
Cytoplasmic tail of coronavirus spike protein has intracellular targeting signals

JIBIN SADASIVAN, MANMEET SINGH and JAYASRI DAS SARMA*

*Department of Biological Sciences,
Indian Institute of Science Education and Research Kolkata,
Mohanpur 741 246, India*

**Corresponding author (Email, dassarmaj@iiserkol.ac.in)*

Intracellular trafficking and localization studies of spike protein from SARS and OC43 showed that SARS spike protein is localized in the ER or ERGIC compartment and OC43 spike protein is predominantly localized in the lysosome. Differential localization can be explained by signal sequence. The sequence alignment using Clustal W shows that the signal sequence present at the cytoplasmic tail plays an important role in spike protein localization. A unique GYQEL motif is identified at the cytoplasmic terminal of OC43 spike protein which helps in localization in the lysosome, and a novel KLHYT motif is identified in the cytoplasmic tail of SARS spike protein which helps in ER or ERGIC localization. This study sheds some light on the role of cytoplasmic tail of spike protein in cell-to-cell fusion, coronavirus host cell fusion and subsequent pathogenicity.

[Sadasivan J, Singh M and Sarma JD 2017 Cytoplasmic tail of coronavirus spike protein has intracellular targeting signals. *J. Biosci.* **42** 231–244]

1. Introduction

Coronaviruses are enveloped positive strand RNA viruses which acquire their membrane envelope by budding into the lumen of the intermediate compartment between the endoplasmic reticulum and Golgi complex (ERGIC) (Krijnse-Locker *et al.* 1994). Matured virus are thought to move through the vesicles via the secretory pathways and exit the cell when the vesicles fuse with the plasma membrane (Holmes *et al.* 1984; Toozé *et al.* 1987). Mature coronavirus virion normally consists of structural proteins, S (spike), M (matrix), and E (envelope) that surrounds a helical nucleocapsid. An additional protein, HE (hemagglutinin esterase) is expressed in some beta-coronaviruses. M is a triple spanning membrane protein believed to play a key role in maintaining viral core structure; whereas E, a small enveloped glycoprotein, is critical for virion assembly. The S protein is a type I

membrane glycoprotein which (Vennema *et al.* 1990) constitutes the characteristic coronavirus spike. HE is also a type I glycoprotein with a single membrane spanning anchor and a short cytoplasmic tail. Detailed studies on virus-like particle formation (VLPs) revealed that only M and E proteins are required for VLP formation.

Most glycoproteins of the enveloped viruses sort to the cellular membrane, where their budding occurs (Garoff *et al.* 1998). Some viral glycoproteins sort to the specialized apical or basolateral plasma membrane of the polarized epithelial cells. For example, glycoprotein gp160 from type 1 human immunodeficiency virus (HIV-1) sorts to the basolateral domain (Owens and Compans 1989), but a prototype glycoprotein, hemagglutinin from the influenza virus, sorts to the apical domain (Mora *et al.* 2002). To mediate budding of infectious viral particles through the intracellular membranes, viral glycoproteins must embed a signal for sorting

Keywords. Coronavirus spike protein trafficking; cytoplasmic tail signal; endoplasmic reticulum–Golgi intermediate complex; lysosome

Abbreviations used: ER, endoplasmic reticulum; ERGIC, endoplasmic reticulum–Golgi intermediate compartment; YFP, yellow fluorescence protein

to an intracellular compartment. For example, Golgi targeting signal (Teasdale and Jackson 1996) in the membrane spanning domains of the M protein of coronaviruses (Machamer and Rose 1987; Swift and Machamer 1991) and the G1 protein of bunya viruses (Matsuoka *et al.* 1991) specify the glycoprotein accumulation at Golgi complex (Andersson and Pettersson 1998). The dilysine signal (DEKKMP) in adenovirus glycoprotein E19 carboxy terminal localizes it to the ER (Paabo *et al.* 1987; Nilsson *et al.* 1989). A similar dilysine signal partitions human foamy virus maturation to intracytoplasmic compartments (Goepfert *et al.* 1999). Proteins containing the dilysine motif have been observed to bind directly to well-characterized coat proteins (COPI) (Cosson and Letourneur 1994; Gaynor *et al.* 1998; Corse and Machamer 2002). The COPI coated vesicles mediate retrograded transport of the dilysine signal-containing proteins from the proximal Golgi compartments to the ER.

Recent studies on coronavirus spike protein signals have revealed a dilysine-based ER localization signal in the infectious bronchitis virus (IBV, a gamma-coronavirus) (figure 1). The IBV S protein also carries a tyrosine-based (YYTF) endocytotic signal in its cytoplasmic tail (Corse and Machamer 2002; Bonifacino and Traub 2003) which is more crucial for the intracellular retention of the S protein than its canonical dilysine-based signal (Winter *et al.* 2008). A similar dibasic motif (KXHXX) in the cytoplasmic tail of S protein from alpha-coronavirus TGEV and from the newly emerging pathogen SARS, an outlier of beta-coronavirus, localizes the S protein in the ERGIC (Teasdale and Jackson 1996; Lontok *et al.* 2004). A tyrosine-based sorting motif (YEPI) has also been shown to be an intracellular localization signal in the cytoplasmic tail of TGEV (Schwegmann-Wessels *et al.* 2004). Very few studies have

convincingly demonstrated a localization signal in the cytoplasmic tail of beta-coronaviruses that would lead to their sequestration within the intracellular compartment. The intracellular transport and targeting of proteins from their site of synthesis to their correct destination is a process instrumental to maintenance of virion integrity. A major effort is underway to understand the interplay of factors that guide the envelope proteins to the intracellular budding compartments avoiding the bulk flow through the secretory pathways (Bonifacino and Traub 2003). There is ample evidence that the localization signalling motifs in the virus mimic those used by the endogenous cellular proteins of the host cell subverting their cellular machinery (Bonifacino and Traub 2003).

We studied the intracellular trafficking and localization of the full-length S protein from two human coronaviruses (HCoV): SARS and OC43 (a beta-coronavirus). Full-length SARS S protein when expressed exogenously is observed to accumulate predominantly in intracellular compartments resembling ER and ERGIC with occasional surface staining. HCoV-OC43 S protein is localized in distinct puncta that could represent the endocytic structure. Exogenously expressed HCoV-SARS S protein occasionally induces cell-to-cell fusion when transported to surface in transfected cells, but HCoV-OC43 S protein fail to induce the same. The intracellular distribution of the S protein differs significantly in the two viruses. We identified a unique tyrosine-based motif (GYQEL, an YXXØ motif where Ø = bulky hydrophobic side chains) at the cytoplasmic tail of HCoV-OC43 (figure 1). YXXØ is known to function as a signal for rapid internalization (Mellman 1996; Kirchhausen *et al.* 1997; Marks *et al.* 1997), lysosomal targeting and localization to basolateral surface in polarized epithelial cells (Mellman 1996). The position of this motif in the protein sequence

- CCSTGCCGCI GCLGSCCHS I CSRRQFENYEP - - - IE - - KVHVH -	TGEV - group 1
- CFSTGCCGCI GCLGSCCHS I CSRRQFENYEP - - - IE - - KVHVH -	FIPV - group 1
- CCSTGCCGFFSCFASS I RGCESTKLPYYD - - - VE - - KIH I Q -	HCoV-229E - group
- CCS - - CLK - GAC - - SCG - SCCK - - - FDEDDSEPV LKGVKLHYT -	HCoV-SARS - group 2
- CCCTGCG - - TSCFKKCG - GCCD - - DYTGHQE - LV I K - - TSHED -	group 4 outlier
- CCCTGCG - - TSCFKKCG - GCCD - - DYTGHQE - LV I K - - TSHED -	BCV - group 2
- CCCTGCG - - SCCFKKCG - NCCD - - EYGGHQDS I V I HN I S SHED -	MHV-A59 - group 2
- CCCTGCG - - TSCFKKCG - GCCD - - DYTGYQE - I V I K - - TSHDD -	HCoV-OC43 - group:
- CCC - GCFG I MPLMSKCG - KKSS - - YTTT I DNDVVTEQYR P K K S V -	IBV - group 3

Figure 1. Alignment of the amino acid sequence in the cytoplasmic tail downstream of the membrane-spanning domain in the coronavirus S proteins. The YXXØ internalization signals are coloured red, and the di-lysine-based motifs, grey. The di-basic motifs from HCoV-229E and FIPV are putative intracellular signals not experimentally verified (Lontok *et al.* 2004). Localization signals experimentally identified from this study are marked in bold. The sequence alignment was done using the program Clustal omega (McWilliam *et al.* 2013). Specific virus, localization signal, cellular localization and the authors describing them are elaborated as follows: (SARS; KLHYT; ER/ERGIC; this study, also Lontok *et al.* 2004 on truncated protein); (OC43; GYQEL; lysosome/internalization; this study); (TGEV; KVHVH, YEPI; ER; (Schwegmann-Wessels *et al.* 2004); (IBV; KKS SV, YYTF; ER/Internalization; (Lontok *et al.* 2004).

relative to the transmembrane domain determines its cellular location. This signal is not present in the cytoplasmic tail of other two beta-coronavirus, BCV (bovine coronavirus) and MHV-A59 (mouse hepatitis virus). Our findings allow rationale for the observed differential localization of the S proteins in different coronaviruses from previous published studies.

2. Materials and methods

2.1 Antisera and reagents

The rabbit polyclonal anti-SARS spike antisera was a kind gift of Jim Wilson (University of Pennsylvania, PA). The rabbit polyclonal anti-BCV S antibody that recognizes HCoV-OC43-S protein was a kind gift from Brenda Hogue (Arizona state university). Rabbit anticalnexin was purchased from Stress gene (San Diego, CA), rabbit anti- β -COP from Affinity Bioreagents (Golden, CO), rabbit anti-MannII (Chemicon), rabbit anti-EEA1 and rabbit anti-Lamp1 from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA). Fluorescent-tagged secondary antibodies were obtained from Jackson ImmunoResearch (West Grove, PA), Triton X-100 from Roche Diagnostics and tissue culture reagents from Invitrogen. Unless otherwise specified, all other reagents were from Sigma.

2.2 Construction of SARS-CoV genes expression plasmids

The S gene of SARS-CoV strain Urbani was amplified by reverse transcription-PCR from viral RNA isolated from cell lysate of infected Vero cells (provided by Dr W Bellini, CDC). Reverse transcription was performed with Superscript II (Invitrogen, Carlsbad, CA) and random primers using 1 μ g of total cellular RNA, as described by the manufacturer. The cDNA was amplified with a mix of Tth DNA polymerase (Roche, Indianapolis, IN) and Vent DNA polymerase (New England Biolabs, Beverly, MA) with primers 5'-GTTAACAACCTAAGAATTCATGTTATTTTC-3' and 5'-AATCTCATAAACCTCGAGTAAAGTTCGTTTATGTG-3' using a hot-start long-PCR consisting of one cycle of 94°C, 2 min and 80°C, 3 min, followed by 30 cycles of 94°C, 30 s; 55°C, 20 s; 72°C, 3 min, and 72°C, 7 min extension. The resulting PCR product was cloned into TOPO-II TA vector (Invitrogen) and its sequence was verified by automated sequencing using BigDye Terminator v3.1 Cycle sequencing kit (Applied Biosystems, Foster City, CA). A wild-type Urbani spike sequence was confirmed by sequence analysis using Macvector (Accelrys, San Diego, CA). The S gene was then inserted between the EcoRI and XmaI sites downstream of the CMV IE enhancer and the chicken β -actin promoter into

the mammalian expression vector pCAGGS-MCS (Niwa *et al.* 1991). OC43 spike in pCDNA3 was used as template.

2.3 Construction of yellow fluorescence-tagged expression vectors encoding either HCoV-SARS or HCoV-OC43 S protein

We engineered the yellow fluorescent protein (YFP)-tagged HCoV-SARS S protein (SARS-S) and HCoV-OC43 S protein (OC43-S) by constructing the full-length cDNA of HCoV-SARS-S and HCoV-OC43-S in commercially available pYFP-N1 vector (Promega). Full-length cDNA of HCoV-SARS-S and HCoV-OC43-S was amplified by PCR amplification in a Robocycler (Stratagene) using Expand Long Template PCR System (Roche Diagnostics, Indianapolis, IN), starting with either HCoV-SARS cDNA or HCoV-OC43 cDNA as a template. The resulting PCR products were restricted with XhoI and BamHI and ligated in pYFP-N1 vector to make in frame fusion protein. Ligated product was transformed into bacterial stocks and DNA was isolated and purified by using Qiagen Midiprep kit according to the manufacturer's instructions. Positive clones of HCoV-SARS-S/pYFP-N1 and HCoV-OC43-S/pYFP-N1 were screened by restriction digestion and sequenced to confirm the in frame fusion of S protein-YFP (S-YFP).

2.4 Construction of pCAGGS expression vector encoding either SARS-S-YFP or OC43-S-YFP fusion protein

For transient transfection in Hela cells, the full-length cDNA encoding HCoV-SARS-S-YFP (SARS-S-Y) and HCoV-OC43-S-YFP (OC43-S-Y) was subcloned into pCAGGS vector. The full-length cDNA of SARS-S-Y and OC43-S-Y was amplified by PCR by using HCoV-SARS-S/pYFP-N1 and HCoV-OC43-S/pYFP-N1 as a template respectively. PCR products were restricted with EcoRI and XmaI and ligated into pCAGGS mammalian expression vector (Niwa *et al.* 1991). Ligated product was transformed and plasmid DNA was isolated and purified as described previously. Positive clones of SARS-S-Y/pCAGGS and OC43-S-Y/pCAGGS constructs were sequenced.

2.5 Site-directed mutation

The dibasic residues (K and H) at the C terminus (KLHYT) of SARS-S protein were mutated using Quick change site directed mutagenesis kit (Stratagene, La Jolla, CA), generating the SARS-S_{2A}-Y/pCAGGS mutant signal (Lys→Ala and His→Ala). Same kit was also used to mutate the critical tyrosine after the glycine residues, a lysosomal targeting signal of HCoV-OC43-S (GYQTI) cytoplasmic tail, to

generate OC43-S_{GtoA}-Y/pCAGGS mutant and HCoV-OC43-S_{YtoA}-Y/pCAGGS mutant.

2.6 Transfection and immunofluorescence

For transient transfection and immunofluorescence, HeLa cells were plated on 25 mm circular coverslips in 35 mm dishes 1 day prior to transfection and transfected with S protein from either wild-type HCoV-SARS, HCoV-OC43, or tagged spike cDNA constructs using Fugene (Roche Diagnostics, Indianapolis, IN) at 1 µg/mL DNA using Fugene/DNA ratios of 6:1 (1 µg). For protein analysis, cells were plated on 100 mm tissue culture dishes and transfected with cDNA constructs as described above.

For immunofluorescence after 48 h of transfection, the cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 10 min at room temperature, then washed 3x with PBS and permeabilized with PBS+0.5% Triton X-100 and then blocked with PBS + 0.5% Triton X-100+ 2% heat-inactivated goat serum (PBS/GS). The cells were incubated with primary antisera diluted into PBS/GS for 1 h, washed, and labelled for 1 h with secondary antisera (Texas Red goat anti-rabbit) diluted into PBS/GS. The cells were then washed with PBS, mounted into Mowiol. Cells were next visualized by fluorescence microscopy using an Olympus IX-81 microscope system with a 60X UPlanApo oil immersion objectives with the iris diaphragm partially closed to the limit the contribution of out of plane fluorescence and filter packs suitable for green (U-MWIBA BP460-490 DM505 BA515-550) and red (U-NMG BP530-550 DM570 BA590-800+) fluorescence. Images were acquired with a Hamamatsu Orca-1 CCD camera and Image Pro images analysis software (Media Cybernetics, Silver Spring, MD).

For staining with anti-calnexin, cells were fixed with MeOH/acetone (1:1) (v/v) instead of paraformaldehyde. To label surface SARS-S-Y and OC43-S-Y fusion proteins, transfected cells were washed with ice-cold PBS and incubated with rabbit anti-SARS-S or rabbit anti-BCV-S antisera for 10 min at 4°C. Cells were then washed 2x with ice cold PBS and fixed by 4% paraformaldehyde for 10 min, followed by washing with 2x PBS and mounted into Mowiol and visualized as described above.

Transfection ability of YFP-tagged spike protein constructs are much more efficient compared to wild-type spike construct, the reasons for which are unclear (data not shown). Because of efficient detection of YFP fluorescence and the limitations of spike specific antibodies, we decided to use the YFP-tagged spike constructs instead of wild-type constructs throughout this study.

The expression of the S protein via the CMV promoter was below the limits of detection, we subcloned S-Y fusion protein and the full-length untagged S into pCAGGS, a mammalian expression vector under the control of a chicken

β-actin promoter that, in the past, has proved useful for the efficient expression of RNA virus glycoprotein (Niwa *et al.* 1991). Transient expression in HeLa cells by using pCAGGS/S-Y resulted in detectable expression S-Y under epifluorescence microscopy. It is not clear why pCAGGS was more efficient at driving S expression compared to CMV promoter containing vectors.

3. Results

3.1 Both HCoV-OC43 and HCoV-SARS YFP-tagged S protein localize in the intracellular compartment

YFP-tagged full-length cDNA of HCoV-SARS spike and HCoV-OC43 spike in pCAGGS constructs were generated as described in materials and methods. To detect the intracellular localization of these constructs, HCoV-SARS-S-Y (SARS-S-Y) and HCoV-OC43-S-Y (OC43-S-Y) were transiently transfected into HeLa cells and then examined by indirect immunofluorescence microscopy, utilizing the YFP autofluorescence properties of the constructs. Full-length HCoV-SARS S (SARS-S) protein was predominantly accumulated in intracellular compartment resembling ER and ERGIC (figure 2c). In contrast full-length HCoV-OC43 S (OC43-S) protein was localized in the puncta that could represent the endocytotic structures (figure 2d). We observed similar intracellular distribution of the S protein in both human kidney (293 cell line; figure 3a, b) and human lung epithelial cells (A549; figure 3c, d). Our observation is consistent with the hypothesis by Vennema *et al.* (Vennema *et al.* 1990) that exogenously expressed coronavirus S protein mostly remain intracellular. When YFP alone plasmid was transfected in HeLa cells, the fluorescence was detected all through the cells (figure 2a, b).

To further confirm the intracellular localization of SARS-S and OC43-S protein, and to rule out that the intracellular retention is not due to YFP, we transfected HeLa cells with wild-type full-length cDNA construct of SARS-S protein and OC43-S protein in pCAGGS. Transfected cells were immunolabelled with respective antisera and examined by immunofluorescence microscopy (figure 2e, f). Wild-type S constructs of both HCoV-SARS and HCoV-OC43 were predominantly localized in intracellular compartments as with YFP-tagged fusion S proteins.

3.2 Confirmation of intracellular localization of YFP-tagged HCoV-SARS and HCoV-OC43 S protein using respective anti-spike antisera

To confirm that antisera is also detecting the intracellular distribution of YFP-tagged spike protein, we immunolabelled the YFP-tagged S protein transfected cells

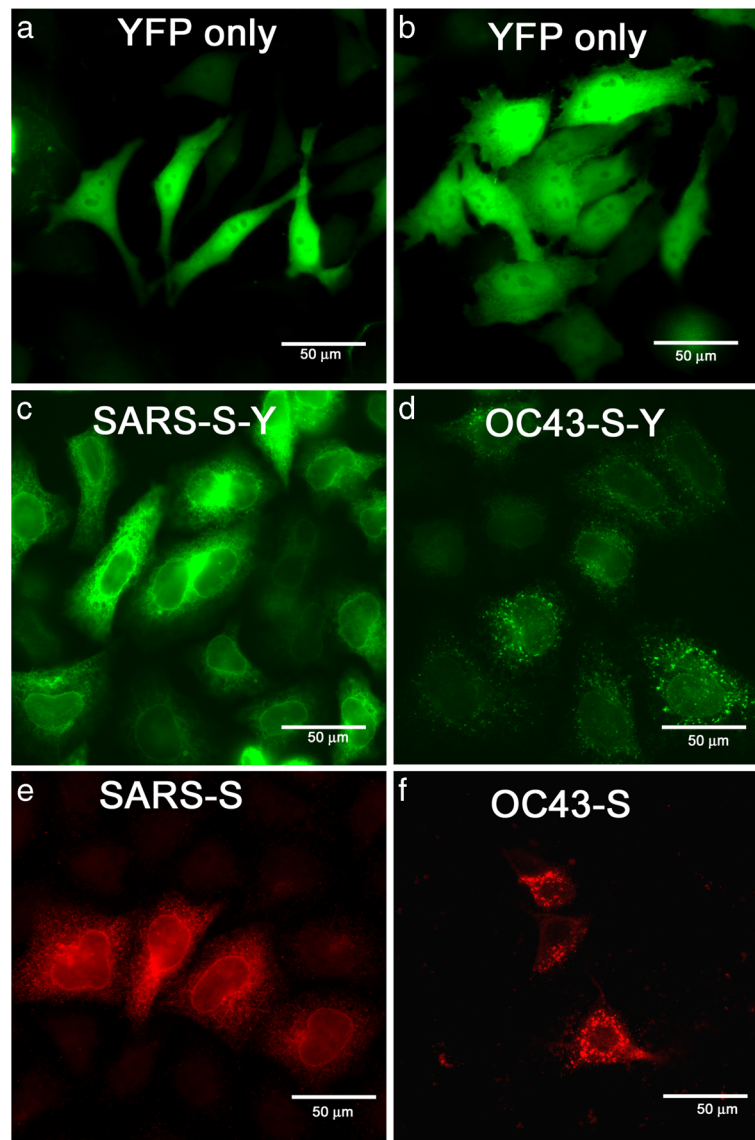


Figure 2. HCoV-OC43 and HCoV-SARS YFP-tagged S protein localizes in the intracellular compartment. HeLa cells were transiently transfected with SARS-S-Y (c) and OC43-S-Y (d). After 48 h of post-transfection, cell were fixed in 4% PFA for 10 min, washed and mounted on glass coverslip and observed under Olympus IX-81 microscope system. YFP fluorescence spread all through the cells when transfected with the YFP alone (a,b). YFP fluorescence in SARS spike can be seen surrounding the nucleus in ER or ERGIC (c), while in OC43, it is seen as a punctate staining in vesicular compartment (d). Wild-type S constructs of both HCoV-SARS and HCoV-OC43 were immunolabelled with respective antisera and were also observed in intracellular compartments (e and f) ($n > 3$).

with respective antisera – rabbit anti-HCoV-SARS S protein antisera for SARS-S-Y construct and rabbit anti-BCV antisera for OC43-S-Y construct. BCV antisera are known to cross-react with HCoV-OC43 (Hogue *et al.* 1984). As shown in figure 4a–f, spike antisera and YFP are significantly co-localizing, demonstrating that both HCoV-SARS (figure 4a–c) and HCoV-OC43 (figure 4d–f) antisera also detects the similar localization of the S fusion protein in the intracellular compartments.

3.3 Reconfirming that SARS-S protein and OC43-S is mainly retained in the intracellular compartment

Localization of the S protein (figure 2) was further confirmed by surface staining and antibody uptake experiments. To distinguish the surface expression from the internally localized YFP auto fluorescence of SARS-S-Y and OC43-S-Y, intact HeLa cells expressing SARS-S-Y and OC43-S-Y were stained with respective rabbit antisera at 4°C for 10 min and then fixed and

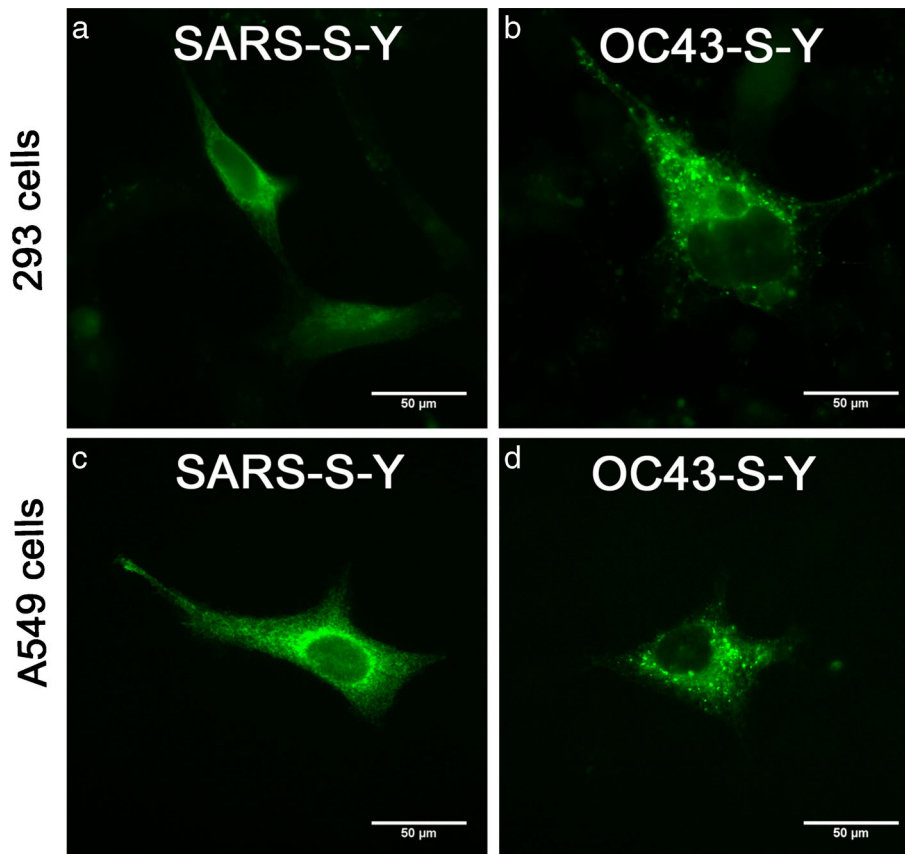


Figure 3. HCoV-OC43 and HCoV-SARS YFP-tagged S protein localizes in the intracellular compartment. Human kidney (293 cell line) (a,b) and human lung epithelial cells (A549) (c,d) were transiently transfected with SARS-S-Y (a,c) and OC43-S-Y (b,d). After 48 h of post-transfection, cells were fixed in 4% PFA for 10 min, washed and mounted on glass coverslip and observed under Olympus IX-81 microscope system. YFP fluorescence in SARS spike can be seen surrounding the nucleus in ER or ERGIC, while in OC43, it is seen as punctate staining in vesicular compartment.

stained with fluorescent-tagged goat anti-rabbit IgG. SARS-S-Y was absent from surface in most of the cells and localized at the intracellular compartments (figure 5a), and OC43-S protein was mainly localized in distinct puncta that could represent endocytic structures following internalization from the plasma membrane (figure 5b).

To test whether SARS-S protein after reaching the surface can be internalized and OC43-S puncta could represent endocytic structures following internalization from the plasma membrane, live cells expressing S protein from either SARS or OC43 were incubated for 15 min at 37°C with either anti-SARS antisera or anti-BCV antisera, and then fixed, permeabilized, and stained with fluorescence conjugated secondary antibodies. In some transfected cells where the SARS-S protein expressed in higher levels, we observed antibody staining mainly restricted to the surface (figure 5c; colour merged). The perinuclear intracellular fluorescent S protein rarely labelled with the exogenously added antibody,

indicating that the portion of the SARS-S protein which reaches the plasma membrane may not efficiently endocytosed after reaching the plasma membrane. Interestingly, OC43-S protein expressing cells did not show any surface antigen staining, and the intracellular puncta containing OC43-S-Y were not labelled by exogenously added antibody (figure 5d; colour merged), indicating that OC43-S-Y could not reach the plasma membrane.

3.4 Steady-state organellar localization of HCoV-SARS and HCoV-OC43 S protein

To determine the specific intracellular localization compartment of SARS and OC43 spike protein, transiently transfected HeLa cells were double labelled with antibodies recognizing various resident proteins localized in different intracellular compartments (calnexin for ER, β cop for ERGIC, MannII for Golgi

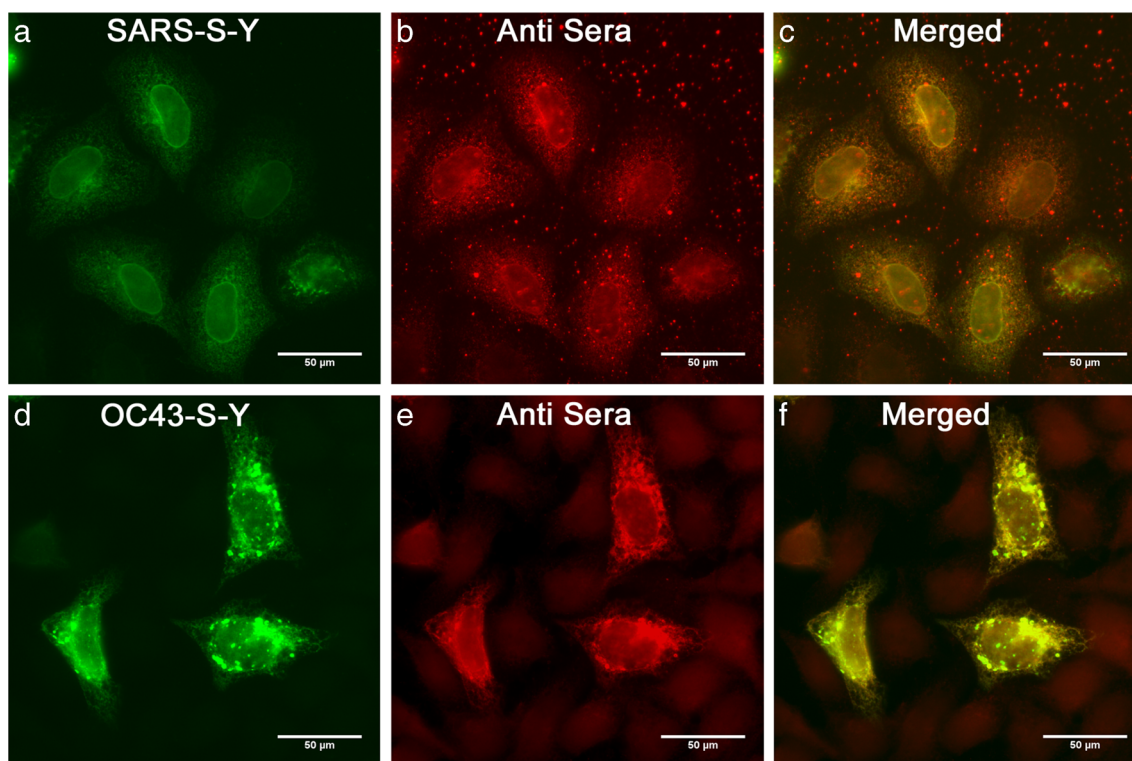


Figure 4. Co-localization of spike antisera with YFP. HeLa cells were transiently transfected with SARS-S-Y (a–c) and OC43-S-Y (d–f). After 48 h of post-transfection, cell was immunolabelled using rabbit anti-HCoV-SARS S protein antisera for SARS-S-Y construct (b) and rabbit anti-BCV antisera for OC43-S-Y construct (e) which were then stained with Texas Red goat anti-rabbit IgG secondary antibodies. YFP is denoting the spike protein (a and c); merged images (c and f) ($n > 3$).

and LAMP1 for lysosome). The localization of SARS-S-Y partially overlapped the ER resident protein calnexin, the ERGIC protein β -COP and *trans*-Golgi marker MannII. The distribution of SARS-S-Y strongly overlapped with the ER (figure 6a–c) indicating its large accumulation in the ER compartment, possibly due to the slow folding of its large luminal domain (Vennema *et al.* 1990). Additionally, the localization of SARS-S-Y in ERGIC compartment (figure 6g–i) is consistent with the presence of the dibasic KXHXX signaling sequence at the extreme C termini of SARS-S as demonstrated by Lontok *et al.* (2004). The distribution of SARS-S-Y also overlapped with the Golgi marker MannII (figure 6m–o). Unlike SARS-S-Y, OC43-S-Y showed partial overlap with both ER (figure 6d–f) and ERGIC marker (figure 6j–l), Golgi marker (figure 6p–r) and mainly localized in the puncta that could represent the endocytotic structures (more description below).

3.5 S protein of HCoV-OC43 possess a lysosomal targeting signal

Examination of the cytoplasmic tail sequence of the S protein from HCoV-OC43 (figure 1) revealed a previously

known motif, GYXX \emptyset , at the 9 residues from the carboxyl terminus. YXX \emptyset signals are known to be much more widely involved in protein sorting, and are required for the rapid internalization from the plasma membrane. These signals also involved in targeting of transmembrane proteins to lysosomes and lysosome-related organelles in addition to endocytosis. The presence of glycine, preceding the critical tyrosine, a characteristic of lysosomal targeting signals may inhibit the flexible or unhindered recognition of the YXX \emptyset residues in close proximity to the membrane.

Full-length OC43 spike protein was unable to demonstrate surface expression either by surface labeling or by antibody uptake experiment. To verify whether the motif signals for OC43-S-Y transport to endocytic organelles, we used immunofluorescence co-localization with the early endosomal marker (EEA1) (figure 7a–c) and late endosomal- and lysosomal- marker (Lamp1) (figure 7d–f). OC43-S-Y intracellular puncta showed significant co-localization with the lysosomal marker Lamp-1 (figure 7d–f) but no significant overlap with early endosomal marker EEA1 (figure 7a–c). This suggests that GYXX \emptyset motif may

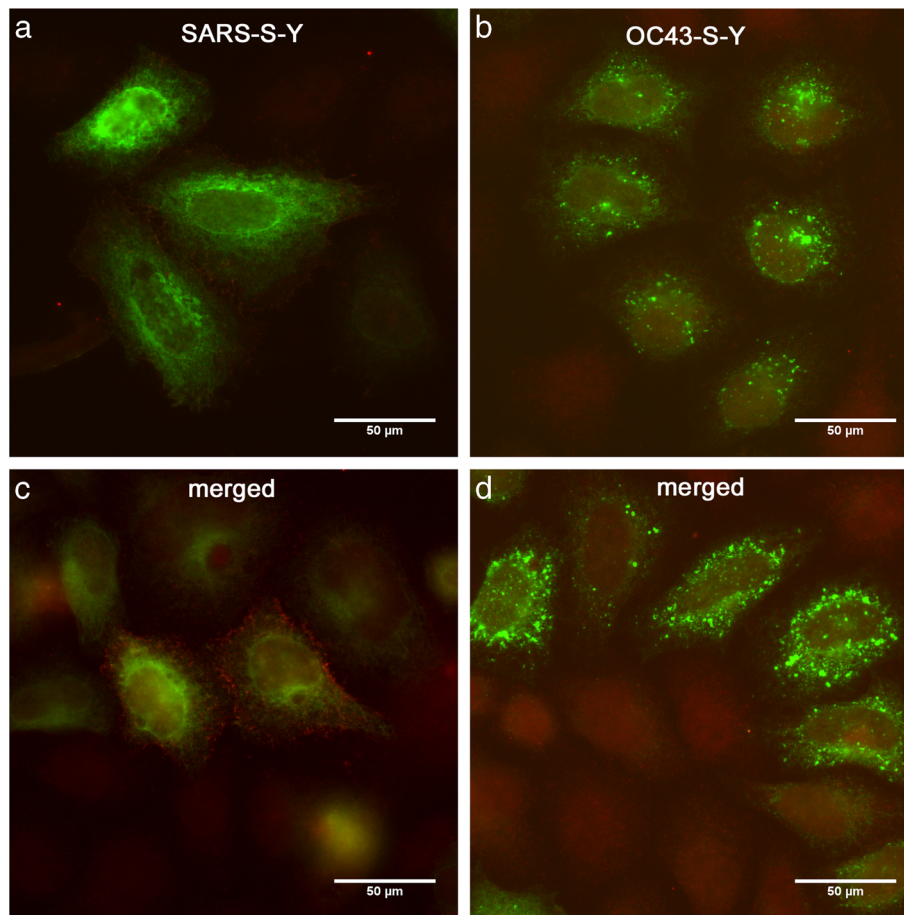


Figure 5. Antibody uptake experiment to demonstrate SARS-S protein occasionally goes to surface due to overexpression (a and c), and OC43-S (b and d) retains mainly in the intracellular compartment. Transfected cells were stained with respective rabbit antisera at 4°C for 10 min and then fixed and stained with fluorescent-tagged goat anti-rabbit IgG ($n > 3$).

be directing the OC43-S-Y in late endosomal or lysosomal compartments with or without recycling from the plasma membrane.

3.6 Mutation of the putative internal dibasic signal of HCoV-SARS-S protein results in surface expression

Examinations of cytoplasmic tail sequence of the S protein from SARS (figure 1) show a unique motif KLHYT, at the C termini. S protein from other alpha-coronavirus (human coronavirus 229E and feline infectious peritonitis virus) also contains the sequence at their C termini (figure 1). The KLHYT fits the criteria for the dibasic KXHXX motif if the histidine residue is protonated. Lontok *et al.*, in their chimeric S protein studies used C terminal 11 amino acids of SARS-S protein attached to the plasma membrane reporter protein VSV-G to show KXHXX motif is an intracellular localization signal for SARS, and the intracellular

distribution closely overlapped with ERGIC. They also demonstrated that mutagenesis of the lysine and histidine residues in the KXHXX signal results in the loss of intracellular localization (Lontok *et al.* 2004). In our full-length SARS spike YFP fusion protein though the KXHXX motif is no more near the C-terminal; it still acts like the KXHXX motif in the C terminal tail of SARS wild-type protein. To test that the internal KXHXX motif could function as a genuine intracellular localization signal, the lysine and histidine residues in SARS-S-Y_{PCAGG} was mutagenized to alanines to construct the SARS-S_{2A}-Y_{PCAGGS}. HeLa cells were transiently transfected with mutated SARS-S_{2A}-Y_{PCAGGS} and were examined by epifluorescence microscopy (figure 8a, b). SARS-S_{2A}-Y_{PCAGGS} fluorescence was observed all through the cell and not localized to any particular organelle. Our studies clearly demonstrated that the KXHXX motif is the major intracellular localization signal of the full-length SARS-S protein and the C-terminal proximity is not essential.

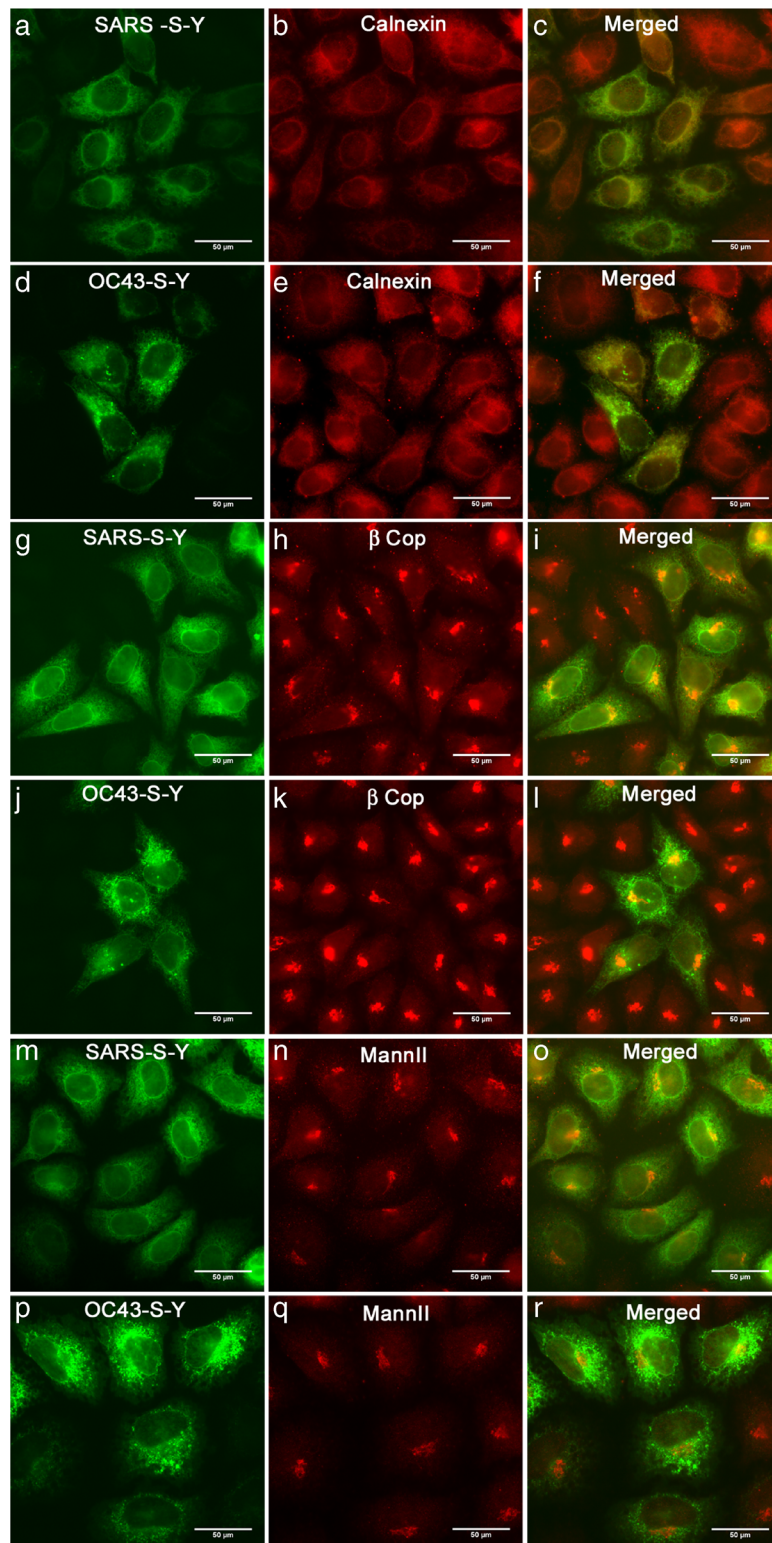


Figure 6. Organellar localization of SARS-S and OC43-S protein. HeLa cells were transiently transfected with SARS-S-Y (**a–c**, **g–i** and **m–o**) and OC43-S-Y (**d–f**, **j–l** and **p–r**), fixed and then immunelabelled using rabbit anti-calnexin (**b** and **e**) or rabbit anti- β -COP (**h** and **k**), rabbit anti-MannII (**n** and **q**), which were stained with Texas Red goat anti-rabbit IgG secondary antibody. Merged images (**c**, **f**, **i**, **l**, **o** and **r**). SARS YFP showed extensive co-localization with the ER marker calnexin, with ERGIC marker β -COP, and *trans*-Golgi marker MannII, whereas OC43 YFP did not show co-localization either with calnexin or with Beta cop or MannII ($n > 3$).

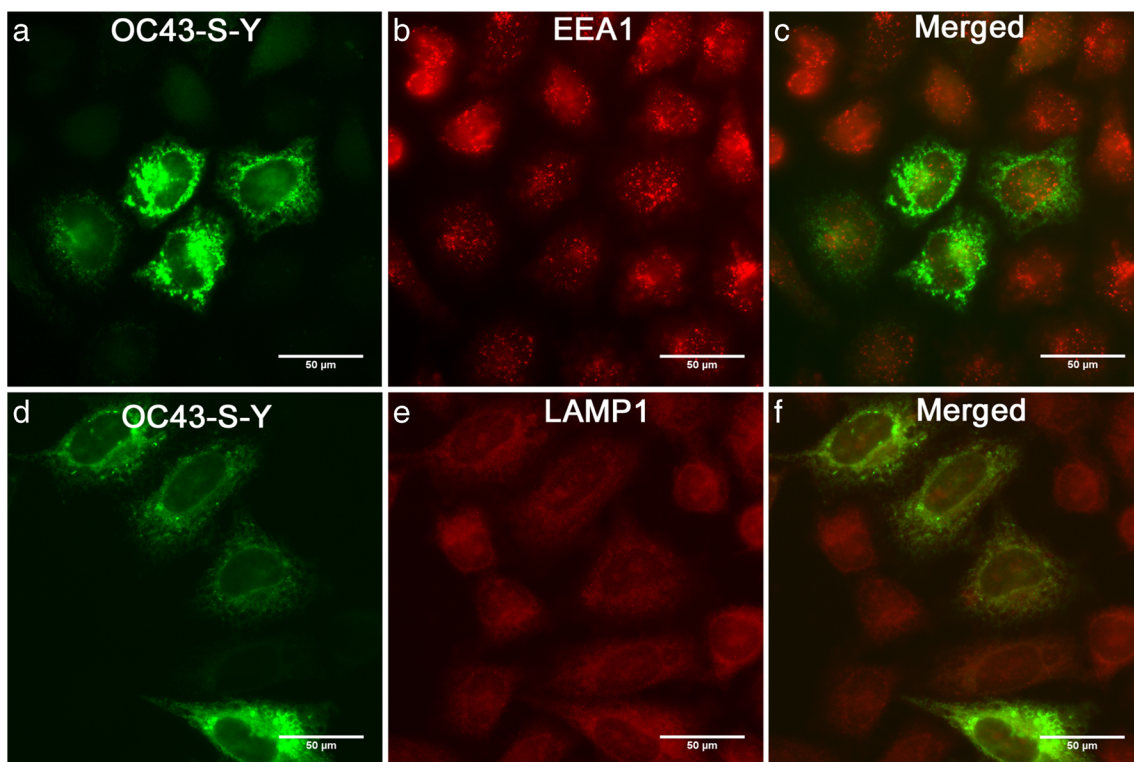


Figure 7. OC43-S protein localized in the lysosomal compartment. OC43 spike protein targeted to lysosomal compartment in transiently transfected cells. HeLa cells were transiently transfected with OC43-S-Y, fixed, and then immunolabelled using rabbit EEA1 (b) or rabbit anti-LAMP1 (e) which were stained with Texas Red goat anti-rabbit IgG secondary antibody. Merged images (c and f) ($n>3$).

3.7 Mutation of critical tyrosine residues of OC43 lysosomal targeting signal causes HCoV-OC43 spike to traffic to cell surface

To resolve whether the YXXØ motif in the cytoplasmic tail of HCoV-OC43 is indeed essential for its intracellular distribution by rapid turnover, we mutated the critical tyrosine residues in the GYXXØ motif to alanines generating the OC43-S_{YtoA}-Y construct. Transiently transfected HeLa cells expressing OC43-S_{YtoA}-Y was examined by immunofluorescence microscopy (figure 8c, d). The majority of cells showed reduction of mutated S protein localization in the intracellular puncta and increased surface expression.

To test whether the glycine residue preceding the critical tyrosine play any role in HCoV-OC43 S protein trafficking, we mutated the glycine residue to alanine residue to generate a construct OC43-S_{GtoA}-Y. If YXXØ in the GYXXØ motif plays a major role in rapid internalization and G facilitates the lysosomal targeting from endocytosed spike protein after reaching the plasma membrane then the mutated construct are expected to continue to be internalized at the same rates but recycle to the plasma membrane, instead of going to the lysosomes. If GYXXØ motif acts as a direct lysosomal targeting signal then the construct

are expected to localize primarily in the perinuclear region (due to its large luminal domain) as well as in the surface (by default bulk flow) like other beta-coronavirus.

When we transfected the cells with the G to A mutated construct interestingly we observed that S protein goes readily to the surface and also retains in the perinuclear region (figure 8e, f). The expression level was much higher in the same experimental condition in compare to the non-mutated wild-type OC43-S-Y construct (data not shown). Due to higher level of expression we observed a blazing signal in the nucleus. G to A construct also induces cell to cell fusion in transfected cells which we never observed before. Most likely G residue plays a critical role in the GYXXØ motif of HCoV-OC43-S tail.

We can conclude that the tyrosine residue may be essential for rapid internalization, and glycine residues is involved in facilitating the distribution of rapidly internalized S protein to lysosome or it playing a role in direct lysosomal sorting.

4. Discussion

The control of protein translocation during trafficking in coronaviruses have features common to other organisms.

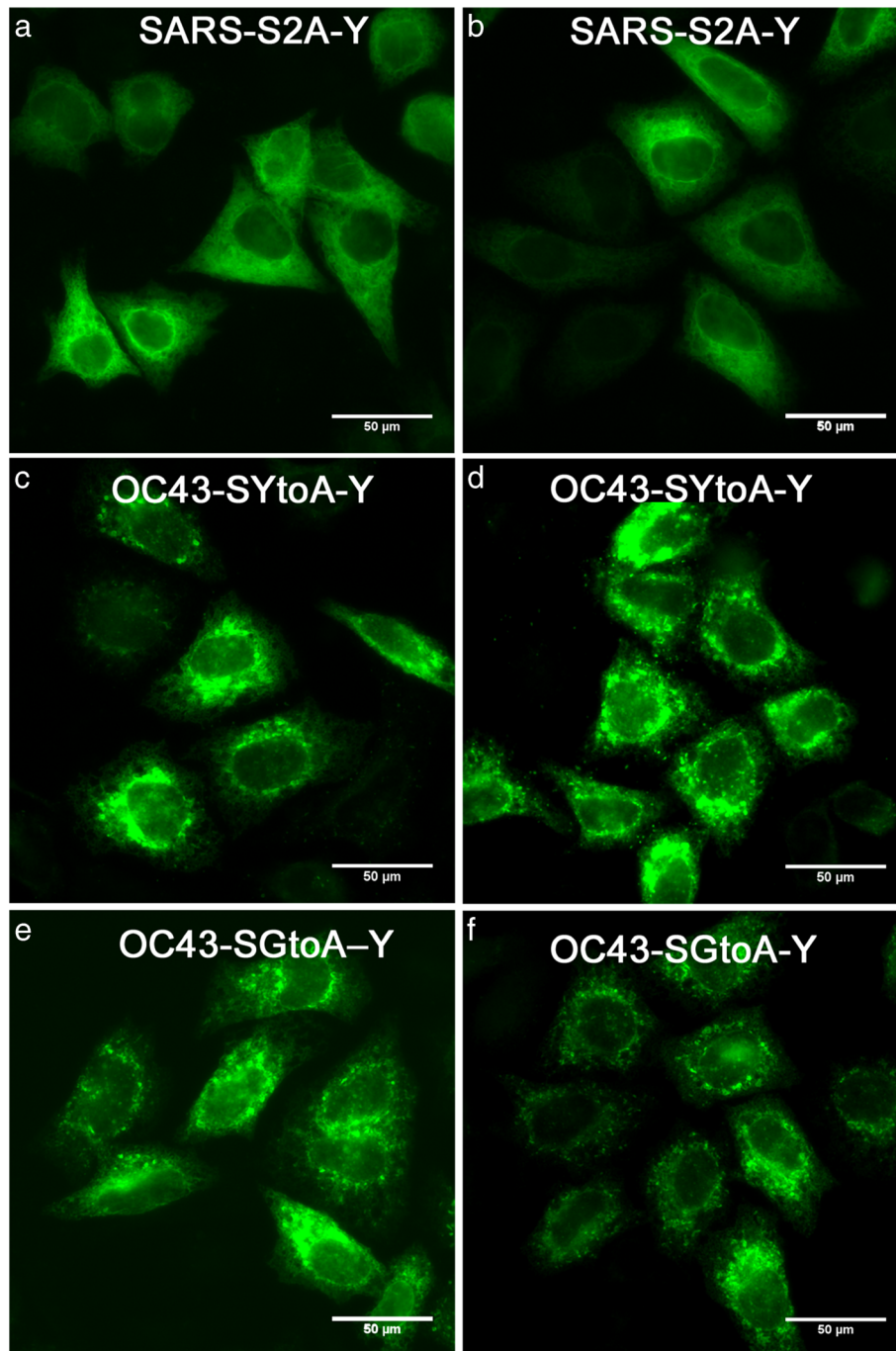


Figure 8. Mutagenesis revealed the targeting signal is required for different organellar localization. Intact HeLa cells expressing HCoV-SARS-S_{2A}-Y, HCoV-OC43-S_{YtoA}-Y and HCoV-OC43-S_{GtoA}-Y were fixed for YFP expression. SARS-S_{2A}-YpCAGGS fluorescence was readily observed all through the cell and not localized to any particular organelle (**a** and **b**). Increased surface expression of OC43-S_{YtoA}-Y can be seen (**c** and **d**) with reduced punctate staining intracellularly (**d**). OC43-S_{GtoA}-Y can be seen on the surface as well as puncta intracellularly (**e** and **f**) (n>3).

Previous studies have suggested that the presence of dibasic residue pair and its specific location relative to the transmembrane domain and the carboxy terminal is critical for ER

localization activity. Specific studies propose a minimum of 5 amino acid residues to be mandatory between the first charged residue of the cytoplasmic tail and the -3 lysine such

that the signal sequence be exposed from the membrane lipids in order to function (Teasdale and Jackson 1996). Our alanine mutation studies on the KXHXX motif confirm the importance of the lysine and histidine in the full-length wild-type HCoV-SARS S protein; the mutant protein showed localization in plasma membrane instead of the usual ER and ERGIC. The carboxy terminal proximity of the motif was not found mandatory. The C-terminal S-YFP-tagged fusion protein also showed ER/ERGIC localization. YFP is a 120-residue protein, and when fused to the C-terminal end along with an artificial linker makes the dibasic motif XXX residues carboxy terminal distal. We have not systematically examined the maximum length of a cytoplasmic domain of S protein on which a lysine-based motif still functions. There is putative evidence of dilysine motif 546 amino acids from the predicted membrane-spanning region and X residues from the carboxy terminal end in the ER localizing HMG CoA reductase (Luskey and Stevens 1985). Other ER localization signals like the RXR motif identified in certain oligomeric ion channels do not require proximity to either the N or C terminus. This suggest carboxy terminal proximity of the dibasic motif is not a general requirement for localization function, although, exposure from the membrane lipids may be necessary.

The dibasic signals are present in the S protein from alpha and gamma and the SARS- coronavirus; there was no such obvious dibasic motif in the cytoplasmic tail of beta-coronavirus. However, we reported that a tyrosine-based signal (GYQEL), at the 23–27 residues of the carboxy terminal tail of HCoV-OC43 S protein, is a motif of the class YXXØ conferring sorting information onto transmembrane proteins. The Y residue is essential for function in most cases. X can be any amino acid and are highly variable but tend to be hydrophilic. The Ø position can accommodate several residues with bulky hydrophobic side chains, although the exact identity of this residue can specify the properties of the signal (Canfield *et al.* 1991). However, the X residues and residues flanking the motif also contribute to the strength and fidelity of the signals. The topological location of the YXXØ motif on the context of a specific protein is also as important as the actual amino acid sequence. Depending on the context of the YXXØ motif in the cytoplasmic tail of the protein, YXXØ motif can function as a rapid internalization from the surface (Mellman 1996; Kirchhausen *et al.* 1997; Marks *et al.* 1997), lysosomal targeting, localization to specialized endosomal-lysosomal organelles such as antigen-processing compartments as well as basolateral surface in polarized epithelial cells. Purely endocytic YXXØ signals are most often situated at 10–40 residues from the transmembrane domains, but not at the carboxy termini of the proteins (Bonifacino and Traub 2003). In contrast, lysosomal targeting YXXØ signals are conspicuous for their presence at 6–9 residues from the transmembrane domain and are often located at the carboxy terminus of short cytoplasmic

tail preceded by a glycine (Bonifacino and Traub 2003). Their effectiveness may be modified by spacing the GYXXØ motif relative to the membrane bilayer (Rohrer *et al.* 1996). Lysosomal-sorting YXXØ motifs tend to have acidic residues at the X positions, and these may also contribute to the efficiency of lysosomal targeting (Rous *et al.* 2002). These signals can be found in all types of transmembrane protein, including Type I (e.g. LAMP-1 and LAMP-2), Type II (e.g., the transferrin and a sialoglycoprotein receptors) and multispinning integral membrane proteins (e.g., CD63 and cystinosin). The importance of the distance from the transmembrane domain has been emphasized by a study showing that changing the spacing of the GYQTI signal from LAMP-1 impairs targeting to lysosomes (Rohrer *et al.* 1996). The mutant proteins continue to be internalized at the same rates but recycle to the plasma membrane, a behavior typical of endocytic receptors (Rohrer *et al.* 1996). These observations indicate that the placement of YXXØ signals at 6–9 residues from the transmembrane domain allows their recognition as lysosomal targeting signals at the TGN and/or endosomes.

Very recently it has been demonstrated that the GYXXØ motifs in their cytosolic tail can take both direct and indirect (via plasma membrane) traffic routes from the *trans*-Golgi network (TGN) to lysosomes. LAMP-1 and LAMP-2 which contain the GYXXØ at the extreme carboxy terminal tail (after 6 residues from the transmembrane domain) were delivered from TGN to lysosomal compartment via an intracellular route. In contrast, Lysosomal acid phosphatase, also a type I membrane protein with a cytosolic tail GYXXØ motif located 7 residues from both transmembrane domain and the carboxy termini (Tm-RMQAQPPEGYRHVADGEDHA) delivered mainly via the cell surface (Braun *et al.* 1989).

The GYXXØ motif in HCoV-OC43 spike cytoplasmic tail could act as an endocytic signal or lysosomal targeting signal. It has been known that the position of actual amino acid sequence of YXXØ signals within the cytoplasmic domain is important (Bonifacino and Traub 2003). As GYXXØ motif is not in close sequential proximity to the transmembrane domain, it predominantly acts as lysosomal targeting signal rather than rapid internalization or endocytic signal. It is also possible that when YXXØ motif acts as a rapid internalization signal, the presence of glycine (due to its special physico-chemical properties) preceding the critical tyrosine may facilitate the targeting of internalized protein to the lysosome.

While HCoV-SARS is a deadly pathogen causing acute respiratory syndrome, HCoV-OC43 causes only benign common cold. It has been reported that this virus HCoV-OC43 has neuroinvasive properties (Arbour *et al.* 2000). The molecular determinants that may account for the dramatic difference in pathogenic potential of the two HCoV are currently unknown. Previous attempts to uncover the basis of pathogenicity have shown several directions. The primary target of such studies

has been the S proteins which are most amenable for interaction with the host cells due to their convenient location on the mature virion surface. Several evidences support a major role for the S protein in determining the tropism and pathogenesis (Casais *et al.* 2003, Das Sarma 2010). The nucleotide sequencing revealed that alterations in virus virulence were more closely associated with differences in the S protein. The S protein is responsible for attachment to the cellular receptor both for virus cell fusion during viral entry and cell to cell fusion later during infection (reviewed in (Belouzard *et al.* 2012). It contains epitopes for viral neutralization and T-cell response. Zinc metalloproteinase, angiotensin converting enzyme 2 (ACE2) has been identified as the functional receptor for HCoV-SARS (Li *et al.* 2003). No functional receptor is yet known for HCoV-OC43. The fate of the S protein during the virion assembly is therefore important in understanding the basis of the host-pathogen interaction and virus infectivity. The differential localization of the S protein in two closely related HCoV with distinct pathogenic potential offers several insights into the pathologic determinants.

The budding and maturation of the virion in the host cell determines the turnover of the coronavirus infection. Unique localization signals in the corona viral envelope proteins permit them to follow different intracellular trafficking pathway to reach the same budding compartment at the same time for proper virion assembly. Although the localization signals vary from protein to protein as well as virus to virus, in most of the coronaviruses, all three envelope proteins M, E and S possess targeting signals that maintain the localization near the budding site (Vennema *et al.* 1990). There is evidence to suggest that the localization signal in S protein is only active in the unfolded structure in ER, and its specific interactions in the folded form between the M protein leads to its incorporation into the virion, where the retention signals are no longer recognized causing excess S protein to go to the surface. When they reach the surface, they induce cell-cell fusion with uninfected cells forming multinucleated syncytia causing rapid spread of the infection. The syncytia formation is a late stage phenomenon of the infection due to the over production of the S protein, which cannot be retained intracellular.

Our studies on differential S protein localization in the intracellular compartment indicates the probable mode of infection to be dependent on signal-sequence-based protein trafficking. In our study exogenously expressed the full-length HCoV-SARS had a steady state distribution in the ER and ERGIC, and HCoV-OC43 was mostly sorted to lysosomes. The predominant intracellular localization or retention of S protein in the early stage of infection have two implications in pathogenesis: (i) delayed cell-damaging effect which may contribute to maximum virus production, and (ii) absence from the cell surface may help to avoid the defense mechanism, e.g. complement activation. At the same time the surface expression could promote direct cell-to-cell

spread of the infection. The low virulence of HCoV-OC43 can be clearly attributed by its restricted surface transport of its S protein during infection. The sorting of the excess S protein of HCoV-OC43 in the lysosomal compartment may hinder its cell-to-cell spread during infection. Viral production is mainly dependent on the intracellular budding. Cell-to-cell fusion cannot help to spread the disease.

Acknowledgements

This work was supported by the National Multiple Sclerosis Society Grant RG3774A2-11 and Research Grant (BT/PR14260/MED/30/437/2010) from Department of Biotechnology, India, to JDS. We thank IISER Kolkata for providing DST Inspire Support to JS, and University Grant commission, India, for providing support to MS.

References

- Andersson AM and Pettersson RF 1998 Targeting of a short peptide derived from the cytoplasmic tail of the G1 membrane glycoprotein of Uukuniemi virus (Bunyaviridae) to the Golgi complex. *J. Virol.* **72** 9585–9596
- Arbour N, Day R, Newcombe J and Talbot PJ 2000 Neuroinvasion by human respiratory coronaviruses. *J. Virol.* **74** 8913–89121
- Belouzard S, Millet JK, Licitra BN and Whittaker GR 2012 Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **4** 1011–1033
- Bonifacino JS and Traub LM 2003 Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.* **72** 395–447
- Braun M, Waheed A and von Figura K 1989 Lysosomal acid phosphatase is transported to lysosomes via the cell surface. *EMBO J.* **8** 3633–3640
- Canfield WM, Johnson KF, Ye RD, Gregory W and Kornfeld S 1991 Localization of the signal for rapid internalization of the bovine cation-independent mannose 6-phosphate/insulin-like growth factor-II receptor to amino acids 24–29 of the cytoplasmic tail. *J. Biol. Chem.* **266** 5682–5688
- Casais R, Dove B, Cavanagh D and Britton P 2003 Recombinant avian infectious bronchitis virus expressing a heterologous spike gene demonstrates that the spike protein is a determinant of cell tropism. *J. Virol.* **77** 9084–9089
- Corse E and Machamer CE 2002 The cytoplasmic tail of infectious bronchitis virus E protein directs Golgi targeting. *J. Virol.* **76** 1273–1284
- Cosson P and Letourneur F 1994 Coatamer interaction with di-lysine endoplasmic reticulum retention motifs. *Science* **263** 1629–1631
- Das Sarma J 2010 A mechanism of virus-induced demyelination. *Interdiscip. Perspect. Infect. Dis.* **2010** 109239
- Garoff H, Hewson R and Opstelten DJ 1998 Virus maturation by budding. *Microbiol. Mol. Biol. Rev.* **62** 1171–1190

- Gaynor EC, Graham TR and Emr SD 1998 COPI in ER/Golgi and intra-Golgi transport: do yeast COPI mutants point the way? *Biochim. Biophys. Acta.* **1404** 33–51
- Goepfert PA, Shaw K, Wang G, Bansal A, Edwards BH and Mulligan MJ 1999 An endoplasmic reticulum retrieval signal partitions human foamy virus maturation to intracytoplasmic membranes. *J. Virol.* **73** 7210–7217
- Hogue BG, King B and Brian DA 1984 Antigenic relationships among proteins of bovine coronavirus, human respiratory coronavirus OC43, and mouse hepatitis coronavirus A59. *J. Virol.* **51** 384–388
- Holmes KV, Frana MF, Robbins SG and Sturman LS 1984 Coronavirus maturation. *Adv. Exp. Med. Biol.* **173** 37–52
- Kirchhausen T, Pines J, Toldo L and Lafont F 1997 Membranes and sorting. Membrane permeability. *Curr. Opin. Cell Biol.* **9** 473
- Krijnse-Locker J, Ericsson M, Rottier PJ and Griffiths G 1994 Characterization of the budding compartment of mouse hepatitis virus: evidence that transport from the RER to the Golgi complex requires only one vesicular transport step. *J. Cell Biol.* **124** 55–70
- Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, Somasundaran M, Sullivan JL, *et al.* 2003 Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426** 450–454
- Lontok E, Corse E and Machamer CE 2004 Intracellular targeting signals contribute to localization of coronavirus spike proteins near the virus assembly site. *J. Virol.* **78** 5913–5922
- Luskey KL and Stevens B 1985 Human 3-hydroxy-3-methylglutaryl coenzyme A reductase. Conserved domains responsible for catalytic activity and sterol-regulated degradation. *J. Biol. Chem.* **260** 10271–10277
- Machamer CE and Rose JK 1987 A specific transmembrane domain of a coronavirus E1 glycoprotein is required for its retention in the Golgi region. *J. Cell Biol.* **105** 1205–12014
- Marks MS, Ohno H, Kirchnausen T and Bonracino JS 1997 Protein sorting by tyrosine-based signals: adapting to the Ys and wherefores. *Trends Cell Biol.* **7** 124–128
- Matsuoka Y, Chen SY and Compans RW 1991 Bunyavirus protein transport and assembly. *Curr. Top. Microbiol. Immunol.* **169** 161–179
- Mcwilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP and Lopez R 2013 Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* **41** W597–600
- Mellman I 1996 Endocytosis and molecular sorting. *Annu. Rev. Cell Dev. Biol.* **12** 575–625
- Mora R, Rodriguez-Boulan E, Palese P and Garcia-Sastre A 2002 Apical budding of a recombinant influenza A virus expressing a hemagglutinin protein with a basolateral localization signal. *J. Virol.* **76** 3544–3553
- Nilsson T, Jackson M and Peterson PA 1989 Short cytoplasmic sequences serve as retention signals for transmembrane proteins in the endoplasmic reticulum. *Cell* **58** 707–718
- Niwa H, Yamamura K and Miyazaki J 1991 Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene.* **108** 193–199
- Owens RJ and Compans RW 1989 Expression of the human immunodeficiency virus envelope glycoprotein is restricted to basolateral surfaces of polarized epithelial cells. *J. Virol.* **63** 978–982
- Paabo S, Bhat BM, Wold WS and Peterson PA 1987 A short sequence in the COOH-terminus makes an adenovirus membrane glycoprotein a resident of the endoplasmic reticulum. *Cell* **50** 311–317
- Rohrer J, Schweizer A, Russell D and Kornfeld S 1996 The targeting of Lamp1 to lysosomes is dependent on the spacing of its cytoplasmic tail tyrosine sorting motif relative to the membrane. *J. Cell Biol.* **132** 565–776
- Rous BA, Reaves BJ, Ihrke G, Briggs JA, Gray SR, Stephens DJ, Banting G and Luzio JP 2002 Role of adaptor complex AP-3 in targeting wild-type and mutated CD63 to lysosomes. *Mol. Biol. Cell* **13** 1071–1082
- Schwegmann-Wessels C, Al-Falah M, Escors D, Wang Z, Zimmer G, Deng H, Enjuanes L, Naim HY, *et al.* 2004 A novel sorting signal for intracellular localization is present in the S protein of a porcine coronavirus but absent from severe acute respiratory syndrome-associated coronavirus. *J. Biol. Chem.* **279** 43661–43666
- Swift AM and Machamer CE 1991 A Golgi retention signal in a membrane-spanning domain of coronavirus E1 protein. *J. Cell Biol.* **115** 19–30
- Teasdale RD and Jackson MR 1996 Signal-mediated sorting of membrane proteins between the endoplasmic reticulum and the Golgi apparatus. *Annu. Rev. Cell Dev. Biol.* **12** 27–54
- Tooze J, Tooze SA and Fuller SD 1987 Sorting of progeny coronavirus from condensed secretory proteins at the exit from the trans-Golgi network of AtT20 cells. *J. Cell Biol.* **105** 1215–1226
- Vennema H, Rottier PJ, Heijnen L, Godeke GJ, Horzinek MC and Spaan WJ 1990 Biosynthesis and function of the coronavirus spike protein. *Adv. Exp. Med. Biol.* **276** 9–19
- Winter C, Schwegmann-Wessels C, Neumann U and Herrler G 2008 The spike protein of infectious bronchitis virus is retained intracellularly by a tyrosine motif. *J. Virol.* **82** 2765–2771

MS received 02 August 2016; accepted 14 February 2017

Corresponding editor: SAUMITRA DAS