Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

ELSEVIER

Structural insights into SARS coronavirus proteins

Mark Bartlam, Haitao Yang and Zihe Rao

The SARS coronavirus was identified as the pathogen of a global outbreak of SARS (severe acute respiratory syndrome) in 2003. Its large RNA genome encodes four structural proteins, sixteen non-structural proteins and eight accessory proteins. The availability of structures of SARS coronavirus macromolecules has enabled the elucidation of their important functions, such as mediating the fusion of viral and host cellular membranes, and in replication and transcription. In particular, the spike protein fusion core and the main protease have been the most extensively studied, with the aim of designing anti-SARS therapeutics. Attention is now being focused on replicase proteins, which should enhance our understanding of the replication and transcription machinery. The structures and functions of most SARS proteins remain unknown, and further structural studies will be important for revealing their functions and for designing potential anti-SARS therapeutics.

Addresses
National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China and Laboratory of Structural Biology, Tsinghua University, Beijing 100084, China

Corresponding author: Rao, Zihe (raozh@xtal.tsinghua.edu.cn)

Current Opinion in Structural Biology 2005, **15**:664–672

This review comes from a themed issue on
Proteins
Edited by Edward N Naker and Guy G Dodson

Available online 2nd November 2005

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2005.10.004

Introduction

In 2003, a previously unknown coronavirus termed SARS-CoV was identified as the causative agent of severe acute respiratory syndrome (SARS), responsible for a worldwide epidemic with approximately 800 deaths [1–4]. The most likely explanation for the emergence of SARS-CoV is animal-to-human interspecies transmission [5]. However, the animal reservoir for SARS-CoV in nature remains to be identified and the mechanism of viral adaptation to human hosts requires further investigation.

SARS-CoV is a plus-strand RNA virus featuring a large single-stranded RNA genome of approximately 29 700 nucleotides [6,7]. The genome is predicted to consist of at least fourteen functional ORFs that encode three classes of proteins: two large polyproteins, pp1a and pp1ab, which are cleaved into sixteen non-structural proteins (nsps) required for viral RNA synthesis (and probably other functions); four structural proteins (the S, E, M and N proteins), essential for viral assembly; and eight accessory proteins, which are thought unimportant in tissue culture but may provide a selective advantage in the infected host (Table 1) [8].

In this review, recent structural studies of SARS-CoV macromolecules (including a conserved RNA motif) are summarised, focusing on those proteins that mediate the fusion of the viral membrane with the host cell membrane, or that are involved in coronavirus genome replication and transcription. The latter have been extensively studied, with the aim of designing anti-SARS therapeutics.

Structural proteins

The SARS coronavirus includes 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 (Table 1) [6,7]. Here, we will focus on the S protein and N protein, whose partial structures have been solved.

SARS spike protein fusion core

Similar to other class I virus fusion proteins, the SARS-CoV S protein can be subdivided into an N-terminal half (S1) and C-terminal half (S2), but without proteolytic cleavage [9••]. S1 is responsible for variations in host range and tissue tropism according to its receptor specificity, whereas S2 is responsible for cell entry following virus and host cell membrane fusion [10]. S1 is responsible for binding to cellular receptors; one potential SARS-CoV receptor has been identified as angiotensin-converting enzyme 2 (ACE2) [11]. S2 contains an internal fusion peptide and has two hydrophobic (heptad) repeat regions, designated HR1 and HR2 [12]. The putative fusion peptide has recently been identified upstream of and near to HR1 [13]. HR2 is located close to the transmembrane region, some 170 amino acids downstream of HR1 [12]. The classical mechanism of enveloped virus and host cell membrane fusion mediated by class I fusion proteins was established by Wiley and colleagues in their comprehensive study of influenza hemagglutinin (HA), with structures of the unprocessed precursor, the cleaved metastable HA1–HA2 heterodimer and post-fusion conformations available [14,15]. In the following years, extensive structural studies of the orthomyxovirus, retrovirus, paramyxovirus and filovirus families have led to a common fusion mechanism [15]. To confirm the value of fusion proteins as anti-viral targets, an HIV-1 membrane fusion inhibitory peptide,

Table 1

Summary of SARS proteins.

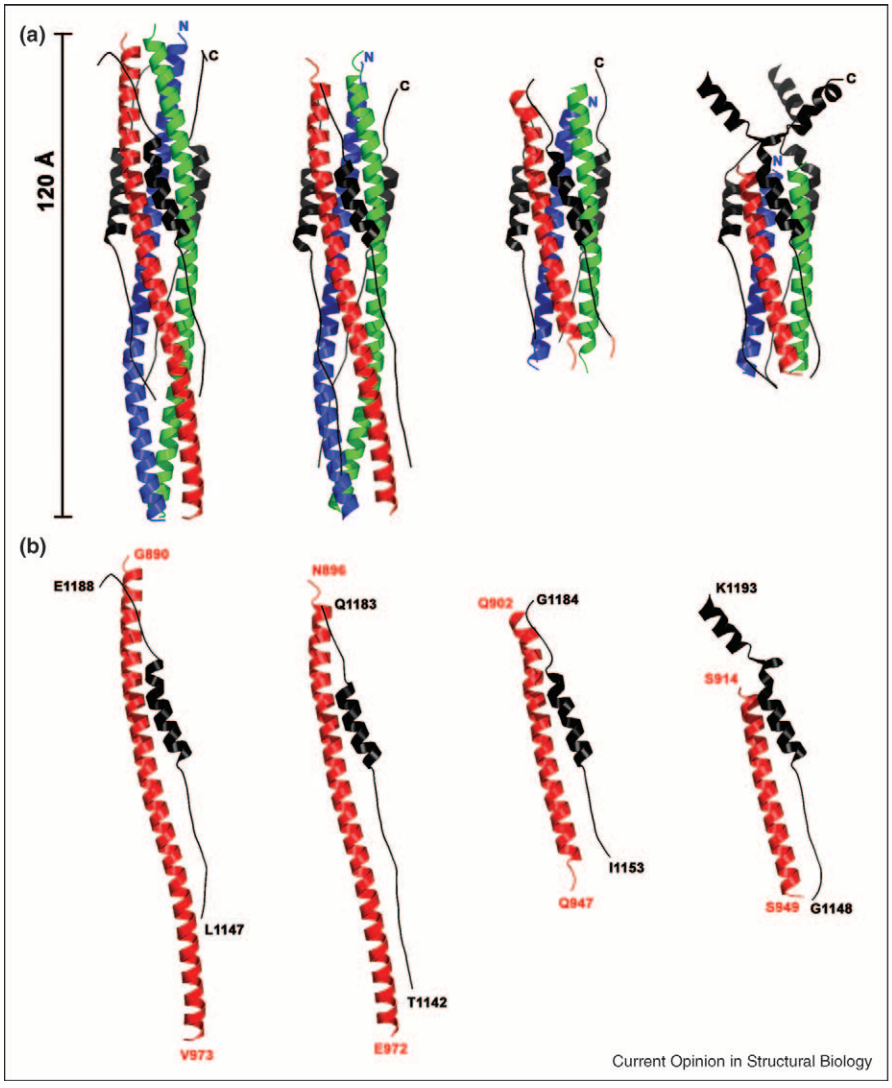
Protein	Protein size (amino acids)	ORF (location in genome sequence)	Putative functional domain(s)	Structure available?
Structural proteins				
Spike (S) protein	1255	ORF2 (21492–25259)		Yes (S protein fusion core)
Envelope (E) protein	76	ORF4 (26117–26347)		No
Membrane (M) protein	221	ORF5 (26398–27063)		No
Nucleocapsid (N) protein	422	ORF9a (28120–29388)		Yes (N-terminal RNA-binding domain)
Non-structural proteins (Nsps)				
Nsp1	180	ORF1a (265–804)		No
Nsp2	638	ORF1a (805–2718)		No
Nsp3	1922	ORF1a (2719–8484)	Ac, X, $PL2^{pro}$, Y (TM1), ADRP	No
Nsp4	500	ORF1a (8485–9984)	TM2	No
Nsp5	306	ORF1a (9985–10902)	M^{pro}	Yes
Nsp6	290	ORF1a (10903–11772)	TM3	No
Nsp7	83	ORF1a (11773–12021)		Yes[a]
Nsp8	198	ORF1a (12022–12615)		Yes[a]
Nsp9	113	ORF1a (12616–12954)	ssRNA binding	Yes
Nsp10	139	ORF1a (12955–13371)	GFL	No
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 homologue)	No
Nsp15	346	ORF1b (19551–20588)	NTD, endoribonuclease (XendoU homologue)	No
Nsp16	298	ORF1b (20589–21482)	2′-*O*-MT	No
Accessory proteins				
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)		No

[a] Structure has been deposited in the PDB, but has not been published. Ac, acidic domain; ADRP, adenosine diphosphate-ribose 1′-phosphatase; ExoN, 3′-5′ exonuclease; GFL, growth-factor-like domain; HEL1, superfamily 1 helicase; M^{pro}, main (or 3C-like cysteine) protease; NTD, nidovirus conserved domain; NTPase, nucleoside triphosphatase; 2′-*O*-MT, *S*-adenosylmethionine-dependent ribose 2′-*O*-methyltransferase; $PL2^{pro}$, papain-like protease 2; RdRp, RNA-dependent RNA polymerase; TM, transmembrane domain; X, Y, domains with unknown or hypothetical function; ZD, putative zinc-binding domain.

T-20 (developed by Trimeris, Research Triangle Park, North Carolina, USA), which targets the prehairpin intermediate, was recently approved by the US Food and Drug Administration as a new anti-HIV drug [16].

In 2004, the structure of the S protein fusion core, consisting of the HR1 and HR2 regions, was determined by two groups in the post-fusion (or fusion-active) state [9••,17•] (Figure 1). Xu and co-workers [17•] constructed a single chain by engineering a linker between the HR1 and HR2 regions to prepare the fusion core (HR1: amino acids 900–948; HR2: amino acids 1145–1184), whereas Supekar and colleagues [9••] individually synthesized longer HR1 and HR2 peptides to prepare the complex (HR1: amino acids 889–972; HR2: amino acids 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. HR2 peptides pack into the hydrophobic grooves of the HR1 trimer in a mixed extended and helical conformation; this represents a stable post-fusion structure, similar to that observed for HIV-1 gp41 [15]. 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 *et al.* [9••] also provided the structure of an S2 fragment consisting of a smaller HR1 peptide (amino acids 919–949) and an HR2 peptide with extra C-terminal residues in proximity to the transmembrane region (amino acids 1149–1193) (Figure 1). The C-terminal part is α helical and points away from the HR1 trimer axis, probably due to the lack of stabilisation by the corresponding HR1 region. The authors consider that this could mimic the conformation of this region before the

Figure 1



The SARS-CoV S protein fusion core. **(a)** Comparison of four 'six-helix bundle' structures. Shown from left to right are S protein fusion cores 1WYY [18], 2BEZ [17•], 1WNC [9••] and 2BEQ [9••]. The central HR1 peptides are shown in ribbon representation, and are coloured red, blue and green. The HR2 peptides are shown in black. The N and C termini are labelled. **(b)** Comparison of four 'HR1+HR2' constructs, corresponding to the structures in (a). The labelled residues correspond to the start and end residues of the HR1 (red) and HR2 (black) peptides. (Figure adapted from [18].)

formation of the final post-fusion hairpins. A later structure was reported by Duquerroy and colleagues [18] (HR1: amino acids 890–973; HR2: amino acids 1145–1190) (Figure 1), in which they emphasized the hydrogen-bonding network formed by conserved asparagine and glutamine residues, together with two possible chlorides, which could stabilise the conformation of post-fusion hairpins.

Fusogenic mechanisms similar to those of other class I fusion proteins have been proposed for SARS-CoV [16•,18,19•]. However, understanding the possible conformational changes of the fusion peptide, HR1 and HR2 during the membrane fusion process needs further structural studies of the native state of the S protein and the prehairpin intermediate that probably results from S1 binding to a receptor (e.g. ACE2).

Several peptides derived from HR1 and HR2 regions of SARS-CoV S proteins have been synthesized to block viral entry, targeting the putative prehairpin intermediate [17•,20,21]. Two groups discovered that only peptides derived from HR2, and not from HR1, inhibited SARS-CoV infection [17•,20]. Moreover, the efficacy of HR2 peptides derived from SARS-CoV S protein is lower than that of corresponding HR2 peptides derived from murine coronavirus mouse hepatitis virus (MHV) in inhibiting MHV infection [20]. Supekar and colleagues considered

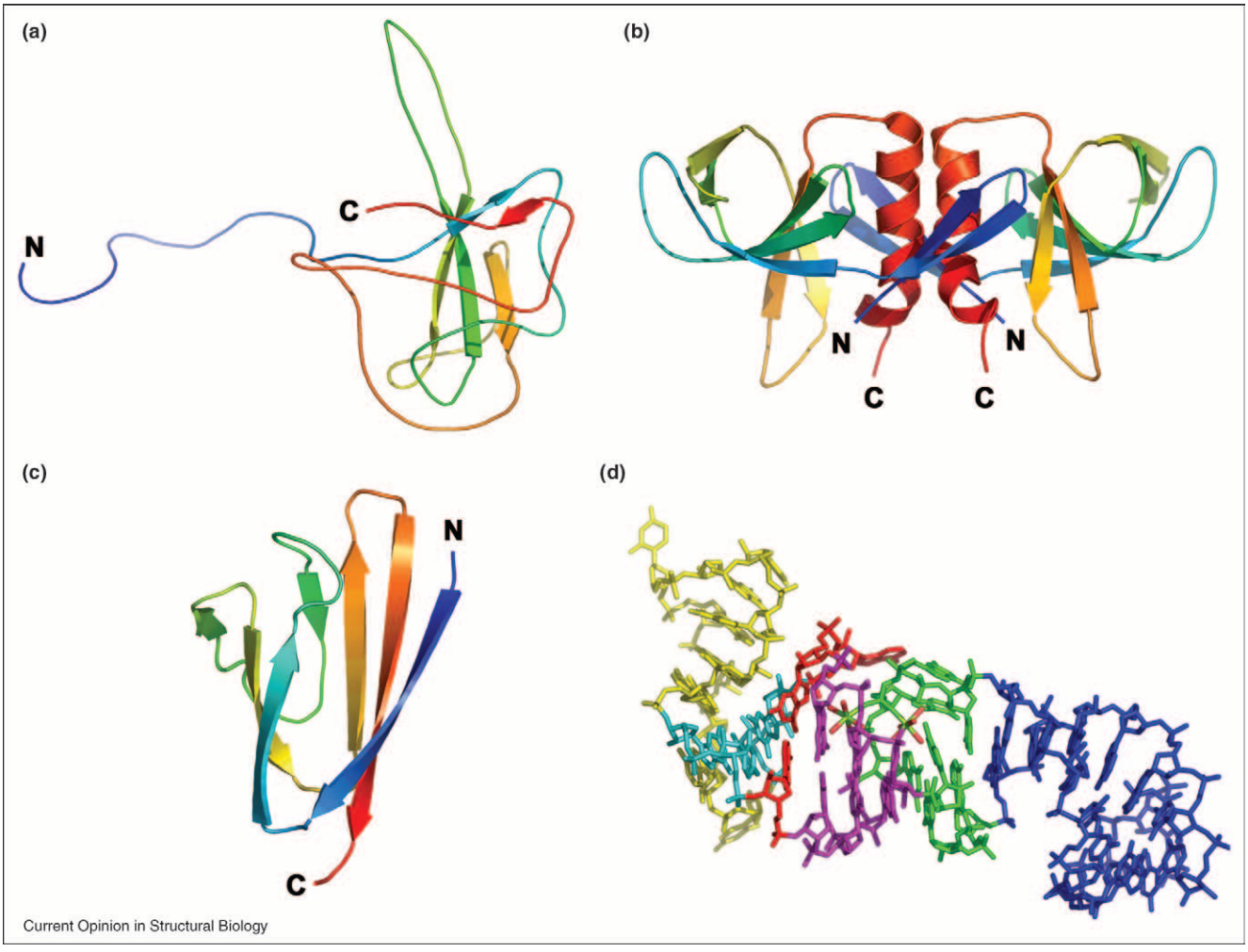
that this might be due to the lower affinity of these peptides for the corresponding HR1 trimer [20], as a larger surface area is buried in the HR1–HR2 interface of MHV S2 than in SARS-CoV S2 [9••]. In any case, determination of the HR1–HR2 fusion core structure will help in the discovery of viral entry inhibitors against SARS.

SARS nucleocapsid protein RNA-binding domain

The N protein, which binds to the genomic RNA via a leader sequence, recognises a stretch of RNA that serves as a packaging signal and leads to the formation of the helical ribonucleoprotein (RNP) complex during assembly [22]. The structure of the RNA-binding domain of the SARS-CoV N protein was determined by NMR spectroscopy in 2004 [23]. It consists of a five-stranded β sheet whose fold is unrelated to that of other RNA-binding proteins (Figure 2a). The authors identified a binding site for single-stranded RNA (ssRNA), using NMR to determine the resonance of residues perturbed by the addition of RNA, and revealed a similar mode of interaction to RNA-binding proteins such as U1A RNP. They also identified small molecules from an NMR-based screen that bind to the RNA-binding domain and might impair its function.

Antigenic peptides of the coronavirus N protein can be recognised by T cells on the surface of infected cells [24,25]. 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 a SARS-specific sequence, has recently been determined to 1.45 Å resolution [26]. 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.

Figure 2



Other structures of SARS-CoV proteins. **(a)** Solution structure of the N-terminal RNA-binding domain of the SARS-CoV N protein (PDB code 1SSK). **(b)** X-ray crystal structure of nsp9, an ssRNA-binding protein (PDB code 1UW7). **(c)** X-ray crystal structure of the accessory protein Orf7a (PDB code 1XAK). **(d)** X-ray crystal structure of the s2m, a rigorously conserved RNA element of the SARS-CoV genome.

Non-structural proteins

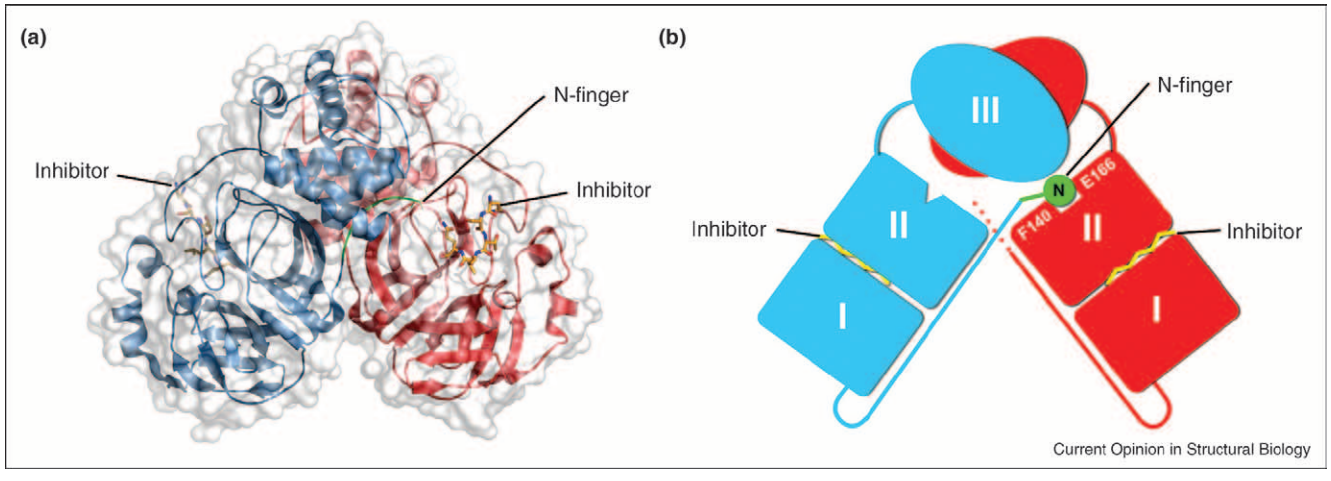
The SARS-CoV replicase gene encodes 16 nsps with multiple enzymatic functions (Table 1) [27]. These are known or are predicted to include types of enzymes that are common components of the replication machinery of plus-strand RNA viruses: an RNA-dependent RNA polymerase activity (RdRp, nsp12); a 3C-like serine protease activity (M^{pro} or $3CL^{pro}$, nsp5); a papain-like protease 2 activity ($PL2^{pro}$, nsp3); and a superfamily-1 helicase activity (HEL1, nsp13) [6,28,29••]. In addition, the replicase gene encodes proteins that are indicative of 3′-5′ exoribonuclease activity (ExoN homologue, nsp14), endoribonuclease activity (XendoU homologue, nsp15), adenosine diphosphate-ribose 1′-phosphatase activity (ADRP, nsp3) and ribose 2′-*O*-methyltransferase activity (2′-*O*-MT, nsp16) [27]. These enzymes are less common in plus-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. The nsps 4, 10 and 16 have been implicated by genetic analysis in the assembly of a functional replicase–transcriptase complex. Here, we detail two nsps with available structures, of which nsp5 is the most extensively characterised.

Nsp5: the main protease (M^{pro} or $3CL^{pro}$) is a target for anti-viral drug design

The replicase polyproteins, pp1a and pp1ab, undergo extensive proteolytic processing by viral proteases to produce multiple functional subunits, which are involved in the formation of the replicase complex that mediates viral replication and transcription. The coronavirus main protease (M^{pro}), also known as the 3C-like protease ($3CL^{pro}$) after the 3C proteases of the *Picornaviridae*, is an ≈33 kDa cysteine protease that cleaves the replicase polyprotein at 11 conserved sites with canonical Leu-Gln↓(Ser, Ala, Gly) sequences. The cleavage process is initiated by the enzyme's own autolytic cleavage from pp1a and pp1ab [30,31]. Its functional importance in the viral life cycle and the lack of closely related cellular homologues make M^{pro} an attractive target for the development of drugs directed not only against SARS but also against other coronavirus infections [29••,30–32].

The first structural models of SARS M^{pro} were homology models built from the crystal structures of M^{pro} from human coronavirus strain 229E (HCoV-229E) and porcine transmissible gastroenteritis virus (TGEV), both group I coronaviruses. These homology models were widely used in the design of anti-SARS inhibitors [32,33]. In 2003, shortly after the peak of the SARS outbreak, Yang *et al.* reported the first crystal structure of SARS-CoV M^{pro} to 1.9 Å resolution, which provided important structural information for rational drug design (Figure 3) [29••]. CoV M^{pro} forms a dimer and each protomer consists of three domains: domains I and II resemble chymotrypsin, whereas domain III has a globular cluster of five mostly antiparallel α helices. The cleft between domains I and II is the location of substrate recognition and catalysis. Domain III and the additional N-terminal finger of domain I (or 'N-finger') are considered to influence the dimerisation and activity of M^{pro} [29••,34–36] (Figure 3). However, one group has reported incongruent results concerning the role of the N-finger in dimerisation [37]. In contrast to common serine proteases, which have a catalytic triad, CoV M^{pro} has a Cys-His catalytic dyad. An exceptionally stable water

Figure 3



Nsp5, the SARS-CoV M^{pro}. **(a)** The crystal structure of SARS-CoV M^{pro} in complex with a CMK inhibitor (PDB code 1UKW). Protomers A and B are shown in ribbon representation, and are coloured red and blue, respectively. The CMK inhibitors are shown in yellow stick representation. The N-finger, residues 1–7 of protomer B, is shown in green. A transparent molecular surface is shown covering the structure. **(b)** Schematic of the SARS-CoV M^{pro} dimer, corresponding to the view in (a). Residue S1 on the N-finger of protomer B forms hydrogen bonds with two residues in protomer A, F140 and E166.

molecule occupies the position of the usual third member of the triad, which might stabilise the protonated histidine in the intermediate state during proteolytic cleavage.

As a prelude to inhibitor design, the structure of SARS-CoV M^{pro} in complex with a substrate analogue (a chloromethyl ketone [CMK] inhibitor, Cbz-VNSTLQ-CMK) was determined in 2003 (Figure 3). The sequence of this substrate analogue was derived from residues P6–P1 of the N-terminal autoprocessing site of TGEV M^{pro} [32]. However, the two protomers of SARS-CoV M^{pro} each exhibited an unexpected binding mode. This most probably resulted from the comparatively weak binding of peptidyl elements derived from the substrate of TGEV M^{pro} and from the highly reactive electrophile CMK, suggesting that nucleophilic attack might have occurred before a stable non-covalently bound enzyme–inhibitor complex was formed [38••].

Following the SARS outbreak, a series of potential inhibitors against SARS-CoV M^{pro} was reported, some of which could prevent viral replication *in vitro* [39–43]. However, complex structures are rarely available to guide further modification of these compounds. A recent study of representative structures from all three groups of the genus *Coronavirus* has indicated that all CoV M^{pro} share a highly conservative substrate-recognition pocket [38••]. Mechanism-based irreversible inhibitors were designed based on this conserved structural region, and further modification of these compounds could possibly lead to the discovery of a single agent with clinical potential against existing and possible future emerging CoV-related diseases [38••].

Nsp9: an ssRNA-binding protein

Crystal structures of nsp9, determined simultaneously in 2004 by Egloff *et al.* (to 2.7 Å resolution) [44••] and Sutton *et al.* (to 2.8 Å resolution) [45••], have established its previously unknown function as an ssRNA-binding protein. Both groups report that the biological unit is a dimer. The core structure of the protein is an open six-stranded β barrel reminiscent of, although unrelated to, the nucleic acid binding OB (oligosaccharide/oligonucleotide-binding) fold (Figure 2b). Instead, nsp9 is structurally homologous to certain subdomains of serine proteases, most notably domain II of SARS-CoV M^{pro}. Based on this similarity to the picornavirus 3C proteases, which feature a conserved RNA-binding motif, it was inferred that nsp9 should bind ssRNA; this was subsequently confirmed by electrophoretic mobility shift assays (EMSAs) [45••] and surface plasmon resonance [44••]. One role of nsp9 may be to stabilise nascent and template RNA strands during replication and transcription, and to protect them against nuclease processing. Besides replication, nsp9 may also be involved in base-pairing-driven processes, such as RNA processing.

In addition to their nsp9 structure, Sutton and colleagues showed evidence of its interaction with nsp8 [45••]. Furthermore, dual-labelling studies of SARS-CoV replicase proteins have demonstrated co-localisation of nsp8 with nsp2 and nsp3 [46], and an interaction between nsp7 and nsp8 has also been found (Z Rao, unpublished; see Update), suggesting that the nsps assemble to form a sophisticated viral replication/transcription machinery. Nsp9 is the first component of the complex with an available three-dimensional structure, providing a starting point to reveal the architecture and underlying functions of the replication/transcription complex.

Accessory proteins

The genomic sequences of numerous SARS-CoV isolates have been determined. The 'conserved' open reading frames (ORFs) of the SARS-CoV genome occur in the same order as and are of similar size to those found in other coronaviruses. However, in addition to the conserved genes, the SARS-CoV genome contains eight novel ORFs at the 3′ end (ORFs 3a, 3b, 6, 7a, 7b, 8a, 8b and 9b) (Table 1) [27]. To date, the functions of these genes remain largely unknown, although their absence from other genomes suggests unique functions that might be advantageous to SARS-CoV replication, assembly or virulence [8]. Only one of these so-called accessory proteins has a known structure and further studies are required to elucidate their precise functions.

The Orf7a accessory protein

Sequence analysis predicted that ORF 7a encodes a type I transmembrane protein of 122 amino acids, consisting of a 15-residue N-terminal signal peptide, an 81-residue luminal domain, a 21-residue transmembrane segment and a 5-residue cytoplasmic tail [27]. The Orf7a sequence has been identified in all isolates of SARS-CoV collected from both human and animal sources, but it appears to be unique to SARS, with no significant similarity to any other viral or non-viral protein. The structure of the luminal domain of the Orf7a accessory protein was determined earlier this year to 1.8 Å resolution. It reveals a compact Ig-like domain with a β-sandwich fold topology (Figure 2c), despite its unusually small size and lack of significant sequence similarity to other members of the Ig superfamily [47•]. This common structural fold occurs in a wide variety of proteins, where it performs a diverse set of functions, making it difficult to predict the functional role of Orf7a from the structure alone. For example, the fold is found in proteins of the extracellular matrix, muscle proteins, proteins of the immune system, cell surface receptors, enzymes, transcription factors and a wide variety of viral proteins [48].

Other structures

The crystal structure of the stem-loop II motif (s2m) RNA element of SARS-CoV was determined in 2005 to 2.7 Å resolution [49••]. s2m is a rigorously conserved motif

located at the 3′ end of SARS and other coronaviruses, as well as astroviruses [50]. The highly structured s2m RNA element includes a remarkable 90° bend of the helix axis (Figure 2d). Several novel longer-range tertiary interactions create a tunnel perpendicular to the main helical axis, whose interior is negatively charged and binds two magnesium ions. These unusual features form probable surfaces for interaction with conserved host cell components or other reactive sites required for virus function. An interesting observation is the possible mimicry by s2m RNA of an rRNA fold, the 530 loop of 16S rRNA [51]. This implies a mechanism for RNA hijacking of host protein synthesis in SARS, similar to that observed in other RNA viruses [52]. The 530 loop of the 30S ribosome binds to prokaryotic proteins S12 and IF-1, further suggesting that s2m may interact with their eukaryotic homologues [49••]. Nevertheless, the high sequence conservation of s2m in an otherwise rapidly mutable RNA genome implies its pathogenic importance and signals that it could be another attractive target for the design of anti-viral therapeutics.

Conclusions

The rapid growth in the availability of SARS-CoV protein structures since the 2003 outbreak has emphasized the importance and strength of structural biology as a tool to address significant health-related issues, including functional annotation of proteins and identification of important drug targets. An important wealth of information and clues for further study have been accumulated from the SARS-CoV macromolecular structures determined so far. The first SARS-CoV structure to be determined, M^{pro}, is an important target for drug design and has been widely used since 2003 as a basis for inhibitor design, with promising results. Similarly, the S protein fusion core has also been confirmed as an important drug target for the design of fusion inhibitor peptides. Future prospects for SARS structural biology include the structures of replicase proteins alone and in protein–protein complexes, with the aim of understanding the sophisticated function and assembly of the replication/transcription machinery, as well as the characterisation of the structural interaction between the SARS-CoV S protein and its possible cellular receptors, for instance, ACE2.

Update

A number of structures of SARS coronavirus proteins have recently been published, including the structure of the SARS coronavirus spike receptor-binding domain (RBD) in complex with the receptor ACE2, determined by Harrison and colleagues [53••]. The authors reveal that the interface between the two proteins shows important residue changes that facilitate efficient cross-species infection and human-to-human transmission, and suggest ways to make truncated disulfide-stabilised RBD variants for use in the design of coronavirus vaccines.

The work referred to in the text as (Z Rao, unpublished) is now in press [54••]. The crystal structure of the hexameric complex between nsp7 and nsp8 to 2.4 Å resolution provides the first insight into the sophisticated architecture of the replication and transcription machinery. The supercomplex is a unique, hollow, cylinder-like structure assembled from eight copies of nsp8 and held tightly together by eight copies of nsp7. The central channel has dimensions and positive electrostatic properties favourable for nucleic acid binding, implying that its role is to confer processivity on RdRp. The structure of nsp7 has also been determined in the free unbound form by NMR [55].

We are also aware that the crystal structure of the ADRP domain of nsp3 has been determined and is currently in press [56].

Acknowledgements

This work was supported by projects 973 and 863 of the Ministry of Science and Technology of China (grant numbers 200BA711A12, G199075600), the National Natural Science Foundation of China (grant numbers 30221003, 20342002, 20321202), 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 and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, Nicholls J, Yee WK, Yan WW, Cheung MT *et al.*: **Coronavirus as a possible cause of severe acute respiratory syndrome**. *Lancet* 2003, **361**:1319-1325.

2. 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.*: **Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome**. *Lancet* 2003, **362**:263-270.

3. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W *et al.*: **A novel coronavirus associated with severe acute respiratory syndrome**. *N Engl J Med* 2003, **348**:1953-1966.

4. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, Rabenau H, Panning M, Kolesnikova L, Fouchier RA *et al.*: **Identification of a novel coronavirus in patients with severe acute respiratory syndrome**. *N Engl J Med* 2003, **348**:1967-1976.

5. Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ *et al.*: **Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China**. *Science* 2003, **302**:276-278.

6. Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattra J, Asano JK, Barber SA, Chan SY *et al.*: **The genome sequence of the SARS-associated coronavirus**. *Science* 2003, **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.*: **Characterization of a novel coronavirus associated with severe acute respiratory syndrome**. *Science* 2003, **300**:1394-1399.

8. Ziebuhr J: **Molecular biology of severe acute respiratory syndrome coronavirus**. *Curr Opin Microbiol* 2004, **7**:412-419.

9. •• Supekar VM, Bruckmann C, Ingallinella P, Bianchi E, Pessi A, Carfi A: **Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein**. *Proc Natl Acad Sci USA* 2004, **101**:17958-17963.
The authors reported the structure of the HR1– HR2 complex, representing the S protein fusion core of SARS-CoV, to a high resolution of 1.6 Å. The authors also presented a second structure comprising shorter HR1 peptides and HR2 peptides with extra residues in proximity to the transmembrane region; this mimics the conformation before the formation of the final post-fusion hairpins. These structures revealed targets for the design of viral entry inhibitors.

10. Gallagher TM, Buchmeier MJ: **Coronavirus spike proteins in viral entry and pathogenesis**. *Virology* 2001, **279**:371-374.

11. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, Luzuriaga K, Greenough TC *et al.*: **Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus**. *Nature* 2003, **426**:450-454.

12. de Groot RJ, Luytjes W, Horzinek MC, van der Zeijst BA, Spaan WJ, Lenstra JA: **Evidence for a coiled-coil structure in the spike proteins of coronaviruses**. *J Mol Biol* 1987, **196**:963-966.

13. Sainz B Jr, Rausch JM, Gallaher WR, Garry RF, Wimley WC: **Identification and characterization of the putative fusion peptide of the severe acute respiratory syndrome-associated coronavirus spike protein**. *J Virol* 2005, **79**:7195-7206.

14. Skehel JJ, Wiley DC: **Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin**. *Annu Rev Biochem* 2000, **69**:531-569.

15. Eckert DM, Kim PS: **Mechanisms of viral membrane fusion and its inhibition**. *Annu Rev Biochem* 2001, **70**:777-810.

16. Liu S, Xiao G, Chen Y, He Y, Niu J, Escalante CR, Xiong H, Farmar J, Debnath AK, Tien P *et al.*: **Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors**. *Lancet* 2004, **363**:938-947.

17. • Xu Y, Lou Z, Liu Y, Pang H, Tien P, Gao GF, Rao Z: **Crystal structure of severe acute respiratory syndrome coronavirus spike protein fusion core**. *J Biol Chem* 2004, **279**:49414-49419.
Using a single chain by engineering a linker between HR1 and HR2 to prepare the fusion core, the authors determined the crystal structure of the SARS S protein fusion core to 2.8 Å resolution.

18. Duquerroy S, Vigouroux A, Rottier PJ, Rey FA, Bosch BJ: **Central ions and lateral asparagine/glutamine zippers stabilize the post-fusion hairpin conformation of the SARS coronavirus spike glycoprotein**. *Virology* 2005, **335**:276-285.

19. • Xu Y, Liu Y, Lou Z, Qin L, Li X, Bai Z, Pang H, Tien P, Gao GF, Rao Z: **Structural basis for coronavirus-mediated membrane fusion. Crystal structure of mouse hepatitis virus spike protein fusion core**. *J Biol Chem* 2004, **279**:30514-30522.
The first reported structure of a coronavirus S protein fusion core was determined for MHV. On the basis of the structure, the authors propose a mechanism for coronavirus-mediated membrane fusion.

20. Bosch BJ, Martina BE, Van Der Zee R, Lepault J, Haijema BJ, Versluis C, Heck AJ, De Groot R, Osterhaus AD, Rottier PJ: **Severe acute respiratory syndrome coronavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides**. *Proc Natl Acad Sci USA* 2004, **101**:8455-8460.

21. Yuan K, Yi L, Chen J, Qu X, Qing T, Rao X, Jiang P, Hu J, Xiong Z, Nie Y *et al.*: **Suppression of SARS-CoV entry by peptides corresponding to heptad regions on spike glycoprotein**. *Biochem Biophys Res Commun* 2004, **319**:746-752.

22. Lai MM, Cavanagh D: **The molecular biology of coronaviruses**. *Adv Virus Res* 1997, **48**:1-100.

23. Huang Q, Yu L, Petros AM, Gunasekera A, Liu Z, Xu N, Hajduk P, Mack J, Fesik SW, Olejniczak ET: **Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein**. *Biochemistry* 2004, **43**:6059-6063.

24. Boots AM, Kusters JG, van Noort JM, Zwaagstra KA, Rijke E, van der Zeijst BA, Hensen EJ: **Localization of a T-cell epitope within the nucleocapsid protein of avian coronavirus**. *Immunology* 1991, **74**:8-13.

25. Bergmann C, McMillan M, Stohlman S: **Characterization of the Ld-restricted cytotoxic T-lymphocyte epitope in the mouse hepatitis virus nucleocapsid protein**. *J Virol* 1993, **67**:7041-7049.

26. Blicher T, Kastrup JS, Buus S, Gajhede M: **High-resolution structure of HLA-A*1101 in complex with SARS nucleocapsid peptide**. *Acta Crystallogr D Biol Crystallogr* 2005, **61**:1031-1040.

27. Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Guan Y, Rozanov M, Spaan WJ, Gorbalenya AE: **Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage**. *J Mol Biol* 2003, **331**:991-1004.

28. Thiel V, Ivanov KA, Putics A, Hertzig T, Schelle B, Bayer S, Weissbrich B, Snijder EJ, Rabenau H, Doerr HW *et al.*: **Mechanisms and enzymes involved in SARS coronavirus genome expression**. *J Gen Virol* 2003, **84**:2305-2315.

29. •• Yang H, Yang M, Ding Y, Liu Y, Lou Z, Zhou Z, Sun L, Mo L, Ye S, Pang H *et al.*: **The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor**. *Proc Natl Acad Sci USA* 2003, **100**:13190-13195.
The first reported structure of any protein from the SARS coronavirus. The authors describe the three-dimensional structure of the SARS main protease, the key enzyme mediating viral replication and transcription. SARS-CoV main protease is the most popularly studied target in the design of anti-SARS drugs and its structure provides important information for rational drug design.

30. Ziebuhr J, Snijder EJ, Gorbalenya AE: **Virus-encoded proteinases and proteolytic processing in the Nidovirales**. *J Gen Virol* 2000, **81**:853-879.

31. Ziebuhr J: **The coronavirus replicase**. *Curr Top Microbiol Immunol* 2005, **287**:57-94.

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

33. Anand K, Palm GJ, Mesters JR, Siddell SG, Ziebuhr J, Hilgenfeld R: **Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain**. *EMBO J* 2002, **21**:3213-3224.

34. Chou CY, Chang HC, Hsu WC, Lin TZ, Lin CH, Chang GG: **Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease**. *Biochemistry* 2004, **43**:14958-14970.

35. Hsu WC, Chang HC, Chou CY, Tsai PJ, Lin PI, Chang GG: **Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease**. *J Biol Chem* 2005, **280**:22741-22748.

36. Shi J, Wei Z, Song J: **Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors**. *J Biol Chem* 2004, **279**:24765-24773.

37. Chen S, Chen L, Tan J, Chen J, Du L, Sun T, Shen J, Chen K, Jiang H, Shen X: **Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations**. *J Biol Chem* 2005, **280**:164-173.

38. •• Yang H, Xie W, Xue X, Yang K, Ma J, Liang W, Zhao Q, Zhou Z, Pei D, Ziebuhr J *et al.*: **Design of wide spectrum inhibitors targeting coronavirus main proteases**. *PLoS Biol* 2005, **3**:e324.
In this paper, a strategy for preventing infection by existing and possible future emerging coronaviruses is presented. The authors discovered that all coronaviruses (about 25 species) share a highly conservative substrate-recognition pocket and designed wide-spectrum mechanism-based irreversible inhibitors that target this conserved region. A series of protease–inhibitor complex structures explains the uniform inhibition mechanism. Further modification of these inhibitors could lead to the discovery of a single agent with clinical potential against all coronavirus-associated diseases.

39. Bacha U, Barrila J, Velazquez-Campoy A, Leavitt SA, Freire E: **Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro**. *Biochemistry* 2004, **43**:4906-4912.

40. Blanchard JE, Elowe NH, Huitema C, Fortin PD, Cechetto JD, Eltis LD, Brown ED: **High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase**. *Chem Biol* 2004, **11**:1445-1453.

41. Jain RP, Pettersson HI, Zhang J, Aull KD, Fortin PD, Huitema C, Eltis LD, Parrish JC, James MN, Wishart DS *et al.*: **Synthesis and evaluation of keto-glutamine analogues as potent inhibitors of severe acute respiratory syndrome 3CLpro**. *J Med Chem* 2004, **47**:6113-6116.

42. Kao RY, Tsui WH, Lee TS, Tanner JA, Watt RM, Huang JD, Hu L, Chen G, Chen Z, Zhang L *et al.*: **Identification of novel small-molecule inhibitors of severe acute respiratory syndrome-associated coronavirus by chemical genetics**. *Chem Biol* 2004, **11**:1293-1299.

43. Wu C-Y, Jan J-T, Ma S-H, Kuo C-J, Juan H-F, Cheng Y-SE, Hsu H-H, Huang H-C, Wu D, Brik A *et al.*: **Small molecules targeting severe acute respiratory syndrome human coronavirus**. *Proc Natl Acad Sci USA* 2004, **101**:10012-10017.

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

The authors report the structure of the SARS replicase protein nsp9, with previously unknown function. Based on their structural analysis and surface plasmon resonance experiments, they confirm nsp9 to be an ssRNA-binding protein.

45. •• Sutton G, Fry E, Carter L, Sainsbury S, Walter T, Nettleship J, Berrow N, Owens R, Gilbert R, Davidson A *et al.*: **The nsp9 replicase protein of SARS-coronavirus, structure and functional insights**. *Structure (Camb)* 2004, **12**:341-353.

The authors report the structure of the SARS replicase protein nsp9, with previously unknown function. Based on their structural analysis and EMSA experiments, they confirm nsp9 to be an ssRNA-binding protein. They also provide evidence showing that nsp9 interacts with nsp8, suggesting that it forms part of the larger viral replication/transcription complex.

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

47. • Nelson CA, Pekosz A, Lee CA, Diamond MS, Fremont DH: **Structure and intracellular targeting of the SARS-coronavirus Orf7a accessory protein**. *Structure (Camb)* 2005, **13**:75-85.

The authors determined the first crystal structure of a SARS accessory protein, Orf7a, and showed it to have an Ig-like fold, providing clues to its function. The authors also provide evidence of the intracellular targeting of Orf7a, showing mainly intracellular retention within the Golgi network mediated by the transmembrane segment and short cytoplasmic tail of the protein.

48. Clarke J, Cota E, Fowler SB, Hamill SJ: **Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway**. *Structure Fold Des* 1999, **7**:1145-1153.

49. •• Robertson MP, Igel H, Baertsch R, Haussler D, Ares M Jr, Scott WG: **The structure of a rigorously conserved RNA element within the SARS virus genome**. *PLoS Biol* 2005, **3**:e5.

The authors report the crystal structure of the stem-loop II motif (s2m) RNA element of the SARS virus, a conserved motif at the 3′ end of the RNA genome. The structure shows an interesting 90° bend in the helix axis, and its similarity to an rRNA fold suggests that s2m might use molecular mimicry in SARS viral RNA hijacking of host protein synthesis.

50. Jonassen CM, Jonassen TO, Grinde B: **A common RNA motif in the 3′ end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus**. *J Gen Virol* 1998, **79**:715-718.

51. Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V: **Structure of the 30S ribosomal subunit**. *Nature* 2000, **407**:327-339.

52. Bushell M, Sarnow P: **Hijacking the translation apparatus by RNA viruses**. *J Cell Biol* 2002, **158**:395-399.

53. •• Li F, Li W, Farzan M, Harrison SC: **Structure of SARS coronavirus spike receptor-binding domain complexed with receptor**. *Science* 2005, **309**:1864-1868.

The long-awaited structure of the SARS spike protein RBD complexed with the receptor ACE2 is a major breakthrough that should prove important for the design of coronavirus vaccines. The complex structure reveals the interface between the two proteins, and identifies residues important for efficient cross-species infection and human-to-human transmission. Furthermore, the authors suggest that glycosylation is unlikely to interfere with major neutralising epitopes in the RBD.

54. •• Zhai Y, Sun F, Li X, Pang H, Xu X, Bartlam M, Rao Z: **Insights into coronavirus transcription and replication from the structure of the SARS-CoV nsp7-nsp8 hexadecamer**. *Nat Struct Mol Biol* 2005, in press.

The structure of the complex between two SARS replicase proteins, nsp7 and nsp8, is noteworthy for several reasons. Firstly, both nsp7 and nsp8 exhibit novel folds, with nsp8 demonstrating a unique 'golf-club'-like fold. Secondly, the complex is the first between two nsps, and provides the first insight into the organisation and sophisticated architecture of the replication and transcription machinery. The complex is a cylindrical hexadecamer composed of eight copies of nsp7 and eight copies of nsp8. A central channel has suitable dimensions and electrostatic properties for encircling double-stranded RNA, suggesting its function might be to confer processivity on the RNA-dependent RNA polymerase (nsp12).

55. Peti W, Johnson MA, Herrmann T, Neuman BW, Buchmeier MJ, Nelson M, Joseph J, Page R, Stevens RC, Kuhn P *et al.*: **Structural genomics of the SARS coronavirus: NMR structure of the protein nsP7**. *J Virol* 2005, **79**:1-9.

56. Saikatendu KS, Joseph JS, Subramanian V, Clayton T, Griffith M, Moy K, Velasquez J, Neuman BW, Buchmeier MJ, Stevens RC, Kuhn P: **Structural basis of severe acute respiratory syndrome coronavirus (SARS-CoV) ADP-ribose-1′-phosphate (Appr-1′-p) dephosphorylation by a conserved domain of nsP3**. *Structure (Camb)* 2005, in press.