The 29-Nucleotide Deletion Present in Human but Not in Animal Severe Acute Respiratory Syndrome Coronavirus Disrupts the Functional Expression of Open Reading Frame 8

Monique Oostra, Cornelis A. M. de Haan, and Peter J. M. Rottier*

Virology Division, Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

Received 26 July 2007/Accepted 27 September 2007

One of the most striking and dramatic genomic changes observed in the severe acute respiratory syndrome coronavirus (SARS-CoV) isolated from humans soon after its zoonotic transmission from palm civets was the acquisition of a characteristic 29-nucleotide deletion. This occurred in open reading frame 8 (ORF8), one of the accessory genes unique to the SARS-CoV. The function of ORF8 and the significance of the deletion are unknown. The intact ORF8 present in animal and some early human isolates encodes a 122-amino-acid polypeptide (8ab\textsuperscript{*}), which we expressed in cells using the vaccinia virus T7 expression system. It was found to contain a cleavable signal sequence, which directs the precursor to the endoplasmic reticulum (ER) and mediates its translocation into the lumen. The cleaved protein became N-glycosylated, assembled into disulfide-linked homomultimeric complexes, and remained stably in the ER. The 29-nucleotide deletion splits ORF8 into two ORFs, 8a and 8b, encoding 39- and 84-residue polypeptides. The 8a polypeptide is likely to remain in the cytoplasm, as it is too small for its signal sequence to function and will therefore be directly released from the ribosome. However, we could not confirm this experimentally due to the lack of proper antibodies. ORF8b appeared not to be expressed in SARS-CoV-infected cells or when expressed from mRNA's mimicking mRNA8. This was due to the context of the internal AUG initiation codon, as we demonstrated after placing the ORF8b immediately behind the T7 promoter. A soluble, unmodified and monomeric 8b protein was now expressed in the cytoplasm, which was highly unstable and rapidly degraded. Clearly, the 29-nucleotide deletion disrupts the proper expression of the SARS-CoV ORF8, the implications of which are discussed.
SARS-CoV replication in cell culture and in a mouse model, though the effect in mice was hard to evaluate as the wild-type virus infection did not elicit clear signs of disease or pathology (47).

Despite their apparent importance in virus-host interactions, the functions encoded by the coronaviral group-specific genes are still largely unknown. For the SARS-CoV, however, some of the group-specific proteins, which show no sequence homology to other (corona)viral or cellular proteins, have been studied quite extensively, and several functions have been assigned to these proteins. Three of the SARS-CoV group-specific proteins—i.e., the 3a (14, 36), 7a (13), and 7b (34) proteins—appeared to be structural proteins as they were incorporated into SARS-CoV virions. Furthermore, overexpression of the 7a and the 3b proteins was found to induce apoptosis and cell cycle arrest at the G0/G1 phase (39, 48, 49). The 3b protein and protein 6 appeared to function as interferon antagonists (16). Interestingly, when incorporated into the genome of an attenuated mouse hepatitis coronavirus, the gene encoding protein 6 significantly enhanced the virulence of the recombinant virus in mice (33).

One of the most intriguing and still unresolved questions regarding the SARS epidemic is the origin of the virus. While there is no doubt that it has been transmitted as a zoonosis, its immediate animal source has not been firmly established. SARS-CoV-like viruses have been isolated from civet cats, raccoon dogs, and wild bats (10, 21) living in the area where the SARS epidemic started, and these animals probably form the reservoir from which the virus crossed the host-species barrier and spread into the human population. As soon as the first comparative sequence data became available, particular attention was drawn by observations in the group-specific ORF8a/b region (10). Viruses isolated from animals as well as some early-stage human isolates were found to possess a single continuous ORF8, while in the middle or late phase of the epidemic the isolated human strains contained a 29-nucleotide (nt) deletion that created two ORFs, designated ORF8a and ORF8b (4, 10, 21) (Fig. 1A). Mutations were also found in the S protein, which is responsible for receptor binding and entry, and these mutations were probably important for the species transmission or were the result of adaptation to the human host. However, only the viruses containing the deletion in the ORF8 region were isolated later during the epidemic, suggesting that only these viruses were able to spread efficiently from human to human.

In view of its potential importance in the epidemiology and pathogenicity of SARS-CoV, the ORF8 genomic region is the focus of the present study. By analyzing the expression, cellular localization, and membrane association of the various ORF8-related products, we elucidated the basic features of the proteins as they are expressed from this region in the context of the different SARS-CoVs. Our results indicate that the full-
length protein, designated 8ab\textsuperscript{\textbullet}, encoded by ORF8 in the animal virus isolates is a functional protein that is lost upon transmission to the human population, leaving two probably nonfunctional proteins 8a and 8b. Our conclusions appear to be inconsistent with observations by others on the biological functions of the 8a and 8b proteins (3, 15, 22).

\section*{MATERIALS AND METHODS}

Cells, viruses, and antibodies. OST7-1 cells (obtained from B. Moss) (7) and Vero E6 cells (obtained from E. Snijder) were maintained as monolayer cultures in Dulbecco's modified Eagle's medium (DMEM) (Cambrex Bio Science) containing 10% fetal calf serum (FCS) (Bodinco V.), 100 IU of penicillin, and 100 \mu g of streptomycin per ml.

Recombinant vaccinia virus encoding the bacteriophage T7 RNA polymerase (vTF7-3) was obtained from B. Moss (8). The SARS-CoV (strain 5088) was kindly provided by B. Haagmans and A. Osterhaus (20).

The rabbit polyclonal antiserum against enhanced green fluorescent protein (EGFP) and calreticulin were obtained from ICL and Sigma, respectively. The goat antiserum recognizing the SARS-CoV membrane (Ms) protein was kindly provided by Y.-J. Tan (40), whereas the ferret antiserum against the complete SARS-CoV 

\begin{table}[ht]
\centering
\caption{Sequence, polarity, and purpose of primers used in this study}
\begin{tabular}{llll}
\hline
Primer no. & Sequence (5' to 3') & Polarity & Purpose \\
\hline
2220 (8a rev SOE) & cacattaagttgtaaagcagtggAACAGGATCTCTAAGCACT & + & 8ab\textsuperscript{\textbullet} \\
2221 (8b for SOE) & ctgtaacactaatgaatAAGTGACAACACTAGGG & + & 8ab\textsuperscript{\textbullet} \\
2267 (8b for) & ggatccATTTGTTCGTTTATTTAAAAC & + & 8b-EGFP \\
2985 (8a for) & gaattcacccATGAACTTCTCATTGTTTT & + & 8a-EGFP, 8ab\textsuperscript{\textbullet}-EGFP, 8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP \\
3043 (8a for -ss) & ccattcaggttggtaaccagtaggACAAGGATCTTCAAGCACAT & + & 8b\textsuperscript{-}\textsuperscript{\textbullet}-EGFP, 8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP \\
3121 (EGFP\textsuperscript{\textbullet} for) & GCGACGTTAAACGGCacCAAGTTCAGCGTG & + & N-glycosylation site in EGFP \\
3123 (EGFP\textsuperscript{\textbullet} rev) & CACGCTGAACTGTTCGTTTATTT & - & 8b\textsuperscript{-}\textsuperscript{\textbullet}-EGFP \\
\hline
\end{tabular}
\end{table}

\textsuperscript{a} for, forward primer; rev, reverse primer.

\textsuperscript{b} Coding sequences are shown in uppercase. Lowercase letters indicate nucleotides added for cloning purposes, with the restriction enzyme recognition sites underlined.

Subsequently, the 8ab\textsuperscript{\textbullet} gene including the stop codon was cloned into the pTUG31 vector in fusion with the EGFP gene. The ORF8 fragments were obtained by restriction with EcoRI and BamHI from the pGEM-T Easy vector, and the sequences were confirmed by sequence analysis. The leader-TRS-ORF8ab\textsuperscript{-EGFP} construct was created by performing site-directed mutagenesis on pEGFP-N3 (Clontech) using BamHI and NotI, with the latter restriction site filled in with Klenow polymerase (Invitrogen). These fragments were cloned together into the EcoRI-Smal-digested pTUG31 vector, creating pTug8ab-EGFP, pTug8b-EGFP, pTug8ab\textsuperscript{-EGFP}, and pTug8b\textsuperscript{-EGFP}, which encode fusion proteins of the different ORF8 products with EGFP.

For the 8ab\textsuperscript{-EGFP} construct containing the leader and transcription-regulatory sequence (TRS) in front of the stop codon, an SOE PCR was performed. The leader-TRS fragment was obtained by RT-PCR on the 5' end of the isolated viral RNA using primers 3207 and 3208 (Table 1), and the TRS-ORF8ab\textsuperscript{-EGFP} fragment was obtained by RT-PCR using primers 3209 and 2986. The two fragments were annealed and amplified by PCR using primers 3207 and 2986. The primers contain 5' extensions introducing either an EcoRI or a BamHI restriction enzyme recognition site (Table 1, underlined), while additionally the stop codon is deleted from the coding sequence. The PCR product was cloned into the pGEM-T Easy vector (Promega), and the sequence was confirmed by sequence analysis. The leader-TRS-ORF8ab\textsuperscript{-EGFP} fragment was obtained by restriction with EcoRI and BamHI from the pGEM-T Easy vector, while the EGFP fragment was excised from the pEGFP-N3 vector (Clontech) using BamHI and NotI, with the latter restriction site filled in with Klenow polymerase (Invitrogen). The two fragments were cloned into the EcoRI-Smal-digested pTUG31 vector, creating pTug8ab-EGFP, pTug8b-EGFP, pTug8ab\textsuperscript{-EGFP}, and pTug8b\textsuperscript{-EGFP}, which encode fusion proteins of the different ORF8 products with EGFP.

The coding sequence for the first 15 amino acids of the 8ab\textsuperscript{\textbullet} sequence were deleted by performing PCR amplification on pTug8ab\textsuperscript{-EGFP} using primers 3043 and 2268 (Table 1). Both primers contain a 5' extension introducing either an EcoRI or a BamHI restriction enzyme recognition site (Table 1, underlined) while additionally a new start codon is created. The PCR product was cloned into the pGEM-T Easy vector (Promega), and the sequence was confirmed by sequence analysis. The 8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP fragment (8ab\textsuperscript{-}\textsuperscript{\textbullet} lacking the 15-amino-acid signal sequence [ss]) was obtained from the pGEM-T Easy vector by restriction with EcoRI and BamHI and cloned into the pTUG31 vector digested with the same enzymes, resulting in construct pTug8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP. An EcoRI-KpnI fragment was obtained from this construct and used to replace the EcoRI-KpnI fragment in the pTUG31 vector, thereby creating pTug8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP construct containing the leader and transcription-regulatory sequence (TRS) in front of the stop codon, an SOE PCR was performed. The leader-TRS fragment was obtained by RT-PCR on the 5' end of the isolated viral RNA using primers 3207 and 3208 (Table 1), and the TRS-ORF8ab\textsuperscript{-EGFP} fragment was obtained by RT-PCR using primers 3209 and 2986. The two fragments were annealed and amplified by PCR using primers 3207 and 2986. The primers contain 5' extensions introducing either an EcoRI or a BamHI restriction enzyme recognition site (Table 1, underlined), while additionally the stop codon is deleted from the coding sequence. The PCR product was cloned into the pGEM-T Easy vector (Promega), and the sequence was confirmed by sequence analysis. The leader-TRS-ORF8ab\textsuperscript{-EGFP} fragment was obtained by restriction with EcoRI and BamHI from the pGEM-T Easy vector, while the EGFP fragment was excised from the pEGFP-N3 vector (Clontech) using BamHI and NotI, with the latter restriction site filled in with Klenow polymerase (Invitrogen). The two fragments were cloned into the EcoRI-Smal-digested pTUG31 vector, creating pTug8ab-EGFP, pTug8b-EGFP, pTug8ab\textsuperscript{-EGFP}, and pTug8b\textsuperscript{-EGFP}, which encode fusion proteins of the different ORF8 products with EGFP.

The coding sequence for the first 15 amino acids of the 8ab\textsuperscript{\textbullet} sequence were deleted by performing PCR amplification on pTug8ab\textsuperscript{-EGFP} using primers 3043 and 2268 (Table 1). Both primers contain a 5' extension introducing either an EcoRI or a BamHI restriction enzyme recognition site (Table 1, underlined) while additionally a new start codon is created. The PCR product was cloned into the pGEM-T Easy vector (Promega), and the sequence was confirmed by sequence analysis. The 8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP fragment (8ab\textsuperscript{-}\textsuperscript{\textbullet} lacking the 15-amino-acid signal sequence [ss]) was obtained from the pGEM-T Easy vector by restriction with EcoRI and BamHI and cloned into the pTUG31 vector digested with the same enzymes, resulting in construct pTug8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP. An EcoRI-KpnI fragment was obtained from this construct and used to replace the EcoRI-KpnI fragment in the pTUG31 vector, thereby creating pTug8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP construct, thereby creating pTug8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP, which encodes the 8ab\textsuperscript{-}\textsuperscript{\textbullet}-EGFP fusion protein without the ss. An overview of all ORFS-related constructs is shown in Fig. 1C.

The EGFP tag containing an N-glycosylation consensus sequence (EGFP\textsuperscript{\textbullet}) was created by performing site-directed mutagenesis on pEGFP-N3 (Clontech) with primers 3212 and 3213 to mutate the histidine at position 26 to a threonine.
were then heated at 95°C for 1 min. a three-times-concentrated solution of LSB was added to the samples, which the final immunoprecipitation pellets were suspended in PBS instead of LSB, 2-glycosidase F (PNGase F) or endoglycosidase H (endo H). To this end, N by SDS-polyacrylamide gel electrophoresis (PAGE) using 15% polyacrylamide [pH 7.4], 150 mM NaCl, 0.1% SDS, 1% sodium deoxycholate). The mixtures were incubated overnight at 4°C. The immune complexes were ad-
mixed with SARS-CoV strain 5688 (20) at a multiplicity of infection of 1 for 1 h, after which the medium was replaced by a transfection mixture consisting of 0.5 ml of DMEM without FCS but containing 10 μl of Lipofectin (Invitrogen) and 5 μg of each selected construct. After a 5-min incubation at room temperature, 0.5 ml of DMEM was added, and incubation was continued at 37°C. Three hours postinfection, the medium was replaced by culture medium and tunicamycin (5 μg/ml) was added to the medium, as indicated in Fig. 4. 

Immunofluorescence microscopy. Vero E6 or OST-71 cells grown on glass coverslips were fixed with 3% paraformaldehyde for 1 h at room temperature at the times postinfection or posttransfection indicated in the figure legends. The fixed cells were washed with PBS and permeabilized using 0.1% Triton X-100 for 10 min at room temperature. The permeabilized cells were washed with PBS and incubated for 15 min in blocking buffer (PBS–10% normal goat serum), followed by a 45-min incubation with antibodies directed against SARS-CoV, 8ab, or calreticulin. After three washes with PBS–0.05% Tween-20, the cells were incubated for 45 min with either fluorescein isothiocyanate-conjugated goat anti-ferret immunoglobulin G antibodies (KPL) or Cy5-conjugated donkey anti-rabbit immunoglobulin G antibodies (Jackson Laboratories). After three washes with PBS–0.05% Tween-20 and one with PBS, the samples were mounted on glass slides in FluorSave (Calbiochem). The samples were examined with a confocal fluorescence microscope (Leica TCS SP2).

Metabolic labeling and immunoprecipitation. Prior to labeling, the cells were starved for 30 min in cysteine- and methionine-free modified Eagle’s medium containing 10 mM HEPES (pH 7.2) and 5% dialyzed FCS. This medium was replaced by 1 ml of similar medium containing 100 μCi of 35S in vitro cell-labeling mixture or 35S-labeled cysteine (Amersham), after which the cells were further incubated for the indicated time periods. After pulse-labeling, where indicated, the radioactivity was chased using culture medium containing a 2 mM concentration (each) of unlabeled methionine and cysteine. After pulse-labeling or chase, the cells were washed once with PBS containing 50 mM Ca2+ and 50 mM Mg2+ and then lysed on ice in 1 ml of lysis buffer (0.5 mM Tris [pH 7.3], 1 mM EDTA, 0.1 M NaCl, 1% Triton X-100) per 10-cm2 dish. The lysates were cleared by centrifugation for 5 min at 15,000 rpm and 4°C.

In vitro transcription and translation reactions were performed using the TNT coupled reticulocyte lysate system from Promega, according to the manufacturer’s instructions, in the presence of 35S in vitro cell-labeling mixture or 35S-labeled cysteine (Amersham), either with or without the use of canine microsomal membranes (Promega).

Radioimmunoprecipitations were essentially performed as described previ-
ously (29); 200-μl aliquots of the cell lysates or 5 μl of in vitro translation reaction mixtures were diluted in 1 ml of detergent buffer (50 mM Tris [pH 8.0], 625 mM EDTA, 1% NP-40, 0.4% sodium deoxycholate, 0.1% sodium dodecyl sulfate [SDS]) containing antibodies (3 μl of rabbit anti-EGFP serum or rabbit anti-M, serum or 25 μl of rabbit anti-Sab-1 serum) containing mixtures were incubated overnight at 4°C. The immune complexes were ad-
sorbed to Pansorbin cells (Calbiochem) for 60 min at 4°C and were subsequently collected by centrifugation. The pellets were washed three times with resuspension and centrifugation using radioimmunoprecipitation assay buffer (10 mM Tris [pH 7.4], 150 mM NaCl, 0.1% SDS, 1% NP-40, 1% sodium deoxycholate). The final pellets were suspended in Laemmli sample buffer (LSB), where indicated without β-mercaptoethanol (Fig. 7), and heated at 95°C for 1 min before analysis by SDS-polyacrylamide gel electrophoresis (PAGE) using 15% polyacrylamide gels.

Where indicated (Fig. 4 and 6), immunoprecipitates were treated with peptide-N-glycosidase F (PNGase F) or endoglycosidase H (endo H). To this end, the immunoprecipitated pellets were suspended in PBS instead of LSB, 2 μl of PNGase F or endo H (New England Biolabs) was added and the samples were incubated at 37°C for 2 h. Before analysis by SDS-PAGE, a 0.5 volume of a three-times-concentrated solution of LSB was added to the samples, which were then heated at 95°C for 1 min.

Sodium carbonate extraction. The sodium carbonate membrane fractionation method was adapted from procedures described previously (44). The proteins were expressed using the TNT coupled reticulocyte lysate system from Promega in the presence of canine microsomal membranes (Promega), after which the samples were mixed 1:1 with either 0.1% Tris-buffered saline (25 mM Tris-HCl [pH 7.5], 137 mM NaCl, 5 mM KCl, 0.7 mM CaCl2, 0.5 mM MgCl2, and 0.6 mM Na2HPO4) or 200 mM Na2CO3 (pH 11). The samples were laid on top of a sucrose gradient consisting of 2 M and 0.2 M sucrose. Microsomal membranes were pelleted by centrifugation in a Beckman TL100 rotor at 65,000 rpm for 30 min at 4°C. The pellet fractions were lysed using 5× lysis buffer (2.5 mM Tris [pH 7.3], 5 mM EDTA, 0.5 M NaCl, 5% Triton X-100) for 15 min at 4°C, after which both the supernatant and pellet fractions were subjected to immunoprecipitation as described above.

RESULTS
Expression of the ORF8 products. In most human SARS-CoV isolates the subgenomic mRNA 8 contains two ORFs at its 5′ end. The first one, ORF8a, is very small, coding for only 39 amino acids. The second ORF, 8b, encodes a protein of 84 amino acids but does not contain a TRS for the production of its own mRNA (41). Its expression from mRNA 8 would re-
quire translation initiation from an internal AUG. To examine the synthesis and properties of the ORF8 products, constructs were made for expression using the recombinant vaccinia virus bacteriophage T7 RNA polymerase (vTF7-3) expression system. These constructs contained either the ORF8 sequence as it is found in viruses isolated from civet cats, designated 8ab+ and containing one continuous ORF, or the sequence as it is found in human virus isolates, designated 8ab− and having a 29-nt deletion giving rise to two ORFs. Constructs were also made that contain only the ORF8a sequence or the ORF8b sequence placed directly behind the T7 promoter. All gene sequences were 3′ terminally fused to the EGFP gene for easy detection. Expressions using this vaccinia virus system are highly efficient. A potential internal ribosomal entry site (IRES) present in front of the ORF8b sequence in human virus isolates would be expected to lead to the expression of the 8b protein from the 8ab− construct in this system.

OST-71 cells were infected with vTF7-3, transfected with each of the above-described constructs, and labeled with 35S-methionine for 1 h starting at 5 h postinfection (p.i.). The cells were lysed and processed for immunoprecipitation with a rabbit polyclonal antiserum directed to the EGFP tag, and the samples were analyzed by SDS-PAGE. The 8ab+, EGFP, 8a-EGFP, and 8b-EGFP products derived from the constructs where these genes were located directly behind the T7 promoter could be readily detected in the gel. They appeared to migrate at a rate slightly slower than or according to their predicted molecular masses, which are 41, 37 and 31 kDa for the 8ab+, 8b, and 8a proteins, respectively (Fig. 2A and data not shown). However, no expression of the 8b-EGFP fusion protein was observed from the construct containing ORF8a in front of ORF8b (8ab−). Also, when the SARS-CoV leader and TRS had been cloned in front of the ORFs in this construct to make the 5′ end of the produced RNA resemble that of sub-
genomic mRNA 8, no expression of 8b-EGFP could be de-
tected (Fig. 2A). The expression of 8ab−-EGFP was not af-
fected by cloning of the leader and TRS in front of this gene (Fig. 2A). It thus appears that the 8ab− sequence as it is found in human virus isolates does not contain an IRES for the expression of the second ORF, ORF8b.
Next, the expression of the ORF8b product was examined in virus-infected cells. Vero E6 cells were infected with SARS-CoV strain 5688, a human isolate having the 29-nt ORF8 deletion. The cells were fixed at 8 h p.i. and processed for immunofluorescence microscopy using a polyclonal serum from a SARS-CoV-infected ferret and rabbit polyclonal antiserum against either the 8b or the M protein. A large number of SARS-CoV-positive cells could be detected at this time point after infection (Fig. 2B). In all infected cells the M protein was also detected, and this staining colocalized with that of the ferret serum. However, no positive signal could be detected using the antiserum against the 8b protein in any of the infected cells. As a positive control the 8b gene product was expressed in parallel from the construct with the ORF8b-EGFP fusion product directly behind the T7 promoter, using the vaccinia virus T7 expression system. The infected and transfected OST7-1 cells were fixed at 6 h p.i. and processed for immunofluorescence microscopy using the rabbit polyclonal antiserum against the 8b protein. Here, all the EGFP-positive cells also stained positive with the 8b antiserum (Fig. 2B), confirming that the antiserum does recognize the protein in this type of assay. Therefore, it can be concluded that the 8b protein is not, or only very inefficiently, expressed in virus-infected cells. This result is consistent with that of the immunoprecipitation carried out on cell lysates after vTF7-3-mediated expression of the 8ab\(^a\) construct.

The 8b gene product could be expressed in the vaccinia virus system only when cloned directly behind the T7 promoter, i.e., without ORF8a in front of it. The expression and stability of this protein were compared to those of the full-length 8ab\(^a\) protein that is encoded by the animal ORF8 sequence. OST7-1 cells were infected with vTF7-3, transfected with the construct containing either the 8b-EGFP or the 8ab\(^a\)-EGFP sequence, and labeled with [\(^{35}\)S]methionine for 1 h starting at 5 h p.i. Subsequently, the cells were either lysed directly or the radioactivity was chased for 2 h before the cells were lysed. The cell lysates were processed for immunoprecipitation with an antiserum directed to the EGFP tag, and the samples were analyzed by SDS-PAGE. Both proteins were expressed at similar levels during the pulse-labeling. After the chase the 8ab\(^a\)-EGFP protein was still detected with similar intensity, whereas the 8b-EGFP protein could hardly be detected (Fig. 2C). It thus appears that the 8b protein is hardly expressed in the normal context in which it is present in the human SARS-CoV isolates and that it is highly unstable when expressed in cells out of this context.

**Localization of the ORF8 products.** The 8ab\(^a\) protein is 122 amino acids long and contains a hydrophobic domain at its N terminus, which likely functions as a signal sequence. This hydrophobic domain is also present in the 8b protein. The 8b protein, however, does not contain any hydrophobic domains and is therefore predicted to be a cytoplasmic protein. We examined whether these predictions could be visualized by a difference in localization between these proteins. The intracellular localization of the different ORF8 proteins was studied by making use of EGFP fusion proteins, which were expressed in OST7-1 cells using the vTF7-3 expression system. The cells were fixed at 6 h p.i. and processed for immunofluorescence microscopy. The 8ab\(^a\)-EGFP and 8a-EGFP fusion proteins showed a quite reticular pattern, reminiscent of the endoplasmic reticulum (ER), whereas for the 8b-EGFP fusion protein the fluorescence was found distributed throughout entire cells, more indicative of a cytoplasmic localization (Fig. 3). To obtain more information on the localization of the proteins, the cells were stained with an antiserum directed to calreticulin, a protein used as a marker for the ER. The calreticulin staining largely colocalized with the fluorescence of 8ab\(^a\)-EGFP and 8a-EGFP but showed no colocalization with the fluorescence of 8b-EGFP (Fig. 3). These results indicate that the 8b protein...
resides in the cytoplasm, whereas the 8ab\(^+\) and 8a proteins localize to the ER and are probably membrane associated.

**Processing of the ORF8 products.** The 8b protein appears to localize cytoplasmically without membrane association, whereas the 8a and 8ab\(^+\) proteins seem to be membrane associated. Therefore, the proteins were next studied for their co- and post-translational processing. The 8ab\(^+\) protein contains an N-terminal hydrophobic domain, apparently functioning as a signal sequence, and one predicted N-glycosylation site at position 81 in the amino acid sequence (http://www.cbs.dtu.dk/services/NetNGlyc/). The 8ab\(^+\)-EGFP fusion protein was expressed by in vitro translation and by using the vTF7-3 expression system. To investigate the N-linked glycosylation, the fusion protein was expressed in the presence and absence of tunicamycin, which is an inhibitor of N-linked glycosylation. OST7-1 cells were infected with vTF7-3, transfected with the 8ab\(^+\)-EGFP-containing plasmid, and labeled with \(^{35}\)S]methionine for 1 h starting at 5 h p.i. Cells were lysed and processed for immunoprecipitation with a rabbit polyclonal antiserum directed to the EGFP tag. In parallel, in vitro translation was performed on the same construct using the TNT coupled reticulocyte lysate system of Promega in the absence of membranes to analyze the electrophoretic mobility of the full-length nonprocessed protein.

vTF7-3-mediated expression of the 8ab\(^+\) fusion protein in the presence of tunicamycin resulted in a product with an electrophoretic mobility that corresponds to the predicted molecular mass of this protein, which is 41 kDa (Fig. 4A). In the absence of tunicamycin the protein migrated at a slower rate, indicating that the 8ab\(^+\) protein was indeed being glycosylated. This was confirmed by treating the protein expressed in the absence of tunicamycin with either PNGase F or endo H, both of which remove N-linked sugars. Treatment with the glycosidases increased the electrophoretic mobility of the protein; the slight mobility difference between the different treatments is explained by PNGase F cleaving off all sugar residues whereas endo H leaves one GlcNAc residue attached to the polypeptide (Fig. 4B). These results together lead to the conclusion that the 8ab\(^+\) protein is N-glycosylated, most likely at Asn81, the only predicted N-glycosylation site in the protein sequence.

The 8ab\(^+\)-EGFP fusion protein expressed using the vTF7-3 system in the presence of tunicamycin migrated at a slightly slower rate in the gel than in the in vitro translated fusion protein (Fig. 4A). This was suspected to be caused by a decrease in the electrophoretic mobility of the protein, but the release of a hydrophobic domain can also result in a decreased binding of SDS and hence a lower electrophoretic mobility. To further study its processing, the 8ab\(^+\) protein was expressed without the EGFP tag using both the vTF7-3 system and in vitro translation. In both cases the proteins were labeled with \(^{35}\)S]cysteine instead of \(^{35}\)S]methionine because the untagged protein contains only two methionine residues. The radiolabeled proteins were immunoprecipitated with a rabbit antiserum directed to the 8ab\(^+\) protein, which is not very sensitive, but the more sensitive antiserum raised against the 8b protein appeared not to recognize the glycosylated form of the 8ab\(^+\) protein. The results of the vTF7-3-mediated protein expression of the 8ab\(^+\) protein in the presence and absence of tunicamycin were similar to what was seen for the tagged protein (Fig. 4C). The electrophoretic mobility of the protein expressed in the presence of tunicamycin was higher than that of the protein expressed in the absence of tunicamycin, indicating the N-glycosylation of the 8ab\(^+\) protein. However, the result of a comparison of the protein expressed using the vTF7-3 system in the presence of tunicamycin to the in vitro-translated protein was different from that for the tagged protein. In this case the vTF7-3-expressed protein migrated faster in the gel than the in vitro translated protein (Fig. 4C). The increase in electrophoretic mobility is consistent with cleavage of the signal sequence. Apparently, the effect of this cleavage is different for the untagged and tagged proteins.

To clarify this discrepancy, the sequence predicted to be cleaved (the first 15 amino acids) was deleted from the nucleotide sequence in both the 8ab\(^+\) and 8ab\(^-\)-EGFP construct. These constructs were expressed using the vTF7-3 system in the presence and absence of tunicamycin, and the products were compared to the full-length proteins expressed either in vitro or using the vTF7-3 system. In the presence of tunicamycin there was no difference in electrophoretic mobility between the protein obtained from the full-length construct and that from the construct in which the signal sequence had been deleted (Fig. 4D). For the 8ab\(^+\) proteins the migration was faster than migration of the in vitro translated product of the full-length construct whereas for 8ab\(^-\)-EGFP it was somewhat slower. This, indeed, confirmed our interpretation that the signal sequence of the 8ab\(^+\) protein is being cleaved and that this influences the electrophoretic mobility of the 8ab\(^-\) and 8ab\(^-\)-EGFP proteins differently. The electrophoretic mobility of the proteins synthesized without a signal sequence was the same in the presence and absence of tunicamycin. No N-glycosylation occurred since these proteins could not be translocated to the lumen of the ER due to the absence of the signal sequence.

![Diagram](https://via.placeholder.com/150)
The 8a protein contains a N-terminal hydrophobic domain that functions as a signal sequence in the context of the 8ab protein. Yet in the 8a protein this signal will not be able to function due to the small (39 amino acids) size of the 8a polypeptide. This is too short to span the large ribosomal subunit and expose the signal peptide for efficient binding by the signal recognition particle (SRP) since this requires a minimal protein length of 50 amino acids (28). In the EGFP-tagged 8a protein, however, the hydrophobic sequence should be recognized by the SRP and the polypeptide should be translocated across the ER membrane. To verify this supposition, the 8a-EGFP protein was expressed by in vitro translation and by using the vTF7-3 expression system. vTF7-3-infected OST7-1 cells were transfected with the 8a-EGFP plasmid construct and labeled with [35S]methionine from 5 to 6 h p.i. Cells were lysed and processed for immunoprecipitation with specific antibodies followed by SDS–15% PAGE. As shown in Fig. 4E the in vitro translation resulted in two protein species, of which the slower-migrating one represents the 8a-EGFP fusion protein while the faster-migrating species represents the EGFP protein, which is efficiently translated from the internal start codon in this system (but not in the vTF7-3 expression system).
The vTF7-3-expressed 8a-EGFP protein migrated slightly faster in the gel than the in vitro translated product, indicating that the signal sequence is indeed cleaved and that the protein is translocated to the lumen of the ER.

To confirm the translocation we introduced an N-glycosylation site within the EGFP tag and expressed this protein using the vTF7-3 expression system in the presence or absence of tunicamycin and by in vitro translation. The expression of the protein in the absence of tunicamycin resulted in a product with a lower electrophoretic mobility than the product expressed in the presence of tunicamycin, which again migrated faster in the gel than the in vitro translated product (Fig. 4E). This result clearly shows that the protein is translocated to the lumen where the tag is being glycosylated. As expected, the expression of the 8a protein carrying the unmodified EGFP tag was not affected by the presence of tunicamycin. These data demonstrate that also for the 8a protein the N-terminal hydrophobic domain can function as a signal sequence to translocate the protein to the lumen of the ER. We similarly tested the 8b protein and observed that, as predicted, this protein was not co- or posttranslationally processed (data not shown).

Membrane association of the 8ab⁺ and 8b proteins. The signal sequence is the only hydrophobic domain in the 8ab⁺ protein, and since it is cleaved off, the protein is probably not an integral membrane protein. The 8b protein does not contain any predicted hydrophobic domains and is thus unlikely to be membrane associated, consistent with its cytoplasmic localization in the immunofluorescence assay. The membrane association of both the 8ab⁺ and 8b proteins was investigated by performing a sodium carbonate extraction. The proteins were expressed using the TNT coupled transcription/translation system (Promega) in the presence of canine microsomal membranes. The in vitro reaction mixtures were treated with either a sodium carbonate buffer of pH 11 or a Tris-buffered saline buffer of pH 7.5 and were separated by centrifugation into soluble and pellet fractions. In the pH 7.5 buffer proteins present in the lumen of the microsomes or attached peripherally to the membranes will be released, since the sodium carbonate treatment opens the membrane sheets while leaving integral membrane proteins anchored in the lipid bilayer. As a control the SARS-CoV M protein, which is an integral membrane protein, was expressed and treated similarly.

As can be seen in Fig. 5, the 8b protein was found in the...
soluble fraction both after treatment with the pH 7.5 buffer and after treatment with the sodium carbonate buffer of pH 11. This confirmed that it is indeed a soluble, cytoplasmic protein. When treated with the pH 7.5 buffer, the glycosylated forms of the M and 8ab" proteins were found solely in the pellet fractions. This indicates that these proteins are indeed membrane associated. Neither of the proteins was fully glycosylated, which is probably caused at least to some extent by an incomplete incorporation of the proteins into the microsomal membranes. Hence, these unglycosylated forms were found in the soluble as well as in the pellet fractions. After treatment with the pH 11 buffer, the glycosylated M protein was still largely found in the pellet fraction, confirming its identity as an integral membrane protein. In contrast, after this same carbonate treatment the 8ab" protein was no longer found in the pellet fraction but had become fully solubilized. Thus, the 8ab" protein is not an integral membrane protein but exists either as a soluble protein in the lumen of the microsomes or is peripherally associated with the inside of the membranes.

These results together with the immunofluorescence data lead to the conclusion that the 8ab" protein localizes to the lumen of the ER, most likely as a soluble protein.

**Secretion of the 8ab" protein.** Having established that the 8ab" protein is translocated into the ER, we wanted to determine its subsequent fate. Since the only hydrophobic domain on the 8ab" protein appears to be cleaved by signal peptidases and since the protein locates in the lumen of microsomes without an integral membrane association, it is possible that the protein is secreted from the cells as a soluble protein. Therefore, its secretion was examined. To this end vTF7-3-infected OST7-1 cells were transfected with constructs expressing either the 8ab"-EGFP fusion protein or the untagged 8ab" protein. The fusion protein is more efficiently labeled and precipitated, due to a higher abundance of methionine and cysteine residues and a better antiserum, but to exclude the possibility that secretion might somehow be hindered by the EGFP tag, the untagged protein was also expressed. The proteins were pulse-labeled from 5 to 6 h p.i. and chased for 2 h. Immunoprecipitations were performed both on the cell lysate and on the culture medium. Both proteins were clearly detected in the cell lysate and on the medium. To support these multimerization data, a coimmunoprecipitation experiment was performed using the EGFP antiserum. The EGFP antiserum did not precipitate the 8ab" protein when 8ab"-EGFP was expressed alone (data not shown). These results confirm that the 8ab" protein does not travel along the secretory pathway through the Golgi compartment for secretion out of the cells but appears to remain in the ER.

**Multimerization of the 8ab" protein.** Having established that the 8ab" protein resides in the lumen of the ER as a soluble protein, we wanted to further investigate the protein’s fate. As the 8ab" protein contains as many as 10 cysteine residues, we examined whether it engages in homologous protein-protein interactions by comparing its electrophoretic mobility under reducing and nonreducing conditions. The 8ab"-EGFP protein and the 8b-EGFP proteins were expressed in parallel using the vTF7-3 expression system and labeled from 5 to 6 h p.i. with 35S-labeled methionine. Cell lysates were processed for immunoprecipitation using the EGFP antiserum. The immunoprecipitated material was suspended in sample buffer either with or without β-mercaptoethanol and after being heated for 1 min was analyzed by SDS-PAGE. In the presence of β-mercaptoethanol both proteins appeared mainly as single bands (Fig. 7A). These same bands were still observed in its absence but some slower-migrating forms were additionally detected in the case of the 8ab"-EGFP protein. Apparently, this protein has a tendency to associate into covalently linked multimeric complexes, which seem to occur as dimers and higher-order assemblies.

To support these multimerization data, a coimmunoprecipitation experiment was performed using the EGFP-tagged and untagged 8ab" proteins. OST7-1 cells were infected with vTF7-3 and cotransfected with constructs encoding the 8ab" or the 8ab"-EGFP proteins. The cells were labeled with 35S-labeled methionine from 5 to 6 h p.i., after which they were lysed and processed for immunoprecipitation using the EGFP or the 8ab" antiserum. The EGFP antiserum did not precipitate the 8ab" protein when it was expressed alone (Fig. 7B). However, when the tagged and untagged 8ab" proteins were expressed together, the 8ab"-EGFP protein was precipitated but also a protein with the same molecular mass as the 8ab" protein that was not detected when 8ab"-EGFP was expressed alone. This indicates that there is an interaction between the 8ab"-EGFP protein and the untagged 8ab" protein and confirms the earlier observations on the multimerization of the 8ab" protein.
DISCUSSION

SARS-CoV is the genetically most complex CoV presently known. In this respect its position among the Coronaviridae is somewhat comparable with that of the lentiviruses within the retrovirus family. SARS-CoVs encode an extended range of accessory proteins shown to be dispensable for replication in cell culture (47) but for which the significance in the animal or human host has not yet been established. However, as has already become clear from work with other CoVs (6, 12, 30), these proteins contribute critically to the clinical outcome of host infection. Intriguingly, when a virus changes its host species, as was the case for the zoonotic transmission of the SARS-CoV, the role of accessory proteins under the new conditions may change as well. This notion arises in particular in relation to ORF8. SARS-CoVs isolated from masked palm civets, which are considered to have been the immediate animal source for transmission to humans, and from humans during the earliest phases of the SARS outbreak contained an intact ORF8. As we showed here, a glycoprotein is expressed from this ORF that is delivered to the lumen of the ER with the aid of an N-terminal signal sequence, where it remains and supposedly has its function. However, the 29-nt deletion in ORF8 acquired very soon after the zoonotic transmission and observed in all subsequent human isolates entirely disrupts the expression of a functional protein. Whether and how this remarkable evolutionary event in the adaptation of SARS-CoV from masked palm civets to humans has contributed to the course and severity of the epidemic by affecting viral pathogenicity and spread are key questions yet to be addressed.

The 29-nt deletion gives rise to two ORFs, both of which encode nonfunctional protein products. The 5’ terminal ORF 8a specifies a 39-amino-acid-long polypeptide of which the first 35 residues are identical to the N-terminal part of the ORF8 primary product; the remaining 4 residues are acquired by translation from another reading frame that is engaged due to the deletion. We have been unable to study the fate of this
polypeptide due to the lack of a proper antiserum. It is quite unlikely, however, that it undergoes the same targeting and processing as the full-length 8ab^+ protein. Given its small size the polypeptide is probably already released from the ribosome before the N-terminal signal sequence has been recognized by the SRP (28). Due to the strict cotranslational nature of the eukaryotic SRP action, the polypeptide will thus not be delivered to the ER, and the protein will remain in its precursor form in the cytosol. Alternatively, if the precursor would somehow succeed in becoming inserted posttranslationally into the ER membrane, cleavage of its signal sequence would release a probably nonfunctional 24-residue fragment of the 107-residue mature 8ab^+ protein. It has been demonstrated that antibodies to the 8a protein can be detected in serum from a small fraction of SARS patients (3), indicating that ORF8a is indeed expressed in the human host. Using expression of an N-terminally hemagglutinin-tagged form of the 8a protein, these authors additionally reported that the polypeptide localizes in mitochondria, induces apoptosis, and enhances SARS-CoV replication. It is, however, unclear to what extent these observations were influenced by the presence of the hydrophilic extension preceding the signal sequence.

In contrast to a previous report (15), we were unable to detect an ORF8b protein in SARS-CoV-infected cells. This was also the case when we used the vaccinia virus T7 expression system to express a construct containing the ORF8b in its viral context, i.e., behind ORF8a. It was even the case when we mimicked the ORF8 mRNA, known to be transcribed in SARS-CoV-infected cells using the TRS occurring directly upstream of ORF8 (41), by providing the construct with the 5’ viral leader sequence. The lack of ORF8b expression did not come as a surprise. Its expression would require translation initiation at an internal AUG codon on the mRNA by one of two possible mechanisms. The first, translation via leaky ribosomal scanning, is mainly seen when the upstream AUG is in a very poor translational context (18). Though the sequence around the start codon of ORF8a may not be in the most optimal Kozak context, it does fulfill the most important requirement of a purine at the −3 position (17). Moreover, the context of the ORF8b start codon is considerably less optimal, and this is the fourth AUG codon that would be encountered. Leaky ribosomal scanning is, however, the manner by which the SARS-CoV ORF7b (34) and the infectious bronchitis virus ORF3b (24) are translated. The second mechanism involves direct ribosomal initiation at the internal AUG, a process also not uncommon in CoV replication. Proteins like the mouse hepatitis coronavirus and infectious bronchitis virus E protein (24, 42) and the transmissible gastroenteritis virus accessory protein 3b (27) are expressed this way. The process is critically dependent on the mRNA secondary structure preceding the AUG. It is extremely unlikely that the 29-nt deletion in ORF8 would have accidentally created such an IRES for which there was no need in the parental virus. Consistent with all these considerations, the occurrence of antibodies against the 8b protein in SARS patients has not been reported. Though this might be due to the protein’s poor immunogenicity, our ability to induce 8b antibodies in rabbits argues against this explanation.

The 8b protein tagged at its C terminus with EGFP appeared to be highly unstable. Expressed using the vaccinia virus system, it was degraded almost completely within 2 h. Due to the lack of a signal sequence, it is mislocalized to the cytoplasm, where it occurs as a soluble protein that does not properly oligomerize and mature and where it is apparently targeted to the proteasome. These characteristics are very different from those of the 8ab^+ protein as it is encoded by the ORF8 of animal and early epidemic human SARS-CoV isolates. For this protein no obvious degradation was observed after 2 h. We demonstrated that the 8ab^+ protein is translocated via a cleavable signal sequence to the lumen of the ER, where it becomes N-glycosylated, and forms homomultimeric complexes. The most likely fate of such a soluble, lumenal ER protein is secretion, as was found, for instance, for the FIPV 7b protein (previously called 6b), another soluble glycoprotein with a cleavable signal sequence (44). This protein appeared to be secreted significantly faster from FIPV-infected cells than when expressed individually (44), indicating that the infection conditions somehow affected the protein’s transport. We could not, however, find any evidence for secretion of the expressed SARS-CoV 8ab^+ protein. Based also on the maturation state of its N-glycans, we conclude that the protein is retained and accumulates in the ER. It is unclear what causes this retention; the protein appears not to exhibit a known ER retention signal in its sequence (32). It remains to be established whether the protein is also retained in the ER in virus-infected cells or whether under these circumstances the protein is secreted. The 8ab^+ protein has been shown to interact with several other viral proteins, such as the S, M, 3a, and 7a proteins (15), and these interactions are likely to affect the fate of the protein in virus-infected cells.

We can only speculate about the function of the 8ab^+ protein. While the interactions with the structural proteins just mentioned are probably advantageous in the animal hosts, as judged from the conservation of ORF8, the acquired deletion indicates that these interactions are not essential in the human host. But whether this disruption of ORF8 was beneficial for the virus and contributed to the selection and predominance of mutants carrying this deletion will probably be hard to establish. In civet cats the clinical signs of SARS-CoV infection did not appear to be affected by the absence or presence of the ORF8 deletion (46). What might be still learned from sequence analyses of human cases is whether the 29-nt deletion really occurred just once and gave rise to the viruses that subsequently spread across the globe or whether there were multiple, independent identical events. It is of note that, later during the SARS epidemic, some human virus isolates have been described carrying additional deletions, i.e., a cluster of viruses with an 82-nt deletion in the same region of ORF8 (4) and at least two viruses with a 415-nt deletion entirely removing ORF8 (4, 5). The observations are consistent with the conclusion that, if there is any effect of the 29-nt deletion, it is more likely caused by the loss of the 8ab^+ protein than by the gain of a new ORF.

ADDITIONAL INFORMATION

After the completion of this work another study appeared about the properties of the proteins expressed from SARS-CoV ORF8 (23). The results of this study are largely consistent with the data presented here.
ACKNOWLEDGMENTS

We thank Matthias Raaben, Mijke Vogels, and Marne Hagemeijer for stimulating discussions; Martijn van Beijnen and Michiel van Gent for help with some of the cloning work; and Gert-Jan Godeke for help with the SARS-CoV infection experiment.

This work was supported by a grant from the European Community (Frame VI, DISsect Project, SP22-CT-2004-511060).

REFERENCES


