Insights into SARS-CoV transcription and replication from the structure of the nsp7–nsp8 hexadecamer

Yujia Zhai[1,2,4], Fei Sun[1,2,4], Xuemei Li[1,3], Hai Pang[1,2], Xiaoling Xu[1,2], Mark Bartlam[1,2] & Zihe Rao[1,2,3]

Coronavirus replication and transcription machinery involves multiple virus-encoded nonstructural proteins (nsp). We report the crystal structure of the hexadecameric nsp7–nsp8 supercomplex from the severe acute respiratory syndrome coronavirus at 2.4-Å resolution. nsp8 has a novel 'golf-club' fold with two conformations. The supercomplex is a unique hollow, cylinder-like structure assembled from eight copies of nsp8 and held tightly together by eight copies of nsp7. With an internal diameter of $\sim$30 Å, the central channel has dimensions and positive electrostatic properties favorable for nucleic acid binding, implying that its role is to confer processivity on RNA-dependent RNA polymerase.

Coronaviruses are enveloped positive-stranded RNA viruses with the largest currently known RNA genomes. Expression of their genomes begins with the translation of two large replicase polyproteins, pp1a (>4,000 residues) and pp1ab (>7,000 residues), which are encoded by the viral replicase gene that comprises open reading frame 1a (orf1a) and orf1b[1]. pp1a and pp1ab are extensively processed by orf1a-encoded proteases to yield 15 or 16 mature nonstructural (replicase) proteins that assemble to form the membrane-associated viral replication and transcription machinery, which is vital to the viral life cycle[2]. Together with a number of cellular factors, this machinery synthesizes not only genome-sized RNA but also a nested set of eight subgenomic mRNAs. These subgenomic mRNAs are predicted to express all ORFs downstream of orf1b, encoding a variety of structural and accessory proteins[3–5]. Knowledge of the structure and organization of coronavirus replication and transcription machinery at the molecular level is limited. However, the extraordinary size of the coronavirus replicative polyproteins, their generally large phylogenetic distance from those of other RNA viruses and the presence of several predicted RNA-processing activities that are not found in other positive-stranded RNA viruses suggest that this machinery is of unparalleled complexity[3,6,7].

Recently, many studies of coronaviruses have been focused on a newly identified coronavirus: severe acute respiratory syndrome coronavirus (SARS-CoV)[8–12], the etiological agent responsible for the 2003 global SARS outbreak[13–15]. Its genome is $\sim$29.7 kilobases long (excluding the 3′ poly(A) tail) and is predicted to contain 14 functional ORFs[3–5], encoding 16 replicase proteins, 4 structural proteins and 8 accessory proteins[3]. The replicase proteins are expected to have multiple enzymatic activities, and some of these have been ascertained experimentally[6]. The activities of a papain-like protease (PL2pro, also known as nsp3), a main protease (Mpro, also known as

3CLpro or nsp5), a single-stranded (ss) RNA–binding protein (nsp9), an RNA-dependent RNA polymerase (RdRp, also known as nsp12), a superfamily 1–like helicase (HEL1, also known as nsp13) and a uridylate-specific endoribonuclease (NendoU, also known as nsp15) were recently established and characterized[3,6,16–20]. In addition, nsp3, nsp14 and nsp16 were predicted to have ADP-ribose 1′-phosphatase, 3′ → 5′ exonuclease and 2′-O-ribose methyltransferase domains, respectively[6]. However, no functions have definitively been assigned to other replicase proteins.

We present here the crystal structure of the hexadecameric nsp7–nsp8 supercomplex of SARS-CoV at 2.4-Å resolution. To our knowledge, it is the first structure to show interactions between coronavirus replicase proteins, and it offers a glimpse of the sophisticated architecture of the coronavirus replication and transcription machinery at the atomic level. Our experiments suggest that the supercomplex could encircle RNA and could function as a general processivity factor for RdRp (nsp12). The structure should help in understanding the replication and transcription mechanisms of SARS-CoV and other coronaviruses such as mouse hepatitis virus (MHV), human coronavirus strain 229E (HCoV-229E) and the recently reported human coronavirus strain HKU1 (HCoV-HKU1)[21].

RESULTS

Structural overview

The structure of the hexadecameric nsp7–nsp8 supercomplex resembles a hollow cylinder with a central channel and two handles protruding from opposite sides (**Fig. 1**). The cylinder has a height of $\sim$90 Å, an internal diameter of $\sim$30 Å and an external diameter of $\sim$95 Å ($\sim$120 Å if the two handles are included). There are 4:4 interacting nsp7 (chains A–D) and nsp8 (chains E–H) molecules
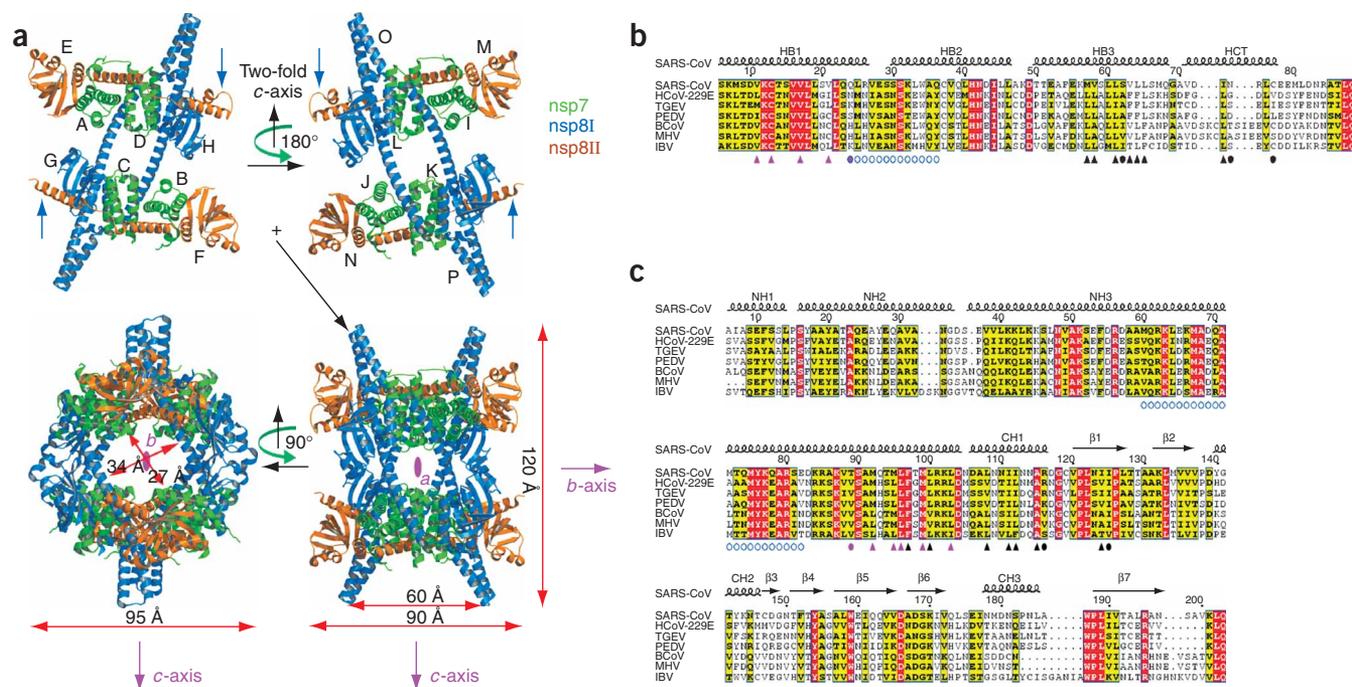
Figure 1 The supercomplex structure and its sequence. (**a**) Overall structure of the nsp7–nsp8 hexadecameric supercomplex. nsp7, nsp8I and nsp8II are colored green, blue and gold respectively. All diagrams for ribbons, sticks and balls were generated by BobScript 2.6b[40]. (**b,c**) Sequence alignment of coronavirus proteins homologous to SARS-CoV nsp7 (NP_828865) and nsp8 (NP_828866*): HCoV-229E (NP_835348, NP_835349*), transmissible gastroenteritis virus (TGEV; NP_840005, NP_840006*), porcine epidemic diarrhea virus (PEDV; NP_839961, NP_839962*), bovine coronavirus (BCoV; NP_742134, NP_742135*), murine hepatitis virus strain A59 (MHV; NP_740612, NP_740613*) and avian infectious bronchitis virus (AIBV; NP_740625, NP_740626*). Accession codes in parentheses are for GenBank; asterisks indicate nsp8 homolog. Residues boxed in red are completely conserved and those in yellow have a conservation of > 70%. Residues marked by solid circles (involved in forming hydrogen bonds) and triangles (involved in hydrophobic interactions) are responsible for interactions between nsp7 and nsp8. Magenta marks residues at site 1, black marks site 2 and empty circles mark residues surrounding the channel. Secondary structure elements are labeled according to the structures of SARS-CoV nsp7 and nsp8. The alignment was generated by ClustalW 1.7 (ref. 41) and colored by ESPript 2.1 (ref. 42).

per asymmetric unit, and the whole hexadecamer comprises two asymmetric units related by the crystallographic two-fold *c*-axis. This, together with two other pseudo two-fold axes parallel to the *a*-axis and *b*-axis, endows the structure with high symmetry. Along the *b*-axis, the supercomplexes are packed together to form a channel.

Structure of nsp7

nsp7 is an all–α-helical protein (**Fig. 2a**). Its central core is an N-terminal helical bundle (HB), with helices HB1, HB2 and HB3 (residues 5–26, 30–47 and 49–68, respectively), forming a triple-stranded antiparallel coiled coil with a right-handed superhelical pitch. A search in DALI[22] did not identify any similar structures. The HB regions of four nsp7 monomers in one asymmetric unit superimpose well (r.m.s. deviation < 0.8 Å), but the short helix HCT (residues 70–78) retains some mobility (**Supplementary Fig. 1** online). HB1–3 interact with one another mainly through hydrophobic residues. Analysis of the sequence conservation among known coronaviruses shows that the HB region is more conserved than the HCT region, an observation presumably related to the HB's role in interacting with nsp8 (**Fig. 1b**).

The 'golf-club' fold of nsp8

The four nsp8 monomers in one asymmetric unit adopt two markedly different conformations: nsp8I (chains G and H) has a 'golf club'–like structure composed of an N-terminal 'shaft' domain and a C-terminal

'head' domain (residues 6–104 and 105–196, respectively; **Fig. 2b**). The shaft domain contains three helices (NH1–3), one of which (NH3) is very long. Another three α-helices (CH1–3) and seven β-strands (β1–7) form the head domain, which has an α/β fold. The seven β-strands form an open β-barrel with two antiparallel β-sheets packed orthogonally. More than half the residues in the C-terminal domain are hydrophobic, and the whole domain forms a tight hydrophobic core.

nsp8II (chains E and F) resembles a golf club with a bent shaft (**Fig. 2c**). Although its head domain is similar to that of nsp8I (r.m.s. deviation < 0.5 Å), the shaft helix NH3 bends into two shorter helices, NH3α and NH3β, linked by a coil, C3 (**Supplementary Fig. 1**). Residues before Leu43 on chain E and residues before Asp55 on chain F could not be assigned from the electron density map. SDS-PAGE analysis of the supercomplex crystals showed that their absence was not due to protein degradation or cleavage; thus, the flexibility of the peptide resulting from crystal packing might be the reason. The presence of two conformations of nsp8 agrees with the results of PONDR analysis[23], which strongly suggests that residues 43–84 of nsp8 are disordered.

Multiple sequence alignment of corresponding coronavirus nsp8 proteins reveals high conservation, with the N-terminal domain more conserved than the C-terminal domain (**Fig. 1c**). This suggests that the N-terminal domain might have an important role in interaction with other molecules and complex assembly. No similar structures were identified from the DALI server[22].
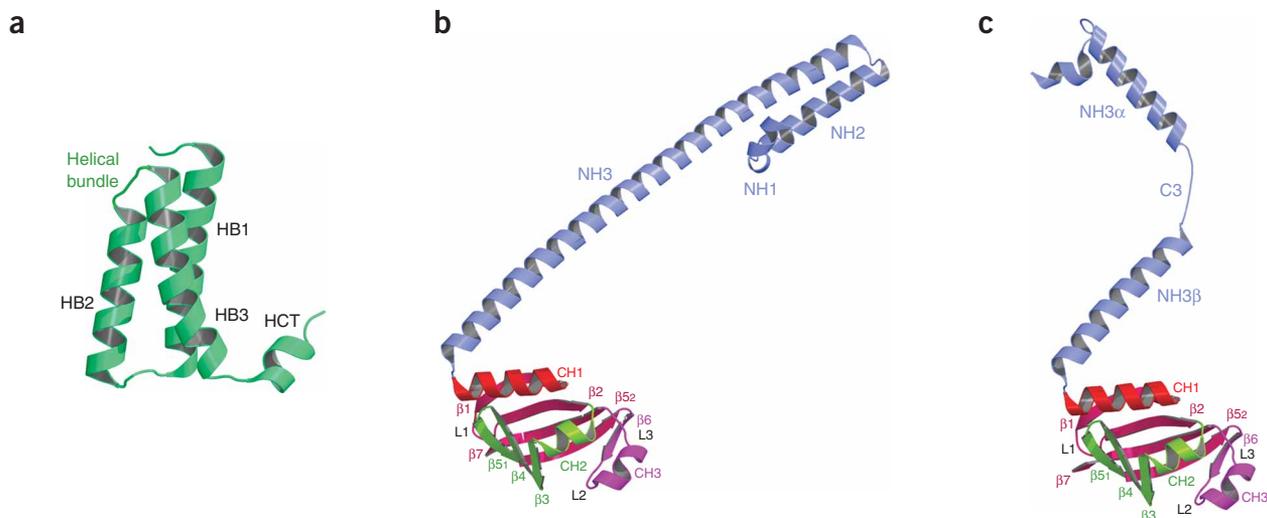
Figure 2 Three-dimensional structure of SARS-CoV nsp7 and nsp8. (**a**–**c**) Ribbon representations of SARS-CoV nsp7 (**a**), nsp8I (**b**) and nsp8II (**c**). Secondary structure elements are labeled as in **Figure 1b,c**.

Interactions between nsp7 and nsp8

Both nsp8I and nsp8II interact tightly with nsp7 to give a buried surface area of $\sim$1,400 Å^2 ($\sim$26% of the whole surface area of nsp7), forming two types of heterodimers, D1 and D2, respectively. nsp8I and nsp8II interact with nsp7 via the same sites (sites 1 and 2, **Figs. 1** and **3** and **Supplementary Fig. 2** online), suggesting that the conformational change of nsp8 is not induced by nsp7 binding. Site 1 is situated in the C-terminal region of the nsp8 shaft domain. Residues on the NH3/NH3β helix of nsp8 (Met92, Met 95, Leu96, Met99 and Leu103) and residues on the HB1 helix of nsp7 (Val11, Cys13, Val17 and Val21) form a hydrophobic core. An additional hydrogen bond is formed between the side chains of nsp8 Thr89 and nsp7 Gln24. Site 2 is located in the CH1 helix of nsp8. Helices HB3 and HCT of nsp7 interact with CH1 of nsp8 at site 2. More specifically, the side chains of residues on CH1 of nsp8 (Phe97, Leu100, Leu108, Ile111, Ile112 and Ala115) are involved in hydrophobic interactions with Met57, Val58, Leu61, Leu64, Leu65 and Ile73 of nsp7. The side chain and the main chain carbonyl group of nsp8 Arg116 form hydrogen bonds with the main chain carbonyl group of nsp7 Cys77 and the side chain of nsp7 Asn74, respectively. Furthermore, another hydrogen bond is formed between the main chain amide group of nsp8 Ile125 and the side chain of nsp7 Ser62.

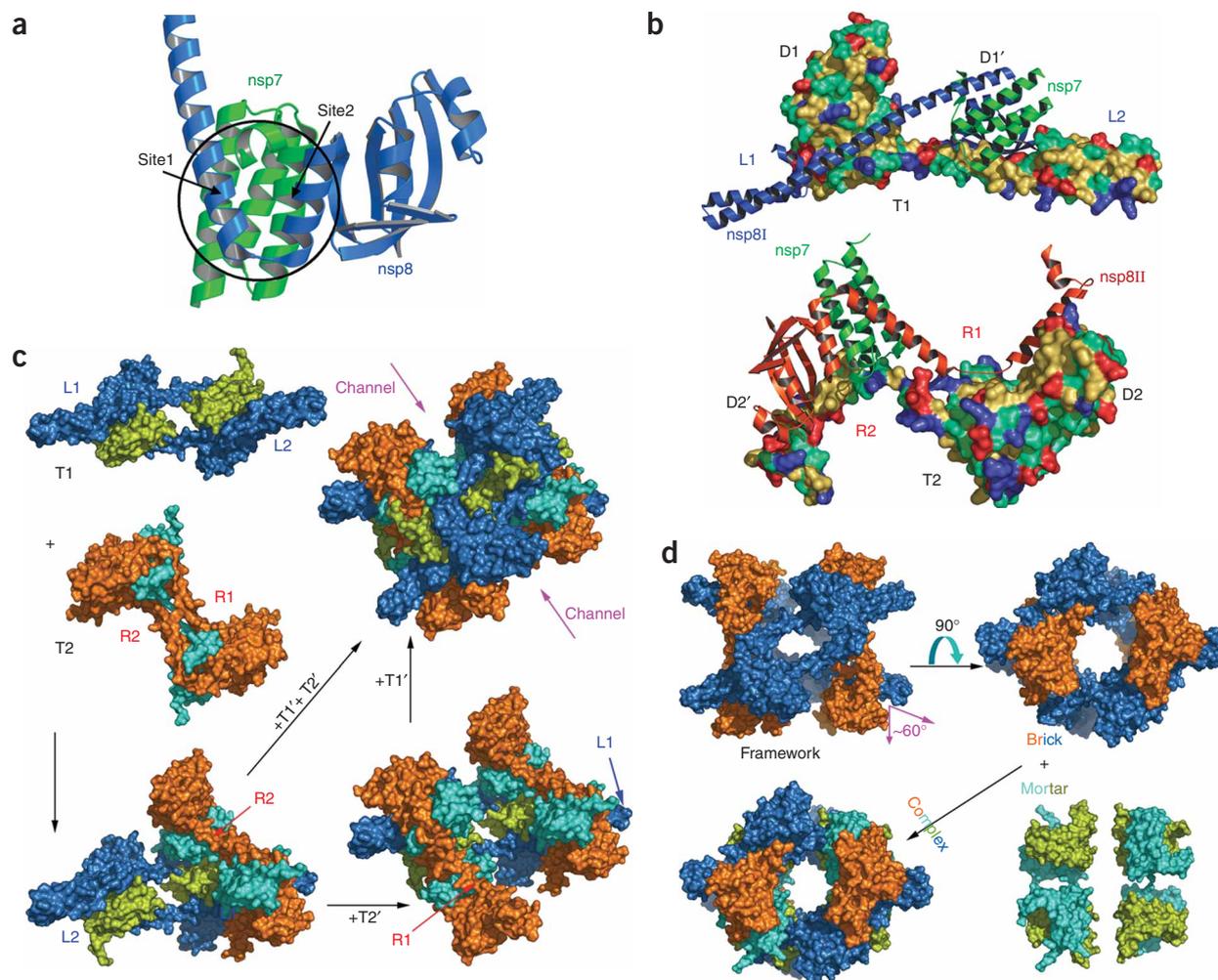### Assembly and architecture of the supercomplex

The D1 and D2 heterodimers can further dimerize to form hetero-tetramers T1 and T2, with a total buried surface area of $\sim$1,500 Å^2 and 1,700 Å^2, respectively (**Fig. 3b**). The legs (L1 and L2) of T1 are clamped by opposite regions (R1 and R2, respectively) of two T2s, with a buried surface area of $\sim$2,000 Å^2. The interactions link two T1s and two T2s together in the order T1-T2-T1′-T2′(-T1) (primes distinguish identical tetramers and parenthetical interaction indicates closing of a ring of four tetramers), enabling the full construction of the hexadecameric supercomplex (**Fig. 3c**).

The four monomers of nsp8II are oriented approximately perpendicular to those of nsp8I. Such an arrangement constitutes the framework of the supercomplex (**Fig. 3d**). Helices NH1 and NH2 and the N-terminal region of NH3 from nsp8I (residues 6–57) form the handles. The central channel is created by the middle parts

(residues 60–82) of four NH3 helices and is surrounded by NH3α helices, with β1 and β2 strands situated on either side (**Fig. 1a**). In the preliminary framework, nsp7 interacts extensively with nsp8 to make this configuration more compact and stable (**Fig. 3d**). Residues 24–36 between HB1 and HB2 of nsp7 also participate in the formation of the channel. nsp8I and nsp8II are equally important for the assembly of the supercomplex. The spatial arrangement of the 16 monomers becomes possible because the long α-helix NH3 of nsp8I bends into NH3α and NH3β in nsp8II. It is notable that most nsp8 residues around the channel are highly conserved among coronaviruses, suggesting that they have biological significance (**Fig. 1c**).

The architecture of the hexadecameric nsp7–nsp8 supercomplex is unique among macromolecular complexes reported so far that contain two kinds of protein (or subunits). In other structures, homologous multimers of one type always stack on those of another type as layers with high degrees of symmetry. The absence of any one kind of protein markedly affects the global shape of the complex. Two such examples are the structures of the multienzyme complex Rubisco[24] and the GroEL–GroES–ATP complex[25]. In contrast, the relationship between nsp7 and nsp8 multimers in the supercomplex is not stacking but cross-linking. nsp8I and nsp8II constitute the framework of the supercomplex as 'bricks,' and nsp7 stabilizes and fills this configuration as 'mortar.' Loss of nsp7 should not markedly change the shape of the structure (**Fig. 3d**). Thus, we conclude that the nsp7–nsp8 hexadecamer demonstrates a new mode of protein architecture in large macromolecular complexes. The relationships between the 16 monomers are shown in **Supplementary Figure 3** online. Overall, each monomer of nsp7, nsp8I or nsp8II interacts with other neighboring monomers to give an average buried surface area of $\sim$2,000 Å^2, 3,200 Å^2 or 3,500 Å^2, respectively, which accounts for 37%, 31.5% or 23.5% of its total solvent-accessible surface area, respectively.

The cross-linking behavior of the mixture of nsp7 and nsp8 further supports the assembly mode of the supercomplex. In cross-linking experiments, both the nsp7–nsp8 heteromultimer and the nsp8 framework were detected (**Supplementary Fig. 4** online). To examine whether the hexadecamer is the natural state of the nsp7–nsp8 complex in solution, we used negative-staining electron microscopy to obtain two-dimensional average images of the supercomplex. The

Figure 3 Architecture and assembly of the hexadecameric nsp7–nsp8 complex. (**a**) The interaction sites between nsp7 and nsp8I/II. Green, nsp7; blue, nsp8. (**b**) T1 and T2 heterotetramer formation. The two dimers from one tetramer are illustrated by the surface representation and ribbon diagram, respectively. Residues: green, polar; yellow, hydrophobic; red, acidic; blue, basic. Ribbons: green, nsp7; blue, nsp8I; orange, nsp8II. (**c**) Possible pathway for assembly of the complex by heterotetramers T1 and T2. (**d**) Hexadecameric supercomplex construction with 'bricks' of nsp8 and 'mortar' of nsp7. The angle between nsp8I and nsp8II is labeled in magenta. In **c** and **d**, nsp7 molecules interacting with nsp8I and nsp8II are colored yellow-green and blue-green, respectively.

images show particles with similar dimensions to the hexadecamer, indicating that the architecture of the nsp7–nsp8 supercomplex crystal structure also exists in solution (**Supplementary Fig. 5** online).

Interaction with dsRNA in an encircling mode

The electrostatic properties and dimensions of the nsp7–nsp8 super-complex imply that its role is to bind nucleic acids. The inner channel is coated by positive potential, whereas the outer surface of the cylinder is mainly covered by negative potential (**Fig. 4a**). This bipartite charge distribution ensures that the phosphate backbone of nucleic acids can pass through the channel without electrostatic repulsions, as with other DNA/RNA-binding proteins such as PCNA and the β subunit[26,27]. Furthermore, the central channel of the supercomplex has an average internal diameter of ∼30 Å and can suitably accommodate duplex DNA/RNA (**Figs. 1** and **4**).

It is widely accepted that coronavirus replication occurs in the cytoplasm of infected cells and that no DNA is involved[28]. A double-stranded (ds) RNA intermediate is required for genomic replication of

all coronaviruses during the RNA synthesis process. The hollow cylindrical structure of the hexadecamer suggests that its function is to encircle and stabilize dsRNA, thus holding the nascent and template strands together to facilitate efficient replication and transcription.

On the basis of this presumption, we constructed a model with a dsRNA fragment inserted into the channel to analyze the possible mode of interaction between them (**Fig. 4b**). We found that the four long helices of nsp8I could insert into the grooves of dsRNA and the positively charged residues on these helices are conserved in all homologous coronavirus nsp8 proteins. To test whether these residues are related to nucleic acid binding, we identified several basic residues located around the channel (**Supplementary Fig. 6** online) and designed three mutants: nsp7m (nsp7 R26A K32A), nsp8m1 (nsp8 K77A R80A) and nsp8m2 (nsp8 K63A R84A R85A). We then performed electrophoretic mobility shift assays (EMSAs) to examine the nucleic acid binding affinity of each mutant. The results showed that the nucleic acid binding affinities of nsp8m1 and nsp8m2 were much weaker than that of the wild-type protein, whereas mutations in nsp7
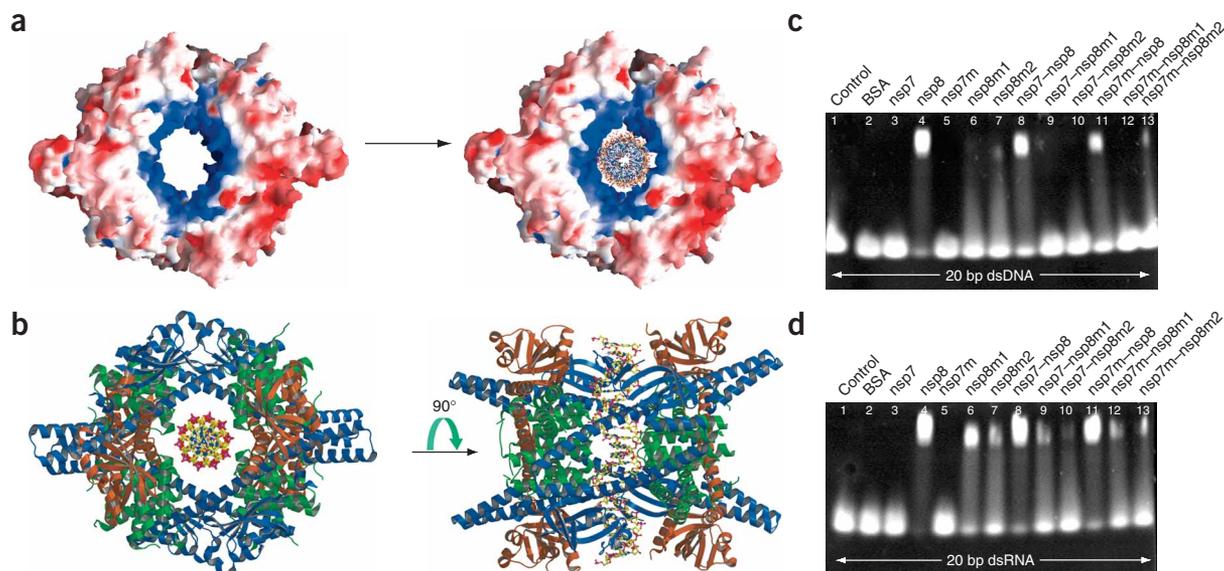
Figure 4 Hypothetical interaction between RNA and the hexadecameric nsp7–nsp8 complex. (**a**) The electrostatic potential surface of the hexadecamer modeled with (right) and without (left) duplex RNA in the positive channel. Blue, positive charge (+10 k_BT); red, negative charge (–10 k_BT). (**b**) Model of the hexadecameric nsp7–nsp8 supercomplex with hypothetical duplex RNA. Left, top view, showing the channel's proper dimensions to accommodate dsRNA. Right, side view, showing a possible mode of interaction where the four nsp8I NH3 helices insert into the dsRNA grooves. (**c,d**) Results of EMSAs of nsp7, nsp8 and their mutants. Each lane contains 75 pmol dsRNA (**c**) or dsDNA (**d**). Lanes 1–7 contain 650 pmol of each protein; lanes 8–13 contain 90 pmol of hexadecamer.

had no effect (**Fig. 4c,d**). With a calculated isoelectric point of 6.5, nsp8 should not bind nucleic acids by electrostatic interaction, unlike basic proteins. As nsp8 contains a total of 22 positively charged residues (lysines and arginines), the change in overall charge caused by the loss of two or three of these should not result in a marked change in affinity. Furthermore, the mutations did not affect super-complex formation and stability, as we ascertained by gel filtration and crystallization experiments (data not shown). The only explanation is that the locations of these mutated residues make them essential for nucleic acid binding. As they are all located around the central channel, we conclude that the channel should encircle nucleic acid. In addition, the results of EMSAs showed that nsp8 mutants have higher affinity for dsRNA than for dsDNA, as they could still bind dsRNA but hardly bound dsDNA. This suggests that dsRNA is the likely natural binding partner of the nsp7–nsp8 supercomplex.

DISCUSSION

The high conservation of nsp7 and nsp8 in known coronaviruses suggests that the hexadecamer should be a general component for all coronaviruses. The electrostatic properties of the hexadecamer and the diameter of its central channel are similar to those of PCNA and the β subunit ring, the processivity factors of DNA polymerase[26,27], which encircle dsDNA and interact with the polymerase to confer high processivity on it[29]. Coincidentally, experiments on MHV have shown that RdRp co-immunoprecipitates with nsp8, nsp9, nsp5 (main protease, also called M^{pro} or $3CL^{pro}$) and nsp13 (helicase)[30], which also implies an interaction between RdRp and the nsp7–nsp8 hexa-decamer. Their remarkably large genomes and putative proofreading activities suggest that coronaviruses may differ from other RNA viruses and share unprecedented similarities with DNA-based life forms in the mechanisms of genome biosynthesis[6]. The hexadecamer might be a factor that binds and trails RdRp, conferring high processivity on it for efficient replication of the extremely large

coronavirus genome. Such a binding mode would give a molar ratio of 8:1 between nsp7–nsp8 and RdRp, which agrees with the natural abundance of orf1a-encoded nsps in the replication and transcription machinery (three- to five-fold in excess of orf1b-encoded nsps such as RdRp, helicase, exonuclease and others) resulting from a ribosomal frameshifting mechanism[1].

The colocalization of nsp8 with nsp7, nsp9 and nsp10 in experi-ments on MHV[31] provides very strong evidence for their interaction in this virus. Analytical ultracentrifugation experiments further indicate that SARS-CoV nsp8 interacts with nsp9, a ssRNA-binding protein[20]. In addition, the disorder of the nsp8 N-terminal region has been seen to decrease upon the addition of nsp9 to nsp8 (ref. 20). On the basis of the nsp7–nsp8 hexadecameric structure, the most probable nsp9-binding site should be in the region formed by the N-terminal 50 residues of nsp8II, which is located at the entrance of the channel and has high flexibility with missing electron density. Wrapping of ssRNA around the nsp9 dimer is suggested by the nsp9 structure and by a study using tryptophan fluorescence quenching[19]. We therefore consider that the function of the nsp9 dimer might be to protect the newly unwinding nascent and template strands emerging from the channel of the nsp7–nsp8 complex, which have not yet formed a stable secondary structure, from nuclease processing.

The crystal structure of the SARS-CoV hexadecameric nsp7–nsp8 supercomplex is the first obtained so far to show atomic inter-actions between coronavirus nonstructural proteins. Sixteen molecules associate tightly with one another to form a handled hollow cylindrical structure in which the coexistence of two conformations (nsp8I and nsp8II) is observed. Its novel architecture and unique mode of assembly should provide new insights in the field of macromolecular complex structures. Designing peptides or nonpeptidyl compounds that mimic the interaction interface between nsp7 and nsp8 is one strategy to block supercomplex formation and interfere with virus replication. Besides M^{pro}, this structure could provide a new candidate

Table 1 Data collection, phasing and refinement statistics

	Native	nsp7(SeMet)–nsp8 crystal		
Data collection				
Space group	P2₁2₁2	P2₁2₁2		
Cell dimensions				
a, b, c (Å)	93.6, 94.0, 150.9	93.6, 94.0, 150.9		
α, β, γ (°)	90, 90, 90	90, 90, 90		
		Peak	Inflection	Remote
Wavelength (Å)	1.0332	0.9791	0.9793	0.9500
Resolution (Å)	2.3	2.8	2.8	2.8
R_{merge}	0.099 (0.476)[a]	0.104 (0.422)	0.092 (0.369)	0.127 (0.604)
$I/\sigma I$	24.0 (1.7)	15.5 (2.0)	15.5 (2.0)	10.1 (1.6)
Completeness (%)	96.8 (82.0)	99.0 (95.2)	97.1 (81.0)	96.0 (90.3)
Redundancy	8.4 (3.1)	5.9 (3.8)	5.4 (3.2)	5.4 (4.1)
Refinement				
Resolution (Å)	50–2.4			
No. reflections	51,504			
R_{work} / R_{free}	21.3/25.1			
No. atoms	7,782			
Protein	7,598			
Ligand/ion	35			
Water	149			
B-factors				
Protein	56.5			
Ligand/ion	76.2			
Water	55.2			
R.m.s. deviations				
Bond lengths (Å)	0.006			
Bond angles (°)	1.14			

[a]Highest-resolution shell is shown in parentheses (2.90–2.80 Å for selenium energies and 2.38–2.30 Å for the native data set).

for drug design targeting those serious diseases caused by SARS-CoV, HCoV-229E and HCoV-HKU1.

METHODS

Protein expression, purification and supercomplex assembly. The coding sequences for SARS-CoV nsp7 and nsp8 were amplified by PCR from the SARS-CoV BJ01 strain (corresponding to 3837Ser–3919Gln and 3920Ala–4117Gln of orf1a replicative polyprotein, respectively) and inserted into the pGEX-6p-1 plasmid using BamHI and XhoI sites. The proteins were expressed in *Escherichia coli* strain BL21 (DE3) as GST fusion proteins. A selenomethionyl (SeMet) derivative of nsp7 was prepared using the method of methionine-biosynthesis pathway inhibition[32]. The GST fusion proteins were first purified by glutathione affinity column. The GST was released by GST–rhinovirus 3C protease (Amersham Biosciences), leading to five additional residues (GPLGS) at the N terminus, and SeMet-nsp7 and nsp8 were further purified by Superdex 200 (10/30) gel filtration column (Amersham Biosciences) in 25 mM sodium HEPES, 150 mM NaCl, 1 mM EDTA and 5 mM DTT (pH 7.5). nsp8 was then mixed with ~1 molar excess of SeMet-nsp7 and passed over the Superdex 200 column in the same buffer. Fractions of the SeMet-nsp7–nsp8 complex were then concentrated and used for crystallization.

Crystallization and data collection. Crystals were grown at 291 K by the hanging drop vapor diffusion method from an ammonium sulfate system with a complex concentration of 5 mg ml⁻¹ in the gel filtration buffer. Crystals were flash-frozen in liquid nitrogen in a crystallization buffer supplemented with 25% glycerol. MAD data were collected to 2.8-Å resolution on beamline 3W1B of the Beijing Synchrotron Radiation Facility. Higher-resolution (2.4 Å) data were collected from a single crystal of the native complex on beamline BL19-ID of the Advanced Photon Source (Argonne

National Laboratory, Argonne, Illinois, USA). Data were processed with the HKL2000 suite of programs[33] (**Table 1**).

Phasing, model building and refinement. MAD data (2.8 Å) were used to locate selenium sites and calculate an initial electron density map for model building. Twelve selenium sites were located using SOLVE[34], which identified noncrystallographic symmetry (four operators), and the initial phases were calculated up to 3.2 Å. Density modification and phase extension to 2.8 Å were performed with RESOLVE[35], and an interpretable electron density map was calculated. Automatic model building was performed using RESOLVE and ~43% of the asymmetric unit was traced, including 166 full residues and 521 residues lacking side chains. The remainder of the model was built manually using O[36]. Finally, four nsp7 molecules and four nsp8 molecules were located in one asymmetric unit. This model was refined to a resolution of 2.8 Å, with $R_{work} = 23.2\%$ and $R_{free} = 28.2\%$ for the MAD data, using CNS[37]. The model was further refined, using the higher-resolution native data, to 2.4 Å, with $R_{work} = 21.5\%$ and $R_{free} = 25.4\%$. The final model was confirmed to have good stereochemistry according to a Ramachandran plot calculated by PROCHECK[38]. Phasing and refinement statistics are summarized in **Table 1**.

Electron microscopy and image processing. For electron microscopy, the mixture of nsp7 and nsp8 was diluted to a final concentration of 0.3 mg ml⁻¹ in 20 mM sodium HEPES (pH 7.5) and 150 mM NaCl. The sample was applied to 400 mesh copper grids using the carbon sandwich technique with 1% uranyl acetate as a negative stain. Each micrograph was recorded with a JEM-100CX transmission electron microscope at a magnification of ×33,000 and an accelerating voltage of 80 kV. The images were processed with EMAN[39] and about 2,500 particles were extracted from the best image after manual screening of the automatic extracted particles. The reference-free alignment and two-dimensional averaging were performed by Boxer[39].

Cross-linking gel assay. nsp7, nsp8 and mixtures of the two were diluted to 5 mg ml⁻¹ in 25 mM HEPES (pH 7.5), 300 mM NaCl, 1 mM EDTA and 1 mM DTT. Ethylene glycolbis (succinimidylsuccinate) was dissolved in DMSO to a concentration of 50 mM and then added to 10 µl of protein sample with a final concentration of 5 mM. After the mixture was incubated on ice for 2 h, the reaction was quenched for 15 min by adding 1 M Tris-HCl (pH 7.5) to a final concentration of 50 mM. An equal volume of 2× SDS-PAGE sample buffer was added and a small amount was analyzed on a 4–15% SDS polyacrylamide gel.

Electrophoretic mobility shift assays. The dsDNA (5′-CTTGCAAAAGACA CAACTGA-3′) was synthesized by BioAsia. The dsRNA (5′-NGGAGACCAU GUGAUUGGCA-3′) was a gift from G. Gao (Chinese Academy of Sciences, Beijing, China). Nucleic acid was incubated with protein in 20 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM EDTA and 5% glycerol for 30 min at room temperature. The samples were run on 4% nondenaturing TBE polyacrylamide gel, and the gel was then stained with ethidium bromide.

Accession codes. Protein Data Bank: Coordinates have been deposited with accession code 2AHM.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

1. Brierley, I., Digard, P. & Inglis, S.C. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57**, 537–547 (1989).
2. Ziebuhr, J., Snijder, E.J. & Gorbalenya, A.E. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**, 853–879 (2000).
3. Thiel, V. *et al.* Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305–2315 (2003).
4. Marra, M.A. *et al.* The Genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399–1404 (2003).
5. Rota, P.A. *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**, 1394–1399 (2003).
6. Snijder, E.J. *et al.* Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **331**, 991–1004 (2003).
7. Gorbalenya, A.E. Big nidovirus genome. When count and order of domains matter. *Adv. Exp. Med. Biol.* **494**, 1–17 (2001).
8. Drosten, C. *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003).
9. Fouchier, R.A. *et al.* Aetiology: Koch's postulates fulfilled for SARS virus. *Nature* **423**, 240 (2003).
10. Ksiazek, T.G. *et al.* A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* **348**, 1953–1966 (2003).
11. Kuiken, T. *et al.* Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* **362**, 263–270 (2003).
12. Peiris, J.S. *et al.* Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**, 1319–1325 (2003).
13. Peiris, J.S., Yuen, K.Y., Osterhaus, A.D. & Stohr, K. The severe acute respiratory syndrome. *N. Engl. J. Med.* **349**, 2431–2441 (2003).
14. Zhao, Z. *et al.* Description and clinical treatment of an early outbreak of severe acute respiratory syndrome (SARS) in Guangzhou, PR China. *J. Med. Microbiol.* **52**, 715–720 (2003).
15. Zhong, N.S. *et al.* Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**, 1353–1358 (2003).
16. Yang, 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* **100**, 13190–13195 (2003).
17. Ivanov, K.A. *et al.* Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J. Virol.* **78**, 5619–5632 (2004).
18. Ivanov, K.A. *et al.* Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Natl. Acad. Sci. USA* **101**, 12694–12699 (2004).
19. Egloff, M.P. *et al.* The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. *Proc. Natl. Acad. Sci. USA* **101**, 3792–3796 (2004).
20. Sutton, G. *et al.* The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure (Camb).* **12**, 341–353 (2004).
21. Woo, P.C. *et al.* Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* **79**, 884–895 (2005).
22. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
23. Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M. & Obradovic, Z. Intrinsic disorder and protein function. *Biochemistry* **41**, 6573–6582 (2002).
24. Taylor, T.C. & Andersson, I. The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate. *J. Mol. Biol.* **265**, 432–444 (1997).
25. Xu, Z., Horwich, A.L. & Sigler, P.B. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* **388**, 741–750 (1997).
26. Kong, X.P., Onrust, R., O'Donnell, M. & Kuriyan, J. Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* **69**, 425–437 (1992).
27. Krishna, T.S., Kong, X.P., Gary, S., Burgers, P.M. & Kuriyan, J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **79**, 1233–1243 (1994).
28. Siddell, S.G. The coronaviridae: an introduction. in *The Coronaviridae* (ed. Siddell, S.G.) 1–9 (Plenum, New York, USA, 1995).
29. Bauer, G.A. & Burgers, P.M. The yeast analog of mammalian cyclin/proliferating-cell nuclear antigen interacts with mammalian DNA polymerase delta. *Proc. Natl. Acad. Sci. USA* **85**, 7506–7510 (1988).
30. Brockway, S.M., Clay, C.T., Lu, X.T. & Denison, M.R. Characterization of the expression, intracellular localization, and replication complex association of the putative mouse hepatitis virus RNA-dependent RNA polymerase. *J. Virol.* **77**, 10515–10527 (2003).
31. Bost, A.G., Carnahan, R.H., Lu, X.T. & Denison, M.R. Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. *J. Virol.* **74**, 3379–3387 (2000).
32. Doublie, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
33. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
34. Terwilliger, T.C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861 (1999).
35. Terwilliger, T.C. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 965–972 (2000).
36. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–9 (1991).
37. Brunger, A.T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
38. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
39. Ludtke, S.J., Baldwin, P.R. & Chiu, W. EMAN: semiautomated software for high-resolution single particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
40. Esnouf, R.M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132–134 (1997).
41. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
42. Gouet, P., Courcelle, E., Stuart, D.I. & Metoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).