

# Structural proteomics of the SARS coronavirus: a model response to emerging infectious diseases

Mark Bartlam · Yuanyuan Xu · Zihe Rao

Received: 19 March 2007 / Accepted: 23 July 2007 / Published online: 7 August 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** A number of structural genomics/proteomics initiatives are focused on bacterial or viral pathogens. In this article, we will review the progress of structural proteomics initiatives targeting the SARS coronavirus (SARS-CoV), the etiological agent of the 2003 worldwide epidemic that culminated in approximately 8,000 cases and 800 deaths. The SARS-CoV genome encodes 28 proteins in three distinct classes, many of them with unknown function and sharing low similarity to other proteins. The structures of 16 SARS-CoV proteins or functional domains have been determined to date. Remarkably, eight of these 16 proteins or functional domains have novel folds, indicating the uniqueness of the coronavirus proteins. The results of SARS-CoV structural proteomics initiatives will have several profound biological impacts, including elucidation of the structure–function relationships of coronavirus proteins; identification of targets for the design of anti-viral compounds against SARS-CoV and other coronaviruses; and addition of new protein folds to the fold space, with further understanding of the structure–function relationships for several new protein families. We discuss the use of structural proteomics in response to emerging

infectious diseases such as SARS-CoV and to increase preparedness against future emerging coronaviruses.

**Keywords** Severe acute respiratory syndrome · Coronavirus · Structural proteomics · Biological impact · Fold discovery · Structure based drug discovery

## Introduction

One of the central aims of Structural Genomics is to determine the structures of proteins with biomedical importance, in order to understand the molecular basis of these diseases via the proteins involved, and thus to improve disease treatment, diagnosis or prevention. A number of Structural Genomics initiatives worldwide are focused on the structures of proteins related to human disease, including various bacterial, protozoan and viral pathogens. These include the TB Structural Genomics Consortium (<http://www.doe-mtb.ucla.edu/TB/>), involving 50 laboratories across 9 countries and aiming to determine 400 structures from *Mycobacterium tuberculosis*. The Structural Genomics of Pathogenic Protozoa initiative (<http://www.sgpp.org/>) is targeting the protozoan species that cause tropical diseases such as malaria, sleeping sickness, leishmaniasis and Chagas' disease. In Europe, the Structural Proteomics IN Europe (SPINE) (<http://www.spineurope.org/>) programme focuses on both bacterial and viral pathogens: the former include *Bacillus anthracis* and *Mycobacterium tuberculosis*, while the latter include poxviruses, herpesviruses and coronaviruses. Also in the area of viral pathogens, the focus of the VIZIER project (<http://www.vizier-europe.org/>) is comparative structural genomics of viral enzymes involved in replication. The specific aim of VIZIER is to identify potential new

---

M. Bartlam · Z. Rao  
College of Life Sciences, Nankai University, Tianjin 300071, China

M. Bartlam · Y. Xu · Z. Rao (✉)  
Laboratory of Structural Biology, Life Sciences Building,  
Tsinghua University, Beijing 100084, China  
e-mail: raozh@xtal.tsinghua.edu.cn

M. Bartlam · Y. Xu · Z. Rao  
National Laboratory of Biomacromolecules, Institute of  
Biophysics, Chinese Academy of Sciences, Beijing 100101, China

anti-viral targets against RNA viruses through targeting their replication machinery. However, VIZIER does not include the SARS virus as part of its sphere of activity.

In 2003, the emergence of a form of pneumonia called severe acute respiratory syndrome (SARS) was attributed to a previously unknown coronavirus termed SARS-CoV [1, 2, 3, 4]. SARS-CoV was the aetiological agent for a worldwide epidemic with approximately 8,000 reported cases and 800 deaths, and its emergence was attributed to an animal-to-human interspecies transmission [5]. Coronaviruses, characterized as enveloped, positive-stranded RNA viruses with the largest known genomes, belong to the genus *Coronavirus* of the family *Coronaviridae* [6, 7]. Approximately 26 species of coronaviruses (CoVs) can be classified into three distinct groups on the basis of genome sequence and serological reaction [8]. Prior to the outbreak, very little attention was paid to the structure–function studies of coronavirus proteins by researchers as this genus of virus predominantly causes severe diseases in animals and comparatively mild diseases in humans. While extensive research had been carried out on model coronaviruses over the previous 20 years or so, little was understood about underlying mechanisms such as viral assembly and viral replication/transcription prior to the SARS outbreak.

The SARS-CoV genome is approximately 29,700 nucleotides and is composed of at least 14 functional open reading frames (ORFs) that encode 28 proteins covering three classes: two large polyproteins (pp)1a and (pp)1ab that are cleaved into 16 non-structural proteins required for viral RNA synthesis (and probably with other functions); four structural proteins (the S, E, M and N-proteins) essential for viral assembly; and eight accessory proteins that are thought unimportant in tissue culture but may provide a selective advantage in the infected host (Table 1, Fig. 1) [9]. Many of the 28 SARS-CoV proteins share low sequence similarity with other proteins, including those from other viruses, indicating their uniqueness and hampering functional assignment based on homology.

In this review, we will focus on the current progress in SARS coronavirus (SARS-CoV) structural proteomics initiatives and assess their biological impact. In addition to several traditional structural biologists, there are currently three major international structural proteomics initiatives focused on SARS-CoV: in China (our group, led by Zihé Rao), USA (The Scripps Research Institute, led by Peter Kuhn) and France (University of Marseilles, led by Bruno Canard). Other SARS-CoV protein structures have been solved by the SPINE consortium led by David Stuart. The strategies adopted by the three groups are similar: to systematically determine the three-dimensional structure of each protein encoded by the SARS coronavirus in order to elucidate their function and

identify potential new therapeutic targets. Drug development strategies targeting SARS-CoV are focused on two main avenues: inhibitors to block virus entry into the host cells, and compounds to prevent viral replication and transcription. The three structural proteomics initiatives have focused more specifically on the replication/transcription machinery formed by the 16 non-structural proteins.

### Non-structural proteins

The SARS-CoV replicase gene encodes 16 non-structural proteins (nsps) with multiple enzymatic functions [10]. These are known or are predicted to include types of enzymes that are common components of the replication machinery of plus-strand RNA viruses: an RNA-dependent RNA polymerase activity (RdRp, nsp12), a 3C-like serine protease activity ( $M^{\text{pro}}$  or  $3CL^{\text{pro}}$ , nsp5), a papain-like protease activity ( $PL2^{\text{pro}}$ , nsp3), and a superfamily 1-like helicase activity (HEL1, nsp13). In addition, the replicase gene encodes proteins that are indicative of 3′-5′ exonuclease activity (ExoN homolog, nsp14), endoribonuclease activity (XendoU homolog, nsp15), adenosine diphosphate-ribose 1′-phosphatase activity (ADRP, nsp3), and ribose 2′-O-methyltransferase activity (2′-O-MT, nsp16) [10]. These enzymes are less common in positive-strand RNA viruses and may therefore be related to the unique properties of coronavirus replication and transcription. Finally, the replicase gene encodes another nine proteins, of which little is known about their structure or function. Here we detail the available structures of non-structural proteins, of which nsp5 is the most widely characterized.

### Nsp1

The non-structural protein nsp1 is the N-terminal cleavage product of the viral replicase polyprotein that mediates RNA replication and processing. Nsp1 lacks any viral or cellular homologs other than in coronaviruses and its precise function remains unknown, although it has been shown to specifically accelerate mRNA degradation with a reduction in cellular protein synthesis. An NMR structure of nsp1 covering residues 13–128 was determined by Kurt Wuthrich and colleagues as part of the US structural proteomics initiative [11] and presents a novel irregular  $\beta$ -barrel fold, indicating an unidentified and possibly unique biological function (Fig. 1). The full-length nsp1 protein, also characterized by Wuthrich and colleagues, has two flexibly disordered polypeptide segments from residues 1–12 and 129–179.

**Table 1** Summary of SARS proteins

Protein	Protein size (a.a.)	ORF (location in genome sequence)	Putative functional domain(s)	Structure available
<i>Structural proteins</i>				
Spike (S) protein	1255	ORF2 (21492–25259)		Yes (fusion core, receptor binding domain)
Envelop (E) protein	76	ORF4 (26117–26347)		No
Membrane (M) protein	221	ORF5 (26398–27063)		No
Nucleocapsid (N) protein	422	ORF9a (28120–29388)		Yes (N-terminal RNA binding domain, C-terminal domain)
<i>Non-structural proteins (Nsp)</i>				
Nsp1	180	ORF1a (265–804)		Yes
Nsp2	638	ORF1a (805–2718)		No
Nsp3	1922	ORF1a (2719–8484)	Ac, X, PL <sup>pro</sup> , Y (TM1), ADRP	Yes (Glu-rich <sup>a</sup> , ADRP, PL <sup>pro</sup> domains)
Nsp4	500	ORF1a (8485–9984)	TM2	No
Nsp5	306	ORF1a (9985–10902)	M <sup>pro</sup>	Yes
Nsp6	290	ORF1a (10903–11772)	TM3	No
Nsp7	83	ORF1a (11773–12021)		Yes
Nsp8	198	ORF1a (12022–12615)		Yes
Nsp9	113	ORF1a (12616–12954)	ssRNA binding	Yes
Nsp10	139	ORF1a (12955–13371)	GFL	Yes
Nsp11	13	ORF1a (13372–13410)		No
Nsp12	932	ORF1b (13398–16166)	RdRp	No
Nsp13	601	ORF1b (16167–17969)	ZD, NTPase, HEL1	No
Nsp14	527	ORF1b (17970–19550)	Exonuclease (ExoN homolog)	No
Nsp15	346	ORF1b (19551–20588)	NTD, endoRNase (XendoU homolog)	Yes
Nsp16	298	ORF1b (20589–21482)	2'-O-MT	No
<i>Accessory proteins</i>				
Orf3a	274	ORF3a (25268–26092)		No
Orf3b	154	ORF3b (25689–26153)		No
Orf6	63	ORF6 (26913–27265)		No
Orf7a	122	ORF7a (27273–27641)	Ig-like	Yes (Luminal domain)
Orf7b	44	ORF7b (27638–27772)		No
Orf8a	39	ORF8a (27779–27898)		No
Orf8b	84	ORF8b (27864–28118)		No
Orf9b	98	ORF9b (28130–28426)		Yes

<sup>a</sup> Indicates that a structure has been deposited in the Protein Data Bank but has not been published

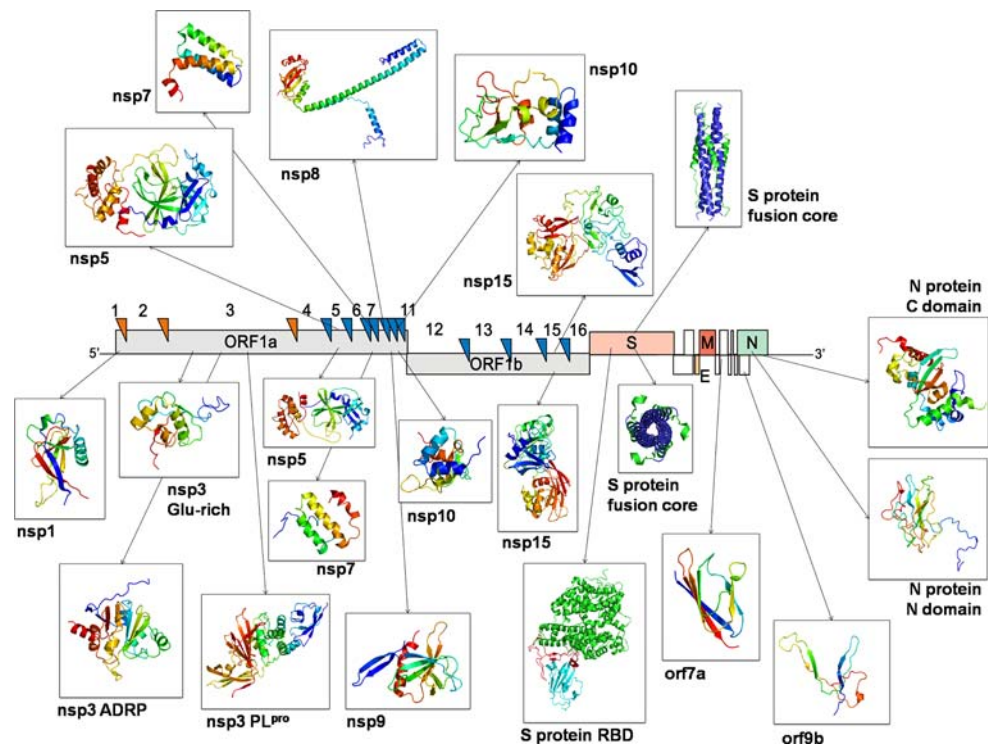
Abbreviations: PL<sup>pro</sup>, papain-like protease; ADRP, adenosine diphosphate-ribose 100-phosphatase; TM, transmembrane domain; M<sup>pro</sup>, main (or 3C-like cysteine) protease; GFL, growth factor-like domain; RdRp, RNA-dependent RNA polymerase; ZD, putative Zinc-binding domain; HEL1, superfamily 1 helicase; NTD, nidovirus conserved domain; ExoN, 30-to-50 exonuclease; 2'-O-MT, S-adenosylmethionine-dependent ribose 2'-O-methyltransferase

### Nsp3 ADRP and PL<sup>pro</sup> domains

One limitation of SARS structural proteomics is the difficulty in expressing soluble, stable and functional proteins. One workaround is to identify the functional domains of individual proteins to increase the chance of successful structure determination. Such an approach was taken in the case of nsp3, which is a large, multidomain

protein yielded by proteolytic cleavage of the ppla polyprotein at two sites by the papain-like protease (PL<sup>pro</sup>). It is comprised of 1,922 amino acids and features conserved sequence motifs for six domains: (1) an N-terminal Glu-rich acidic domain; (2) an 'X' domain with predicted Appr-100-p processing activity; (3) a SUD domain (SARS-specific unique domain); (4) a peptidase C-16 domain that contains the PL<sup>pro</sup>; (5) a transmembrane

**Fig. 1** Summary of SARS-CoV protein structures to date. The SARS-CoV genome is shown surrounded by the available structures of SARS-CoV proteins (drawn in ribbon representation): nsp1, nsp3 (Glu-rich, ADRP and PL<sup>pro</sup> domains), nsp5, nsp7, nsp8, nsp9, nsp10, nsp15, Spike protein (receptor binding domain and fusion core), N-protein (N-terminal RNA-binding domain and C-terminal dimerization domain), orf7a and orf9b. Orange and blue triangles represent PL<sup>pro</sup> (nsp3) and M<sup>pro</sup> (nsp5) cleavage sites, respectively. Structures shown above the genome (nsp5, nsp7, nsp8, nsp10, nsp15, S-protein fusion core) were solved by Zihao Rao and colleagues in China. Representative structures shown below the genome were solved by other groups. Structures are not drawn to scale



domain; and (6) the ‘Y’ domain. Peter Kuhn and colleagues in Scripps determined the crystal structures of two functional domains of nsp3: the ‘X’ or ADP-ribose-1<sup>st</sup>-phosphate dephosphorylation (ADRP) domain [12] and the papain-like protease (PL<sup>pro</sup>) domain [13]. A third NMR structure from the Scripps consortium is available in the Protein Data Bank for the N-terminal Glu-rich acidic domain. The French consortium of Bruno Canard and colleagues have also reported a structure–function study of the ADRP domain [14].

The structure of the ‘X’ domain, also known as the ADRP domain, reveals a close structural relationship with macro-H2A-like fold proteins (Fig. 1). Furthermore, the ‘X’ domain shares sequence homology with Poa1p from *Saccharomyces cerevisiae*, which is known to be a highly specific phosphatase that removes the 1<sup>st</sup> phosphate group of ADP-ribose-1<sup>st</sup>-phosphate (Appr-1<sup>st</sup>-p) in the tRNA splicing pathway. Using *in vitro* assays, the authors confirm that the nsp3 ‘X’ domain does indeed remove the 1<sup>st</sup> phosphate group of ADP-ribose-1<sup>st</sup>-phosphate (Appr-1<sup>st</sup>-p).

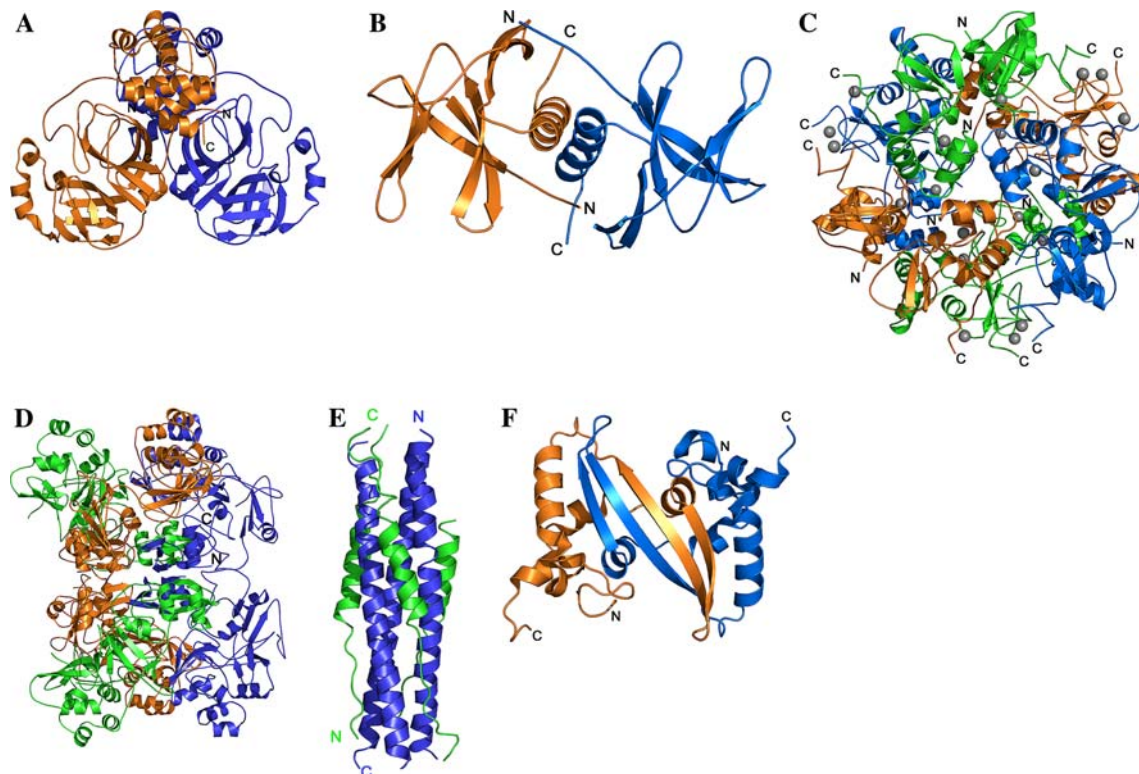
The structure of the PL<sup>pro</sup> domain of nsp3 was determined in 2006 and found to possess a “thumb-palm-fingers” fold related to known deubiquitinating enzymes (Fig. 1). However, certain key features of nsp3 PL<sup>pro</sup>, including a zinc-binding motif and a ubiquitin-like N-terminal domain, separate it from other characterized deubiquitinating enzymes. The availability of the nsp3 PL<sup>pro</sup> structure now provides a clearer understanding of the

proteolytic processing at the consensus (LXGG) cleavage site and provides details at the molecular level for the mechanism of deubiquitination, suggesting an important dual role for this enzyme.

At the time of writing, the structure of a third domain of nsp3, the Glu-rich acidic domain, has been deposited in the Protein Data Bank with accession number 2GRI yet remains unpublished. Determined by the Scripps group using NMR, the solution structure has a globular  $\alpha$ -helical fold (Fig. 1). A DALI search for structural similarity shows no significant structural homologs.

### Nsp5, the SARS-CoV main protease

The replicase polyproteins pp1a and pp1ab undergo extensive proteolytic processing by viral proteases to produce multiple functional subunits, which are involved in formation of the replicase complex to mediate viral replication and transcription. The coronavirus main protease (M<sup>pro</sup>), also known as the 3C-like protease (3CL<sup>pro</sup>) after the 3C proteases of the *Picornaviridae*, is a  $\approx$ 33 kDa cysteine protease that cleaves the replicase polyprotein at 11 conserved sites involving canonical Leu–Gln↓(Ser, Ala, Gly) sequences. The cleavage process is initiated by the enzyme’s own autolytic cleavage from pp1a and pp1ab [15, 16]. Its functional importance in the viral life cycle and the lack of closely related cellular homologs makes the M<sup>pro</sup> an attractive target for the development of drugs directed not



**Fig. 2** Functional oligomers of SARS-CoV proteins. **(A)** Nsp5, the main protease ( $M^{pro}$ ). SARS-CoV  $M^{pro}$ , shown in ribbon representation, is active as a dimer. **(B)** Nsp9, the ssRNA binding protein. SARS-CoV nsp9, shown in ribbon representation, functions as a dimer. **(C)** Nsp10, a zinc finger protein. SARS-CoV nsp10, shown in ribbon representation, can exist as a dodecamer in solution. The active form of nsp10 remains to be determined. Zinc ions are shown as grey

spheres. **(D)** Nsp15, the endoribonuclease. Nsp15, shown in ribbon representation, is active as a hexamer. **(E)** The S-protein fusion core. The HR1 and HR2 peptides together form a six-helix bundle characteristic of class I viral fusion proteins. **(F)** The N-protein dimerization domain. The C-terminal domain of the N-protein functions as a dimer

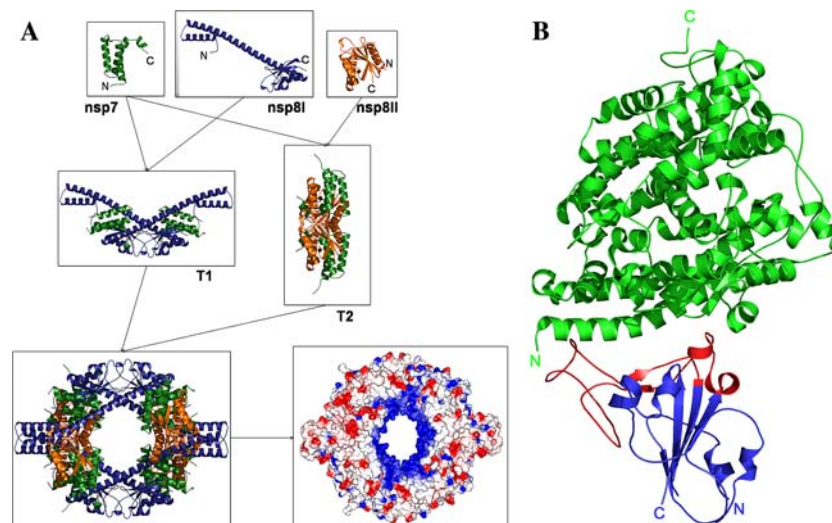
only against SARS, but also against other coronavirus infections.

The crystal structure of SARS-CoV  $M^{pro}$  was determined in 2003, mere months after the emergence of the epidemic, by our group in Tsinghua University, Beijing [17], and by the San Diego-based company Structural GenomiX (Fig. 1). Structural analysis confirmed that the functional unit of the  $M^{pro}$  is a dimer, with the first seven N-terminal residues (called the “N-finger”) important for stabilizing the active pocket of the neighbouring monomer (Fig. 2A). The availability of the  $M^{pro}$  structures in the Protein Data Bank enabled other researchers worldwide to design inhibitors targeting this important replication enzyme, thus speeding up drug development in case of the re-emergence of SARS. Prior to this, homology models constructed from the crystal structures of the  $M^{pro}$  from human coronavirus strain 229E (HCoV-229E) and porcine transmissible gastroenteritis virus (TGEV) [18, 19], both group I coronaviruses, were widely used to design anti-SARS inhibitors. However, comparison between the SARS-CoV  $M^{pro}$  structure and a homology model

constructed from HCoV-229E and TGEV  $M^{pro}$  (PDB ID: 1P9T) [19] showed a root-mean-square deviation of 3.8 Å [17]. There have since been widespread reports of various strategies used to design inhibitors targeting the SARS-CoV  $M^{pro}$  (see [20] for a review). In 2005, our group confirmed that the  $M^{pro}$  is significantly conserved among all three coronavirus antigenic groups and, moreover, that inhibitors designed to target the SARS-CoV  $M^{pro}$  can be effective ‘broad spectrum’ inhibitors against all coronavirus  $M^{pro}$  [21].

### Nsp7 and nsp8

In 2005, our group in Tsinghua University identified the interaction between two non-structural proteins, nsp7 and nsp8, by GST pulldown experiments. From the subsequent determination of the crystal structure of the nsp7–nsp8 protein–protein complex, eight copies of nsp7 and eight copies of nsp8 were observed to form an intricate hollow cylindrical scaffold (Fig. 3A) [22]. The inner dimensions



**Fig. 3** SARS-CoV protein-protein complexes. **(A)** The structure of the nsp7–nsp8 supercomplex. The complex assembly is formed by eight copies of nsp7 and nsp8. Nsp8 exists simultaneously in two conformations, termed nsp8I and nsp8II. Nsp7, nsp8I and nsp8II are shown in ribbon representation (top) and colored blue, green and orange respectively. The complex (below) is assembled from two tetramers: T1, formed between nsp7 and nsp8I (center, left); and T2, formed between nsp7 and nsp8II (center, right). A surface representation showing the charge distribution is also shown (below right),

with positive charge colored in blue and negative charge colored in red. The positive charge distributed around the central channel of the nsp7–nsp8 complex is favourable for the passage of RNA. **(B)** The SARS-CoV S-protein receptor binding domain complexed with the receptor ACE2. The complex structure is shown in ribbon representation with the ACE2 receptor colored in green, the S-protein receptor binding domain (RBD) colored in blue and the S-protein receptor binding motif (RBM) colored in red

and electrostatic properties of the cylindrical nsp7–nsp8 structure enable it to encircle nucleic acid, and an interaction was demonstrated with dsRNA by EMSA and mutagenesis. The architecture and electrostatic properties are reminiscent of PCNA or the  $\beta$ -subunit ring, the processivity factors of DNA polymerase, leading us to postulate that the nsp7–nsp8 complex should be a processivity factor for the RNA-dependent RNA polymerase (nsp12). Interestingly, both nsp7 and nsp8 were found to possess novel folds: nsp7 is an  $\alpha$ -helical bundle, while nsp8 has a so-called ‘golf club’ fold with an N-terminal  $\alpha$ -helical ‘shaft’ domain and a C-terminal mixed  $\alpha/\beta$  ‘head’ domain (Fig. 1). Within the complex framework, nsp8 exists simultaneously in two conformations: one with an extended  $\alpha$ -helical ‘shaft’ domain, and the other with a bent ‘shaft’ domain. The solution structure of nsp7 alone, also determined in 2005 by the Scripps consortium, adopts the same  $\alpha$ -helical bundle observed in the crystal structure [23].

In a follow-up study by Imbert and colleagues from the French consortium [24], it was reported that nsp8 constitutes a second RNA-dependent RNA polymerase (RdRp) in addition to nsp12, which includes an RdRp domain conserved in all RNA viruses. Distant structural homology was found between nsp8 and the catalytic palm subdomain of RNA virus RdRps. Further activity assays confirmed that nsp8 recognizes specific short sequences in the ssRNA coronavirus genome to catalyze the synthesis of <6

nucleotides with low fidelity. The properties of nsp8 indicate that it most likely functions as a primase to catalyze the synthesis of RNA primers for the primer-dependent nsp12, which is a unique characteristic of coronaviruses. It is worth noting that nsp8 alone can form a complex in solution and possesses similar activity to the nsp7–nsp8 complex, but has poor thermal stability as predicted from our crystal structure. Nsp7 therefore serves as ‘mortar’ to stabilize the nsp8 scaffold.

### Nsp9, a single-stranded RNA binding protein

Crystal structures of nsp9 were determined simultaneously in 2004 by the French consortium (to 2.7 Å resolution) [25] and by the SPINE consortium (to 2.8 Å resolution) [26], and established its previously unknown function as a single-stranded RNA binding protein whose biological unit is a dimer (Fig. 2B). The core structure of the protein is an open 6-stranded  $\beta$ -barrel reminiscent of, yet unrelated to, the nucleic acid binding OB (oligosaccharide/oligonucleotide binding) fold (Fig. 1). Searches for structural homology revealed that nsp9 shares similarity with certain subdomains of serine proteases, including domain II of the SARS-CoV M<sup>Pro</sup>. Based on the similarity to the picornavirus 3C proteases, which feature a conserved RNA binding motif, it was inferred that nsp9 should bind ssRNA,

and subsequently confirmed by EMSA assay and surface plasmon resonance. One role of nsp9 may be to stabilize nascent and template RNA strands during replication and transcription and protect them against nuclease processing. Besides replication, it is believed that nsp9 may also be involved in base-pairing driven processes such as RNA processing.

### Nsp10, a novel zinc-finger protein

An international collaborative effort between the Chinese and American groups led to the determination of SARS-CoV nsp10 as both a dodecamer [27] and monomer [28], respectively. The monomer structure, possessing a novel fold, contains two zinc-fingers with the sequence motifs C-(X)<sub>2</sub>-C-(X)<sub>5</sub>-H-(X)<sub>6</sub>-C and C-(X)<sub>2</sub>-C-(X)<sub>7</sub>-C-(X)-C (Fig. 1). These zinc finger motifs are strictly conserved among the three coronavirus antigenic groups, implying an essential function for nsp10 in all coronaviruses. A PFAM search identified a match for nsp10 with the HIT-type zinc finger proteins, which had previously not been structurally characterized. While zinc finger proteins often play a role in transcription, the precise function of nsp10 in the viral life cycle remains to be determined. Nsp10 is located next to nsp8 and nsp9 in the SARS-CoV genome; both nsp8 and nsp9 are known to interact with RNA, and nsp10 features a large patch of positive charge distributed on its surface, all of which suggest that nsp10 should also interact with nucleic acid. However, our experiments and those of Joseph and colleagues found only weak affinity between nsp10 and both ssRNA and dsRNA. Further work is also needed to ascertain the significance of the oligomeric state of SARS-CoV nsp10 (Fig. 2C). The monomer structure has an intact second zinc-finger which appears to stabilize the C-terminal tail of nsp10. However, in the dodecamer structure, the second zinc-finger lacks the last cysteine residue and the remainder of the C-terminal tail is disordered.

### Nsp15, an endoribonuclease

The crystal structures of nsp15 have been determined from SARS-CoV by the French consortium [29] and mouse hepatitis virus (MHV) by the Chinese consortium [30]. Both SARS-CoV and MHV belong to the antigenic group II of the genus *Coronavirus*. The function of nsp15 is an XendoU ribonuclease and the active biological unit is a hexamer (Fig. 2D). Nsp15 has a novel fold and is the first member of the XendoU family of endoribonucleases to be characterized, providing the first structural and mechanistic characteristics for this family of enzymes. It also represents

the first crystal structure of an endoribonuclease from the genus *Coronavirus*. The nsp15 monomer structure consists of three subdomains: a small N-terminal formed by two  $\alpha$ -helices packed against a three stranded  $\beta$ -sheet; a middle domain comprising of a mixed  $\beta$ -sheet, two smaller  $\beta$ -sheets and two short  $\alpha$ -helices; and a C-terminal domain made up of two  $\beta$ -sheets and five  $\alpha$ -helices. Each of the three subdomains in turn has a novel fold (Fig. 1).

Only the hexameric form of nsp15 is known to bind RNA, and the affinity of interaction can be increased by Mn<sup>2+</sup> ions. The US consortium recently determined the crystal structure of SARS-CoV nsp15 in a shortened monomeric form as a means of understanding the relationship between hexamer formation and activity (P. Kuhn, personal communication). In the absence of monomer-monomer interactions, the catalytic loop of nsp15 flips back to occupy the active site cleft. Given the critical importance of nsp15 in the viral life cycle, it is therefore an attractive target for anti-viral drug design. Strategies for inhibitor design therefore include the design of active site inhibitors, non-peptidyl compounds that mimic the catalytic loop of nsp15, and compounds that disrupt formation of the hexamer species.

### Structural proteins

The SARS-CoV genome encodes four structural proteins that are required to drive cytoplasmic viral assembly: the spike (S) protein, the membrane (M) protein, the nucleocapsid (N) protein and the envelope (E) protein. The S-protein is mainly responsible for binding to the host cell and for subsequent cell entry by virus-cell membrane fusion. We will focus on the S-protein and N-protein, whose partial structures have been solved.

### SARS-CoV spike protein fusion core

Similar to other class I virus fusion proteins, the SARS-CoV S-protein can be divided into an N-terminal half (S1) and C-terminal half (S2), but without proteolytic cleavage [31]. S1 and S2 are individually responsible for variations in host range and tissue tropism by its receptor specificity and cell entry by virus-cell membrane fusion [32]. S1 is responsible for binding to cellular receptors, and one potential SARS-CoV receptor has been identified as angiotensin-converting enzyme 2 (ACE2) [33]. S2 contains an internal fusion peptide and has two hydrophobic (heptad) repeat regions designated HR1 and HR2 [34]. The putative fusion peptide has recently been identified upstream close to HR1 [35]. HR2 is located close to the transmembrane region some 170 amino acids (aa)

downstream of HR1 [34]. Don Wiley and colleagues first established the classical mechanism of class I fusion proteins for mediating enveloped virus and host-cell membrane fusion from their comprehensive study of influenza hemagglutinin (HA) [36, 37]. In subsequent years, a common fusion mechanism has been established from extensive structural studies on the viral families of orthomyxovirus, retrovirus, paramyxovirus, and filovirus [36].

In 2004, the spike (S) protein fusion core was determined by two groups in the postfusion (or fusion-active) state, albeit by employing slightly different strategies [31, 38]. The Chinese structural proteomics initiative utilized a single chain by engineering a linker between the HR1 and HR2 domains to prepare the fusion core (HR1: 900–948, HR2: 1145–1184), while Supekar and colleagues individually synthesized longer HR1 and HR2 peptides (HR1: 889–972, HR2: 1142–1185). Both structures exhibit a six-helix bundle in which three HR1 helices form a central coiled-coil surrounded by three HR2 helices in an oblique, antiparallel manner (Figs. 1, 2E). HR2 peptides pack into the hydrophobic grooves of the HR1 trimer in a mixed extended and helical conformation, representing a stable postfusion structure similar to that for HIV-1 gp41 [36]. The N-terminus of HR1 and the C terminus of HR2 are located at the same end of the six-helix bundle, which would place the fusion peptide and transmembrane region close together. Supekar and colleagues also provided a structure of S2 fragment consisting of a smaller peptide of HR1 (919–949) and a peptide of HR2 (1149–1193) with extra C-terminal residues in proximity to the transmembrane region [31]. The C-terminal part is  $\alpha$ -helical and points away from the HR1 trimer axis, probably resulting from the lack of stabilization by the corresponding HR1 region, and may mimic the conformation of this region before the formation of the final postfusion hairpins. A later structure reported by Duquerroy and colleagues (HR1: 890–973, HR2: 1145–1190) emphasized the hydrogen-bonding network formed by conserved asparagine and glutamine, together with two possible chlorides, which could stabilize the conformation of postfusion hairpin [39]. Fusogenic mechanisms mediated by SARS-CoV were proposed according to those of other class I fusion proteins, although the possible conformational changes of the HR1 and HR2 fusion peptides during the membrane fusion process need further structural studies in the native state of S-protein and the pre-hairpin intermediate probably resulting from S1 binding to a receptor (e.g. ACE2).

Several peptides derived from HR1 and HR2 regions of SARS-CoV spike proteins have been demonstrated to block viral entry by targeting the putative pre-hairpin intermediate [40, 41, 42]. For instance, peptides derived from HR2, and not from HR1, are sufficient to inhibit

SARS-CoV infection [40, 41]. Interestingly, the efficacy of HR2 peptides derived from the SARS-CoV spike protein is lower than those of corresponding HR2 peptides of MHV in inhibiting MHV infection [40]. This might be explained by the larger surface area buried in the HR1–HR2 interface of MHV S2 than in SARS-CoV S2, this resulting in a higher affinity of the MHV peptides for the corresponding HR1 trimer [40], since a larger surface area is buried by the MHV S2 HR1–HR2 interface than by the SARS-CoV S2 [31]. In any case, the availability HR1–HR2 fusion core structure will help in the discovery of viral entry inhibitors against SARS.

### SARS-CoV spike protein receptor binding domain

An important part of the structure–function studies of any virus is to characterize its interaction with possible host cellular receptors. In the case of SARS-CoV, one known cellular receptor is ACE2 [33]. In 2005, Stephen Harrison and colleagues determined the structure of the SARS-CoV S-protein receptor-binding domain (RBD, covering residues 318 to 510 of the S-protein) with the ACE2 receptor (Fig. 1) [43]. The RBD is the critical determinant of virus-receptor interaction and thus of viral host range and tropism.

The specific recognition of ACE2 by the SARS-CoV RBD occurs through surface complementarity (Fig. 3B). The interface between the RBD and the ACE2 receptor is well defined, while the opposite face of the RBD, which would interact with the rest of the spike protein, is more disordered. As revealed by the authors, the interface between the two proteins shows important residue changes that facilitate efficient cross-species infection and human-to-human transmission. ACE2 is highly conserved in mammals and birds, and its receptor activity for SARS-CoV can be markedly affected by only a few amino acid substitutions at the virus binding site. Subtle changes in the RBD residues at positions 479 and 487 in human coronaviruses can increase affinity for human ACE2. Palm civet coronaviruses have lysine in position 479 and serine in position 487, which reduce affinity for human but not palm civet ACE2. The authors further suggest ways to make truncated disulfide-stabilised RBD variants for use in the design of coronavirus vaccines.

### SARS-CoV nucleocapsid protein RNA binding domain

Specific packaging of the viral genome into the virion is a critical step in the life cycle of an infectious virus. The nucleocapsid protein (N-protein) plays an important role by binding to the genomic RNA via a leader sequence,



recognizing a stretch of RNA that serves as a packaging signal and leading to the formation of the helical ribonucleoprotein (RNP) complex during assembly. The structure of the RNA binding domain from the SARS-CoV N-protein, consisting of a five-stranded  $\beta$ -sheet whose fold is unrelated to other RNA binding proteins, has been determined by NMR (covering residues 45–181) [44] and two X-ray crystallographic studies (covering residues 45–175) (Fig. 1) [45]. The authors of the NMR study identified a binding site for single stranded RNA (ssRNA) using NMR to determine the resonance of residues perturbed by the addition of RNA. The RNA binding groove in the N-terminal domain of the N-protein is shallow and should be able to bind both single- and double-stranded RNA in infected cells. The structure of the N-protein RNA binding domain exhibits a similar mode of interaction with RNA binding proteins such as U1A RNP. The more recent X-ray crystal structures of the N-terminal RNA binding domain of the N-protein are similar overall to the NMR structure and to two structures from avian infectious bronchitis virus (IBV) [46], a group III coronavirus. It was suggested that the SARS-CoV and IBV structures imply a common mode of RNA recognition, but homology modelling predicts this is not necessarily the case for related coronavirus N-proteins. The discovery of small molecules that bind to the RNA binding domain, as identified from an NMR-based screen by Huang and colleagues, might impair the function of the nucleocapsid [44].

### SARS-CoV nucleocapsid protein dimerization domain

The full-length N-protein is known to form a dimer in solution via its C-terminal domain. A crystal structure of this so-called dimerization domain, covering residues 270–370, was reported in 2006 (Fig. 1) [47]. The structure was determined as a dimer and featured extensive interactions between the two protomers, consistent with the dimeric nature of the full-length protein (Fig. 2F). Sequence alignments suggest that the core dimerization domain is conserved among the three coronavirus antigenic groups. A DALI search for structural similarity did not yield any results, but nevertheless the authors found common structural features shared by the nucleocapsid protein of an arterivirus, porcine reproductive and respiratory syndrome virus (PRRSV). The coronaviruses and arteriviruses both belong to the *Nidovirales* and, from a structural basis, it is suggested that they are evolutionarily linked. From a functional aspect, the structure of the N-protein dimerization domain helps to explain the self-association of the N-protein to form a large helical nucleocapsid core. Dimerization is believed to bring the N-terminal RNA binding domains of N proteins into close proximity, thus enabling

them to interact with the viral RNA and effectively package the large viral genome into the virion.

It is also worth noting that antigenic peptides of the coronavirus N-protein can be recognized on the surface of infected cells by T cells [48, 49]. The structure of the MHC-I molecule HLA-A\*1101 in complex with such a peptide derived from the SARS-CoV N-protein, a nonamer with SARS specific sequence, was determined to 1.45 Å resolution in 2005 [50]. Although it is similar with other MHC-I molecules and shows a similar peptide binding mode, the structure adds to the growing library of MHC-I structures and could be used as a template for peptide-based vaccine design.

### Supramolecular architecture of S, M and N structural proteins

While not strictly part of the structural proteomics remit, it is worth including the 2006 work by the Scripps consortium using cryo-electron microscopy to study the supramolecular architecture of the S, M and N structural proteins [51]. Their resulting model shows interactions between S–M, M–M and M–N near the viral membrane in accord with previous observations. Proteins located close to the viral membrane are arranged in overlapping lattices and surrounding a disordered core. The trimeric glycoprotein spikes appear to be in register with densities for four underlying ribonucleoproteins. The spikes were dispensable for ribonucleoprotein lattice formation, and ribonucleoprotein particles exhibited coiled shapes following release from the viral membrane. The overall results suggest that lattice formation by structural proteins is integral to coronavirus budding.

### Accessory proteins

In addition to the structural and non-structural proteins, the SARS-CoV genome encodes a further eight so-called “accessory” proteins unique to this coronavirus. Viruses frequently make use of alternative open reading frames to achieve greater output from their limited genomes. Out-of-frame translation is initiated from a start codon within an existing gene and results in a distinct protein product. These accessory proteins are poorly characterized structurally and their functions are largely unknown. They are believed to be unimportant in tissue culture but may provide the virus with a selective advantage in the infected host. The structures of two accessory proteins have been determined to date: orf7a and orf9b.

The crystal structure of the SARS-CoV orf7a luminal domain was reported in 1995 by Nelson and colleagues

[52]. At the time, significant progress had been made in understanding the structure–function relationships of SARS-CoV proteins with essential replication or structural roles. However, the functions of the accessory proteins which are coronavirus group-specific were poorly understood. The structure of the first accessory protein from SARS-CoV therefore provided important new information. The orf7a luminal domain is an all- $\beta$  structure comprising seven  $\beta$ -strands in two  $\beta$ -sheets (Fig. 1). Fold assignment indicates the orf7a luminal domain is similar to I-set Ig proteins and places it as a member of the Ig superfamily, despite low sequence identity with other Ig-like proteins. The function of Ig-like proteins is diverse, but subcellular localization experiments confirm that orf7a is expressed and retained intracellularly. Furthermore, the short cytoplasmic tail and transmembrane domain are implicated in trafficking orf7a in the endoplasmic reticulum and Golgi network. It follows that possible functions of orf7a might include roles in viral assembly or SARS-specific budding events, or as a secondary attachment protein within the virion analogous to the hemagglutinin-esterase (HE) protein.

The SARS-CoV orf9b crystal structure, a new fold, was solved by the SPINE consortium [53]. It has a dimeric  $\beta$  structure with an amphipathic surface and a central hydrophobic tunnel which is confirmed to bind lipid molecules (Fig. 1). SARS-CoV orf9b most likely involves in membrane attachment and further functional studies confirmed that orf9b associates with intracellular vesicles in mammalian cells. The authors propose that SARS-CoV orf9b may interact with compartments of the ER-Golgi network to act as an accessory protein during the assembly of the SARS virion.

### Biological implications of SARS-CoV structural proteomics

Since the emergence of SARS in 2003, a substantial number of full-length SARS-CoV proteins or functional domains have been determined by X-ray crystallography or NMR. Structures are now available for half of the 16 non-structural proteins involved in viral replication and transcription, providing us with a much greater understanding of the inner workings of this large and sophisticated machinery. The three SARS-CoV structural proteomics initiatives operate independently but there is good communication and co-operation between them, and overlaps are generally avoided even when groups are working on the same protein targets. For instance, the Chinese and American initiatives joined forces in 2006 to report the structure of SARS-CoV nsp10 [28, 27]; the Chinese group reported an nsp10 dodecamer structure while the American

group reported the monomer structure. In the case of nsp15, the French group reported the structure of the active hexameric form from SARS-CoV [29]; the Chinese group reported the active hexameric form of nsp15 from MHV [30]; and the American group reported a shortened and inactive monomeric form of nsp15 from SARS-CoV (P. Kuhn, personal communication). The different perspectives offered by the three structural proteomics initiatives can provide deeper, more penetrating insights into the structure–function relationships of SARS-CoV proteins.

One interesting and significant outcome of the SARS-CoV structural proteomics initiatives is the prevalence of new protein folds. Remarkably, of the 16 SARS-CoV proteins or functional domains with known structure to date, eight of them possess new folds, representing a fold discovery of about 50% (Fig. 1). This is in contrast to current estimates which put the discovery of new folds by structural genomics efforts targeting other organisms at somewhere between 5 and 7%. The overall rate of fold discovery is currently estimated at around 10%. This is perhaps not surprising as viruses are the most biodiverse of all biological entities. One of the principal aims of structural genomics is completion of the protein fold space, and in this regard the SARS-CoV structural proteomics initiatives have been successful. The addition of new folds to the Protein Data Bank should improve understanding of the structure–function relationships of several new families of proteins.

At the time of the 2003 outbreak, there were no therapeutic agents against SARS-CoV or indeed against any other coronavirus. Coronavirus research up to that point had been limited, largely due to the lack of medical or economic incentives as human coronaviruses were considered relatively harmless. Until the emergence of SARS, coronaviruses had been known to cause predominantly severe diseases in animals and only comparatively mild diseases in humans. Coronaviruses account for a significant percentage of upper and lower respiratory tract infections in humans, including common colds, bronchiolitis and pneumonia, and are also implicated in otitis media, exacerbations of asthma, diarrhoea, myocarditis and neurological disease [54–56, 57, 58, 59]. Anti-coronavirus drug discovery strategies to date have generally been focused in two main areas: blocking viral entry into the host cell, or inhibiting viral replication and transcription. In the case of the former, the availability of SARS-CoV spike protein fusion core structures will enable the design of inhibitors that block viral entry by targeting the pre-fusion hairpin intermediate [60]. In the latter case, three major conserved targets have been identified among the SARS non-structural proteins: nsp5, the main protease; nsp12, the RNA dependent RNA polymerase; and the RNA helicase [21].

While SARS was brought under control by effective global public health measures and is no longer in circulation among humans, there is still a possibility that it could re-emerge. The recent discovery of animal reservoirs for a SARS-like coronavirus has prompted new public health fears [61, 62]. Furthermore, the human coronaviruses HCoV-NL63 and HCoV-HKU1 were identified in the wake of SARS [58, 59]. Several key factors controlling the host spectrum and viral pathogenicity are highly variable among CoVs, including the requirement of different host receptors for cellular entry, poorly conserved structural proteins (antigens), and diverse accessory genes in their 3'-terminal genome regions that most likely contribute to the pathogenicity of CoVs in specific hosts [63, 64, 8, 65, 33, 6, 7, 59]. This structural and functional diversity presents a significant obstacle for the design of a versatile compound against all CoVs. For instance, a fusion peptide inhibitor derived from the MHV spike protein cannot prevent SARS-CoV replication in cell culture [40]. Identification of conserved structural targets among the coronaviruses will provide an opportunity for the development of broad-spectrum inhibitors against all CoV-related diseases.

### Concluding remarks

The emergence of SARS in 2003 had a particularly devastating impact, both to human health and to the global economy, and demonstrated how rapidly viruses can spread around the world. The outbreak also provided a stark warning of how ill-prepared we were at the time against a newly emerging infectious disease such as SARS. The paucity of available scientific data for coronaviruses was a considerable disadvantage, but scientists mounted a rapid international response to the threat of SARS. For instance the SARS coronavirus was quickly identified and its genome was sequenced within weeks [1, 2, 3, 4]. Ultimately, however, the disease was only brought under control by effective public health control measures. Since then, considerable efforts have been made by researchers around the world to understand the origins of the virus, its inner workings and its interaction with host cells.

The accumulated structural and functional data from the SARS-CoV structural proteomics initiatives will have many obvious benefits. First, the available structural information will provide a starting point for understanding important viral mechanisms. Specifically, the structures of the non-structural proteins will help elucidate their functions, many of which were previously unknown, and provide a vital starting point for understanding the unique and complex mechanism of coronavirus replication and transcription. Second, the new fold information provided by SARS-CoV structures will aid the understanding of the

structure–function relationships of several new protein families. Third, the availability of SARS-CoV structures provides targets for the structure-based discovery of antiviral compounds for therapeutic intervention. In the event of another emerging coronavirus, a stockpile of anti-coronaviral agents could provide an effective first line of defence.

Regarding the future prospects of SARS-CoV structural proteomics, significant challenges still lie ahead. All of the structural proteomics initiatives have experienced difficulties in expressing stable and functional SARS-CoV proteins. Furthermore, the SARS-CoV proteins that remain to be structurally characterized include several membrane proteins. While some progress has been made towards understanding the functions of the various SARS-CoV proteins, there is still a long way to go towards discovering how the proteins interact with each other, with the viral RNA and with host proteins. The complex structure of the S-protein RBD with the cellular receptor ACE2 is a significant step towards understanding the mechanisms of host recognition [43]. For the replicase proteins, we are slowly learning how they interact with one another within the replication machinery. Our group has already determined the complex structure between nsp7 and nsp8 [22]. In addition to their nsp9 structure, Sutton and colleagues showed evidence for its interaction with nsp8 [26]. Furthermore, dual-labeling studies of SARS-CoV replicase proteins have demonstrated colocalization of nsp8 with nsp2 and nsp3 [5]. The available three-dimensional structures of nsp7, nsp8 and nsp9 provide a starting point to reveal the architecture and underlying functions of the replication/transcription complex.

**Acknowledgments** This work was supported by Project 973 of the Ministry of Science and Technology of China (Grant numbers 2006CB806503, 2007CB914301), the International Cooperation Project of the Ministry of Science and Technology (Grant number 2006DFB32420), the NSFC (Grant number 30221003), the Chinese Academy of Sciences Knowledge Innovation Project (Grant number KSCX1-YW-R-05), the Sino-German Center (Grant number GZ236(202/9)), and the “Sino-European Project on SARS Diagnostics and Antivirals” (SEPSDA) of the European Commission (Grant number 003831).

### References

1. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA et al (2003) *N Engl J Med* 348:1967–1976
2. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W et al (2003) *N Engl J Med* 348:1953–1966
3. Kuiken T, Fouchier RA, Schutten M, Rimmelzwaan GF, van Amerongen G, van Riel D, Laman JD, de Jong T, van Doornum G, Lim W et al (2003) *Lancet* 362:263–270

4. Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, Nicholls J, Yee WK, Yan WW, Cheung MT et al (2003) *Lancet* 361:1319–1325
5. Prentice E, McAuliffe J, Lu X, Subbarao K, Denison MR (2004) *J Virol* 78:9977–9986
6. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattri J, Asano JK, Barber SA, Chan SY et al (2003) *Science* 300:1399–1404
7. Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH et al (2003) *Science* 300:1394–1399
8. Lai MMC, Holmes KV (2001) *Coronaviridae: the viruses and their replication*. Lippincott Williams & Wilkins
9. Ziebuhr J (2004) *Curr Opin Microbiol* 7:412–419
10. Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Guan Y, Rozanov M, Spaan WJ, Gorbalenya AE (2003) *J Mol Biol* 331:991–1004
11. Almeida MS, Johnson MA, Herrmann T, Geralt M, Wuthrich K (2007) *J Virol* 81:3151–3161
12. Saikatendu KS, Joseph JS, Subramanian V, Clayton T, Griffith M, Moy K, Velasquez J, Neuman BW, Buchmeier MJ, Stevens RC, Kuhn P (2005) *Structure* 13:1665–1675
13. Ratta K, Saikatendu KS, Santarsiero BD, Barretto N, Baker SC, Stevens RC, Mesecar AD (2006) *Proc Natl Acad Sci U S A* 103:5717
14. Egloff MP, Malet H, Putics A, Heinonen M, Dutartre H, Frangeul A, Gruez A, Campanacci V, Cambillau C, Ziebuhr J et al (2006) *J Virol* 80:8493–8502
15. Ziebuhr J (2005) *Curr Top Microbiol Immunol* 287:57–94
16. Ziebuhr J, Snijder EJ, Gorbalenya AE (2000) *J Gen Virol* 81:853–879
17. Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, Sun L, Mo L, Ye S, Pang H et al (2003) *Proc Natl Acad Sci U S A* 100:13190–13195
18. Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R (2002) *Embo J* 21:3213–3224
19. Anand K, Ziebuhr J, Wadhvani P, Mesters JR, Hilgenfeld R (2003) *Science* 300:1763–1767
20. Yang H, Bartlam M, Rao Z (2007) *Curr Pharm Design* 12:4573–4590
21. Yang H, Xie W, Xue X, Yang K, Ma J, Liang W, Zhao Q, Zhou Z, Pei D, Ziebuhr J et al (2005) *PLoS Biol* 3:e324
22. Zhai Y, Sun F, Li X, Pang H, Xu X, Bartlam M, Rao Z (2005) *Nat Struct Mol Biol* 12:980–986
23. Peti W, Johnson MA, Herrmann T, Neuman BW, Buchmeier MJ, Nelson M, Joseph J, Page R, Stevens RC, Kuhn P, Wuthrich K (2005) *J Virol* 79:12905–12913
24. Imbert I, Guillemot JC, Bourhis JM, Bussetta C, Coutard B, Egloff MP, Ferron F, Gorbalenya AE, Canard B (2006) *Embo J* 25:4933–4942
25. Egloff MP, Ferron F, Campanacci V, Longhi S, Rancurel C, Dutartre H, Snijder EJ, Gorbalenya AE, Cambillau C, Canard B (2004) *Proc Natl Acad Sci U S A* 101:3792–3796
26. Sutton G, Fry E, Carter L, Sainsbury S, Walter T, Nettleship J, Berrow N, Owens R, Gilbert R, Davidson A et al (2004) *Structure (Camb)* 12:341–353
27. Su D, Lou Z, Sun F, Zhai Y, Yang H, Zhang R, Joachimiak A, Zhang XC, Bartlam M, Rao Z (2006) *J Virol* 80:7902–7908
28. Joseph JS, Saikatendu KS, Subramanian V, Neuman BW, Brooun A, Griffith M, Moy K, Yadav MK, Velasquez J, Buchmeier MJ et al (2006) *J Virol* 80:7894–7901
29. Ricagno S, Egloff MP, Ulferts R, Coutard B, Nurizzo D, Campanacci V, Cambillau C, Ziebuhr J, Canard B (2006) *Proc Natl Acad Sci U S A* 103:11892–11897
30. Xu X, Zhai Y, Sun F, Lou Z, Su D, Xu Y, Zhang R, Joachimiak A, Zhang XC, Bartlam M, Rao Z (2006) *J Virol* 80:7909–7917
31. Supekár VM, Bruckmann C, Ingallinella P, Bianchi E, Pessi A, Carfi A (2004) *Proc Natl Acad Sci U S A* 101:17958–17963
32. Gallagher TM, Buchmeier MJ (2001) *Virology* 279:371–374
33. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC et al (2003) *Nature* 426:450–454
34. de Groot RJ, Luytjes W, Horzinek MC, van der Zeijst BA, Spaan WJ, Lenstra JA (1987) *J Mol Biol* 196:963–966
35. Duquerooy S, Rausch JM, Gallaher WR, Garry RF, Wimley WC (2005) *J Virol* 79:7195–7206
36. Eckert DM, Kim PS (2001) *Annu Rev Biochem* 70:777–810
37. Skehel JJ, Wiley DC (2000) *Annu Rev Biochem* 69:531–569
38. Xu Y, Lou Z, Liu Y, Pang H, Tien P, Gao GF, Rao Z (2004) *J Biol Chem* 279:49414–49419
39. Duquerooy S, Vigouroux A, Rottier PJ, Rey FA, Bosch BJ (2005) *Virology* 335:276–285
40. Bosch BJ, Martina BE, Van Der Zee R, Lepault J, Haijema BJ, Versluis C, Heck AJ, De Groot R, Osterhaus AD, Rottier PJ (2004) *Proc Natl Acad Sci U S A* 101:8455–8460
41. Liu S, Xiao G, Chen Y, He Y, Niu J, Escalante CR, Xiong H, Farmar J, Debnath AK, Tien P, Jiang S (2004) *Lancet* 363:938–947
42. Yuan K, Yi L, Chen J, Qu X, Qing T, Rao X, Jiang P, Hu J, Xiong Z, Nie Y et al (2004) *Biochem Biophys Res Commun* 319:746–752
43. Li F, Li W, Farzan M, Harrison SC (2005a) *Science* 309:1864–1868
44. Huang Q, Yu L, Petros AM, Gunasekera A, Liu Z, Xu N, Hajduk P, Mack J, Fesik SW, Olejniczak ET (2004) *Biochemistry* 43:6059–6063
45. Saikatendu KS, Joseph JS, Subramanian V, Neuman BW, Buchmeier MJ, Stevens RC, Kuhn P (2007) *J Virol* 81(8):3913–3921
46. Jayaram H, Fan H, Bowman BR, Ooi A, Jayaram J, Collisson EW, Lescar J, Prasad BV (2006) *J Virol* 80:6612–6620
47. Yu IM, Oldham ML, Zhang J, Chen J (2006) *J Biol Chem* 281:17134–17139
48. Bergmann C, McMillan M, Stohlman S (1993) *J Virol* 67:7041–7049
49. Boots AM, Kusters JG, van Noort JM, Zwaagstra KA, Rijke E, van der Zeijst BA, Hensen EJ (1991) *Immunology* 74:8–13
50. Blicher T, Kastrup JS, Buus S, Gajhede M (2005) *Acta Crystallogr D Biol Crystallogr* 61:1031–1040
51. Neuman BW, Adair BD, Yoshioka C, Quispe JD, Orca G, Kuhn P, Milligan RA, Yeager M, Buchmeier MJ (2006) *J Virol* 80:7918–7928
52. Nelson CA, Pekosz A, Lee CA, Diamond MS, Fremont DH (2005) *Structure (Camb)* 13:75–85
53. Meier C, Aricescu AR, Assenberg R, Aplin RT, Gilbert RJC, Grimes JM, Stuart DI (2006) *Structure* 14:1157–1165
54. Gagneur A, Sizun J, Vallet S, Legr MC, Picard B, Talbot PJ (2002) *J Hosp Infect* 51:59–64
55. Holmes KV (2001) *Coronaviruses*. Lippincott Williams & Wilkins
56. Myint SH (1995) In: Siddell SG (ed) *The Coronaviridae*. Plenum Press, New York, pp 389–401
57. Siddell SG, Ziebuhr J, Snijder EJ (2005) In: Mahy BWJ, ter Meulen V (eds) *Topley & Wilson's microbiology and microbial infections*, 10th edn. Hodder Arnold, London, pp 823–856
58. van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B (2004) *Nat Med* 10:368–373
59. Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, Wong BH, Poon RW, Cai JJ, Luk WK et al (2005) *J Virol* 79:884–895
60. Bartlam M, Yang H, Rao Z (2005) *Curr Opin Struct Biol* 15:664–672

61. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY (2005) *Proc Natl Acad Sci U S A* 102:14040–14045
62. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H et al (2005b). *Science* 310:676–679
63. de Haan CA, Masters PS, Shen X, Weiss S, Rottier PJ (2002) *Virology* 296:177–189
64. Hofmann H, Pyrc K, van der Hoek L, Geier M, Berkhout B, Pohlmann S (2005) *Proc Natl Acad Sci U S A* 102(22):7988–7993
65. Jeffers SA, Tusell SM, Gillim-Ross L, Hemmila EM, Achenbach JE, Babcock GJ, Thomas WD Jr., Thackray LB, Young MD, Mason RJ et al (2004) *Proc Natl Acad Sci U S A* 101:15748–15753