence (Li *et al.,* 1999) and hnRNP A1 binds to the minus-strand complement to the leader-adjacent TRS motif (Li *et al.,* 1997). In addition, several other host and viral proteins are anticipated to be part of the 5´-stem loop-binding region.

The 3′-UTR region varies from 288 to 506 nt in length among the coronaviruses (excluding the poly(A) tail). An essential bulged-stem loop, an overlapping pseudoknot and the Hyper Variable Region (HVR), which contains the highly conserved octanucleotide motif 5′-GGAAGAGC-3′ that is predicted to be present in the 3′-UTR region of all group II coronaviruses (Goebel *et al.,* 2007). Group I and III coronaviruses also contain the octanucleotide in the HVR region and a pseudoknot, however these regions exhibit varying length.

Up to now, proteins such as poly(A)-binding protein (Spagnolo & Hogue, 2000), mitochondrial aconitase (Nanda & Leibowitz, 2001), polypyrimidine tract-binding protein (Huang & Lai, 1999) are known to interact with the 3′-UTR region or its minus-strand counterpart. Higher-order structures present in the 3′-UTR region are proposed to be involved in the initiation and/or regulation of plus-strand synthesis, perhaps via genome circularization with the use of poly(A)-binding protein (reviewed in Brian & Baric, 2005).

1.5.3 Translation

Coronavirus genomic RNA is released into the cytoplasm of the infected cells. Positive-strand RNA viruses do not encapsulate the polymerase within the virus. Therefore, translation of viral RNA is a prerequisite for the RNA to be replicated. The coronavirus genome is capped, polycistronic and positive-sense. So once the RNA is released into the cytoplasm, the genome is presumably translated by a cap-dependent translation mechanism. The eukaryotic cap-dependent translation mechanism is well studied and reviewed (Pestova *et al.,* 2001). Briefly, translation occurs in three steps, initiation, elongation, and termination. Initiation is the time-consuming step in translation; all eukaryotic mRNAs have a 5′-terminal cap structure, eukaryotic initiation factor 4 F (eIF4F), a hetero-trimeric protein, recognizes the 5′-cap structure, specifically the eIF4E domain of the eIF4F protein. After eIF4F binds to mRNA, several initiation factor proteins and the 40S ribosomal subunit join to form the initiation complex. Once this initiation complex is formed, the 40S ribosomal subunit scans the mRNA for the initiation codon, AUG. After encountering the first AUG codon, the 60S ribosomal

subunit combines with the initiation complex, and elongation starts. Since the coronavirus genome also has the same polarity and resembles the eukaryotic mRNA structure with 5′-cap and 3′-poly(A) tail, it is very likely translated in a similar fashion as the eukaryotic mRNA. ORF1a is directly translated into polyprotein 1a (pp1a), with a calculated molecular mass of ~ 454 kDa. Involving a (-1) ribosomal frame-shift, translation of ORF1a and ORF1b together yields the giant polypeptide 1ab (pp1ab), with a calculated molecular mass of ~ 754 kDa. The signal mediating the frame-shift consists of a 'slippery' sequence, UUUAAC, and a downstream pseudoknot structure (Namy *et al.,* 2006; *for review* Plant & Dinman, 2008). The coronavirus genome is polycistronic and two or three proteases are responsible for the proteolytic cleavage of pp1a and pp1ab to achieve 16 mature non-structural proteins (15 Nsps in case of IBV), which together with the help of host proteins form the replicase/transcriptase complexes (Fig. 1.2).
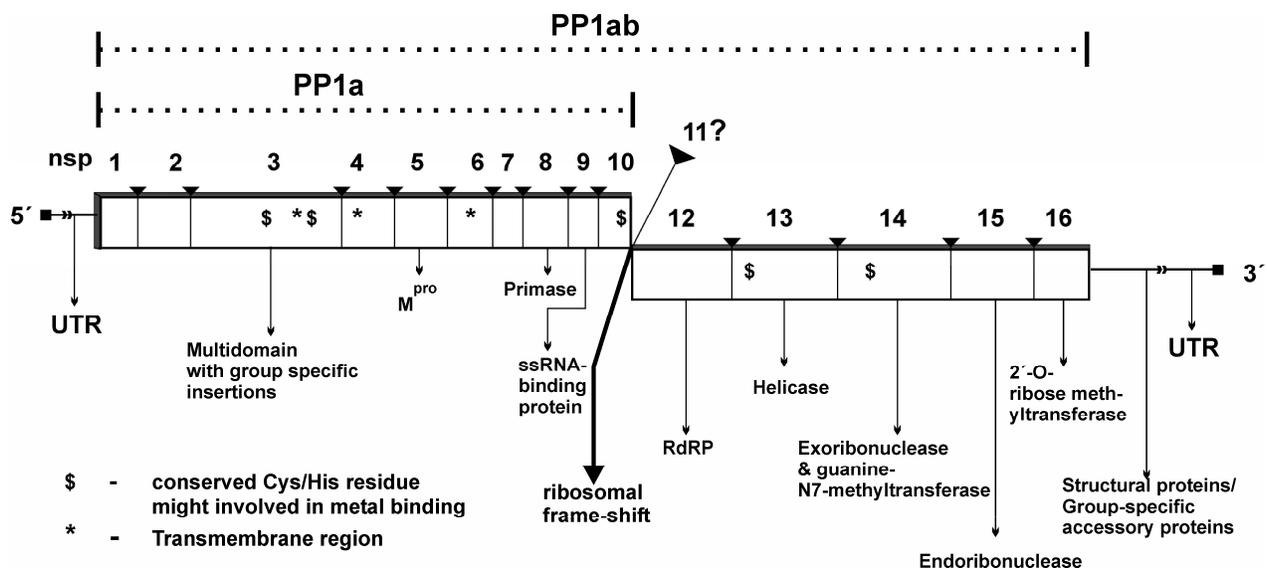
Fig. 1.2 Genome organization of coronaviruses

However, it is not yet clear how the coronavirus genome battles with the host mRNA, for recruiting the initiation factor proteins. Different viruses evolve very different mechanism to compete or to shut-off the host protein translation system. For instance, viruses encode a specific secondary structure in the 5′-UTR region such as Internal Ribosome Entry Sites (IRES) (Vagner *et al.,* 2001), viral proteins specifically interacting with initiation factor proteins (for review Prévot *et al.,* 2003), viral proteases specifically cleaving the initiation factor proteins and poly(A)-binding proteins (Rivera & Lloyd, 2008). First

evidence for the race between the viral genome and the host mRNA in the coronavirus family comes from SARS-CoV. Nsp1 promotes host-mRNA degradation and thereby suppresses host-gene expression, including that of genes involved in host innate immune response (Kamitani *et al.,* 2006). The N-terminal domain of the spike protein from IBV and SARS-CoV specifically interacts with eIF4F and modulates host-gene expression (Xiao *et al.,* 2008). There are several other proteins that are expected to be involved in modulating host mRNA translation.

1.5.4 Replication and transcription

All studied +RNA viruses usurp and modify cytoplasmic membranes for the formation of functional sites for polyprotein processing and RNA synthesis as described in recent reviews (Mackenzie, 2005; Miller & Krijnse-Locker, 2008; Novoa *et al.,* 2005; Ahlquist, 2006). Coronaviruses hijack their double-membrane vesicles from the endoplasmic reticulum or late endosomes (Gosert *et al.,* 2002; Prentice *et al.,* 2004; Snijder *et al.,* 2006; van Hemert *et al.,* 2008). The components of the coronavirus Replicase/Transcriptase Complex (RTC) are released from pp1a and pp1ab by two or three virally encoded proteases; papain-like cysteine proteases (PL1^{pro} and PL2^{pro}) residing in the multi-domain Nsp3, and the main proteinase (M^{pro}; Nsp5). M^{pro} is responsible for the majority of the polyprotein cleavages, which occur in a distinct and sequential manner (Ziebuhr & Siddell 1999; Anand *et al.,* 2002b, 2003). The proteinase-mediated regulation of replication has been reported for other positive-stranded RNA viruses, such as switching off minus-strand synthesis in alphavirus (Lemm *et al.,* 1994). However, in coronaviruses it is not yet clear whether the uncleaved polyprotein has any specific role in replication and/or transcription. Once Nsps are released from the polyprotein, the replicase proteins are used to produce all subsequent genomic (replication) and sub-genomic (transcription) RNA species. The replication of genomic RNA (gRNA) is presumably performed by uninterrupted synthesis involving both positive- and minus-stand full-length RNA. Unlike replication, transcription occurs by discontinuous synthesis for the production of subgenomic messenger RNAs (sgmRNAs). These sgmRNAs carry a common 5′-leader sequence derived from the 5′-end of gRNA (Lai & Cavanagh, 1997). The fusion of the 5′-leader sequence to each sgmRNA occurs

through the base paring between the leader and the body sequence in the region of the TRS. Two different models have been proposed for the discontinuous transcription. Both models suggest that base pairing between the leader and the body TRS is a hallmark of a discontinuous transcription. The models differ with the occurrence of a discontinuous step during minus- or positive-strand synthesis. Compelling evidence has indicated that minus-strand synthesis is the point in replication, during which the discontinuous step occurs (Baric & Yount, 2000; van Marle *et al.,* 1999; Pasternak *et al.,* 2001; Sawicki *et al.,* 2001; Sawicki & Sawicki, 1995; Sethna *et al.,* 1989; Zuniga *et al.,* 2004; for reviews Pasternak *et al.,* 2006; Sawicki *et al.,* 2007). The current minus-strand model proposes that the RTC transcribes minus-strand copies of the genome and attenuates at one of the various body TRSs. Together with the nascent strand the RTC translocates to the leader TRS, guided by the complementarity between the 3´-end of the nascent minus strand and the first TRS motif from the 5´-end of the genome, to copy the leader sequence. The resulting minus-strand of the particular sgmRNA is subsequently transcribed to a translatable positive-sense sgmRNA. Unfortunately, the replication and transcription models are yet to be validated in detail.

Several of the coronaviral Nsps being replicative enzymes such as RNA-dependent RNA polymerase (RdRp; Nsp12) and helicase (Nsp13), which are always present with positive-strand RNA viruses. But there is also a set of intriguing enzymes lacking in other RNA viruses, such as 3´- to 5´-exonuclease (N-terminal domain of Nsp14), an uridylate-specific endonuclease (Nsp15), and S-adenosylmethionine (AdoMet)-dependent N7- and 2´-O-methyltransferases (C-terminal domain of Nsp14 and Nsp16, respectively); these are distantly related to cellular RNA-processing enzymes. Moreover, the 3´-region of pp1a comprises a set of relatively small polypeptide domains, Nsp7 to Nsp11. These polypeptides are very unique and their functional role in the coronaviruses is yet unclear. Though, most of the three-dimensional structures of coronavirus Nsps are available for this part of the coronavirus genome. The fold of these proteins is as unique as their role in the coronaviruses. Nsp8 of SARS-CoV has the function of an RNA primase (Imbert *et al.,* 2006); its 8:8 complex with Nsp7 has a three-dimensional structure reminiscent of the β_2 "sliding clamp" of bacterial DNA polymerase, with a central channel suitable for double-

stranded RNA binding (Zhai *et al.,* 2005). Nsp9 from SARS-CoV and HCoV-229E bind to ss-RNA/DNA without any sequence specificity, and their fold is a variant of the oligonucleotide/oligosaccharide binding (OB) fold (Egloff *et al.,* 2004; Sutton *et al.,* 2004; Ponnusamy *et al.,* 2006, 2008). Nsp10 of MHV is a double-stranded RNA-binding zinc-finger protein (Matthes *et al.,* 2006) and two different oligomeric forms (monomeric and dodecameric) of SARS-CoV Nsp10 have been elucidated by X-ray crystallography (Joseph *et al.,* 2006; Su *et al.,* 2006). Nsp11 is present in the 5′-proximal region of the frame-shift between pp1a and pp1ab, its existence as a mature protein is still questionable. The crystal structure of Nsp10 including Nsp11 at the C-terminal failed to reveal any ordered structure for Nsp11 (Su *et al.,* 2006). Very likely, these non-structural proteins are involved directly in the replication complex built around the RdRp (Nsp12), which is anchored to the intracellular membrane using the three membrane anchoring domains residing in Nsp3, Nsp4, and Nsp6 (Oostra *et al.,* 2008). Three independent genome-wide analyses elucidated the interaction between the SARS-CoV Nsps. von Brunn *et al.* (2007) reported 70 pairs of interactions; while, Imbert *et al.* (2008) reported only 17 pairs of interactions, about half of which were related to Nsp3. Surprisingly, none of the interaction revealed in the two studies were overlapping, although both employed the yeast two-hybrid system for the screening. In contrast, though Pan *et al.* (2008) studied the interactions using a mammalian system, they identified 9 pairs of interactions overlapping with von Brunn *et al.* (2007) and 4 pairs of interactions overlapping with Imbert *et al.* (2007). Meanwhile, van Hemert *et al.* (2008) successfully isolated the active SARS-CoV RTC from the virus-infected cells, but very little is known about the composition of the RTC complex. The isolated active RTCs are membrane associated and required the addition of a cytoplasmic host factor. Unfortunately, the molecular mechanism of the RTC is still ambiguous and it is not understood in detail.

1.6 Aim and objective of this thesis

The continuous (re-)emergence of life-threatening RNA viruses, such as SARS-CoV and Influenza virus, posses a serious problem to human health. Studying proteins from various coronaviruses, those that infect humans and those that so far have been restricted

to animals but may cross the species-barrier by zoonotic transition, will accelerate the discovery and development of antiviral therapeutics in case of a new pandemic.

There is a set of intriguing coronaviral replicase proteins, Nsp7 to Nsp11, present in the C-terminal region of polyprotein 1a (pp1a), which are poorly understood. Deletion of any of the regions encoding Nsp7 to Nsp10 from MHV-A59 was lethal (Deming *et al.,* 2007). The aim of this thesis is to study the Nsp8 and Nsp9 from different coronaviruses by characterizing these proteins both structurally and biochemically. The main objective is to understand the role of these proteins in the coronaviral life-cycle.

Herein, Nsp8 from HCoV-229E and SARS-CoV were characterized and the structural data elucidated using bioinformatics and various biophysical techniques such as CD-spectroscopy, DLS and 1D-NMR. The ability of Nsp8 to bind nucleic-acid and other Nsps was investigated and is discussed in detail. Further, the crystal structures of Nsp9 from HCoV-229E and TGEV are presented. The novel disulfide-linked dimer of HCoV-229E Nsp9, very different from the previously reported SARS-CoV Nsp9 dimer (Sutton *et al.,* 2004, Egloff *et al.,* 2004), and the Nsp9 Cys69Ala dimer are compared at length. The various oligomerization modes of Nsp9 with respect to RNA binding are elucidated. In order to understand the Nsp9 structure and its function, four mutant proteins were produced. Wild type and mutant proteins were biochemically characterized.

The results generated herein are aimed towards shedding light on numerous aspects of coronaviral replication, which may facilitate the development of novel antivirals.

2. Materials and Methods

2.1 Materials

2.1.1 Equipment

A list of laboratory instruments and devices used for this work is given in Table 2.1 below.

Table 2.1: General laboratory devices and their manufacturers

Equipment	Manufacturer
Cloning, expression, purification and characterization	
Plasmid Maxi/Midi prep Kit	Qiagen (Hilden, Germany)
GeneJET^{TM} Plasmid Miniprep Kit	Fermentas (St. Leon-Rot, Germany)
Thermocycler	Biometra (Göttingen, Germany)
Centrifuge - Biofuge	Heraeus (Hanau, Germany)
Gel-electrophoresis Mini-Sub$^®$ cell GT cell	Bio-Rad (München, Germany)
CARY 50 UV-Vis Spectrophotometer	Varian (Darmstadt, Germany)
Wizard$^®$ SV Gel and PCR clean-up system	Promega (Mannheim, Germany)
Analytical balance - Sartorius 2024 MP6	Sartorius (Göttingen, Germany)
Table balance - PB3002	Mettler Toledo (Giessen, Germany)
pH Meter - InLab$^®$406	Mettler Toledo (Giessen, Germany)
His-Trap^{TM} HP column (1ml)	GE Healthcare (München, Germany)
Gel-filtration superdex^{TM} 75	GE Healthcare (München, Germany)
ÄKTA Prime	GE Healthcare (München, Germany)
French$^®$ press	Thermo IEC (Schwerte, Germany)
Ultracentrifuge Centrikon T-2070	Kontron (Zürich, Switzerland)
Amicon Stirred Cell -50/10/3 ml	Millipore (Schwalbach, Germany)
Dialysis membrane - Spectra/Por$^®$	Spectrum (Breda, The Netherlands)
Blot membrane - Roti$^®$-NC	Roth (Karlsruhe, Germany)
Dynamic light-scattering	RiNA (Berlin, Germany)
CD - Jasco J-715 a spectropolarimeter	Jasco (Groß-Umstadt, Germany)
Biacore	GE Healthcare (Freiburg, Germany)

CARY Eclipse - Fluorescence Spect.	Varian (Darmstadt, Germany)
Crystallization	
Incubator	Viessmann (Allendorf, Germany)
Microscope - SZH10 binocular	Olympus (Hamburg, Germany)
Crystallization - Phoenix robot	Dunn Labortechnik (Thelenberg, Germany)
Crystallization image system - Rock imager	Formulatrix (Amsterdam, Netherlands)
Crystallization plate - 24 wells	Hampton Research (Laguna Niguel, USA)
Crystallization Intelli plate - 96 wells	Art Robbins (CA, USA)
Data collection	
X-ray generator - Xcalibur PX Ultra	Oxford Diffraction (Oxfordshire, UK)
Cryojet XL/HT	Oxford Cryosystems (Oxfordshire, UK)
Onyx CCD detector	Oxford Diffraction (Oxfordshire, UK)
Synchrotron - University of Hamburg - University of Lübeck - EMBL beamline X13	DESY (Hamburg, Germany)
Detector - CCD 165 mm	Mar Research (Hamburg, Germany)
Cryostream - Oxford 700 series	Oxford Cryosystems (Oxfordshire, UK)

2.1.2 Chemicals

Unless otherwise specified, chemicals and reagents were purchased in the highest available quality from Sigma (München, Germany), Merck KGaA (Darmstadt, Germany), Fluka (Seelze, Germany), or Roche Diagnostics (Mannheim, Germany). Enzymes and buffers for cloning were from New England Biolabs (Beverly, USA) or Fermentas (St. Leon-Rot, Germany).

2.1.3 Vector and *E. coli* expression strains

Vectors pET-15b and pET-11a: Novagen (Darmstadt, Germany). DH5α - cloning purpose: Invitrogen (Karlsruhe, Germany); Tuner[TM](DE3)pLacI and B834 (DE3) expression strains: Novagen (Darmstadt, Germany).

2.1.4 Crystallization materials and cryo-tools

Glass sample capillaries: GLAS (Berlin, Germany). Highly liquid paraffin oil: Merck (Darmstadt). Vaselin: Weißes (Frankfurt, Germany). Magnetic basic crystal caps, mounted cryo-loops, 22 mm circular siliconized coverslips, sealing wax, crystal storage vials, cryo-canes, magnetic crystal wands, curved vial clamps and microtools: Hampton Research (Laguna Niguel, USA). Crystallization basic and extension screening kits: Sigma-Aldrich (München, Germany)

2.2 Methods

2.2.1 Cloning

The regions coding for Nsp9 of HCoV 229E and SARS-CoV were amplified by the polymerase chain reaction from virus-derived cDNA fragments. The HCoV-229E nsp9 PCR product was cloned into the pET-15b vector, resulting in pETHCoV-229E/nsp9, which contained the full-length gene (coding for pp1a residues 3825-3933) with an N-terminal extension (MHHHHHHVKLQ), including a hexahistidine tag for protein purification and a SARS-CoV main-proteinase (M^{pro}) cleavage site (VKLQ$^{\downarrow}$NNE...) for tag removal. The same approach was chosen for SARS-CoV Nsp9, which comprises pp1a residues 4118-4230, resulting in the construct pETSARS-CoV/nsp9. Four mutants were prepared using single-site mutagenesis. The corresponding codon of the amino acid to be mutated was encoded in the primer (Table 2.2). The pETHCoV-229E/nsp9 or pETSARS-CoV/nsp9 plasmids were used as a template in the PCR reaction. The PCR product corresponding to a size of ~6.05 kb was purified and the template was removed using the restriction enzyme DpnI. The restricted product was transformed into *E. coli* XL1-blue supercompetent cells. The correctness of the mutations (Cys69→Ala and Cys69→Ser for HCoV 229E, and Cys73→Ala, Cys73→Ser for SARS-CoV) was confirmed by DNA sequencing. The primers and restriction enzymes used for cloning Nsp9 genes are tabulated in Table 2.2.

Table 2.2: Oligonucleotide primers

Name[*]	Sequence	Restriction enzyme
229E_Nsp9_wt_S	GAA TTG GAA CCA CCT TGC AGA TTT GTT ATA GAC	XhoI
229E_Nsp9_wt_R	GTC TAT AAC AAA TCT GCA AGG TGG TTC CAA TTC	BamHI
229E_Nsp9_Cys69Ala_S	GAA TTG GAA CCA CCT **GCC** AGA TTT GTT ATA GAC	§
229E_Nsp9_Cys69Ala_R	GTC TAT AAC AAA TCT **GGC** AGG TGG TTC CAA TTC	§
229E_Nsp9_Cys69Ser_S	GAA TTG GAA CCA CCT **AGC** AGA TTT GTT ATA GAC	§
229E_Nsp9_Cys69Ser_R	GTC TAT AAC AAA TCT **GCT** AGG TGG TTC CAA TTC	§
229E_Nsp7_wt_S	GGG ATT CTC GAG TCT ACT TTG CAG TCT AAA TTG ACT GAT CTT AAG TGC	XhoI
229E_Nsp7_wt_R	GGG TTG GGA TCC CTA TTG CAA AAT GGA GTC GTT CTC AAA	BamHI
SARS_Nsp9_wt_S	GGG AGT CTC GAG CTT AAA TTG CAG AAT AAT GAA CTG	XhoI
SARS_Nsp9_wt_R	GGG TTG GGA TCC TTA CTG AAG ACG TAC TGT AGC AGC TAA	BamHI
SARS_Nsp9_Cys73Ala_S	GAA CTG GAA CCA CCT **GCT** AGG TTT GTT ACA GAC	§
SARS_Nsp9_Cys73Ala_R	GTC TGT AAC AAA CCT **AGC** AGG TGG TTC CAG TTC	§
SARS_Nsp9_Cys73Ser_S	GAA CTG GAA CCA CCT **AGT** AGG TTT GTT ACA GAC	§
SARS_Nsp9_Cys73Ser_R	GTC TGT AAC AAA CCT **ACT** AGG TGG TTC CAG TTC	§

§ - Primers used for single-site mutagenesis; restriction enzymes were the same as in case of the wild-type proteins (XhoI/BamHI). The codon corresponding to the mutations is in bold and underlined. * - 229E - HCoV-229E; SARS - SARS-CoV; Nsp - Non Structural Protein; wt - wild-type; Cys69Ala, Cys69Ser, Cys73Ala, and Cys73Ser - are the mutated residues and their position in the protein sequence. S - Forward sense primer; R - Reverse sense primer.

Table 2.3: Proteins and its expression strains

Protein[$]	*E.Coli* expression strain	Vector	Antibiotics[§]
HCoV-229E Nsp7_N-H**	Tuner[TM](DE3)pLacI	PET-15b	Amp and Cm
HCoV-229E Nsp8_C-H*	B834(DE3)	pET-11a	Amp
HCoV-229E Nsp9_N-H	Tuner[TM](DE3)pLacI	PET-15b	Amp and Cm
HCoV-229E Nsp9 Cys69Ala_N-H	Tuner[TM](DE3)pLacI	PET-15b	Amp and Cm
HCoV-229E Nsp9 Cys69Ser_N-H	Tuner[TM](DE3)pLacI	PET-15b	Amp and Cm
SARS-CoV Nsp8_C-H*	B834(DE3)	pET-11a	Amp
SARS-CoV Nsp9_N-H	Tuner[TM](DE3)pLacI	PET-15b	Amp and Cm
SARS-CoV Nsp9 Cys73Ala_N-H	Tuner[TM](DE3)pLacI	PET-15b	Amp and Cm
SARS-CoV Nsp9 Cys73Ser_N-H	Tuner[TM](DE3)pLacI	PET-15b	Amp and Cm

[$] - Last three characters corresponding to the presence of N- or C- terminal His-tag. [§] - Amp - Ampicillin; Cm - chloramphenicol. Expression clones were kindly provided by ***** Prof. John Ziebuhr, Institute of Virology and Immunology, University of Würzburg, Germany (Present address: The Queen's University of Belfast, Northern Ireland); ****** Dr. Ralf Moll, Institute of Biochemistry, University of Lübeck, Germany.

2.2.2 Gene expression

Proteins encoding plasmids were transformed in the corresponding competent cells (Table 2.3). Cultures were grown in TY medium at 37°C until cells reached an optical density of 0.4 at 660 nm. Recombinant gene expression was induced by addition of 1 mM IPTG to the liquid cultures and grown for a further 4 h for Nsp9s and 5 h for Nsp8s at 37°C. The cells were then harvested by centrifugation at 5500 rpm for 30 min at 4°C. The resulting pellets were frozen at -20°C.

2.2.3 Protein purification

The cell pellets were resuspended in the lysis buffer (Table 2.4). Cells were broken by French press after adding one Complete EDTA-free protease inhibitor cocktail tablet (Roche Diagnostics, Mannheim, Germany). To optimize the solubility of overproduced Nsp9, a sparse-matrix screen of buffer composition was applied (Jancarik *et al.,* 2004).

Table 2.4: Protein purification buffer system

Proteins	Wash buffer	Lysis buffer
HCoV-229E Nsp8	50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 7.5 (25°C)	Wash buffer + 1% (v/v) glycerol, 0.01% (w/v) n-octyl-β-glucoside
SARS-CoV Nsp8	50 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 8.0 (25°C)	Wash buffer + 1% (v/v) glycerol, 0.01% (w/v) n-octyl-β-glucoside
HCoV-229E Nsp7/8 (Co-purification)	20 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, pH 7.6 (25°C)	Wash buffer + 10% (v/v) glycerol
HCoV-229E Nsp9 wild-type, Cys69Ala mutant, Cys69Ser mutant.	20 mM Tris-HCl, 300 mM NaCl, 5 mM DTT, 20 mM imidazole, pH 7.5 (25°C)	Wash buffer + 10% (v/v) glycerol
SARS-CoV Nsp9 wild-type, Cys73Ala mutant, Cys73Ser mutant.	20 mM Tris-HCl, 300 mM NaCl, 5 mM DTT, 20 mM imidazole, pH 8.0 (25°C)	Wash buffer + 10% (v/v) glycerol

The sample was ultracentrifuged at 30,000 rpm for 1 h at 4°C. The supernatant was applied to a His Trap HP column (1 ml, GE Healthcare, Freiburg, Germany) with a flow rate of 1 ml/min. After washing with wash buffer (Table 2.4), the protein was eluted with a linear gradient ranging from 20 mM up to 500 mM imidazole. For SARS-CoV Nsp9 and SARS-CoV Nsp8 purification, 8 µl of Benzonase (Merck KGaA, Darmstadt, Germany) was added after breaking the cells for 15 min at 37°C, in order to hydrolyze contaminations by *E. coli* nucleic acids. Protein was blotted and detected with an anti-tetrahistidine antibody (Dianova, Hamburg, Germany) and anti-mouse IgG-alkaline phosphatase conjugate (Sigma, Munich, Germany).

2.2.4 Cleavage of the N-terminal His-tag

Before cleaving the N-terminal His-tag, proteins were dialysed against a buffer containing 20 mM Tris, 200 mM NaCl, pH 7.5 (25°C). SARS-CoV M^{pro} carrying a C-terminal His$_6$ tag was added in a molar ratio of 1:100 and cleavage was allowed to continue for 16 h at 37°C in the presence of 5 mM DTT. The protein solution was applied to a His Trap HP column (1 ml, GE Healthcare, Freiburg, Germany) with a flow rate of 1 ml/min. His-tag cleaved Nsp9 passed through the column whereas uncleaved Nsp9 and M^{pro} remained bound to it.

Table 2.5: Protein storage buffer

Protein	Storage buffer
HCoV-229E Nsp8	15 mM Tris-HCl, 200 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 7.5 (pH 25°C)
SARS-CoV Nsp8	15 mM Tris-HCl, 200 mM NaCl, 1 mM DTT, 0.1 mM EDTA, pH 8.0 (pH 25°C)
HCoV-229E Nsp7/8 (Co-purified)	25 mM Na HEPES, 150 mM NaCl, 5 mM DTT, 1 mM EDTA, pH 7.5 (25°C)
HCoV-229E and SARS-CoV wild-type and mutant Nsp9s	20 mM Tris-HCl, 200 mM NaCl, 5 mM DTT, pH 7.5 (25°C)
TGEV Nsp9	10 mM HEPES, 300 mM NaCl, pH 7.5 (25°C)

2.2.5 Reduction of disulfide bonds with DTT

To determine the minimum concentration of DTT required to reduce disulfide bonds, SDS gel electrophoresis (without β-mercaptoethanol) was used. DTT solutions were freshly prepared and different concentrations of DTT (increasing in 1-mM steps) were added to the HCoV-229E Nsp9 wild-type preparation. After incubation for 30 min at room temperature, 2% loading buffer was added and SDS-PAGE was started.

2.2.6 Characterization of purified proteins

2.2.6.1 Dynamic light-scattering

Measurements were taken using a Spectroscatter 201 (RiNA GmbH, Berlin, Germany) using 10 µl of protein solution in its respective storage buffer (Table 2.1.2). Unless otherwise mentioned, all the measurements were carried out at ~10 mg/ml protein concentration. Results were analysed using the software provided by the manufacturer. Experimental errors were estimated as standard deviations calculated from 10 measurements for each sample.

2.2.6.2 Electrophoretic mobility shift assay

Zone Interference Gel Electrophoresis (ZIGE) is a method to analyze weak protein/nucleic-acid interactions (Abrahams *et al.,* 1988; Matthes *et al.,* 2006). Gel shift assays were performed using a 1% agarose horizontal gel system at pH 8.3 in TBE buffer (20 mM Tris, 50 mM boric acid, 0.1 mM Na EDTA, pH 8.3 (25˚C) adjusted by addition of acetic acid). The protein and DNA samples were mixed with dimethylsulfoxide (DMSO) with a final concentration of 10% (v/v) and a trace of bromophenolblue (BPB) dye. After the electrophoresis, the gel was fixed in 3.5% (w/v) α-sulfosalicilic acid, 10% (w/v) trichloroacetic acid, until the dye turned yellow. At this stage, zones with higher concentrations of ligands such as oligonucleotide can show up as a whitish area. For the detection of protein bands, the gel was washed for 15 min in 15% (v/v) ethanol, 8% (v/v) acetic acid, and stained for 30 min with 0.25% (w/v) Coomassie Brilliant Blue in the same solution, containing an additional 10% (v/v) methanol. The gel was washed in 15% (v/v) ethanol, 8% (v/v) acetic acid, and stored in 10% (v/v) acetic acid. In case of Nsp8/tRNA interactions, 100 µl of samples with increasing tRNA concentration were

applied to the extended zone slot and 10 µl of Nsp8 solution at a final concentration of 10 µM were applied to the small slot. Gels were run in electrophoresis buffer at 200 mA for 2 hrs at 4°C.

The gel was then processed as described (Abrahams *et al.,* 1988). K_D values were calculated using the equation:

$$[(d_{exp} - d_M) / [L] = - (d_{exp} - d_{ML}) / K_D].$$

(d_{exp} – migration distance of protein with varying ligand concentrations, d_{ML} – migration distance of the complex between macromolecule (M) and ligand (L), d_M – migration distance of macromolecule, [L] – ligand concentration in µM). ($d_{exp} - d_M$) / [L] was plotted against d_{exp}, yielding -1/ K_D as the slope.

For the Nsp9/nucleic-acid interaction, the ZIGE was slightly modified because of the high isoelectric point of Nsp9. All the Nsp9s under investigation have a theoretical pI above 9.0; thus, the protein is positively charged at the pH of the running buffer (8.3), whereas the nucleic acids are negatively charged and will move towards the opposite pole during electrophoresis. Pictorial representation of the modified version of ZIGE was shown at different time points during the run (Fig. 2.2.1). The oligonucleotides used for this experiment are tabulated (Table 2.6).

Table 2.6: Oligonucleotide sequences used for the electrophoretic-mobility-shift assay

Name	Oligonucleotide (5´-3´ direction)
6-mer	GATGCT
13-mer	TTAGCTATGGTGC
18-mer	TTAGCCTCTAGATTTCAT
24-mer	GATAGCCGGGCACGGTTGCTGTTT
30-mer	CCCTTGGGATCCCTACAATGAAGAGAATAA
36-mer	GGGAGTCTCGAGGTTAAATTGCAGAATAATGAACTG
40-mer	TAATACGACTCACTATAGGAGATAGTCGTGGGTTCCCTTT
45-mer	TTATTAGTGATGGTGATGGTGATGGCTCTGCTTGTCGGTGAAGTA
50-mer	TTATTAGTGATGGTGATGGTGATGGCTCTGCTTGTCGGTGAAGTATTTTT
55-mer	GATTTTTATTAGTGATGGTGATGGTGATGGCTCTGCTTGTCGGTGAAGTAT TTTT
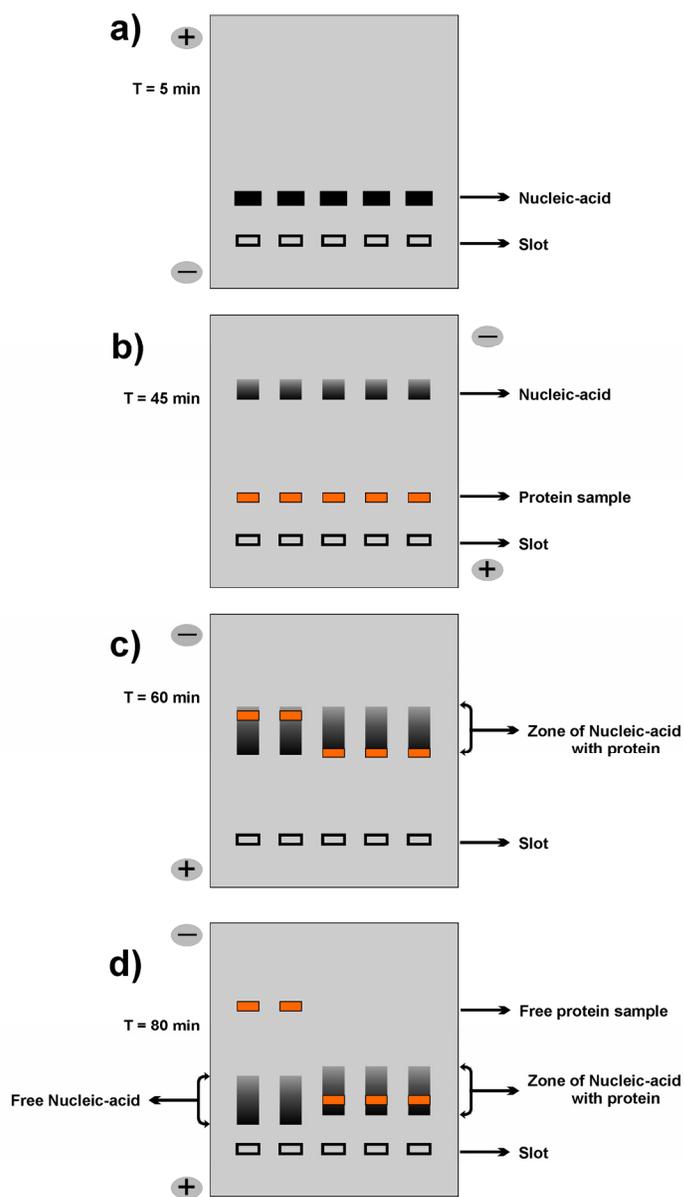
Fig. 2.2.1 Modified version of Zone-Interference Gel Electophoresis
(a) 100 µM of oligonucleotide was loaded into the sample slots. Nucleic acid was moving towards the positive pole. (b) After a 40-min run, the gel was stopped and 50 µM of protein sample was loaded into the respective slots. Directly after the protein samples had been loaded, the poles of the electrodes were interchanged. (c) Due to the nucleic-acid migration in the opposite direction, the samples were diffused into a zone like formation. With this design of the experiment, the protein and the nucleic acid run in opposite directions and meet in the middle of the agarose gel. (d) If there is interaction between the two, the protein will move in the same direction as the nucleic acid, whereas if there is no interaction, the protein will just pass the nucleic acid. T – Time points; approximate position of the samples at different time points.

2.2.6.3 Fluorescence measurements

Fluorescence measurements were performed with a Cary Eclipse fluorescence spectrometer (Varian). The binding of 1,8-anilino-naphtalene sulfonate (1,8-ANS)-Mg^{2+} to the protein was monitored in 15 mM Tris, 200 mM NaCl, pH 8.0, at final concentrations of 5 µM Nsp8 protein and 10 µM 1,8-ANS-Mg^{2+}. Reference fluorescence

spectra for 10 µM 1,8-ANS-Mg^{2+} were recorded with 10%, 50% and 90% (v/v) methanol diluted with the buffer mentioned above, in order to compare the blueshift and the quantum-yield increase of the 1,8-ANS fluorescence in the presence of growing concentrations of an organic solvent less polar than water.

2.2.6.4 Circular dichroism

CD spectra were recorded using a Jasco J-715 spectropolarimeter equipped with a temperature-controlled quartz cell of 1 cm path length. Spectra were measured between 190 and 250 nm. The molar ellipticity was calculated on the basis of the mean residue mass (Andrade *et al.,* 1993) of His-tagged SARS-CoV Nsp8 (205 residues, 22.8 kDa) and HCoV-229E Nsp8 (202 residues, 22.6 kDa). Thermal transition curves were obtained at a fixed wavelength of 220 nm by increasing the temperature at a rate of 30°C/h using a Gilford temperature control unit. The sample concentration was 200 µg/ml. Spectra were recorded in a sample buffer containing 50 mM NaH_2PO_4, 50 mM NaCl, pH 8.0.

2.2.6.5 One dimensional-Nuclear Magnetic Resonance

1D-NMR spectra of protein samples were acquired in 600 µl of buffer containing 20 mM NaH_2PO_4, 200 mM NaCl, pH 7.2 and 5% d_6-DMSO at 5°C with 512 scans and 32768 complex data points and with a spectral width of 7000 Hz on a Bruker DRX500 spectrometer. The spectrometer is equipped with a 5 mm triple resonance probehead (HCN) with z-axis gradients. The water signal was suppressed using the DPFGSE (double-pulsed field gradient spin echo) method and binomal w5 pulses (Liu *et al.,* 1998), the deuterons of the DMSO were used as the lock substance. 1D-NMR spectra measurement was performed by Dr. Andrew Benie, Institute of Chemistry, Lübeck, Germany.

2.2.6.6 Crosslinking

Chemical cross-linking experiments were performed for wild-type HCoV-229E Nsp9, the Cys69Ala mutant, TGEV Nsp9, and SARS-CoV Nsp9 with glutaraldehyde. Different concentrations of Nsp9 were incubated in 20 µl final volume of cross-linking buffer (50 mM NaH_2PO_4 (pH 7.6), 50 mM NaCl) for 5 min at 20°C. After addition of 0.01% (v/v)

glutaraldehyde, the reaction tubes were incubated for 5 min at 20˚C, before the reactions were stopped by the addition of 100 mM Tris. In the case of the Nsp9-DNA complex, oligonucleotides and proteins were incubated for 2 hr at 20°C prior to glutaraldehyde crosslinking. The samples were analyzed by SDS-PAGE after adding 10 µl of 2% sample-loading buffer and incubation at 70˚C for 10 min.

2.2.6.7 Surface plasmon resonance analysis

The streptavidin-coated sensor chip (SA chip, Biacore, GE Healthcare, Freiburg, Germany) was preconditioned before immobilization, by treating it three times with 1 M NaCl in 50 mM NaOH and 0.05% SDS for 60 s at a flow rate of 5 µl/min. 5′-Biotinylated 50-mer oligonucleotide (5′-TTATTAGTGATGGTGATGGTGATGGCTCTGCTT GTCGGTGAAGTATTTTT-3′) was purchased from MWG-Biotech AG, Ebersberg, Germany. For immobilization, 2 nM of the oligonucleotide was injected in the manual injection mode until the desired amount of DNA on the surface was reached. To reduce non-specific binding of protein to the surface of the SA chip, the chip was activated with 50 mM N-hydroxysuccinimide and 200 mM 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (NHS/EDC) and deactivated with 1 M ethanolamine for 5 times. Saturation of residual free streptavidin-binding sites was achieved with a 0.4% solution of biotin. Unless specifically mentioned, all the injections were done at a flow rate of 10 µl/min in the running buffer consisting of 10 mM HEPES pH 7.4, 150 mM NaCl (HBS-N). For determining the K_D value, the oligonucleotide was immobilized up to 300 RU. (For the set of experiments analyzing the interactions between oligonucleotide and the reduced or oxidized forms of HCoV-229E Nsp9, this value was 88 RU). HCoV-229E Nsp9 wild-type and SARS-CoV Nsp9 were injected into the flow cell for 22 and 30 min, respectively. The concentration used for determining the K_D value ranged from 2 µM to 85 µM. K_D values were calculated using a steady-state equilibrium analysis in a 1:1 Langmuir single-state binding model or a two-state binding model.

Single-state binding model:

$$R_{eq} = \frac{\left(K_A \times C \times R_{max}\right)}{\left(K_A \times C \times n + 1\right)}$$

Two-state binding model:

$$R_{eq} = \frac{\left(K_{A1} \times C \times R_{1max}\right)}{\left(K_{A1} \times C \times n + 1\right)} + \frac{\left(K_{A2} \times C \times R_{2max}\right)}{\left(K_{A2} \times C \times n + 1\right)}$$

where K_A is the equilibrium association constant, C is the concentration of injected protein, R_{max} is the maximum analyte-binding capacity and n is the steric interference factor.

Regeneration was achieved using 0.05% SDS for 60 s. The efficiency of the regenerated SA chip was tested with multiple injections after the regeneration. The data were analyzed with the BIAevaluation software, version 3.2 (GE Healthcare, Freiburg, Germany).

2.2.7 Crystallization

Crystallization screening was performed using a Phoenix robot (Dunn Labortechnik, Thelenberg, Germany) employing the sitting-drop vapor diffusion method. Initial hits were optimized by manually setting up 2-µl hanging drops consisting of 1 µl of protein and 1 µl of reservoir solution.

HCoV-229E Nsp7/8: The Nsp7 and Nsp8 proteins were co-purified and concentrated to 11 mg/ml for crystallization screening. Most of the drops stayed clear even after several days of storage at 10°C; only few crystalline precipitates were obtained after one week. Eventually, few conditions led to spherulites and phase separation. The crystallization condition contains a special phosphate citrate buffer, which was prepared by mixing equal concentrations of sodium dihydrogen phosphate and citric acid. Further optimization of these conditions was not successful in obtaining any three-dimensional crystals.

HCoV-229E Nsp9 wild-type and Cys69Ala mutant: Both HCoV-229E wild-type and mutant protein crystals appeared within two days at 10˚C and a protein concentration of 6

- 8 mg/ml in the storage buffer (Table 2.5). The most useful precipitants were ammonium sulfate at pH 4.0 - 4.5 for wild-type Nsp9, and polyethylene glycol monomethyl ether 2000 (pH 4.6) for the mutant.

SARS-CoV Nsp9: The SARS-CoV Nsp9 N-terminal His-tag version crystallized readily with the Sigma basic and extension screening kits within overnight storage at 10°C and a protein concentration of 11 mg/ml in the storage buffer (Table 2.5). The most useful precipitant was ammonium sulfate at pH 4.6. Only minor optimization was needed to obtain good three-dimensional crystals. Unlike the N-terminal His-tag version, the protein with authentic chain termini version did not yield any crystals.

TGEV Nsp9: Several crystalline substances were obtained within one day at 10°C for TGEV Nsp9 using the Sigma basic and extension screening kits and a protein concentration of 10 mg/ml in the storage buffer (Table 2.5). The most useful precipitants were PEG 4000, 8000, and many alcohols at a wide range of pH (5.6 - 8.5). Attempts at reproducing the crystals in similar crystallization conditions often led to thin fragile inter-grown crystals. Crucial factors in obtaining the thick three-dimensional needles were to increase the protein concentration from 10 to 12 mg/ml and to use only the main fractions eluted from the gel-filtration column.

All the optimized crystallization conditions are given in the results section.

2.2.8 Data acquisition and structure determination

The diffraction quality of crystals was tested and cryocondtions were optimized using an in-house X-ray generator source. Well-diffracting quality crystals were then flash-cooled for synchrotron diffraction experiments. X-ray diffraction data were collected at 100 K using synchrotron radiation of wavelength 0.8075 Å, provided by the University of Hamburg - University of Lübeck - EMBL beamline X13 at Deutsches Elektronen-Synchrotron (DESY), Hamburg. TGEV Nsp9 crystals alone were collected at 100 K using synchrotron radiation of wavelength 1.2548 Å, provided by the Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY), Berlin. The

cryoprotectant was 25% glycerol for the HCoV-229E Nsp9 wild-type protein and SARS-CoV Nsp9, and 25% polyethylene glycol 400 for the HCoV-229E Nsp9 Cys69Ala mutant. TGEV Nsp9 crystal was cryoprotected using 30% of ethylene glycol; annealing protocol by Harp *et al.,* (1998) was followed to obtain higher resolution data set. Data were processed with DENZO (Otwinowski & Minor, 1997) and scaled by using SCALA (CCP4, 1994). The structures were solved by using the molecular replacement program Phaser (McCoy *et al.,* 2005). HCoV-229E wild-type and Cys69Ala mutant Nsp9s were solved with a monomer of SARS coronavirus Nsp9 (1QZ8) as the search model. The initial solutions had R factors of 51.0% (wild-type) and 50.8% (mutant). TGEV Nsp9 was solved with a monomer of HCoV-229E Nsp9 Cys69Ala mutant (2j98) as the search model. Manual reconstructions of search model were needed for the successful molecular replacement solution. The initial solution had R factor of 48.7%. Data collection statistics are tabulated in Table 7.1 – 7.4 (Appendix)

2.2.9 Model refinement

After initial rigid-body refinements, models were built into electron density by using COOT (Emsley & Cowtan, 2004) and refined by REFMAC5 (Murshudov *et al.,* 1997). The final model for HCoV-229E Nsp9 wild-type includes 70 water molecules, two sulfate-ion sites, and one MPD molecule. The structural model for the HCoV-229E Cys69Ala mutant comprises 90 water molecules and one DTT molecule. The final steps of Cys69Ala mutant structure refinement incorporated TLS refinement with residues 4 – 109 of monomer A as one group and residues 6 – 106 of monomer B as a second group. The TGEV Nsp9 structural model includes 28 water molecules, three chloride-ion sites, and one ethylene glycol molecule. Four TLS groups (residues 2-4, 15-57, 58-83, and 84-111) were defined and included for the final step of refinement. TLS groups were identified using *TLSMD* web server (Painter & Merritt, 2006). The overall geometric quality of the models was checked using PROCHECK (Laskowski *et al.,* 1993). Final refinement statistics are tabulated in Table 7.1 – 7.3 (Appendix). The surface area buried upon dimerization was calculated using AREAIMOL (Lee & Richards, 1971). Structure superimposition and calculation of r.m.s. deviations were carried out using ALIGN (Cohen, 1997). Figures were created using Pymol (DeLano, 2002).

3. Results

3.1 HCoV-229E and SARS-CoV Nsp8

3.1.1 Protein production

SARS-CoV and HCoV-229E Nsp8 were successfully produced under heterologous conditions, at about 10-20 mg of protein per litre expression culture. Both proteins exhibited apparent molecular masses of about 23 kDa under denaturing conditions in SDS-PAGE (Fig. 3.1.1, (a1) and (b1)).

Owing to the attached His-tag, the Nsp8s were readily detected in immunoblots using anti-histidine antibodies (Fig. 3.1.1, (a2) and (b2)). SARS-CoV Nsp8 was rather instable. The protein was cleaved to several peptide fragments smaller than 12 kDa (as judged by SDS-PAGE) within three days at room temperature. It was stable at 4°C or in the presence of protease inhibitors (not shown). Limited tryptic cleavage experiments did not reveal any stable domains within Nsp8.
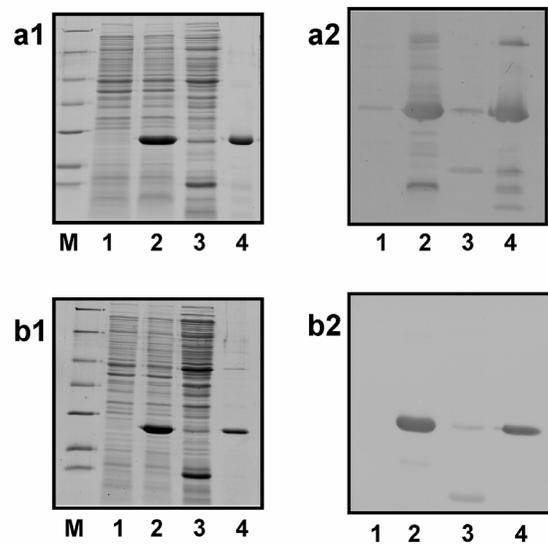
Fig. 3.1.1 Overproduction of His-tagged SARS-CoV and HCoV-229E Nsp8
(a1 and b1): SDS-PAGE of SARS-CoV and HCoV-229E Nsp8; Lane M: marker proteins with apparent molecular masses of 116 (most upper band), 66.2, 45, 35, 25, 18.4, 14.4 kDa (lowest band). Lane 1: *E. coli* lysate prior to induction, lane 2: lysate after induction, lane 3: flow-through of Ni-NTA chromatography, lane 4: Nsp8 eluted with an imidazole gradient. (a2 and b2): Western blots of corresponding fractions using anti-Histidine antibodies.

3.1.2 Spectroscopic investigation of Nsp8s

3.1.2.1 Circular dichroism

From Circular Dichroism (CD) spectra of SARS-CoV and HCoV-229E Nsp8 (Fig. 3.1.2 (a)), the secondary structure was deduced to consist, respectively, of 26% and 29% α-helix, 14% and 12% β-sheet, and 60% and 59% non-repetitive secondary structure (e.g. loops). Thus, the content of secondary structure elements was almost identical for the two proteins, indicating that they very likely have the same fold. The CD results were not compatible with the secondary structure prediction (PHD program), which proposed an α-helix content of 50% as well as 30% "loop". Melting points determined by CD spectroscopy were 42°C for HCoV-229E Nsp8 and 49°C for SARS-CoV Nsp8 (Fig. 3.1.2 (b)).
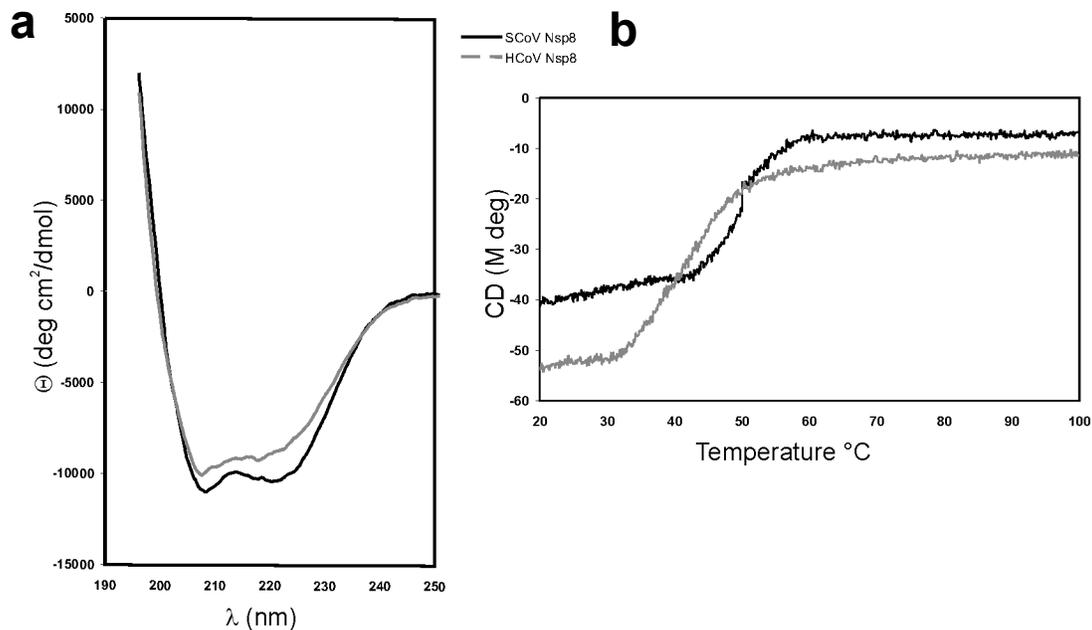
Fig. 3.1.2 Circular Dichroism spectra of SARS-CoV and HCoV-229E Nsp8
(a) θ molar ellipticity. (b) Melting curve of Nsp8s; measured at 220 nm using CD.

3.1.2.2 One dimensional-Nuclear Magnetic Resonance

1D-NMR spectra were recorded for SARS-CoV Nsp8 at 500 MHz (Fig. 3.1.3). In the amide proton region (7-9 ppm), as well as in the aliphatic region (1-4 ppm), the spectra contained relatively few sharp peaks, and those that were present were of significantly

higher intensity than the rest of the signals (Fig. 3.1.3 (a)). In particular, the lack of signals beyond 8.5 ppm (8.5-11 ppm) and below 0 ppm indicated that the protein under study was either aggregated or partially disordered (Rehm *et al.,* 2002). As a positive control experiment, 1D-NMR spectra of SARS-CoV Nsp9, a protein that has been crystallized (Egloff *et al.,* 2004; Sutton *et al.,* 2004) and has a folded state, were recorded. Fig. 3.1.3 (b) shows that the two signature regions at > 8.5 ppm and < 0 ppm contain many more signals and the difference in the intensity between the most intense signals and the majority is significantly smaller. This indicates that this protein is present in a predominantly folded form in solution.
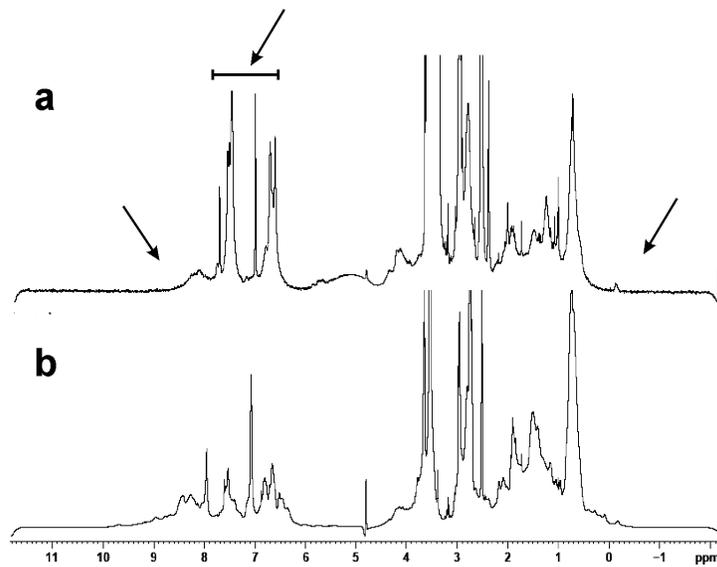
Fig. 3.1.3 1D-NMR spectra recorded at 12.1 T of SARS-CoV Nsp8
(a) A representative spectrum of the protein Nsp8, which is thought to be partially unfolded, and an example of a folded protein (b), Nsp9, showing more signals above 8.5 ppm and below 0 ppm as well as improved signal dispersion and a more uniform intensity profile.

3.1.2.3 Fluorescence quenching experiment

In the C-terminal region of SARS-CoV and HCoV-229E Nsp8s, there are two absolutely conserved tryptophan residues in positions 154 and 182. The environments of the tryptophans were assessed by fluorescence spectroscopy. Both Nsp8s exhibited an emission maximum at 332 nm with an excitation wavelength of 280 nm. Using the same

protein concentration and operating parameters, the quantum yield of SARS-CoV Nsp8 was twice as high as for HCoV-229E Nsp8. Therefore, one or both the tryptophans of the SARS-CoV protein were protected from water quenching and located in a properly ordered, hydrophobic environment.

Remarkably, 1,8-anilino-naphthalene-sulfonate (1,8-ANS), a fluorescence dye that is used as a probe binding non-covalently to hydrophobic pockets within proteins, can bind to both the proteins. Upon interacting, the emission maximum of 1,8-ANS undergoes a blueshift in the fluorescence spectrum and a strong increase of the quantum yield (Creighton, 1992). Fig. 3.1.4 (a) and (b) indicate that both Nsp8 proteins likely bind 1,8-ANS at a similar non-polar site. Emission maxima of 482 nm (HCoV-229E) and 484 nm (SARS-CoV) were observed using an excitation wavelength of 378 nm, with
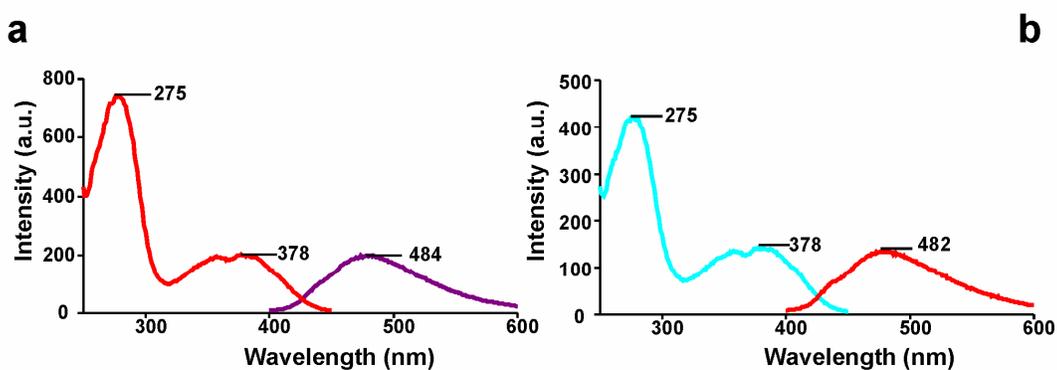
Fig. 3.1.4 Fluorescence spectra of 1,8-ANS in the presence of SARS-CoV (a) and HCoV-229E (b) Nsp8
Emission spectra (right side of Fig. a & b) were recorded with excitation wavelength of 378 nm. Excitation spectra (left side of Fig. a & b) were obtained with emission wavelength of 480 nm. Signals were amplified with 650 V and the filter slit width was set to 10 nm. Spectra were obtained at room temperature. The Y-axis scale is different between (a) and (b).

absorption maxima of 275 and 378 nm (wavelength corresponding to the absorption maxima of the chromophore ANS). The quantum yield of the excited fluorophore was higher for SARS-CoV Nsp8, probably due to the decreased quenching. The observed emission maximum of 1,8-ANS upon protein binding was the same as in a 1,8-ANS solution in 90% (v/v) methanol in the absence of the protein. Increasing the polarity of

the solvent by decreasing the methanol concentration to zero led to a red shift of the 1,8-ANS fluorescence up to 530 nm with an concomitant decrease in quantum yield almost to zero (not shown). Therefore, the 1,8-ANS binding site of Nsp8 appeared to be strongly hydrophobic.

3.1.3 Oligomeric state of Nsp8s

SARS-CoV and HCoV-229E Nsp8s tend to aggregate as demonstrated by native PAGE. The native PAGE profile contained three distinct protein bands indicating different oligomeric states for SARS-CoV and HCoV-229E Nsp8 (Fig 3.1.5). Furthermore, up to six SARS-CoV Nsp8 bands were visualized with isoelectric points from 7.0 to 8.1 in the isoelectric focussing gel under native conditions (not shown). This heterogeneity may correspond to different oligomers of Nsp8. Hence, the hydrodynamic particle sizes of the oligomeric Nsp8 species were assessed using Dynamic Light-Scattering (DLS) (Fig. 3.1.6).
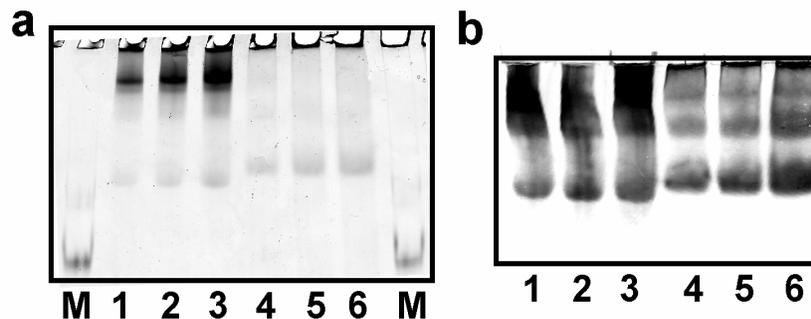
Fig. 3.1.5 Native PAGE (a) and Western Blot (b) of SARS and HCoV-229E Nsp8
M- BSA as a Marker protein. Lane 1, 2 and 3 correspond to 20, 15 and 10 μg of SARS-CoV Nsp8, Lane 4, 5 and 6 correspond to 20, 15 and 10 μg of HCoV Nsp8.

The calculated hydrodynamic radius for a monomeric protein molecule of spherical shape with the molecular mass of Nsp8 (23 kDa) would be 2.07 nm. As evident from Fig. 3.1.6 (a) and (c), both Nsp8 proteins exhibited a high degree of oligomerization with a hydrodynamic radius of 300 to 500 nm in a polydisperse state, when measured in Tris buffer (see Fig. 3.1.6 legend). Using DTT and EDTA as combined additives, the observed species showed broad, but more monodisperse distributions at lower hydrodynamic radii, centred at 13 nm (SARS-CoV) and 20 nm (HCoV-229E) (Fig. 3.1.6

(b) and (d)). Nevertheless, assuming spherical shape, several monomers must be present in these aggregates. The hydrodynamic radius could not be shifted to lower values by decreasing the protein concentration from 6.0 to 0.5 mg/ml. Therefore, substantial amount of low-molecular weight oligomers might be present only below 0.5 mg/ml (22 µM). In size-exclusion chromatography, the predominant oligomerization state of SARS-CoV and HCoV-229E Nsp8s was a dimer at concentration of 0.3 mg/ml (13 µM) (not shown).
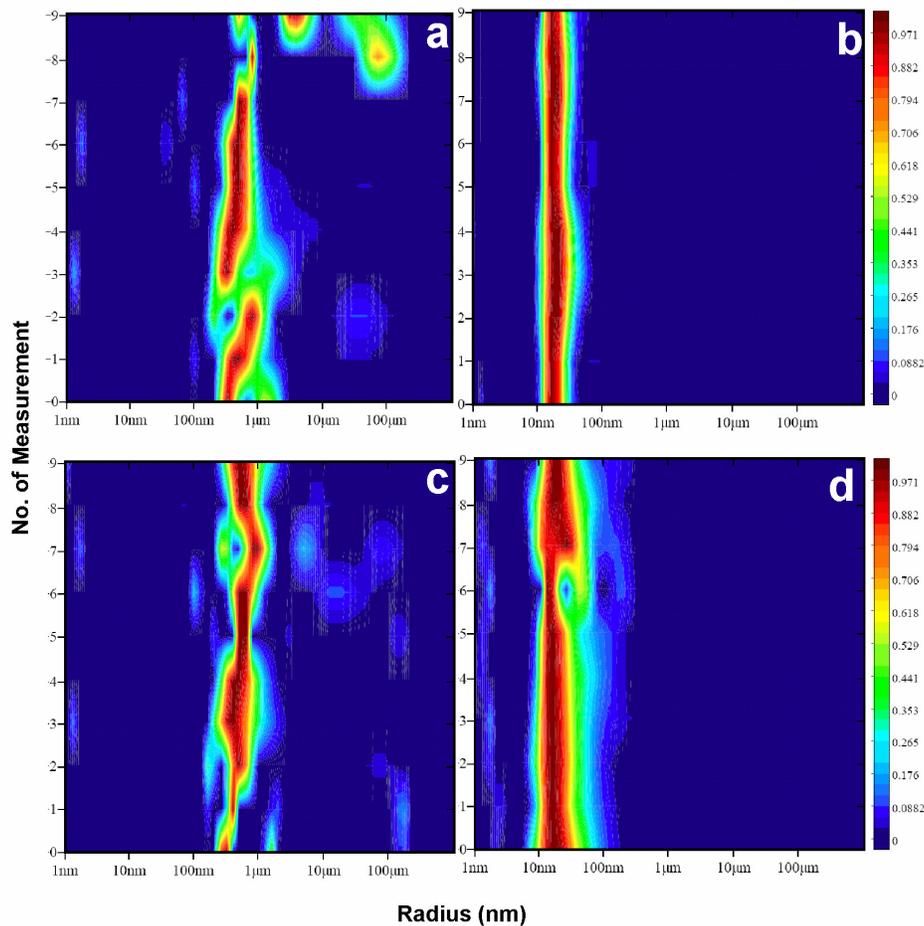
Fig. 3.1.6 Dynamic Light-Scattering of SARS-CoV and HCoV-229E Nsp8 proteins
SARS-CoV Nsp8 before (a) and after (b) adding additives; HCoV-229E Nsp8 before (c) and after (d) adding additives. SARS-CoV Nsp8 was dissolved at 5.5 mg/ml in 15 mM Tris-HCl, 200 mM NaCl, pH 8.0 (a), complemented with 1 mM DTT and 0.1 mM EDTA (final concentrations) (b). HCoV-229E Nsp8 was dissolved at 9.84 mg/ml in 15 mM Tris-HCl, 200 mM NaCl, pH 7.5 (c), complemented with 2 mM DTT and 0.1 mM EDTA (d).

3.1.4 HCoV-229E Nsp8 interaction partners

3.1.4.1 Dynamic Light-Scattering

HCoV-229E Nsp8 oligomerized to higher molecular-mass species as seen by DLS. The Nsp8 hydrodynamic radius appeared to decrease in the presence of one or more interaction partners. HCoV-229E Nsp8 and Nsp7 separately showed broad monodispersed hydrodynamic radii centered at 20 and 15.6 nm, respectively (Fig. 3.1.7 (a) and (b)). By mixing HCoV-229E Nsp7 and Nsp8 in equimolar amounts, the hydrodynamic radius decreased to 11 nm (Fig. 3.1.7 (c)). Moreover, the 11-nm radius of the Nsp7-8 protein complex were further reduced to a monodisperse hydrodynamic radius centered at 7 nm by the addition of an equimolar amount of Nsp9 indicating that Nsp7-8-9 forms a tight complex (Fig. 3.1.7 (d)).
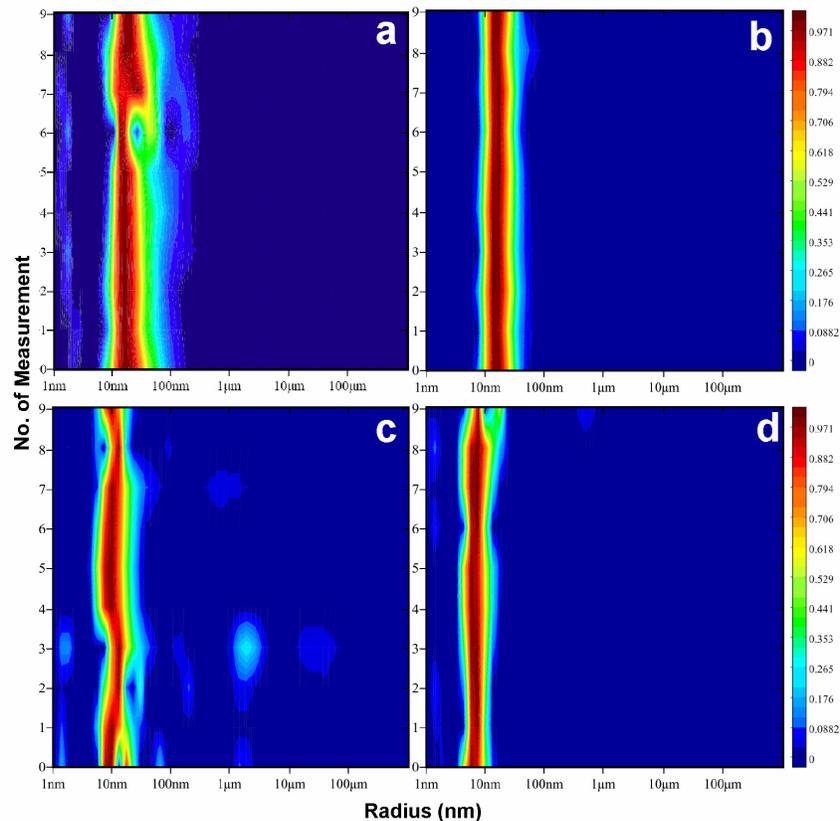
Fig. 3.1.7 HCoV-229E Nsp8 interaction with Nsp7 and Nsp9 analysed using Dynamic Light-Scattering
All the measurements were done at 25°C in a buffer consisting of 15 mM Tris-HCl, 200 mM NaCl, pH 7.5 (25°C), 2 mM DTT and 0.1 mM EDTA. (a) HCoV-229E Nsp8 alone (b) HCoV-229E Nsp7 alone (c) equimolar ratio of Nsp7 and Nsp8 (d) equimolar ratio of Nsp7, Nsp8 and Nsp9.

3.1.4.2 Co-purification

The solubility of HCoV-229E Nsp8 was very low (Fig 3.1.8 (a)), whereas that of HCoV-229E Nsp7 was modest (Fig 3.1.8 (b)). Equal amounts of HCoV-229E Nsp7 and Nsp8 pellets were combined and resuspended in the lysis buffer (for details see Materials and Methods) and the cells were broken by French Press. The N-terminally His-tagged HCoV-229E Nsp7 and Nsp8 were then co-purified using a Ni-NTA column followed by gel-filtration. The combination of pellets during the purification led to a higher solubility of both (Fig 3.1.8 (c)), indicating complex formation between the two proteins. In the gel-filtration profile, Nsp7 and Nsp8 eluted together with a molecular mass of 41 kDa. This corresponds to a ratio of ~ 1:1 or 2:1 of Nsp7 (monomer MW 12 kDa) and of Nsp8 (monomer MW 23 kDa) respectively (Figure not shown).
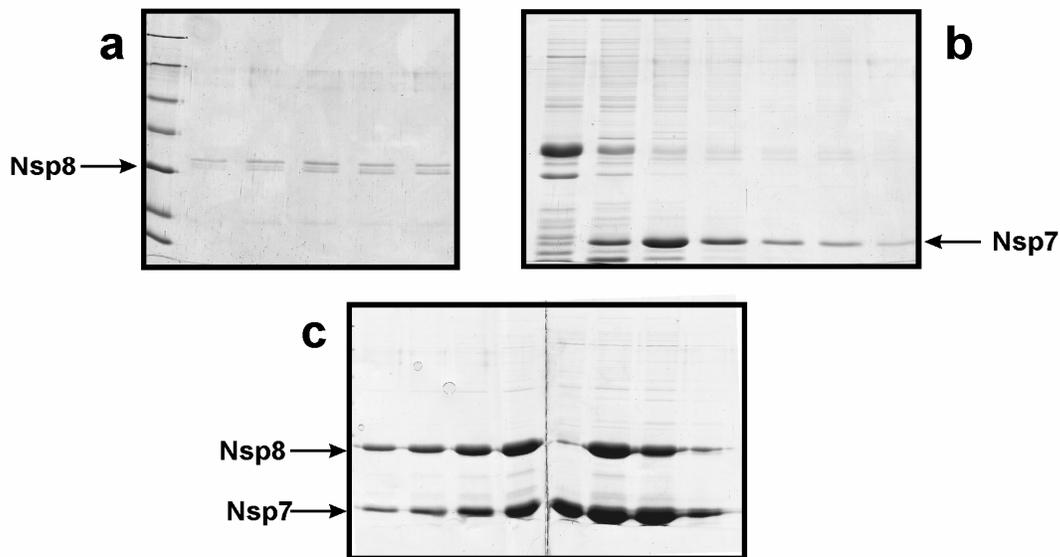
Fig. 3.1.8 Co-purification of HCoV-229E Nsp7 and Nsp8 proteins
HCoV-229E Nsp7 and Nsp8 were purified by Ni-NTA and gel-filtration chromatography; individual protein elution profiles from the Ni-NTA column with increasing imidazole concentration are shown for (a) Nsp8 and (b) Nsp7. (c) Elution profile of HCoV-229E Nsp7 and Nsp8 from the gel-filtration chromatography; solubility of both the proteins increased dramatically upon co-purification.

3.1.5 Nsp8 – nucleic acid interaction

Nsp8 interaction with nucleic acid was demonstrated by Zone-Interference Gel Electrophoresis (ZIGE), a technique used to analyze weak protein-nucleic acid

complexes in the μM-affinity range (Abrahams *et al.,* 1988). Binding of the negatively charged tRNA to the two Nsp8 proteins led to a considerable increase of the electrophoretic mobility of the RNA-protein complexes compared to the protein alone. The complex was formed in a concentration-dependent manner (Fig. 3.1.9 (a)). tRNA was used as a probe ligand, because of the presence of both single and double stranded regions. *E. coli* elongation factor EF-Tu was used as a negative control and dissociated rapidly from deacetylated tRNA in the gel shift assay as reported by
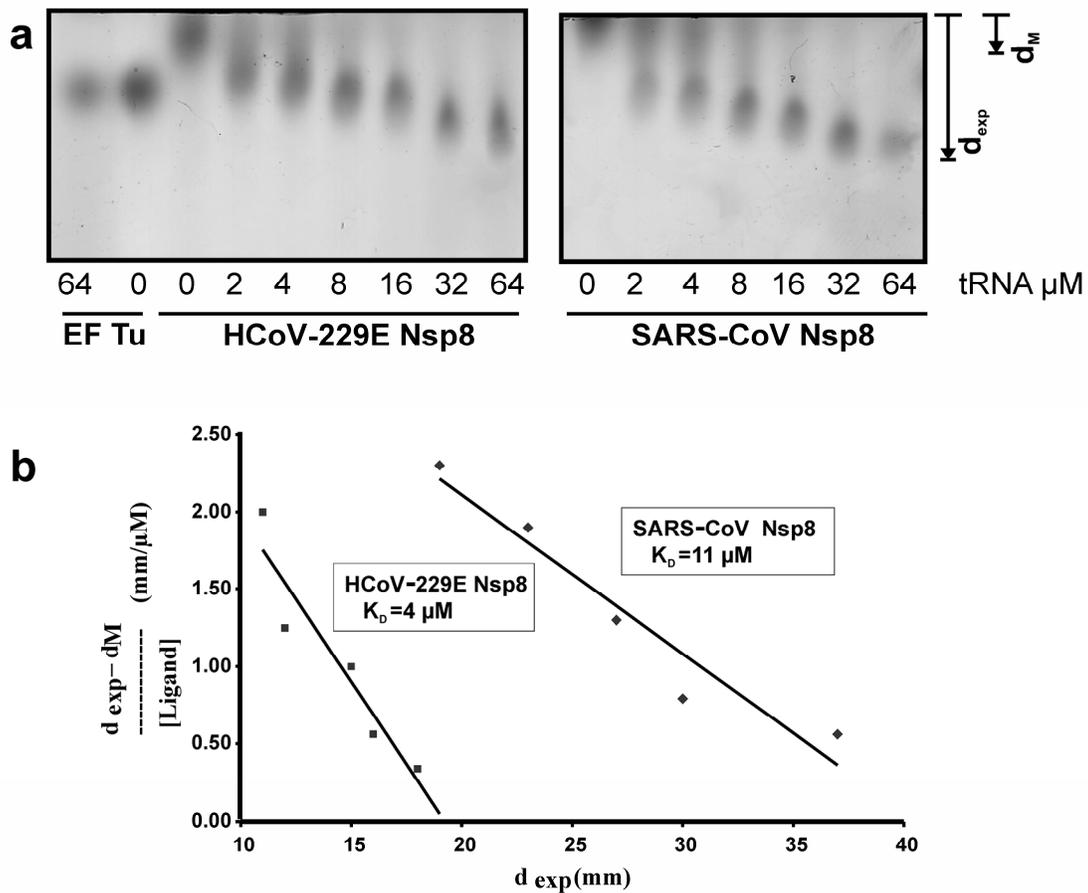
Fig. 3.1.9 Zone-Interference Gel Electrophoresis of HCoV-229E and SARS-CoV Nsp8s with tRNA
(a) Different concentrations of tRNA were used to probe the nucleic acid interaction with Nsp8 proteins; EF-Tu was used as a negative control. (b) K_D values were determined from the experimental data (Fig. 3.1.9 (a)) using the equation $[(d_{exp} - d_M) / [L] = - (d_{exp} - d_{ML}) / K_D]$ (see Materials and Methods for detail) and plotted for the HCoV-229E- and SARS-CoV Nsp8-tRNA complexes (b).

Abrahams *et al.* (1988), resulting in no net increase of electrophoretic mobility. K_D values of the SARS-CoV Nsp8/tRNA and HCoV-229E Nsp8/tRNA complexes were determined as 11 µM and 4 µM, respectively (Fig. 3.1.9 (b)). Since the dissociation constants for the two proteins were in the same range, a comparable mode of tRNA binding can be assumed.

To identify the sequence specificity of Nsp8 in nucleic-acid binding, several polynucleotides were used in ZIGE with HCoV-229E and SARS-CoV Nsp8. Both proteins bound to poly(U)-poly(A), poly(U), poly(C) and, to lesser extent, poly(G). Poly (A) was a poor substrate where almost no shift was obtained (Fig. 3.1.10).
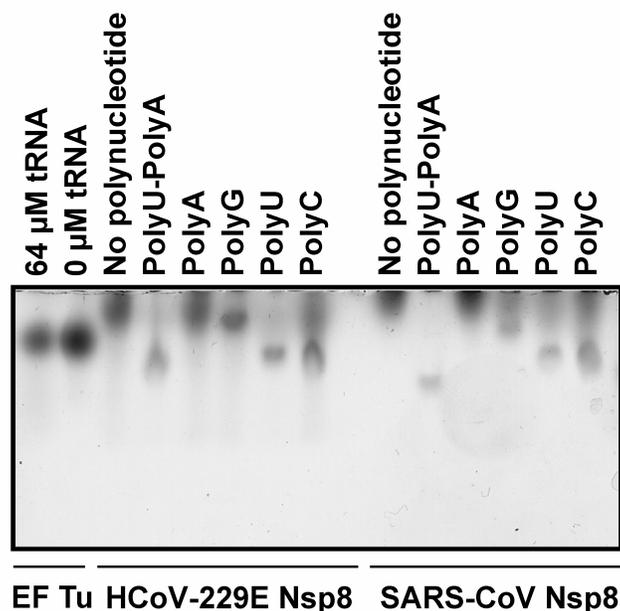
Fig. 3.1.10 Nsp8 interaction with various polynucleotides demonstrated using Zone-Interference Gel Electrophoresis
To identify the sequence specificity of Nsp8 proteins, poly(U)-poly(A) (156 µM), poly(A) (35 µM), poly(G) (35 µM), poly(U) (35 µM), and poly(C) (58 µM) were probed for interaction with HCoV-229E and SARS-CoV Nsp8 proteins. Except poly(A), all the polynucleotides showed interaction with Nsp8 to various extents. EF-Tu was used as a negative control.

3.1.6 Crystallization trials

Nsp8 from HCoV-229E and SARS-CoV were concentrated to 10 mg/ml. Initial crystallization trials were made using commercial crystallization screening kits. No

crystal was obtained. 1D-NMR, CD spectroscopy and DLS experiments demonstrated that at least part of the Nsp8 was not properly folded and that was interacting with several other Nsps. Therefore, one of the strong interaction partners, HCoV-229E Nsp7 was co-purified with HCoV-229E Nsp8 as explained in section 3.1.4.2 and the complex was concentrated to 11 mg/ml. Initial crystallization trials were made with HCoV-229E Nsp7 and Nsp8 using commercial crystallization screening kits. Unfortunately, only microcrystals and spherulites were obtained under the optimized conditions given in the legend Fig. 3.1.11. Crystallization conditions are yet to be optimized to produce well-diffracting crystals.
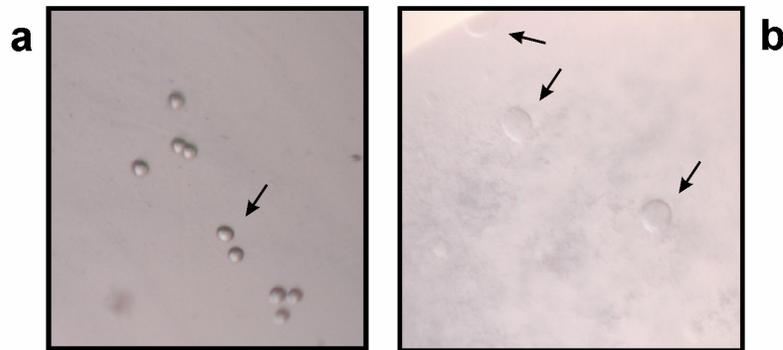
Fig. 3.1.11 Crystallization trials with the HCoV-229E Nsp7-8 complex
1 µl of the HCoV-229E Nsp7-8 complex was mixed with 1 µl precipitant using the sitting drop vapor diffusion method. After one week of storage at 20°C, (a) spheres of the size of 0.2 µm in diameter appeared in 1.0 M Na-acetate, 0.05 M Cd-sulfate, 0.1 M HEPES Na-salt (pH 7.5). (b) disk like "spherulites" were seen after two days in 1.6 M tri-sodium citrate (pH 6.5).

3.2 HCoV-229E and SARS-CoV Nsp9

3.2.1 Protein production

Wild-type HCoV-229E Nsp9 and its Cys69Ala and Cys69Ser mutants were cloned with a His-tag connected to the N-terminus of the protein *via* the linker sequence VKLQ. The latter peptide sequence corresponds to the C-terminus of SARS-CoV Nsp8 (as well as HCoV-229E Nsp8) and therefore introduces a cleavage site for the main proteinase (M^{pro}) of SARS-CoV. After purification of the His-tagged protein using Ni-NTA chromatography, cleavage with the M^{pro} yielded Nsp9 with an authentic N-terminus. In a similar manner to HCoV-229E Nsp9, wild-type SARS-CoV Nsp9 and its Cys73Ala and Cys73Ser mutants were cloned, produced and purified.

In total, six variants of Nsp9s were successfully produced under heterologous conditions, at about 15 mg of protein per liter expression culture. All the proteins exhibited apparent molecular masses of about 12 kDa under denaturing conditions in SDS-PAGE (Fig. 3.2.1). The apparent impurity present in lanes 1-3 of Fig. 3.2.1 at a molecular mass of around 26 kDa might correspond to the dimeric form of HCoV-229E Nsp9.
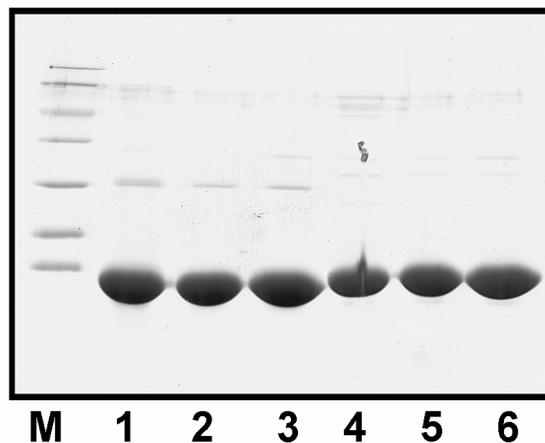
Fig. 3.2.1 Overproduced and purified Nsp9s
SDS-PAGE of Nsp9s from HCoV-229E and SARS-CoV; Lane M: Marker proteins with apparent molecular masses of 116 (most upper band), 66.2, 45, 35, 25, 18.4, 14.4 kDa (lowest band). Lane 1: HCoV-229E Nsp9 wild-type, Lane 2: HCoV-229E Nsp9 Cys69Ala mutant, Lane 3: HCoV-229E Nsp9 Cys69Ser mutant, Lane 4: SARS-CoV Nsp9 wild-type, Lane 5: SARS-CoV Nsp9 Cys73Ala mutant, Lane 6: SARS-CoV Nsp9 Cys73Ser mutant. Each lane contained 20 µg of Nsp9.

3.2.2 Crystallization of recombinant coronavirus Nsp9

All initial crystallization trails were carried out using a Phoenix robot (Dunn Labortechnik, Thelenberg, Germany) employing the sitting-drop vapor diffusion method. Initial hits were optimized by manually setting up 2-µl hanging drops consisting of 1 µl of protein and 1 µl of reservoir solution.

Wild-type HCoV-229E Nsp9: The protein solution containing the storage buffer (Table 2.5) was screened for potential crystallization conditions using commercially available screening kits from Emerald BioSystems (Wizard 1 and 2), Molecular Dimensions Limited (MD 1 and 2) and Jena Bioscience (JBScreen classic 1-4). The authentic and N-terminally His-tagged versions of wild-type HCoV-229E Nsp9 were subjected to crystallization trials. No crystalline substance was seen with the N-terminal His-tag version of wild-type HCoV-229E Nsp9, whereas the authentic version yielded crystals in several conditions using the screening kits. Most useful precipitants were ammonium sulfate and polyethylene glycol monomethyl ether (PEG-MME) 5000 at acidic pH. The crystalline substances were further optimized with the addition of various additives. Hexagonal crystals were obtained with 2-methyl-2,4-pentanediol (MPD) as an additive (Fig. 3.2.2 (a)). The optimized crystallization condition was 0.1 M sodium acetate pH 4.0-4.5, 1.8-2.1 M ammonium sulfate, and 5% MPD at 20°C.

HCoV-229E Nsp9 Cys69Ala mutant: In the case of the Cys69Ala mutant, multiple thin plates and needles were obtained using crystallization screening kits from Sigma-Aldrich (basic and extension kits). The conditions identified for crystallization of the wild-type Nsp9 failed to yield crystals of the mutant. Instead, the following crystallization conditions were established: 0.2 M sodium acetate pH 4.6, 0.2 M ammonium sulfate, and 30% PEG-MME 2000 at 20°C. Addition of 10 mM DTT to the reservoir solution was crucial for obtaining single thick needles (Fig. 3.2.2 (b)).

Wild-type SARS-CoV Nsp9: Two versions of SARS-CoV Nsp9 wild-type (N-terminal His-tag and authentic version) were subjected to crystallization trials. Initial screening experiments were performed using crystallization screening kits from Emerald

BioSystems (Wizard 1 and 2) and Molecular Dimensions Limited (MD 1 and 2). Unlike wild-type HCoV-229E Nsp9, the authentic version of wild-type SARS-CoV Nsp9 did not yield any crystals. Only the N-terminal His-tag version gave crystals (Fig. 3.2.2 (c)). The optimized crystallization condition was 0.1 M phosphate citrate buffer pH 4.2 and 2.1 M ammonium sulfate at 20°C, which is very similar to the crystallization condition reported earlier (Sutton *et al.,* 2004).
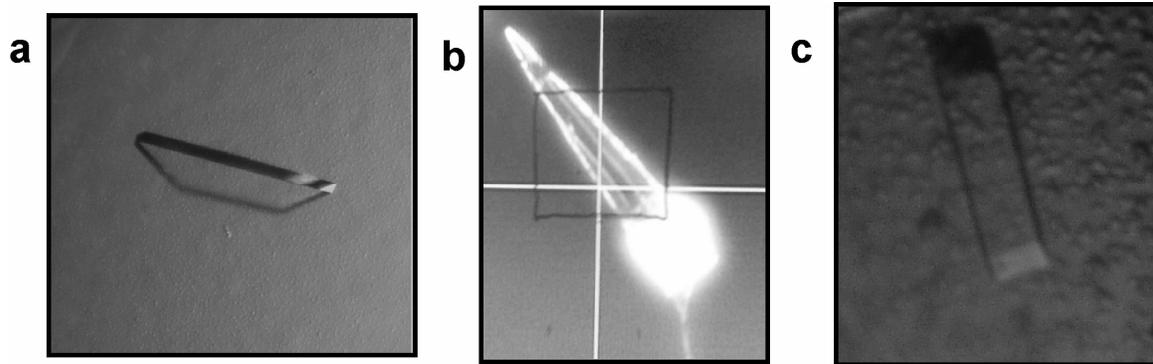
Fig. 3.2.2 Crystals of Nsp9s
Three-dimensional crystals were obtained after one day. (a) wild-type HCoV-229E Nsp9, (b) HCoV-229E Nsp9 Cys69Ala mutant, (c) wild-type SARS-CoV Nsp9 (N-His-tag).

3.2.3 Structure elucidation

3.2.3.1 Data acquisition

In addition to the optimized precipitant condition, wild-type HCoV-229E Nsp9 and wild-type SARS-CoV Nsp9 crystals were rinsed in 25% glycerol and a HCoV-229E Nsp9 Cys69Ala mutant crystal was rinsed in 25% PEG 400 for cryoprotection. The crystals were immediately cryo-cooled in liquid nitrogen. Diffraction data sets from crystals were collected at 100 K on the X13 beam line at Deutsches Elektronen-Synchrotron (DESY), Hamburg.

3.2.3.2 X-ray diffraction data

Wild-type HCoV-229E Nsp9: Crystals displayed space group P622 with unit cell dimensions $a = b = 85.63$ Å, $c = 48.69$ Å, and one monomer in the asymmetric unit. The average $I/\sigma(I)$ was 13.7 for the resolution range 40.0-1.75 Å and 4.45 in the highest resolution shell, with an overall completeness of 99.9%. A total of 129,656 measurements were made, representing 11,317 independent reflections, with an R_{merge} of 8.3%. The

Matthews coefficient for one molecule per asymmetric unit was 2.1 A^3/Da and the solvent content was 42.0% (Matthews, 1968). Diffraction data are summarized in Table 7.1 (Appendix).

HCoV-229E Nsp9 Cys69Ala mutant: Crystals displayed space group $P2_12_12_1$ with unit cell dimensions a = 26.40 Å, b = 61.38 Å, c = 107.31 Å, and a dimer in the asymmetric unit. The average $I/\sigma(I)$ was 19.7 for the resolution range 30.0-1.80 Å and 3.96 in the highest resolution shell, with an overall completeness of 99.4%. A total of 139,726 measurements were made, representing 16,843 independent reflections, with an R_{merge} of 8.9%. The Matthews coefficient for two molecules per asymmetric unit was 1.9 A^3/Da and the solvent content was 33.5% (Matthews, 1968). Diffraction data are summarized in Table 7.2 (Appendix).

SARS-CoV Nsp9: Crystals displayed space group P4 with unit cell dimensions a=b= 40.11 Å, c = 95.24 Å, and one monomer in the asymmetric unit. The average $I/\sigma(I)$ was 14.0 for the resolution range 31.75-2.80 Å and 6.90 in the highest resolution shell, with an overall completeness of 99.8%. A total of 12,233 measurements were made, representing 3,753 independent reflections, with an R_{merge} of 8.5%. The Matthews coefficient for one molecule per asymmetric unit was 2.95 A^3/Da and the solvent content was 58.3% (Matthews, 1968). Diffraction data are summarized in Table 7.4 (Appendix). The crystals were of the same habit as previously reported by Sutton *et al.* (2004); therefore, no further model building and refinement were carried out with this data set.

3.2.3.3 Molecular replacement solution and quality of the structural models

Wild-type HCoV-229E Nsp9: The structure was determined by molecular replacement using a monomer of SARS-CoV Nsp9 (Egloff *et al.,* 2004) as the search model, and refined to 1.75 Å resolution. Residues 1-7 and 33-36 could not be modeled due to lack of electron density. Alternate conformations were detected in the electron density for the side chains of Met9 and Lys82. The final R factor for the structural model was 19.0% and the R_{free} was 22.4%; 96.8% of the amino-acid residues are in the most-favored regions of the Ramachandran plot and the remainder in the additionally allowed regions (Laskowski

et al., 1993). A Ramachandran plot for the HCoV-229E Nsp9 wild-type structure is available in Fig. 7.1 (Appendix). Final model refinement statistics are listed in Table 7.1 (Appendix).

HCoV-229E Nsp9 Cys69Ala mutant: The structure was determined by molecular replacement, using a dimer of the SARS-CoV Nsp9 (Egloff *et al.,* 2004) as the search model. Residues 1-2 and 109 of monomer A have not been modeled due to lack of electron density; the same is true for residues 1-4 and 107-109 of monomer B. The segment comprising residues 53-56 could be built into electron density but proved to be very flexible. The structure was refined to a resolution of 1.80 Å, with R = 22.1% and R_{free} = 28.1% (see Fig. 3.2.3). 96.6% of the amino-acid residues in the structural model are in the most favored regions of the Ramachandran plot and the remainder in the additionally allowed regions (Laskowski *et al.,* 1993). A Ramachandran plot for the HCoV-229E Nsp9 Cys69Ala mutant structure is available in Fig. 7.2 (Appendix). Final model refinement statistics are listed in Table 7.2 (Appendix).
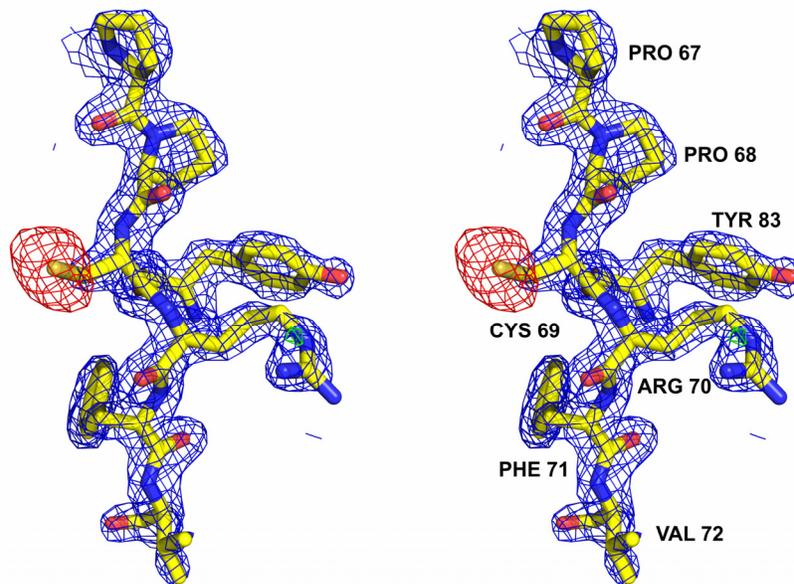
Fig. 3.2.3 Stereo view of the electron density map of the HCoV-229E Nsp9 Cys69Ala mutant
The 2|Fo|-|Fc| electron density map is contoured at 1.2 σ and the |Fo|-|Fc| is contoured at 3 σ above the mean. Alanine 69 was replaced with cysteine as in the wild-type Nsp9 and the structure was refined using REFMAC5 (Murshudov *et al.,* 1997).

3.2.4 Overall structure of the Nsp9 monomer

Crystals of wild-type HCoV-229E Nsp9 contain one monomer per asymmetric unit, which forms a homodimer due to the crystallographic two-fold axis (see below). The fold of the monomer belongs to the family of oligonucleotide/oligosaccharide-binding modules (OB fold). This fold is characteristic of proteins binding to single-stranded nucleic acids (Theobald *et al.,* 2003) and occurs, for example, in single-stranded DNA-binding proteins from bacteria (Webster *et al.,* 1997) to man (Bochkarev *et al.,* 1997) as well as in viruses (Mapelli *et al.,* 2005).
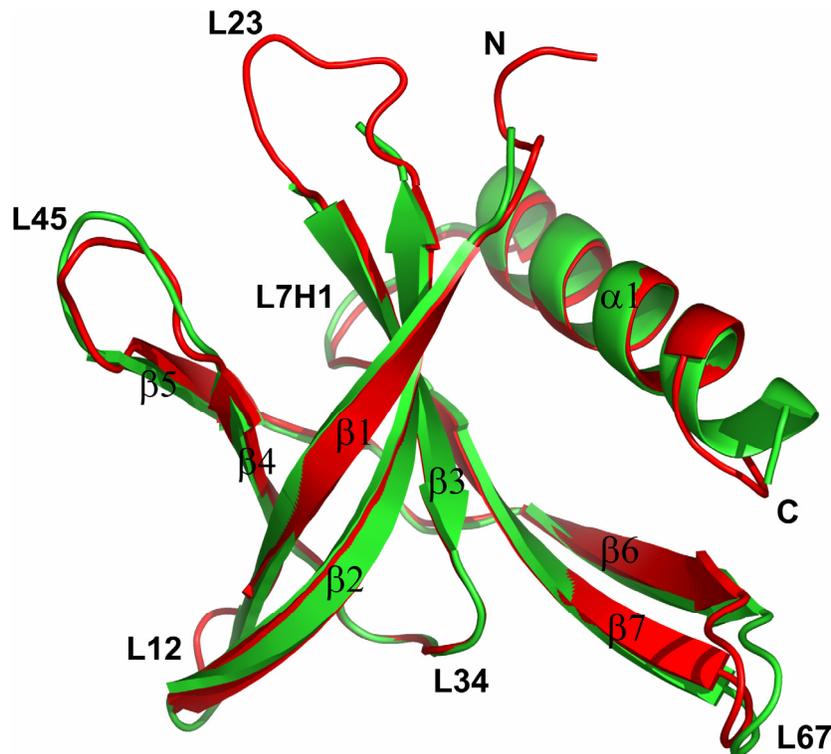
Fig. 3.2.4 Superimposition of monomers
Ribbon representation of HCoV-229E Nsp9 wild-type and Cys69Ala mutant monomers, superimposed with a C^{α} rms deviation of 0.71 Å. Green, HCoV-229E Nsp9 wild-type and red, Cys69Ala mutant. The loop L23 for HCoV-229E Nsp9 wild-type could not be built due to the lack of electron density.

The canonical OB fold comprises five antiparallel β-strands that form a partial barrel, and an α-helix that packs against the bottom of the barrel, usually in an orientation along the long axis of the β-barrel cross-section (Theobald *et al.,* 2003). In the classical OB fold, the α-helix is interspersed between β-strands 3 and 4, but in Nsp9, the helix is appended

at the C-terminal of the polypeptide chain (residues 92-108). Also, Nsp9 has two extra β-strands (strands 6 and 7) forming a long hairpin (L67). Some of the loops connecting the β-strands, *e.g.* L12, L23, L45, and L67 (see Fig. 3.2.4), are very flexible.

In the electron density map for wild-type HCoV-229E Nsp9, an MPD molecule was unequivocally located, which fills a space between strand β2 and the C-terminal α-helix in the monomer, very much in agreement with the commonly observed binding pattern for this amphiphilic additive (Anand *et al.,* 2002a). The hydroxyl groups of the MPD make hydrogen bonds with Asn27 and a (half-occupied) sulfate ion, which is in turn interacting with one of two alternative side-chain conformers of Lys82 and the main-chain amide of Asn27. The hydrophobic side of the MPD interacts with Leu29 and packs
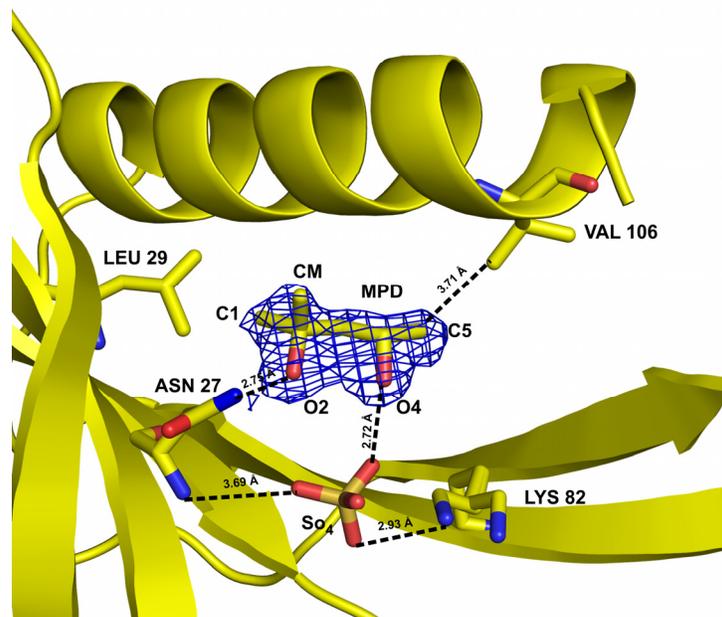
Fig. 3.2.5 An MPD molecule bound to the wild-type HCoV-229E Nsp9
The MPD molecule was bound in an R enantiomeric form. Hydrogen bonds and hydrophobic interactions are indicated by broken lines. The MPD molecule is shown with 2|Fo|-|Fc| electron density map contoured at 1.0 σ above the mean.

against the helix near Val106 (Fig. 3.2.5). This observation nicely explains why 5% MPD was an essential additive in the crystallization of wild-type HCoV-229E Nsp9, in addition to the $(NH_4)_2SO_4$.

The structure of the monomer of the HCoV-229E Nsp9 Cys69Ala mutant displays an r.m.s. deviation of 0.71 Å from the wild-type monomer (for 92 C^α atoms of chain A of the mutant; the corresponding values for chain B are 87 C^α atoms and 0.67 Å; see Fig. 3.2.4). In this calculation, residues 1-7, 33-36 (loop L23), and 107-109 have been omitted because of weak or non-visible electron density in one or both of the two structures. Larger-than-average deviations occur in loops L12 (residues 19-22) and L45 (residues 55-60; the tip of this loop at residue Ser58 deviates by 4.36 Å and 2.89 Å, respectively, between wild-type and mutant molecules A and B). The r.m.s. deviation between monomers A and B of the Cys69Ala mutant is 0.96 Å (for 99 C^α atoms; see table 7.6 (Appendix)).

In contrast to the wild-type Nsp9, MPD was not a useful additive in crystallization experiments with the mutant protein. However, DTT was absolutely essential. Again, an extra electron density between Asn27 and the α-helix was located (this time near Ile102) and attributed this to a DTT molecule, although the assignment was not as unambiguous as was the identification of MPD in a near-by location in the wild-type protein.

3.2.4.1 Comparison of the Nsp9 monomer with SARS-CoV Nsp9

The HCoV-229E Nsp9 monomer is also very similar to the monomer of SARS-CoV Nsp9 (Egloff *et al.,* 2004; PDB code: 1QZ8), with an r.m.s. deviation of 0.75 Å for 84 C^α atoms of wild-type Nsp9 (for chain A of 1QZ8; the value for chain B is 0.66 Å). The values for the mutual comparisons between the individual chains of the Cys69Ala mutant and those of 1QZ8 are between 0.76 and 1.23 Å. Interestingly, the other available crystal structure for SARS-CoV Nsp9 (Sutton *et al.,* 2004; PDB code: 1UW7) is significantly more distant in terms of r.m.s. deviations, with 1.75 Å for 94 C^α atoms of the wild-type HCoV-229E Nsp9, and 1.39 Å for 94 C^α atoms of the Cys69Ala mutant.

3.2.5 Structure of the Nsp9 dimer

Wild-type HCoV-229E Nsp9 forms a disulfide-linked homodimer, with the two-fold crystallographic axis of symmetry running through the disulfide bond formed between the Cys69 residues of each monomer (Fig. 3.2.6 (a)). The α-helix of each monomer is

also involved in dimerization through formation of two hydrogen bonds between the Asn92 side chain and the main chain of residue 74 (β-strand 6), as well as one H-bond between the Asn92 side-chain amide and the C-terminal carboxylate (residue 109; Fig. 3.2.6 (b)). Among the sequenced coronaviral Nsp9 proteins, this asparagine is only present in HCoV 229E and SARS-CoV; other coronaviruses have either Thr or Ser at this position. There is a fourth hydrogen bond donated by the $N^{\eta 1}$ atom of totally conserved Arg95 (helix H1; see Fig. 4.2.3) to the main-chain carbonyl of residue 72 (β-strand 6). Because of the crystallographic two-fold symmetry, all of these interactions are duplicated in the dimer, so that there are a total of eight intersubunit hydrogen bonds (Fig. 3.2.6 (b)). In spite of the presence of these favorable interactions, there seem to be a number of less ideal ones. For example, Arg70 does not have an interaction partner proper; its side chain makes contacts with Pro67 and Pro68. Its intersubunit contacts are in fact determined by the nearby disulfide bond (Cys A69 - Cys B69). There is a weak, but favorable interaction between the side chains of Phe A71 and Phe B86. The two α-helices (residues 92-108) that are part of the interface are in an antiparallel orientation (describing an angle of 167°, Fig. 3.2.6 (a)) with a close approach of 3.77 Å between the C^{α} atoms of Gly A96 and Gly B103. In between these two glycines of the α-helix, there is a third one, Gly 100, which also facilitates the close approach of the two helices (distance to its symmetry-mate is 4.13 Å, see Fig. 4.2.2). Gly96 and Gly100 are completely conserved among the coronaviruses and Gly103 is present in group I coronaviruses (Fig. 4.2.3). The surface area per Nsp9 monomer buried through dimer formation (Lee & Richards, 1971) is 985 Å^2.
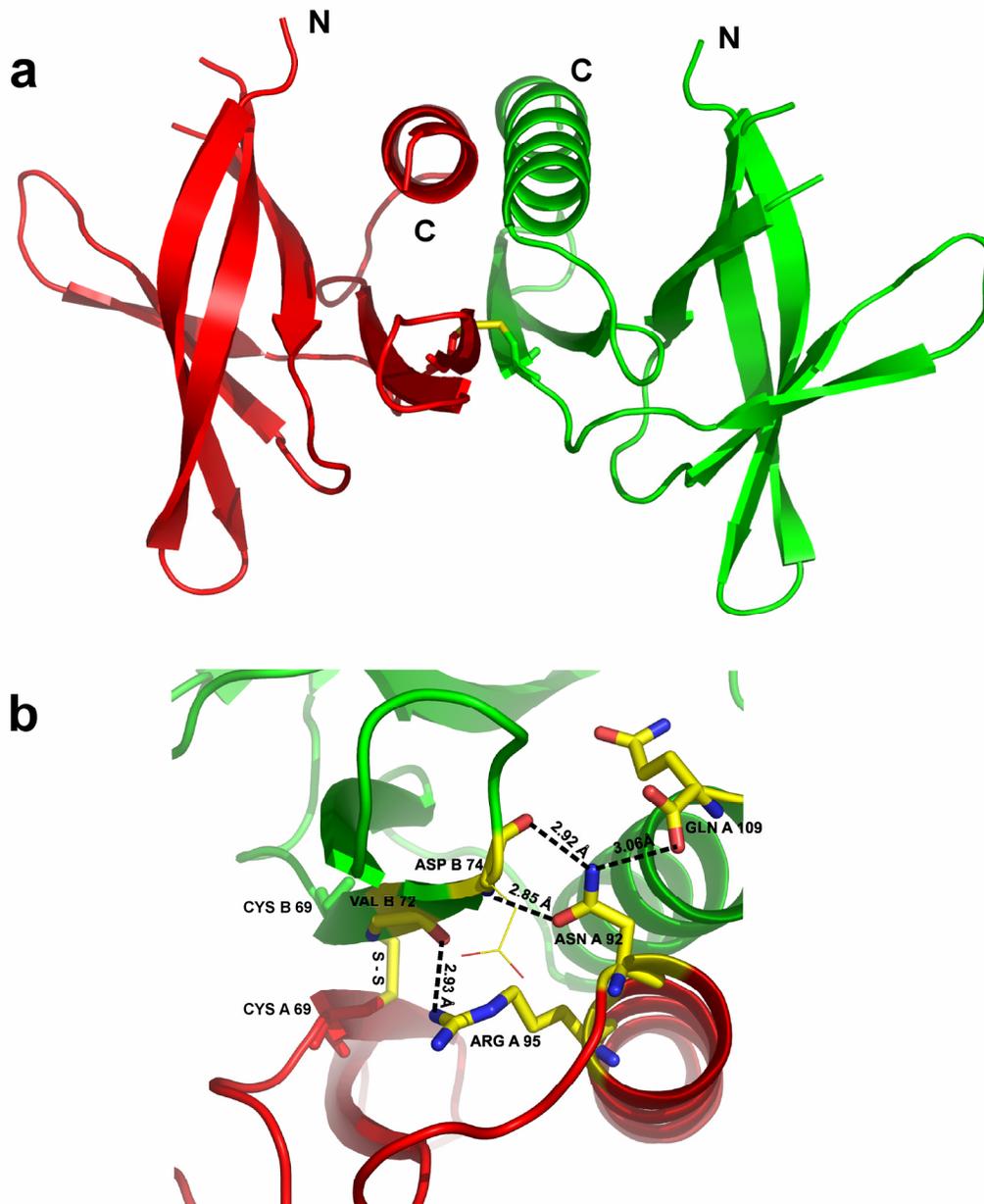
Fig. 3.2.6 Structural features of the wild-type HCoV-229E Nsp9
The two monomers are colored in red and green, respectively. (a) Ribbon representation of wild-type HCoV-229E Nsp9, a disulfide bridge mediated homo-dimer. (b) Residues involved in the dimer interface of wild-type Nsp9, the disulfide bridge is shown in yellow sticks and the few hydrogen bonds are indicated by broken lines.

Although the residue responsible for disulfide formation in HCoV-229E Nsp9, Cys69, is conserved in SARS-CoV Nsp9, and the sequence identity is as high as 45% between the two proteins (see Fig. 4.2.3), the mode of dimerization in the latter is very different from the HCoV-229E Nsp9. A disulfide bond is not formed, and the dimerization interface mainly involves the α-helix of each subunit, but in a parallel rather than anti-parallel orientation with an angle of 167° (Egloff *et al.*, 2004; Sutton *et al.*, 2004). In contrast to HCoV-229E Nsp9, which was prepared with authentic N- and C-termini, the SARS-CoV protein used by Sutton *et al.* (2004) for crystallization carried 30 additional amino-acid residues at the N-terminus, due to the cloning procedure. From the structure (Sutton *et al.*, 2004), it is evident that the additional N-terminal segment leads to formation of a β-hairpin involving residues (-7) to 8, as well as an intermolecular salt bridge between GluA(-7) and ArgB111. This additional interaction might favor the dimerization mode seen in the Sutton *et al.* (2004) structure. However, the structure published by Egloff *et al.* (2004) for SARS-CoV Nsp9 was derived from a protein that carried only six additional histidine residues at its N-terminus (B. Canard, personal communication). These residues were not seen in the electron density maps, presumably due to disorder, and certainly are not involved in intersubunit interactions. Yet, this structure still features a mode of dimerization that is highly similar to that described by Sutton *et al.* (2004) and completely different from the one observed for the HCoV-229E Nsp9.
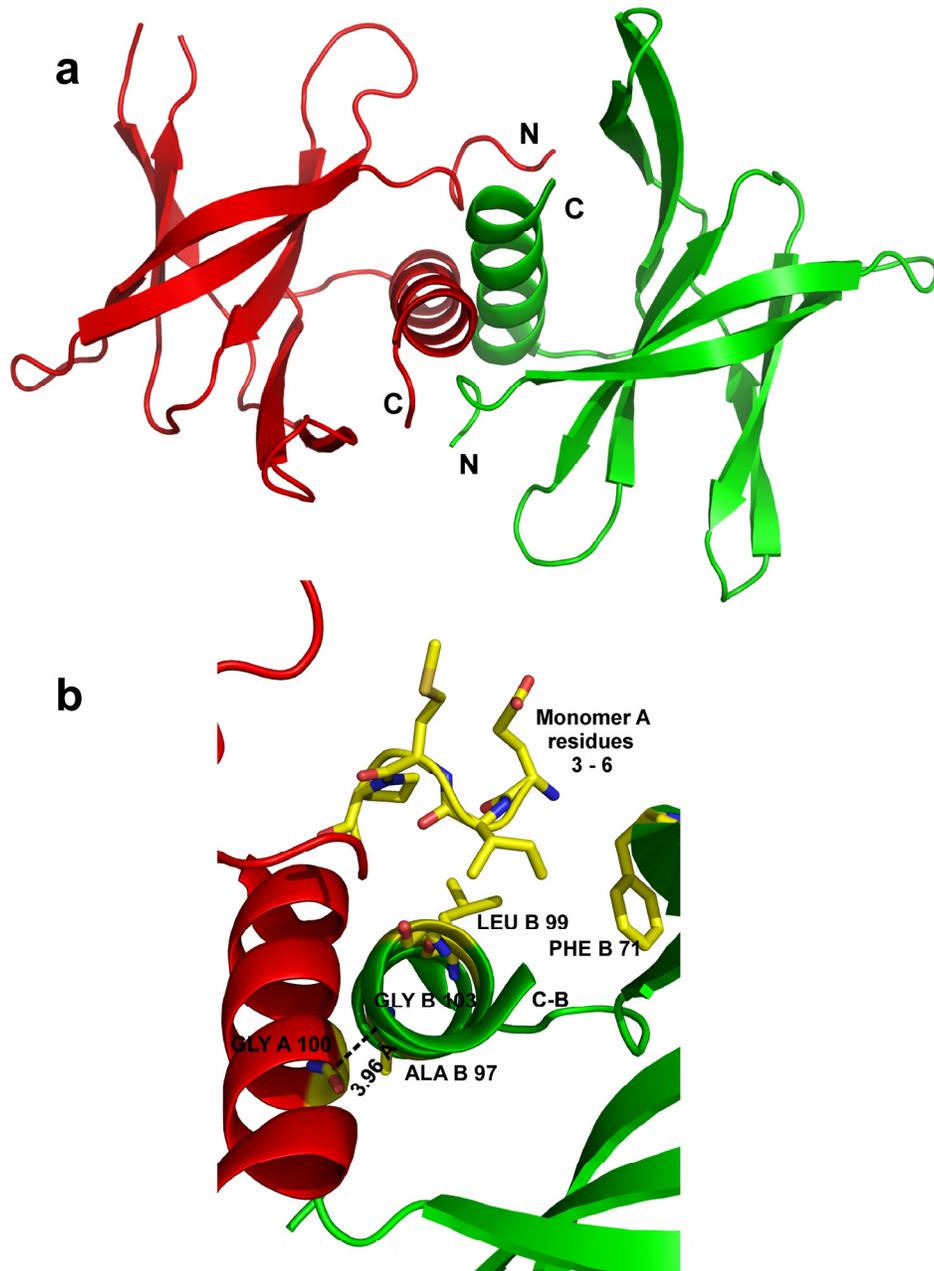
Fig. 3.2.7 Structural features of the HCoV-229E Nsp9 Cys69Ala mutant
The two monomers are colored in red and green, respectively. (a) Ribbon representation
of the HCoV-229E Nsp9 Cys69Ala mutant dimer. (b) The dimer interface, the N-terminal
of monomer A makes hydrophobic interactions with monomer B. The closest approach
between the C-terminal α-helices, between the Cα atoms of Gly A100 and Ala B97, is
indicated by a dotted line.

Disulfide bonds are rare in proteins that exist in the cytosol, where the environment is of reducing character, and the possibility of the dimerization mode seen in the structure of HCoV-229E Nsp9 being an artifact of disulfide-bond formation has to be discussed. It is noteworthy that 5 mM DTT present throughout protein preparation and crystallization (higher concentrations of DTT prevented the crystallization of the protein). In order to assess the role of the disulfide bond in dimer formation, Cys69 was replaced in HCoV-229E Nsp9 by alanine. The structure of the mutated protein (Fig. 3.2.7 (a)) revealed a dimer which is grossly different from that of wild-type HCoV-229E Nsp9: When superimposing monomer A of the mutant structure onto the same monomer of the wild-type protein, the centroid of monomer B deviates from its position in the wild-type protein by 23.5 Å, and the angle of rotation between the two positions of monomer B is 72°. However, this dimerization mode is identical to that of SARS-CoV Nsp9 (cf. Fig. 3.2.6, Fig. 3.2.7 and Fig. 4.2.1). The dimer of the HCoV-229E Cys69Ala mutant can be superimposed onto the dimer of the SARS-CoV Nsp9 protein (1QZ8; Egloff *et al.,* 2004) with an r.m.s. deviation of 0.99 Å for 175 C^{α} atom pairs (see Fig. 4.2.1). The r.m.s. deviation is much higher (2.7 Å for 191 C^{α} atom pairs) for the SARS-CoV Nsp9 structure described by Sutton *et al.* (2004; 1UW7); this is very likely due to the disturbances of the latter structure by the N-terminal tag residues. In the HCoV-229E Nsp9 mutant, dimerization appears to rely on a few interactions only. There is not a single proper hydrogen bond between the two monomers, and only a few hydrophobic contacts mediate the interaction (Fig. 3.2.7 (b)). The immediate N-terminus is disordered, but residues 3 to 6 and 5 to 7 of the A and B chain, respectively, lie over the other monomer and weakly interact with conserved Phe71 (strand β6) and with the C-terminal α-helix near residues Leu99 and Gly103. The majority of the interactions between the two monomers is made by the two helices, one from each monomer, that run largely in parallel in this dimer, crossing at an angle of 48° and a closest approach of 3.96 Å (C^{α}-C^{α}) between Ala97 and Gly100 (Fig. 3.2.7 (b)). But again, the hydrophobic contacts between the helices are only weak ones. Importantly, neither of the two helices deviates from its ideal geometry, including the intrahelical hydrogen bonds, for the benefit of the intermolecular contacts. The surface area buried upon dimer formation (Lee & Richards, 1997) is 687 $Å^{2}$ for the mutant protein. This value is significantly lower than that observed for the dimerization

mode seen for the wild-type HCoV-229E Nsp9 (see above). Even if a few more intermolecular interactions were made by the disordered N-terminal residues not seen in the electron density map (residues 1-2 of chain A and 1-4 of chain B), this value would not increase dramatically, and such interactions would very likely not be strong (otherwise these residues would not be disordered). In summary, the monomer-monomer interface of the Cys69Ala mutant of HCoV-229E Nsp9 is far from ideal and appears to be much weaker than the interface seen in the wild-type protein.

3.2.6 Higher oligomers in the crystal of wild-type Nsp9

In the crystal structure of wild-type HCoV-229E Nsp9, three dimers are arranged to form a hexamer with 32 symmetry (Fig. 3.2.8). There are multiple interactions between monomers across the hexamer, which will be briefly discussed here according to the color scheme used in the figure. At the center of the hexamer, there are two disordered sulfate ions sitting on the crystallographic threefold axis, 19.0 Å apart. They make ionic interactions (2.58 Å from the closest sulfate oxygen) with the Lys50 side chains (N^ζ atom) of the red, green and blue monomers (upper layer of the hexamer) and the yellow, cyan, and magenta monomers (lower layer), respectively. Also, Asp19 of each monomer forms a 2.90 Å *inter*molecular hydrogen bond with the Lys50 residue (e.g., red-green), as well as a 4.76 Å *intra*molecular salt bridge with Lys50 of the same monomer (e.g., red-red). As a result, a ring of ionic interactions formed by Asp19 (loop L12) and Lys50 (L34) residues along both the upper and the lower rim of the central cylinder inside the trimer of dimers (Fig. 3.2.8). Asp19 is present in most coronavirus Nsp9 sequences (not in SARS-CoV and IBV), whereas Lys50 is highly conserved across the family (see Fig. 4.2.3). The cylindrical hole running along the three-fold axis is only about 4.4 Å wide, i.e. large enough for sulfate ions, but too small to accommodate single-stranded nucleic acid at the center of the hexamer (Fig. 3.2.8).
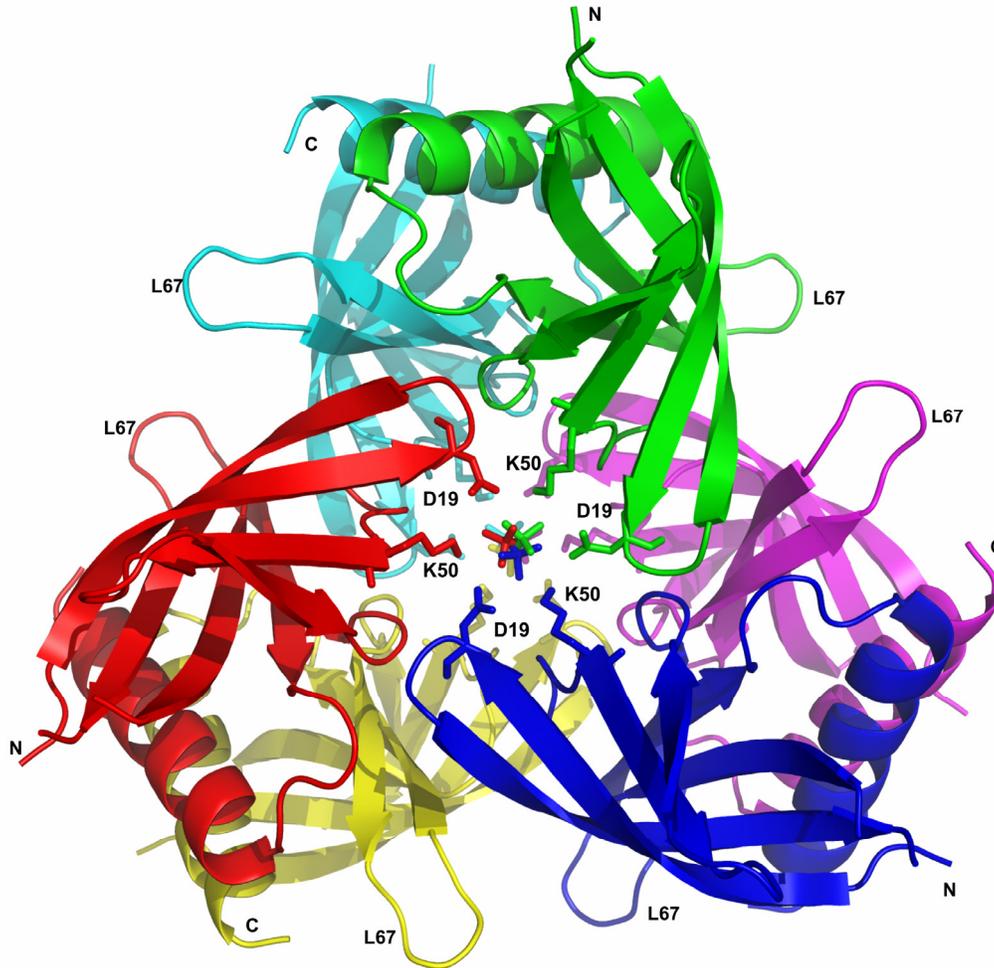
Fig. 3.2.8 Ribbon representation of the wild-type HCoV-229E Nsp9 hexamer
Three dimers of the protein form a hexamer through the 32 axis of symmetry. The threefold axis is at the center of the hexamer. Nsp9 monomers in the upper layer are colored red, blue and green, and those in the lower layer are colored yellow, cyan and magenta. The two sulfate ions on the threefold axis are indicated in the same colors. Each sulfate is three-fold disordered. The twofold axes run between the monomers.

Additional intermolecular interactions within the Nsp9 hexamer are listed in Table 3.2.1.

Through the 6-fold axis of the crystal, hexamers are arranged into 36-mers (Fig. 3.2.9).
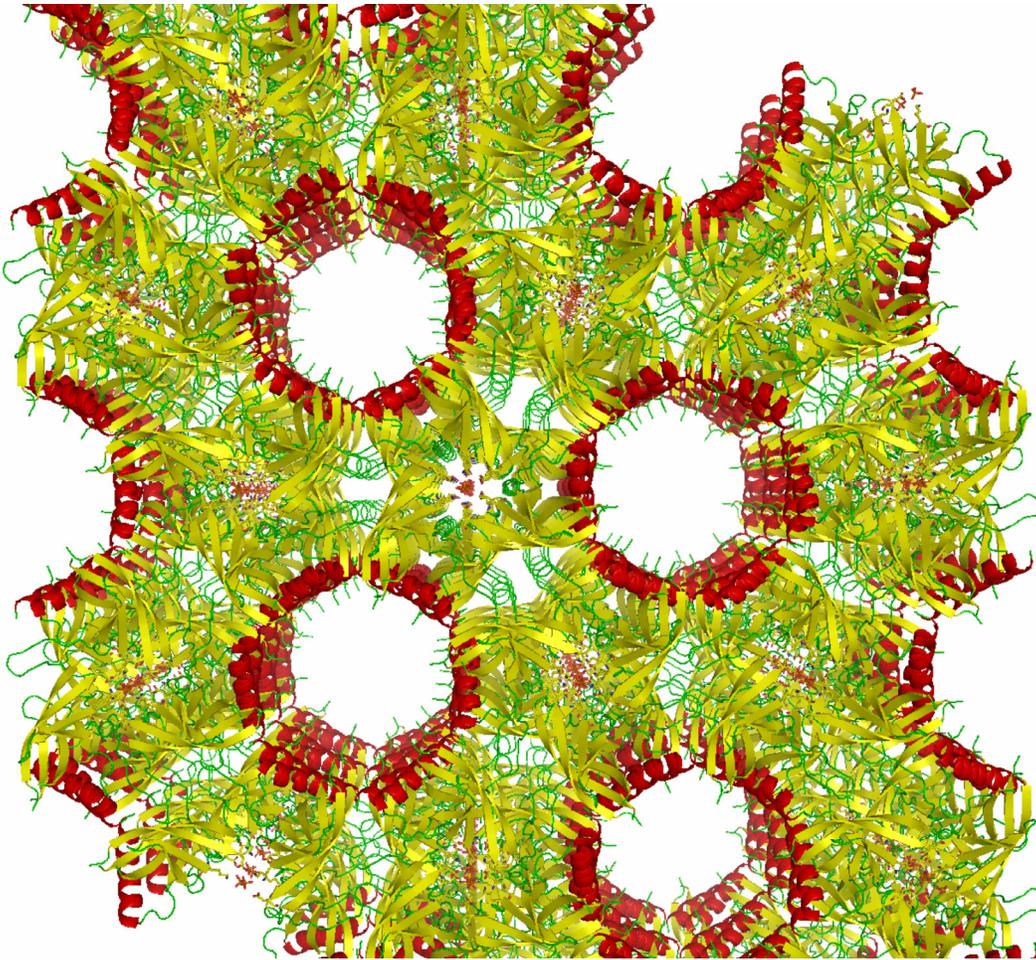
Fig. 3.2.9 Ribbon representation of several hexamers forming a 36-mer of wild-type HCoV-229E Nsp9
Six copies of the wild-type HCoV-229E Nsp9 hexamer seen in the crystal form a 36-mer due to the hexagonal space-group symmetry. The flexible N-termini and the L23 loops are oriented towards the big solvent channel at the center of the hexamer of hexamers.

3.2.7 Nsp9 polymers in the crystal of the Cys69Ala mutant

In the crystal structure of the Cys69Ala mutant dimer of HCoV-229E Nsp9, a second protein-protein interface (Fig. 3.2.9 (a); Table 3.2.1) is formed through the close approach of strands β5 of neighboring molecules, although only one hydrogen bond is formed (Ser58 O...Glu64 N, 2.62 Å). In addition, there are two hydrogen bonds donated by the guanidinium group of Arg94 (in the helix) to the carbonyl oxygen of Gly34 (L23; 2.99 and 3.14 Å). Furthermore, there is yet another H-bond between the side-chains of Asn89 (loop L7H) and Asp57 (L45; 2.99 Å). The surface area buried by formation of this dimer is 450 $Å^2$. Together, the interfaces mediated by the α-helix (described above) and

the β-sheet lead to the polymerization of the protein (see Discussion section 4.2, Fig. 4.2.5). Yet another protein-protein interface (β6 and β7) with quasi-twofold symmetry, but not involved in polymer formation, is mentioned in Fig. 3.2.10 (b); Table 3.2.1.
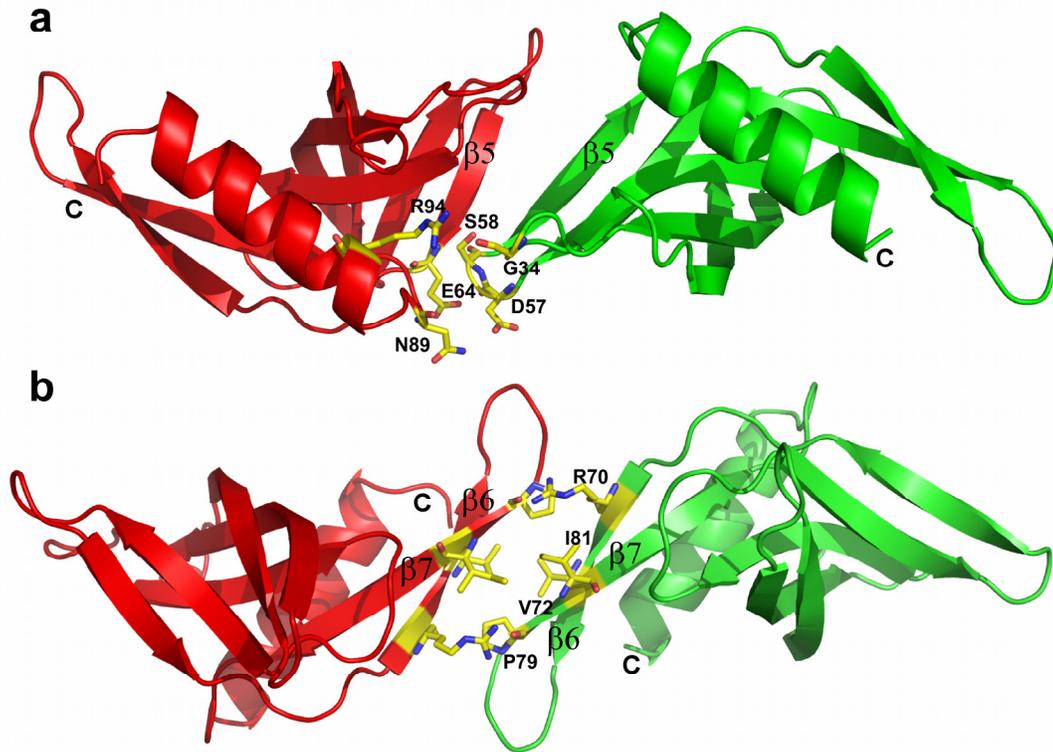
Fig. 3.2.10 Dimer interfaces of the HCoV-229E Nsp9 Cys69Ala mutant
In the crystal structure of the HCoV-229E Nsp9 Cys69Ala mutant, other than the main interface mediated by the α-helices (Fig. 3.2.7), two additional dimer interfaces are also seen. (a) A β-sheet-mediated interface forms through the close approach of the β5 strands (b) an interface forms between the sheet β67 and its quasi-twofold symmetry mate. The two monomers are colored red and green, respectively. Residues involved in the interfaces are shown using sticks.

Table 3.2.1: Protein-protein interfaces seen in the HCoV-229E Nsp9 wild-type (W) and Cys69Ala mutant (M) structure: Intermolecular interactions, buried dimerization surface area and shape complementarity values. Colors given under "dimer interface" refer to the colors of the monomers in Fig. 3.2.9. Distances or residue names given in italics are those involving monomer B rather than A of the mutant structure (which has a dimer in the asymmetric unit)

Dimer interface (W)	Hydrogen bonds	Other interactions	BSA*	Sc†
Red-yellow (hexamer, W1)	(α helix) Arg 95 Nη1 … (β6) Val 72 O (2.93 Å) (α helix) Asn 92 Nδ2 ... (β6) Asp 74 O (2.92 Å) (α helix) Asn 92 Nδ2 ... (α helix) Gln 109 O (3.06 Å) (β6) Asp 74 N ... (α helix) Asn 92 Oδ1 (2.85 Å)	(β6) Cys 69 ... (β6) Cys 69 disulfide bond (β6) Phe 71 ... (β7) Phe 86 (α helix) Gly 96 … (α helix) Gly 103	985	0.56
Red-green (hexamer, W2)	(L12) Asp 19 Oδ2... (L34) Lys 50 Nζ (2.90 Å) (L12) Gly 20 N … (L56) Glu 64 Oε2 (2.67 Å)	(β2) Gly 21, Ile 22 … (L56) Glu 64, Gly 66, Pro 67	244	0.76
Red-cyan (hexamer, W3)	(L34) Lys 46 Nζ … (L67) Pro 79 O (2.88 Å)	(β2) Ile 22 … (L67) Pro 79	420	0.62
36-mer (W4)	(α helix) Tyr 101 Oη … (α helix) Leu 108 O (2.58 Å)	(β1) Val 11, Ala 13 ….. (β2) Glu 25, Gly 26, Asn 27 (α helix) Gln 109 … (L67) Pro 76	551	0.74

Dimer interface (M)	Hydrogen bonds	Other interaction	BSA*	Sc†
α-helix-mediated	-	(N-ter) Glu 3, Ile 4, Met 5 *(Met 5, Pro 6, Gly 7)*… (β6) Phe 71, (α helix) Leu 99, Gly 103. (α helix) Ala 97 … (α helix) Gly 100	687	0.67
β-sheet-mediated	(β5) Ser58 O... (β5) Glu 64 N (2.62 Å) (α helix) Arg 94 Nη1 … (L34) Gly 34 O (2.99 Å) *(3.14 Å)* (L7H) Asn89 and (L45) Asp57 (2.99 Å)	-	450	0.59
β67-sheet-mediated	(β6) Arg 70 Nη1 … (β7) Pro 79 O (2.90 Å) *(2.70 Å)* (β6) Arg 70 Nη2 … (β7) Pro 79 O (2.77 Å) *(2.53 Å)*	(β6) Val 72 … (β7) Ile 81	269	0.68

* BSA - Buried dimerization surface area per monomer (Å^2)   † Sc - Shape complementarity

3.2.8 Oligomeric state in solution

In order to determine the oligomeric state of wild-type and mutant HCoV-229E Nsp9 in solution, a number of biophysical and biochemical techniques was applied. For both wild-type HCoV-229E Nsp9 (fresh preparation) and the Cys69Ala mutant, Dynamic Light-Scattering (DLS) yielded a monodisperse peak centered at a hydrodynamic radius of 28 ($\pm$ 1.4) Å, indicating that the homodimer is the prevalent species in solution. A similar result was obtained by analytical gel filtration, which showed a single peak corresponding to a molecular mass of ~26 kDa for both wild-type and mutant (data not shown). In agreement with the crystal structure, the DLS experiment revealed the presence of higher oligomers upon addition of small amounts of sulfate ions (up to 9.7 mM; data not shown).

Glutaraldehyde crosslinking was carried out for HCoV-229E Nsp9 wild-type, the Cys69Ala mutant, and SARS-CoV Nsp9. 0.01% glutaraldehyde was used with different protein concentrations ranging from 10 to 100 µM. The HCoV-229E Nsp9 Cys69Ala mutant and SARS-CoV wild-type Nsp9 showed similar crosslinking products corresponding to monomers, dimers, trimers, tetramers, and higher oligomers (Fig. 3.2.10). In HCoV-229E Nsp9, the wild-type showed only monomers, dimers, and trimers. The increasing presence of higher-molecular mass species correlated with increasing protein concentration. This pattern did not change in the presence of 36-mer or 51-mer single-stranded DNA of random sequence (not shown).

3.2.8.1 Oxidation state of Cys69 in solution

By titration of free sulfhydryl groups with Ellman's reagent (Ellman, 1959), the wild-type HCoV-229E Nsp9 has no free cysteine in solution, i.e. the disulfide bond exists in solution as well. However, in a more recent preparation of the wild-type protein, reaction with Ellman's reagent immediately after purification of wild-type Nsp9 did indicate the presence of one free cysteine per mole of protein. Yet, crystallization of this sample yielded crystals overnight that were of the same habit as the original ones obtained for the wild-type protein, with unit cell parameters $a = b = 85.4$ Å and $c = 48.8$ Å in space group P622. When this phenomenon was further investigated, the formation of the disulfide

bond seemed to depend on the age of the protein preparation. It is possible that in the presence of oxygen, gradual oxidation of the protein (probably correlated with the oxidation of DTT) may lead to formation of the disulfide bond, resulting in the dimerization mode visualized by X-ray crystallography. The concentration of DTT required to reduce the disulfide bond completely was determined as 10 mM by SDS gel electrophoresis (see Materials & Methods). With DTT concentrations up to 4 mM, the dimer was the dominant species visible on the gel, whereas above, the monomer was more pronounced. The dimer band vanished completely at 10 mM DTT. This result was the same independent of the presence or absence of a heating step (95°C for 5 min).
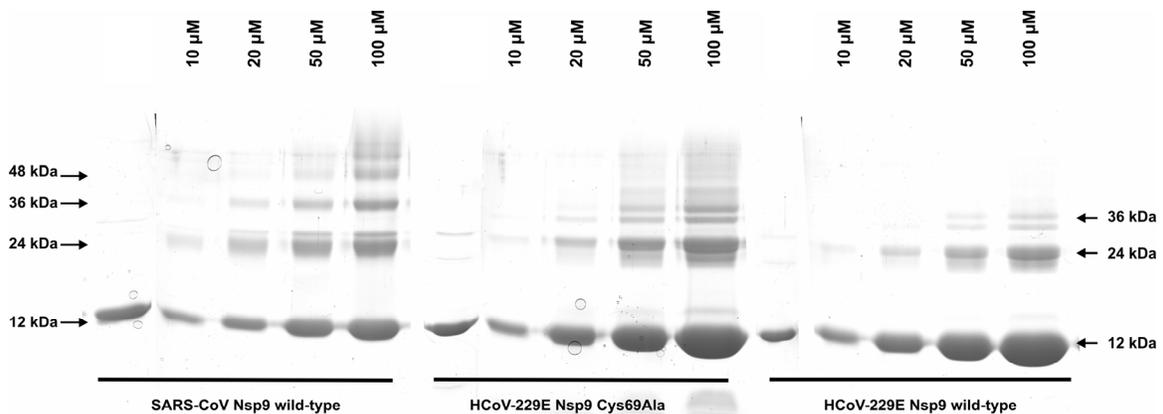
Fig. 3.2.10 Nsp9 crosslinking using glutaraldehyde
Crosslinking was carried out with different protein concentrations from 10 µM to 100 µM using 0.01% of glutaraldehyde. The respective molecular mass is labeled with arrows. SARS-CoV Nsp9 and HCoV-229E Nsp9 Cys69Ala mutant form higher oligomers with the highest protein concentration, presumably with a similar interaction as seen in the crystal structure. Instead, wild-type HCoV-229E Nsp9 does not form any higher oligomers as seen in the SARS-CoV Nsp9 and in the HCoV-229E Nsp9 Cys69Ala mutant.

In contrast to HCoV-229E Nsp9, the SARS-CoV Nsp9, which has three cysteine residues and which was prepared in the same way as its HCoV-229E orthologue (i.e., with authentic N- and C-termini), had three free sulfhydryl groups per mole in solution even after several weeks of storage as determined by using Ellman's reagent.

3.2.9 Binding of nucleic acids

3.2.9.1 Gel mobility-shift assay

Nucleic-acid binding with Nsp9 was demonstrated using a gel mobility-shift assay (a modified version of zone-interference gel electrophoresis, ZIGE (Abrahams *et al.,* 1988; Matthes *et al.,* 2006; Fig. 3.2.11)), the wild-type HCoV-229E Nsp9 bound to single-stranded oligodeoxynucleotides (6-mers to 50-mers; Fig. 3.2.11 (a), lanes 3 to 10) and, to a very limited extent (if at all), to a double-stranded oligodeoxynucleotide (24-mer; Fig. 3.2.11 (a), lane 2). Nsp9 of SARS-CoV also bound to both single-stranded and, again very weakly, to double-stranded oligonucleotides (Fig. 3.2.11 (b)). However, the effect on the gelshift did not increase smoothly with oligonucleotide length; rather, there was a step-wise increase from the 13-mer (Fig. 3.2.11 (b), lane 3; no shift) via the 18- and 24-mers (lanes 4, 5) and the 30- to 45-mers (lanes 6-9) to the 50-mer (lane 10). In contrast, nucleic-acid binding by the HCoV-229E Cys69Ala mutant was not detectable with this method except for a very weak shift with the 55-mer (Fig. 3.2.11 (c), lane "HAM 3"). As the apparent inability of the Cys69Ala mutant to bind nucleic acids could depend on the lack of a direct (hydrogen-bonding) interaction between the cysteine and the oligonucleotides, the cysteine was replaced also by serine. The Cys69Ser mutant did not bind short oligonucleotides either, but did show some gel shift with the 55-mer (Fig. 3.2.11 (c), lane "HSM 3"). Next, the corresponding cysteine in SARS-CoV Nsp9 was replaced by alanine and serine. Both the Cys73Ala and Cys73Ser mutants displayed a similar shift in the presence of the 55-mer oligodeoxynucleotide as the wild-type protein (Fig. 3.2.11 (c), lanes "SAM 3" and "SSM 3"). Reduction (by 50 mM DTT) or oxidation (by 17.5% H_2O_2) of the disulfide-containing wild-type HCoV-229E Nsp9 did not change the gelshift pattern of the protein in the presence of nucleic acids (not shown).

3.2.9.2 Surface plasmon resonance

Nsp9 binding to ssDNA was analyzed using surface plasmon resonance (SPR) experiments. For this, a 5′-biotinylated 50-mer oligonucleotide was immobilized on a streptavidin-coated chip (SA chip). Freshly prepared Nsp9 was treated with 5 mM DTT directly before injection. A signal was observed for Nsp9 interaction with the oligonucleotide when protein concentrations were in the μM range (Fig. 3.2.12).

Apparent K_D values for the wild-type Nsp9 from HCoV 229E and SARS-CoV were determined as 28 µM (χ^2 = 1.29) and 29 µM (χ^2 = 8.52 for a single-state binding model), respectively (The χ^2 value indicates the goodness of the model fit to the data points). In fact, the SARS-CoV Nsp9 binding profile is better explained by a two-state
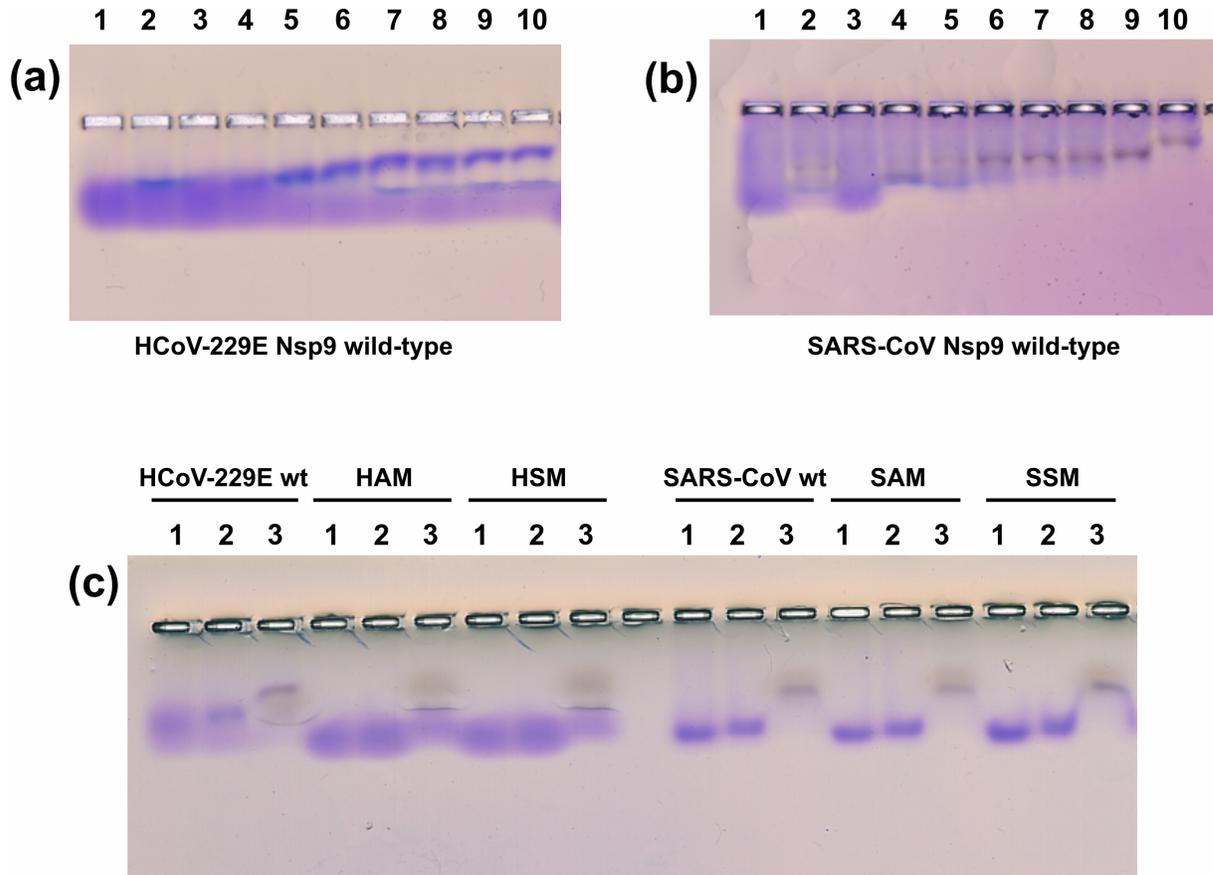
Figure 3.2.11 Gel mobility-shift assay with Nsp9 and DNA
Gel mobility-shift assay (zone-interference gel electrophoresis, 1% agarose; see Materials and Methods) probing oligonucleotide binding to Nsp9. (a) Wild-type HCoV-229E Nsp9; (b) wild-type SARS-CoV Nsp9. Lane 1, protein without ssDNA; lane 2, 24-mer dsDNA; lanes 3–10, various lengths of ssDNA from 6-mer to 50-mer. Wild-type HCoV-229E Nsp9 displays a linear increase of the shift with increasing length of ssDNA, whereas the increase is step-wise for SARS-CoV Nsp9. (c) Gel mobility-shift analysis for mutant proteins, compared to the corresponding wild-type. Lanes 1, protein without ssDNA; lanes 2, 24-mer, and lanes 3, 55-mer ssDNA with protein. The HCoV-229E Nsp9 Cys69Ala mutant (HAM) and the Cys69Ser mutant (HSM) do not show any shift with the 24-mer (lane 2) and only a small shift with the 55-mer (lane 3), whereas the SARS-CoV Nsp9 Cys73Ala mutant (SAM) and the Cys73Ser mutant (SSM) exhibit shifts with the 55-mer oligonucleotide that are similar to wild-type SARS-CoV Nsp9. The upper bands (gray) in lanes 3 for HAM and HSM correspond to precipitated, unbound 55-mer oligonucleotide (not stained by Coomassie brilliant blue; see Materials and Methods).

binding model ($\chi^2 = 0.73$). This is not true for HCoV-229E Nsp9, the binding profile of which agrees well with a single-state binding model. However, concentrations higher than 35 µM and 85 µM could not be used for the HCoV-229E Nsp9 and SARS-CoV Nsp9, respectively, because non-specific binding appeared to govern the profile above this value and saturation was not reached. The K_D values could also not be derived for the oxidized form of wild-type HCoV-229E Nsp9 nor for the SARS-CoV and HCoV-229E Nsp9 mutants because of the same phenomenon.
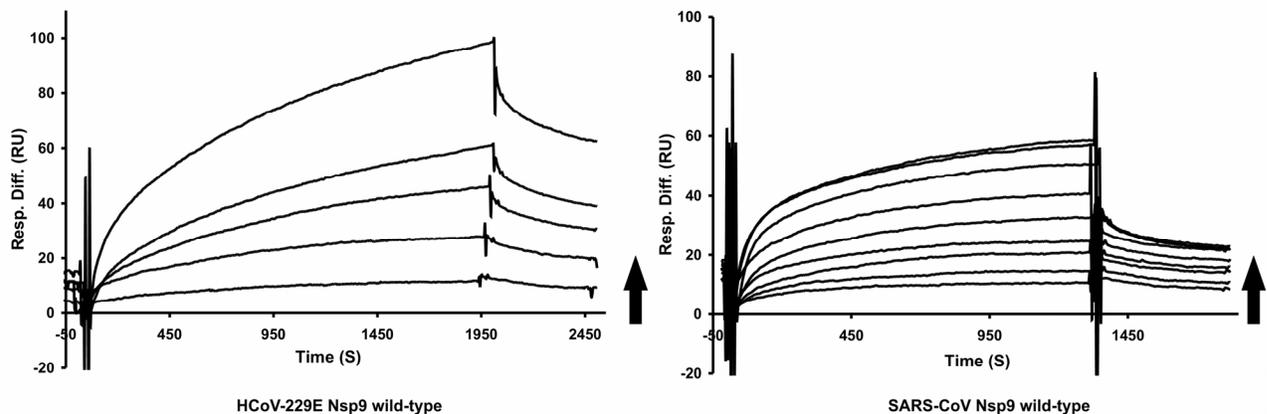
Fig. 3.2.12 Nsp9 binding to ss-DNA: sensograms from the surface plasmon resonance experiments
5′-biotinylated 50-mer oligonucleotide was immobilized on a SA chip up to 300 RU. The right-side arrow indicates the direction of the increasing protein concentration. HCoV-229E Nsp9 wild-type concentration varies from 2 µM to 35 µM and SARS-CoV Nsp9 wild-type concentration varies from 2 µM to 85 µM.

The reduced and oxidized state (*i.e.,* presence and absence of 5 mM DTT, resp.) of wild-type HCoV-229E Nsp9 with respect to binding the oligonucleotide was also compared in the surface plasmon resonance experiment. For this, the oligonucleotide was again immobilized on the chip. 20 µM of freshly prepared (less than three days old, containing one free thiol group per mole) wild-type HCoV-229E Nsp9 and the HCoV-229E Cys69Ala mutant were injected into the flow cell, with the constant presence of DTT in the running buffer. Also, 20 µM of aged preparation (more than two weeks old, no free cysteine) of wild-type HCoV-229E Nsp9 was injected without the presence of DTT in the running buffer. Both freshly prepared wild-type (in the presence of DTT) and Cys69Ala mutant protein showed a similar binding curve with the oligonucleotide and gave a

maximum response (R$_{max}$) of 33 RU. In contrast, the aged preparation of wild-type HCoV-229E Nsp9 displayed an R$_{max}$ of 83 RU, indicating much stronger binding to the nucleic acid than observed for the fresh preparation or the mutant (Fig. 3.2.13).
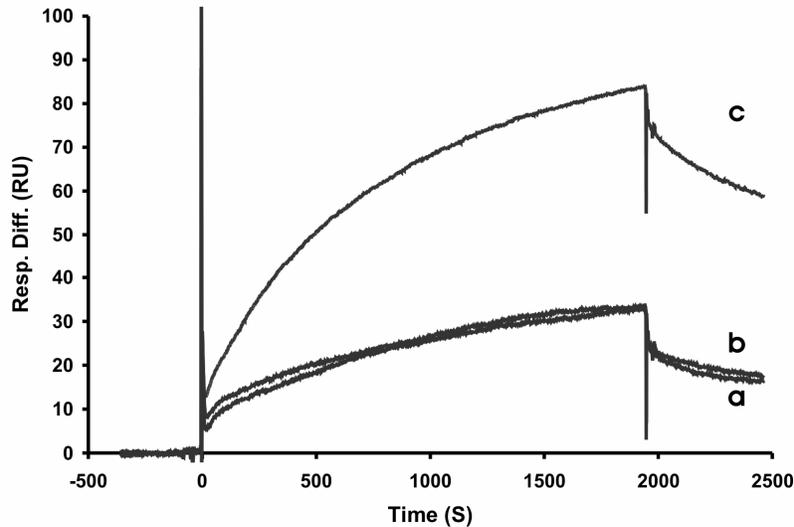
Figure 3.2.13 HCoV-229E Nsp9 binding to ssDNA analyzed by surface plasmon resonance, in the presence and absence of DTT
A 5′-biotinylated 50-mer oligonucleotide was immobilized on an SA chip up to 88 RU. (**a**) and (**b**) are the binding curves for 20 µM of a fresh preparation of the Nsp9 Cys69Ala mutant and for wild-type Nsp9, respectively, both injected in the presence of 5 mM DTT. (**c**) is the binding curve for 20 µM aged preparation of wild-type Nsp9, injected in the absence of 5 mM DTT.

3.3 TGEV Nsp9

3.3.1 Protein production

The vector consisting of the gene coding for the His-tag and the TGEV Nsp9 was provided through the VIZIER consortium (B. Coutard, Marseille). The N-terminally His-tagged protein was purified using a Ni-NTA column followed by gel-filtration. The protein exhibited an apparent molecular mass of about 14 kDa under denaturing conditions in SDS-PAGE (Fig. 3.3.1).
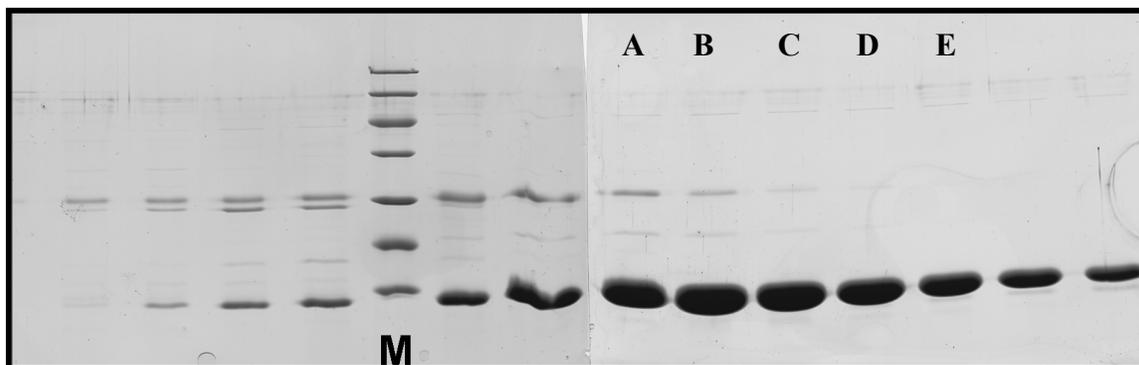
Fig. 3.3.1 Gel-filtration profile of TGEV Nsp9
M: Marker proteins with apparent molecular masses of 116 (most upper band), 66.2, 45, 35, 25, 18.4, 14.4 kDa (lowest band). 10 µl of TGEV Nsp9 solution from each fraction were loaded directly from the gel-filtration chromatography. Main peak fractions were labeled as A to E.

3.3.2 Crystallization of TGEV Nsp9

The protein solution containing the storage buffer (Table 2.5) was screened for potential crystallization conditions using commercially available screening kits from Sigma-Aldrich (basic and extension kits). Initial crystallization screening trials were carried out using a Phoenix robot (Dunn Labortechnik, Thelenberg, Germany) employing the sitting-drop vapor diffusion method. Initial hits were optimized by manually setting up 2-µl hanging drops consisting of 1 µl of protein and 1 µl of reservoir solution. Fragile and often inter-grown thin needles were observed from the screening kits. They were obtained in a wide range of pH, 5.6 - 8.5, with PEG 4000, 8000, and in many alcohols. Using only the gel-filtration main peak fractions (labeled as A to E in Fig. 3.3.1) was crucial for obtaining individual thick needles. The crystallization conditions were optimized as follows: 0.1 M HEPES pH 7.5, 11-12% iso-propanol, and 4% PEG 4000 at 20°C (Fig. 3.3.2).
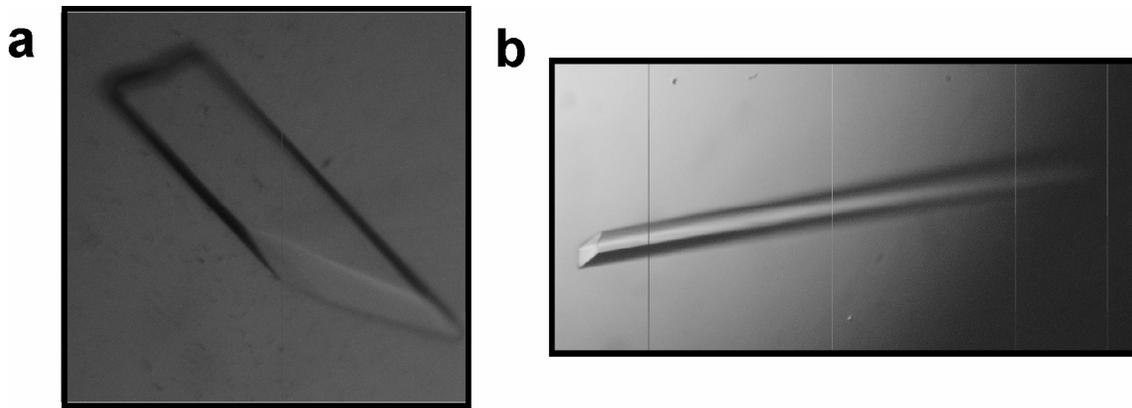
Fig. 3.3.2 Crystals of TGEV Nsp9
Three-dimensional crystals were obtained after one day. (a), (b) crystals of TGEV Nsp9 grown under similar crystallization conditions.

3.3.3 Structure elucidation

3.3.3.1 Data acquisition

Thick needles diffracted X-rays very weakly to a Bragg spacing of 6.0 Å with severe ice rings. Therefore, 30% of ethylene glycol in the precipitation solution was used as a cryoprotectant. Ethylene glycol was vital and could not be replaced with either isopropanol or glycerol as a cryoprotectant. The macromolecular crystal annealing protocol by Harp *et al.* (1998) was employed for extending the diffraction limit from 6.0 Å to 3.2 Å using the in-house X-ray source. Briefly, three steps were followed: first, the cryocooled crystal was removed from the cryogas stream; secondly, the crystal was re-equilibrated in the cryosolution for one minute and last, the crystal was recooled again in the cryostream and checked for the diffraction limit. The crystals were then cryo-cooled in liquid nitrogen, and X-ray diffraction data were measured to 2.41 Å at 100 K on beam line 14.2 at Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY), Berlin.

3.3.3.2 X-ray diffraction data

Crystals displayed space group I422 with unit cell dimensions a=b= 89.25 Å, c = 74.85 Å, and one monomer in the asymmetric unit. The average $I/\sigma(I)$ was 24.6 for the resolution range 44.63-2.41 Å and 5.2 in the highest resolution shell, with an overall completeness of 99.0%. A total of 56,655 measurements were made, representing 6,074 independent

reflections, with an R_{merge} of 7.9%. The Matthews coefficient for one molecule per asymmetric unit was 2.66 A^3/Da and the solvent content was 53.8% (Matthews, 1968). Diffraction data are summarized in Table 7.3 (Appendix).

3.3.3.3 Molecular replacement solution and quality of the structural models

The structure was determined by molecular replacement, using a monomer of the HCoV-229E Nsp9 Cys69Ala mutant (PDB code: 2j98; Ponnusamy *et al.,* 2008) as the search model. Truncation of strand β5 and flexible regions in the search model was required for a successful solution. TGEV Nsp9 consists of 111 amino-acid residues and the gene construct used coded for an additional 23 residues (including a His_6 sequence) at the N-terminus. All extra residues and the first residue of Nsp9 could not be modeled due to lack of electron density. The final R-factor for the structural model is 19.5% and the R_{free} is 25.0%; 96.3% of the amino-acid residues are in the most-favored region of the Ramachandran plot and the remainder in the additionally allowed region (Laskowski *et al.,* 1993). A Ramachandran plot for the TGEV Nsp9 structure is available in Fig. 7.3 (Appendix). Final model refinement statistics are listed in Table 7.3 (Appendix). During refinement of the structure, difference density for one ethylene glycol molecule and three chloride ions emerged.

3.3.4 Overall monomer structure

The overall monomer structure of TGEV Nsp9 is very similar to the known HCoV-229E and SARS-CoV Nsp9 structures (Fig. 3.3.3). The fold of the monomer is a variant of oligonucleotide/oligosaccharide-binding module (OB fold). In TGEV Nsp9, the β-barrel is made up of the first five antiparallel β-strands (β1- β5), flanked by a two extra β-strands (β6 and β7) forming a long hairpin (L67) parallel to the C-terminal α-helix.

The smallest overall monomer r.m.s. deviations of TGEV Nsp9 with its orthologues from HCoV-229E and SARS-CoV are 0.89 Å (for 89 C^α atoms) and 0.86 Å (for 93 C^α atoms), respectively (see Table 7.6, Appendix). Higher degrees of deviation between the TGEV Nsp9 and other homologous Nsp9s are seen in the β-barrel; four out of five antiparallel β-strands (β2 - β5) are positioned differently, in response, loops L23 and L45 open the β-

barrel much wider than other Nsp9s (Fig. 3.3.3). The tips of the L23 and L45 loops contain residues Ser36 and Asn59, respectively; the distance between the C^α atoms of these residues is 22.4 Å, whereas the average distance in other orthologues is 12.0 Å. The open β-barrel conformation observed in TGEV Nsp9 is due to crystal packing. The residues present in the loop L45 of the symmetry mate is negatively charged and packs into the positively charged β-barrel cleft (Fig. 3.3.4). Another major difference between TGEV Nsp9 and its orthologous can be seen in loop L12. TGEV Nsp9 L12 is two residues longer compared to HCoV-229E Nsp9 and the tip of L12 contains the acidic
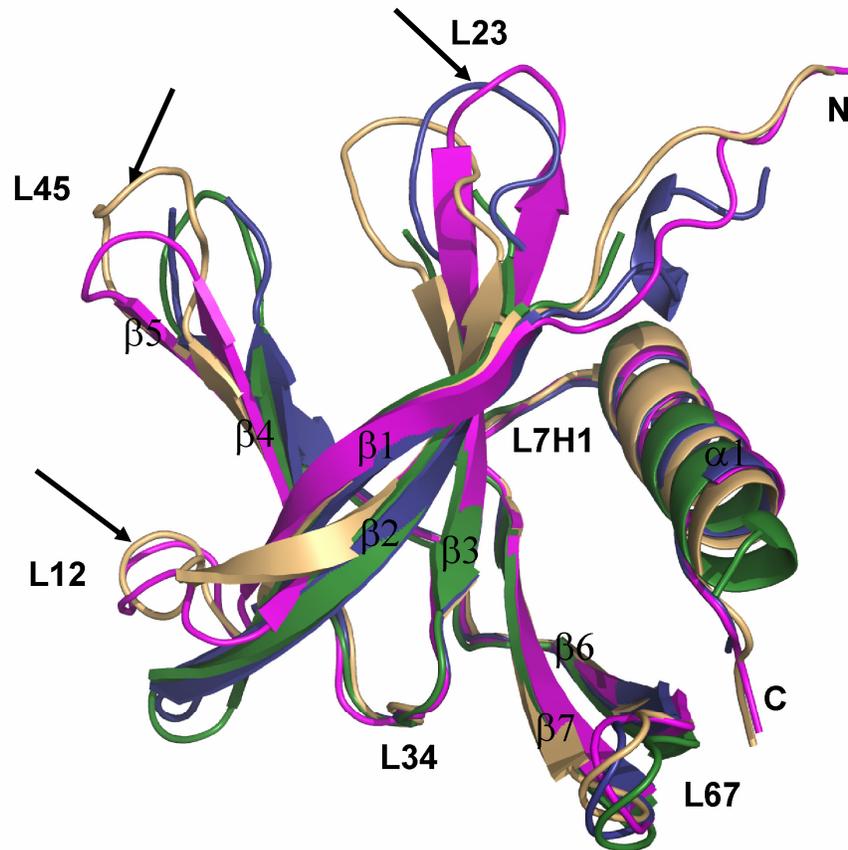
Fig. 3.3.3 Superimposition of TGEV Nsp9 with HCoV-229E and SARS-CoV Nsp9
Ribbon representation of TGEV Nsp9 (violet), HCoV-229E Nsp9 wild-type (green), HCoV-229E Nsp9 Cys69Ala mutant (blue), and SARS-CoV Nsp9 (pale orange; PDB code: 1QZ8) superimposed with C^α r.m.s deviations smaller than 1.0 Å (see Table 7.6, (Appendix) for detailed r.m.s deviations). A high degree of variation is seen in L12, L23 and L45.

residue Asp21, which is crucial for the formation of one dimer interface (see below). Variations in the sequence lengths among the coronaviral Nsp9s are mainly due to insertions in loop L12. SARS-CoV Nsp9 is also longer than HCoV-229E Nsp9 by two residues in L12 but the acidic residue is missing.
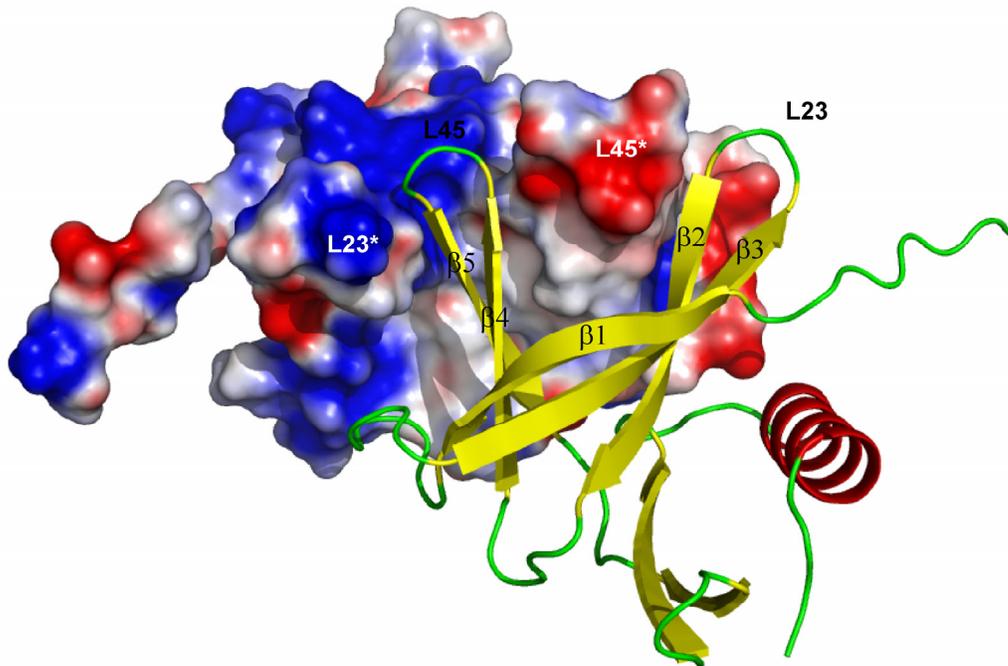
Fig. 3.3.4 TGEV Nsp9 monomer with its symmetry mate
Ribbon representation of TGEV Nsp9 monomer; its loop L45 is inserted into the β-barrel cleft of the symmetry-related molecule displayed according to electrostatic surface potential. The surface is colored according to the electrostatic potential ranging from deep blue (positive charge +10 k_BT) to red (negative charge -10 k_BT). The electrostatic potential was calculated using APBS tools implemented in the Pymol software (Baker *et al.,* 2001).

One ethylene glycol and three chloride ions were modeled into the electron density map of TGEV Nsp9. Ethylene glycol locates near the two-fold symmetry axis close to the β5 strand. Therefore, the hydroxyl group of ethylene glycol comes close to its own symmetry mate and forms a hydrogen bond. The same hydroxyl group makes another hydrogen bond to the main-chain amide (residue 66). The other hydroxyl group of ethylene glycol makes two more hydrogen bonds to the hydroxyl group of Tyr53 present in the same molecule and to the main-chain carbonyl atom (residue 90) of the symmetry related molecule (Fig. 3.3.5). Ethylene glycol fills the cavity near the β5 strands and

makes several interactions with the protein molecule within the same asymmetric unit and also with a symmetry-related molecule, thereby acting as a bridge and allowing the β5 strands to adopt specific positions. The cleft created by the β5 strands is narrow and only ethylene glycol can fill the cavity. This observation nicely explains why ethylene glycol was crucial as a cryoprotectant.
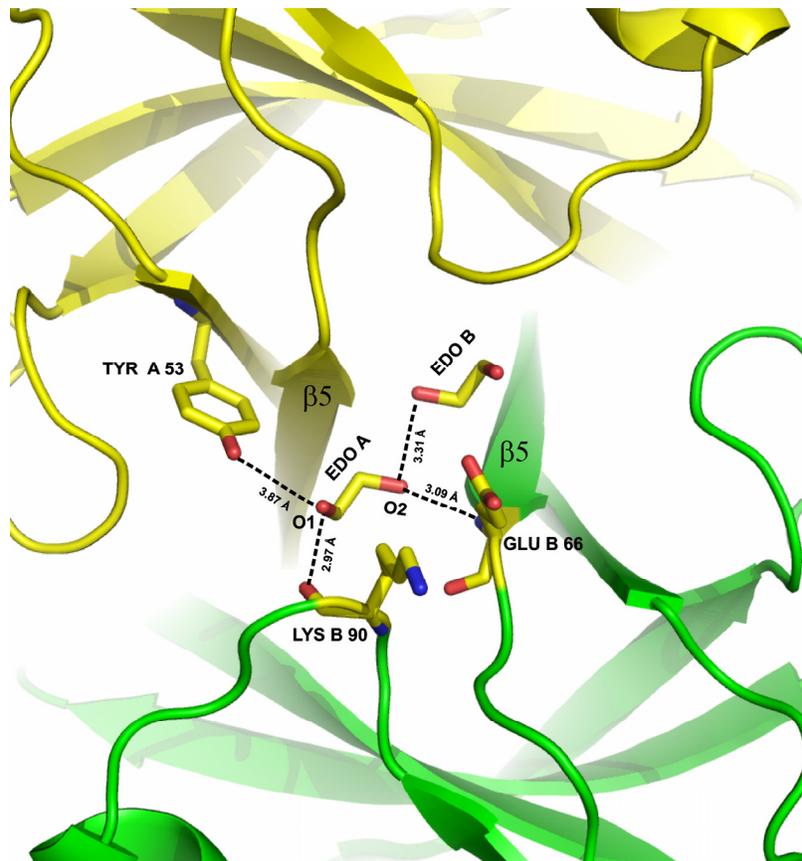
Fig. 3.3.5 Ethylene glycol bound to TGEV Nsp9
A cross-section of an interface with the ethylene glycol making several hydrogen bonds indicated by broken lines. The ethylene glycol was used as a cryoprotectant; it fills the space between the β5-strands.

3.3.5 TGEV Nsp9 dimer structure

Crystal packing in TGEV Nsp9 reveals two potential dimer interfaces. The first of these is mainly mediated by the C-terminal α-helix and N-terminal residues. The 23 additional tag residues at the N-terminus and the first residue of the protein molecule do not show any visible electron density and are expected to be flexible. The second residue, Asn2, which is completely conserved among coronaviruses, makes two hydrogen bonds with

the main-chain atoms of residues 70 and 72 (β-strand 6). Yet another hydrogen bond is formed between the main-chain amide of Ile4 and the carbonyl atom of residue 72 (β-strand 6). A fourth hydrogen bond is between the side-chain atom of Lys8 (N^ζ) and the main-chain carbonyl of residue 106 (α-helix). Because of the crystallographic two-fold symmetry, all of these interactions are duplicated in the dimer, so in total there are eight intersubunit hydrogen bonds in this interface. All eight hydrogen bonds are made by the N-terminal residues. There are also several hydrophobic residues buried due to this dimerization mode; N-terminal hydrophobic residues Ile4, Pro6 and Gly7 lie over the β6 strand and the C-terminal α-helix of the symmetry mate. The C-terminal α-helices of the two monomers bury a number of hydrophobic residues by crossing at an angle of 48° with a closest approach of 3.97 Å between the C^α atoms of Ala99 and Gly102 (Fig. 3.3.6 (b)). This interface buries a surface area of 1110 $Å^2$ per monomer, with a shape complementarity of 0.71 (Lee & Richards, 1971; Lawrence & Colman, 1993). A similar dimerization mode is also seen in the other homologous Nsp9 structures from SARS-CoV and HCoV-229E Cys69Ala mutant (Egloff *et al.,* 2004; Sutton *et al.,* 2004; Ponnusamy *et al.,* 2008).
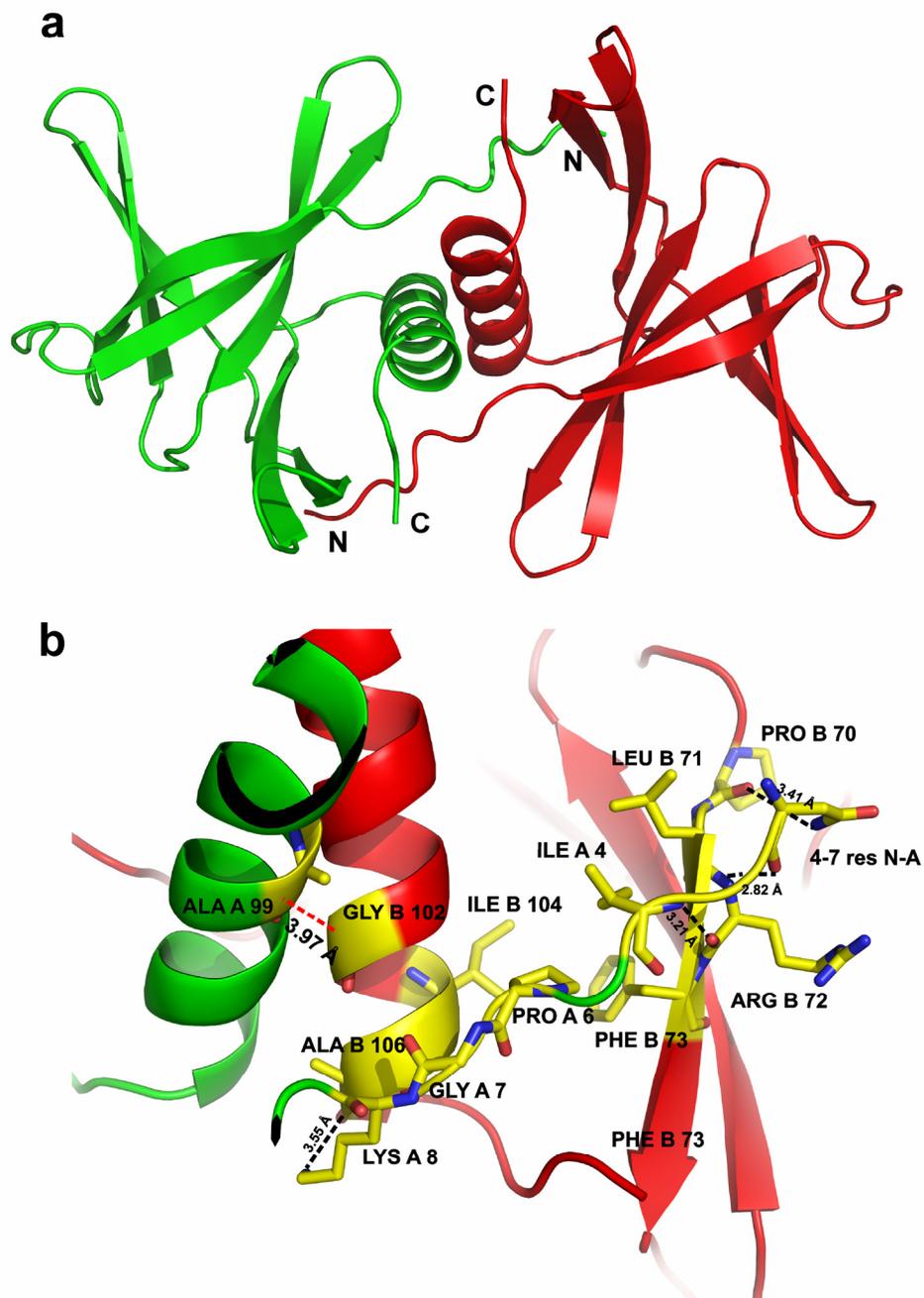
Fig. 3.3.6 Structural features of the TGEV Nsp9 helix-mediated dimer
The two monomers are colored in green and red respectively. (a) Ribbon representation of the TGEV Nsp9 homodimer. (b) A cross-section of the dimer interface with residues involved in interface formation shown using sticks and the few hydrogen bonds indicated by broken lines. The closest C^α-C^α approach of the helices is indicated by the broken orange line.

The second dimer interface seen in the TGEV Nsp9 crystal structure is unique and not observed in the other homologous structures (Fig. 3.3.7). This interface is mediated by several hydrogen bonds and other electrostatic interactions. The β45 sheet comprises several negatively charged residues and inserts into the positively charged β-barrel (Fig. 3.3.8). Therefore, several salt bridges and hydrogen bonds are made by the residues present in L45. One of the crucial interactions is the salt bridge between Arg12 and Asp61; however, these residues are not conserved among coronaviruses. Another salt bridge exists between $Arg96N^{\eta 2}$ and $Asp21O^{\delta 1}$. Again Arg96 is present only in some coronaviruses, whereas Asp21 is conserved in most of these except IBV and SARS-CoV, in which this amino-acid residue is replaced by Glu and Thr, respectively. Also, there are five hydrogen bonds observed in this interface. The side-chain atom $N^{\delta 2}$ of Asn93 donates a hydrogen bond to $O^{\delta 2}$ of Asp21. A second hydrogen bond is present between the main-chain atoms of residues 62 (L7H) and 39 (L12). All these interactions are duplicated due to the two-fold crystallographic symmetry. The last hydrogen bond in this interface is created directly between the symmetry-related side-chain atoms O^{γ} of Ser58. Again, Ser58 is not conserved among the coronaviruses. The buried dimer surface area of this interface is 955 $Å^2$ per monomer, with a shape complementary of 0.57 (Lee & Richards, 1971; Lawrence & Colman, 1993). Although there are several interactions in this interface, the residues involved in the interface are not conserved among the coronaviruses.
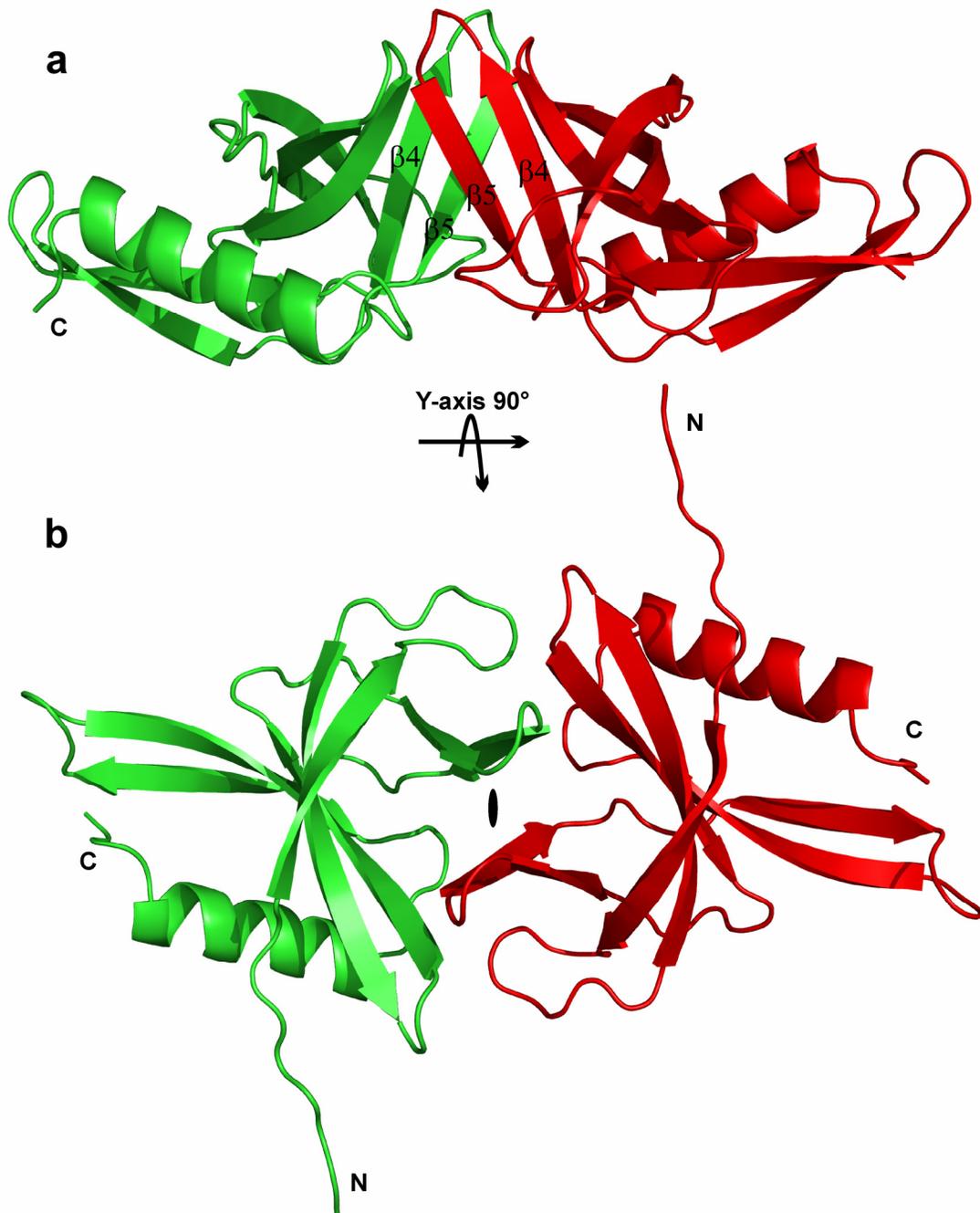
Fig. 3.3.7 Cartoon representation of the TGEV Nsp9 β-sheet mediated dimer
The two monomers are colored in green and red, respectively. (a) β45-sheets crossing each other in a parallel orientation and thereby making extensive salt bridges and hydrogen bonds between the monomers. (b) Top-view of the above picture. The two-fold axis is running in-between the sheets.
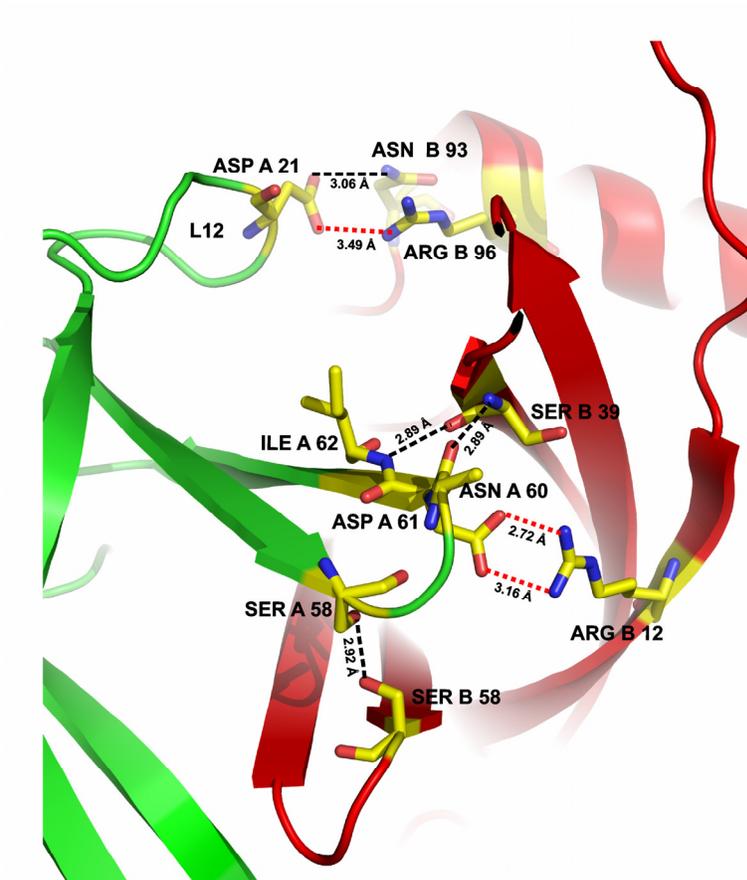
Fig. 3.3.8 Residues involved in the β-sheet-mediated dimer interface
A cross-section of the dimer interface with residues involved in the interface is shown
using sticks. Hydrogen bonds are shown as black dotted lines and salt bridges as red
dotted lines.

3.3.6 Higher oligomers in the crystal of TGEV Nsp9

In the crystal structure of TGEV Nsp9, polymers are formed by the two different dimer
interfaces explained above. The polymers run in a parallel orientation with a two-fold
symmetry axis forming a double-stranded helix with pitch of ~150 Å and diameter of ~80
Å. Per turn, this right-handed super helix comprises eight monomers in each strand (Fig.
3.3.9).
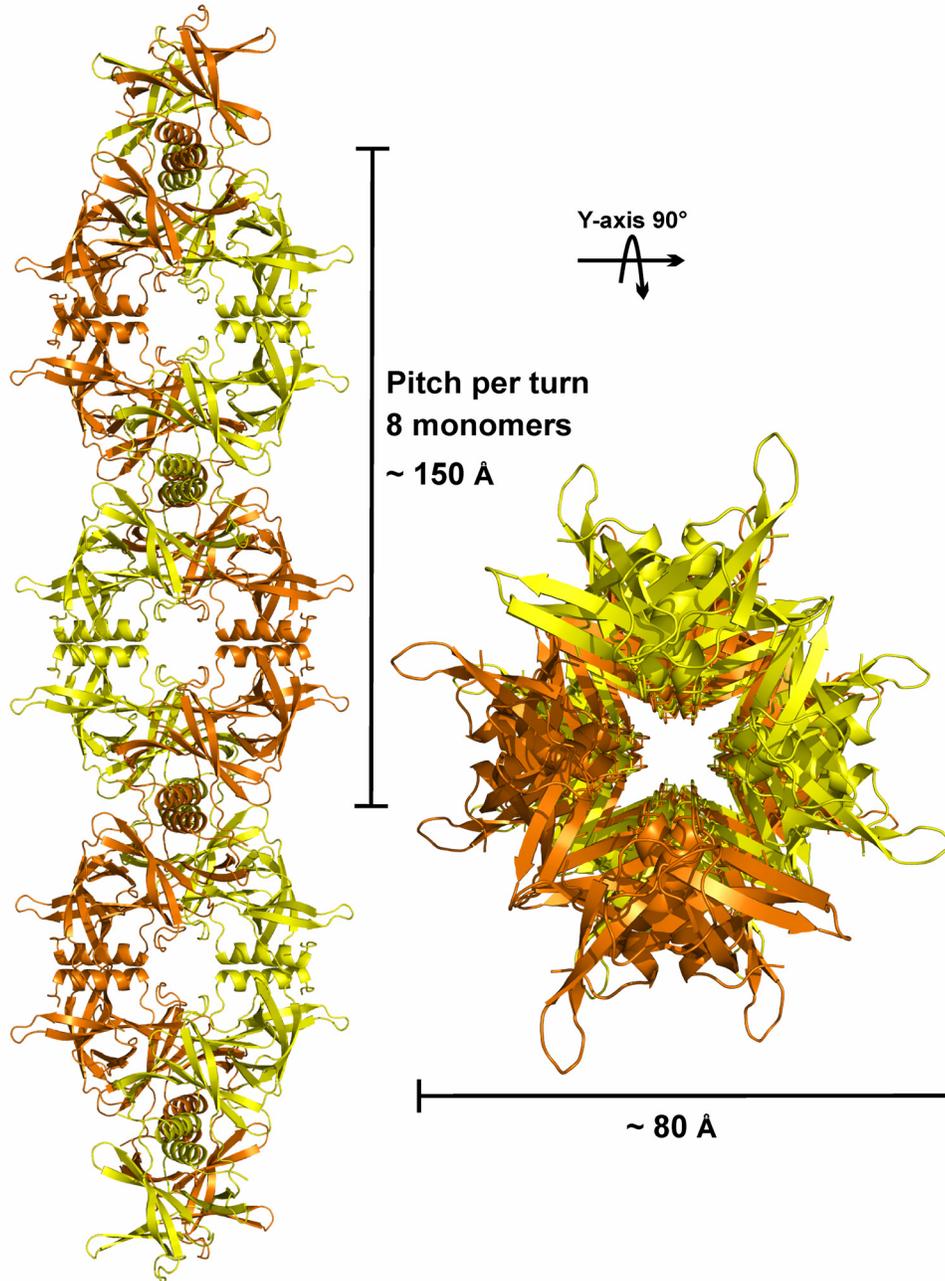
Fig. 3.3.9 TGEV Nsp9 forms a double-stranded super-helix in the crystal
TGEV Nsp9 oligomerizes in the crystal using two different dimer interfaces. The oligomers form a double-stranded helix with two polymers of TGEV Nsp9 running in parallel orientation; shown in yellow and orange.

Table 3.3.1: Protein-protein interfaces seen in the TGEV Nsp9 structure: intermolecular interactions, buried dimerization surface area, and shape complementarity values

Dimer interface	Hydrogen bonds	Other interactions	BSA*	Sc†
α-helix	(N-ter) Asn 2 O… (β6) Arg 72 N (2.82 Å) (N-ter) Asn 2 Nδ2 ... (L56) Pro 70 O (3.41 Å) (N-ter) Ile 4 N ... (β6) Arg 72 O (3.21 Å) (N-ter) Lys 8 Nζ ... (α-helix) Ala 106 O (3.55 Å)	Hydrophobic interactions (N-ter) Ile 4, Pro 6, Gly 7 … (β6) Leu 71, Arg 72, Phe 73 (α-helix) Ala 106, Leu 103. (α-helix) … (α-helix) (Several hydrophobic residues)	1110	0.71
β-sheet	(β3) Ser 39 N … (L45) Asn 60 O (2.89 Å) (L45) Ser 58 Oγ … (L45) Ser 58 Oγ (2.92 Å) (β5) Ile 62 N ... (β3) Ser 39 O (2.89 Å) (L7H1) Asn 93 Nδ2 ... (L12) Asp 21 Oδ2 (3.06 Å)	Salt bridges (β1) Arg 12 Nη1 … (β5) Asp 61 Oδ1 (2.72 Å) (β1) Arg 12 Nη2 … (β5) Asp 61 Oδ2 (3.16 Å) (α-helix) Arg 96 Nη2 … (L12) Asp 21 Oδ1 (3.49 Å)	955	0.57

* BSA - Buried dimerization surface area per monomer (Å^2) † Sc - Shape complementarity

3.3.7 Oligomeric state in solution

TGEV Nsp9 was proven to be a dimer in solution using Dynamic Light Scattering (DLS) and gel-filtration chromatography. In DLS, the purified TGEV Nsp9 showed a monodisperse peak centered at a hydrodynamic radius of 26 (± 1.6) Å, indicating that the dimer was the prevalent species in solution. A similar result was also obtained by size-exclusion chromatography, which showed a single peak corresponding to a molecular mass of ~ 28 kDa (molecular mass was calculated using five different marker proteins; data not shown).

TGEV Nsp9 also oligomerizes in a concentration-dependent manner with the chemical crosslinker glutaraldehyde, which was used with protein concentrations ranging from 10 µM to 100 µM. The TGEV Nsp9 showed crosslinking products corresponding to dimers, trimers, tetramers, and higher oligomers as indicated by SDS-polyacrylamide gel. This pattern did not change in the presence of 13-mer, 35-mer or 40-mer ssDNA of random sequence (data not shown). Similar oligomerization with the help of glutaraldehyde was also seen with SARS-CoV and HCoV-229E Cys69Ala mutant Nsp9s (Ponnusamy *et al.,* 2008; cf. Fig. 3.2.10 in section 3.2.7)
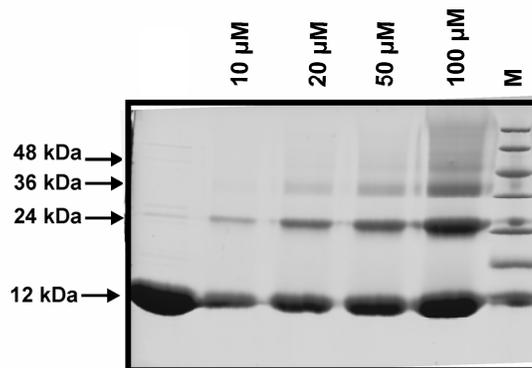
Fig. 3.3.10 TGEV Nsp9 crosslinking using glutaraldehyde
Crosslinking was carried out with different protein concentrations from 10 µM to 100 µM using 0.01% of glutaraldehyde. Respective molecular mass is labeled with arrows. Higher oligomers are seen with the highest protein concentration. M: Marker proteins with apparent molecular masses of 116 (most upper band), 66.2, 45, 35, 25, 18.4, 14.4 kDa (lowest band).

4. Discussion

4.1 Nsp8 contribution to viral replication

Nsp8 is indispensable in the coronavirus life cycle (Deming *et al.,* 2007). The crystal structure of SARS-CoV Nsp7-8 revealed a hexadecameric supercomplex (Zhai *et al.,* 2005). Recently, Imbert *et al.* (2006) have shown that SARS-CoV Nsp8 possesses primase activity. Moreover, SARS-CoV Nsp8 seems to interact with an unusually large number of interaction partners (Prentice *et al.,* 2004; von Brunn *et al.,* 2007; Kumar *et al.,* 2007; Imbert *et al.,* 2008; Pan *et al.,* 2008). Hence, Nsp8 is apparently essential for the coronavirus. However, biophysical and biochemical information on Nsp8 alone, without its interaction partners, is largely unavailable. Therefore, herein a crystallographic and biochemical approach is undertaken to characterize and compare the Nsp8 from SARS-CoV (group II) and HCoV-229E (group I) and seek to understand their role in the coronavirus life cycle.

Despite little sequence identity between SARS-CoV and HCoV-229E Nsp8, the two proteins are very similar in their biophysical characteristics as probed by electrophoresis, fluorescence spectroscopy and dynamic light-scattering. Both the proteins bind 1,8-ANS, a fluorescent compound probing for hydrophobic binding sites. The C-terminal part of the protein sequence is largely hydrophobic and folded (Fig. 4.1 (a)), thus it is tempting to predict that both the SARS-CoV and HCoV-229E Nsp8 bind 1,8-ANS in this region. Both the proteins bind nucleic acids with a similar affinity and sequence specificity.

In the crystal structure, SARS-CoV Nsp7 and Nsp8 form a hexadecamer, in which eight Nsp7 and eight Nsp8 molecules form a supercomplex with two different conformations of Nsp8 (Zhai *et al.,* 2005). This supercomplex forms a central positively charged cylinder-like structure with an internal diameter of ~30 Å, easily capable to encapsulate dsRNA. The authors propose that the Nsp7-8 complex might confer processivity to the synthesis of large RNAs in coronavirus Nsp12 RdRp-mediated replication and transcription. However, Imbert *et al.* (2006) have shown that SARS-CoV Nsp8 is a polymerase and able to synthesize <6 nucleotides, like a primase. These primers might be

utilized by Nsp12 for primer-dependent RdRp activity (Cheng *et al.,* 2005; Imbert *et al.,* 2006). Addition of SARS-CoV Nsp7 does not influence the primase activity of Nsp8.

The crystal structure of SARS-CoV Nsp7-8 revealed the interaction between Nsp7 and Nsp8 (Zhai *et al.,* 2005). However, Nsp8 interacts with several other non-structural proteins, Nsp2 and Nsp3 (Prentice *et al.,* 2004); Nsp2, Nsp5, Nsp6, Nsp7, Nsp8, Nsp9, Nsp12, Nsp13 and Nsp14 (von Brunn *et al.,* 2007); ORF6 (Kumar *et al.,* 2007); Nsp2, PLP2, Nsp5, Nsp12, Nsp13, Nsp14, Nsp15 and Nsp16 (Imbert *et al.,* 2008); Nsp3 (C-terminal region) and ORF 3b (Pan *et al.,* 2008); Nsp7 and Nsp9 (this study). These interactions were demonstrated by different techniques, and it will be difficult to evaluate which of these interactions are biologically relevant, because of the limitation of the techniques used. Nevertheless, it is apparent that Nsp8 has a central role in the coronavirus life cycle. In addition to the primase activity, Nsp8 acts as a hub by its promiscuous binding with several Nsps, most likely at different stages of the coronavirus life cycle.

4.1.1 Is Nsp8 intrinsically disordered?

How does Nsp8 interact with unusually large number of interaction partners? FoldIndex (Prilusky *et al.,* 2005) and PONDR (Romero *et al.,* 2001) (programs to identify intrinsically unfolded regions) predicted that SARS-CoV and HCoV-229E Nsp8s have a long disordered region of about 40 residues covering the sequence residues 40-80 within the 198 residues (Fig. 4.1). In the PONDR analysis, this region is depicted as an α-MoRE region. The α-MoRE regions are characterized as intrinsically disordered and forming a helix upon binding to their interaction partners (Oldfield *et al.,* 2005). As predicted by PONDR analysis, the α-MoRE region of Nsp8 residues 40-80 displays two conformations in the crystal structure of the SARS-CoV Nsp7-Nsp8 supercomplex. One conformation of this region is completely unfolded, whereas the other one forms a helix (Fig. 4.2). The spatial arrangement of 16 monomers (Nsp7 and Nsp8 – each eight monomers) becomes possible because of the unfolded region in Nsp8. The two Nsp8 conformations seen in the crystal structure are equally important for the formation of the unique hexadecameric supercomplex with the central positively charged channel
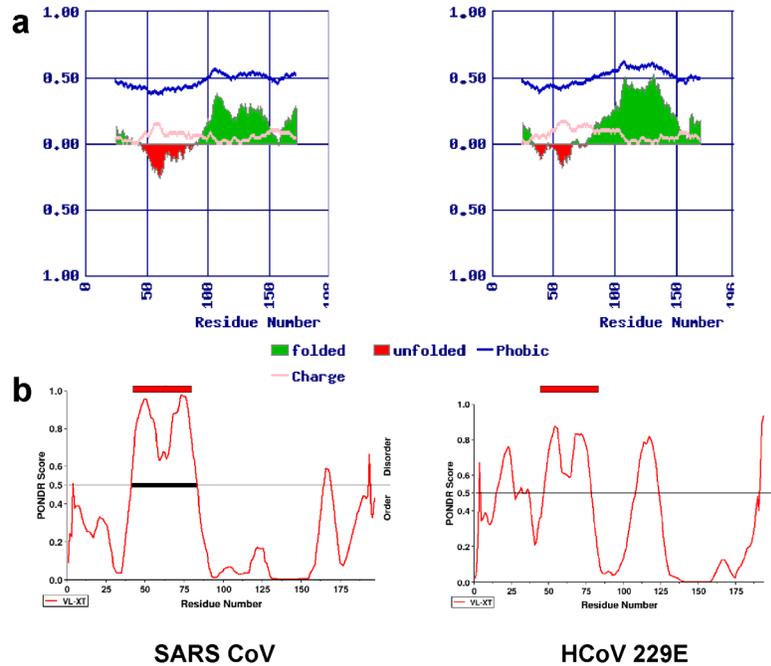
Fig. 4.1 FoldIndex and PONDR prediction of Nsp8
Nsp8 from SARS-CoV and HCoV-229E showed signals for the presence of disordered regions when using the programs FoldIndex and PONDR. (a) FoldIndex predicted that the Nsp8 region around residues 40-80 is disordered, whereas the C-terminal domain is folded and hydrophobic (shown by the blue line). (b) PONDR analysis suggested that the same region around residues 40-80 is disordered and has a signature for α-MoRE indicated by the red bar.

favourable for nucleic-acid binding. This region of the Nsp8 sequence is more conserved among the coronaviruses compared to the other regions (Fig. 4.3). CD spectroscopy results suggest that the protein exhibits 60% "loops" (without any secondary structure). In 1D-NMR studies, SARS-CoV Nsp8 did not show any signals beyond 8.5 ppm and below 0 ppm, indicating that the protein is either aggregated or partially disordered. Furthermore, Nsp8 protein is completely digested within 5 min by trypsin treatment (not shown), demonstrating that the protein is disordered and flexible. Altogether, based on the above experiments, it is evident that part of the Nsp8 is intrinsically unfolded or disordered. Its disordered state is the key feature, which enables the protein to interact with several interaction partners.
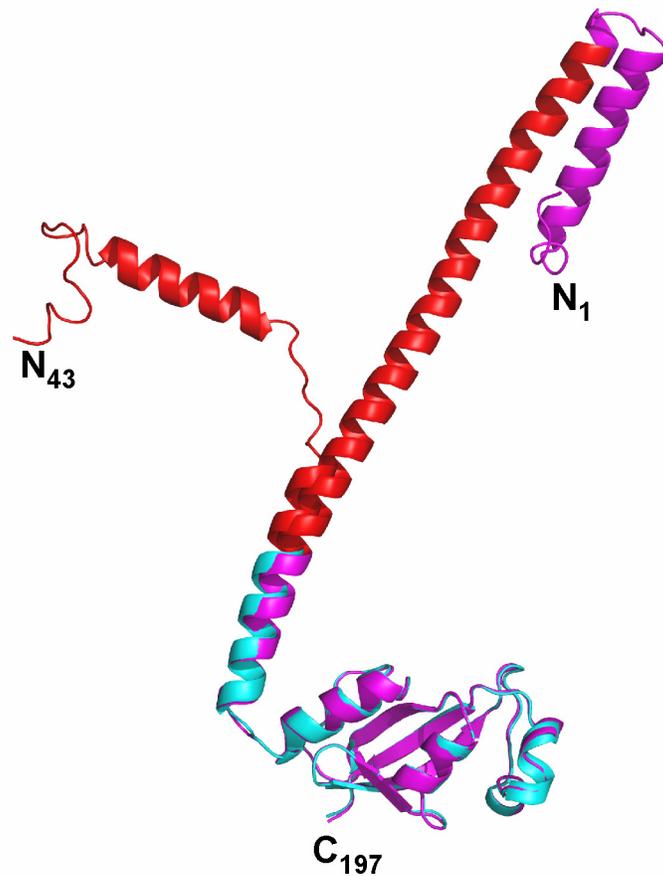
Fig. 4.2 Ribbon representation of SARS-CoV Nsp8
Two different conformations were superimposed in two different colors, cyan and magenta. The α-MoRE region is colored red, where in one conformation the α-MoRE region was unfolded and in the other conformation, the α-MoRE region forms a helix. The N-terminal residues 1-42 of one conformation were flexible and could not be identified due to the crystal packing (Zhai *et al.,* 2005).

It is not very odd for a viral protein to be intrinsically disordered; in-fact, there are several examples for the presence of long disordered regions in eukaryotic and viral proteins (Chen *et al.,* 2006a; 2006b). It must be emphasized that the majority of intrinsically disordered proteins undergo a disorder-to-order transition upon functioning (Pontius, 1993; Spolar & Record 1994; Rosenfeld *et al.,* 1995; Plaxco & Gross 1997; Dunker *et al.,* 1997, 2001; Wright & Dyson 1999), and are able to bind several different targets (Wright and Dyson 1999; Dunker *et al.,* 2001; Romero *et al.,* 1998b). An advantage of intrinsically disordered regions is to overcome steric restrictions, enabling

essentially larger interaction surfaces in the complex than from rigid partners (Meador *et al.,* 1992; Choo and Schwabe, 1998; Dunker *et al.,* 2001).

```
                    η1          α1                                                                              α2
SARS-CoV/1-198      1        10        20          30           40         50        60

SARS-CoV/1-198    AIASEFSSLPSYAAYATAQEAYEQAVA...NGDS.EVVLKKLKKSLNVAKSEFDRDAAMQRKLEKMADQA
229E/1-195        SVASSFVGMPSFVAYETARQEYENAVA...NGSS.PQIIKQLKKAMNVAKAEFDRESSVQKKINRMAEQA
NL63/1-195        SVASSFVSMPSYIAYENARQAYEDAIA...NGSS.SQLIKQLKRAMNIAKSEFDHEISVQKKINRMAEQA
PEDV/1-195        SVASTYVGLPSYVIYENARQQYEDAVN...NGSP.PQLVKQLRHAMNVAKSEFDREASTQRKLDRMAEQA
FCoV/1-195        SVASAYAALPSWIAYEKARADLEEAKK...NDVS.PQLLKQLTKACNIAKSEFEREASVQKKLDKMAEQA
OC43/1-197        ALQSEFVNMASFVEYEVAKKNLDEARF...SGSANQQQLKQLEKACNIAKSAYERDRAVAKKLERMADLA
BCoV/1-197        ALQSEFVNMASFVEYEVAKKNLDEARS...SGSANQQQLKQLEKACNIAKSAYERDRAVARKLERMADLA
MHV-A59/1-197     ALQSEFVNMASFVEYELAKKNLDEAKA...SGSANQQQIKQLEKACNIAKSAYERDRAVARKLERMADLA
HKU-1/1-197       ALQSEFVNMASFVEYEVAKKNLADAKN...SGSVNQQQIKQLEKACNIAKSVYERDKAVARKLERMADLA
TGEV/1-210        SVTQEFSYIPSYAEYERAKNLYEKVLADSKNGGVTQQELAAYRKAANIAKSVFDRDLAVQKKLDSMAERA
IBV/1-210         SVTQEFSHIPSYAEYERAKNLYEKVLVDSKNGGVTQQELAAYRKAANIAKSVFDRDLAVQKKLDSMAERA

                                                              α3        β1            β2
SARS-CoV/1-198      70        80        90       100        110       120       130

SARS-CoV/1-198    MTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNIIPLTTAAKLMVVVPDYG
229E/1-195        AAAMYKEARAVNRKSKVVSAMHSLLFGMLRRLDMSSVDTILNMARNGVVPLSVIPATSAARLVVVVPDHD
NL63/1-195        ATQMYKEARSVNRKSKVISAMHSLLFGMLRRLDMSSVETVLNLARDGVVPLSVIPATSASKLTIVSPDLE
PEDV/1-195        AAQMYKEARAVNRKSKVVSAMHSLLFGMLRRLDMSSVDTILNLAKDGVVPLSVIPAVSATKLNIVTSDID
FCoV/1-195        AASMYKEARAVDRKSKIVSAMHSLLFGMLKKLDMSSVNTIIEQARNGVLPLSIIPAASATRLIVVTPNLE
OC43/1-197        LTNMYKEARINDKKSKVVSALQTMLFSMVRKLDNQALNSILDNAVKGCVPLNAIPSLAANTLNIIVPDKS
BCoV/1-197        LTNMYKEARINDKKSKVVSALQTMLFSMVRKLDNQALNSILDNAVKGCVPLNAIPSLAANTLTIIVPDKS
MHV-A59/1-197     LTNMYKEARINDKKSKVVSALQTMLFSMVRKLDNQALNSILDNAVKGCVPLNAIPSLTSNTLTIIVPDKQ
HKU-1/1-197       LTNMYKEARINDKKSKVVSALQTMLFSMVRKLDNQALNSILDNAVKGCVPLSAIPALAANTLTIVIPDKQ
TGEV/1-210        MTTMYKEARVTDRRAKLVSSLHALLFSMLKKIDSEKLNVLFDQASSGVVPLATVPIVCSNKLTLVIPDPE
IBV/1-210         MTTMYKEARVTDRRAKLVSSLHALLFSMLKKIDSEKLNVLFDQASSGVVPLATVPIVCSNKLTLVIPDPE

                  α4  β3    β4        β5              η2      β6
SARS-CoV/1-198      140       150       160       170        180       190

SARS-CoV/1-198    TYKNTCDGNTFTYASALWEIQQVVDADSKIVQLSEINM.....DNSPNLAWPLIVTALRANS.A...VKLQ
229E/1-195        SFVKMMVDGFVHYAGVVWTLQEVKDNDGKNVHLKDVTK.....ENQEILVWPLILTCERV......VKLQ
NL63/1-195        SYSKIVCDGSVHYAGVVWTLNDVKDNDGRPVHVKEITK.....ENVETLTWPLILNCERV......VKLQ
PEDV/1-195        SYNRIQREGCVHYAGTIWNIIDIKDNDGRPVHVKEVTA.....QNAESLSWPLVLGCERI......VKLQ
FCoV/1-195        VLSKVRQENNVHYAGAIWSIVEVKDANGAQVHLKEVTA.....ANELNITWPLSITCERT......TKLQ
OC43/1-197        VYDQVVDNVYVTYAGNVWQIQTIQDSDGTNKQLNEISD.....DCN....WPLVIIANRYNE.VSATVLQ
BCoV/1-197        VYDQVVDNVYVTYAGNVWQIQTIQDSDGTNKQLNEISD.....DCN....WPLVIIANRHNE.VSATVLQ
MHV-A59/1-197     VFDQVVDNVYVTYAGNVWHIQFIQDADGAVKQLNEIDV.....NST....WPLVIAANRHNE.VSTVVLQ
HKU-1/1-197       VFDKVVDNVYVTYAGSVWHIQTVQDADGINKQLTDISV.....DSN....WPLVIIANRYNE.VANAVMQ
TGEV/1-210        TWVKCVEGMHVTYSTVVWNIDTVIDADGTELHPTSVGSGLTYCISSDNIAWPLKVNLTRNGHNKVDAALQ
IBV/1-210         TWVKCVEGVHVTYSTVVWNIDTVIDADGTELHPTSTGSGLTYCISGANIAWPLKVNLTRNGHNKVDVVLQ
```

Fig. 4.3 Multiple sequence alignment of SARS-CoV Nsp8
Alignment was done using CLUSTALW (Thomson *et al.,* 1994) for SARS-CoV Nsp8 with its homologous coronavirus proteins. Residues boxed in red are completely conserved. Secondary structures were denoted using the SARS-CoV Nsp7-8 supercomplex (pdb code 2AHM; chain G) (Zhai *et al.,* 2005). The α-MoRE region predicted using the PONDR analysis is labeled as a red line.

In summary, Nsp8 from SARS-CoV (group II) and HCoV-229E (group I) coronavirus are very similar to one another and correlate very well with their biophysical and biochemical characteristics. Strikingly, both the proteins show the feature of an intrinsically disordered protein. The 1D-NMR studies proves that the SARS-CoV Nsp8 is partly unfolded. Both the proteins form large aggregates and exist in different oligomeric states when they are not interacting with their partners. The proteins bind to nucleic acids and several other Nsps. Based on the above results, the Nsp8s are likely intrinsically disordered, which facilitates the Nsp8s to interact with several other Nsps and nucleic acids. Their disordered regions are expected to become ordered upon binding to RNA and/or their interaction partners.

Due to the importance and conservation of Nsp8 in coronaviruses, the Nsp8 would be a very intriguing anti-viral target. The hot spot residues "40-80" in the protein sequence described above are conserved, intrinsically disordered and expected to be vital for the protein-RNA/protein interaction. Consequently, targeting this region of the sequence for anti-viral compounds would be very effective. One of the significances to probe the intrinsically disordered region is its weak interactions per unit of surface area, because the binding energy gained over the complex formation is spent to organize the disordered partner. Therefore, a small molecule could easily compete with the large protein-protein interaction surface. There are very few examples in the literature for probing the intrinsically disordered region, for instance in p53-MDM2 interaction, MDM2 binds to a disordered region of p53, and in turn the disordered region forms an amphipathic helix (Kussie *et al.,* 1996). Targeting this interaction, Vassilev *et al.* (2004a, 2004b) have demonstrated that series of small molecules called nutlins, appear to act by mimicking the shape and physicochemical characteristics of a disordered p53 protein fragment after it has transitioned to an ordered helical structure, as a result of interaction with its binding partner MDM2. Similar to p53-MDM2, characteristic interaction features are screened at entire human proteome level and several new targets have been identified (Cheng *et al.,* 2006). In coronaviruses, Nsp8 would be the ideal target and its intrinsically disordered region could be targeted by new anti-viral compounds.

4.2 Nsp9 contribution to viral replication

Nsp9 is one of the 15-16 Nsps produced from the cleavage of the polyproteins pp1a and pp1ab. Nsp9 is present in the C-terminal region of pp1a, among a set of several small Nsps (Nsp7-Nsp10). The proteins Nsp7-Nsp10 are highly conserved among the coronaviruses, but their exact role in the coronavirus life cycle is largely unknown. The crystal structures of SARS-CoV Nsp9 were solved by two independent groups (Egloff *et al.,* 2004; Sutton *et al.,* 2004), and proven to bind ssRNA/DNA without any sequence specificity. Nsp9 from MHV-A59 has been shown through immunofluorescence studies to colocalize with, among others, the helicase (Nsp13), the nucleocapsid protein (N-protein), and M^{pro} (Nsp5) (Bost *et al.,* 2000; Brockway *et al.,* 2003). Nsp9 also co-localizes to late endosomes at sites of replication with Nsp7, Nsp8 and Nsp10, and likely is a member of replication complex (Bost *et al.,* 2000). A Nsp9 knockout in MHV-A59 is not viable, mutation in the cleavage site of Nsp9-10 leads to the fusion of a Nsp9-10 polyprotein that is viable, but attenuated in growth, suggesting that the mature form of Nsp9 plays a critical role in viral replication (Deming *et al.,* 2007). Recently, Züst *et al.* (2008) have shown for the first time that Nsp8 and Nsp9 are cis-acting genomic RNA elements. All together, Nsp9 is unequivocally one of the putative components of the replication complex. However, there are no structural or biochemical data available for Nsp9 of group I and group III coronaviruses. Therefore herein, a crystallographic and biochemical analysis is carried out with the Nsp9 from HCoV-229E and TGEV of group I coronaviruses and compared with Nsp9 of SARS-CoV, a group II coronavirus.

The crystal structure of HCoV-229E Nsp9 reveals a dimerization mode very different from the one previously seen for SARS-CoV Nsp9 (Sutton *et al.,* 2004; Egloff *et al.,* 2004), in spite of a sequence identity of 45% between the two proteins. In HCoV-229E Nsp9, dimerization is mediated by a disulfide bridge, a few hydrogen bonds, and by hydrophobic interactions between the C-terminal helix of each monomer.

One major difference between the preparation of HCoV-229E Nsp9 and that of SARS-CoV Nsp9 by both Egloff *et al.* (2004) and Sutton *et al.* (2004) is that HCoV-229E protein has an authentic chain termini, whereas the SARS-CoV protein used by those

authors has N-terminal extensions from the cloning procedure (a hexahistidine tag in case of the Egloff *et al.* structure, and an extra 30 residues in the Sutton *et al.* structure). Interestingly, the differences of the N-terminal extensions between the two reported structures of SARS-CoV Nsp9 lead to deviations in the dimer in detail, resulting in a rather high r.m.s. deviation of 2.07 Å (for C^α atoms) between the two models. Residues (-7) to (-2) of the N-terminal tag present in the Sutton *et al.* structure form an extra antiparallel β-sheet with residues 3 to 8 of the protein, thereby pushing away the β6-β7 hairpin (L67) and also causing the C-terminal part of the α-helix to kink. In any case, whether or not the presence of the extra residues at the N-terminus of the SARS-CoV Nsp9 preparations used for structure determination results in artifacts, the observation of a completely different, disulfide-linked dimer in HCoV-229E Nsp9 is remarkable.

The occurrence of a disulfide bond in a viral protein located in the cytoplasm of the infected cell is unexpected, because here the overall milieu is reductive and disulfide bonds are rare, although a few cytosolic proteins containing them have been described (Bessette *et al.,* 1999; Parks *et al.,* 1997). Therefore, the possibility for the formation of the disulfide being an artifact of the conditions of protein preparation should be taken into account. In order to probe the effect of the disulfide bridge, Cys69 of HCoV-229E Nsp9 was mutated to alanine. Surprisingly, the crystal structure of the mutant displays the same dimerization mode as SARS-CoV Nsp9 and is thus very different from the wild-type HCoV-229E dimer. Furthermore, Nsp9 from TGEV, a group I coronavirus lacking the corresponding cysteine residue in the sequence, has also been crystallized. The crystal structure of TGEV Nsp9 is very similar to those of the HCoV-229E Cys69Ala mutant and of SARS-CoV Nsp9 (Fig. 4.2.1). To summarize, only the wild-type HCoV-229E Nsp9 dimerization mode is very different from the other known Nsp9 structures.
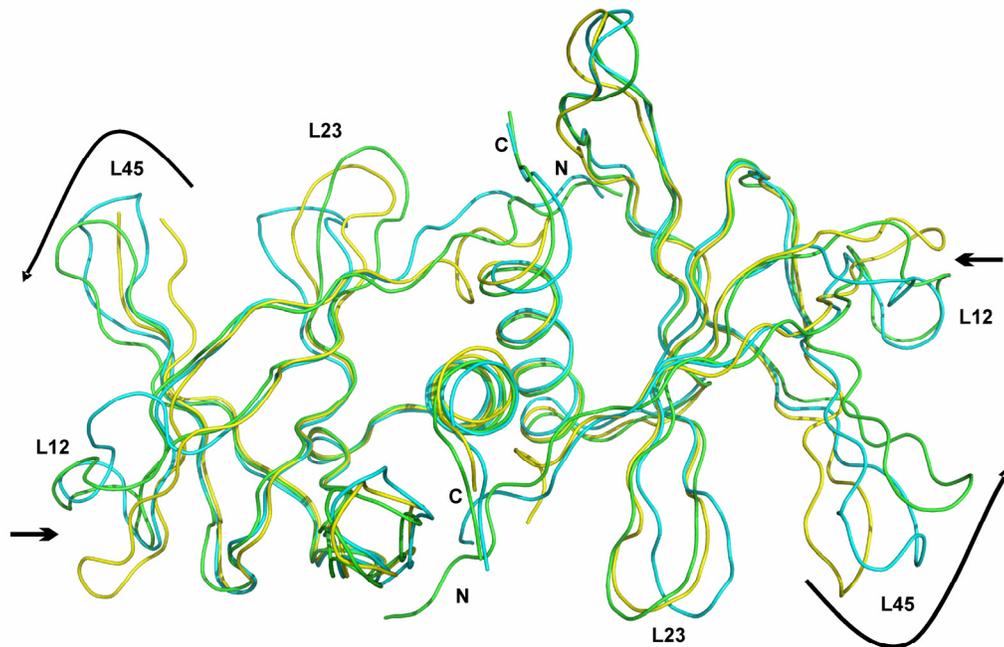
Fig. 4.2.1 Superimposition of HCoV-229E Cys69Ala mutant, SARS-CoV and TGEV Nsp9s.
HCoV-229E Nsp9 Cys69Ala mutant (yellow), SARS-CoV Nsp9 (cyan) and TGEV Nsp9 (green) have very similar dimerization modes mediated by the C-terminal helices. A higher degree of deviations is seen in L12, L23 and L45. In TGEV Nsp9, the β-barrel opens wider compared to the other two Nsp9s.

Using the surface area buried upon dimerization (Lee & Richards, 1971) and the shape complementarity (Lawrence & Colman, 1993), the relevance of the two different dimers seen in HCoV-229E Nsp9 wild-type and the Cys69Ala mutant, which is similar to the TGEV and SARS-CoV Nsp9 structures, is examined (see Tables 3.2.1 & 3.3.1). The shape complementarity of the monomer-monomer interface in the wild-type structure is as low as 0.56, but that of the mutant is not much better (0.63). The TGEV Nsp9 shape complementarity is 0.71, which is slightly higher; this may be due to the contribution of N-terminal residues to the interface. For comparison, this value is 0.67 and 0.70, respectively, for the two crystal structures of SARS-CoV Nsp9 (Sutton *et al.,* 2004; Egloff *et al.,* 2004), which show the same dimerization mode as the Cys69Ala mutant of HCoV-229E Nsp9 and TGEV Nsp9 (the artificial N-terminal tag, which also makes intersubunit contacts, has been removed from the Sutton *et al.* structure (PDB code 1UW7) in this calculation).

4.2.1 Role of the GXXXG motif in Nsp9s

Both the wild-type and the mutant dimer of HCoV-229E Nsp9 contacts are mainly mediated by the C-terminal helix of one monomer interacting with its (quasi-)symmetry mate in the other. However, the helices pack against one another in different orientations. The amino-acid residues of the helix that are involved in the interaction are highly conserved and small (GX_3GX_2GA), allowing helix packing according to the "ridges-into-grooves" model (X- stands for any amino acid; Fig. 4.2.2; see also Fig. 4.2.3; Chothia *et al.*, 1981). The GXXXG motif is actually a common feature of the association of transmembrane helices (Russ & Engelman, 2000; Senes *et al.*, 2000) but occurs also often in the helix-helix dimer association of soluble proteins (Kleiger *et al.*, 2002). In wild-type HCoV-229E Nsp9, residues 1, 4 and 7 of the helix sequence given above are involved in the interaction, allowing a close approach of the helices in an *anti-parallel*
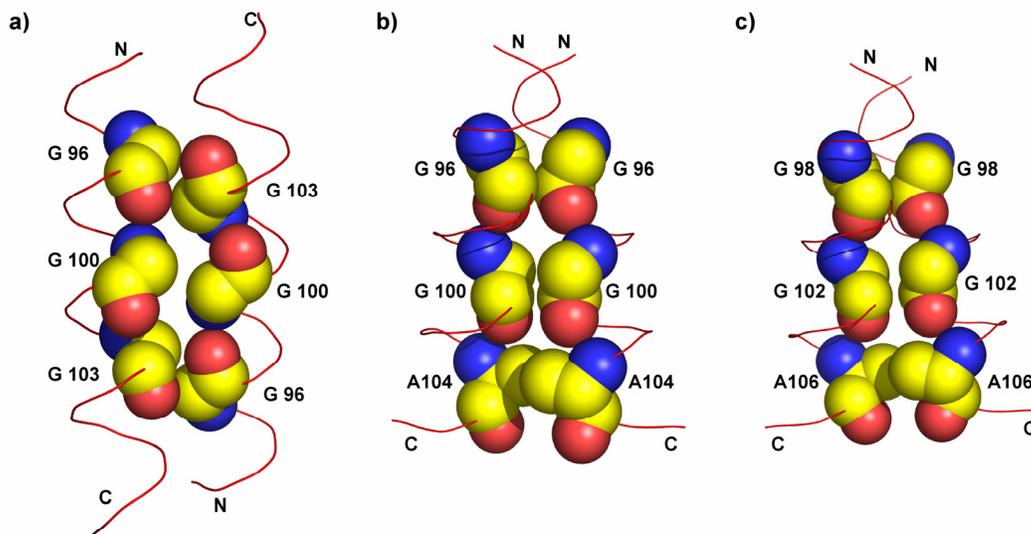
Fig. 4.2.2. Ribbon and space-filling representation of the C-terminal α-helix dimer interface
(a) HCoV-229E Nsp9-wild-type helices running in *antiparallel* orientation; Three glycine residues are in close contact. Glycine 96 and 100 are completely conserved, and glycine 103 is replaced with small amino acids (Ser/Ala) among the coronaviruses. (b) HCoV-229E Nsp9 Cys69Ala-mutant helices are crossing each other at 48° in a *parallel* orientation and also use the two conserved glycines 96 and 100. Instead of glycine 103, alanine 104 makes a close contact in the Cys69Ala mutant and is conserved as a small amino acid among the coronaviruses (Ser/Ala, with one exception, Asn in IBV). (c) TGEV Nsp9 helices crossing each other in a similar manner as the HCoV-229E Cys69Ala mutant Nsp9.

orientation with an angle of 167°. In contrast, in the mutant structure, residues 1, 4, and 8 are involved in the stabilization of a *parallel* orientation with a crossing angle of 48°. The TGEV and SARS-CoV Nsp9 helices are also stabilized in a similar manner as in the HCoV-229E Nsp9 Cys69Ala mutant (Fig. 4.2.2). The fact that the amino-acid sequence of the C-terminal helix of Nsp9 allows stabilization of both forms of the dimer seen in HCoV-229E may support the idea that both forms are indeed biologically relevant (see below). Recently, two independent groups (Chen *et al.,* 2009; Miknis *et al.,* 2009) studied the effect of the GXXXG motif with respect to Nsp9 dimerization and subsequently, to viral replication. Chen *et al.* (2009) investigated IBV Nsp9 (group III), whereas Miknis *et al.* (2009) worked on SARS-CoV Nsp9 (group II). Any of the glycine residues from the GXXXG motif mutated to charged residues (Glu or Asp) had a drastic effect on Nsp9 dimerization and on coronavirus replication. However, the Nsp9 RNA-binding activity was 5-12 fold decrease in affinity, but not completely abolished. But, replacing any glycine residue in the GXXXG motif with valine or alanine had a minor effect on Nsp9 dimerization and on coronavirus replication; in one case, the wild-type protein sequence is even reverted back through a codon transversion (Miknis *et al.,* 2009). Nevertheless, the mutation studies did not help discriminate the two different dimer interfaces seen in HCoV-229E Nsp9 because both the wild-type and Cys69Ala mutant use the C-terminal helix in their respective dimerization modes.

4.2.2 Are the N-terminal residues locking the helix-helix dimerization mode?

The N-terminal residues (1-6) "NNEI/LXP" of Nsp9 are strictly conserved among the coronaviruses (see Fig. 4.2.3). Their exact role in the coronavirus life cycle is not yet clear. The Nsp9 crystal structures, not all of the structures have a well-defined electron density for the N-terminal residues. The HCoV-229E Nsp9 wild-type disulfide-bridged crystal structure completely lacks any visible electron density for the seven N-terminal residues. Presumably, the N-terminal residues are flexible and may not be involved in the wild-type dimer interface of HCoV-229E Nsp9. In the HCoV-229E Cys69Ala mutant, the first two and the first four residues, respectively, in monomer A and B, are not seen in the electron density. However, the conserved hydrophobic residues Ile4 and Pro6 in monomer A lie over the α-helix of monomer B, whereas monomer B Pro6 lies over the α-

helix of monomer A. In the TGEV Nsp9 crystal structure, only the first residue is not seen in the electron density. Three hydrogen bonds, out of four seen in this interface, are made by the N-terminal residues Asn2 and Ile4 with residues Arg72 and Pro74 present in β6 and L56, respectively. Again, in TGEV Nsp9, very similar to HCoV-229E Nsp9 Cys69Ala mutant, the hydrophobic interactions by Ile4 and Pro6 are observed. In the crystal structure of SARS-CoV Nsp9 described by Egloff *et al.* (2004), the first two residues of monomer A and the first three residues of monomer B are not seen in the electron density. There is one hydrogen bond formed between the main-chain carbonyl atom of Leu4 of monomer A and the thiol of Cys73 (β6) of monomer B. Furthermore, two hydrogen bonds formed by Glu3 and Ser5 of monomer B with Arg74 (β6) of monomer A. Also, as seen in the other crystal structures, SARS-CoV Nsp9 hydrophobic N-terminal residues Leu4 and Pro6 are involved in the interface. The N-terminal residues in the crystal structure of SARS-CoV Nsp9 described by Sutton *et al.* (2004) make extensive interactions across the interface, but these N-terminal residues are influenced by the presence of 23 additional tag residues attached to them (see above). TGEV Nsp9 as well carries the same number of additional tag residues in the N-terminus but none of these have any visible electron density, and therefore may not influence the N-terminus of TGEV Nsp9. It is noteworthy that the SARS-CoV Nsp9 structure described by Egloff *et al.* (2004), the N-terminal residues carry a hexahistidine tag (B. Canard, personal communication). Altogether, there are several inconsistencies regarding the hydrogen bonds donated by the N-terminal residues of Nsp9s to mediate the helix-helix dimer interface. Although the residues are completely conserved, the hydrogen bonds are not seen in many of the known crystal structures, and in some structures, there is not even any visible electron density for these residues. So the contribution of hydrogen bonding by the N-terminal residues to the helix-helix dimer interface is ambiguous. Strikingly, the N-terminal hydrophobic residues Leu/Ile4 and Pro6 of one monomer lie over the hydrophobic cavity created between the β6-strand and α-helix of the other monomer in almost all the structures known so far. Therefore, the two conserved hydrophobic residues may further contribute to stabilize the hydrophobic helix-helix dimer interface found in HCoV-229E Cys69Ala mutant, TGEV, and SARS-CoV Nsp9s. Similar to Nsp9, M^{pro} from the coronaviruses uses its N-terminal residues as "N-finger" to stabilize the

active dimer necessary for protease activity (Anand *et al.,* 2002b, 2003; Yang *et al.,* 2003). Without the N-terminal residues, M^{pro} is catalytically inactive, however it forms a different dimer in solution with the help of the C-terminal domains (Zhong *et al.,* 2008).

It is very likely that the helix-helix dimer interface seen in Nsp9 crystal structures, except the disulfide-bridged HCoV-229E Nsp9 dimer, is the biologically relevant form. This dimerization mode is crucial for the formation of the replication complex, rather than just RNA binding (Chen *et al.,* 2009; Miknis *et al.,* 2009). However, the question regarding the disulfide-bridge formation in HCoV-229E Nsp9 wild-type remains to be understood. In addition to the helix-helix ('parent dimer') dimerization mode that was identified in all Nsp9 structures, a number of additional protein-protein interfaces that are seen in the crystal structures has to be considered (see Table 3.2.1 & 3.3.1).

4.2.3 Other interfaces in wild-type HCoV-229E Nsp9

In the wild-type protein, three disulfide-bonded dimers form a trimer of dimers, or hexamer, involving interfaces W2 and W3. Hexamers are assembled into 36-mers through interface W4 (Fig. 3.2.9; Table 3.2.1).

4.2.4 Dimer interface mediated by strand β5

Although the helix-helix dimer interface discussed above is the parent mode, the dimer interface formed through the β5-strands is also observed in all the Nsp9 structures solved so far (except the disulfide-bridged wild-type HCoV-229E Nsp9 dimer). In the HCoV-229E Nsp9 Cys69Ala mutant, the interface arises through some limited interactions between strands β5 of neighboring molecules in the crystal (Fig. 3.2.9 (a)). The β-sheet-mediated interaction comprises four hydrogen bonds between the main-chain atoms and is reminiscent of intersubunit interactions involving β-strands in the single-stranded DNA-binding protein (SSB) from *E. coli,* a prototype OB-fold protein (Webster *et al.,* 1997). This same alternative dimerization mode has also been discussed by Sutton *et al.* (2004) for their structure of the SARS-CoV Nsp9, but considered irrelevant. However, this interface is also found in the SARS-CoV Nsp9 structure described by Egloff *et al.* (2004), even though the space group of these crystals is different. In the crystal structure

of TGEV Nsp9, the interface mediated by the β5-strand is quite different compared to HCoV-229E and SARS-CoV Nsp9s (Fig. 3.3.7). There is an extensive interaction between the β56-sheets to its symmetry mate rather than the limited interaction seen in the other Nsp9 structures (Fig. 3.3.8).

Fig. 4.2.3 Multiple sequence alignment of Nsp9
Alignment was done using CLUSTALW (Thomson *et al.,* 1994) for Nsp9 from different groups of coronaviruses. Residues boxed in red are completely conserved. Secondary structures are denoted using the HCoV-229E Nsp9 structure (top) and the TGEV Nsp9 structure (bottom). The position of the cysteine residue involved in disulfide-bridge formation in HCoV-229E Nsp9 is labeled with a blue asterisk.

4.2.5 Why does TGEV Nsp9 adopt the unique β-sheet interface?

The β-sheet interface seen in SARS-CoV and HCoV-229E Cys69Ala mutant Nsp9 structures are mainly mediated by the close approach of the β5-strands from different monomers, thereby forming several hydrogen bonds between the main-chain atoms. In TGEV Nsp9, the close approach of β-strands is not possible due to the gross negative charge in L45 and L12. TGEV and SARS-CoV Nsp9s L12 is longer by two residues

compared to HCoV-229E (Fig. 4.2.3). But SARS-CoV Nsp9 lacks the acidic residue Asp21, which makes a crucial salt bridge in the TGEV Nsp9 β-sheet interface. These features favour the TGEV Nsp9 to adopt its unique β-sheet interface. If this interface has any biological implication, then it may be unique to TGEV among the coronaviruses because the residues involved in the interface are not conserved.
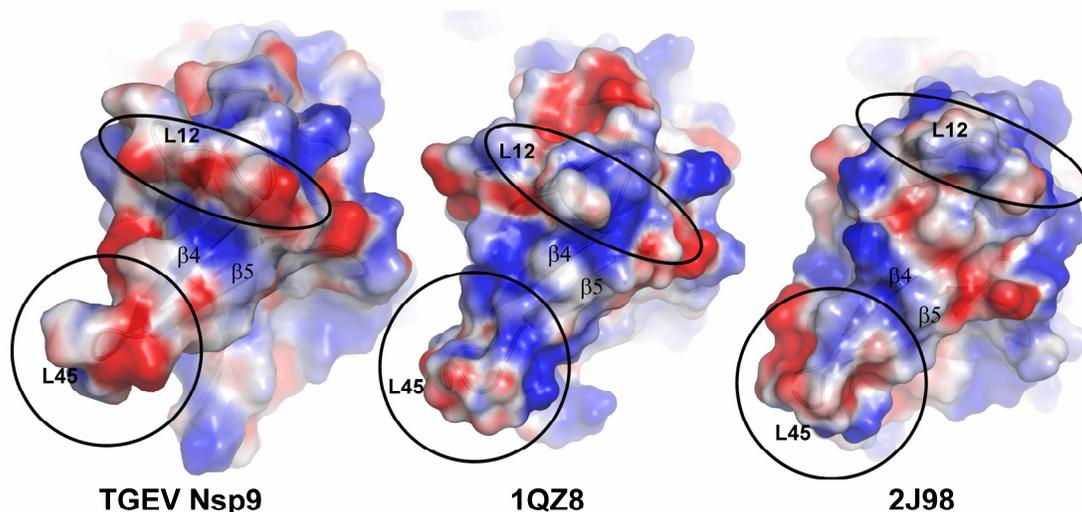
Fig. 4.2.4 Cross-section of the β-sheet interface with electrostatic potential surface
The electrostatic potential surface of TGEV Nsp9 β-sheet interface is compared with SARS-CoV and HCoV-229E Cys69Ala mutant Nsp9s. The surface is colored according to the electrostatic potential ranging from deep blue (positive charge +10 k_BT) to red (negative charge -10 k_BT). The electrostatic potential was calculated using APBS tools implemented in the Pymol software (Baker *et al.*, 2001).

4.2.6 Common oligomerization mode with some plasticity

All the Nsp9 crystal structures solved so far tend to aggregate as polymers in the crystal. The existence of such oligomers in solution is supported by glutaraldehyde crosslinking experiments, which revealed the presence of monomers, dimers, trimers, and higher oligomers for SARS-CoV Nsp9, TGEV Nsp9 and HCoV-229E Cys69Ala mutant Nsp9 in the SDS-PAGE under reducing conditions (Fig. 3.2.10 & Fig. 3.3.9). Interestingly, for the wild-type HCoV-229E protein, only monomers, dimers, and to a limited extent, trimers were seen by this method. This is consistent with the observed crystal structures: when in the disulfide-linked state, wild-type 229E Nsp9 cannot normally form oligomers larger than hexamers (Fig. 3.2.8), whereas the HCoV-229E Cys69Ala mutant, TGEV as well as SARS-CoV Nsp9s can form polymers (Fig. 4.2.5). Very likely these polymers are formed
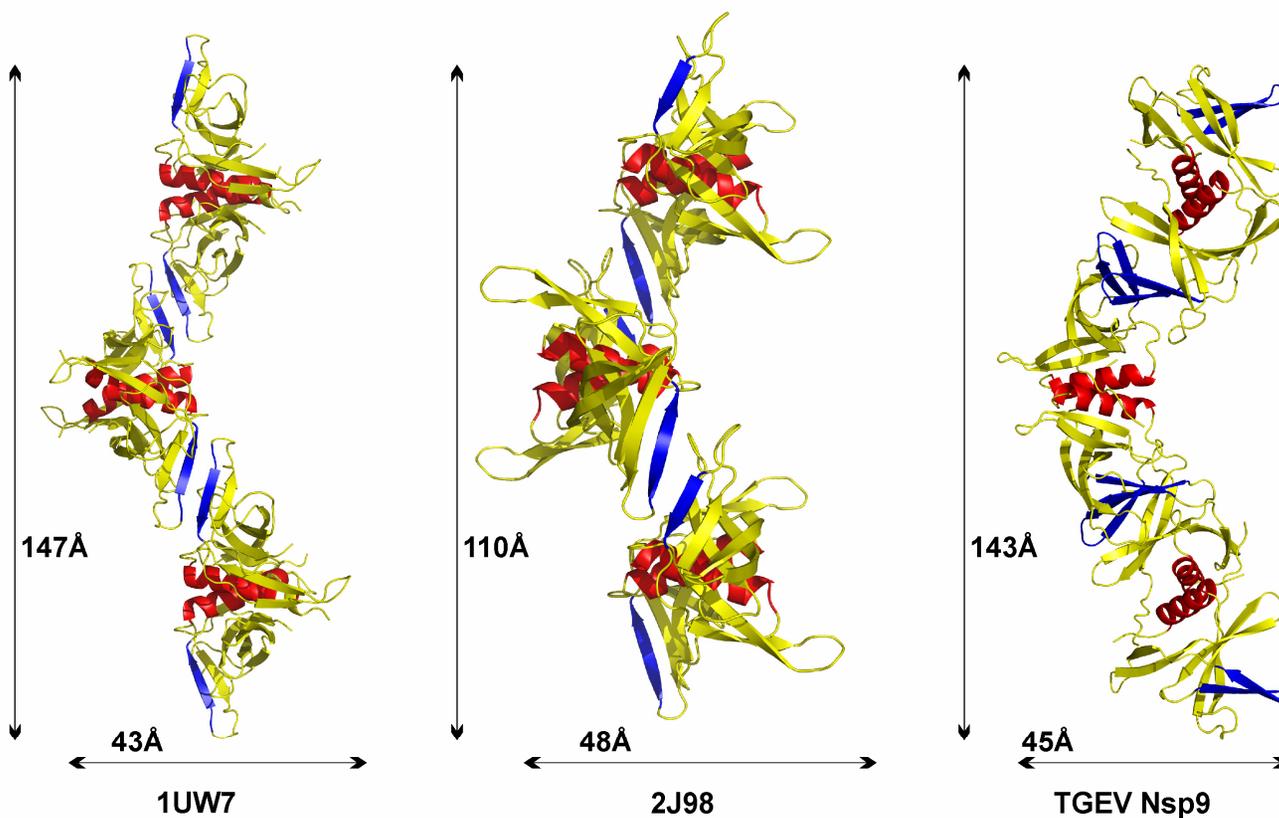
Fig.4.2.5 Higher oligomers of Nsp9s
Independent of different space groups (SARS-CoV Nsp9; PDB Code: 1UW7 - P4$_3$22, HCoV-229E Nsp9 Cys69Ala mutant; PDB Code: 2J98 - P2$_1$2$_1$2$_1$, TGEV Nsp9 - I422), two common dimer interfaces are present in all the crystal structure of Nsp9. The first interface is mainly meditated by the C-terminal α-helix colored in red and the second interface is formed by the β-sheet colored in blue. However, the TGEV Nsp9 oligomerization mode is slightly different compared to the other Nsp9 structures.

with the two common dimer interfaces seen in the crystal structure of Nsp9s. The parent mode is mediated by C-terminal helices and the β-strand interface is especially used for the oligomerization. However, the TGEV Nsp9 β-strand interaction differs from other Nsp9 β-strand interactions. Nevertheless, the oligomerization based on β-strand interaction is again seen with TGEV Nsp9. In all the crystal structures of Nsp9s, except the wild-type Nsp9 of HCoV-229E, oligomerization is the common feature and indeed may reflect their biological role. Interestingly, two dimer interfaces are commonly used for the oligomerization, independent of the discrepancies in the mode of β5 sheet usage. This quaternary plasticity may reflect the different groups of coronaviruses.

Yet another interface formed by the β6-β7 hairpin in the HCoV-229E Cys69Ala mutant Nsp9 is noted (Fig. 3.2.9 (b)). This interface involves mainly hydrophobic interactions between side chains that are not conserved (Table 3.2.1), although a similar interface exists in the Egloff *et al.* (2004) structure of SARS-CoV Nsp9 but not in the Sutton *et al.* (2004) or TGEV Nsp9 structure.

4.2.7 Model for ssRNA binding to Nsp9

The Nsp9 dimer unit mediated by the helix-helix interface was investigated to identify the ssRNA binding mode by using the electrostatic surface potential (Baker *et al.,* 2001). The theoretical pI of these proteins are above approximately 9.0, which is clearly reflected in their electrostatic surface potential (Fig. 4.2.6). Therefore, only based on the electrostatic surface potential a ssRNA path could not be identified. Very likely, the ssRNA may wrap around the Nsp9 molecule as observed in the *E.coli* single-stranded DNA binding protein (Ragunathan *et al.,* 2000). Moreover, with the help of the bound sulfate ions present in one of the SARS-CoV Nsp9 structures and with the electrostatic surface potential (Egloff *et al.,* 2004; PDB code: 1QZ8) a model for ssRNA binding was proposed (Fig. 4.2.7). In this crystal structure, three sulfate ions are located near one of the two monomers, two of them in the vicinity of the completely conserved lysine residues 50 (52; the HCoV-229E numbering scheme is used, with numbers for SARS-CoV in brackets) and 88(92), and one interacting with residue 46(48; Lys46 in 229E, His48 in SARS-CoV). By superimposition with the structure of the wild-type HCoV-
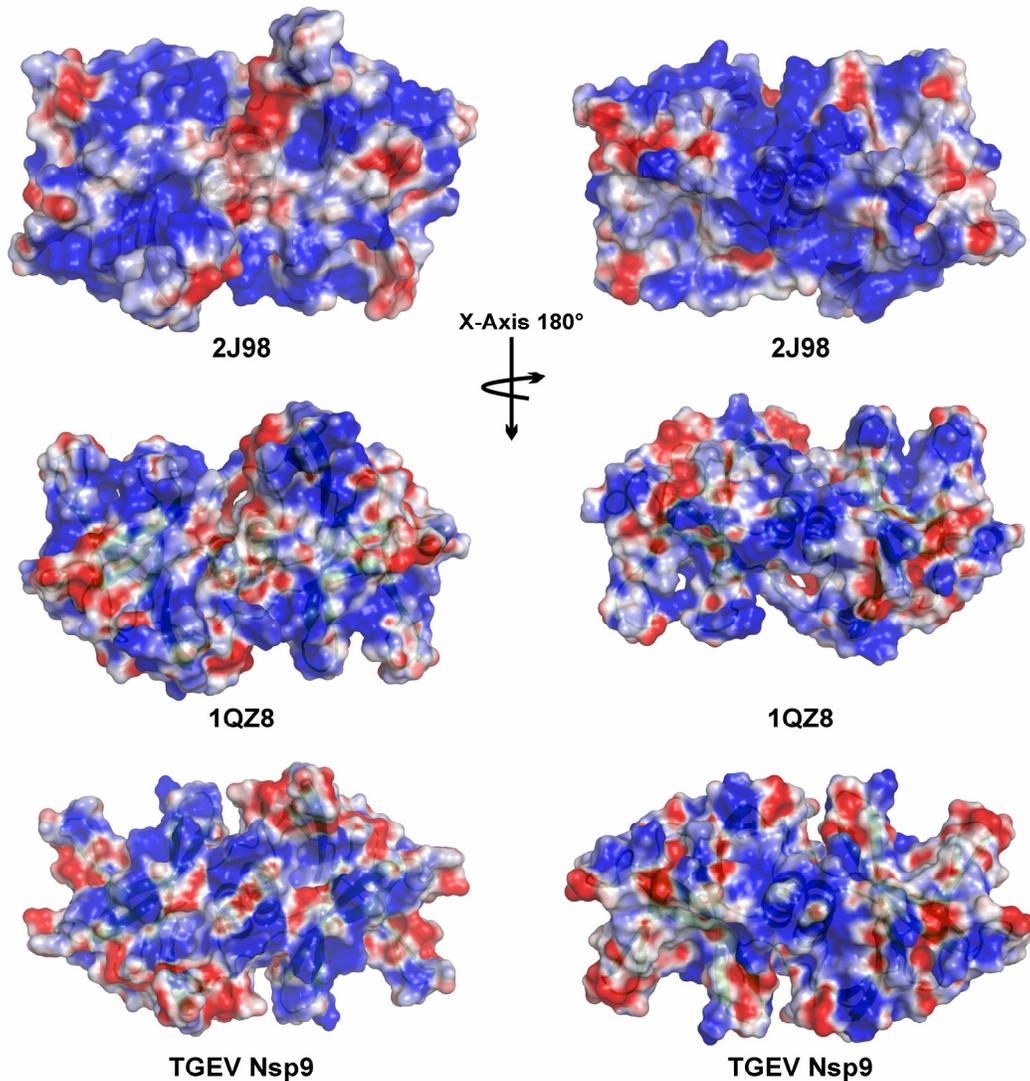
Fig. 4.2.6 Electrostatic potential surface of Nsp9 helix-mediated dimers
Comparison of the electrostatic potential surface of helix-helix dimer structures (a) HCoV-229E Cys69Ala mutant (2j98), (b) SARS-CoV Nsp9 (1QZ8) and (c) TGEV Nsp9. The surface is colored according to the electrostatic potential ranging from deep blue (positive charge +10 k_BT) to red (negative charge -10 k_BT). The electrostatic potential was calculated using the APBS tools implemented in the Pymol software (Baker *et al.,* 2001).

229E Nsp9, which was crystallized from sulfate, two further sulfate-binding sites are revealed. One is also near Lys50(52), but in a different position, and the other one interacts with Lys82(86). The resulting five independent sulfate positions were used to define a path for ssRNA on the surface of the monomer and subsequently, the polymer. The residues proposed to interact with the ssRNA on the basis of this model (Lys10, Lys50, Tyr51, Arg70, Tyr83, Lys88, and Arg107) are better conserved than is the polypeptide sequence on average. In the crude model, the ssRNA forms a left-handed helix wrapping around the Nsp9 polymer, similar to the model recently proposed for the nucleocapsid protein of SARS-CoV interacting with ssRNA (Chen *et al.,* 2007). Based on the model, approximately 40 nucleotides can be bound per Nsp9 dimer within the extended polymer for SARS-CoV Nsp9 (Fig. 4.2.7 (d)). In case of TGEV Nsp9, the extended polymer could not be superimposed onto the SARS-CoV Nsp9 polymer, due to the *parallel* β-sheet mode of interface in TGEV Nsp9 rather than *anti-parallel* as seen in the SARS-CoV and HCoV-229E Cys69Ala mutant Nsp9s. The arrangement of the β-sheet interface positions the helix-helix dimer unit in very different orientation in TGEV Nsp9 polymerization compared to SARS-CoV and HCoV-229E Cys69Ala mutant Nsp9s (Fig. 4.2.5). Therefore, at this point it is difficult to rationalize the ssRNA binding mode for Nsp9s from different groups of coronaviruses. The formation of the double-stranded super helix seen in TGEV Nsp9 may be an option for the ssRNA to wrap around the Nsp9 molecule, when one of the strands is absent. It is also noted that there are almost no interactions between the two long Nsp9 polymer strands. Very recently, similar long polymer formation was seen with the Herpes simplex virus type I SSB protein (ICP8), and the authors have proposed that these long polymers could bind two ssDNA molecules continuously (Makhov *et al.,* 2009). In symptom of this hypothesis, one of the TGEV Nsp9 polymer strand in the double-stranded helical model is shown with a mesh and the other strand in the ribbon representation (Fig. 4.2.8).

4.2.8 Nucleic-acid interaction of Nsp9s

Nsp9 interaction with nucleic acids was visualized using a slightly modified version of Zone Interference Gel Electrophoresis (ZIGE; Abrahams *et al.,* 1988; see Materials and Methods for details). HCoV-229E and SARS-CoV Nsp9 bind to single-stranded and, to a

limited extent, double-stranded deoxyoligonucleotides without any sequence specificity. Also, TGEV Nsp9 is able to bind nucleic acids in a similar manner to SARS-CoV Nsp9 (not shown). Therefore, TGEV Nsp9 was left out for the further comparison studies. Deoxyribonucleotides were used instead of ribonucleotides because they showed identical behavior in test runs. In agreement with the structural data discussed above, surface-exposed positively charged residues could interact with the sugar-phosphate backbone of the nucleic acid. However, there is a strong correlation between the length of the oligonucleotide and the gel-shift observed (Fig. 3.2.11). Although the dimerization modes of HCoV-229E wild-type and SARS-CoV Nsp9 wild-type proteins are very different, their binding profile to nucleic-acid in the gel-shift experiment is similar (neglecting the step-wise rather than linear increase of the gel shift with oligonucleotide lengths in case of the SARS-CoV protein). On the other hand, the HCoV-229E Cys69Ala mutant has a dimerization mode similar to the wild-type SARS-CoV Nsp9 and yet, it does not show binding of the nucleic acid in the gel-shift experiment (except for the small shift seen for the longest oligonucleotide tested, the 55-mer). This inability to bind oligonucleotides could, in principle, be due to a direct interaction of the Cys69 with the nucleic acid. Therefore, an additional three mutants were prepared, namely HCoV-229E Cys69Ser and the corresponding mutants of SARS-CoV Nsp9, Cys73Ala and Cys73Ser. Similar to HCoV-229E Cys69Ala, the Cys69Ser mutant did not exhibit a significant shift in the ZIGE experiment either. However, the homologous SARS-CoV Cys73Ala and Cys73Ser mutants did show a shift of the same magnitude as the SARS-CoV Nsp9 wild-type (Fig. 3.2.11 (C)). It remains to be understood why the HCoV-229E Cys69Ala and Cys69Ser mutants apparently do not bind nucleic acids, whereas the corresponding SARS-CoV mutants do.
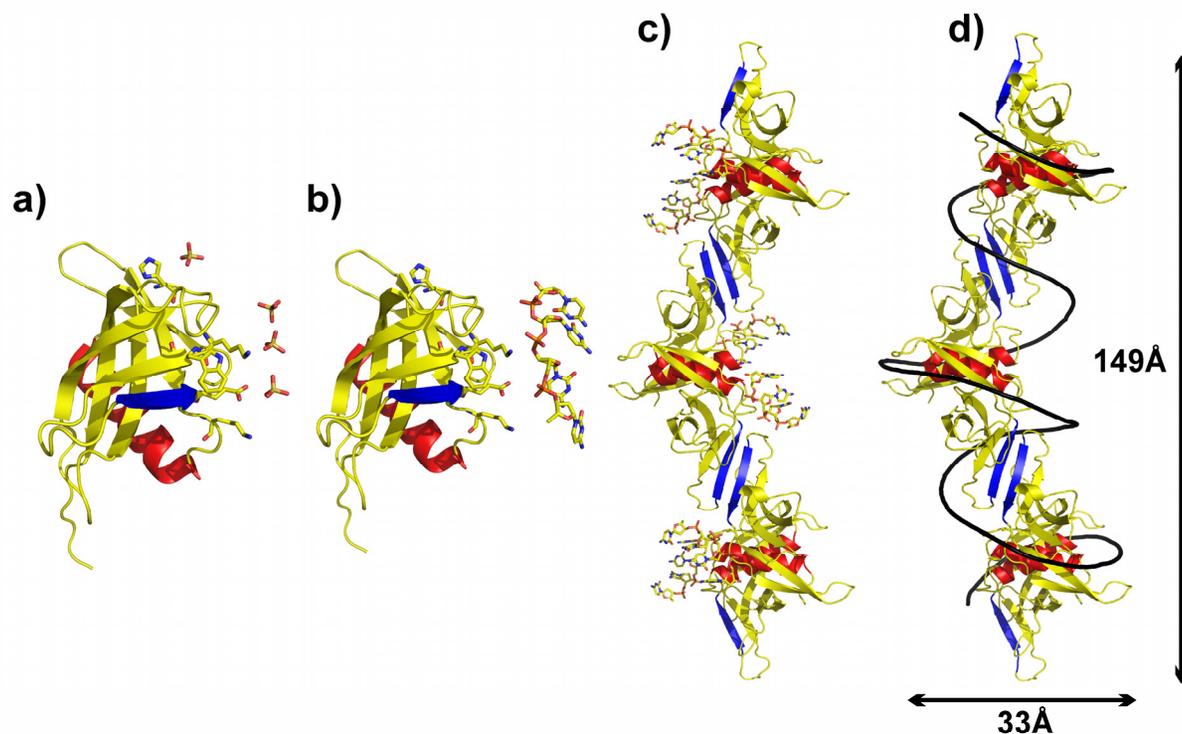
Fig. 4.2.7 ssRNA binding model derived for SARS-CoVNsp9 using the bound sulfate ions in the crystals
a) Ribbon representation of SARS-CoV Nsp9 (1QZ8 - $P6_122$) monomer with two bound sulfate ions; another two sulfate ions were identified by superimposing the monomer of HCoV-229E Nsp9 wild-type onto the SARS-CoV Nsp9 monomer. Residues in close contact with the sulfate ions are shown using sticks. (b) Three nucleotides and their backbone phosphates were superimposed onto the bound sulfate ions manually. (c) The monomer structure carrying the manually built nucleotide bases were superimposed onto the Nsp9 oligomers. (d) Furthermore, the nucleotide bases were joined by following the sequence conservation and the two other bound sulfate ions. ssRNA represented as a black line forming a left-handed helix with approximately 40 nucleotides are bound per Nsp9 dimer.
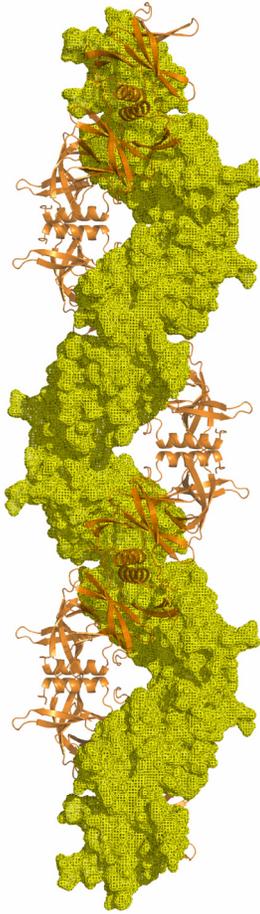
Fig. 4.2.8 Polymer formation in the crystal structure of TGEV Nsp9. Two polymers of TGEV Nsp9 form a double-stranded helix. One strand of polymer is shown as a surface mesh colored in yellow, and the other polymer as a ribbon representation colored orange.

To further analyze the interaction between nucleic acids and the different dimeric forms of HCoV-229E Nsp9, surface plasmon resonance was employed. An "aged" preparation (two weeks old) of wild-type HCoV-229E Nsp9 showed a strong signal for the binding with a 50-mer oligonucleotide under non-reducing conditions. Freshly prepared wild-type HCoV-229E Nsp9 gave a much weaker binding signal under reducing conditions. In case of the HCoV-229E Nsp9 Cys69Ala mutant, the signal was equally small, irrespective of whether or not DTT was present. These relatively weak signals still indicate significant binding to the 50-mer, albeit much weaker than found for the wild-type HCoV-229E protein in its oxidized state. This is in agreement with the gel shifts, where only a weak shift was observed for the HCoV-229E mutants in case of the longest oligonucleotide examined (the 55-mer), but not with shorter oligonucleotides.

Why does the wild-type HCoV-229E Nsp9 form a disulfide-linked homodimer, but not the SARS-CoV protein, even though it has Cys69(73) conserved? The chemical environment of the cysteine residues is the same in the two structures, *i.e.* there is no special structural feature in HCoV-229E Nsp9 that would cause a particular reactivity of Cys69. However, in contrast to HCoV-229E, the SARS-CoV protein has two additional cysteine residues, no. 14 and 23. All three cysteines are in the free form, as determined using Ellman's reagent (not shown). If one of these cysteines was involved in an intermolecular disulfide bond, the latter would probably be reduced by the remaining ones, so that a disulfide-bonded dimer would be unlikely to be the dominant species. In agreement with this argument, there are only few proteins that contain a disulfide bond in addition to a free cysteine (Petersen *et al.,* 1999). (An exception are the cysteine proteases of the papain family, where the active-site cysteine has special properties such as an unusual pK_a value).

The question remains then, whether the disulfide-bonded form of wild-type HCoV-229E Nsp9 is an artifact that may have occurred during protein preparation. As mentioned before, the freshly prepared sample of this protein gave a reaction with Ellman's reagent, but not so after one day. Apparently, there is a correlation between the age of the sample and disulfide formation, even though the reducing agent dithiothreitol (DTT; 5 mM) was added at regular intervals. It is known that DTT is subject to oxidation itself; its half-life at 20°C is 10 h and 40 h at pH 7.5 and 6.5, respectively (Stevens *et al.,* 1983). The concentration of DTT required to fully reduce the Nsp9 disulfide was determined as 10 mM; however, at this concentration, the protein would not crystallize.

As the disulfide-bonded form of HCoV-229E Nsp9 binds oligonucleotides with much higher affinity than the reduced form, it may indeed have a biological role, possibly in response to the oxidative stress induced by the viral infection of the host cell. There are several earlier reports suggesting the regulation of DNA/RNA-binding proteins by redox processes. Thus, the redox state has been shown to determine the interaction with DNA of the multifunctional eukaryotic SSB protein RPA (You *et al.,* 2000). Also, many transcription factors including Fos, Jun, NF-κB, PaX, FNR, OxyR (see Bauer *et al.,* 1999

for a review) are regulated by the redox state of the environment. Another example is the p53 tumor suppressor protein, which binds to DNA with sequence specificity only in the reduced state. Disulfide formation in the oxidized state alters the conformation and the protein binds DNA without any sequence specificity (Parks *et al.,* 1997).

Several viruses have been reported to induce oxidative stress in the infected cell. Among the coronavirus family, TGEV was shown to induce apoptosis in the infected cell via oxidative stress (Eleouet *et al.,* 1998). Similarly, rhinovirus (Kaul *et al.,* 2000) and baculovirus (Wang *et al.,* 2001) also induce oxidative stress in the infected cell. More specifically, LEF3 (the SSB of baculovirus) shows a DNA-annealing effect in its oxidized state, whereas in the reduced state, its DNA-unwinding activity is favored. Cys214 is apparently involved in DNA binding; when mutated to serine, both DNA-binding and unwinding activities are reduced (Mikhailov *et al.,* 2005). It has been hypothesized that this cysteine could form an intermolecular disulfide bridge and thus results in LEF3 oligomers in solution. This could allow more DNA to bind in closer proximity, thus favoring the annealing effect (Mikhailov *et al.,* 2005). The E2 protein of bovine papilloma virus type 1 and ICP8 of herpes simplex virus type 1 also show DNA-binding activity depending on the redox state of the environment (McBride *et al.,* 1992; Sampson *et al.,* 2000; Knipe *et al.,* 1982; Dudas & Ruyechan, 1998).

What is the relevance of these *in-vitro* studies to the situation in the infected host cell? For several RNA viruses, including mouse hepatitis (corona)virus (MHV) and SARS-CoV, it has been shown that viral replication is localized to double-membrane vesicles that have been hijacked from the endoplasmic reticulum or late endosomes (Gosert *et al.,* 2002; Prentice *et al.,* 2004; Snijder *et al.,* 2006; van Hemert *et al.,* 2008). These double-membrane vesicles are around 200-350 nm in diameter and present alone or as clusters in the cytosol (Prentice *et al.,* 2004; Knoops *et al.,* 2008). The milieu inside these vesicles or at their surface is unknown, but it is well possible that it is partially oxidative. Since it is here that replicase proteins are produced at high levels, it is conceivable that the oxidized form of HCoV-229E Nsp9 is the dominant species. According to the findings, this form binds to single-stranded RNA more tightly than does the reduced form, and could

therefore promote replication of the viral genome. This speculation is supported by the recent report by Wu *et al.* (2008) who have shown that oxidative stress in the host cell promotes HCoV-229E infectivity. Very likely, SARS-CoV will also induce oxidative stress in the infected host cell (Imai *et al.,* 2008), even though its Nsp9 does not seem to form disulfide-linked dimers, at least not *in vitro*. In case of TGEV Nsp9, there is not even a single cysteine residue. The replicase of these viruses may have other mechanisms to deal with an oxidative environment. It is worth noting that the number of cysteine residues is above average in coronavirus replicase proteins; in SARS-CoV and TGEV, their shares are 3.9% and 3.3%, respectively (HCoV 229E: 3.3%), whereas only 1.25% of residues in human cytosolic proteins are cysteines (Miseta & Csutora, 2000). Some of these could perhaps scavenge oxygen radicals, thereby undergoing oxidation to sulfenic, sulfinic, or even sulfonic acid.

5. Summary

The Nsp8 gene of HCoV-229E and SARS-CoV were cloned, expressed, and the proteins were purified using affinity and size-exclusion chromatography. The purified proteins were largely aggregated, as elucidated by native-gel electrophoresis and DLS analysis. A small amount of reducing and chelating agents dissolved the Nsp8 aggregates although, the hydrodynamic radius was still much larger compared to the calculated monomeric or lower oligomeric states. Bioinformatics programs, such as FoldIndex and PONDR, predicted the presence of intrinsically disordered region. Furthermore, PONDR spotted residues 40 to 80 as an α-MoRE region, which is known to be involved in protein-protein interaction. Using CD spectroscopy, the secondary structure of HCoV-229E and SARS-CoV Nsp8 consisted of 59% and 60% of loops (without any secondary structure), respectively. In line with the previous finding, 1D-NMR studies on the SARS-CoV Nsp8 indicated the protein was either aggregated or partially disordered. Simultaneously, trypsin was able to digest the protein completely within 5 min, indicating the protein is partially disordered. The ability of the fluorescence probe 1,8-ANS to bind to both Nsp8 protein molecules, allowed to speculate that part of the proteins were folded, in agreement with CD spectroscopy (~40% of secondary structure elements). Moreover, HCoV-229E and SARS-CoV Nsp8 were able to bind nucleic acid with μM affinity as determined by zone-interference gel electrophoresis (ZIGE). Both proteins exhibited little sequence specificity although, polyA and polyG were really poor substrates. In the presence of HCoV-229E Nsp7 and Nsp9, Nsp8 showed a reduced hydrodynamic radius in the DLS measurement, indicating specific protein-protein interaction. The solubility of HCoV-229E Nsp8 was greatly increased in the presence of Nsp7. Therefore, the Nsp7 and Nsp8 proteins were co-purified. The crystallization trials of the Nsp7-Nsp8 complex yielded only spheres and spherulites. Further optimization is needed for obtaining X-ray quality crystals. All attempts to crystallize HCoV-229E and SARS-CoV Nsp8 alone remained unsuccessful.

Seven different variants of Nsp9 genes, wild-type HCoV-229E Nsp9, Cys69Ala mutant, Cys69Ser mutant, wild-type SARS-CoV Nsp9, Cys73Ala mutant, Cys73Ser mutant, and

wild-type TGEV Nsp9, were cloned, expressed, and the proteins were purified using affinity and size-exclusion chromatography. In total, four different crystal structures of Nsp9s were determined using the molecular replacement technique. The crystal structure of HCoV-229E Nsp9 reveals a novel disulfide-linked homodimer, which is very different from the previously reported Nsp9 dimer of SARS coronavirus (Egloff *et al.,* 2004; Sutton *et al.,* 2004). In contrast, the structure of the Cys69Ala mutant of HCoV-229E Nsp9 shows the same dimer organization as the SARS-CoV protein. The previously reported SARS-CoV Nsp9 structures carried additional residues at the N-terminus and therefore, authentic protein was produced. However, crystallization of the authentic version of SARS-CoV Nsp9 was unsuccessful. Only SARS-CoV Nsp9 carrying additional residues at its N-terminus could be crystallized. These crystals were of the same habit as previously reported (Sutton *et al.,* 2004) and thus not included for further analysis. Discrepancies in the crystal structures of Nsp9 were further clarified by crystallizing the TGEV Nsp9 that does not contain any cysteine residue in its sequence. The TGEV Nsp9 dimer structure turned out to be similar to the ones of SARS-CoV Nsp9 and HCoV-229E Cys69Ala mutant. All together, only the wild-type HCoV-229E Nsp9 dimer structure is very different from the other known Nsp9 structures. In the crystal, the wild-type HCoV-229E Nsp9 forms a trimer of dimers whereas the SARS-CoV, TGEV, and HCoV-229E Cys69Ala Nsp9 are organized in rod-like polymers. Chemical cross-linking suggests similar modes of aggregation in solution. From the crystal structures, models for single-stranded RNA binding by Nsp9 are deduced.

In order to study the Nsp9 interaction with nucleic acids, the ZIGE technique was modified. The modified version of ZIGE is suitable to study any weak protein-nucleic-acid interaction in case the protein displays a basic isoelectric point. The wild-type HCoV-229E Nsp9 was able to bind oligonucleotides with relatively high affinity in contrast to the mutant Nsp9s. The Cys69Ala and Cys69Ser Nsp9 mutants were only able to bind rather long oligonucleotides. The corresponding mutations in SARS-CoV Nsp9 do not hamper nucleic-acid binding. Both the wild-type HCoV-229E and SARS-CoV Nsp9 bind nucleic acid in the lower µM range as determined by surface plasmon resonance. It is plausible that both the dimer forms are biologically relevant; the

occurrence of the disulfide-bonded form may be correlated with oxidative stress induced in the host cell upon viral infection.

6. References

Abrahams, J.P., Kraal, B. & Bosch, L. (1988): Zone-interference gel electrophoresis: a new method for studying weak protein-nucleic acid complexes under native equilibrium conditions. *Nucleic Acids Res.* **16**, 10099-10108.

Ahlquist, P. (2006): Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nat. Rev. Microbiol.* **4**, 371-382.

Anand, K., Pal, D. & Hilgenfeld, R. (2002a): An overview on 2-methyl-2,4-pentanediol in crystallization and in crystals of biological macromolecules. *Acta Cryst.* **D58**, 1722-1728.

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

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R. & Hilgenfeld, R. (2003): Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763-1767.

Andino, R., Rieckhof, G.E., Achacoso, P.L. & Baltimore, D. (1993): Poliovirus RNA synthesis utilizes an RNP complex formed around the 5′-end of viral RNA. *EMBO J.* **12**, 3587-3598.

Andino, R., Rieckhof, G.E. & Baltimore, D. (1990): A functional ribonucleoprotein complex forms around the 5′end of poliovirus RNA. *Cell* **63**, 369-380.

Andrade, M.A., Chacón, P., Merelo, J.J. & Morán, F. (1993): Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng.* **6**, 383-390.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. (2001): Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037-10041.

Baric, R.S. & Yount, B. (2000): Subgenomic negative-strand RNA function during mouse hepatitis virus infection. *J. Virol.* **74**, 4039-4046.

Bauer, C.E., Elsen, S. & Bird, T.H. (1999): Mechanisms for redox control of gene expression. *Annu. Rev. Microbiol.* **53**, 495-523.

Bessette, P.H., Aslund, F., Beckwith, J. & Georgiou, G. (1999): Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proc. Natl. Acad. Sci. USA* **96**, 13703-13708.

Bochkareva, E., Belegu, V., Korolev, S. & Bochkarev, A. (2001): Structure of the major single-stranded DNA-binding domain of replication protein A suggests a dynamic mechanism for DNA binding. *EMBO J.* **20**, 612-618.

Bosch, B.J., Martina, B.E.E., Van Der Zee, R., Lepault, J., Haijema, B.J., Versluis, C., Heck, A.J.R., De Groot, R., Osterhaus, A.D.M.E. & Rottier, P.J.M. (2004): Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc. Natl. Acad. Sci. USA* **101**, 8455-8460.

Bost, A.G., Carnahan, R.H., Lu, X.T. & Denison, M.R. (2000): Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. *J. Virol.* **74**, 3379-3387.

Brian, D.A. & Baric, R.S. (2005): Coronavirus genome structure and replication. *Curr.*

Top. Microbiol. Immunol. **287**, 1-30.

Brockway, S.M., Clay, C.T., Lu, X.T. & Denison, M.R. (2003): Characterization of the expression, intracellular localization, and replication complex association of the putative mouse hepatitis virus RNA-dependent RNA polymerase. *J. Virol.* **77**, 10515-10527.

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

Cavanagh, D. (2005). Coronaviruses in poultry and other birds. *Avian Pathol.* **34**, 439-448.

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

Chen, C., Chang, C., Chang, Y., Sue, S., Bai, H., Riang, L., Hsiao, C. & Huang, T. (2007): Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J. Mol. Biol.* **368**, 1075-1086.

Chen, J.W., Romero, P., Uversky, V.N. & Dunker, A.K. (2006a): Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *J. Proteome Res.* **5**, 879-887.

Chen, J.W., Romero, P., Uversky, V.N. & Dunker, A.K. (2006b): Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder. *J. Proteome Res.* **5**, 888-898.

Cheng, Y., LeGall, T., Oldfield, C.J., Mueller, J.P., Van, Y.Y., Romero, P., Cortese, M.S., Uversky, V.N. & Dunker, A.K. (2006): Rational drug design via intrinsically disordered protein. *Trends Biotechnol.* **10**, 435-442.

Cheng, A., Zhang, W., Xie, Y., Jiang, W., Arnold, E., Sarafianos, S.G. & Ding, J. (2005): Expression, purification, and characterization of SARS coronavirus RNA polymerase. *Virology* **335**, 165-176.

Choo, Y. & Schwabe, J.W. (1998): All wrapped up. *Nat. Struct. Biol.* **5**, 253-255.

Chothia, C., Levitt, M. & Richardson, D. (1981): Helix to helix packing in proteins. *J. Mol. Biol.* **145**, 215-250.

Cohen, G. E. (1997). ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallog.* **30**, 1160-1161.

CCP4. (1994): The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.

Creighton, T. E.: Protein folding. (W. H. Freeman), New York, 1992, pp. 284-285.

DeLano, W. L. (2002): The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.

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

van Dinten, L.C., Rensen, S., Gorbalenya, A.E. & Snijder, E.J. (1999): Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and Is essential for virus replication. *J. Virol.* **73**, 2027-2037.

Drosten, C., Günther, S., Preiser, W., van der Werf, S., Brodt, H., Becker, S., Rabenau, H.,

Panning, M., Kolesnikova, L., Fouchier, R.A.M., Berger, A., Burguière, A., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J., Müller, S., Rickerts, V., Stürmer, M., Vieth, S., Klenk, H., Osterhaus, A.D.M.E., Schmitz, H. & Doerr, H.W. (2003): Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967-1976.

Dudas, K.C. & Ruyechan, W.T. (1998): Identification of a region of the herpes simplex virus single-stranded DNA-binding protein involved in cooperative binding. *J. Virol.* **72**, 257-265.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., Ausio, J., Nissen, M.S., Reeves, R., Kang, C., Kissinger, C.R., Bailey, R.W., Griswold, M.D., Chiu, W., Garner, E.C. & Obradovic, Z. (2001): Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26-59.

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

Eleouet, J.F., Chilmonczyk, S., Besnardeau, L. & Laude, H. (1998): Transmissible gastroenteritis coronavirus induces programmed cell death in infected cells through a caspase-dependent pathway. *J. Virol.* **72**, 4918-4924.

Ellman, G.L. (1959): Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **82**, 70-77.

Emsley, P. & Cowtan, K. (2004): Coot: model-building tools for molecular graphics. *Acta Cryst.* **D60**, 2126-2132.

Fang, Y., Kim, D., Ropp, S., Steen, P., Christopher-Hennings, J., Nelson, E.A. & Rowland, R.R.R. (2004): Heterogeneity in Nsp2 of European-like porcine reproductive and respiratory syndrome viruses isolated in the United States. *Virus Res.* **100**, 229-235.

Fauquet, C.: Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses. CA (Elsevier Academic Press), San Diego, 2005.

Galán, C., Sola, I., Nogales, A., Thomas, B., Akoulitchev, A., Enjuanes, L. & Almazán, F. (2009): Host cell proteins interacting with the 3´end of TGEV coronavirus genome influence virus replication. *Virology* **391**, 304-314.

Goebel, S.J., Miller, T.B., Bennett, C.J., Bernard, K.A. & Masters, P.S. (2007): A hypervariable region within the 3' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis. *J. Virol.* **81**, 1274-1287.

Gorbalenya, A.E., Snijder, E.J. & Spaan, W.J.M. (2004): Severe acute respiratory syndrome coronavirus phylogeny: toward consensus. *J. Virol.* **78**, 7863-7866.

Gosert, R., Kanjanahaluethai, A., Egger, D., Bienz, K. & Baker, S.C. (2002): RNA replication of mouse hepatitis virus takes place at double-membrane vesicles. *J. Virol.* **76**, 3697-3708.

Graham, R.L., Sims, A.C., Baric, R.S. & Denison, M.R. (2006): The nsp2 proteins of mouse hepatitis virus and SARS coronavirus are dispensable for viral replication. *Adv. Exp. Med. Biol.* **581**, 67-72.

Groneberg, D.A., Hilgenfeld, R. & Zabel, P. (2005): Molecular mechanisms of severe acute respiratory syndrome (SARS). *Respir. Res.* **6**, 8-23.

de Haan, C.A.M. & Rottier, P.J.M. (2005): Molecular interactions in the assembly of coronaviruses. *Adv. Virus Res.* **64**, 165-230.

Harp, J.M., Timm, D.E. & Bunick, G.J. (1998): Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography. *Acta Cryst.* **D54**, 622-628.

van Hemert, M.J., de Wilde, A.H., Gorbalenya, A.E. & Snijder, E.J. (2008): The in vitro RNA synthesizing activity of the isolated arterivirus replication/transcription complex is dependent on a host factor. *J. Biol. Chem.* **283**, 16525-16536.

Herold, J., Gorbalenya, A.E., Thiel, V., Schelle, B. & Siddell, S.G. (1998): Proteolytic processing at the amino terminus of human coronavirus 229E gene 1-encoded polyproteins: identification of a papain-like proteinase and its substrate. *J. Virol.* **72**, 910-918.

Herold, J., Raabe, T., Schelle-Prinz, B. & Siddell, S.G. (1993): Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* **195**, 680-691.

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

Hofmann, H., Pyrc, K., van der Hoek, L., Geier, M., Berkhout, B. & Pöhlmann, S. (2005): Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proc. Natl. Acad. Sci. USA* **102**, 7988-7993.

Horowitz, P., Prasad, V. & Luduena, R.F. (1984): Bis(1,8-anilinonaphthalenesulfonate). A novel and potent inhibitor of microtubule assembly. *J. Biol. Chem.* **259**, 14647-14650.

Huang, P. & Lai, M.M. (1999): Polypyrimidine tract-binding protein binds to the complementary strand of the mouse hepatitis virus 3' untranslated region, thereby altering RNA conformation. *J. Virol.* **73**, 9110-9116.

Imai, Y., Kuba, K., Neely, G.G., Yaghubian-Malhami, R., Perkmann, T., van Loo, G., Ermolaeva, M., Veldhuizen, R., Leung, Y.H.C., Wang, H., Liu, H., Sun, Y., Pasparakis, M., Kopf, M., Mech, C., Bavari, S., Peiris, J.S.M., Slutsky, A.S., Akira, S., Hultqvist, M., Holmdahl, R., Nicholls, J., Jiang, C., Binder, C.J. & Penninger, J.M. (2008): Identification of oxidative stress and Toll-like receptor 4 signaling as a key pathway of acute lung injury. *Cell* **133**, 235-249.

Imbert, I., Guillemot, J., Bourhis, J., Bussetta, C., Coutard, B., Egloff, M., Ferron, F., Gorbalenya, A.E. & Canard, B. (2006): A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J.* **25**, 4933-4942.

Imbert, I., Snijder, E.J., Dimitrova, M., Guillemot, J., Lécine, P. & Canard, B. (2008): The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. *Virus Res.* **133**, 136-148.

Jancarik, J., Pufan, R., Hong, C., Kim, S.H. & Kim, R. (2004): Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. *Acta Cryst.* **D60**, 1670-1673.

Joseph, J.S., Saikatendu, K.S., Subramanian, V., Neuman, B.W., Brooun, A., Griffith, M., Moy, K., Yadav, M.K., Velasquez, J., Buchmeier, M.J., Stevens, R.C. & Kuhn, P. (2006): Crystal structure of nonstructural protein 10 from the severe acute respiratory syndrome coronavirus reveals a novel fold with two zinc-binding motifs.

J. Virol. **80**, 7894-7901.

Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., Kubo, H. & Makino, S. (2006): Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc. Natl. Acad. Sci. USA* **103**, 12885-12890.

Kaul, P., Biagioli, M.C., Singh, I. & Turner, R.B. (2000): Rhinovirus-induced oxidative stress and interleukin-8 elaboration involves p47-phox but is independent of attachment to intercellular adhesion molecule-1 and viral replication. *J. Infect. Dis.* **181**, 1885-1890.

Kleiger, G., Grothe, R., Mallick, P. & Eisenberg, D. (2002): GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry* **41**, 5990-5997.

Knipe, D.M., Quinlan, M.P. & Spang, A.E. (1982): Characterization of two conformational forms of the major DNA-binding protein encoded by herpes simplex virus 1. *J. Virol.* **44**, 736-741.

Knoops, K., Kikkert, M., Worm, S.H.E.V.D., Zevenhoven-Dobbe, J.C., van der Meer, Y., Koster, A.J., Mommaas, A.M. & Snijder, E.J. (2008): SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol.* **6**, e226.

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

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

Kumar, P., Gunalan, V., Liu, B., Chow, V.T.K., Druce, J., Birch, C., Catton, M., Fielding, B.C., Tan, Y. & Lal, S.K. (2007): The nonstructural protein 8 (nsp8) of the SARS coronavirus interacts with its ORF6 accessory protein. *Virology* **366**, 293-303.

Kussie, P.H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A.J. & Pavletich, N.P. (1996): Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **274**, 948-53.

Lai, M.M. & Cavanagh, D. (1997): The molecular biology of coronaviruses. *Adv. Virus Res.* **48**, 1-100.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993): PROCHECK - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.

Lau, S.K.P., Woo, P.C.Y., Li, K.S.M., Huang, Y., Tsoi, H., Wong, B.H.L., Wong, S.S.Y., Leung, S., Chan, K. & Yuen, K. (2005): Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. USA* **102**, 14040-14045.

Lawrence, M.C. & Colman, P.M. (1993): Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946-950.

Lee, B. & Richards, F.M. (1971): The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

Lemm, J.A., Rümenapf, T., Strauss, E.G., Strauss, J.H. & Rice, C.M. (1994): Polypeptide requirements for assembly of functional Sindbis virus replication complexes: a model for the temporal regulation of minus- and plus-strand RNA synthesis. *EMBO J.* **13**, 2925-2934.

Li, H.P., Huang, P., Park, S. & Lai, M.M. (1999): Polypyrimidine tract-binding protein binds to the leader RNA of mouse hepatitis virus and serves as a regulator of viral transcription. *J. Virol.* **73**, 772-777.

Li, H.P., Zhang, X., Duncan, R., Comai, L. & Lai, M.M. (1997): Heterogeneous nuclear ribonucleoprotein A1 binds to the transcription-regulatory region of mouse hepatitis virus RNA. *Proc. Natl. Acad. Sci. USA* **94**, 9544-9549.

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

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., Choe, H. & Farzan, M. (2003): Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-454.

Lin, Y.J., Liao, C.L. & Lai, M.M. (1994): Identification of the cis-acting signal for minus-strand RNA synthesis of a murine coronavirus: implications for the role of minus-strand RNA in RNA replication and transcription. *J. Virol.* **68**, 8131-8140.

Lindwall, G., Chau, M., Gardner, S.R. & Kohlstaedt, L.A. (2000): A sparse matrix approach to the solubilization of overexpressed proteins. *Protein Eng.* **13**, 67-71.

Liu, M., Mao, X., Ye, C., Huang, H., Nicholson, J.K. & Lindon, J.C. (1998): Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy. *J. Magn. Reson.* **132**, 125-129.

Mackenzie, J. (2005): Wrapping things up about virus RNA replication. *Traffic* **6**, 967-977.

Makhov, A.M., Sen, A., Yu, X., Simon, M.N., Griffith, J.D. & Egelman, E.H. (2008): The bipolar filaments formed by herpes simplex virus type 1 SSB/Recombination Protein (ICP8) Suggest a Mechanism for DNA Annealing. *J. Mol. Biol.* **386**, 273-279.

Mapelli, M., Panjikar, S. & Tucker, P.A. (2005): The crystal structure of the herpes simplex virus 1 ssDNA-binding protein suggests the structural basis for flexible, cooperative single-stranded DNA binding. *J. Biol. Chem.* **280**, 2990-2997.

van Marle, G., van Dinten, L.C., Spaan, W.J., Luytjes, W. & Snijder, E.J. (1999): Characterization of an equine arteritis virus replicase mutant defective in subgenomic mRNA synthesis. *J. Virol.* **73**, 5274-5281.

Matthes, N., Mesters, J.R., Coutard, B., Canard, B., Snijder, E.J., Moll, R. & Hilgenfeld, R. (2006): The non-structural protein Nsp10 of mouse hepatitis virus binds zinc ions and nucleic acids. *FEBS Lett.* **580**, 4143-4149.

Matthews, B.W. (1968): Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

McBride, A.A., Klausner, R.D. & Howley, P.M. (1992): Conserved cysteine residue in the

DNA-binding domain of the bovine papillomavirus type 1 E2 protein confers redox regulation of the DNA-binding activity in vitro. *Proc. Natl. Acad. Sci. USA* **89**, 7531-7535.

McCoy, A.J., Grosse-Kunstleve, R.W., Storoni, L.C. & Read, R.J. (2005): Likelihood-enhanced fast translation functions. *Acta Cryst.* **D61**, 458-464.

Meador, W.E., Means, A.R. & Quiocho, F.A. (1992): Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science* **257**, 1251-1255.

Mihindukulasuriya, K.A., Wu, G., St Leger, J., Nordhausen, R.W. & Wang, D. (2008): Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J. Virol.* **82**, 5084-5088.

Mikhailov, V.S., Okano, K. & Rohrmann, G.F. (2005): The redox state of the baculovirus single-stranded DNA-binding protein LEF-3 regulates its DNA binding, unwinding, and annealing activities. *J. Biol. Chem.* **280**, 29444-29453.

Miknis, Z.J., Donaldson, E.F., Umland, T.C., Rimmer, R.A., Baric, R.S. & Schultz, L.W. (2009): SARS-CoV nsp9 Dimerization is Essential for Efficient Viral Growth. *J. Virol.* **83**, 3007-3018.

Miller, S. & Krijnse-Locker, J. (2008): Modification of intracellular membrane structures for virus replication. *Nat. Rev. Microbiol.* **6**, 363-374.

Miseta, A. & Csutora, P. (2000): Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Mol. Biol. Evol.* **17**, 1232-1239.

Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997): Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53**, 240-255.

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

Nanda, S.K. & Leibowitz, J.L. (2001): Mitochondrial aconitase binds to the 3' untranslated region of the mouse hepatitis virus genome. *J. Virol.* **75**, 3352-3362.

Nelson, G.W., Stohlman, S.A. & Tahara, S.M. (2000): High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. *J. Gen. Virol.* **81**, 181-188.

Neuman, B.W., Adair, B.D., Yoshioka, C., Quispe, J.D., Orca, G., Kuhn, P., Milligan, R.A., Yeager, M. & Buchmeier, M.J. (2006): Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J. Virol.* **80**, 7918-7928.

Neuman, B.W., Joseph, J.S., Saikatendu, K.S., Serrano, P., Chatterjee, A., Johnson, M.A., Liao, L., Klaus, J.P., Yates, J.R., Wüthrich, K., Stevens, R.C., Buchmeier, M.J. & Kuhn, P. (2008): Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J. Virol.* **82**, 5279-5294.

Novoa, R.R., Calderita, G., Arranz, R., Fontana, J., Granzow, H. & Risco, C. (2005): Virus factories: associations of cell organelles for viral replication and morphogenesis. *Biol. Cell* **97**, 147-172.

Oldfield, C.J., Ulrich, E.L., Cheng, Y., Dunker, A.K. & Markley, J.L. (2005): Addressing the intrinsic disorder bottleneck in structural proteomics. *Proteins* **59**, 444-453.

Oostra, M., Hagemeijer, M.C., van Gent, M., Bekker, C.P.J., te Lintelo, E.G., Rottier, P.J.M. & de Haan, C.A.M. (2008): Topology and membrane anchoring of the

coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J. Virol.* **82**, 12392-12405.

Otwinowski, Z. & Minor, W. (1997): Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 301-326.

Painter, J. & Merritt, E.A. (2006): TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallog.* **39**, 109-111.

Pan, J., Peng, X., Gao, Y., Li, Z., Lu, X., Chen, Y., Ishaq, M., Liu, D., Dediego, M.L., Enjuanes, L. & Guo, D. (2008): Genome-wide analysis of protein-protein interactions and involvement of viral proteins in SARS-CoV replication. *PLoS ONE* **3**, e3299.

Parks, D., Bolinger, R. & Mann, K. (1997): Redox state regulates binding of p53 to sequence-specific DNA, but not to non-specific or mismatched DNA. *Nucleic Acids Res.* **25**, 1289-1295.

Pasternak, A.O., van den Born, E., Spaan, W.J. & Snijder, E.J. (2001): Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. *EMBO J.* **20**, 7220-7228.

Pasternak, A.O., Spaan, W.J.M. & Snijder, E.J. (2006): Nidovirus transcription: how to make sense...? *J. Gen. Virol.* **87**, 1403-1421.

Peiris, J.S.M., Chu, C.M., Cheng, V.C.C., Chan, K.S., Hung, I.F.N., Poon, L.L.M., Law, K.I., Tang, B.S.F., Hon, T.Y.W., Chan, C.S., Chan, K.H., Ng, J.S.C., Zheng, B.J., Ng, W.L., Lai, R.W.M., Guan, Y. & Yuen, K.Y. (2003): Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**, 1767-1772.

Pestova, T.V. & Hellen, C.U. (2001): Functions of eukaryotic factors in initiation of translation. *Cold Spring Harb. Symp. Quant. Biol.* **66**, 389-396.

Pestova, T.V., Kolupaeva, V.G., Lomakin, I.B., Pilipenko, E.V., Shatsky, I.N., Agol, V.I. & Hellen, C.U. (2001): Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. USA* **98**, 7029-7036.

Petersen, M.T., Jonson, P.H. & Petersen, S.B. (1999): Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein Eng.* **12**, 535-548.

Piotrowski, Y., van der Hoek, L., Pyrc, K., Berkhout, B., Moll, R. & Hilgenfeld, R. (2006): Nonstructural proteins of human coronavirus NL63. *Adv. Exp. Med. Biol.* **581**, 97-100.

Piotrowski, Y., Ponnusamy, R., Glaser, S., Daabach, A., Moll, R. & Hilgenfeld, R. (2008): Production of coronavirus nonstructural proteins in soluble form for crystallization. *Methods Mol. Biol.* **454**, 139-159.

Plant, E.P. & Dinman, J.D. (2008): The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front. Biosci.* **13**, 4873-4881.

Plaxco, K.W. & Gross, M. (1997): Cell biology. The importance of being unfolded. *Nature* **386**, 657-659.

Ponnusamy, R., Mesters, J.R., Ziebuhr, J., Moll, R. & Hilgenfeld, R. (2006): Non structural proteins 8 and 9 of human coronavirus 229E. *Adv. Exp. Med. Biol.* **581**, 49-54.

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

Pontius, B.W. (1993): Close encounters: why unstructured, polymeric domains can

increase rates of specific macromolecular association. *Trends Biochem. Sci.* **18**, 181-186.

Prentice, E., Jerome, W.G., Yoshimori, T., Mizushima, N. & Denison, M.R. (2004): Coronavirus replication complex formation utilizes components of cellular autophagy. *J. Biol. Chem.* **279**, 10136-10141.

Prentice, E., McAuliffe, J., Lu, X., Subbarao, K. & Denison, M.R. (2004): Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. *J. Virol.* **78**, 9977-9986.

Prevot, D., Darlix, J. & Ohlmann, T. (2003): Conducting the initiation of protein synthesis: the role of eIF4G. *Biol. Cell* **95**, 141-156.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I. & Sussman, J.L. (2005): FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**, 3435-3438.

Raghunathan, S., Kozlov, A.G., Lohman, T.M. & Waksman, G. (2000): Structure of the DNA binding domain of E. coli SSB bound to ssDNA. *Nat. Struct. Biol.* **7**, 648-652.

Raman, S., Bouma, P., Williams, G.D. & Brian, D.A. (2003): Stem-loop III in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. *J. Virol.* **77**, 6720-6730.

Rehm, T., Huber, R. & Holak, T.A. (2002): Application of NMR in structural proteomics: screening for proteins amenable to structural analysis. *Structure* **10**, 1613-1618.

Rivera, C.I. & Lloyd, R.E. (2008): Modulation of enteroviral proteinase cleavage of poly(A)-binding protein (PABP) by conformation and PABP-associated factors. *Virology* **375**, 59-72.

Roberts, A., Vogel, L., Guarner, J., Hayes, N., Murphy, B., Zaki, S. & Subbarao, K. (2005): Severe acute respiratory syndrome coronavirus infection of golden Syrian hamsters. *J. Virol.* **79**, 503-511.

Robertson, M.P., Igel, H., Baertsch, R., Haussler, D., Ares, M. & Scott, W.G. (2005): The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol.* **3**, e5.

Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J., Garner, E., Guilliot,, S. & Dunker, A. (1998): Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* **3**, 437-448.

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J. & Dunker, A.K. (2001): Sequence complexity of disordered protein. *Proteins* **42**, 38-48.

Rosenfeld, R., Vajda, S. & DeLisi, C. (1995): Flexible docking and design. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 677-700.

Russ, W.P. & Engelman, D.M. (2000): The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* **296**, 911-919.

Sampson, D.A., Arana, M.E. & Boehmer, P.E. (2000): Cysteine 111 affects coupling of single-stranded DNA binding to ATP hydrolysis in the herpes simplex virus type-1 origin-binding protein. *J. Biol. Chem.* **275**, 2931-2937.

Sawicki, D., Wang, T. & Sawicki, S. (2001): The RNA structures engaged in replication and transcription of the A59 strain of mouse hepatitis virus. *J. Gen. Virol.* **82**, 385-396.

Sawicki, S.G. & Sawicki, D.L. (1995): Coronaviruses use discontinuous extension for

synthesis of subgenome-length negative strands. *Adv. Exp. Med. Biol.* **380**, 499-506.

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

Senes, A., Gerstein, M. & Engelman, D.M. (2000): Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. *J. Mol. Biol.* **296**, 921-936.

Sethna, P.B., Hung, S.L. & Brian, D.A. (1989): Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc. Natl. Acad. Sci. USA* **86**, 5626-5630.

Siddell, S.G., Ziebuhr, J. & Snijder, E.J.: Coranaviruses, Toroviruses and Arteriviruses. In: *Topley and Wilson's Microbiology and Microbial Infections* (Cox, F. *et al.,* eds.), Edward Arnold, London, 2005, pp. 823-856.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L.M., Guan, Y., Rozanov, M., Spaan, W.J.M. & Gorbalenya, A.E. (2003): Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **331**, 991-1004.

Snijder, E.J., van der Meer, Y., Zevenhoven-Dobbe, J., Onderwater, J.J.M., van der Meulen, J., Koerten, H.K. & Mommaas, A.M. (2006): Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. *J. Virol.* **80**, 5927-5940.

Spagnolo, J.F. & Hogue, B.G. (2000): Host protein interactions with the 3' end of bovine coronavirus RNA and the requirement of the poly(A) tail for coronavirus defective genome replication. *J. Virol.* **74**, 5053-5065.

Spolar, R.S. & Record, M.T. (1994): Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-784.

Stanhope, M.J., Brown, J.R. & Amrine-Madsen, H. (2004): Evidence from the evolutionary analysis of nucleotide sequences for a recombinant history of SARS-CoV. *Infect. Genet. Evol.* **4**, 15-19.

Stevens, R., Stevens, L. & Price, N.C. (1983): The stabilities of various thiol compounds used in protein purifications. *Biochem. Educ.* **11**, 70.

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

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

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

Theobald, D.L., Mitton-Fry, R.M. & Wuttke, D.S. (2003): Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 115-133.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994): CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.

Tusell, S.M., Schittone, S.A. & Holmes, K.V. (2007): Mutational analysis of aminopeptidase N, a receptor for several group 1 coronaviruses, identifies key determinants of viral host range. *J. Virol.* **81**, 1261-1273.

Vagner, S., Galy, B. & Pyronnet, S. (2001): Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites. *EMBO Rep.* **2**, 893-898.

Vassilev, L.T. (2004a): Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics. *Cell Cycle* **3**, 419-421.

Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., Fotouhi, N. & Liu, E.A. (2004b): In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844-848.

Wang, Y., Oberley, L.W. & Murhammer, D.W. (2001): Evidence of oxidative stress following the viral infection of two lepidopteran insect cell lines. *Free Radic. Biol. Med.* **31**, 1448-1455.

Wang, H., Yang, P., Liu, K., Guo, F., Zhang, Y., Zhang, G. & Jiang, C. (2008): SARS coronavirus entry into host cells through a novel clathrin- and caveolae-independent endocytic pathway. *Cell Res.* **18**, 290-301.

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

Webster, G., Genschel, J., Curth, U., Urbanke, C., Kang, C. & Hilgenfeld, R. (1997): 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 Lett.* **411**, 313-316.

Williams, R.K., Jiang, G.S. & Holmes, K.V. (1991): Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc. Natl. Acad. Sci. USA* **88**, 5533-6.

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

Wright, P.E. & Dyson, H.J. (1999): Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321-331.

Wu, Y., Tseng, C., Cheng, M., Ho, H., Shih, S. & Chiu, D.T. (2008): Glucose-6-phosphate dehydrogenase deficiency enhances human coronavirus 229E infection. *J. Infect. Dis.* **197**, 812-816.

Xiao, H., Xu, L.H., Yamada, Y. & Liu, D.X. (2008): Coronavirus spike protein inhibits host cell translation by interaction with eIF3f. *PLoS ONE* **3**, e1494.

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G.F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003): The crystal structures of severe acute respiratory syndrome virus main protease and its

complex with an inhibitor. *Proc. Natl. Acad. Sci. USA* **100**, 13190-13195.

You, J.S., Wang, M. & Lee, S.H. (2000): Functional characterization of zinc-finger motif in redox regulation of RPA-ssDNA interaction. *Biochemistry* **39**, 12953-12958.

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

Zhong, N., Zhang, S., Zou, P., Chen, J., Kang, X., Li, Z., Liang, C., Jin, C. & Xia, B. (2008): Without its N-finger, the main protease of severe acute respiratory syndrome coronavirus can form a novel dimer through its C-terminal domain. *J. Virol.* **82**, 4227-4234.

Ziebuhr, J. & Siddell, S.G. (1999): Processing of the human coronavirus 229E replicase polyproteins by the virus-encoded 3C-like proteinase: identification of proteolytic products and cleavage sites common to pp1a and pp1ab. *J. Virol.* **73**, 177-185.

Zuniga, S., Sola, I., Alonso, S. & Enjuanes, L. (2004): Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J. Virol.* **78**, 980-994.

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

7. Appendix

7.1 Wild-type HCoV-229E Nsp9
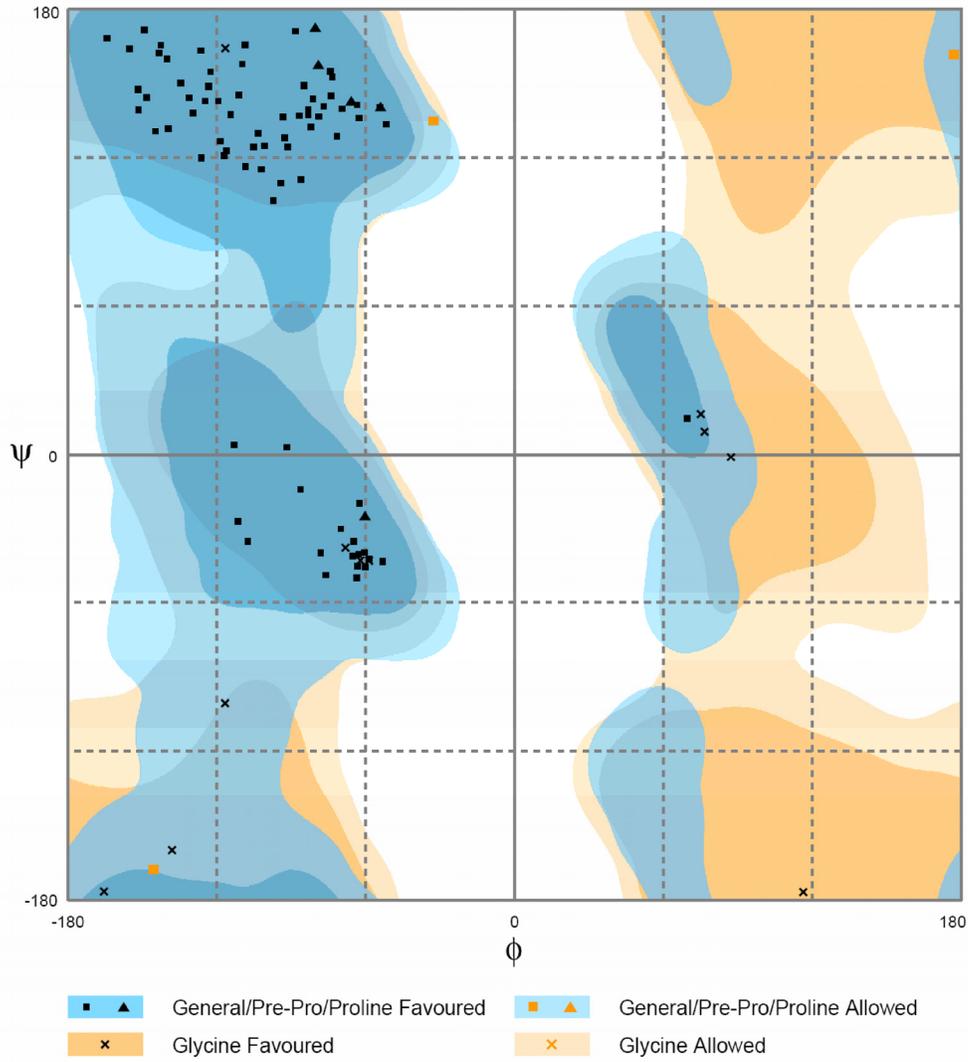
Fig. 7.1 Ramachandran plot of the final wild-type HCoV-229E Nsp9 model.

Table 7.1 Data collection and refinement statistics of wild-type HCoV-229E Nsp9

| Data collection	Wild-type HCoV-229E Nsp9
Wavelength (Å)	0.8075
Resolution (Å)	40.0-1.75 (1.79-1.75)
Space group	P622
Unit-cell parameters	
a (Å)	85.63
b (Å)	85.63
c (Å)	48.69
Solvent content	42.3%
Overall reflections	129,656
Unique reflections	11,317 (730)
Multiplicity	11.5 (11.5)
Completeness (%)	99.9 (100.0)
R_{merge}[1] (%)	8.3 (60.3)
$I/\sigma(I)$	13.7 (4.45)
Refinement	
Resolution (Å)	40.0-1.75
R_{cryst}[2]	0.190
R_{free}[2]	0.224
R.m.s.d. from ideal geometry	
bonds (Å)	0.013
angles (Å)	1.417
Protein atoms	778
Solvent atoms	65
MPD	1
Sulfate	2
Ramachandran plot	
most favoured (%)	96.8
additionally allowed (%)	3.2
disallowed regions (%)	0

7.2 HCoV-229E Nsp9 Cys69Ala mutant
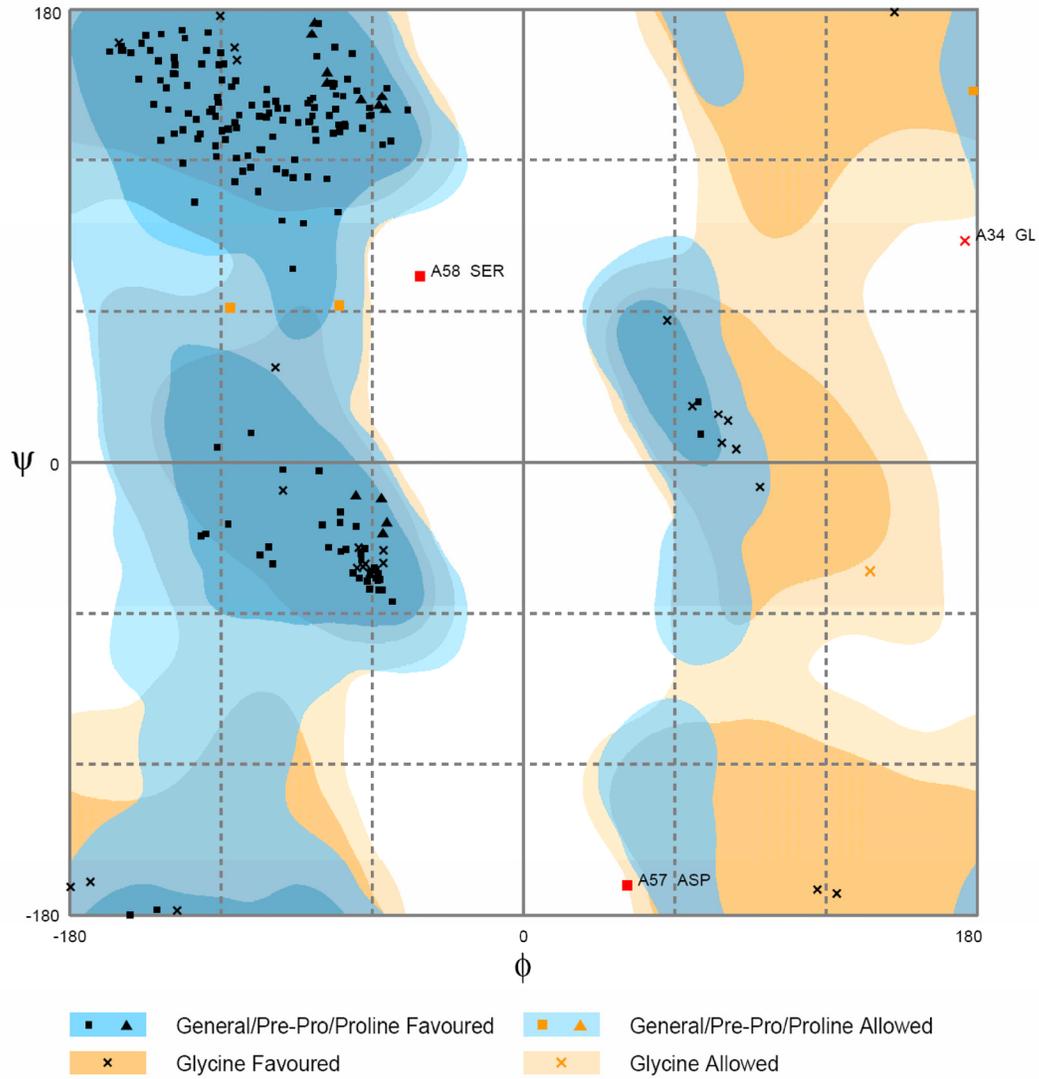
Fig. 7.2 Ramachandran plot of the final HCoV-229E Nsp9 Cys69Ala mutant model

Table 7.2 Data collection and refinement statistics of HCoV-229E Nsp9 Cys69Ala mutant

| Data collection	Nsp9 Cys69Ala mutant
Wavelength (Å)	0.8075
Resolution (Å)	30.0-1.80 (1.86-1.80)
Space group	$P2_12_12_1$
Unit-cell parameters	
a (Å)	26.40
b (Å)	61.38
c (Å)	107.31
Solvent content	31.5%
Overall reflections	139,726
Unique reflections	16,842 (1648)
Multiplicity	8.3 (4.7)
Completeness (%)	99.4 (99.7)
R_{merge}[1] (%)	8.9 (35.1)
$I/\sigma(I)$	19.7 (3.96)
Refinement	
Resolution (Å)	30.0-1.80
R_{cryst}[2]	0.221
R_{free}[2]	0.281
R.m.s.d. from ideal geometry	
bonds (Å)	0.017
angles (Å)	1.962
Protein atoms	1604
Solvent atoms	74
DTT	1
Ramachandran plot	
most favoured (%)	96.6
additionally allowed (%)	2.0
disallowed regions (%)	1.5

7.3 TGEV Nsp9
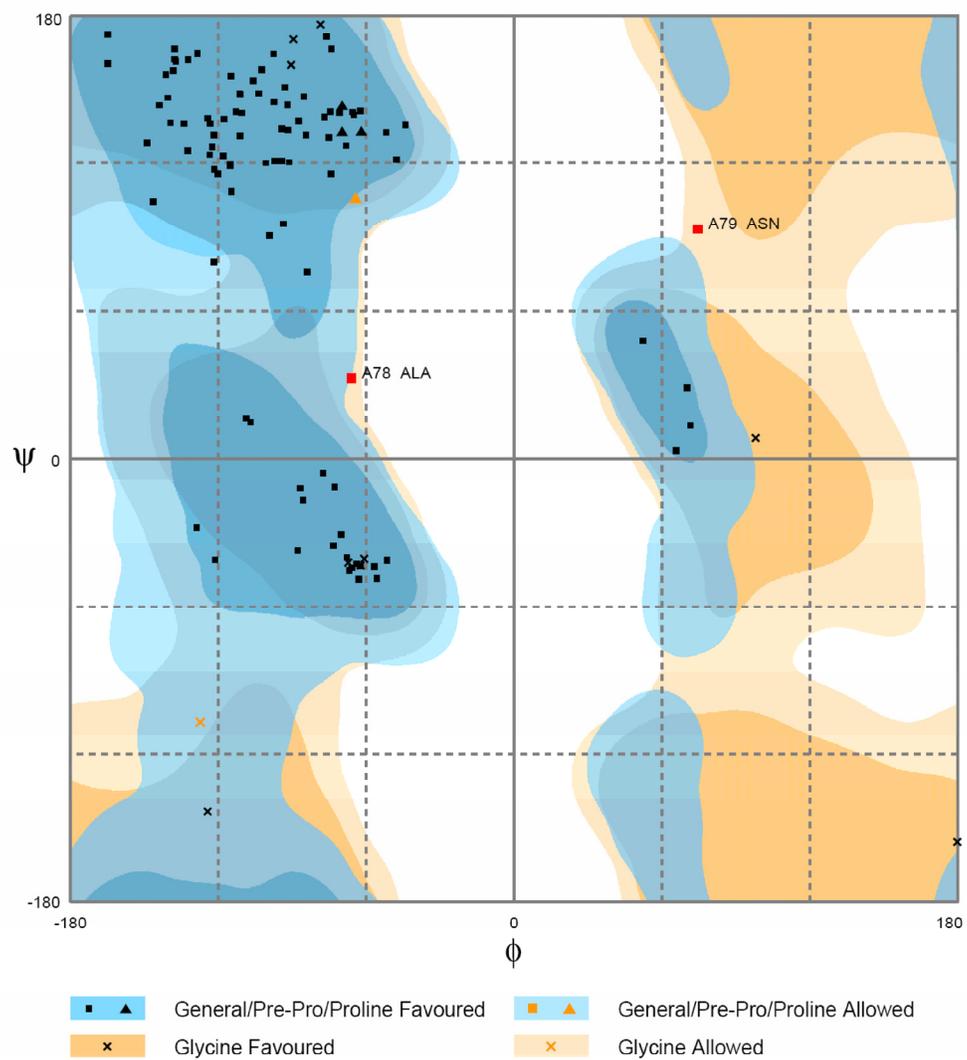
Fig. 7.3 Ramachandran plot of the final TGEV Nsp9 model

Table 7.3 Data collection and refinement statistics of TGEV Nsp9

Data collection	TGEV Nsp9
Wavelength (Å)	1.2549
Resolution (Å)	44.63-2.41 (2.54-2.41)
Space group	I422
Unit-cell parameters	
a (Å)	89.25
b (Å)	89.25
c (Å)	74.85
Solvent content	58.5%
Overall reflections	56,655
Unique reflections	6,076 (880)
Multiplicity	9.3 (9.4)
Completeness (%)	99.7 (100.0)
R_{merge}^{1} (%)	7.9 (61.1)
I/σ(I)	24.6 (5.2)
Refinement	
Resolution (Å)	44.63-2.41
R_{cryst}^{2}	0.195
R_{free}^{2}	0.250
R.m.s.d. from ideal geometry	
bonds (Å)	0.018
angles (Å)	1.598
Protein atoms	857
Solvent atoms	35
EDO	1
CL	3
Ramachandran plot	
most favoured (%)	96.3
additionally allowed (%)	1.9
disallowed regions (%)	1.9

7.4 SARS-CoV Nsp9

Table 7.4 Data collection statistics of SARS-CoV Nsp9

Data collection	SARS-CoV Nsp9
Wavelength (Å)	0.8075
Resolution (Å)	31.75-2.80 (2.95-2.80)
Space group	P4
Unit-cell parameters	
a (Å)	40.11
b (Å)	40.11
c (Å)	95.24
Solvent content	58.3%
Overall reflections	12,233
Unique reflections	3,753 (546)
Multiplicity	3.3 (3.3)
Completeness (%)	99.8 (100.0)
R_{merge}[1] (%)	8.5 (24.6)
I/σ(I)	14.0 (6.9)

7.5 Footnotes

Values in parentheses are for the highest resolution shell.

[1] $R_{\mathrm{merge}} = \sum_{hkl} \sum_i | I(hkl)_i - \langle I(hkl) \rangle | / \sum_{hkl} \sum_i I(hkl)_i$, where *I(hkl)* is the intensity of reflection hkl and $\langle I(hkl) \rangle$ is the average intensity over all equivalent reflections.

[2] $R_{cryst} = \sum_{hkl} |F_o(hkl) - F_c(hkl)| / \sum_{hkl} F_o(hkl)$. R_{free} was calculated for a test set of reflections (5%-10%) omitted from the refinement.

7.6 r.m.s deviation of Nsp9s (ALIGN Program; Cohen, 1997)

a) Monomer structure

Structure (C^α)	C69A mut A	C69A mut B	SARS 1QZ8 A	SARS 1QZ8 B	1UW7	TGEV Nsp9
HCoV-229E Nsp9wt	0.71 Å (92)	0.67 Å (87)	0.75 Å (84)	0.66 Å (84)	1.75 Å (94)	1.05 Å (86)
HCoV-229E C69A mut A		0.96 Å (99)	0.86 Å (87)	0.76 Å (92)	1.39 Å (94)	0.89 Å (89)
HCoV-229E C69A mut B			1.23 Å (92)	0.99 Å (92)	1.97 Å (98)	1.13 Å (90)
SARS 1QZ8 monomer-A				0.64 Å (100)	1.16 Å (108)	0.86 Å (93)
SARS 1QZ8 monomer-B					0.91 Å (96)	1.22 Å (99)
1UW7						1.46 Å (101)

b) α-helix mediated dimer structure

Structure (C^α)	SARS-CoV 1QZ8	SARS-CoV 1UW7	TGEV Nsp9
HCoV-229E C69Amut	0.99 Å (175)	2.76 Å (197)	1.20 Å (180)
SARS-CoV 1QZ8		2.08 Å (206)	1.06 Å (187)
SARS-CoV 1UW7			2.13 Å (202)

c) β-sheet mediated dimer structure

Structure (C^α)	SARS-CoV 1QZ8	SARS-CoV 1UW7
HCoV-229E C69Amut	4.29 Å (171)	3.85 Å (166)
SARS-CoV 1QZ8		1.62 Å (213)

Lebenslauf

Persönliche Daten

Name:	Rajesh Ponnusamy
Geburtsdatum und -ort:	28.05.1982 in Thirukkovilur, Indien
Familienstand:	ledig

Schulische Ausbildung

Juni 1986 bis Mai 1990:	Samad Secondary School, Tiruchirappalli, Indien
Juni 1990 bis Mai 1995:	Vailankanni Matriculation Higher Secondary School, Tiruchirappalli, Indien
Juni 1995 bis Mai 1997:	Kajamian Higher Secondary School, Tiruchirappalli, Indien
Juni 1997 bis Mai 1999:	Periyar Centenary Memorial Matriculation Higher Secondary School, Tiruchirappalli, Indien

Studium

August 1999 bis Mai 2002:	St. Joseph´s College, (Bharathidasan University), Tiruchirappalli, Indien Abschluß: Bachelor of Science in Physics
August 2002 bis Mai 2004:	University of Madras, Chennai, Indien Abschluß: Master of Science in Biophysics
Seit August 2004:	Anfertigung der vorliegenden Dissertation unter der Anleitung von Prof. Dr. Dr. h.c. Rolf Hilgenfeld am Institut für Biochemie, Universität zu Lübeck, Germany.

Publications

Asanithi, P., Saridakis, E., Govada, L., Jurewicz, I., Brunner, E.W., Ponnusamy, R., Cleaver, J.A.S., Dalton, A.B., Chayen, N.E. and Sear, R.P. (2009): Carbon-Nanotube-Based Materials for Protein Crystallization. *ACS Appl. Mater. Interfaces.* **6**, 1203-1210.

Ponnusamy, R., Moll, R., Weimar, T., Mesters, J.R. and Hilgenfeld, R. (2008): Variable oligomerization modes in coronavirus non-structural protein 9. *J. Mol. Biol.* **383**, 1081-1096. *(Feature on the cover of the JMB)*

Piotrowski, Y., Ponnusamy, R., Glaser, S., Daabach, A., Moll, R. and Hilgenfeld, R. (2008): Production of coronavirus nonstructural proteins in soluble form for crystallization. In *Methods in Molecular Biology* **454**: SARS- and other coronaviruses (ed. D. Cavanagh), pp. 139-159. Humana Press, Totowa, NJ.

Ponnusamy, R., Mesters, J.R., Ziebuhr, J., Moll, R. and Hilgenfeld, R. (2006): Non structural proteins 8 and 9 of human coronavirus 229E. *Adv. Exp. Med. Biol.* **581**, 49-54.

Workshops and Conferences

1. Workshop on characterization of protein complexes in structural biology. 30/06/2009 – 03/07/2009. Hamburg, Germany. *(Poster presentation)*

2. Deutschen Gesellschaft für kristallographie. 09/03/2009 – 12/03/2009. Hannover, Germany. *(Poster presentation)*

3. The XIth International Nidovirus symposium. 22/06/2008 – 27/06/2008. St.Catherine´s College, Oxford, United Kingdom. *(Poster presentation)*

4. Third European Congress of Virology. 01/09/2007 – 05/09/2007. Nürnberg, Germany. *(Poster presentation)*

5. EMBO workshop RNA viruses: replication, evolution and drug design. 27/08/2007 – 30/08/2007. Vienna, Austria. *(Poster presentation)*

6. Workshop on diffraction data collection using synchrotron radiation. 23/08/2007 – 25/08/2007. Bessy, Berlin, Germany. *(Poster presentation)*

7. CCP4 study weekend 2007. Topic: Molecular Replacement. 05/01/2007 – 06/01/2007. University of Reading, United Kingdom.

8. European crystallography Meeting 23. 06/08/2006 – 11/08/2006. Leuven, Belgium. *(Poster presentation)*

9. International Workshop on Discovery of Antiviral compounds. 25/04/2006 – 30/04/2006. Lübeck, Germany. *(Poster presentation)*

10. Trends in macromolecular structure and function. 18/01/2006 – 20/01/2006. Chennai, India. *Oral presentation:* Structure of non-structural protein 9 from human coronavirus 229E.

11. CCP4 study weekend 2006. Topic: Crystallography of Complexes. 06/01/2006 – 07/01/2006. University of Leeds, United Kingdom.