Protein Cell 2010, 1(2): 198–204
DOI 10.1007/s13238-010-0028-8

Protein & Cell

RESEARCH ARTICLE

New nsp8 isoform suggests mechanism for tuning viral RNA synthesis

Shuang Li[1*], Qi Zhao[2*], Yinjie Zhang[3*], Yang Zhang[2], Mark Bartlam[3], Xuemei Li[1], Zihe Rao[1,2,3] (✉)

[1] National Laboratory of Macromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[2] Structural Biology Laboratory, Tsinghua University, Beijing 100084, China
[3] Tianjin Key Laboratory of Protein Science, College of Life Sciences, Nankai University, Tianjin 300071, China

✉ Correspondence: raozh@xtal.tsinghua.edu.cn
Received December 26, 2009; accepted January 6, 2010

ABSTRACT

During severe acute respiratory syndrome coronavirus (SARS-CoV) infection, the activity of the replication/transcription complexes (RTC) quickly peaks at 6 hours post infection (h.p.i) and then diminishes significantly in the late post-infection stages. This "down-up-down" regulation of RNA synthesis distinguishes different viral stages: primary translation, genome replication, and finally viron assembly. Regarding the nsp8 as the primase in RNA synthesis, we confirmed that the proteolysis product of the primase (nsp8) contains the globular domain (nsp8C), and indentified the resectioning site that is notably conserved in all the three groups of coronavirus. We subsequently crystallized the complex of SARS-CoV nsp8C and nsp7, and the 3-D structure of this domain revealed its capability to interfuse into the hexadecamer super-complex. This specific proteolysis may indicate one possible mechanism by which coronaviruses to switch from viral infection to genome replication and viral assembly stages.

KEYWORDS nsp8, SARS-CoV, RNA primase, viral life cycle

INTRODUCTION

Coronaviruses are positive-strand RNA viruses whose replication machinery is primarily translated by the host ribosome. For positive-strand RNA viruses, the replication of the viral genome requires an RNA-dependent RNA polymerase (RdRp), together with a varied priming strategy. Also, functional non-structural proteins (nsp) utilized as replication machinery are generated via post translational processing.

Determining the specific function of these nsp proteins remains one of the most difficult challenges in virology. However, in the light of severe acute respiratory syndrome coronavirus (SARS-CoV) genome sequencing, bioinformatical studies have predicted the function of nsp12 (RNA-dependent RNA polymerase), nsp5 (main protease), nsp13 (endo-ribonuclease), nsp14 (helicase), nsp15 (exoribonuclease) and nsp16 (methyltransferase), and domains of nsp3 (ADRP and papain-like protease) (Prentice et al., 2004; Xu et al., 2006; Yang et al., 2003). These analyses also indicate that the C-terminal domain of SARS-CoV nsp8 contains an RNA binding motif that is widely shared by helicases, telomere-associated proteins and other single strand nucleotide binding proteins (Imbert et al., 2006). Nsp8 can also interact with nsp7, presumably for stabilization purposes, and the crystal structure of the nsp7-nsp8 super complex provided the first glance into the machinery of SARS replication and transcription complex (Zhai et al., 2005). The ability of the nsp7-nsp8 complex to bind nucleic acid was first characterized by Zhai et al. (2005), using random RNA and DNA sequences as substrates. In the same study, mutagenesis experiments targeting several positively charged residues demonstrated that the end of the extended substrate tail is crucial for nucleic acid affinity. Subsequent experimental

*These authors contributed equally to the work

assays concerning RNA with different sequences (Imbert et al., 2006) confirmed the role of the long helix motif of nsp8 in nucleic acid binding.

In another species of positive-strand RNA viruses, the picornaviruses, both the cellular factors and the non-structural proteins are orchestrated for tuning the viral infection cycle (Brandt, 2005; Perera et al., 2007). The cellular factor poly(rC) binding protein (PCBP2, also known as HnRNP K2) is reported to participate in both the initiation of translation and RNA replication (Graff et al., 1998). In addition, the mechanism of switching from viral protein synthesis to RNA replication was found to occur via proteolysis of the key regulator PCBP2 (Perera et al., 2007). As a result of the action of the 3CD proteinase, the third domain of PCBP2 is truncated and viral translation is thus switched off. For coronaviruses, viral development in infected cells can follow similar stages subsequent to virus entry (Stertz et al., 2007). By analyzing the viral RNA synthesis, it is reported that the plus strand synthesis of SARS-CoV in cells declines to undetectable levels by 12 hours post-infection (h.p.i.) (Sawicki et al., 2007), when the infection stage moves from RNA synthesis to viral assembly.

The native form of nsp8 in complex with nsp7 suggested that functional nsp8 interacts with nsp7 to form a channel-like hexadecamer, with the positively-charged electrostatic potential around the central channel significant for interaction with the negatively-charged RNA backbone (Zhai et al., 2005). Based on primase assays and structure comparison, Imbert and colleagues identified this nsp7-nsp8 complex as a second RNA-dependent RNA polymerase (RdRp) in SARS-CoV, generating 6 nt RNA primers for further elongation carried out by the major RdRp, nsp12 (Imbert et al., 2006). The indispensability of this secondary RdRp/nsp8 was subsequently confirmed by Deming et al. (2007). In contrast, nsp2 was shown to be dispensable for the viability of coronavirus (Graham et al., 2006), and thus a less-important factor for regulating the function of the coronavirus RTC.

In early 2004, shortly after the outbreak of the SARS epidemic, Prentice et al. (2004) characterized the SARS-CoV non-structural proteins in infected cells by immunoblotting. Interestingly, two of the 14 nsps (nsp2 and nsp8) were found to migrate as two species of different sizes featuring a lower molecular weight isoform. Nsp2 has an isoform that is 15 kDa shorter than the predicted full-length protein, while nsp8 has a second band that is 8 kDa shorter. Regarding the crucial role of nsp8, we hypothesized that the smaller band could be the isoform of nsp8, which might play a crucial role in viral RNA synthesis.

To verify our hypothesis, we have incubated the native full-length nsp8 with RNA to gain a similar proteolytic product (also 8 kDa smaller). By X-ray crystallography, we find that the previously observed “golf club-like” architecture of nsp8 was reduced to a single globular domain (nsp8c). The three-dimensional structure of nsp7-nsp8c suggests the possibility that nsp8c could interfere with its activity, possibly through incorporating into the nsp7-nsp8 supercomplex.

RESULTS

SARS-CoV nsp8 undergoes resectioning in the cell

During early attempts at characterizing the non-structural proteins by Western blot, an unexpected lower molecular weight band for nsp8 was discovered, but its origin was unknown. After discounting the possibility of an artefact, Prentice et al. proposed that nsp8 might undergo further processing (Prentice et al., 2004). Different groups reported the inherent instability of nsp8 in SARS-CoV (Imbert et al., 2006; Prentice et al., 2004) and MHV (Brockway et al., 2003; van der Meer et al., 1999), but also suggested that the presence of nsp7 might increase its stability. In order to determine the residual nsp8 domain after host proteolysis in the cell, we performed Western blotting using antibodies prepared from the N-terminus and C-terminus of recombinant nsp8, respectively (Fig. 1).

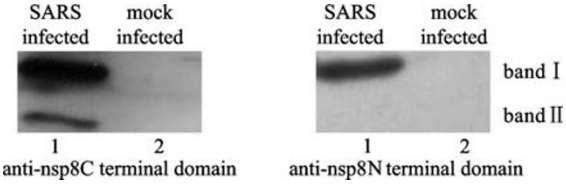

Figure 1. Immunological mapping of band II. The gel was blotted using anti-nsp8C (left) or anti-nsp8N (right) as the primary antibody. Lane 1 is mock infected negative control, and lane 2 is the sample infected with SARS-CoV. The band of interest is marked as “band I”.

This mapping shows that the smaller isoform of nsp8 contains the 120 a.a. C-terminal domain of the protein but not the 80 a.a. N-terminal domain. As this specific degradation leads to the loss of the N-terminal head, this process could be conceived as a resectioning of nsp8. Since no conventional viral protease recognition site is located inside nsp8, the accumulation of nsp8C beginning from the early stages of viral infection could be largely attributed to host defensive proteases,or its inherent instability or both.

Our *in vitro* results also indicate that nsp8 is unstable, and N-terminal sequencing of nsp8C shows that the first consecutive amino acids are “$M^{75}Y^{76}K^{77}Q^{78}A^{79}$”. We can therefore deduce that the resectioning site occurs at the methionine in position 70 of nsp8. Unfortunately, due to the scarcity of viral nsp8 in infected cells, the precise position of resectioning under physiological conditions is unknown. However, we were able to make predictions from the molecular weight, immunological mapping results, near-physiological studies and structural clues, as discussed below.

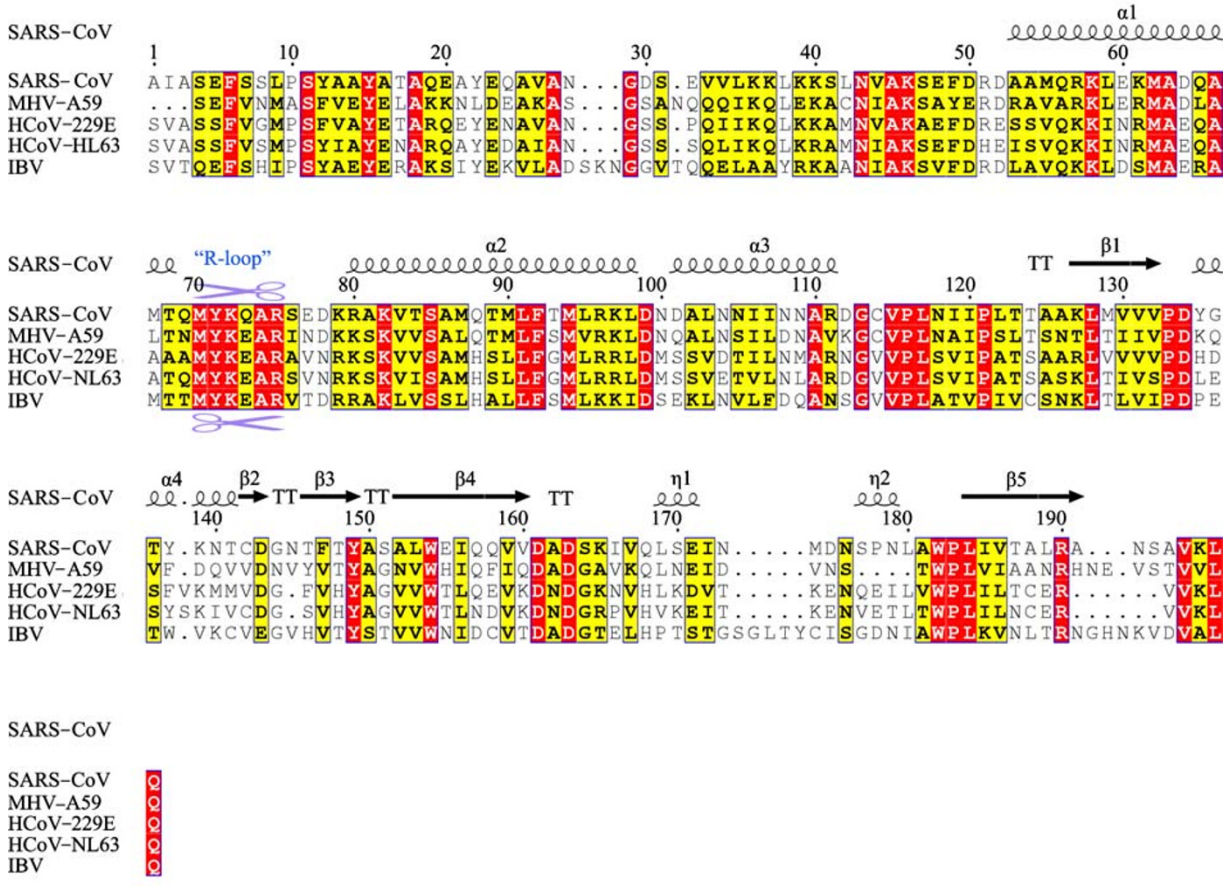

Figure 2. Sequence alignment of nsp8 from four typical coronaviruses. Colors indicate different levels of conservation (white, no conservation; yellow, homologous substitution; red, absolute conservation). The resectioning point ("R-loop") is marked using a symbolic blue scissor.

Coronaviruses include members from three groups: group 1 (e.g., human coronavirus 229E), group 2A (e.g., human coronavirus HKU1) and 2B (e.g., SARS-CoV), and group 3 (e.g., avian infectious bronchitis virus). We aligned the nsp8 sequences from all of these groups and discovered features, including a "resectioning marker" (R-Loop), which is notably conserved in one of the most variable parts of the entire nsp8 sequence (Fig. 2). Based on this, we hypothesized that proteolysis of nsp8 might occur in the cells infected by CoVs from other groups.

This discovery prompted us to further investigate the functional implication of the resectioning process. To characterize its structural and biochemical features, we overexpressed and purified the nsp8 C-terminal domain (nsp8C) using a truncated construct. The structure of the nsp7-nsp8C complex was then obtained through X-ray crystallography (Fig. 3A, Supplemental Table 1).

Structural view of the resectioning site and implications on the biphasic activity

In the intact structure of the channel-like nsp7-nsp8 hexadecamer, the position of the resectioning site was mapped to the loop that bends the long "shaft" in half of the nsp8 subunits in the nsp7-nsp8 super-complex (Fig. 3B). Only half of the nsp8s (termed nsp8II, or "bent golf club") possess such loops, while the four straight nsp8s (nsp8I, or "regular golf club") molecules adopt a straight helical "shaft" in the corresponding region. The unstructured loop presumably indicates instability. We hypothesized that resectioning might first occur at these regions, and then by compromising the integrity of the channel, nsp8I could be destabilized thereafter. While the central hole of the channel is lined solely by the N-termini of nsp8I (Fig. 4), any resectioning of nsp8I would be detrimental to the whole architecture that is thought to be crucial for RNA

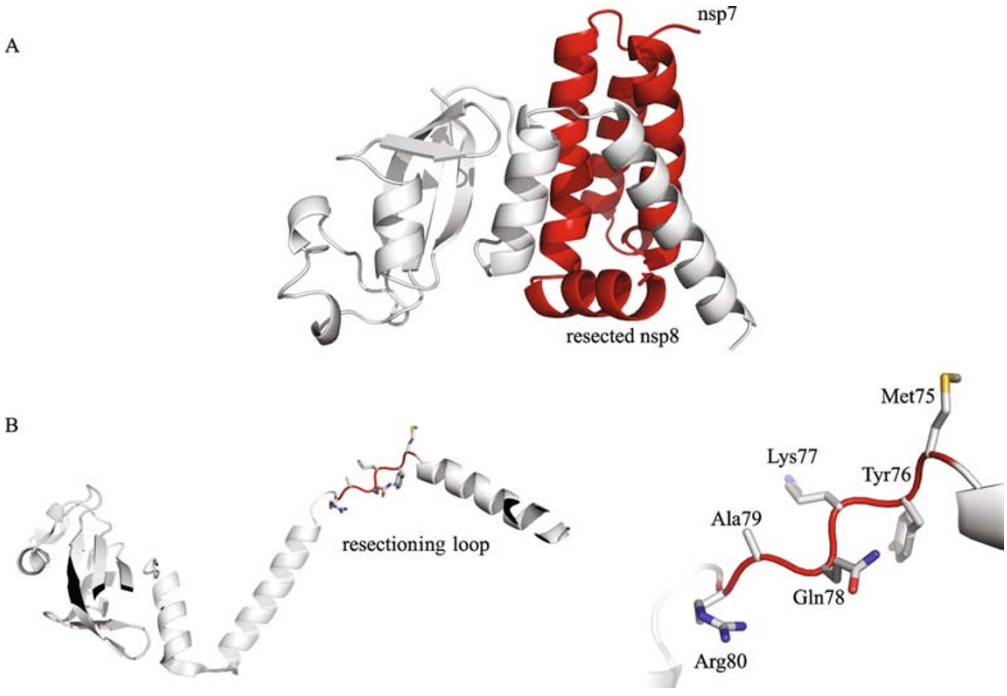

Figure 3. Crystal structure of the nsp7-nsp8C complex and structural view of the resectioning site. (A) Crystal structure of the nsp7-nsp8C complex featuring a third nsp8 isoform, nsp8C. Nsp7 is displayed in red cartoon, and nsp8C is displayed in white cartoon. (B) The position of the resectioning loop. The loop is shown in red stick representation and the whole nsp8I structure is represented in white cartoon.

binding. On the contrary, due to the insignificant role of the N-termini of nsp8II, resectioning might lead to much milder consequences.

DISCUSSION

The alignment of the structures of all three nsp8 isoforms reveals a conserved C-ternimal domain (Fig. 5A), which is not only the major interface that forms the hexadecamer structure as is discussed above but also might be responsible for its activity. As is proposed by Imbert et al. (2006) the gross fold of C-terminal domain of the nsp8 resembles the structure of hnrnp A1 (PDBID 2UP1), which is a single nucleotide binding protein (Fig. 5B) in human cells. Therefore, it is possible that the nsp8C isoform is able to preserve the nucleotide binding affinity to interfere with the full-lengh nsp7-nsp8 complex, indicating a mechanism to regulate the initiation of viral RNA synthesis.

The timeline of SARS-CoV infection can be interpreted as a process during which the virus balances the apoptotic effects and viral self-replication in the cell. We have identified a new SARS-CoV nsp8 isoform (nsp8C), which functions as a possible primase and could be one of the indicators and/or parameters for coronaviruses to tune their biological priorities during the viral life cycle (Fig. 6). The replication of coronaviruses can be divided into two steps: replication of the genome and virus assembly. The regulatory tilting point between the two steps is ambiguous; however, we here propose a possible mechanism by which the virus could regulate the switching point. The amount of viral RNA is directly related to the RNA polymerization reaction that requires both the primase and RdRp. After proteolysis, nsp8C may regulate RNA synthesis through remodeling the components of primase.

The proteolysis of viral and/or host proteins commonly requires host and/or viral proteolytic equipment post infection. Proteosome- and endosome-based protein degradations are two of the major pathways for the cell to process the relevant proteins. However, proteases in these pathways usually degrade proteins to oligo-peptides or single residues. In addition, apoptosis-related caspases are responsible for a large amount of protein degradation but in a more sequence-specific manner. To determine the relevance between the host caspase and nsp8 processing, we performed a series of experiments in which infected cells were incubated with the presence of caspase 3 inhibitors in media. However, no inhibition of nsp8 resectioning was detected, suggesting that the caspase pathway may not directly cause nsp8 proteolysis (Supplemental Fig. 1). Further experiments are required to evaluate other possible proteolytic agents. However, it should be noted that the *in vitro* degradation of nsp8 in the absence of known protease indicates that the inherent instability of nsp8 itself should not be neglected.

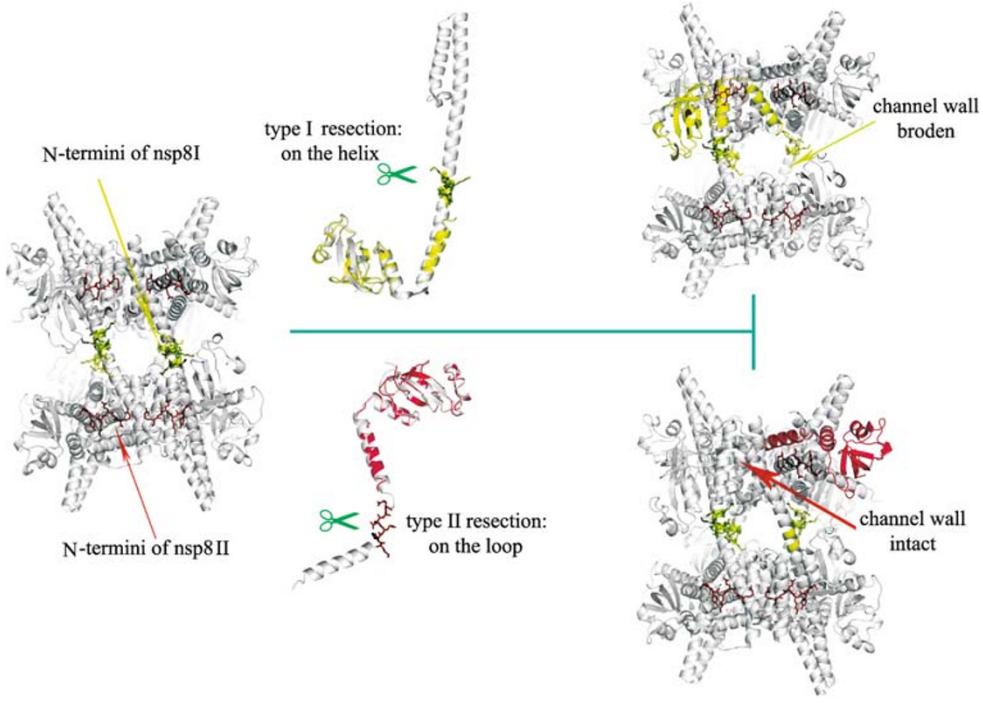

Figure 4. Two types of resectioning on the full-length channel and their consequences for the preservation of the supercomplex.

MATERIAL AND METHODS

Cells and culturing

Immunological mapping of nsp8C from SARS-CoV-infected cells were collected at 12 h.p.i., and the cells were adjusted to 3 million. Cells were pelleted by centrifugation at 5000 rpm for 10 min, and cell lysates were extracted with a buffer containing 50 mM sodium phosphate (pH 7.3), 150 mM NaCl, 10 mM KCl, 1% NP40, 2% SDS and a cocktail of protease inhibitors. The supernatant was obtained by centrifugation at 12,000 rpm, and then 5% of the sample (10 μL) was mixed with 2 μL 6 × loading dye and loaded onto a 15% SDS-PAGE gel. The negative controls (mock infected cells) for each sample were treated according to the same protocol. The gel was transferred onto PVDF membrane and blocked for 1 h with 10% skim milk (Sigma), then incubated for 1 h with 1:1000 New Zealand rabbit anti-serum prepared in 2% skim milk. The membrane was washed 3 times with washing buffer (PBS with 0.5% Tween-20) for 15 min each. After incubating with the primary antibody, the membrane was incubated with the secondary goat anti-rabbit antibody in the ratio 1:5000 for 1 h, and then washed following the same protocol. Finally, the proteins were detected with the Merck fluorescence-based Western blot detection kit and visualized by Kodak imaging film. All procedures were performed at room temperature. The N-terminal tracing experiment was also conducted using an anti-nsp8N antibody as the primary antibody while other conditions were unchanged.

In vitro degradation and N-terminal sequencing

The native protein was incubated with RNA to mimic physiological conditions. After 2 days, the 22-kDa native protein was reduced to a 14-kDa product. Peptide fingerprints and N-terminal sequencing of the proteolysis product were performed by transferring the protein onto a PVDF membrane and cutting out the band of interest. The first five residues of this band were determined as $M^{75}Y^{76}K^{77}Q^{78}A^{79}$. The C-terminus was then sub-cloned (see below) and its structure solved. In the electron density, the structure could be seen beginning from Lys^{82}, possibly as a result of structural flexibility.

Cloning and expression

A sub-clone based on information derived from N-terminal sequencing of the *in vitro* degradation product was constructed by inserting a *Bam*HI restriction site before Asp78. The end of this sequence was flanked by an *Xho*I restriction site immediately after a stop codon. The first 60 residues of nsp8 were cloned using the same strategy. The PCR product was digested with commercial *Bam*HI and *Xho*I restriction enzymes (Takara) and inserted into linearized plasmid pGEX 6p-1. The validity of this construct was confirmed by duel enzyme digestion and DNA sequencing. The plasmid was first transformed into a commercial *E. coli* strain BL21 (DE3) Rosetta (Invitrogen). After incubation at 37°C overnight on an Amp^+ algae LB plate, fresh transformants were inoculated into 5 mL LB media in the presence of 100 μg/mL ampicillin. After 12 h of growth, the incubation system was scaled up to 1 L LB media, with the same concentration of antibiotics in a 2 L flask, and was shaken vigorously at 37°C until the $OD_{600\,nm}$ reached 0.6. Protein expression was induced by 0.5 mM IPTG (Sigma) at 16°C overnight. Cells expressing nsp7 were treated identically and then mixed with cells expressing nsp8C prior to harvesting. Cell pellets were harvested by centrifugation, re-suspended in 40 mL PBS buffer with

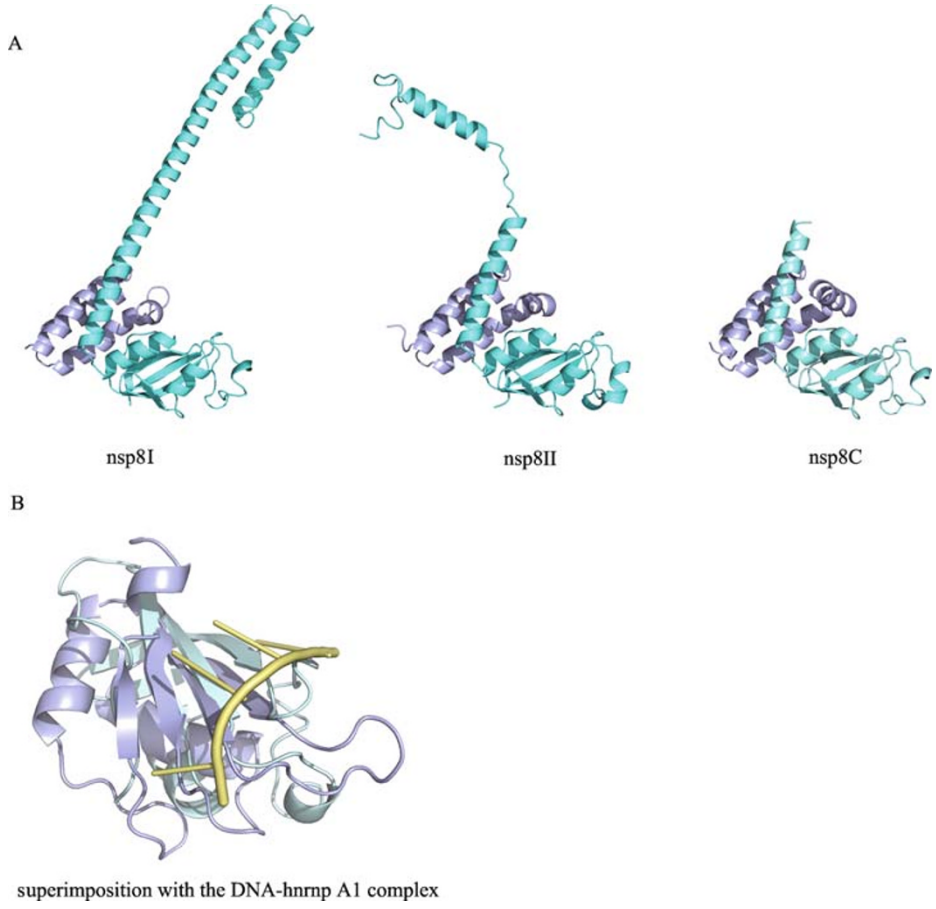

Figure 5. The alignment of nsp8 isoforms in complex with nsp7. (A) The alignment of three nsp8 isoforms in complex with nsp7. The protein structure was represented with cartoon model with nsp8 colored in cyan and nsp7 molecule colored in light blue. (B) The superimposition of nsp8 and DNA-hnrnp A1 complex. The nsp8C (with the N-helix truncated), hnrnp A1 and the DNA nucleotide were represented with cartoon model in cyan, light blue and yellow, respectively.

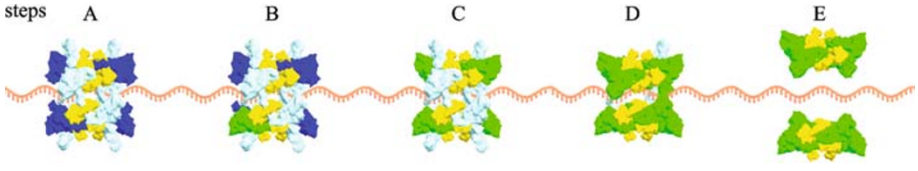

Figure 6. A model for nsp7-nsp8 complex during the viral life cycle. Nsp8I, cyan; nsp8II, blue; nsp7, yellow; nsp8C, green; viral genome, orange.

7 mM β-ME, and sonicated on ice for 25 min. The supernatant was collected after centrifugation of the sonicant at 15,000 rpm for 20 min. Affinity purification was achieved by two passes of the supernatant through 2 mL GST affinity media. In-column digestion lasted for 16 h at 4°C with PreScission Protease (GE Healthcare), and the protein of interest was harvested and concentrated to 20 mg/mL by an Amicon concentrator (cutoff value: 10 kDa). The nsp7-nsp8C complex was finally purified by gel filtration using a Superdex 200 column (GE Healthcare). Protein concentration was adjusted to 20 mg/mL for crystallization trials. The nsp8N protein was purified separately.

Crystallization and structure determination

Crystals of nsp7-nsp8C were grown in 0.2 M magnesium acetate and 12%–20% PEG 3350 by the hanging-drop vapor diffusion method at 16°C. Synchrotron X-ray diffraction data were collected on Beamline BL-5A of the Photon Factory (Tsukuba, Japan) and processed to 2.5 Å resolution using HKL2000 (Otwinowski and Minor, 1997) for data indexing and scaling. Molecular replacement using the nsp7-nsp8 supercomplex as a template was performed with PHASER (Terwilliger, 2000; Terwilliger and Berendzen, 1999). Manual rebuilding of the structure was performed using Coot (Emsley and Cowtan,

2004), and the structure was refined using REFMAC in the CCP4 suite (1994). Final modification was carried out using CNS (Brunger et al., 1998).

ACKNOWLEDGEMENTS

This work was supported by Project 973 of the Ministry of Science and Technology of China (Nos. 2006CB806503, 2007CB914301), and the National Natural Science Foundation of China (Grant Nos. 30221003, 30730022).

ABBREVIATIONS

a.a., amino acid; h.p.i., hours post infection; nsp, non-structural protein; PCBP2, poly(rC) binding protein; RdRp, RNA-dependent RNA polymerase; RTC, replication/transcription complex; SARS-CoV, severe acute respiratory syndrome coronavirus

REFERENCES

Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50, 760–763.

Brandt, C.R. (2005). The role of viral and host genes in corneal infection with herpes simplex virus type 1. Exp Eye Res 80, 607–621.

Brockway, S.M., Clay, C.T., Lu, X.T., and Denison, M.R. (2003). 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.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 54, 905–921.

Deming, D.J., Graham, R.L., Denison, M.R., and Baric, R.S. (2007). Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. J Virol 81, 10280–10291.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60, 2126–2132.

Graff, J., Cha, J., Blyn, L.B., and Ehrenfeld, E. (1998). Interaction of poly(rC) binding protein 2 with the 5' noncoding region of hepatitis A virus RNA and its effects on translation. J Virol 72, 9668–9675.

Graham, R.L., Sims, A.C., Baric, R.S., and Denison, M.R. (2006). The nsp2 proteins of mouse hepatitis virus and SARS coronavirus are dispensable for viral replication. Adv Exp Med Biol 581, 67–72.

Imbert, I., Guillemot, J.C., Bourhis, J.M., Bussetta, C., Coutard, B., Egloff, M.P., Ferron, F., Gorbalenya, A.E., and Canard, B. (2006). A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. EMBO J 25, 4933–4942.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In Macromolecular Crystallography, part A, C.W. Carter Jr., and R.M. Sweet, eds. (Academic Press), pp. 307–326.

Perera, R., Daijogo, S., Walter, B.L., Nguyen, J.H., and Semler, B.L. (2007). Cellular protein modification by poliovirus: the two faces of poly(rC)-binding protein. J Virol 81, 8919–8932.

Prentice, E., McAuliffe, J., Lu, X., Subbarao, K., and Denison, M.R. (2004). Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. J Virol 78, 9977–9986.

Sawicki, S.G., Sawicki, D.L., and Siddell, S.G. (2007). A contemporary view of coronavirus transcription. J Virol 81, 20–29.

Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martinez-Sobrido, L., Garcia-Sastre, A., Weber, F., and Kochs, G. (2007). The intracellular sites of early replication and budding of SARS-coronavirus. Virology 361, 304–315.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr D 56, 965–972.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr D 55, 849–861.

van der Meer, Y., Snijder, E.J., Dobbe, J.C., Schleich, S., Denison, M. R., Spaan, W.J., and Locker, J.K. (1999). Localization of mouse hepatitis virus nonstructural proteins and RNA synthesis indicates a role for late endosomes in viral replication. J Virol 73, 7641–7657.

Xu, X., Zhai, Y., Sun, F., Lou, Z., Su, D., Xu, Y., Zhang, R., Joachimiak, A., Zhang, X.C., Bartlam, M., *et al*. (2006). New antiviral target revealed by the hexameric structure of mouse hepatitis virus nonstructural protein nsp15. J Virol 80, 7909–7917.

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., *et al*. (2003). The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc Natl Acad Sci U S A 100, 13190–13195.

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

