Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 30 August 2007
Accepted 30 October 2007

The search for a structural basis for therapeutic intervention against the SARS coronavirus

Mark Bartlam,[a,b,c] Xiaoyu Xue[b,c] and Zihe Rao[a,b,c]*

[a]College of Life Sciences, Nankai University, Tianjin 300071, People's Republic of China, [b]Tsinghua–Nankai–IBP Joint Research Group for Structural Biology, Laboratory of Structural Biology, Life Sciences Building, Tsinghua University, Beijing 100084, People's Republic of China, and [c]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China. Correspondence e-mail: raozh@xtal.tsinghua.edu.cn

The 2003 outbreak of severe acute respiratory syndrome (SARS), caused by a previously unknown coronavirus called SARS-CoV, had profound social and economic impacts worldwide. Since then, structure–function studies of SARS-CoV proteins have provided a wealth of information that increases our understanding of the underlying mechanisms of SARS. While no effective therapy is currently available, considerable efforts have been made to develop vaccines and drugs to prevent SARS-CoV infection. In this review, some of the notable achievements made by SARS structural biology projects worldwide are examined and strategies for therapeutic intervention are discussed based on available SARS-CoV protein structures. To date, 12 structures have been determined by X-ray crystallography or NMR from the 28 proteins encoded by SARS-CoV. One key protein, the SARS-CoV main protease (M^{pro}), has been the focus of considerable structure-based drug discovery efforts. This article highlights the importance of structural biology and shows that structures for drug design can be rapidly determined in the event of an emerging infectious disease.

1. Introduction

In 2003, a previously unidentified coronavirus, termed SARS coronavirus (SARS-CoV), was the aetiological agent for the worldwide epidemic responsible for approximately 8000 reported cases and 800 deaths (Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Kuiken *et al.*, 2003; Peiris *et al.*, 2003), and its emergence was attributed to an animal-to-human interspecies transmission (Prentice *et al.*, 2004). Coronaviruses are characterized as enveloped positive-stranded RNA viruses with the largest known genomes and belong to the genus *Coronavirus* of the family *Coronaviridae* (Marra *et al.*, 2003; Rota *et al.*, 2003). Approximately 26 species of coronaviruses (CoVs) have been identified to date and can be classified into three distinct groups on the basis of genome sequence and serological reaction (Lai & Holmes, 2001; Spaan & Cavanagh, 2004). Prior to the SARS 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 but comparatively mild diseases in humans, such as common colds caused by human coronaviruses. 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 before the SARS outbreak. No licensed drugs are currently available and strategies against coronavirus infection relied mainly on vaccines prior to the outbreak of SARS.

The global epidemic of severe acute respiratory syndrome (SARS) in 2003 had profound social and economic impacts all over the world, but particularly in China where the outbreak originated. Increased levels of support have been made available by governments and funding agencies; great efforts have been made by researchers to understand the origins of the SARS coronavirus, its interactions with the host, and the mechanisms of coronavirus replication and transcription; and considerable work has been made towards developing vaccines or anti-viral compounds to prevent SARS-CoV infection. Structural biology has so far played an important role in providing information for functional assignment of SARS-CoV proteins and for anti-viral drug discovery (Bartlam *et al.*, 2005).

As with many researchers in China, our group began work on SARS-CoV once the severity of the outbreak became apparent. Adopting a structural proteomics approach, a large project was initiated with strong support from the Chinese government and funding bodies. In the wake of the outbreak and the increased public awareness, other large projects such

Table 1
Summary of SARS-CoV proteins.

Protein†	Protein size (a.a.)	ORF (location in genome sequence)	Putative functional assignment(s)	Structure available
Structural proteins				
Spike (S) protein	1255	ORF2 (21492–25259)	ACE2 recognition, viral entry into infected cells, major virion coat glycoprotein	Yes (fusion core, receptor binding domain)
ORF3a	274	ORF3a (25268–26092)	Minor structural protein	No
Envelop (E) protein	76	ORF4 (26117–26347)	Structural protein, induces apoptosis in cells	No
Membrane (M) protein	221	ORF5 (26398–27063)	Viral structural glycoprotein	No
ORF7a	122	ORF7a (27273–27641)	Integral membrane protein	Yes (luminal domain)
Nucleocapsid (N) protein	422	ORF9a (28120–29388)	Genomic RNA packaging	Yes (N-terminal RNA binding domain, C-terminal domain)
Non-structural proteins (nsp)				
nsp1	180	ORF1a (265–804)		Yes
nsp2	638	ORF1a (805–2718)		No
nsp3	1922	ORF1a (2719–8484)	UB1, AC, ADRP, SUD, PL^{pro}, TM1-4, Zn finger, Y	Yes (Glu-rich‡, ADRP, PL^{pro} domains)
nsp4	500	ORF1a (8485–9984)	Transmembrane	No
nsp5	306	ORF1a (9985–10902)	M^{pro}	Yes
nsp6	290	ORF1a (10903–11772)	Transmembrane	No
nsp7	83	ORF1a (11773–12021)	RNA primase	Yes
nsp8	198	ORF1a (12022–12615)		Yes
nsp9	113	ORF1a (12616–12954)	ssRNA binding	Yes
nsp10	139	ORF1a (12955–13371)	RNA binding	Yes
nsp11	13	ORF1a (13372–13410)		No
nsp12	932	ORF1b (13398–16166)	RdRp	No
nsp13	601	ORF1b (16167–17969)	Helicase	No
nsp14	527	ORF1b (17970–19550)	3′→5′ exonuclease	No
nsp15	346	ORF1b (19551–20588)	Uridylate specific endonuclease	Yes
nsp16	298	ORF1b (20589–21482)	Putative methyltransferase	No
Accessory proteins				
ORF3b	154	ORF3b (25689–26153)		No
ORF6	63	ORF6 (26913–27265)		No
ORF7b	44	ORF7b (27638–27772)		No
ORF8a	39	ORF8a (27779–27898)		No
ORF8b	84	ORF8b (27864–28118)		No
ORF9b	98	ORF9b (28130–28426)	Lipid binding, putative membrane attachment	Yes

† Bold letters for the protein indicates a three-dimensional structure is available in the Protein Data Bank. ‡ Structure has been deposited in the Protein Data Bank but has not been published. Abbreviations. UB1: ubiquitin-like; AC: acidic Glu-rich domain; ADRP: adenosine diphosphate-ribose 100-phosphatase; SUD: SARS-CoV: unique domain; TM: transmembrane domain; PL^{pro}: papain-like protease; M^{pro}: main (or 3C-like cysteine) protease; RdRp: RNA-dependent RNA polymerase.

as SEPSDA (Sino-European Project on SARS Diagnostics and Antivirals, http://www.sepsda.org/) funded by the European Union and FSPS (Functional–Structural Proteomics of SARS CoV Related Proteins, http://visp.scripps.edu/SARS/) funded by NIAID and NIH have been established. A large part of their sphere of activity includes structural biology, aided by high-throughput technologies developed for structural genomics/proteomics. In this review, we will focus on achievements made by the structure–function studies of the SARS coronavirus proteins, and subsequent strategies for therapeutic intervention against SARS-CoV and other coronaviruses.

2. The SARS coronavirus

The SARS-CoV genome is approximately 29.7 kbp 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 (see Table 1)] 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) (Marra *et al.*, 2003; Rota *et al.*, 2003; Ziebuhr, 2004). 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. Of these 28 SARS-CoV proteins, 12 protein structures (X-ray crystallography or NMR) are available from the Protein Data Bank, thus providing a starting point for therapeutic intervention against the SARS coronavirus.

3. The replicase complex

The SARS-CoV replicase gene encodes 16 non-structural proteins (nsps) with multiple enzymatic functions (Snijder *et*

al., 2003), which are known or predicted to include types of enzymes that are common components of the replication machinery of plus-strand RNA viruses. These enzymes are typically not available or accessible in the host cell and are thus identified as potential targets for anti-SARS drug design. They include: an RNA-dependent RNA polymerase (RdRp, nsp12), a 3C-like cysteine protease (M^{pro} or $3CL^{pro}$, nsp5), a papain-like protease (PL^{pro}, nsp3), and a superfamily 1-like helicase (HEL1, nsp13). The replicase gene also encodes proteins less commonly found in positive-strand RNA viruses, which are indicative of 3′–5′ exoribonuclease 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) (Snijder *et al.*, 2003). These enzymes 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. Given the vital role of the replicase proteins in the viral life cycle, elucidating their function and how they interact to form the replicase complex is essential for determining strategies for the design of anti-viral compounds. However, further structural and functional studies of the replicase complex are still required for the discovery of anti-CoV therapeutics.

Nsp5, more commonly known as the main protease (M^{pro}) or $3CL^{pro}$ (for its similarity to 3C proteases), is the most widely investigated of the SARS-CoV proteins. Its crystal structure was reported in 2003, mere months after the outbreak, by our group (Yang *et al.*, 2003) and by the San Diego based company Structural GenomiX (Fig. 2*a*), although the first coronavirus M^{pro} structure was determined from transmissible gastroenteritis virus (TGEV) in 2002 (Anand *et al.*, 2002). The M^{pro} acts on 11 of the 14 cleavage sites on the replicase polyprotein to release the individual components of the replicase complex, and is therefore a critically important target for the discovery of anti-viral therapeutics. The coronavirus M^{pro} structures are characterized by two chymotrypsin-like β-barrel domains, similar to other viral proteases, and an additional C-terminal globular α-helical domain (Anand *et al.*, 2002, 2003; Yang *et al.*, 2003). The M^{pro} functions as a dimer and relies on the C-terminal domain for dimerization (Fig. 3*a*) (Shi *et al.*, 2004).

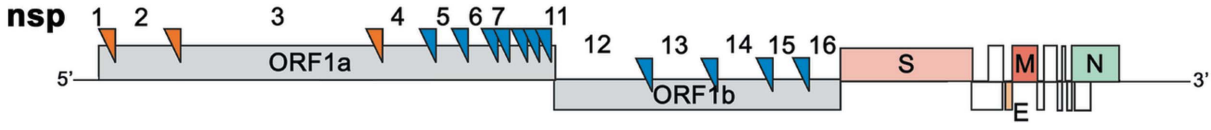

Figure 1
The SARS-CoV genome. Orange and blue triangles represent cleavage sites for the PL^{pro} and M^{pro}, respectively.

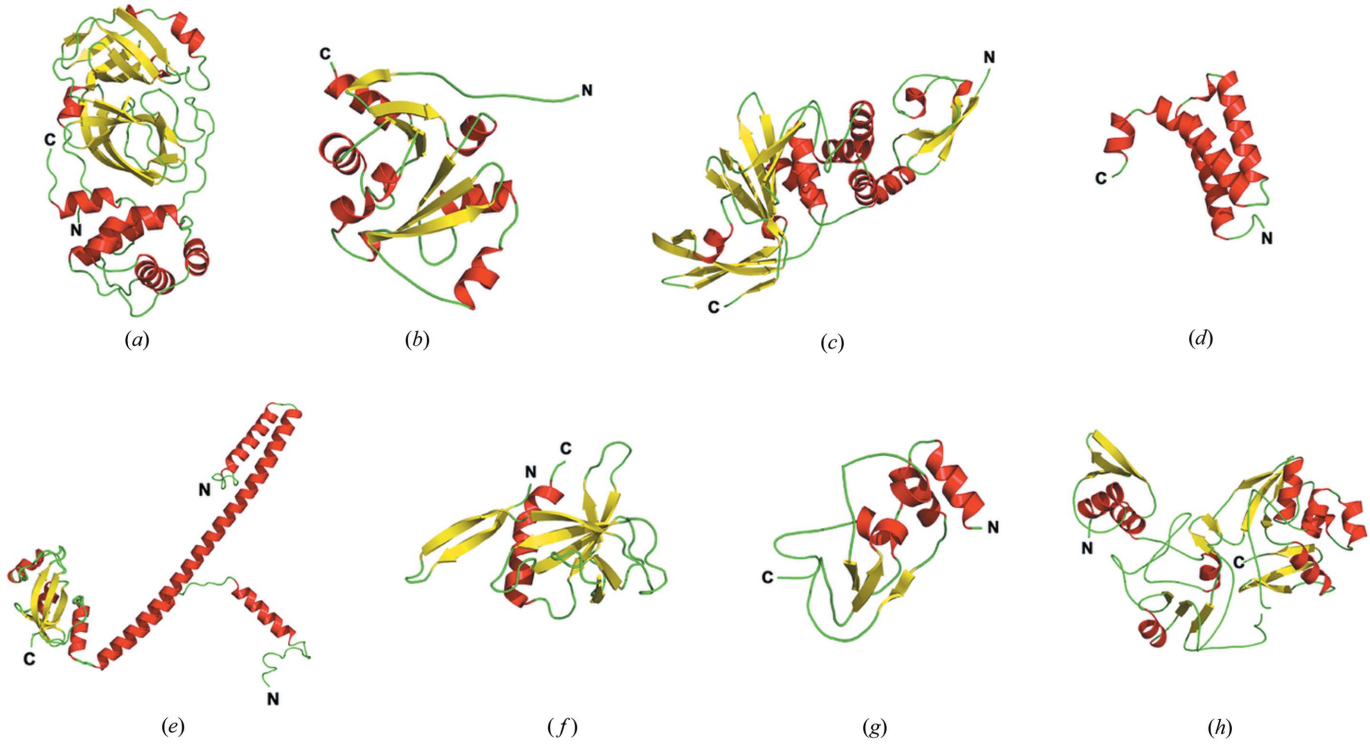

Figure 2
SARS-CoV replicase protein structures. (*a*) nsp5, the M^{pro}; (*b*) nsp3 ADRP domain; (*c*) nsp3 PL^{pro} domain; (*d*) nsp7; (*e*) nsp8; (*f*) nsp9; (*g*) nsp10; and (*h*) nsp15. All structures are shown in ribbon representation and coloured according to secondary structure (α-helix red; β-strand yellow; coil green). Nsp8 is shown with two conformations superimposed.

A catalytic Cys–His dyad and the substrate binding sites are located in a cleft between domains I and II.

The substrate specificity of SARS-CoV M^{pro} has been well characterized, both biochemically and structurally (Hegyi & Ziebuhr, 2002; Anand *et al.*, 2002, 2003; Yang *et al.*, 2003, 2005). All coronavirus M^{pro}s have an absolutely conserved Gln residue at the P1 position, whereas small residues such as Ala, Ser and Gly are conserved at the P1′ position (Ziebuhr *et al.*, 2000). At the P2 position of SARS-CoV M^{pro}, Leu is strongly preferred, although other hydrophobic residues, such as Phe, Met and Val, also occupy this position occasionally. No side-chain specificity is required at the P3 position since the side chain of P3 orients toward the bulk solvent. Small residues, such as Ala, Val, Thr and Pro are preferred at the P4 position. The structural information provided for coronavirus M^{pro} to date will prove useful for researchers to design inhibitors targeting SARS-CoV M^{pro}.

A number of strategies have been used to discover inhibitors targeting the coronavirus M^{pro} with nanomolar or low micromolar binding affinities [see Yang *et al.* (2007) for a review]. Our group has designed peptidomimetic ester inhibitors based on the natural N-terminal autocleavage substrate of the SARS-CoV M^{pro} and optimized using a structure-based approach (Fig. 3*a*) (Yang *et al.*, 2005). Furthermore, owing to the remarkable conservation of the active sites across all three coronavirus antigenic groups, our compounds have broad-spectrum activity against all coronavirus M^{pro}. Other classes of compounds found to have activity against coronavirus M^{pro} include anilides (Shie *et al.*, 2005), hexachlorophene and its analogues (Hsu *et al.*, 2004; Liu *et al.*, 2005), natural polyphenols (Chen, Lin *et al.*, 2005), isatin derivatives (Chen, Wang *et al.*, 2005), cinanserin (a serotonin agonist) (Chen, Gui *et al.*, 2005), interferons (Tan *et al.*, 2004), keto-glutamine analogues (Jain *et al.*, 2004), zinc conjugated compounds (Hsu *et al.*, 2004), aryl boronic acid compounds (Bacha *et al.*, 2004), quercetin-3-b-galactoside and its synthetic derivatives (Chen *et al.*, 2006), plant terpenoids and lignoids (Wen *et al.*, 2007), benzotriazole esters (Wu *et al.*, 2006), coumarin derivative (Hamill *et al.*, 2006), and other compounds (Kaeppler *et al.*, 2005; Lu *et al.*, 2006; Tsai *et al.*, 2006). In addition to active site

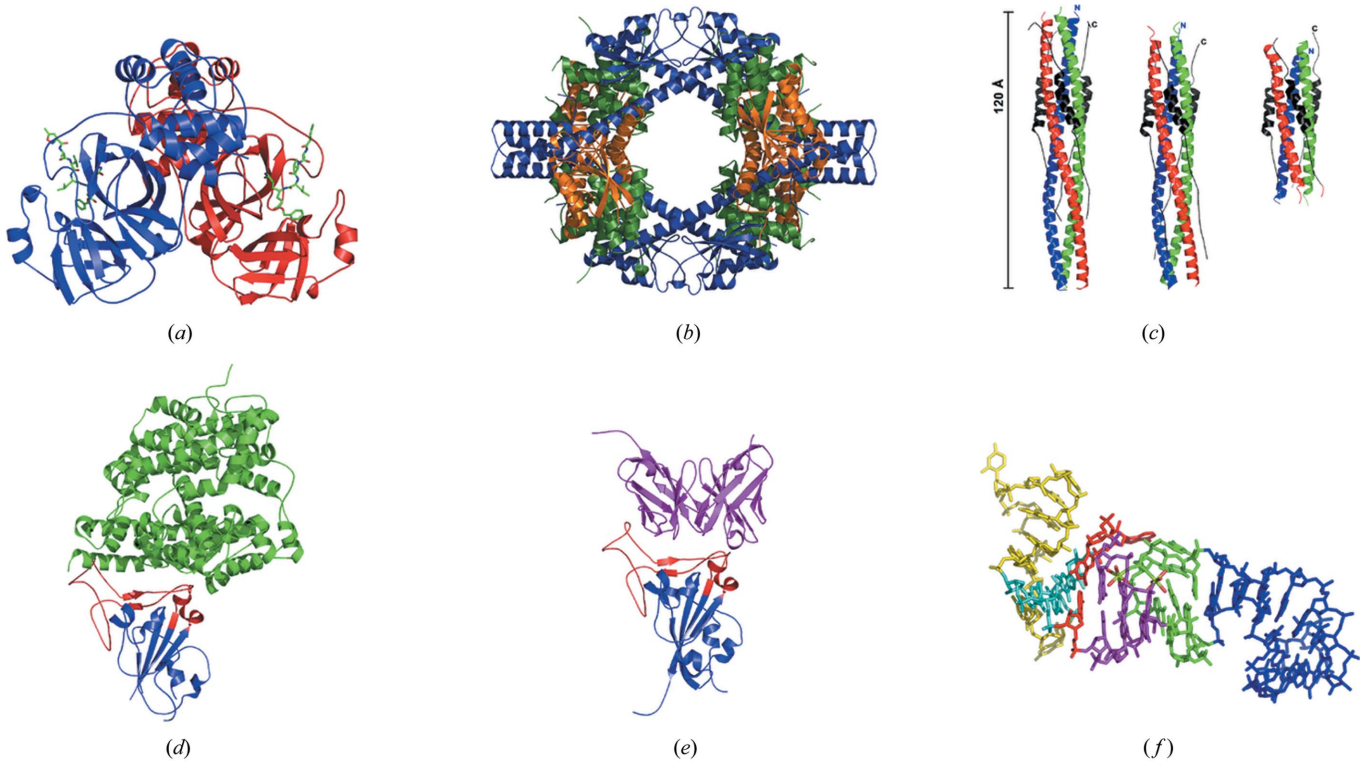

Figure 3
(*a*) The M^{pro} dimer with bound peptidomimetic ester inhibitor N3. M^{pro} monomers are shown in ribbon representation and coloured red and blue. The peptidomimetic ester inhibitors (one per monomer) are shown in stick representation and coloured green. (*b*) The nsp7–nsp8 super-complex. The nsp7–nsp8 complex is shown in ribbon representation. Nsp7 is coloured green, one conformation of nsp8 is coloured blue and a second conformation of nsp8 is coloured orange. (*c*) The S protein fusion core. Shown from left to right are 1WYY (Duquerroy *et al.*, 2005), 2BEZ (Supekar *et al.*, 2004) and 1WNC (Xu *et al.*, 2004). The central HR1 peptides are shown in ribbon representation and coloured red, blue and green. The HR2 peptides are shown in black. The N- and C-termini are labelled. (*d*) The S protein RBD bound to the cellular receptor ACE2. The complex structure is shown in ribbon representation with the ACE2 receptor coloured green, the S protein receptor binding domain (RBD) coloured blue and the S protein receptor binding motif (RBM) coloured red. (*e*) The S protein RBD bound to the 80R antibody. The complex structure is shown in ribbon representation with the 80R antibody coloured magenta, the S protein RBD coloured blue and the S protein RBM coloured red. The orientation of the S protein RBD is the same as for (*d*). (*f*) S2m, a rigorously conserved RNA element in the SARS-CoV genome. S2m is shown in stick representation and coloured according to the following scheme: GNRA-like pentaloop, is yellow; A-form RNA helices are blue and magenta; three-purine asymmetric bulge is red; seven-nucleotide bubble is green.

inhibitors, a second strategy is to inhibit the dimerization of the M^{pro} and thus abolish its activity. This approach was first suggested in 2004 (Shi *et al.*, 2004) and one such inhibitor, an octapeptide designed on the basis of the SARS-CoV M^{pro} N-terminal sequence, was later reported (Wei *et al.*, 2006). This work, together with our own, suggests that the design of peptidomimetics is one valid approach for the design of antiviral therapeutics targeting the coronavirus M^{pro}.

Nsp3 is a large multidomain protein of 1922 amino acids that is yielded by proteolytic cleavage of the pp1a polyprotein at two sites by the papain-like protease (PL^{pro}). Two crystal structures of the functional enzymatic domains of nsp3 have been determined: the 'X' domain with proposed ADP-ribose-1″-phosphate dephosphorylation (ADRP) activity (Saikatendu *et al.*, 2005; Egloff *et al.*, 2006) and the papain-like protease (PL^{pro}) domain (Ratta *et al.*, 2006). The 'X' domain, also known as the ADRP domain, is conserved among all CoV (Putics *et al.*, 2005) and is structurally related to macro-H2A-like fold proteins (Fig. 2*b*). Interestingly, the work by Egloff and colleagues suggests that this 'X' domain actually has poor ADRP activity and efficiently binds poly(ADP-ribose) instead (Egloff *et al.*, 2006), and its role in the viral life cycle remains unclear. Coronaviruses generally feature two papain-like protease (PL^{pro}) domains in nsp3, termed $PL1^{pro}$ and $PL2^{pro}$. However, SARS-CoV encodes only one PL^{pro} domain. Its structure was found to possess a 'thumb–palm–fingers' fold related to known deubiquitinating enzymes (Fig. 2*c*). However, certain key features of nsp3 PL^{pro} include a zinc-binding motif and a ubiquitin-like N-terminal domain, separating it from other characterized deubiquitinating enzymes. The availability of the nsp3 PL^{pro} structure helps to delineate 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.

Our group identified the interaction between two non-structural proteins, nsp7 and nsp8, and subsequent determination of the crystal structure of the nsp7–nsp8 protein–protein complex showed formation of an intricate hollow cylindrical scaffold comprised of eight copies of nsp7 and eight copies of nsp8 (Zhai *et al.*, 2005). Nsp7 (Fig. 2*d*), nsp8 (Fig. 2*e*) and the nsp7–nsp8 complex (Fig. 3*b*) all have novel structures, and nsp8 exists in two distinct conformations in the structure. The inner dimensions and electrostatic properties of the cylindrical nsp7–nsp8 structure enable it to encircle nucleic acid, suggesting that the nsp7–nsp8 complex might be a processivity factor for the RNA-dependent RNA polymerase (nsp12). A follow-up study by Imbert and colleagues (Imbert *et al.*, 2006) 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. Further activity assays confirmed that nsp8 recognizes specific short sequences in the ssRNA coronavirus genome and 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. Interestingly, a recent study has also shown that nsp8 can interact with the orf6 accessory protein (Kumar *et al.*, 2007), implying that the replication of SARS-CoV involves a rather complicated network of many proteins.

Crystal structures of nsp9 were reported in 2004 (Egloff *et al.*, 2004; Sutton *et al.*, 2004) and established its previously unknown function as a single-stranded RNA binding protein whose biological unit is a dimer. The core structure of the protein is an open six-stranded β-barrel reminiscent of, yet unrelated to, the nucleic acid binding OB (oligosaccharide/oligonucleotide binding) fold (Fig. 2*f*). Searches for structural homology revealed that nsp9 shares similarity with certain subdomains of serine proteases, including domain II of the SARS-CoV M^{pro}. Their similarity to the picornavirus 3C proteases, which feature a conserved RNA binding motif, indicated that nsp9 should bind also ssRNA. In addition to its role in the viral replication cycle, possible functions for nsp9 may be in stabilizing nascent and template RNA strands during replication and transcription to protect them against nuclease processing, or in base-pairing-driven processes such as RNA processing.

SARS-CoV nsp10 has been determined both as a dodecamer (Su *et al.*, 2006) and as a monomer (Joseph *et al.*, 2006). The monomer structure has a novel fold and contains two zinc fingers with the sequence motifs C–$(X)_2$–C–$(X)_5$–H–$(X)_6$–C and C–$(X)_2$–C–$(X)_7$–C–(X)–C (Fig. 2*g*). These zinc finger motifs are strictly conserved among the three coronavirus groups, implying an essential function for nsp10 in all coronaviruses, and a PFAM search yields a match with the previously uncharacterized HIT-type zinc finger proteins. While zinc finger proteins often play a role in transcription, the precise function of nsp10 in the viral life cycle remains to be determined. The location of nsp10 next to the RNA-interacting proteins nsp8 and nsp9 in the SARS-CoV genome would 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 determined by our group. We used a construct of nsp10 and nsp11 for crystallization, although nsp11, an 11 amino acid peptide, was not observed in the subsequent structure (Su *et al.*, 2006). The exact function of nsp11 in viral replication and transcription remains largely unknown.

Nsp15, an XendoU ribonuclease, has been determined from SARS-CoV (Ricagno *et al.*, 2006) and mouse hepatitis virus (MHV) (Xu *et al.*, 2006) in the active hexameric form, and from SARS-CoV as an inactive monomer (Joseph *et al.*, 2007). Nsp15 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. The nsp15 monomer structure has a novel fold and consists of three subdomains: a small N-terminal domain formed by two α-helices packed against a three-stranded β-sheet; a middle domain comprising a mixed β-sheet, two smaller β-sheets and two short α-helices; and a C-terminal domain made up of two β-sheets and five α-helices (Fig. 2*h*). In the shortened monomeric structure of nsp15 reported by Joseph and colleagues,

the catalytic loop flips back to occupy the active site cleft due to the absence of monomer–monomer interactions. Given the critical importance of nsp15 in the viral life cycle, it is therefore an attractive target for anti-viral drug design. Potential strategies for inhibitor design include active site inhibitors, peptidomimetics or non-peptidyl compounds that mimic the catalytic loop of nsp15, and compounds that disrupt formation of the hexamer species.

4. Structural proteins

While much of the focus of SARS structural biology work has been on the non-structural proteins, which include several conserved targets that are attractive for the design of therapeutics, other studies have been focused on the 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. More recently, two proteins originally labelled as accessory proteins have been reclassified as structural proteins. Orf3a is believed to be a minor structural protein with three membrane-spanning helices (Ito *et al.*, 2005; Shen *et al.*, 2005) and reports suggest it interacts with the spike protein and may influence its trafficking in the host cell (Tan, 2005). Orf7a, an integral membrane protein expressed on the membrane surface of host cells infected with the SARS virion, has also been suggested to be a structural protein (Huang *et al.*, 2006). The structure of the soluble luminal domain of orf7a has been determined, although the function of the full-length protein remains unclear (Nelson *et al.*, 2005).

Don Wiley and colleagues used their comprehensive study of influenza hemagglutinin (HA) to propose the classical mechanism of class I fusion proteins for mediating enveloped virus and host-cell membrane fusion (Skehel & Wiley, 2000; Eckert & Kim, 2001). A common fusion mechanism has since been established from extensive structural studies on the viral families of orthomyxovirus, retrovirus, paramyxovirus and filovirus (Eckert & Kim, 2001). The SARS-CoV S protein is typical of class I virus fusion proteins in that it can be divided into an N-terminal half (S1), which binds the host cellular receptor, and a C-terminal half (S2), responsible for cell entry *via* virus-cell membrane fusion (Gallagher & Buchmeier, 2001; Supekar *et al.*, 2004).

S2 contains two hydrophobic (heptad) repeat regions, HR1 and HR2 (de Groot *et al.*, 1987), which assemble into a six-helix bundle with three HR2 helices surrounding a central coiled coil of three HR1 helices in an oblique and antiparallel manner. Structures of the spike (S) protein fusion core have been reported by three groups in the post-fusion (or fusion-active) state (Fig. 3*c*) (Supekar *et al.*, 2004; Xu, Lou *et al.*, 2004; Duquerroy *et al.*, 2005). The N terminus of HR1 and the C terminus of HR2 locate at the same end of the six-helix bundle, which places the fusion peptide and transmembrane region in close proximity. Fusogenic mechanisms mediated by SARS-CoV were proposed from these structures according to those of other class I fusion proteins, although further structural studies are needed to determine the possible conformational changes of the HR1 and HR2 fusion peptides during the membrane fusion process.

Fusion peptides have successfully been used to develop anti-viral drugs. For instance, the membrane fusion-inhibitor peptide T-20 targets the HIV pre-fusion intermediate and was recently licensed by the US Food and Drug Administration as an anti-HIV drug. The CoV S protein fusion core has a stable post-fusion structure similar to HIV-1 gp41 (Eckert & Kim, 2001). In the case of SARS-CoV, several peptides derived from HR1 and HR2 regions of SARS-CoV spike proteins block viral entry by targeting the putative pre-hairpin intermediate (Bosch *et al.*, 2004; Liu *et al.*, 2004; Yuan *et al.*, 2004). Peptides derived from HR2 are sufficient to inhibit SARS-CoV infection (Liu *et al.*, 2004; Bosch *et al.*, 2004). 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 (Bosch *et al.*, 2004), which may be explained by the larger surface area buried in the HR1–HR2 interface of MHV S2 compared with that in SARS-CoV S2 (Xu, Liu *et al.*, 2004; Xu, Lou *et al.*, 2004; Supekar *et al.*, 2004).

An important part of the structure–function studies of any virus is to characterize the interaction with possible host cellular receptors. In the case of SARS-CoV, the S1 region of the S protein binds to cellular receptors, including the known receptor angiotensin-converting enzyme 2 (ACE2) (Li *et al.*, 2003). 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 known cellular receptor ACE2 (Fig. 3*d*) (Li, Li *et al.*, 2005). The RBD is the critical determinant of virus–receptor interaction and thus of viral host range and tropism. ACE2 specifically recognizes the SARS-CoV RBD by surface complementarity *via* a well defined interface; the opposite face of the RBD which interacts 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, for instance, which reduce affinity for human but not palm civet ACE2. The authors further suggest engineering truncated disulfide-stabilized RBD variants for use in the design of coronavirus vaccines.

80R is a potent neutralizing human monoclonal antibody against the S1 RBD and binds with nanomolar affinity (Sui *et al.*, 2004). It is known to block the binding of S1 to the ACE2 receptor, prevent the formation of syncytia *in vitro* (Sui *et al.*, 2004) and inhibit viral replication *in vivo* (Sui *et al.*, 2005). A crystal structure of 80R in complex with the SARS-CoV RBD shows that the 80R binding epitope overlaps with the ACE2

binding site, thus providing a structural basis for the strong binding and neutralizing ability of the 80R antibody (Fig. 3*e*) (Hwang *et al.*, 2006). The availability of a SARS-CoV RBD structure in complex with 80R should facilitate the design of immunotherapeutics targeting SARS-CoV.

Crystal structures have been determined for two domains from the SARS-CoV nucleocapsid protein, which 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. First, the structure of the RNA binding domain from the SARS-CoV N protein consists of a five-stranded β-sheet whose fold is unrelated to other RNA binding proteins (Huang *et al.*, 2004; Saikatendu *et al.*, 2007). The structure of the N protein RNA binding domain might constitute another significant target, since the discovery of small molecules that bind to the RNA binding domain should impair the function of the nucleocapsid (Huang *et al.*, 2004). Since specific packaging of the viral genome into the virion is a critical step in the life cycle of an infectious virus, this RNA binding domain might be a viable target for the design of anti-viral therapeutics. Second, the full-length N protein forms a dimer *via* its C-terminal domain, and a crystal structure of this so-called dimerization domain covering residues 270–370 has been reported (Yu *et al.*, 2006). The structure was determined as a dimer and features extensive interactions between the two protomers, consistent with the dimeric nature of the full length protein.

5. Accessory proteins

The SARS coronavirus genome encodes eight so-called accessory proteins with unclear or unknown function, but which might provide a selective advantage in the infected host. These accessory proteins are poorly characterized structurally and their functions are largely unknown, and so it is not clear if the accessory proteins may be viable targets for anti-viral drug discovery. However, it was recently suggested that the accessory proteins orf3a and orf7a should be reclassified as structural proteins (Ito *et al.*, 2005; Shen *et al.*, 2005; Huang *et al.*, 2006). As the accessory proteins vary among different coronaviruses, they almost certainly would not be targets for the design of broad-spectrum anti-virals. Two accessory protein structures have been determined to date: the orf7a luminal domain (Nelson *et al.*, 2005) and orf9b, a lipid-binding protein (Meier *et al.*, 2006).

6. Other targets

In addition to SARS-CoV protein structures, the crystal structure of the stem-loop II motif (s2m) RNA element from SARS-CoV was determined to 2.7 Å resolution (Fig. 3*f*) (Robertson *et al.*, 2005). S2m is a rigorously conserved motif located at the 3′ end of SARS-CoV and the genomes of other viral pathogens (Jonassen *et al.*, 1998) but is not found in the human genome. The highly structured s2m RNA element includes a striking 90° bend in the helix axis. Several longer-range tertiary interactions create a tunnel perpendicular to the main helical axis, with a negatively charged interior that binds two Mg ions. These unusual features form likely surfaces for interaction with conserved host-cell components or other reactive sites required for virus function. The authors suggest that s2m RNA is a functional macromolecular mimic of the 530 loop of 16S rRNA, a ribosomal RNA fold (Wimberly *et al.*, 2000), suggesting a mechanism for RNA hijacking of host protein synthesis similar to other RNA viruses (Bushell & Sarnow, 2002). The 530 loop of the 30S ribosome binds to the prokaryotic proteins S12 and IF-1, further suggesting that s2m may interact with their eukaryotic homologs (Robertson *et al.*, 2005). Nevertheless, the high sequence conservation of s2m in an otherwise rapidly changing RNA genome implies its pathogenic importance. The structural features of s2m, coupled with the fact that it is not found in the human genome, signals that it could be another attractive target for the design of antiviral therapeutics. Compounds designed to bind to s2m might disrupt the structure and thus inhibit SARS-CoV pathogenesis.

7. Vaccines against SARS

Vaccines provide another means of therapeutic intervention against SARS-CoV, and drew particular attention immediately after the SARS outbreak. Several strategies have been used to develop vaccines, including inactivated viruses, subunit vaccines, virus-like particles (VLPs), DNA vaccines, heterologous expression systems, and vaccines derived from SARS-CoV genome by reverse genetics [see Gillim-Ross & Subbarao (2006) for a recent review]. As described above, the S protein RBD could be used as a starting point for the development of a vaccine, since neutralizing antibodies against SARS-CoV recognize epitopes in the RBD. As suggested by the crystal structure, a candidate vaccine could be made by engineering the SARS-CoV RBD to improve stability (Li, Li *et al.*, 2005). In another example, the antigenic peptides of the coronavirus N protein are accessible on the surface of infected cells for T-cell recognition (Boots *et al.*, 1991; Bergmann *et al.*, 1993). Furthermore, in 2005, the crystal structure of the human MHC-I (major histocompatibility complex I) molecule HLA-A*1101 in complex with a nine amino acid peptide (KTFPPTEPK) derived from the SARS-CoV N-protein, was determined by X-ray crystallography to 1.45 Å resolution (Blicher *et al.*, 2005). Although it is similar to 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.

Another recent report suggests that the non-structural protein nsp1, encoded at the 5′ end of the replicase gene, is a major pathogenicity factor and could provide the basis for design of coronavirus vaccines (Zust *et al.*, 2007). Nsp1, whose structure was recently characterized by NMR (Almeida *et al.*, 2007), is the first mature viral protein expressed in the host cell cytoplasm (Ziebuhr, 2005) and may be involved with host cell mRNA degradation and counteracting innate immune responses (Kamitani *et al.*, 2006). A deletion in the nsp1

coding sequence in MHV was found to strongly reduce cellular gene expression, while low doses of nsp1 mutant MHV elicited potent cytotoxic T-cell responses (Zust *et al.*, 2007). Furthermore, mice inoculated with the nsp1 mutant MHV were protected against homologous and heterologous virus challenge. Nsp1 is conserved in all coronaviruses, and so this strategy for the development of coronavirus vaccines may prove effective for the majority of mammalian coronaviruses.

8. Future prospects

Structure–function studies of SARS-CoV proteins have significantly advanced our understanding of coronaviruses and should accelerate structure-based discovery of anti-viral therapeutics. However, a number of important targets remain to be elucidated, most notably among the replicase proteins. These include several membrane proteins and large multi-domain proteins and their structures will be challenging to determine. Foremost among these protein targets is undoubtedly nsp12, the RNA-dependent RNA polymerase (RdRp). Canonical polymerase sequence motifs can be identified in the C-terminal part of the RdRp, while coronaviral RdRp feature a unique N-terminal region of 380 amino acids with unknown function (Xu *et al.*, 2003). Despite numerous attempts by several groups, the SARS-CoV RdRp has proven to be difficult to produce in sufficient quantities for crystallization. The RdRp of other RNA viruses have been major targets for antiviral compounds. For instance, NS5A is the RdRp in hepatitis C virus (HCV) and a major target for non-nucleoside inhibitors (NNI) (Biswal *et al.*, 2005, 2006). The binding sites for thiophene-based NNIs are located in the 'thumb' domain of NS5B, in close proximity to the allosteric GTP binding site and approximately 35 Å from the active site. This part of the 'thumb' domain apparently has an important regulatory function that is modulated by GTP and NNIs. Interestingly, nsp8 is also believed to form a second RdRp for the synthesis of short RNA primers for nsp12. Since this function is unique to coronaviruses, the nsp8 primase may be an effective and specific target for anti-coronavirus therapeutics. Another major target is nsp13, the helicase, whose role is to unwind double-stranded genomic and subgenomic RNA during the replication process and whose three-dimensional structure remains to be determined. Nsp12 and nsp13, together with nsp5 (M^{pro}), share the highest sequence conservation among the three coronavirus groups and should be the focus of broad-spectrum anti-viral drug discovery (Yang *et al.*, 2005).

9. Conclusions

Viral proteins are notoriously difficult to work with, especially with regard to crystallization. Not every SARS-CoV protein may necessarily be a target for therapeutic intervention, but gaining an understanding of the underlying mechanisms of viral replication and host infection will help to identify and prioritize potential SARS-CoV targets. The advent of structural genomics/proteomics has considerably advanced progress in the structure–function studies of SARS-CoV proteins, thus providing a substantial increase in our understanding of coronaviruses.

To date, the structure-based discovery of anti-coronavirus therapeutics has been focused in two main areas: blocking viral entry into the host cell or inhibiting viral replication and transcription *via* the replicase complex. In the former case, the availability of SARS-CoV S protein fusion core structures will enable the design of inhibitors that block viral entry by targeting the pre-fusion hairpin intermediate. Structural differences between the SARS-CoV and MHV S protein fusion cores suggest that inhibitors designed to target the SARS-CoV S protein fusion core should be less efficient against other coronaviruses. In the latter case, three highly conserved proteins have been identified thus far among coronaviruses: nsp5, the M^{pro}; nsp12, the RdRp; and nsp13, the helicase. Targeting these three proteins should enable the design of anti-coronavirus therapeutics with broad-spectrum activity. In the event of a new emerging coronavirus, the availability of broad-spectrum inhibitors should provide a first line of defence until vaccines become available. However, at the time of writing, no anti-coronavirus drugs are available, either on the market or in pre-clinical or clinical trials. SARS-CoV may have been brought under control through effective surveillance and public health measures, but it should be noted that two human coronaviruses, HCoV-NL63 and HCoV-HKU1, have been isolated in the wake of SARS (van der Hoek *et al.*, 2004; Woo *et al.*, 2005) and animal reservoirs for a SARS-like coronavirus have also been identified (Lau *et al.*, 2005; Li, Shi *et al.*, 2005).

This work was supported by Project 973 of the Ministry of Science and Technology of China (Grant Nos. 2006CB806503, 2007CB914301), the International Cooperation Project of the Ministry of Science and Technology (Grant No. 2006DFB32420), the NSFC (Grant No. 30221003), the Chinese Academy of Sciences Knowledge Innovation Project (Grant No. KSCX1-YW-R-05), the Sino-German Center [Grant No. GZ236(202/9)], and the 'Sino-European Project on SARS Diagnostics and Antivirals' (SEPSDA) of the European Commission (Grant No. 003831).

References

Almeida, M. S., Johnson, M. A., Herrmann, T., Geralt, M. & Wuthrich, K. (2007). *J. Virol.* **81**, 3151–3161.
Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). *EMBO J.* **21**, 3213–3224.
Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. (2003). *Science*, **300**, 1763–1767.
Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A. & Freire, E. (2004). *Biochemistry*, **43**, 4906–4912.
Bartlam, M., Yang, H. & Rao, Z. (2005). *Curr. Opin. Struct. Biol.* **15**, 664–672.
Bergmann, C., McMillan, M. & Stohlman, S. (1993). *J. Virol.* **67**, 7041–7049.
Biswal, B. K., Cherney, M. M., Wang, M., Chan, L., Yannopoulos, C. G., Bilimoria, D., Nicolas, O., Bedard, J. & James, M. N. (2005). *J. Biol. Chem.* **280**, 18202–18210.

Biswal, B. K., Wang, M., Cherney, M. M., Chan, L., Yannopoulos, C. G., Bilimoria, D., Bedard, J. & James, M. N. (2006). *J. Mol. Biol.* **361**, 33–45.
Blicher, T., Kastrup, J. S., Buus, S. & Gajhede, M. (2005). *Acta Cryst.* D**61**, 1031–1040.
Boots, A. M., Van Lierop, M. J., Kusters, J. G., Van Kooten, P. J., Van der Zeijst, B. A. & Hensen, E. J. (1991). *Immunology*, **72**, 10–14.
Bosch, B. J., Martina, B. E., Van Der Zee, R., Lepault, J., Haijema, B. J., Versluis, C., Heck, A. J., De Groot, R., Osterhaus, A. D. & Rottier, P. J. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 8455–8460.
Bushell, M. & Sarnow, P. (2002). *J. Cell Biol.* **158**, 395–399.
Chen, C. N., Lin, C. P., Huang, K. K., Chen, W. C., Hsieh, H. P., Liang, P. H. & Hsu, J. T. (2005). *Evid. Based Complement. Alternat. Med.* **2**, 209–215.
Chen, L., Gui, C., Luo, X., Yang, Q., Gunther, S., Scandella, E., Drosten, C., Bai, D., He, X., Ludewig, B., Chen, J., Luo, H., Yang, Y., Yang, Y., Zou, J. *et al.* (2005). *J. Virol.* **79**, 7095–7103.
Chen, L., Li, J., Luo, C., Liu, H., Xu, W., Chen, G., Liew, O. W., Zhu, W., Puah, C. M., Shen, X. & Jiang, H. (2006). *Bioorg. Med. Chem.* **14**, 8295–8306.
Chen, L. R., Wang, Y. C., Lin, Y. W., Chou, S. Y., Chen, S. F., Liu, L. T., Wu, Y. T., Kuo, C. J., Chen, T. S. & Juang, S. H. (2005). *Bioorg. Med. Chem. Lett.* **15**, 3058–3062.
Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A., Berger, A., Burguiere, A. M., Cinatl, J., Eickmann, M., Escriou, N. *et al.* (2003). *N. Engl. J. Med.* **348**, 1967–1976.
Duquerroy, S., Vigouroux, A., Rottier, P. J., Rey, F. A. & Bosch, B. J. (2005). *Virology*, **335**, 276–285.
Eckert, D. M. & Kim, P. S. (2001). *Annu. Rev. Biochem.* **70**, 777–810.
Egloff, M. P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E. J., Gorbalenya, A. E., Cambillau, C. & Canard, B. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 3792–3796.
Egloff, M. P., Malet, H., Putics, A., Heinonen, M., Dutartre, H., Frangeul, A., Gruez, A., Campanacci, V., Cambillau, C., Ziebuhr, J., Ahola, T. & Canard, B. (2006). *J. Virol.* **80**, 8493–8502.
Gallagher, T. M. & Buchmeier, M. J. (2001). *Virology*, **279**, 371–374.
Gillim-Ross, L. & Subbarao, K. (2006). *Clin. Microbiol. Rev.* **19**, 614–636.
Groot, R. J. de, Luytjes, W., Horzinek, M. C., van der Zeijst, B. A., Spaan, W. J. & Lenstra, J. A. (1987). *J. Mol. Biol.* **196**, 963–966.
Hamill, P., Hudson, D., Kao, R. Y., Chow, P., Raj, M., Xu, H., Richer, M. J. & Jean, F. (2006). *Biol. Chem.* **387**, 1063–1074.
Hegyi, A. & Ziebuhr, J. (2002). *J. Gen. Virol.* **83**, 595–599.
Hoek, L. van der, Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J., Wolthers, K. C., Wertheim-van Dillen, P. M., Kaandorp, J., Spaargaren, J. & Berkhout, B. (2004). *Nat. Med.* **10**, 368–373.
Hsu, J. T., Kuo, C. J., Hsieh, H. P., Wang, Y. C., Huang, K. K., Lin, C. P., Huang, P. F., Chen, X. & Liang, P. H. (2004). *FEBS Lett.* **574**, 116–120.
Huang, C., Ito, N., Tseng, C. T. & Makino, S. (2006). *J. Virol.* **80**, 7287–7294.
Huang, Q., Yu, L., Petros, A. M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S. W. & Olejniczak, E. T. (2004). *Biochemistry*, **43**, 6059–6063.
Hwang, W. C., Lin, Y., Santelli, E., Sui, J., Jaroszewski, L., Stec, B., Farzan, M., Marasco, W. A. & Liddington, R. C. (2006). *J. Biol. Chem.* **281**, 34610–34616.
Imbert, I., Guillemot, J. C., Bourhis, J. M., Bussetta, C., Coutard, B., Egloff, M. P., Ferron, F., Gorbalenya, A. E. & Canard, B. (2006). *EMBO J.* **25**, 4933–4942.
Ito, N., Mossel, E. C., Narayanan, K., Popov, V. L., Huang, C., Inoue, T., Peters, C. J. & Makino, S. (2005). *J. Virol.* **79**, 3182–3186.
Jain, R. P., Pettersson, H. I., Zhang, J., Aull, K. D., Fortin, P. D., Huitema, C., Eltis, L. D., Parrish, J. C., James, M. N., Wishart, D. S. & Vederas, J. C. (2004). *J. Med. Chem.* **47**, 6113–6116.
Jonassen, C. M., Jonassen, T. O. & Grinde, B. (1998). *J. Gen. Virol.* **79**, 715–718.
Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Brooun, A., Griffith, M., Moy, K., Yadav, M. K., Velasquez, J., Buchmeier, M. J., Stevens, R. C. & Kuhn, P. (2006). *J. Virol.* **80**, 7894–7901.
Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C. & Kuhn, P. (2007). *J. Virol.* **81**, 6700–6708.
Kaeppler, U., Stiefl, N., Schiller, M., Vicik, R., Breuning, A., Schmitz, W., Rupprecht, D., Schmuck, C., Baumann, K., Ziebuhr, J. & Schirmeister, T. (2005). *J. Med. Chem.* **48**, 6832–6842.
Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., Kubo, H. & Makino, S. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 12885–12890.
Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J. *et al.* (2003). *N. Engl. J. Med.* **348**, 1953–1966.
Kuiken, T., Fouchier, R. A., Schutten, M., Rimmelzwaan, G. F., van Amerongen, G., van Riel, D., Laman, J. D., de Jong, T., van Doornum, G., Lim, W., Ling, A. E., Chan, P. K., Tam, J. S., Zambon, M. C., Gopal, R. *et al.* (2003). *Lancet*, **362**, 263–270.
Kumar, P., Gunalan, V., Liu, B., Chow, V. T., Druce, J., Birch, C., Catton, M., Fielding, B. C., Tan, Y. J. & Lal, S. K. (2007). *Virology*. In the press.
Lai, M. M. C. & Holmes, K. V. (2001). *Coronaviridae: the Viruses and their Replication*. Baltimore: Lippincott Williams and Wilkins.
Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Tsoi, H. W., Wong, B. H., Wong, S. S., Leung, S. Y., Chan, K. H. & Yuen, K. Y. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 14040–14045.
Li, F., Li, W., Farzan, M. & Harrison, S. C. (2005). *Science*, **309**, 1864–1868.
Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H. & Farzan, M. (2003). *Nature (London)*, **426**, 450–454.
Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B. T. *et al.* (2005). *Science*, **310**, 676–679.
Liu, S., Xiao, G., Chen, Y., He, Y., Niu, J., Escalante, C. R., Xiong, H., Farmar, J., Debnath, A. K., Tien, P. & Jiang, S. (2004). *Lancet*, **363**, 938–947.
Liu, Y. C., Huang, V., Chao, T. C., Hsiao, C. D., Lin, A., Chang, M. F. & Chow, L. P. (2005). *Biochem. Biophys. Res. Commun.* **333**, 194–199.
Lu, I. L., Mahindroo, N., Liang, P. H., Peng, Y. H., Kuo, C. J., Tsai, K. C., Hsieh, H. P., Chao, Y. S. & Wu, S. Y. (2006). *J. Med. Chem.* **49**, 5154–5161.
Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., Khattra, J., Asano, J. K., Barber, S. A., Chan, S. Y., Cloutier, A., Coughlin, S. M., Freeman, D., Girn, N., Griffith, O. L. *et al.* (2003). *Science*, **300**, 1399–1404.
Meier, C., Aricescu, A. R., Assenberg, R., Aplin, R. T., Gilbert, R. J. C., Grimes, J. M. & Stuart, D. I. (2006). *Structure*, **14**, 1157–1165.
Nelson, C. A., Pekosz, A., Lee, C. A., Diamond, M. S. & Fremont, D. H. (2005). *Structure (Camb.)*, **13**, 75–85.
Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W., Nicholls, J., Yee, W. K., Yan, W. W., Cheung, M. T., Cheng, V. C., Chan, K. H., Tsang, D. N., Yung, R. W., Ng, T. K. *et al.* (2003). *Lancet*, **361**, 1319–1325.
Prentice, E., McAuliffe, J., Lu, X., Subbarao, K. & Denison, M. R. (2004). *J. Virol.* **78**, 9977–9986.
Putics, A., Filipowicz, W., Hall, J., Gorbalenya, A. E. & Ziebuhr, J. (2005). *J. Virol.* **79**, 12721–12731.
Ratta, K., Saikatendu, K. S., Santarsiero, B. D., Barretto, N., Baker, S. C., Stevens, R. C. & Mesecar, A. D. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 5717.

Ricagno, S., Egloff, M. P., Ulferts, R., Coutard, B., Nurizzo, D., Campanacci, V., Cambillau, C., Ziebuhr, J. & Canard, B. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 11892–11897.
Robertson, M. P., Igel, H., Baertsch, R., Haussler, D., Ares, M. Jr & Scott, W. G. (2005). *PLoS Biol.* **3**, e5.
Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K., Chen, M. H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L. *et al.* (2003). *Science*, **300**, 1394–1399.
Saikatendu, K. S., Joseph, J. S., Subramanian, V., Clayton, T., Griffith, M., Moy, K., Velasquez, J., Neuman, B. W., Buchmeier, M. J., Stevens, R. C. & Kuhn, P. (2005). *Structure*, **13**, 1665–1675.
Saikatendu, K. S., Joseph, J. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C. & Kuhn, P. (2007). *J. Virol.* **81**, 3913–3921.
Shen, S., Lin, P. S., Chao, Y. C., Zhang, A., Yang, X., Lim, S. G., Hong, W. & Tan, Y. J. (2005). *Biochem. Biophys. Res. Commun.* **330**, 286–292.
Shi, J., Wei, Z. & Song, J. (2004). *J. Biol. Chem.* **279**, 24765–24773.
Shie, J. J., Fang, J. M., Kuo, C. J., Kuo, T. H., Liang, P. H., Huang, H. J., Yang, W. B., Lin, C. H., Chen, J. L., Wu, Y. T. & Wong, C. H. (2005). *J. Med. Chem.* **48**, 4469–4473.
Skehel, J. J. & Wiley, D. C. (2000). *Annu. Rev. Biochem.* **69**, 531–569.
Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J. & Gorbalenya, A. E. (2003). *J. Mol. Biol.* **331**, 991–1004.
Spaan, W. J. M. & Cavanagh, D. (2004). *Coronaviridae, Virus Taxonomy*, VIIIth Report of the ICTV, pp. 945–962. London: Elsevier–Academic Press.
Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X. C., Bartlam, M. & Rao, Z. (2006). *J. Virol.* **80**, 7902–7908.
Sui, J., Li, W., Murakami, A., Tamin, A., Matthews, L. J., Wong, S. K., Moore, M. J., Tallarico, A. S., Olurinde, M., Choe, H., Anderson, L. J., Bellini, W. J., Farzan, M. & Marasco, W. A. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 2536–2541.
Sui, J., Li, W., Roberts, A., Matthews, L. J., Murakami, A., Vogel, L., Wong, S. K., Subbarao, K., Farzan, M. & Marasco, W. A. (2005). *J. Virol.* **79**, 5900–5906.
Supekar, V. M., Bruckmann, C., Ingallinella, P., Bianchi, E., Pessi, A. & Carfi, A. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 17958–17963.
Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., Siddell, S., Poon, L. L., Diprose, J., Alderton, D., Walsh, M. *et al.* (2004). *Structure (Camb.)*, **12**, 341–353.
Tan, E. L., Ooi, E. E., Lin, C. Y., Tan, H. C., Ling, A. E., Lim, B. & Stanton, L. W. (2004). *Emerg. Infect. Dis.* **10**, 581–586.
Tan, Y. J. (2005). *Virol. J.* **2**, 5.
Tsai, K. C., Chen, S. Y., Liang, P. H., Lu, I. L., Mahindroo, N., Hsieh, H. P., Chao, Y. S., Liu, L., Liu, D., Lien, W., Lin, T. H. & Wu, S. Y. (2006). *J. Med. Chem.* **49**, 3485–3495.
Wei, P., Fan, K., Chen, H., Ma, L., Huang, C., Tan, L., Xi, D., Li, C., Liu, Y., Cao, A. & Lai, L. (2006). *Biochem. Biophys. Res. Commun.* **339**, 865–872.
Wen, C. C., Kuo, Y. H., Jan, J. T., Liang, P. H., Wang, S. Y., Liu, H. G., Lee, C. K., Chang, S. T., Kuo, C. J., Lee, S. S., Hou, C. C., Hsiao, P. W., Chien, S. C., Shyur, L. F. & Yang, N. S. (2007). *J. Med. Chem.* **50**, 4087–4095.
Wimberly, B. T., Brodersen, D. E., Clemons, W. M. Jr, Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). *Nature (London)*, **407**, 327–339.
Woo, P. C., Lau, S. K., Chu, C. M., Chan, K. H., Tsoi, H. W., Huang, Y., Wong, B. H., Poon, R. W., Cai, J. J., Luk, W. K., Poon, L. L., Wong, S. S., Guan, Y., Peiris, J. S. & Yuen, K. Y. (2005). *J. Virol.* **79**, 884–895.
Wu, C. Y., King, K. Y., Kuo, C. J., Fang, J. M., Wu, Y. T., Ho, M. Y., Liao, C. L., Shie, J. J., Liang, P. H. & Wong, C. H. (2006). *Chem. Biol.* **13**, 261–268.
Xu, X., Liu, Y., Weiss, S., Arnold, E., Sarafianos, S. G. & Ding, J. (2003). *Nucleic Acids Res.* **31**, 7117–7130.
Xu, X., Zhai, Y., Sun, F., Lou, Z., Su, D., Xu, Y., Zhang, R., Joachimiak, A., Zhang, X. C., Bartlam, M. & Rao, Z. (2006). *J. Virol.* **80**, 7909–7917.
Xu, Y., Liu, Y., Lou, Z., Qin, L., Li, X., Bai, Z., Pang, H., Tien, P., Gao, G. F. & Rao, Z. (2004). *J. Biol. Chem.* **279**, 30514–30522.
Xu, Y., Lou, Z., Liu, Y., Pang, H., Tien, P., Gao, G. F. & Rao, Z. (2004). *J. Biol. Chem.* **279**, 49414–49419.
Yang, H., Bartlam, M. & Rao, Z. (2007). *Curr. Pharm. Design*, **12**, 4573–4590.
Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., Hilgenfeld, R., Yuen, K. Y., Wong, L., Gao, G., Chen, S. *et al.* (2005). *PLoS Biol.* **3**, e324.
Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 13190–13195.
Yu, I. M., Oldham, M. L., Zhang, J. & Chen, J. (2006). *J. Biol. Chem.* **281**, 17134–17139.
Yuan, K., Yi, L., Chen, J., Qu, X., Qing, T., Rao, X., Jiang, P., Hu, J., Xiong, Z., Nie, Y., Shi, X., Wang, W., Ling, C., Yin, X., Fan, K. *et al.* (2004). *Biochem. Biophys. Res. Commun.* **319**, 746–752.
Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M. & Rao, Z. (2005). *Nature Struct. Mol. Biol.* **12**, 980–986.
Ziebuhr, J. (2004). *Curr. Opin. Microbiol.* **7**, 412–419.
Ziebuhr, J. (2005). *Curr. Top. Microbiol. Immunol.* **287**, 57–94.
Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. (2000). *J. Gen. Virol.* **81**, 853–879.
Zust, R., Cervantes-Barragan, L., Kuri, T., Blakqori, G., Weber, F., Ludewig, B. & Thiel, V. (2007). *PLoS Pathog.* **3**, e109.