Characterization of viral proteins encoded by the SARS-coronavirus genome

Yee-Joo Tan*, Seng Gee Lim, Wanjin Hong

Institute of Molecular and Cell Biology, 61 Biopolis Drive, Proteos 138673, Singapore

Received 7 September 2004; accepted 20 October 2004

Abstract

A new disease, termed severe acute respiratory syndrome (SARS), emerged at the end of 2002 and caused profound disturbances in over 30 countries worldwide in 2003. A novel coronavirus was identified as the aetiological agent of SARS and the 30 kb viral genome was deciphered with unprecedented speed in a coordinated manner by the global community. Since then, much progress has been made in the virological and molecular characterization of the proteins encoded by SARS-coronavirus (SARS-CoV) genome, which contains 14 potential open reading frames (ORFs). These investigations can be broadly classified into three groups: (a) studies on the replicase 1a/1b gene products which are important for viral replication, (b) studies on the structural proteins, spike, nucleocapsid, membrane and envelope, which have homologues in all coronaviruses, and are important for viral assembly and (c) expression and functional studies of the “accessory” proteins that are specifically encoded by SARS-CoV. A comparison of the properties of these three groups of SARS-CoV proteins with the knowledge that coronavirologists have generated over more than 30 years of research can help us in the prevention and treatment of SARS in the event of the re-emergence of this new infectious disease.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Severe acute respiratory syndrome (SARS); Coronavirus; Viral proteins

1. Introduction

The recent severe acute respiratory syndrome (SARS) epidemic, which affected over 30 countries across five continents, has profoundly disturbed social and economic activities regionally, as well as globally. By the end of the epidemic, which was essentially controlled by isolation, more than 8000 cases have been reported with more than 800 fatalities (World Health Organization, http://www.who.int/csr/sars/country/en/). The concerted efforts of the scientific community led to an extraordinarily rapid identification of a novel coronavirus as the aetiological agent of SARS and the full genome sequencing of the virus (Drosten et al., 2003; Fouchier et al., 2003; Ksiazek et al., 2003; Marra et al., 2003; Peiris et al., 2003; Rota et al., 2003; Ruan et al., 2003). The SARS-coronavirus (SARS-CoV) genome is ~30 kb in length and contains 14 potential open reading frames (ORFs) (Marra et al., 2003; Thié et al., 2003).

Coronaviruses are positive-strand RNA viruses and the virion consists of a nucleocapsid core surrounded by an envelope containing three membrane proteins, spike (S), membrane (M) and envelope (E) that are common to all members of the genus (Siddell, 1995; Lai and Holmes, 2001). The RNA is packaged by the nucleocapsid (N) protein into a helical nucleocapsid. The S protein, which forms morphologically characteristic projections on the virion surface, mediates binding to host receptors and membrane fusion. The M protein is a triple-spanning integral membrane protein with a short ectodomain and a large carboxyl-terminus endodomain. More recently, the E protein was shown to play a major role in coronavirus assembly (Bos et al., 1996; Vennema et al., 1996). The genes for these structural proteins and the replicase 1a/1b gene, that is located at the 5' end of the genome and constitute 2/3 of it, are conserved among the subgroups of coronavirus so is their relative posi...
2. Replicase gene (ORFs 1a and 1b)

Analogous to other coronaviruses, the first 2/3 of the SARS-CoV genome encodes the viral replicase genes (ORFs 1a and 1b), which translates into two large polyproteins, pp1a (486 kDa) and pp1ab (790 kDa) (Thiel et al., 2003). Expression of the ORF 1b-encoded region of pp1ab involves ribosomal frameshifting into the -1 frame just upstream of the ORF 1a translation termination codon (Thiel et al., 2003). Proteolytic processing of these polyproteins is mediated by viral cysteine proteinases and produces a minimum of 13 non-structural proteins (also called nsp’s), some of which are responsible for replicating the viral genome and/or generating a nested set of subgenomic mRNAs to express all the ORFs downstream of ORF 1b (Ziebuhr et al., 2000). Unlike most coronaviruses, which uses three proteinases for polyprotein processing (Ziebuhr et al., 2000; Gorbunyova, 2001), SARS-CoV is predicted only to have two proteinases, which are a papain-like (accessory) cysteine proteinase (termed PL2\textsuperscript{pp} or PL\textsuperscript{pp}), which cleaves at 3 sites, and a 3C-like (main) proteinase (termed 3CL\textsuperscript{pp} or M\textsuperscript{pp}), which cleaves at 11 sites (Rota et al., 2003; Gao et al., 2003a; Snijder et al., 2003; Thiel et al., 2003). As a result, 16 non-structural proteins (Fig. 1A) are predicted but which of these are essential for the replication of the virus remain to be determined. The proteinase activity of 3CL\textsuperscript{pp} was also experimentally demonstrated as purified 3CL\textsuperscript{pp} was shown to cleave peptides covering all the 11 predicted cleavage sites (Fan et al., 2004). In addition, the three-dimensional structure of 3CL\textsuperscript{pp} was also solved by both crystallography and NMR spectroscopy (Yang et al., 2003; Shu et al., 2004). Both studies reported that 3CL\textsuperscript{pp} exists as a dimer and revealed fine conformational details of its interaction with substrates, thus providing a basis for rational drug design.

Another protein that is likely to be important for viral replication is the SARS-CoV helicase (also called nsp13 in Snijder et al., 2003 or nsp10 in Gao et al., 2003a and Tanner et al., 2003). Recombinant SARS-CoV helicase has multiple enzymatic activities, including RNA helicase, DNA helicase, NTPase, dNTPase and an RNA 5’-triphosphatase activities (Tanner et al., 2003; Thiel et al., 2003; Ivanov et al., 2004). In addition, the crystal structure of SARS-CoV nsp9, which has no designated function, has been solved and it has been shown to bind RNA as well as another non-structural protein, SARS-CoV nsp8 (Campanacci et al., 2003; Sutton et al., 2004). The SARS-CoV nsp9 may have a similar function as the nsp9 protein of mouse hepatitis virus (MHV), a Group 2 coronavirus, which colocalized and interacted with other proteins of the replication complex (Host et al., 2000; Brockway et al., 2003). For the remaining non-structural proteins produced from pp1a or pp1ab, putative activities have been predicted based on the presence of functional domains in their sequences or by their structural similarities to other proteins (Gao et al., 2003a; Snijder et al., 2003; von Grothuss et al., 2003; Fig. 1A).

3. Structural proteins (S, E, M and N)

The S protein of coronavirus is important for binding to cellular receptor and for mediating the fusion of viral and host membranes, both of these processes being critical for virus entry into host cells (Gallagher and Buchmeier, 2001). As such, S is known to be responsible for inducing host immune responses and virus neutralization by antibodies (Holmes, 2003; Navas-Martin and Weiss, 2003). For SARS-CoV, it has been demonstrated that prior infection provided protective immunity in a mouse model and the passive transfer of neutralizing antibodies to naive mice also protected them from infection (Subbarao et al., 2004). Importantly, there was no enhancement of SARS-CoV infection in mice upon re-infection or after the administration of immune serum, unlike the case for one coronavirus, the feline infectious peritonitis virus (Olsen, 1993). A DNA vaccine encoding the S protein alone induced T cell and neutralizing antibody responses and protected mice from SARS-CoV infection (Yang et al., 2004), suggesting the S is indeed the primary target for viral neutralization in SARS-CoV infection. This finding was also confirmed by several studies that use surrogate/carrier viruses to express S in mice or primates (Gao et al., 2003b; Bisht et al., 2004; Buchholz et al., 2004; Bukreyev et al., 2004). From these studies, it is clear that humoral response against S plays an important role in controlling and clearing SARS-CoV infection.

In addition, a host cell receptor, the carboxypeptidase angiotensin-converting enzyme-2 (ACE-2), which is an essential regulator of heart function, has also been identified (Li et al., 2003). At least three independent laboratories subsequently showed that a domain in the N-terminus of S, approximately 300–510 amino acids, is the receptor binding domain (Xiao et al., 2003; Babcock et al., 2004; Wong et al., 2004). Importantly, syncytia formation/membrane fusion and viral replication can be specifically inhibited by an anti-ACE-2 an-
The coronavirus S protein is a class I virus fusion protein and contains two regions with a 4,3 hydrophobic (heptad) repeat in S2 domain or C-terminus half of the protein (de Groot et al., 1987; Bosch et al., 2003). These domains (termed as HR1 and HR2) are thought to play an important role in defining the oligomeric structure of S and hence mediating the fusion between viral and cellular membranes (Eckert and Kim, 2001). For the SARS-CoV, HR2 is located close to the transmembrane anchor (1148–1193 amino acids) and HR1 is ∼140 amino acids upstream of it (900–1005 amino acids) (Ingallinella et al., 2004). Biochemical studies have shown that peptides corresponding to the HR1 and HR2 of SARS-CoV S protein can associate into an anti-parallel six-helix bundles with structural features typical of class I fusion proteins, suggesting that the membrane fusion and cell entry mechanisms exploited by SARS-CoV are similar to that for other coronaviruses (Bosch et al., 2004; Ingallinella et al., 2004; Liu et al., 2004b; Tripet et al., 2004; Yuan et al., 2004; Zhu et al., 2004). In the full-length S protein, the HR1–HR2 structure brings the fusion peptide, predicted to be near the N-terminus of HR1 (Bosch et al., 2004), in close proximity to the transmembrane domain and this drives the fusion between viral and cellular membranes, and allows the virus to enter the cell. Indeed, peptides from HR1 or HR2 can inhibit...
viral replication in Vero E6 culture, presumably by interfering with the formation of the six-helix bundle (Bosch et al., 2004; Liu et al., 2004b; Yuan et al., 2004; Zhu et al., 2004).

Studies on the profile of antibodies in SARS patients showed that antibodies against M and E are generally low or not present in SARS patient’s sera (Wang et al., 2003; Guo et al., 2004; Leung et al., 2004; Tan et al., 2004b) as these proteins are embedded in the viral envelope. However, it is clear that M and E are important for viral assembly as demonstrated by the formation of virus-like particles in insect cells expressing these proteins (Ho et al., 2004). This result is consistent with previous studies on coronaviruses, which showed that M and E are sufficient for the assembly of viral particles (Bos et al., 1996; Vennema et al., 1996). By using a proteomic approach, a novel phosphorylated site of M was also identified (Zeng et al., 2004a), but the importance of this for the function of M has not been defined.

The N protein has been shown to be very abundant in SARS-CoV infected Vero E6 cells (Krokhin et al., 2003; Rota et al., 2003) and several independent studies have shown that >90% of sera obtained from convalescent SARS patients have antibodies against N (Shi et al., 2003; Wang et al., 2003; Guo et al., 2004; Leung et al., 2004; Tan et al., 2004b). In addition, as N is not glycosylated, easily expressed in bacteria and highly immunogenic, it is an ideal candidate for development of enzyme-linked immunosorbent assays for the detection of SARS-CoV infection, either for detection of anti-N antibodies (Shi et al., 2003; Guan et al., 2004a) or for direct antigen detection (Che et al., 2004; Lau et al., 2004). In addition, it was reported that the SARS-CoV N can induce specific T-cell responses (Gao et al., 2003b; Kim et al., 2004), as have been observed with other coronaviruses (Siddell, 1995), but how important is this for protective immunity remains to be determined.

Other molecular aspects of N have also been reported, including self-dimerization (He et al., 2004; Surjit et al., 2004a), RNA-binding capabilities (Huang et al., 2004), cleavage by caspase 3 (Ying et al., 2004) and its ability to activate signal transduction pathways (He et al., 2003). In addition, the N protein of SARS-CoV was shown to induce apoptosis and actin reorganization in mammalian cells under stressed conditions (Surjit et al., 2004b). Interestingly, Mizutani et al. (2004) showed that the p38MARK pathway is activated in SARS-CoV infected Vero E6 cells, but it is not clear if this is directly/entirely due to the expression of N. The N proteins of other coronaviruses, including avian infectious bronchitis virus, porcine transmissible gastroenteritis virus and mouse hepatitis virus, are localized to both cytoplasm and nucleolus, and the presence of N in the nucleolus may be important for the synthesis of viral RNA (Hiscox et al., 2001; Wurm et al.,...
Y.-J. Tan et al. / Antiviral Research 65 (2005) 69–78

2001). For SARS-CoV N protein, it has been reported to be found in the cytoplasm and nucleus of SARS-CoV infected cells (Chang et al., 2004; Zeng et al., 2004a).

4. Group-specific genes

Eight subgenomic mRNAs are produced in SARS-CoV infected Vero E6 cells and these are used to express the ORFs besides the replicase 1a/1b (Snijder et al., 2003; Thiel et al., 2003). These include the S (ORF 2), E (ORF 4), M (ORF 5) and N (ORF 9) and another eight ORFs that encode putative proteins with no significance sequence homology to viral structural proteins, S, hemagglutinin-esterase (HE) and E, have been shown to be transported to the plasma membrane/cell surface (Kienzle et al., 1990; Parker et al., 1990; Smith et al., 1990; Vennema et al., 1990). Indeed, Zeng et al. (2004b) could detect disulfide-linked complexes of S and 3a in the medium of SARS-CoV infected cells, indicating that 3a was secreted together with S, possibly through the formation of virus particles. However, it is necessary to confirm this finding with highly purified SARS-CoV virions as viral proteins could also be released into the medium through cell lysis.

Mutations in 3a were observed in a Singapore isolate after three passages in cell culture, resulting in several different forms of 3a in infected cells (Tan et al., 2004c). This may not necessarily be a cell culture adaptation, but may indicate the presence of quasi-species, as comparative analysis of the different isolates of SARS-CoV also showed high frequency of mutations in the 3a gene (Chen et al., 2003; Tan et al., 2004b; Yeh et al., 2004; Zeng et al., 2004b). In fact, sequence comparison of isolates from different clusters of infection showed that both S and 3a showed positive selections during virus evolution, implying that these proteins play important roles in the virus life cycle and/or disease development (3a was termed as X1 in Guan et al., 2004b; Yeh et al., 2004; Zeng et al., 2004b).

The other ORF that has been shown to be expressed in SARS-CoV infected cells is 7a, which contains a cleavable signal peptide at the N-terminus and a transmembrane domain near the C-terminus (7a was also known as ORF 8 in Marra et al., 2003, as X4 in Rota et al., 2003 and as U122 in Fielding et al., 2004; Tan et al., 2004c). An endoplasmic reticulum (ER) retrieval motif (KRKTE), which is important for transport of proteins back to the ER (Teasdale and Jackson, 1996), is located at the C-terminus cytoplasmic domain of 7a and mediates the recycling of 7a between the ER and Golgi apparatus such that 7a is present in the intermediate compartments, where coronaviruses are known to assemble and bud (Fielding et al., 2004). Interestingly 7a can also interact with 3a, which can interact with M, E and S, suggesting that these viral proteins may form complexes during infection (Tan et al., 2004c).

In addition, the over-expression of 7a induces apoptosis via a caspase-dependent pathway, and in cell-lines derived from different organs, including lung, kidney and liver (Tan et al., 2004a). Although there are other factors that contribute to the induction of apoptosis during SARS-CoV infection, the
et al., 1990; Weiss et al., 1993). Shen et al. (2003) also reported an insertion of six nucleotides in 3b while in others, ORF 2a is completely deleted (Schwarz 2003b, while in others, ORF 2a is completely deleted (Schwarz et al., 1995). Recent studies also showed that some of these group-specific genes are not essential for viral replication in cell culture systems but may be important for viral-host interactions and thus contribute to viral stability and/or pathogenesis in vivo. For example, although the 7b gene of feline coronavirus is easily lost upon virus adaptation to cell culture, it is strictly maintained in naturally occurring strains and its loss was correlated with reduced virulence (Herrwegh et al., 1995). Recent studies also showed that some of these group-specific genes are not essential for viral replication in cell culture, but their deletion, by reverse genetics, is attenuating in the natural host (de Haan et al., 2002; Ortego et al., 2003). It has not yet been established which of these SARS-CoV unique ORFs are essential for viral replication and/or for viral-host interactions.

5. Future directions

Thus far, characterization of the three groups of viral proteins of the SARS-CoV revealed that for the first two groups of proteins, i.e. the replicase gene produced from cleavage of the polyprotein pp1a/pp1ab and structural proteins (S, M, E and N), there are significant similarities in their properties when compared to their homologues in other coronaviruses. Nevertheless, detailed analyses have revealed important differences but whether these contributed to severe clinical manifestations of SARS-CoV infection in contrast to the mild diseases caused by most human or animal coronaviruses, remains to be determined.

As for the third group of viral proteins, i.e. the group-specific genes encoded by the sequences between S and N, ability of 7a to induce apoptosis in different cell types is consistent with the clinical observation of apoptosis in different organs infected by SARS-CoV and suggests that the expression of 7a during infection may be one of the underlying mechanisms for the pathogenesis of SARS-CoV infection.

Group-specific genes in coronaviruses, also called “accessory” proteins, are usually dispensable for viral replication in cell culture systems but may be important for viral-host interactions and thus contribute to viral stability and/or pathogenesis in vivo. For example, although the 7b gene of feline coronavirus is easily lost upon virus adaptation to cell culture, it is strictly maintained in naturally occurring strains and its loss was correlated with reduced virulence (Herrwegh et al., 1995). Recent studies also showed that some of these group-specific genes are not essential for viral replication in cell culture, but their deletion, by reverse genetics, is attenuating in the natural host (de Haan et al., 2002; Ortego et al., 2003). It has not yet been established which of these SARS-CoV unique ORFs are essential for viral replication and/or for viral-host interactions.

However, it is interesting to note that the part of viral genome which encodes these “accessory” proteins appears to be prone to large insertions or deletions. For example, in some strains of MHV, ORF 4 is interrupted and becomes ORFs 4a and 4b, while in others, ORF 2a is completely deleted (Schwarz et al., 1990; Weiss et al., 1993). Shen et al. (2003) also reported an insertion of six nucleotides in 3b gene of the infectious bronchitis virus after continuous passages in Vero E6 cells and this resulted in a truncated 3b protein of 34 amino acids with only the first 17 amino acids being homologous to the original full-length 3b protein of 64 amino acids. For SARS-CoV, Guan et al. (2003) analyzed SARS-CoV isolates obtained from animals in a live-market in Guangdong and found that all the animal isolates contain a 29-nucleotides sequence which is absent in most human isolates. As a result of this, the ORF 8a (termed as ORF 10 in Guan et al., 2003) and 8b (termed as ORF 11 in Guan et al., 2003) in the human isolates becomes one ORF encoding a protein of 122 amino acids, whose N-terminus is identical to ORF 8a and C-terminus is identical to ORF 8b. Another extensive study of 63 SARS-CoV isolates obtained from early, middle and late phases of the SARS epidemic in China also showed that there are major deletions in this region of the viral genome (The Chinese SARS Molecular Epidemiology Consortium, 2004). Interestingly, the clustering of patients with different patterns of deletion was correlated with the different phases of the epidemic. Although these mutations do not appear to have any adverse effect on the survival of the virus, it is conceivable that the different variants of ORF 8a/8b have different stabilities and/or functions, and hence would contribute differently to viral pathogenesis in vivo. Another study also reported an in-frame deletion of 45 nucleotides in ORF 7b in the Frankfurt isolate after three passages in tissue culture (Theel et al., 2003). Since SARS-CoV appeared to have only recently crossed species from animal to human and that it has been observed that the profiles of these mutations correlated with clusters of infection, it is likely that the properties of these unique ORFs are changing as the virus undergoes adaptive evolution and these mutations could cause SARS-CoV infection to become more benign but more persistent.

As a full-length infectious clone of SARS-CoV has been assembled (Yount et al., 2003), the use of reverse genetics would certainly reveal more about the contributions of individual viral proteins. Although SARS-CoV caused little or no disease in mice, it is able to replicate in the respiratory tract of the mice to a high level (Subbarao et al., 2004; Wentworth et al., 2004). Together with the technologies to create transgenic or knockout mice, the mouse model will contribute to the developments of vaccines and anti-viral therapeutics against SARS-CoV infection and may also help us understand why certain cohorts of patients are more vulnerable to the disease while others only develop mild symptoms. In addition, experimentally SARS-CoV infected cats, ferrets and primates can develop at least some of the clinical symptoms observed in SARS-CoV infected patients (Martina et al., 2003; Kuiken et al., 2004). Thus, these animal models will be essential for studying virus-host interactions and for delineating the precise contributions of the viral proteins to SARS-CoV infection, replication and pathogenesis.

Acknowledgements

We thank members of the Collaborative Anti-Viral Research group, Institute of Molecular and Cell Biology, for critically reading the manuscript. We apologize to any investigators whose work we have inadvertently omitted.

References

coronavirus spike protein are required for interaction with receptor. J. Virol. 78, 4325–4329.


