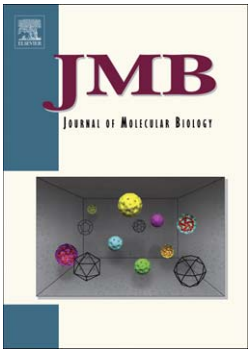

Variable oligomerization modes in coronavirus non-structural protein 9

Rajesh Ponnusamy, Ralf Moll, Thomas Weimar, Jeroen R. Mesters, Rolf Hilgenfeld

PII: S0022-2836(08)00940-6
DOI: doi: 10.1016/j.jmb.2008.07.071
Reference: YJMBI 60680

To appear in: *Journal of Molecular Biology*

Received date: 13 April 2008
Revised date: 17 July 2008
Accepted date: 24 July 2008

Please cite this article as: Ponnusamy, R., Moll, R., Weimar, T., Mesters, J.R. & Hilgenfeld, R., Variable oligomerization modes in coronavirus non-structural protein 9, *Journal of Molecular Biology* (2008), doi: 10.1016/j.jmb.2008.07.071

Variable oligomerization modes in coronavirus non-structural protein 9

Rajesh Ponnusamy[1], Ralf Moll[1], Thomas Weimar[2], Jeroen R. Mesters[1] and Rolf Hilgenfeld[1,3]*

Institutes of [1]Biochemistry and [2]Chemistry, Center of Structural and Cell Biology in Medicine, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany; [3]Laboratory for Structural Biology of Infection and Inflammation, c/o DESY, Building 22a, Notkestr. 85, 22603 Hamburg, Germany.

Author for correspondence:
Prof. Rolf Hilgenfeld
Institute of Biochemistry
University of Lübeck
Ratzeburger Allee 160
23538 Lübeck
Germany
phone: +49-451-500-4060
fax: +49-451-500-4068
e-mail: hilgenfeld@biochem.uni-luebeck.de

Running title: Coronavirus Nsp9

Summary

Non-structural protein 9 (Nsp9) of coronaviruses is believed to bind single-stranded RNA in the viral replication complex. The crystal structure of Nsp9 of Human Coronavirus (HCoV) 229E reveals a novel disulfide-linked homodimer, which is very different from the previously reported Nsp9 dimer of SARS coronavirus. In contrast, the structure of the Cys69Ala mutant of HCoV-229E Nsp9 shows the same dimer organization as the SARS-CoV protein. In the crystal, the wild-type HCoV-229E protein forms a trimer of dimers, whereas the mutant and SARS-CoV Nsp9 are organized in rod-like polymers. Chemical cross-linking suggests similar modes of aggregation in solution. In zone-interference gel electrophoresis assays and surface plasmon resonance experiments, the HCoV-229E wild-type protein is found to bind oligonucleotides with relatively high affinity, whereas binding by the Cys69Ala and Cys69Ser mutants is only observed for the longest oligonucleotides. The corresponding mutations in SARS-CoV Nsp9 do not hamper nucleic-acid binding. From the crystal structures, a model for single-stranded RNA-binding by Nsp9 is deduced. We propose that both forms of the Nsp9 dimer are biologically relevant; the occurrence of the disulfide-bonded form may be correlated with oxidative stress induced in the host cell by the viral infection.

Keywords: Nsp9, dimerization, nucleic-acid binding, disulfide bond, oxidative stress

Abbreviations used: SARS-CoV, severe acute respiratory syndrome coronavirus; HCoV-229E, human coronavirus 229E; Nsp, non-structural protein; DTT, dithiothreitol;

MPD, 2-methyl-2,4-pentanediol; DLS, dynamic light scattering; M^{pro}, main proteinase; r.m.s., root-mean-square; SSB, single-stranded DNA-binding protein; SPR, surface plasmon resonance

Introduction

Since a coronavirus was identified as the causative agent of the 2003 outbreak of severe acute respiratory syndrome (SARS),[1-4] scientific interest in this family of viruses has increased dramatically.[5] Coronaviruses are enveloped, positive-strand RNA viruses that cause a wide spectrum of disease in humans and animals. Based on the genome organization and on phylogenetic analysis, these viruses are divided into three distinct groups. Human coronavirus 229E (HCoV 229E) belongs to group 1 and causes a mild form of the common cold. Group 1 also includes the recently discovered human coronavirus NL63 [6] and the porcine coronavirus, Transmissible Gastroenteritis Virus (TGEV). Human coronaviruses belonging to group 2 are OC43 and HKU1, the latter also having been discovered very recently.[7] The SARS coronavirus has been classified as an outlier of group 2.[8] Coronaviruses infecting birds have been identified as a separate group (group 3).[9]

The genome of HCoV 229E consists of 27,277 nucleotides, comprising a total of eight open reading frames. The entire replicase complex of the virus is encoded within two large overlapping open reading frames, ORF 1a and ORF 1b. ORF 1a codes for polyprotein 1a (pp1a) with a calculated molecular mass of 454 kDa. Involving a (-1) ribosomal frameshift, translation of ORF 1a and ORF 1b together yields the giant polypeptide 1ab (pp1ab) with a calculated molecular mass of 754 kDa.[10,11] These polyproteins are processed by two virus-encoded papain-like proteases ($PL1^{pro}$ and $PL2^{pro}$)[12] and the main proteinase (M^{pro}, also called 3C-like protease, $3CL^{pro}$),[13] resulting

in 16 non-structural proteins (Nsps). The crystal structure of HCoV-229E M^{pro} has been determined by our group[14] and shown to be similar to that of the homologous enzyme from TGEV.[15] The structure of the SARS-CoV M^{pro} [16,17] is also very similar.

The C-terminal region of pp1a comprises a set of relatively small polypeptide domains, Nsp6 to Nsp11. In preliminary experiments, we have shown that Nsp10 from mouse hepatitis (corona)virus (MHV) is a double-stranded RNA-binding zinc-finger protein,[18] and that HCoV-229E Nsp8 and Nsp9 interact with nucleic acids.[19] Also, it has been proposed recently that Nsp9 might specifically interact with the stem-loop II motif (s2m), a well-defined RNA secondary-structure element at the 3' end of many coronavirus genomes.[20] However, the s2m does not seem to be conserved in HCoV 229E (http://athena.bioc.uvic.ca). Nsp8 of SARS-CoV has the function of an RNA primase;[21] 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 dsRNA binding.[22] By analytical ultracentrifugation, it has been shown that Nsp8 also interacts with Nsp9,[23] although according to our own measurements using surface plasmon resonance, this interaction is very weak, if at all existent (R. Ponnusamy, unpublished data). In MHV, colocalization of Nsp7, Nsp8, Nsp9, and Nsp10 was observed.[24] Very likely, these non-structural proteins are directly involved in the replication complex built around the RNA-dependent RNA polymerase (Nsp12).

Here we describe the crystal structures of wild-type HCoV-229E Nsp9 and its Cys69Ala mutant at 1.75 Å and 1.80 Å resolution, respectively. In spite of 45% sequence identity

between SARS-CoV and HCoV-229E Nsp9, the wild-type structure of the latter exhibits a mode of homodimerization that is entirely different from what has been observed in the crystal structure of the former.[23,25] To probe the effect of the observed intermolecular disulfide bridge on the formation of this dimer, Cys69 was mutated to alanine. The crystal structure of this Nsp9 mutant shows a dimerization mode similar to the one observed in SARS-CoV Nsp9.[23,25] However, gel-shift assays and surface plasmon resonance measurements indicate that only the wild-type HCoV-229E Nsp9, not the Cys69Ala mutant, binds strongly to single-stranded RNA and single-stranded DNA. In order to assess a possible direct role of Cys69 in nucleic-acid binding, this residue was also replaced by serine. Again, the mutant showed little or no affinity to ssDNA. Finally, the corresponding residue (Cys73) of SARS-CoV Nsp9, which did not form a disulfide bond even after several weeks of storage, was replaced by alanine and serine. Both mutants showed wild-type affinity to single-stranded oligonucleotides. It is therefore concluded that Nsp9 of HCoV 229E is substantially different from its orthologue in SARS-CoV.

Results

Structure elucidation and quality of the structural models

Wild-type HCoV-229E Nsp9 and its Cys69Ala and Cys69Ser mutants were cloned with a $(His)_6$ tag connected to the N-terminus of the protein *via* the linker sequence VKLQ. The latter tetrapeptide corresponds to the C-terminus of SARS-coronavirus 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. The wild-type Nsp9 was crystallized using a reservoir containing 1.8-2.1 M ammonium sulfate, 0.1 M sodium acetate pH 4.0-4.5, and 5% 2-methyl-2,4-pentanediol (MPD). Crystals were of space group P622, with a monomer in the asymmetric unit (Table 1). The structure was determined by molecular replacement, using a monomer of the SARS-CoV Nsp9 [25] as the search model, and refined to 1.75 Å resolution. Residues 1-7 and 33-36 could not be modelled 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 is 19.0% and the R_{free} is 22.4%; 97.2% of the amino-acid residues are in the most-favored regions of the Ramachandran plot and the remainder in the additionally allowed regions.[26]

The Cys69Ala mutant of HCoV-229E Nsp9 was prepared by single-site PCR mutagenesis from the wild-type plasmid. Preparation of the protein was identical to wild-type Nsp9. 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 ammonium sulfate, 0.2 M sodium acetate (pH 4.6), 30% polyethylene glycol monomethyl ether (PEG-MME) 2000. The crystals displayed space group $P2_12_12_1$, with a dimer of the Nsp9 mutant per asymmetric unit (Table 1). Residues 1-2 and 109 of monomer A have not been modelled 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 Supplementary Fig. S1).

92.5% 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.[26]

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 is related to the oligonucleotide/oligosaccharide-binding modules (OB fold). This fold is characteristic of proteins binding to single-stranded nucleic acids[27] and occurs, for example, in single-stranded DNA-binding proteins from bacteria[28] to man[29] as well as in viruses.[30] 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.[27] 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. 1), are very flexible.

In the electron density map for wild-type HCoV-229E Nsp9, we unequivocally located an MPD molecule that 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.[31] 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 against the helix near Val106 (Supplementary Fig. S2). This observation nicely explains why 5% MPD was an essential additive in the crystallization of HCoV-229E wild-type 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. 1). 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).

In contrast to the wild-type Nsp9, MPD was not a useful additive in crystallization experiments with the mutant protein. However, dithiothreitol (DTT) was absolutely essential. Again, we located extra electron density between Asn27 and the α-helix (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.

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 [25] (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 [23] (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.

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. 2a). 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. 2b). 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 Supplementary Fig. S3 for a sequence alignment) 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. 2b). 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. 2a) 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 Supplementary Fig. S4). Gly96 and Gly100 are completely conserved among the coronaviruses and Gly103 is present in group 1 coronaviruses (Supplementary Fig. S3). The surface area per Nsp9 monomer buried through dimer formation[32] is 985 $Å^2$.

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 Supplementary Fig. S3), the mode of dimerization in the latter is very different from what we observe in our structure. 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.[23,25] In contrast to HCoV-229E Nsp9, which we had prepared with authentic N- and C-termini, the SARS-CoV protein used by Sutton *et al.*[23] for crystallization carried 30 additional amino-acid residues at the N-terminus, due to the cloning procedure. From the structure[23], 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.* structure. However, the structure published by Egloff *et al.*[25] for SARS-CoV Nsp9 was derived from a protein that carried only six additional histidine residues at its N-terminus (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.*[23] and completely different from the one observed by us for HCoV-229E Nsp9.

Disulfide bonds are rare in proteins that exist in the cytosol, where the environment is of reducing character, and we therefore have to discuss the possibility that the dimerization mode seen in our structure of HCoV-229E Nsp9 is an artifact of disulfide-bond formation, in spite of the presence of 5 mM DTT 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, we replaced Cys69 in HCoV-229E Nsp9 by alanine. The structure of the mutated protein (Fig. 2c) 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^{o}. However, this dimerization mode is identical to that of SARS-CoV Nsp9 (cf. Fig. 2c and Supplementary Fig. S5). The dimer of the HCoV-229E Cys69Ala mutant can be

superimposed onto the dimer of the SARS-CoV Nsp9 protein (1QZ8;[25]) with an r.m.s. deviation of 0.99 Å for 175 C^{α} atom pairs (see Supplementary Fig. S5). 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.*[23] (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. 2d). 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^{o} and a closest approach of 3.96 Å (C^{α}-C^{α}) between Ala97 and Gly100 (Fig. 2d). 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[32] 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.

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). There are multiple interactions between monomers across the hexamer, which we will briefly discuss 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, we find 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). Asp19 is present in most coronavirus Nsp9 sequences (not in SARS-CoV and IBV), whereas Lys50 is highly conserved across the family (see Supplementary Fig. 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).

Additional intermolecular interactions within the Nsp9 hexamer are listed in Supplementary Table S1. Through the 6-fold axis of the crystal, hexamers are arranged into 36-mers (Supplementary Fig. S6).

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 (termed M2; Supplementary Table S1) 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 with interface M1, M2 leads to the polymerization of the protein (see Discussion, Fig. 7c). Yet another protein-protein interface (M3) with quasi-twofold symmetry, but not involved in polymer formation, is mentioned in Supplementary Table S1.

Oligomeric state in solution

In order to determine the oligomeric state of wild-type and mutant HCoV-229E Nsp9 in solution, we applied a number of biophysical and biochemical techniques. For both wild-type HCoV-229E Nsp9 (fresh preparation) and the Cys69Ala mutant, Dynamic Light-Scattering (DLS) revealed a monodisperse peak centered at a hydrodynamic radius of 28 (± 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. 4). 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).

Oxidation state of Cys69 in solution

By titration of free sulfhydryl groups with Ellman's reagent,[33] we found that 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 we investigated this phenomenon further, we observed that 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^{o}C for 5 min).

In contrast to HCoV-229E Nsp9, we could show by using Ellman's reagent that SARS-CoV Nsp9, which has three cysteine residues and which we had prepared 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.

Binding of nucleic acids

Gelshift assay

Using a gelshift assay (a modified version[18] of zone-interference gel electrophoresis, ZIGE[34]; Fig. 5), we found that wild-type HCoV-229E Nsp9 bound to single-stranded oligodeoxynucleotides (6-mers to 50-mers; Fig. 5a, lanes 3 to 10) and, to a very limited extent (if at all), to a double-stranded oligodeoxynucleotide (24-mer; Fig. 5a, lane 2).

Nsp9 of SARS-CoV also bound to both single-stranded and, again very weakly, to double-stranded oligonucleotides (Fig. 5b). However, the effect on the gel shift did not increase smoothly with oligonucleotide length; rather, there was a step-wise increase from the 13-mer (Fig. 5b, 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. 5c, 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, we also replaced the cysteine by serine. The Cys69Ser mutant did not bind short oligonucleotides either, but did show some gel shift with the 55-mer (Fig. 5c, lane "HSM 3"). Next, we replaced the corresponding cysteine in SARS-CoV Nsp9 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. 5c, 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).

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. We observed a signal for Nsp9 interaction with the

oligonucleotide when protein concentrations were in the μM range (Supplementary Fig. S7). 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. In fact, the SARS-CoV Nsp9 binding profile is better explained by a two-state binding model (χ^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, we could not use HCoV-229E Nsp9 concentrations higher than 35 μM (SARS-CoV Nsp9: 85 μM), because non-specific binding appeared to govern the profile above this value and saturation was not reached. We could also not derive K_D values 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.

We also compared 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 in the surface plasmon resonance experiment. For this, the oligonucleotide was again immobilized onto 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. 6).

Discussion

In this study, we observed different dimerization modes for HCoV-229E Nsp9 by X-ray crystallography. The wild-type protein exhibits a homodimer that is very different from the one previously seen for SARS-CoV Nsp9,[23,25] 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 our preparation of HCoV-229E Nsp9 and that of SARS-CoV Nsp9 by both Egloff *et al.*[25] and Sutton *et al.*[23] is that we have worked with a protein with 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; Supplementary Fig. S5) 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.[35,36] Therefore, we have to take the possibility into account that formation of the disulfide is an artifact of the conditions of protein preparation. 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.

We tried to assess the relevance of the two different dimers seen in our structures by calculating the surface area buried upon dimerization as well as the shape complementarity[37] (see Supplementary Table S1). 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). For comparison, this value is 0.67 and 0.70, respectively, for the two crystal structures of SARS-CoV Nsp9,[23,25] which show the same dimerization mode as the Cys69Ala mutant of HCoV-229E 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).

Both wild-type and mutant dimer 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 (Supplementray Fig. S4).[38] The GXXXG motif is actually a common feature of the association of transmembrane helices.[39,40] 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 *antiparallel* 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 fact that the amino-acid sequence of the C-terminal helix of Nsp9 allows stabilization of both forms of the dimer may support the idea that both forms are indeed biologically relevant (see below).

In addition to the dimerization modes that we identified in the 'parent' dimer of the wild-type protein and the Cys69Ala mutant, we have to also consider a number of additional protein-protein interfaces that are seen in the crystal structures (see Supplementary Table S1). 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 (Supplementary Fig. S6).

In case of the Cys69Ala mutant, there are two other monomer-monomer interfaces in the crystal, in addition to the 'parent mode' (M1). M3 is formed by the β6-β7 hairpin and

involves mainly hydrophobic interactions between side chains that are not conserved (Supplementary Table S1), although we note that a similar interface exists in the Egloff *et al.*[25] structure of SARS-CoV Nsp9. The other interface, M2, arises through some limited interaction between strands β5 of neighboring molecules in the crystal. M2 comprises four hydrogen bonds between 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.[28] This same alternative dimerization mode has also been discussed by Sutton *et al.*[23] 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.*,[25] even though the space group of these crystals is different. In summary, then, both dimerization modes M1 and M2 occur in all crystal structures of Nsp9 described so far (except that of the HCoV-229E wild-type protein), even in a second crystal form briefly mentioned by Sutton *et al.*[23] (for which no data have been deposited in the PDB), *i.e.* in a total of four different crystalline environments. We therefore propose that Nsp9 oligomerization, in particular in the presence of single-stranded RNA, is mediated through these two protein-protein interaction surfaces. The existence of such oligomers also in solution is supported by our glutaraldehyde crosslinking experiments, which revealed the presence of monomers, dimers, trimers, and higher oligomers for wild-type SARS-CoV Nsp9 and HCoV-229E Cys69Ala Nsp9 in the SDS-PAGE under reducing conditions (Fig. 4). Interestingly, for the wild-type 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), whereas the Cys69Ala mutant as well as SARS-CoV Nsp9 can form polymers (Fig. 7).

We used the sulfate ions present in one of the SARS-CoV Nsp9 structures[25] (PDB ID 1QZ8) to propose a binding mode for ssRNA (Fig. 7a). 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; we use the HCoV-229E numbering scheme here, 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-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. We also note that the residues that we propose 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 our 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.[41] In our model, approximately 40 nucleotides can be bound per Nsp9 dimer within the extended polymer that we propose for SARS-CoV Nsp9 (Fig. 7a).

Nsp9 interaction with nucleic acids was visualized using a slightly modified version of Zone Interference Gel Electrophoresis[34] (ZIGE; 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 (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. 5a and b). Although the dimerization modes of HCoV-229E 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 with 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, we prepared an additional three mutants, 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. 5c). 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.

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 we had observed a weak shift 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 are 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 we could determine using Ellman’s reagent (not shown). If one of these cysteines were 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.[42] (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, we observed that a 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^{o}C is 10 h and 40 h at pH 7.5 and 6.5, respectively.[43] We determined the concentration of DTT required to fully reduce the Nsp9 disulfide 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, we believe that 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.[44] Also, many transcription factors including Fos, Jun, NF-κB, PaX, FNR, OxyR (see Bauer *et al.*[45] 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.[36]

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.[46] Similarly, rhinovirus[47] and baculovirus[48] 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.[49] It has been hypothesized that this cysteine could form an intermolecular disulfide bridge and thus results in LEF3 oligomers in solutions. This could allow more DNA to bind in closer proximity, thus favoring the annealing effect.[49] 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.[50-53]

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.[54-57] These double-membrane vesicles are around 200-350 nm in diameter and present alone or as clusters in the cytosol.[55] 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 our 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.*,[58] 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[59], even though its Nsp9 does not seem to form disulfide-linked dimers, at least not *in vitro*. The replicase of this virus may have other mechanisms to deal with an oxidative environment. We note that the number of cysteine residues is above average in coronavirus replicase proteins; in SARS-CoV, their share is 3.9% (HCoV 229E: 3.3%), whereas only 1.25% of residues in human cytosolic proteins are cysteines[60]. In the SARS-CoV main proteinase (M^{pro}) alone, there are 12 cysteine residues (3.9% of all residues), whereas this number is 8 (2.6%) in HCoV-229E M^{pro}. Some of these could perhaps scavenge oxygen radicals, thereby undergoing oxidation to sulfenic, sulfinic, or even sulfonic acid.

Materials and methods

Cloning, expression and purification

The regions coding for Nsp9 of HCoV 229E and SARS-CoV were amplified by polymerase chain reaction from virus-derived cDNA fragments. The HCoV-229E nsp9 PCR product was cloned into the pET15b 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 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. 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 wild-type Nsp9 of HCoV 229E and of SARS-CoV and the mutant proteins were recombinantly produced and purified in a similar manner. Nsp9-encoding plasmids were transformed in the competent *E. coli* Tuner (DE3) pLacI strain (Merck KGaA, Darmstadt, Germany). Cultures were grown in TY medium at 37°C until cells reached an optical density of 0.4 at 660 nm. Cells were then induced with 1 mM IPTG and grown for a further 4 h 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. For lysis, the cell pellets were resuspended in 20 mM Tris-HCl, 300 mM NaCl, 5 mM DTT, and 20 mM imidazole, pH 7.5 (25°C). Cells were broken by French press after adding 10% (v/v) glycerol and 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.[61] 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 20 mM Tris-HCl, 300 mM NaCl, 5 mM DTT, and 20 mM imidazole, pH 7.5 (25°C), the protein was eluted with a linear gradient ranging from 20 mM up to 500 mM imidazole. For SARS-CoV Nsp9 purification alone, the buffers were adjusted to pH 8.0 (25°C) and 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 anti-tetrahistidine antibody (Dianova, Hamburg, Germany) and anti-mouse IgG-alkaline phosphatase conjugate (Sigma, Munich, Germany).

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.

To determine the minimum concentration of DTT required to reduce the disulfide bond, 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 Nsp9 preparation. After incubation for 30 min at room temperature, 2% loading bufffer was added and SDS-PAGE was started.

Crystallization, structure determination, and refinement

Crystallization screening was performed for both the HCoV-229E Nsp9 wild-type protein and the Cys69Ala mutant 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 protein and 1 µl of reservoir solution. Both wild-type and mutant protein crystals appeared within two days at 10˚C and a protein concentration of 6 - 8 mg/ml, in a buffer containing 20 mM Tris, 200 mM NaCl, 1 mM DTT, pH 7.5. 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 (see Results for optimized conditions). 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. The cryoprotectant was 25% glycerol for the wild-type protein and 25% polyethylene glycol 400 for the mutant. Data were processed with DENZO[62] and scaled by using SCALA.[63] The structures were solved by using the molecular replacement program Phaser[64], 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). Models were built into electron density by using COOT[65] and refined by REFMAC5.[66] The final model for wild-type HCoV-229E Nsp9 includes 70 water molecules, two sulfate-ion sites, and one MPD molecule. The structural model for the mutant comprises 90 water molecules and one DTT molecule. The final steps of 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 overall geometric quality of the models was checked using PROCHECK.[26] The surface area buried upon dimerization was calculated using AREAIMOL.[32] Structure superimposition and calculation of r.m.s. deviations were carried out using ALIGN.[67] Figures were created using Pymol.[68]

Dynamic light scattering

Measurements were taken using a Spectroscatter 201 (RiNA GmbH, Berlin, Germany) using 10 µl solution of Nsp9 in 20 mM Tris-HCl, 200 mM NaCl, and 1 mM DTT, pH 7.0 (25^{o}C). Unless otherwise mentioned, all the measurements were done 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.

Electrophoretic mobility shift assay

Zone Interference Gel Electrophoresis (ZIGE) is a method to analyze weak protein/nucleic-acid interactions.[18,34] Because of the high isoelectric point of Nsp9, this method was slightly modified. HCoV-229E Nsp9 and SARS-CoV Nsp9 have a theoretical pI of 9.3 and 9.1, respectively; 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. Therefore, only the oligonucleotides (concentration of 100 µM) were loaded into the sample slots and run for 40 min at 100 mA and 4˚C. Subsequently, 50 µM of protein was loaded into the

respective lanes and the run was continued for another 40 min, with the poles of the electrodes interchanged. 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. 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. 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 a 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 stowed in 10% (v/v) acetic acid.

Crosslinking

Chemical cross-linking experiments were performed for wild-type HCoV-229E Nsp9, the Cys69Ala mutant, 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% (w/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 h 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.

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 was purchased from MWG-Biotech AG, Ebersberg, Germany. For immobilization, 2 nM of the oligonucleotide was injected in 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, 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:

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

Two-state binding model:

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

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.

Protein Data Bank accession codes

Atomic coordinates of wild-type HCoV-229E Nsp9 and the Cys69Ala mutant have been deposited with the RCSB Protein Data Bank, accession code 2J97 and 2J98, respectively.

Acknowledgments

This work was, in part, supported by the Sino-European Project on SARS Diagnostics and Antivirals (SEPSDA, contract no. SP22-CT-2004-003831; www.sepsda.eu) and by

VIZIER (contract no. LSHG-CT-2004-511960; www.vizier-europe.org), both funded by the European Commission. We also acknowledge support from the Sino-German Center for the Promotion of Research, Beijing, the Schleswig-Holstein Innovation Fund, and the DFG Cluster of Excellence "Inflammation at Interfaces". We thank Prof. J. Ziebuhr (Queen´s University of Belfast) for providing a cDNA for HCoV-229E Nsp7-10 polyprotein, Dr. K.H.G. Verschueren for help with data collection, and Silke Schmidtke for expert technical assistance. R.H. thanks the Fonds der Chemischen Industrie for continuous support.

References

1. Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., *et al.* (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967-1976.
2. Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., *et al.* (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953-1966.
3. Kuiken, T., Fouchier, R. A., Schutten, M., Rimmelzwaan, G. F., van Amerongen, G., van Riel, D., *et al.* (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet,* **362**, 263-270.
4. Peiris, J. S., Chu, C. M., Cheng, V. C., Chan, K. S., Hung, I. F., Poon, L. L., *et al.* (2003). Clinical progression and viral load in a community outbreak of

coronavirus-associated SARS pneumonia: a prospective study. *Lancet*, **361**, 1767-1772.

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

6. van der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J., Wolthers, K. C., *et al.* (2004). Identification of a new human coronavirus. *Nature Med.* **10**, 368-373.

7. Woo, P. C., Lau, S. K., Chu, C. M., Chan, K. H., Tsoi, H. W., Huang, Y., *et al.* (2005). Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* **79**, 884-895.

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

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

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

11. 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.

12. 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.

13. 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.

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

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

16. Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., *et al.* (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.

17. Tan, J., Verschueren, K. H., Anand, K., Shen, J., Yang, M., Xu, Y., *et al.* (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.

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

19. 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.

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

21. Imbert, I., Guillemot, J. C., Bourhis, J. M., Bussetta, C., Coutard, B., Egloff, M. P., *et al.* (2006). A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J.* **25**, 4933-4942.

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

23. Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., *et al.* (2004). The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure*, **12**, 341-353.

24. 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.

25. Egloff, M. P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., *et al.* (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.

26. 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. Crystallog.* **26**, 283-291.

27. 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.

28. 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.

29. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature,* **385**, 176-181.

30. 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.

31. Anand, K., Pal, D. & Hilgenfeld, R. (2002). An overview on 2-methyl-2,4-pentanediol in crystallization and in crystals of biological macromolecules. *Acta Crystallog. Sect. D,* **58**, 1722-1728.

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

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

34. 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.

35. 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.

36. 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.

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

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

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

40. 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.

41. Chen, C. Y., Chang, C. K., Chang, Y. W., Sue, S. C., Bai, H. I., Riang, L., *et al.* (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.

42. 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.

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

44. 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.

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

46. 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.

47. 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.

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

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

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

57. van Hemert, M. J., van den Worm, S. H. E., Knoops, K., Mommaas, A. M., Gorbalenya, A. E. & Snijder, E. J. (2008). *PLoS Pathogens*, **4**, e1000054.

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

59. Imai, Y., Kuba, K., Neely, G. G., Yaghubian-Malhami, R., Perkmann, T., van Loo, G., et al. (2008). Identification of oxidative stress and Toll-like receptor 4 signaling as a key pathway of acute lung injury. *Cell*, **133**, 235-249.

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

61. 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.

62. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

63. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programmes for protein crystallography. *Acta Crystallog. Sect. D,* **50**, 760-763.

64. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallog. Sect. D,* **61**, 458-464.

65. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. Sect. D,* **60**, 2126-2132.

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

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

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

Legends to Figures

Figure 1. Superimposition of monomers. Ribbon representation of HCoV-229E Nsp9 wild-type (green) and Cys69Ala mutant (red) monomers, superimposed with a C^{α} r.m.s. deviation of 0.71 Å. Loop L23 of wild-type HCoV-229E Nsp9 could not be built due to lack of electron density.

Figure 2. Structural features of the homodimers of wild-type HCoV-229E Nsp9 and the Cys69Ala mutant. The two monomers are colored in red and green respectively; the disulfide is shown in yellow where present. N and C denote the amino and carboxy termini, respectively, of the polypeptide chains. (a) Ribbon representation of the disulfide-linked wild-type HCoV-229E Nsp9 dimer. (b) Residues involved in the dimer interface of wild-type Nsp9 (sticks; red, oxygen; blue, nitrogen; yellow, carbon). Intermolecular hydrogen bonds are indicated by dashed lines. (c) HCoV-229E Cys69Ala mutant dimer. (d) Residues involved in the dimer interface of the mutant Nsp9. Color code is the same as in (b). The closest approach between the C-terminal α-helices, between the C^{α} atoms of Gly A100 and Ala B97, is indicated by a dotted line.

Figure 3. 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 3-fold 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 yellow, cyan and magenta. The two sulfate ions on the

3-fold axis are indicated in the same colors. Each sulfate is three-fold disordered. The 2-fold axes are running between the monomers.

Figure 4. Nsp9 crosslinking using glutaraldehyde. Crosslinking was carried out with different protein concentrations (10 μM to 100 μM) using 0.01% glutaraldehyde. The molecular mass of the cross-linking products is indicated. Wild-type SARS-CoV Nsp9 and the HCoV-229E Nsp9 Cys69Ala mutant form higher oligomers at a protein concentration of 100 μM, presumably involving similar interactions as seen in the crystal structure. In contrast, wild-type HCoV-229E Nsp9 does not form oligomers higher than trimers.

Figure 5. Gel mobility shift assay (zone-interference gel electrophoresis, 1% agarose; see Materials & 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-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 (grey) in lanes 3 for HAM and HSM correspond to precipitated, unbound 55-mer oligonucleotide (not stained by Coomassie Brilliant Blue; see Materials & Methods).

Figure 6. 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 to 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.

Figure 7. Oligomers of SARS-CoV Nsp9 (PDB codes 1QZ8, 1UW7, (**a**), (**b**)) and the HCoV-229E Nsp9 Cys69Ala mutant (**c**). Independent of space group symmetry (1QZ8 - $P6_122$, 1UW7 - $P4_322$, HCoV-229E Nsp9 Cys69Ala mutant - $P2_12_12_1$), two common dimer interfaces are present in these crystal structures of Nsp9. One interface is formed mainly by the C-terminal α-helix (red), and the other by strand β5 (blue). We propose that the ssRNA (black) could wrap around the Nsp9 polymer by forming a left-handed helix (**a**), with approximately 40 nucleotides bound per Nsp9 dimer.

Table 1: Data collection and refinement statistics

Data collection	Nsp9 wild-type	Nsp9 mutant
Wavelength (Å)	0.8075	0.8075
Resolution (Å)	40.0-1.75 (1.79-1.75)	30.0-1.80 (1.86-1.80)
Space group	P622	$P2_12_12_1$
Unit-cell parameters		
a (Å)	85.63	26.40
b (Å)	85.63	61.38
c (Å)	48.69	107.31
Solvent content	42.3%	31.5%
Overall reflections	129,656	139,726
Unique reflections	11,317 (730)	16,842 (1648)
Multiplicity	11.5 (11.5)	8.3 (4.7)
Completeness (%)	99.9 (100.0)	99.4 (99.7)
R_{merge}[1] (%)	8.3 (60.3)	8.9 (35.1)
I/σ(I)	13.7 (4.45)	19.7 (3.96)
Refinement		
Resolution (Å)	40.0-1.75	30.0-1.80
R_{cryst}[2]	0.190	0.221
R_{free}[2]	0.224	0.281
R.m.s.d. from ideal geometry		
bonds (Å)	0.013	0.017
angles (Å)	1.417	1.962
Protein atoms	778	1604
Solvent atoms	65	74
MPD	1	-
DTT	-	1
Sulfate	2	-
Ramachandran plot		
most favoured (%)	97.2	92.5
additionally allowed (%)	2.8	7.5
generously allowed (%)	0	0
disallowed regions (%)	0	0

Values in parentheses are for the highest resolution shell.

[1] $R_{\text{merge}} = \sum_{hkl}\sum_i | I(hkl)_i - \langle I(hkl)\rangle | / \sum_{hkl}\sum_i I(hkl)_i$, where *I(hkl)* is the intensity of reflection hkl and ⟨*I(hkl)*⟩ 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%) omitted from the refinement.