The C-terminal domain of the MERS coronavirus M protein contains a trans -Golgi network localization signal

Anabelle Perrier, Ariane Bonnin, Lowiese Desmarets, Adeline Danneels, Anne Goffard, Yves Rouillé, Jean Dubuisson, Sandrine Belouzard

The C-terminal domain of the MERS coronavirus M protein contains a *trans*-Golgi network localization signal

Running title: MERS-CoV M protein TGN localization

*Anabelle Perrier, Ariane Bonnin, Lowiese Desmarets, Adeline Danneels, Anne Goffard, Yves Rouillé, Jean Dubuisson and Sandrine Belouzard**

Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL- Center for Infection and Immunity of Lille, F-59000 Lille, France

*To whom correspondence should be addressed: sandrine.belouzard@ibl.cnrs.fr

Keywords: Viral protein, protein motif, protein sorting, intracellular trafficking, plasma membrane, endoplasmic reticulum, coronavirus, Middle East respiratory syndrome, trans-Golgi network localization

ABSTRACT:

Coronavirus M proteins represent the major protein component of the viral envelope. They play an essential role during viral assembly by interacting with all the other structural proteins. Coronaviruses bud into the endoplasmic reticulum (ER)–Golgi intermediate compartment (ERGIC), but the mechanisms by which M proteins are transported from their site of synthesis, the ER, to the budding site remain poorly understood. Here, we investigated the intracellular trafficking of the Middle East respiratory syndrome coronavirus (MERS-CoV) M protein. Subcellular localization analyses revealed that the MERS-CoV M protein is retained intracellularly in the *trans*-Golgi network (TGN), and we identified two motifs in the distal part of the C-terminal domain as being important for this specific localization. We identified the first motif as a functional diacidic DxE ER export signal, since substituting Asp-211 and Glu-213 with alanine induced retention of the MERS-CoV M in the ER. The second motif, $_{199}$KxGxYR$_{204}$, was responsible for retaining the M protein in the TGN. Substitution of this motif resulted in MERS-CoV M leakage toward the plasma membrane. We further confirmed the role of $_{199}$KxGxYR$_{204}$ as a TGN retention signal by using chimeras between MERS-CoV M and the M protein of infectious bronchitis virus (IBV). Our results indicated that the C-terminal domains of both proteins determine their specific localization, namely, TGN and ERGIC/*cis*-Golgi for MERS-M and IBV-M, respectively. Our findings indicate that MERS-CoV M protein localizes to the TGN because of the combined presence of an ER export signal and a TGN retention motif.

Coronaviruses are widespread pathogens that can infect a wide variety of species among mammals and birds (1, 2), including humans, causing mostly respiratory and enteric symptoms. There are 6 known coronaviruses infecting humans. The first human coronaviruses HCoV-229E

and HCoV-OC43 were isolated in the 60s from patients suffering from a common cold. Administration of these viruses to volunteers rapidly confirmed their harmless character. Therefore, research on coronaviruses had been mostly of veterinary interest, but this changed recently with the emergence of two highly pathogenic human coronaviruses causing severe pneumonia epidemics. First, the severe acute respiratory syndrome coronavirus (SARS-CoV) appeared in 2002, then the Middle East respiratory syndrome coronavirus (MERS-CoV) appeared in 2012. Both viruses have a zoonotic origin showing that this virus family is a reservoir of emerging pathogens, especially because of their high interspecies transmission (3).

Coronaviruses are enveloped positive single-stranded RNA viruses, with a very large genome of 25-30kb, belonging to the *Coronaviridae* Family in the *Nidovirales* order. The viral particle is composed of a lipid envelope in which at least three structural proteins are anchored: the *spike* protein (S), the *envelope* protein (E) and the *membrane* protein (M). Inside the particle, the viral RNA is associated with the *nucleocapsid* protein (N) forming a helical capsid. The S protein triggers viral entry by binding to the cellular receptor and mediating fusion of the viral envelope with the host cell membrane (4, 5). The E protein is a small protein with multiple roles during infection (6). The SARS-CoV E protein plays an important role in viral pathogenesis, and this role can be linked to the ion channel activity of the protein (7). The E protein is also involved in viral assembly, trafficking and egress of virions, by promoting membrane curvature, viral fission, and by inducing morphological changes of the compartments of the secretory pathway (8, 9).

The M protein is the most abundant protein of the envelope (10). Its length ranges from 217 to 230 amino acid residues in most coronaviruses, but it can go up to 270 residues in some coronaviruses (Bottlenose Dolphin Coronavirus). It is a protein with three membrane-spanning hydrophobic segments, a small N-terminal domain located outside the virion (or inside the lumen of intracellular organelles) and a large C-terminal domain that makes up half of the protein, inside the virion (or in the cytoplasm of infected cells) (10). M proteins of some *Alphacoronaviruses* contain an additional hydrophobic segment that functions as a signal peptide. The M protein is invariably glycosylated on its amino-terminal domain. However, there are differences in the type of glycosylation. The murine hepatitis virus (MHV) and some other *Betacoronavirus* M proteins are O-glycosylated whereas *Alpha-* and *Deltacoronavirus* M proteins are modified with N-linked sugars. The glycosylation is dispensable for virus assembly (11). The M protein is considered to be the motor of the assembly of viral particles since it is able to interact with all of the other structural proteins (12–14). Important M-M interactions have also been demonstrated during assembly (15, 16). For many coronaviruses, including the transmissible gastroenteritis virus (TGEV), MHV and IBV, the co-expression of M and E proteins in cells is sufficient to induce the production of virus-like particles (VLPs) (17–19), indicating that these proteins are essential and sufficient to the assembly step. Only for SARS-CoV, the nucleocapsid protein has been shown to be additionally required for the production of VLPs (20). Electron microscopy observations have shown that coronaviruses bud inside the ER-Golgi intermediate compartment (ERGIC) (21, 22) and then travel through the Golgi. Although the assembly step of coronaviruses occurs in the ERGIC compartment, it has been shown for several coronaviruses that the M protein expressed alone in cells can go beyond the assembly site in the secretory pathway (21). However, the subcellular localization of the M protein varies between the different coronavirus species. Indeed MHV M protein can reach the trans-Golgi network (TGN) whereas IBV-M protein is retained in the ERGIC and one or two cisternae of the cis-Golgi (21, 23, 24). The intracellular retention of IBV-M has been

attributed to the first membrane-spanning segment and particularly to polar residues located within this domain (25, 26). For MHV, both the TM1 and the last 22 amino acids seem to be important determinants for the intracellular localization of the protein (27). The aim of this study was to investigate the subcellular localization of the MERS-M protein and to characterize the signals involved in its trafficking.

Results

MERS-CoV M localizes in the Trans-Golgi network.

The M protein is composed of a short N-terminal domain followed by 3 membrane-spanning segments and a long C-terminal domain (Figure 1A). In order to analyze the intracellular trafficking of the MERS-CoV M protein, we transfected a vector expressing the MERS-CoV M protein with a HA tag fused at its N-terminus (HA-MERS-M) in HeLa cells, and compared the protein localization with different compartment markers using immunofluorescence and confocal microscopy (Figure 1B). As observed in Figure 1B, MERS-CoV M shows a good colocalization with TGN46 that localizes in the trans-Golgi network. We confirmed this colocalization by co-transfection of the M protein with the GFP fused to the transmembrane domain and cytosolic tail of the cation-independent mannose-6-phosphate receptor (GFP-CI-MPR). This reporter has been shown to be localized in the TGN in HeLa cells (28). The images were then analyzed using ImageJ to calculate the Pearson Correlation Coefficient (PCC) for each co-staining (Figure 1C). The PCC measures the pixel-by-pixel covariance of the signal levels of two images (29). The PCC values range from -1 to 1. A PCC value of 1 is obtained for two images whose intensities of fluorescence are linearly and perfectly related whereas a value of -1 means that the intensities are inversely related to one another. Values near 0 mean that the intensities are uncorrelated. Our results confirm the localization of MERS-CoV M in the TGN with a PCC of 0.878 (+/-0.014) and 0.852 (+/-0.012) for the MERS-CoV M with TGN46 and GFP-CI-MPR markers, respectively. The PCC for MERS-CoV M and CD4 (a cell surface marker), or the calreticulin (an endoplasmic reticulum (ER) marker) are below 0.4. However, the PCC between MERS-CoV M and the ERGIC marker, ERGIC-53 is higher (0.524 +/-0.03).
To facilitate the detection of the M protein and particularly in co-staining experiments, different M protein constructs were generated with different tags either at the N-terminus (HA or V5) or at the C-terminus (V5 or VSVG). This offers the advantage of using a panel of antibodies raised in different species. In order to use these tools, we analyzed their intracellular localization to confirm that adding a tag, regardless of their sequence or the position of the insertion, had no effect on MERS-CoV M localization. We also generated a polyclonal antibody raised against a C-terminal peptide of the M protein to detect the untagged protein. This antibody was used to control the effect of the tags on the subcellular localization of the M protein. Cells expressing M or the different tagged versions of this protein were processed for double-label immunofluorescent detection of the M protein and TGN46 (Figure 1D). For each protein, the colocalization level with the TGN46 marker was also quantified by calculating the PCC (Figure 1E). As observed with HA-M, the untagged form of the protein presented a strong colocalization with the TGN46 marker (Figure 1D and 1E) with a PCC of 0.857 (+/-0.015). Similar results were obtained with the other tagged proteins. Altogether, these results indicate that the MERS-CoV M protein expressed alone is located in the TGN and that the tag added to our constructs does not alter this localization.

The last 20 residues of the C-terminal domain of the MERS-CoV M protein are important for the MERS-CoV M intracellular trafficking.

To study the role of the C-terminal domain of MERS-CoV M in its intracellular

trafficking, we constructed serial deletion mutants of 20 residues. The deletion of the last 20 residues of the MERS-CoV M protein (MΔ20) produced mutants that were no longer localized in the TGN (figure 2A). Moreover it induced a different intracellular localization depending on the tag used. Indeed, HA-MΔ20 co-localizes with calreticulin, the ER marker, as shown by the double labeling in Figure 2B; whereas MΔ20-VSVG was located at the cell surface as shown by the double labeling of MΔ20-VSVG and CD4 (Figure 2C). We measured the extent of colocalization for HA-MΔ20 or MΔ20-VSVG with the ER, TGN or cell surface marker by measuring the PCC (Figure 2D). The wild-type protein presented a PCC of 0.905 (+/-0.007) for TGN46 and of 0.296 (+/-0.014) and 0.239(+/-0.017) for CRT and CD4, respectively. The HA-MΔ20 protein showed a decrease of the PCC for TGN46 (0.369+/-0.019) associated with an increase of the PCC with CRT (0.855+/-0.011). The HA-MΔ20 also showed a moderate increase in cell surface localization with a PCC for CD4 of 0.539(+/-0.018). The PCC between MΔ20-VSVG and TGN46 was also strongly decreased (0.331+/-0.025), however the MΔ20-VSVG protein presented a strong increase of the PCC for CD4 (0.79+/-0.016) and a moderate increase of the PCC for CRT (0.475+/-0.018). These results suggest that important intracellular trafficking motifs are present in the distal part of the C-terminal domain of MERS-CoV M. Analysis of the sequence of the last 20 residues showed that the deleted sequence contains a potential di-acidic ER export signal at position 211-213 ($_{211}$DIE$_{213}$). This signal was first characterized for the glycoprotein of VSV and is indeed present in the VSVG tag (YTDIEMNRLGK) used in our experiments. It is likely that a blockade of ER export arose from the deletion in HA-MΔ20 and the loss of the DxE signal, however the addition of the VSVG tag in the MΔ20-VSVG protein restored the signal and rescued the ER export. Moreover, the MΔ20-VSVG protein was mainly located at the cell surface instead of being retained in TGN, suggesting that the distal part of the C-terminal domain of the M protein also contains a determinant responsible for its localization in the TGN.

DxE is a functional ER export signal for MERS-CoV M.

To confirm the role of the DxE signal in the ER export of the MERS-M protein, we mutated the aspartic acid and glutamic acid into alanine (D211A, E213A) in the M protein fused with an N-terminal HA tag (HA-M-DxE) or with a C-terminal VSVG tag (M-DxE-VSVG), and analyzed the subcellular localization of the mutants in confocal microscopy. Similarly to HA-MΔ20, HA-M-DxE was mainly localized in the ER, whereas M-DxE-VSVG was localized in the TGN (Figure 2B and 2A), confirming that the VSVG tag is able to compensate the D211A-E213A mutation. This result demonstrates that the DxE signal present in the C-terminal domain of MERS-CoV M protein is a functional ER export signal involved in the trafficking of the protein.

Four residues in the C-terminal domain mediate MERS-CoV M localization to the TGN.

The presence of the DxE signal explained the exit of the M protein from the ER but not its retention in the trans-Golgi, since it is commonly accepted that in non-polarized cells, the constitutive secretory pathway leads to the plasma membrane by default (i.e in absence of specific addressing/retention signals). Considering that and the fact that the MΔ20-VSVG protein migrates to the cell surface, we looked for the presence of another signal in the last 20 amino acids of the cytosolic tail which could be involved in the retention of MERS-CoV M in the TGN compartment. For this purpose, we constructed three smaller C-terminal deletion mutants lacking 5, 10 or 15 residues, but keeping the VSVG tag to rescue the ER export. These mutants were called MΔ5, MΔ10 and MΔ15. We then compared the subcellular localization of these mutants to

the wild-type and MΔ20 M proteins (Figure 3A, 3B and 3C). Similar to what was observed for the wild-type M and in contrast to MΔ20 protein, the MΔ5, MΔ10 and MΔ15 proteins co-localized with the TGN46 marker, indicating that the 5 amino acids sequence AGNYR, located between the Δ20 and Δ15 deletions is likely involved in the specific localization of the protein in the TGN. To identify the residues involved in the TGN localization of the protein, each amino acid (except the alanine) was mutated individually into alanine in the MΔ15 protein. MΔ15-G201A, MΔ15-N202A, MΔ15-Y203A and MΔ15-R204A were expressed in HeLa cells and the subcellular localization of the proteins was analyzed by fluorescence microscopy (Figure 4). The proteins carrying the mutations G201A, Y203A or R204A showed a reduced colocalization with the TGN46 marker compared to the MΔ15 (Figure 4A and 4C), with partial export of the protein to the cell surface (Figure 4B and 4C). This difference in intracellular trafficking is illustrated by a decrease of PCC between the M mutants and TGN46 associated with an increase of the PCC with CD4. These results indicate a role of these three residues in the localization of the protein in the TGN. The mutation N202A had no effect on the subcellular localization of the protein compared to the wild type. To confirm these results, the mutations G201A, Y203A or R204A were also inserted in the full-length protein. Analysis of the subcellular localization of these mutants is shown in Figure 5. M-G201A, M-Y203A and M-R204A showed a reduced co-localization with TGN46 compared to the wild-type M-VSVG and an increase of cell surface expression resulting in an increase of the PCC between the mutant and CD4. To ensure that we identified the full motif involved in the TGN localization of the protein, we also mutated 3 conserved residues (Y195, R197 or K199) among the *Betacoronaviruses* located upstream of the Δ20 deletion. Interestingly, mutation of the residue K199 also resulted in an increase of the cell surface expression of the protein (Figure 5).

We also constructed a quadruple mutant protein, M-K199A, G201A, Y203A, R204A (M-KGYR). The extent of colocalization of the quadruple mutant (M-KGYR) with TGN46 and CD4 was in the same range than those of the single mutants.

To confirm the cell surface expression of the M protein when the residues K199, G201, Y203 and R204 are mutated, we performed a cell surface biotinylation assay. The MERS-M protein contains a N-glycosylation site in its N-terminal domain (MSNMTQLTE), consequently the migration profile of the MERS-M protein in immunoblot renders the quantification of the protein amount difficult (see Figure 6C), so we also mutated the N-glycosylation site in the different mutants. First, we verified that introducing the mutation N3Q had no effect on the intracellular localization of the different proteins (M, MΔ20, M-DxE, M-K199A, M-G201A, M-Y203A, M-R204A and M-KGYR; data not shown). Plasma membrane proteins in cells expressing the different mutants were labeled with non-permeable biotin, then biotinylated proteins were precipitated with streptavidin-conjugated agarose beads and analyzed in immunoblot (Figure 6A and 6B). The M protein is only weakly expressed at the cell surface, with less than 1% of the total amount of protein expression at the cell surface. Mutation of the residues K199, G201, Y203 and R204 alone or in combination induced an increase in cell surface detection of the M protein with approximately 13% of the total amount of N3Q-M-KGYR located at the cell surface. Single mutations induced only moderate increases of the cell surface expression compare to the quadruple mutant. The N3Q-M-DxE was barely detected at the cell surface. As seen in immunofluorescent colocalization assay (Figure 6B), the expression of the N3Q-MΔ20 mutant at the cell surface was slightly increased compared to the wild-type protein. Together, these results indicated that the residues K199, G201, Y203 and R204 of MERS-CoV M are

involved in its specific localization in the TGN.

We also investigated the N-glycosylation status of the DxE and KxGxYR mutants. In immunoblot, the wild-type protein migrated as 3 bands (Figure 6C). A first band of around 20-25 kDa corresponds to the unglycosylated M protein, confirmed by the migration profile of the protein in which the N-glycosylation site was abolished by mutation (N3Q-M) and by treatment with PNGase F. A second band of approximately 30 kDa and a third more diffuse band migrating more slowly were also observed. As expected, the second band is sensitive to Endoglycosidase H treatment, showing that this band corresponds to M proteins glycosylated in the ER that have not reached the Golgi yet. The more diffuse band was not sensitive to EndoH treatment, suggesting that this form is further modified in the Golgi. Only one band was observed with the DxE mutant with a high-intensity at 30 kDa sensitive to EndoH . The EndoH sensitivity of the 30 kDa band likely reflects the accumulation of this protein in the ER due to the lack of export signal.
It is worth noticing that the migration profile of the M-KGYR mutant in western blot showed an increased N-glycosylation consistent with a better trafficking through the Golgi (Figure 6C).
All together these results confirm the presence of two intracellular trafficking motifs in the C-terminal domain of the MERS-CoV M protein: first the DxE motif is responsible for the ER export of the protein, then a second motif KxGxYR is involved in its retention in the TGN.

KxGxYR is a not an internalization signal and is not involved in M oligomerization.
Intracellular trafficking is a dynamic process with proteins that can undergo cycles of cell surface expression and internalization. At steady state, the localization of proteins at the plasma membrane results from an equilibrium between anterograde intracellular trafficking and retrieval of protein by endocytosis. Any inhibition of endocytosis would result in protein accumulation at the cell surface. To test if the KxGxYR motif is an endocytosis signal, we analyzed the endocytosis of the M protein in a biotinylation assay. Proteins expressed at the plasma membrane of cells expressing N3Q-M or N3Q-M-KGYR were labeled at 4°C with a non-permeable cleavable biotin, then endocytosis was allowed by incubating the cells at 37°C for 30 min. Non-internalized biotin was then cleaved with glutathione and internalized proteins were detected in immunoblot. As shown in Figure 6E and 6F, we did not detect any difference in endocytosis levels between the wild-type protein and M-KGYR.
It has been proposed that oligomerization of MHV-M protein could be involved in its TGN retention. Indeed mutants that do not form oligomers were detected at the cell surface. In cell lysates, the mutant N3Q-M-KGYR forms dimer in comparable amount to the wild-type protein (Figure 6D). We also analyzed the formation of N3Q-M-KGYR oligomers at the cell surface. To do so, the cell proteins at the cell surface were biotinylated and cross-linked. The formation of multimers was detected in immunoblot (Figure 6D). As previously shown, expression of N3Q-M-KGYR at the cell surface was increased. We detected the formation of dimers with a strong band migrating at 40 kDa and also the formation of higher oligomers for both proteins, suggesting that the increased cell surface expression conferred by the mutation of the motif KxGxYR is not due to a defect in M-M interactions.
These results indicate that the KxGxYR signal is not involved in M oligomerization and is not an internalization signal, and that the localization of MERS-CoV M in the TGN is likely due to a mechanism of retention, preventing the cell surface expression of the protein.

IBV C-terminal domain is involved in its

ERGIC localization.

To confirm the role of the KxGxYR as a retention signal in the TGN, we tried to transfer this signal on a protein expressed at the cell surface. We used CD4 as a reporter, however the chimeric proteins that were constructed presented folding defects (data not shown). CD4 is a type-I transmembrane protein whereas the M protein has a very different architecture with 3 transmembrane segments. Therefore, we constructed chimeras between MERS-CoV M and the M protein of another coronavirus, IBV. This way, we were able to construct chimeras that conserved the transmembrane domain structure of the protein, which is likely important for the folding and localization of the protein. The IBV-M protein expressed alone in cells is located in the ERGIC and cis-Golgi (26), so its localization can be distinguished from the one of MERS-CoV M, using specific compartment markers. In addition, it has been shown that the first transmembrane segment of IBV-M is involved in the intracellular retention of the protein. The amino acid sequences of the MERS- and IBV-M C-terminal extremity are not conserved and are rather different (Figure 7A).

First, we constructed chimeras in which we switched the C-terminal domains of the proteins: MERS-M/IBV-M and IBV-M/MERS-M or IBV-M/MERS-M-KGYR. We also replaced the first transmembrane segment of the MERS-M-KGYR with the one of IBV-M to test if the first transmembrane segment of IBV-M can retain MERS-M-KGYR intracellularly (TM1-IBV/MERS-M-KGYR). Finally, we replaced the first transmembrane segment of IBV-M with the one of MERS-CoV M. Schematic drawings of the different chimeras that were constructed are presented in Figure 7B. The subcellular localization of the chimeric proteins was then analyzed by fluorescence microscopy. As shown in Figure 8, the immunofluorescent staining of IBV-M and MERS-CoV M differed in their pattern, with a compact perinuclear staining for MERS-CoV M and a punctuated staining for IBV-M. Co-localization assessment of IBV-M with TGN46 and ERGIC-53 showed that the protein mainly localizes within the ERGIC. The MERS-M/IBV-M protein colocalized with the ERGIC-53 marker and not the TGN46 marker, and is thus located in the ERGIC (Figure 8C), whereas the IBV-M/MERS-M protein colocalized with the TGN46 marker and not with the ERGIC-53 marker, and is thus localized in the TGN. In other words, the switch of the C-terminal domains of the IBV-M and MERS-CoV M proteins caused a switch of their specific localizations, respectively to the ERGIC and the TGN. Interestingly, the IBV-M/MERS-KGYR protein localized to the cell surface, confirming the role of the KxGxYR signal in the specific localization of the MERS-M protein to the TGN. This result also suggests that the first transmembrane segment of IBV-M is not able to retain MERS-M-KGYR intracellularly. In accordance with this result, the chimera TM1-IBV/MERS-M-KGYR was also located at the cell surface and the chimera TM1-MERS/IBV-M was located in the ERGIC compartment. These results are unexpected based on previous reports on the role of the first transmembrane segment of IBV-M in its intracellular retention. Moreover, these results indicate that for both MERS-CoV M and IBV-M, the presence of the C-terminal domain is critical to induce the specific localization of the protein. Furthermore, we also confirmed the involvement of the KxGxYR signal in the specific retention of MERS-CoV M to the TGN, even in the chimeric context.

Discussion

Viruses divert the intracellular trafficking machinery and studying the intracellular trafficking of viral membrane proteins often helps to decipher the mechanisms of protein sorting and leads to uncovering new sorting motifs. We investigated the intracellular trafficking of the MERS-CoV M protein and identified a well-known ER export signal. In addition, we identified a novel TGN retention motif.

Specific targeting of viral structural proteins to the assembly site in the cell is crucial for viral egress and spreading. The three envelope proteins of coronaviruses E, M and S are synthesized in the ER. Protein exit from the ER toward the Golgi occurs at specific sites called ER exit sites (ERES) and for most of the proteins relies on the coat protein complex II (COPII). Assembly of the coat starts with the activation of the GTPase Sar1 by Sec12, an integral ER guanine nucleotide exchange factor. This allows the recruitment of the complex Sec23/24 forming the inner layer of the coat followed by the recruitment of the outer layer of the coat formed by Sec13/31 (30). ER export signals, including LxxLE, diacidic DxE, YNNSNP or triple R, can interact directly with Sec24 or Sar1 and lead to the recruitment of the COPII carriers (31). We identified a functional DxE motif in the C-terminal part of the MERS-CoV M protein, and this signal is also present in the M protein of the porcine epidemic diarrhea virus (PEDV). Some *Alpha-* and *Gammacoronavirus* M proteins, such as HCoV-229E, HCoV-NL63, FIPV or IBV, contain a diacidic ExE motif, however it remains to be determined whether these signals can also act as ER export signal since, in the yeast protein Sys1p, ExE cannot compensate for the DxE signal (32).

After its exit from the ER, the MERS-CoV M protein reaches the Golgi where the protein undergoes further modification of its N-glycans as shown in our glycosidase-resistance assays. As previously shown for MHV-M protein, we and others (33) found that ectopically expressed MERS-CoV M protein is mainly located in the TGN at steady state. In the cells, the TGN is a sorting station where proteins are either sent to the cell surface or diverted toward other endomembrane compartments. The localization of proteins in specific biosynthetic compartments generally results from an equilibrium between anterograde and retrograde movements of the proteins. For example, in immunofluorescent labeling of the protein TGN38/46, the protein is located in the TGN but this localization is the consequence of a very dynamic process in which the protein is transported to the plasma membrane and recycled back to the TGN after internalization and sorting in endosomes. The SDYQRL motif in the C-terminal domain of the protein is important for its retrieval from the cell surface (34). In addition, the transmembrane domain of the protein also participates in the TGN localization of the protein by mediating some retention in the TGN (35). An increase of the transport of the protein to the cell surface can lead to the saturation of the retrieval mechanism of the protein and its accumulation at the cell surface.

Here, we identified a KxGxYR motif in the MERS-CoV M protein which is involved in the TGN localization of the protein. The mutation of any residue of this motif leads to the accumulation of the protein at the plasma membrane. This signal is highly conserved in the M proteins of *Betacoronavirus*. Interestingly, Armstrong *et al*. previously reported that deletion of the last 18 residues of the C-terminal domain of the MHV-M protein induced a shift of the protein localization toward the cell surface (36). Interestingly, this deletion is in the middle of the KxGxYR signal, leaving only the K and G residues on the truncated protein. In another study, Krijnse-Locker *et al*. also reported a deletion of 22 residues of the protein leading to the accumulation of the protein at the cell surface (27). Furthermore, in their study about the structure requirements of MHV-M protein, De Haan *et al*. mentioned that one of their mutant containing a mutation of the KxGxYR motif (Y211G) leaked to the cell surface (11). These data suggest that TGN localization of M proteins may be a general feature of the *Betacoronaviruses* and that the KxGxYR motif is involved in this localization. It has been reported that the SARS-CoV M protein is located in the Golgi, unfortunately its precise localization within the Golgi remains unclear (37, 38).

Locker *et al*. reported that oligomerization of MHV-M protein is involved in the TGN

retention of the protein, however our results show that it is unlikely that the KxGxYR motif is involved in the oligomerization of the protein (39). As a small proportion of the MERS-CoV M protein could be detected at the cell surface (less than 1% of the total protein, see Fig 6A and 6B), we tested if the KxGxYR motif could be an endocytosis motif. The MERS-CoV M protein is retrieved from the cell surface by endocytosis, however we could not detect any defect of endocytosis of the M protein when the KxGxYR motif was mutated (Fig 6E and 6F). This result argues against an endocytic function of this motif and suggests a role as a retention signal. Indeed, it was previously reported that the MHV-M protein does not cycle between the plasma membrane of the cell and the TGN but rather acts as a TGN resident protein. Another hypothesis is a cycle of the M between the ER and the TGN with the KxGxYR motif acting as a retrieval signal from the TGN to the ER. The mutation of this signal would inhibit the retrieval of the protein allowing its trafficking to the cell surface. The retrograde trafficking ensures the constant recycling of proteins and lipids from the Golgi to the ER to maintain the their steady-state distribution and the composition and function of the organelles themselves. Two distinct mechanisms are responsible for this retrograde transport. The first one is dependent on the coat protein complex I (COPI) and the second one is the COPI-independent pathway that is less characterized, involves the Rab6 GTPase and is composed of tubular rather than vesicular carriers (40). The formation of the COPI-coated vesicles starts with the recruitment en bloc of the coatomer composed of seven subunits by the Arf1 GTPase. Cargos carry specific signals in their cytosolic-exposed domain mediating their recruitment by COPI. The best characterized motifs are the di-lysine motifs KKxx or KxKxx that are recognized by the α-COP or β'-COP subunit of the coatomer. Multimeric proteins, such as receptors or channels, contains arginine-based sorting signal (ϕRxR with ϕ representing any hydrophobic amino acid). A common feature of the motif involved in the protein sorting toward the COPI-dependent pathway is the presence of basic residues. Because of its content in basic residues, the KxGxYR motif might be recognized by the COPI machinery to prevent its cell surface expression.

Nevertheless, the mechanism of action of the KxGxYR motif remains to be further elucidated, particularly in the context of the viral infection. Indeed, the retention of the protein may be important for the proper assembly of the viral particle by promoting interaction with the other viral membrane components. How the interaction with E or S may mask the KxGxYR retention signal or how these protein complexes may further traffic through the biosynthetic pathway remain to be clarified.

Furthermore, we cannot exclude that other domains of the protein may also be implicated in the TGN localization of the protein. As mentioned above, the transmembrane domain and an endocytic motif cooperate for the proper localization of TGN38/46. The Golgi apparatus is a compartmentalized structure where glycosylation occurs in an ordered process (41). The successful completion of glycosylation relies in part on the proper distribution of glycosyltransferases along the Golgi that will permit their action in a sequential manner. Most of the glycosyltransferases are type II transmembrane proteins. They consist of a short N-terminal domain exposed in the cytosol, a transmembrane domain, a stem region and an enzymatic domain. The non-uniform distribution of these enzymes in the Golgi is likely maintained by a combination of retention and recycling mechanisms (42). The mechanisms that ensure the localization of glycosyltransferases in the Golgi are numerous and diverse. These include the oligomerization status of the enzyme and complex formation, the length of the transmembrane domain but also its composition that may affect the way it is interacting with the different lipid

compositions that the protein encounters in the different Golgi cisternae. The cytosolic domains also contain motifs involved in the localization of the enzymes (42).

To confirm the role of the KxGxYR motif as a retention signal in the TGN we attempted to transfer the motif in CD4, a protein expressed at the plasma membrane. Unfortunately, the different chimeras that we constructed were not folded properly. This is likely due to the difference of protein structures, CD4 being a type I transmembrane protein, whereas MERS-CoV M is a triple membrane-spanning protein. Therefore, to further confirm the role of the KxGxYR motif, we constructed chimeras between MERS-CoV M and IBV-M as they show clear differences in subcellular localization at steady-state. Indeed, IBV-M is mainly expressed in the ERGIC and cis-Golgi compartment. Previous studies reported the role of the first transmembrane segment in the intracellular retention of the protein (26). This was shown by the replacement of the transmembrane domain of the glycoprotein G of the vesicular stomatitis virus (VSVG) with the first transmembrane segment of IBV-M which resulted in the intracellular retention of the protein, and four polar residues were shown to mediate this retention (25). Surprisingly, when we replaced the first membrane-spanning domain of MERS-CoV M with the one of IBV-M in the context where the KxGxYR motif is mutated into alanine residues (TM1-IBV/MERS-M-KGYR), the protein was not located in the ERGIC but was still transported to the cell surface. This difference may lie in the use of a full-length coronavirus M protein instead of a reporter protein such as VSVG. Interestingly, when we swapped the C-terminal domains of the proteins, we also switched their specific localizations, suggesting that the IBV-M C-terminal domain contains signal(s) for the ERGIC localization of the protein, and not the first membrane segment as previously reported.

Altogether, our results suggest that the C-terminal domain of coronavirus M proteins dictates their specific localization, the Betacoronavirus M proteins being addressed to the TGN where they are retained by the action of the KxGxYR motif. The motif mediating IBV-M localization in the ERGIC and cis-Golgi compartment remains however to be determined. At this state of knowledge, we cannot exclude that the membrane-spanning segments of the protein additionally participate in the intracellular retention of these proteins and cooperates with the C-terminal domain to prevent the expression of the protein at the cell surface.

Experimental procedures

Plasmids.

The coding sequence of the M protein was cloned in the pCDNA3.1(+) vector, with or without a sequence coding for different tags, including HA, VSVG and V5. Total RNA from blood samples of infected patient were extracted by using the Nucleospin RNA kit (Macherey-Nagel) according to the manufacturer's instructions. Then, reverse transcription was performed using the high capacity cDNA reverse transcription kit (Applied Biosystems) and the M protein sequence was amplified by two successive PCRs. First, the sequence was amplified by using the two following primers: 5'-gacgagtgggtttaacgaact-3' and 5'-ggggatgccataacaatgaaa-3'. Then, to insert the sequence in expression vectors, the sequence was amplified with 5'-tcggatccaccatgtctaatatgacgcaactcactg-3' (primer A) and 5'cagaattcctaagctcgaagcaatgcaa-3' (primer B ; untagged protein) or by combination of primer A and 5'-tagaattcagctcgaagcaatgcaagttcaat-3' (primer C ; C-terminal tagged protein) or with 5'-acggatccaatatgacgcaactcactgagg-3' (primer D) with primer B (N-terminal tagged protein). PCR products were inserted between the BamH1 and EcoR1 restriction sites of the different vectors.

M protein deletion mutants were generated by PCR by using either primer A or D in combination with a reverse primer annealing at different positions of the M sequence, with

or without a stop codon with an EcoR1 restriction site. M protein point mutants were generated by site-directed mutagenesis using PCR. Overlapping primers containing the mutation(s) of interest were designed and used for PCR on a MERS-M wild type template. PCRs were then gel-purified, digested and cloned into a pCDNA3.1-VSVG or pCDNA3.1-HA plasmid. The chimeric constructions were generated by fusion PCR. For IBV-M/MERS-M-Ct, the sequence of IBV-M corresponding to residues 1 to 101 were amplified with the forward primer 5'-ttaagctttccatgcccaacgagacaaattg-3' and the reverse primer 5'-taaacagccgaatactctggatccaataac-3', containing 10 bases complementary to the MERS-M sequence at its 5' extremity. The MERS-M sequence between the residues 100 to 219 were amplified by PCR using the forward primer 5'-ccagagtATTcggctgtttatgagaactgg-3' containing 10 bases complementary to the IBV-M sequence, and with the primer D. Then the two PCR products were mixed and amplified using the forward primer that anneals the IBV-M sequence and the primer D. Using the same strategy of overlapping sequences for the internal primers, we constructed MERS-M/IBV-M-Ct composed of the MERS-M_{1-102} fused to IBV-$M_{105-225}$. The first transmembrane segment of the MERS-M-KGYR was replaced by the first transmembrane segment of IBV-M by fusing IBV-M_{1-42} to MERS-M-KGYR$_{41-219}$ (TM1-IBV/MERS-M-KGYR) and the first transmembrane segment of IBV-M was replaced by the first transmembrane segment of MERS-M by fusing MERS-M_{1-40} to IBV-M_{43-225} (TM1-MERS/IBV-M). The PCR products were gel-purified and digested by BamH1 and EcoR1, and then inserted into the pCDNA3.1-V5 expression vector. All the constructs were verified by DNA sequencing.

The plasmid coding the ERGIC-53 protein fused to the GFP and coding the GFP-CI-MPR were kindly provided by Dr Hauri (University of Basel, Switzerland) and Dr Hoflack (University of Dresden, Germany) respectively. The plasmid encoding the CD4-GFP fusion protein was constructed by amplification of the GFP with the two following primers 5'-AGACATGTAGCCCCATTGTGAGCAAG GGCGAGGAGCT-3' and 5'-GGGTCGACTCACTTGTACAGCTCGTC CATGC-3' and then inserted into the PCI-CD4 between the AflIII and SalI restriction sites.

Cell culture and transfection.

HeLa cells were maintained in MEM (*Minimum Essential Medium*) supplemented with 10% fetal calf serum (FCS) and 1% Glutamax. 24h before transfection, HeLa cells were plated in 24-well plates on coverslips or in 6 well plates. The next day, plasmids encoding wild-type M protein or M mutant protein were transfected into HeLa cells using *TransIT®-LT1* Transfection Reagent (Mirus Bio).

Immunofluorescence and confocal microscopy.

At 18h post-transfection, cells were rinsed with PBS, fixed with 3% PFA and processed for immunofluorescence analysis. Cells were permeabilized with 0,1% Triton X100 in PBS for 5 min and then blocked containing 10% goat or horse serum in PBS for 10 min. M protein was detected using anti-M pAbs (rabbit, Proteogenix) or anti-tag antibodies: anti-HA mAbs (3F10, Sigma Aldrich), anti-VSVG mAbs (P5D4, produced in the lab of the authors) or anti-V5 mAbs (ThermoFisher Scientific). For co-localization experiments, cells were double-labeled for M-proteins and cellular marker, anti-calreticulin pAbs (CRT) for endoplasmic reticulum (ER) and anti-TGN46 pAbs for TGN (Biorad). Primary antibodies were diluted in blocking buffer. In some cases, intracellular compartments were stained by transfecting an expression vector for a marker fused with the green fluorescent protein (GFP). For ERGIC and TGN

compartments, cells were co-transfected with M proteins and expression vectors for ERGIC53 and M6PR fused to GFP, respectively. For cell surface staining, cells were transfected with a vector expressing CD4 fused to GFP. After a 30 min incubation with primary antibodies, cells were washed 3 times for 5 min with PBS. Then the cells were incubated with fluorescent secondary antibodies (cyanine-3 conjugated goat anti-mouse IgG; cyanine-3 conjugated goat anti-rabbit IgG, alexa488 conjugated donkey anti-rat IgG; cyanine-3 conjugated donkey anti-sheep IgG; alexa488 conjugated donkey anti-mouse IgG; alexa555 conjugated goat anti-rat IgG) and 1 μg/ml of 4',6-diamidino-2-phenylindole (DAPI).

Images were acquired using a laser scanning confocal microscope LSM 880 (Zeiss) using a 63x oil immersion objective. Signals were sequentially collected using single fluorescence excitation and acquisition settings to avoid crossover.

The extent of colocalization was quantified by calculating the Pearson's correlation coefficient (PCC) using the JACoP plugin of *ImageJ*. The PCC examines the relationship between the intensities of the pixels of two channels in the same image. For each calculation, at least 15 images were analyzed to obtain a PCC mean. A PCC of 1 indicates perfect correlation, 0 no correlation, and -1 perfect anti-correlation.

Biotinylation and internalization assay.

HeLa cells were seeded in 6-well plates and transfected the next day with pCDNA3.1-V5-N3Q-M, pCDNA3.1-V5-N3Q-M Δ 20, pCDNA3.1-V5-N3Q-M-D211A,E213A, pCDNA3.1-V5-N3Q-M-K199A,G201A,Y203A,R204A, PCDNA3.1-V5-N3Q-M-K199A, pCDNA3.1-V5-N3Q-M-G201A, pCDNA3.1-V5-N3Q-M-Y203A or pCDNA3.1-V5-N3Q-M-R204A. At 24h post-transfection cells were rinsed on ice with ice-cold PBS, and incubated twice with 250 μg/mL of EZ-Link™ Sulfo-NHS-SS-Biotin (Pierce) diluted in PBS for 15 minutes in order to label cell surface proteins. Unfixed biotin was then quenched by two sequential incubations of the cells for 10 minutes with 50 mM Glycine/PBS.

For internalization assays, cells were biotinylated 48h post-transfection and then incubated at 37°C for 30 min. The biotin of non-endocytosed proteins was then cleaved upon three incubations of 20 min with glutathione buffer (50mM reduced glutathione, 75 mM NaCl, 75 mM NaOH, 10% FCS) followed by two incubations of 15 min with iodoacetamide buffer (50mM iodoacetamide, 1% BSA, PBS).

Cells were then lysed with B1 buffer (50 mM Tris pH 7.5, 100 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, protease inhibitors cocktail) on ice. Lysates were centrifuged at 14,000 rpm at 4°C for 5 min to remove cellular debris, and were then incubated with 30 μL of streptavidin-conjugated agarose beads (Sigma) for 2h. Beads were then washed serially with 1mL of buffers B1, B2 (50 mM Tris pH 7.5, 100 mM NaCl, 2 mM EDTA, 0.1% Triton X-100, 0.5% SDS, 0.5% DOC), B3 (50 mM Tris pH 7.5, 500 mM NaCl, 2 mM EDTA, 0.1% Triton X-100) and B4 (50 mM Tris pH 7.5, 100 mM NaCl, 2 mM EDTA). Proteins were resuspended in Laemmli loading buffer and detected by immunoblotting. Samples were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and proteins were transferred on a nitrocellulose membrane (Amersham). Membrane-bound M proteins were then detected using a monoclonal anti-V5 antibody and horseradish peroxidase-conjugated secondary antibody. Detection was carried out by chemiluminescence (Pierce).

Glycosidases treatment.

HeLa cells were transfected with vectors expressing V5-M, V5-M-DxE, V5-M-KGYR or V5-N3Q-M proteins. 24h later, cells were lysed in B1 buffer. Then, 30 μl of lysates were mock-treated or treated with PNGase F or Endoglycosidase H according to the manufacturers instructions. Then proteins were separated on SDS-PAGE and detected by immunoblotting.

M-M interactions assay.
HeLa cells were seeded in 10 cm dishes and transfected with vectors expressing V5-N3Q-M or V5-N3Q-M-KGYR. The next day, cell surface proteins were biotinylated at 4 °C and cross-linked with 0.8% PFA in PBS for 10 minutes. Then, PFA was quenched by washing the cells with 50 mM NH_4CL/PBS twice. Cells were lyzed with B1 buffer and lysates were processed for streptavidin precipitation as previously described. Proteins were resuspended in non-reducing Laemmli loading buffer without heating and detected by immunoblotting.

Acknowledgments
We thank Hans-Peter Hauri, Bernard Hoflack and Gary Whittaker for providing reagents. The immunofluorescence analyses were performed with the help of the imaging core facility of the BioImaging Center Lille Nord-de-France.
This work was supported by a grant Visionn-AIRR from Region Hauts-de-France. A Perrier was supported by a fellowship from the University of Lille and the Region Hauts-de-France and L Desmarets is supported by a fellowship from the Region Hauts-de-France

Conflict of interest
The authors declare that they have no conflicts of interest with the contents of this article

Author contributions
Conceived and designed the experiments: SB, JD,YR, AG
Performed the experiments: AP, AB, AD, LD
Analyzed the data: SB, LD, YR, JD
Wrote the paper: AP, SB, JD

REFERENCES

1. Masters, P. S. (2006) The molecular biology of coronaviruses. *Adv. Virus Res.* **66**, 193–292
2. Perlman, S., and Netland, J. (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.* **7**, 439–450
3. Corman, V. M., Muth, D., Niemeyer, D., and Drosten, C. (2018) Hosts and Sources of Endemic Human Coronaviruses. *Adv. Virus Res*, **100,** pp. 163–188
4. Belouzard, S., Millet, J. K., Licitra, B. N., and Whittaker, G. R. (2012) Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses.* **4**, 1011–1033
5. Heald-Sargent, T., and Gallagher, T. (2012) Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses.* **4**, 557–580
6. Ruch, T. R., and Machamer, C. E. (2012) The coronavirus E protein: assembly and beyond. *Viruses.* **4**, 363–382
7. Nieto-Torres, J., Verdiá-Báguena, C., Castaño-Rodriguez, C., Aguilella, V., and Enjuanes, L. (2015) Relevance of Viroporin Ion Channel Activity on Viral Replication and Pathogenesis. *Viruses.* **7**, 3552–3573
8. Westerbeck, J. W., and Machamer, C. E. (2019) The Infectious Bronchitis Coronavirus Envelope Protein Alters Golgi pH to Protect Spike Protein and Promote Release of Infectious Virus. *J. Virol.* 10.1128/JVI.00015-19

9. Westerbeck, J. W., and Machamer, C. E. (2015) A Coronavirus E Protein Is Present in Two Distinct Pools with Different Effects on Assembly and the Secretory Pathway. *J. Virol.* **89**, 9313–9323
10. de Haan, C. A. M., and Rottier, P. J. M. (2005) Molecular interactions in the assembly of coronaviruses. *Adv. Virus Res.* **64**, 165–230
11. de Haan, C. A., Kuo, L., Masters, P. S., Vennema, H., and Rottier, P. J. (1998) Coronavirus particle assembly: primary structure requirements of the membrane protein. *J. Virol.* **72**, 6838–6850
12. Opstelten, D. J., Raamsman, M. J., Wolfs, K., Horzinek, M. C., and Rottier, P. J. (1995) Envelope glycoprotein interactions in coronavirus assembly. *J. Cell Biol.* **131**, 339–349
13. de Haan, C. A., Smeets, M., Vernooij, F., Vennema, H., and Rottier, P. J. (1999) Mapping of the coronavirus membrane protein domains involved in interaction with the spike protein. *J. Virol.* **73**, 7441–7452
14. Arndt, A. L., Larson, B. J., and Hogue, B. G. (2010) A Conserved Domain in the Coronavirus Membrane Protein Tail Is Important for Virus Assembly. *J. Virol.* **84**, 11418–11428
15. Tseng, Y.-T., Wang, S.-M., Huang, K.-J., Lee, A. I.-R., Chiang, C.-C., and Wang, C.-T. (2010) Self-assembly of severe acute respiratory syndrome coronavirus membrane protein. *J. Biol. Chem.* **285**, 12862–12872
16. de Haan, C. A. M., Vennema, H., and Rottier, P. J. M. (2000) Assembly of the Coronavirus Envelope: Homotypic Interactions between the M Proteins. *J. Virol.* **74**, 4967–4978
17. Baudoux, P., Carrat, C., Besnardeau, L., Charley, B., and Laude, H. (1998) Coronavirus Pseudoparticles Formed with Recombinant M and E Proteins Induce Alpha Interferon Synthesis by Leukocytes. *J Virol*. **72**, 8636–8643
18. Vennema, H., Godeke, G. J., Rossen, J. W., Voorhout, W. F., Horzinek, M. C., Opstelten, D. J., and Rottier, P. J. (1996) Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* **15**, 2020–2028
19. Lim, K. P. (2001) The Missing Link in Coronavirus Assembly. Retention of the avian coronavirus infectious brochitis virus envelope protein in the pre-Golgi compartments and physical interaction between the envelope and membra,e proteins. *J. Cell Biol.* **276**, 17515–17523
20. Siu, Y. L., Teoh, K. T., Lo, J., Chan, C. M., Kien, F., Escriou, N., Tsao, S. W., Nicholls, J. M., Altmeyer, R., Peiris, J. S. M., Bruzzone, R., and Nal, B. (2008) The M, E, and N Structural Proteins of the Severe Acute Respiratory Syndrome Coronavirus Are Required for Efficient Assembly, Trafficking, and Release of Virus-Like Particles. *J. Virol.* **82**, 11318–11330
21. Klumperman, J., Locker, J. K., Meijer, A., Horzinek, M. C., Geuze, H. J., and Rottier, P. J. (1994) Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding. *J. Virol.* **68**, 6523–6534
22. Salanueva, I. J., Carrascosa, J. L., and Risco, C. (1999) Structural maturation of the transmissible gastroenteritis coronavirus. *J. Virol.* **73**, 7952–7964
23. Machamer, C. E., Mentone, S. A., Rose, J. K., and Farquhar, M. G. (1990) The E1 glycoprotein of an avian coronavirus is targeted to the cis Golgi complex. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 6944–6948
24. Locker, J. K., Griffiths, G., Horzinek, M. C., and Rottier, P. J. (1992) O-glycosylation of the coronavirus M protein. Differential localization of sialyltransferases in N- and O-linked glycosylation. *J. Biol. Chem.* **267**, 14094–14101
25. Machamer, C. E., Grim, M. G., Esquela, A., Chung, S. W., Rolls, M., Ryan, K., and Swift, A.

M. (1993) Retention of a cis Golgi protein requires polar residues on one face of a predicted alpha-helix in the transmembrane domain. *Mol. Biol. Cell*. **4**, 695–704
26. Machamer, C. E., and Rose, J. K. (1987) A specific transmembrane domain of a coronavirus E1 glycoprotein is required for its retention in the Golgi region. *J. Cell Biol.* **105**, 1205–1214
27. Locker, J. K., Klumperman, J., Oorschot, V., Horzinek, M. C., Geuze, H. J., and Rottier, P. J. (1994) The cytoplasmic tail of mouse hepatitis virus M protein is essential but not sufficient for its retention in the Golgi complex. *J. Biol. Chem.* **269**, 28263–28269
28. Waguri, S., Dewitte, F., Borgne, R. L., Rouille, Y., Uchiyama, Y., Dubremetz, J.-F., and Hoflack, B. (2003) Visualization of TGN to Endosome Trafficking through Fluorescently Labeled MPR and AP-1 in Living Cells□V. *Mol. Biol. Cell*. **14**, 14
29. Dunn, K. W., Kamocka, M. M., and McDonald, J. H. (2011) A practical guide to evaluating colocalization in biological microscopy. *Am. J. Physiol.-Cell Physiol.* **300**, C723–C742
30. Venditti, R., Wilson, C., and De Matteis, M. A. (2014) Exiting the ER: what we know and what we don't. *Trends Cell Biol.* **24**, 9–18
31. Geva, Y., and Schuldiner, M. (2014) The Back and Forth of Cargo Exit from the Endoplasmic Reticulum. *Curr. Biol.* **24**, R130–R136
32. Votsmeier, C. (2001) An acidic sequence of a putative yeast Golgi membrane protein binds COPII and facilitates ER export. *EMBO J.* **20**, 6742–6750
33. Yang, Y., Zhang, L., Geng, H., Deng, Y., Huang, B., Guo, Y., Zhao, Z., and Tan, W. (2013) The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. *Protein Cell*. **4**, 951–961
34. Bos, K., Wraight, C., and Stanley, K. K. (1993) TGN38 is maintained in the trans-Golgi network by a tyrosine-containing motif in the cytoplasmic domain. *EMBO J.* **12**, 2219–2228
35. Ponnambalam, S. (1994) The TGN38 glycoprotein contains two non-overlapping signals that mediate localization to the trans-Golgi network. *J. Cell Biol.* **125**, 253–268
36. Armstrong, J., and Patel, S. (1991) The Golgi sorting domain of coronavirus E1 protein. *J. Cell Sci.* **98 (Pt 4)**, 567–575
37. Nal, B. (2005) Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. *J. Gen. Virol.* **86**, 1423–1434
38. McBride, C. E., and Machamer, C. E. (2010) A Single Tyrosine in the Severe Acute Respiratory Syndrome Coronavirus Membrane Protein Cytoplasmic Tail Is Important for Efficient Interaction with Spike Protein. *J. Virol.* **84**, 1891–1901
39. Locker, J. K., Opstelten, D. J., Ericsson, M., Horzinek, M. C., and Rottier, P. J. (1995) Oligomerization of a trans-Golgi/trans-Golgi network retained protein occurs in the Golgi complex and may be part of its retention. *J. Biol. Chem.* **270**, 8815–8821
40. Heffernan, L. F., and Simpson, J. C. (2014) The trials and tubule-ations of Rab6 involvement in Golgi-to-ER retrograde transport. *Biochem. Soc. Trans.* **42**, 7
41. Puthenveedu, M. A., and Linstedt, A. D. (2005) Subcompartmentalizing the Golgi apparatus. *Curr. Opin. Cell Biol.* **17**, 369–375
42. Tu, L., and Banfield, D. K. (2010) Localization of Golgi-resident glycosyltransferases. *Cell. Mol. Life Sci.* **67**, 29–41

FIGURES

Figure 1: Schematic drawing of the MERS-M protein with the sequence of residues 149-219 of the C-terminal domain (1A). Subcellular localization of the MERS-CoV M protein (1B and 1C). Cells expressing HA-tagged M protein in combination with the GFP-CI-MPR, ERGIC-53-GFP or CD4 fused with GFP were labeled with an anti-HA antibody. GFP-CI-MPR is a TGN marker, ERGIC-53-GFP is an ER-Golgi intermediate compartment marker and CD4 is a protein expressed at the cell surface. To detect the ER compartment or the TGN, cells were double-labeled for HA and calreticulin (CRT) or TGN46 as indicated. Bars indicate 20 μm. Pearson's correlation coefficients were calculated for each combination of co-staining.
N-terminal or C-terminal added-tags have no effect on the TGN localization of the M protein (1D and 1E). Cells expressing untagged M protein (M), N- or C-terminally V5-tagged M protein (V5-M and M-V5 respectively), N-terminally HA-tagged M (HA-M) or C-terminally VSVG-tagged M (M-VSVG) were double-labeled with anti-M antibody together with an anti-TGN46 antibody. Pearson's correlation coefficients were calculated.

Figure 2: The distal part of MERS-M C-terminal domain contains motif(s) involved in its subcellular localization. Cells expressing the M protein (HA-M and M-VSVG), or the M protein lacking its last 20 residues (HA-M Δ 20 or M Δ 20-VSVG) or the M protein with D211 and E213 mutated into A (HA-M-DxE or M-DxE-VSVG) were processed for detection of M protein using an anti-tag antibody. The TGN was detected by using an anti-TGN46 (2A) and the ER by using an anti-CRT antibody (2B). The plasma membrane was labeled by co-expression of CD4 fused to GFP together with the M protein (2C). Bars indicate 20 μm. Pearson's correlation coefficients were calculated for each combination of co-staining (2D).

Figure 3: Subcellular localization of mutants with serial deletions of the distal part of the MERS-CoV M protein. M protein with a C-terminal VSVG tag (M) or M protein deleted of 5 (M Δ 5), 10 (M Δ 10), 15 (M Δ 15) or 20 amino acid residues (M Δ 20) were expressed in HeLa cells and their localization either in the TGN or at the plasma membrane was investigated by double labeling with TGN46 (3A) or by co-expressing CD4 fused to GFP (3B). Bars indicate 20 μm. Pearson's correlation coefficients were calculated for each combination of co-staining (3C).

Figure 4: Identification of the MERS-CoV M Δ 15 motif involved in its TGN localization. Amino acid residues G201, N202, Y203 and R204 located in the last 5 residues of M Δ 15 (located in TGN) were mutated individually into alanine and the subcellular localization of the mutants was analyzed as described in Figure 3. Bars indicate 20 μm.

Figure 5: The KxGxYR motif is involved in MERS-M protein localization in TGN. The following mutations K199A, G201A, Y203A, R204A were introduced individually or together in the context of the full-length M protein with a C-terminal VSVG tag. The subcellular localization of the mutants was analyzed as described in Figure 3. Bars indicate 20 μm.

Figure 6: Cell surface expression of M protein mutants. Plasma membrane proteins of cells expressing the different M protein mutants were labeled with non-permeable biotin. Biotinylated proteins were purified using streptavidin-conjugated agarose beads. Biotinylated M proteins and total M proteins in cell lysates were detected in immunoblot (6A) and quantified (6B). Results are expressed as the percentage of total M protein expressed at the cell surface and are expressed as the mean of five independent experiments. Error bars represent the standard error of the means (SEM). Results were analysed by using an ANOVA

test (* P<0.1; ** P<0.01; *** P<0.001; **** P< 0.0001). Glycosylation of M protein mutants. Lysates of cells expressing the M protein or the M protein with the ER export signal mutated (M-DxE) or the TGN localization motif mutated (M-KGYR) were treated with EndoH or PNGase F. A lysate of cells expressing the M protein with its N-glycosylation site mutated (N3Q) was left untreated. N-terminal tagged proteins were detected by Western Blotting with an anti-V5 antibody (6C).
Endocytosis of M protein and M-KGYR. Cells were transfected with vectors expressing the M and M-K199A-G210A-Y203A-R204A (M-KGYR) proteins. Then cell surface proteins were labeled with non-permeable biotin at 4°C. Endocytosis was allowed by incubating the cells for 30 min at 37°C. Biotin of non-internalized proteins was cleaved with glutathione. Internalized M protein was detected after purification with streptavidin-conjugated agarose beads in immunoblot (6E). In each experiment, each condition was performed in duplicate. For cell surface associated protein, only 25% of the sample was loaded on the gel. For the controls of glutathione cleavage (without any internalization, 0 min) and for the samples internalized (30 min), the totality of the samples were loaded on the gel. Internalized M protein was quantified. The results are expressed as the percentage of cell surface-associated M protein and are expressed as the mean of three independent experiments. Error bars represent the standard error of the means (SEM) (Panels 6E and 6F).

Figure 7: MERS-M and IBV-M sequence alignment (7A) and schematic drawings of the different IBV-M and MERS-M chimeras that were constructed (7B). First, the C-terminal domain of MERS-CoV M was replaced with the one of IBV-M (MERS-M/IBV-M) and the C-terminal domain of IBV-M was replaced by the one of MERS-CoV M with or without the mutation of the KxGxYR motif (IBV-M/MERS-M-KGYR and IBV-M/MERS-M, respectively). The first membrane-spanning segment of MERS-CoV M was replaced with the one of IBV-M in the context of the MERS-M-KGYR (TM1-IBV/MERS-M-KGYR) and the first membrane-spanning segment of IBV-M was replaced by the one of MERS-CoV M (TM1-MERS/IBV-M). All the chimeras were tagged at their C-terminal extremity with a VSVG epitope.

Figure 8: Subcellular localization of IBV-M and MERS-CoV M proteins chimeras. The localization of the different chimeras in the TGN or in the ERGIC compartments was investigated by immunofluorescent double labeling by using an anti-VSVG antibody and an anti-TGN46 (8A) or by expressing the ERGIC-53 marker conjugated with GFP (8B). Bars indicate 20 μm. Pearson's correlation coefficients were calculated for each combination of co-staining (8C).

Figure 1:

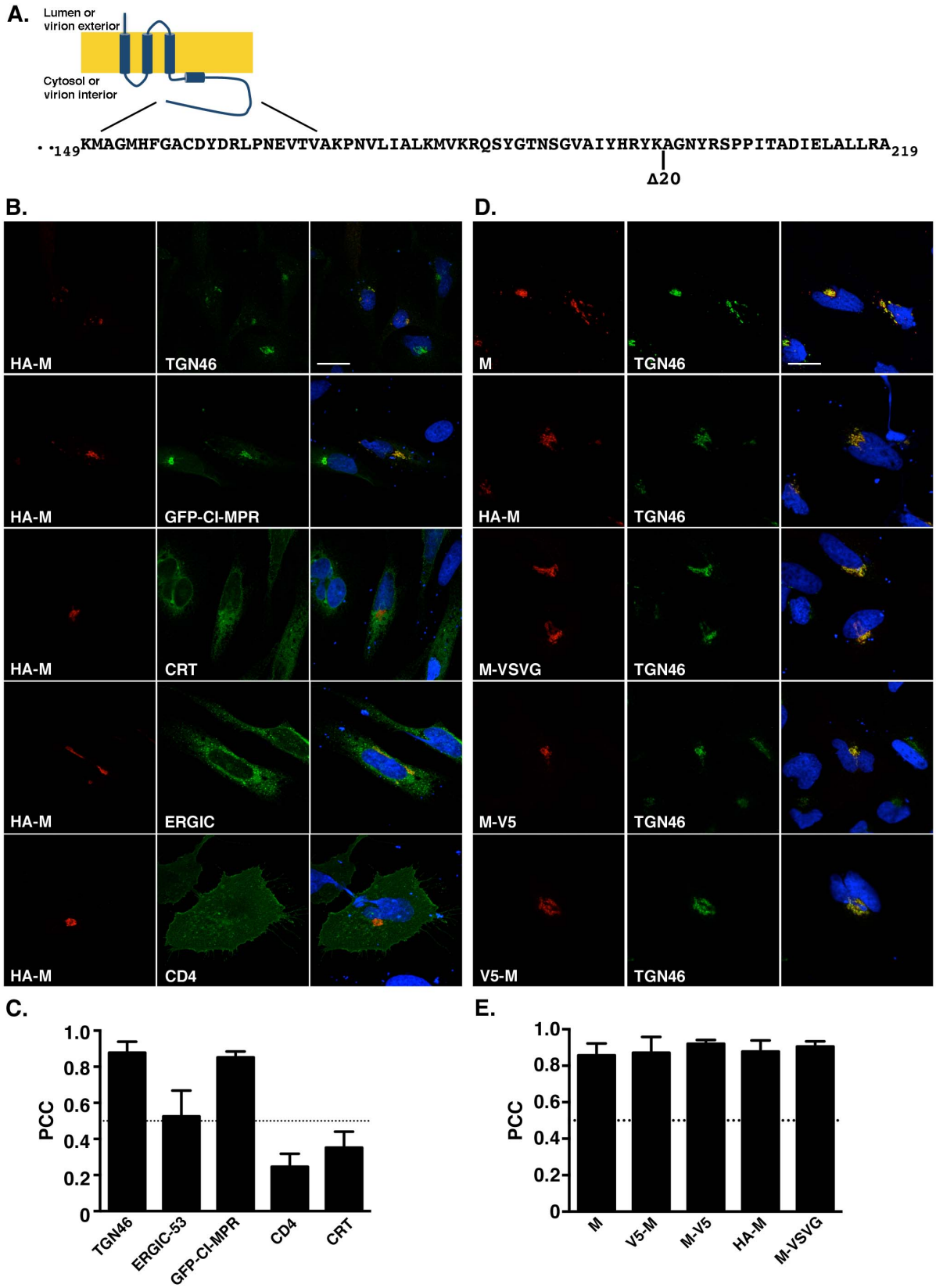
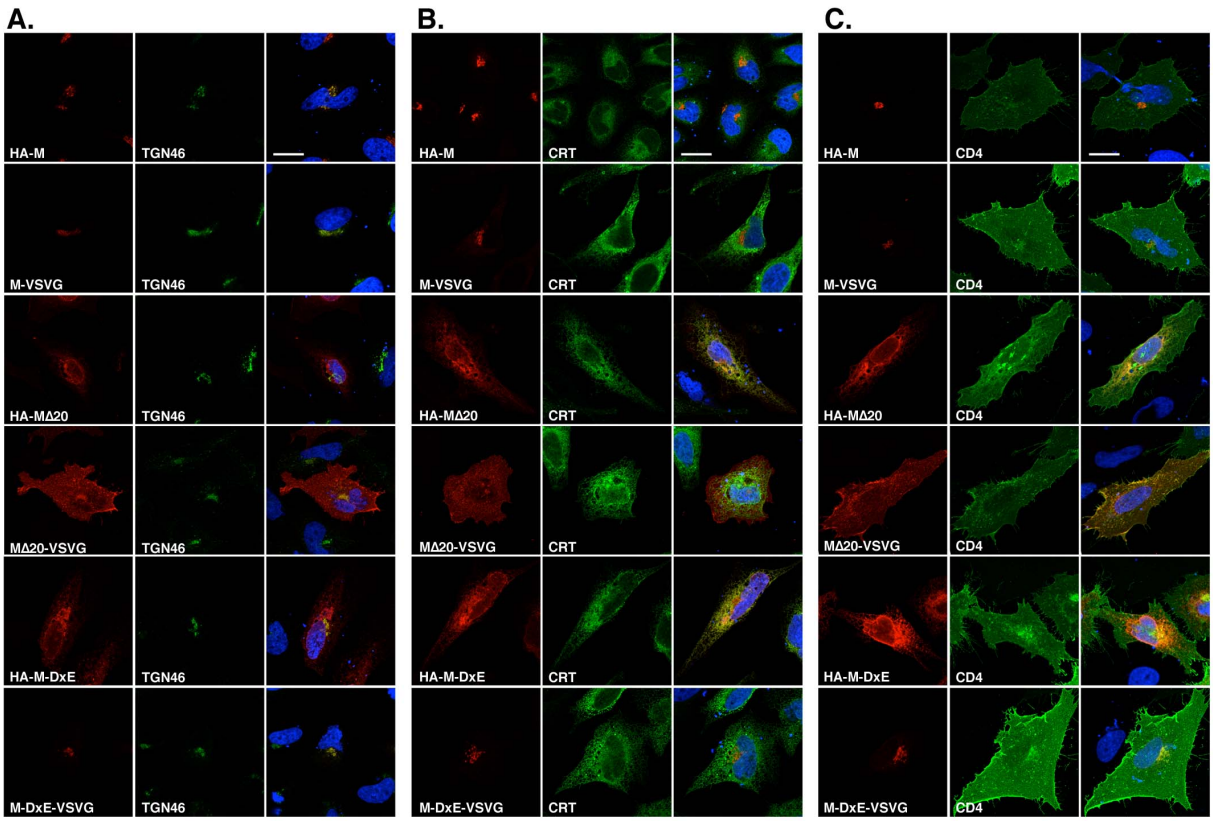

A.
Lumen or virion exterior
Cytosol or virion interior
149KMAGMHFGACDYDRLPNEVTVAKPNVLIALKMVKRQSYGTNSGVAIYHRYKAGNYRSPPITADIELALLRA219
Δ20
B.
HA-M
TGN46
HA-M
GFP-CI-MPR
HA-M
CRT
HA-M
ERGIC
HA-M
CD4
D.
M
TGN46
HA-M
TGN46
M-VSVG
TGN46
M-V5
TGN46
V5-M
TGN46
C.
PCC
1.0
0.8
0.6
0.4
0.2
0
TGN46
ERGIC-53
GFP-CI-MPR
CD4
CRT
E.
PCC
1.0
0.8
0.6
0.4
0.2
0
M
V5-M
M-V5
HA-M
M-VSVG

Figure 2:


A.
B.
C.
HA-M
TGN46
CRT
CD4
M-VSVG
HA-MΔ20
MΔ20-VSVG
HA-M-DxE
M-DxE-VSVG

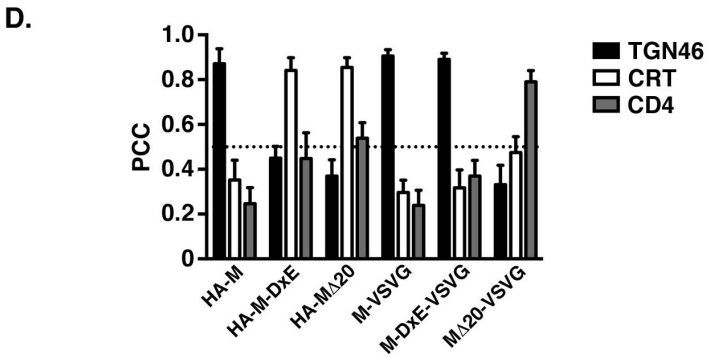

D.
PCC
1.0
0.8
0.6
0.4
0.2
0
TGN46
CRT
CD4
HA-M
HA-M-DxE
HA-MΔ20
M-VSVG
M-DxE-VSVG
MΔ20-VSVG

Figure 3:

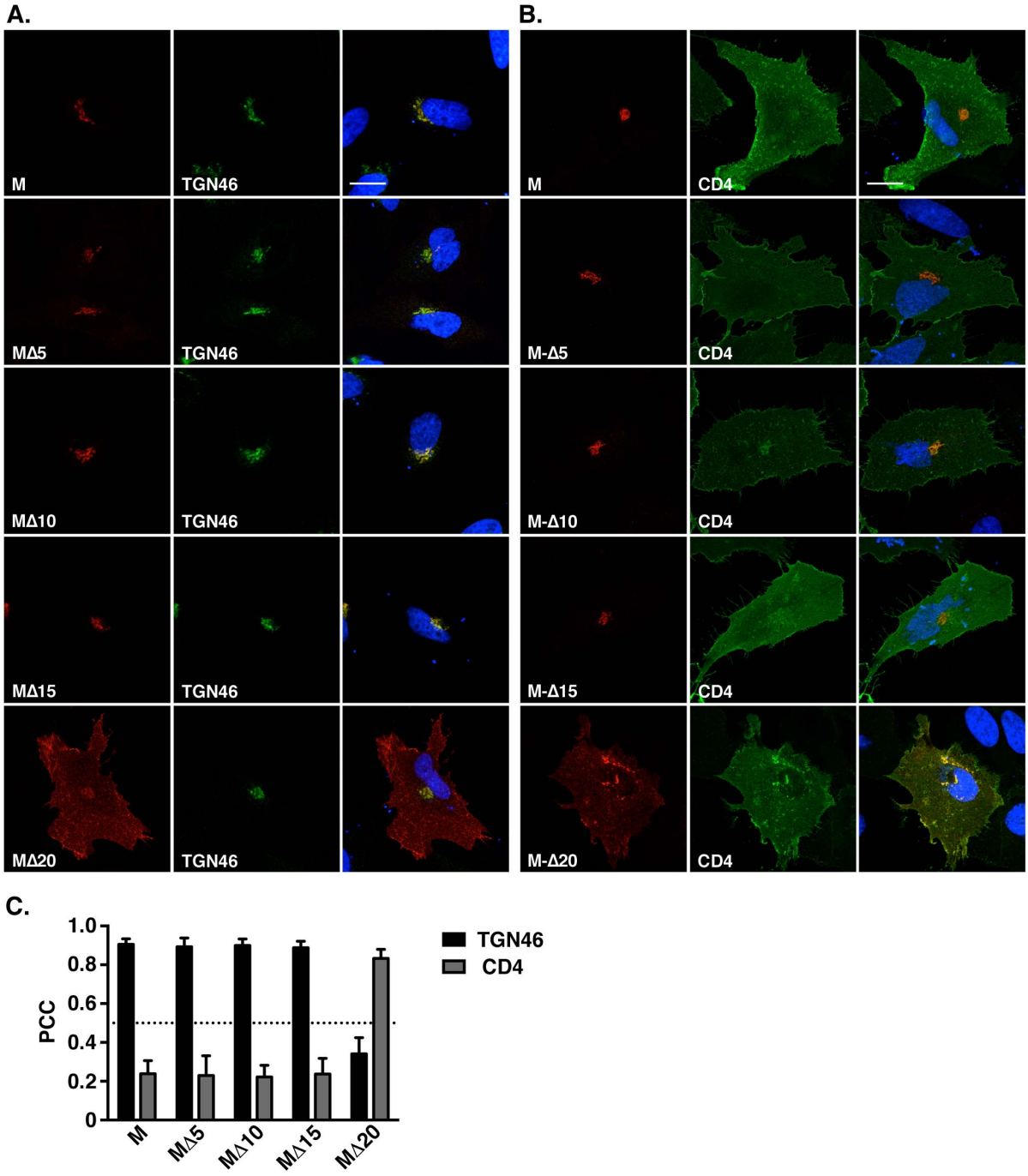

A.
B.
M
TGN46
MΔ5
MΔ10
MΔ15
MΔ20
M-Δ5
M-Δ10
M-Δ15
M-Δ20
CD4
C.
PCC
1.0
0.8
0.6
0.4
0.2
0
TGN46
CD4
M
MΔ5
MΔ10
MΔ15
MΔ20

Figure 4:

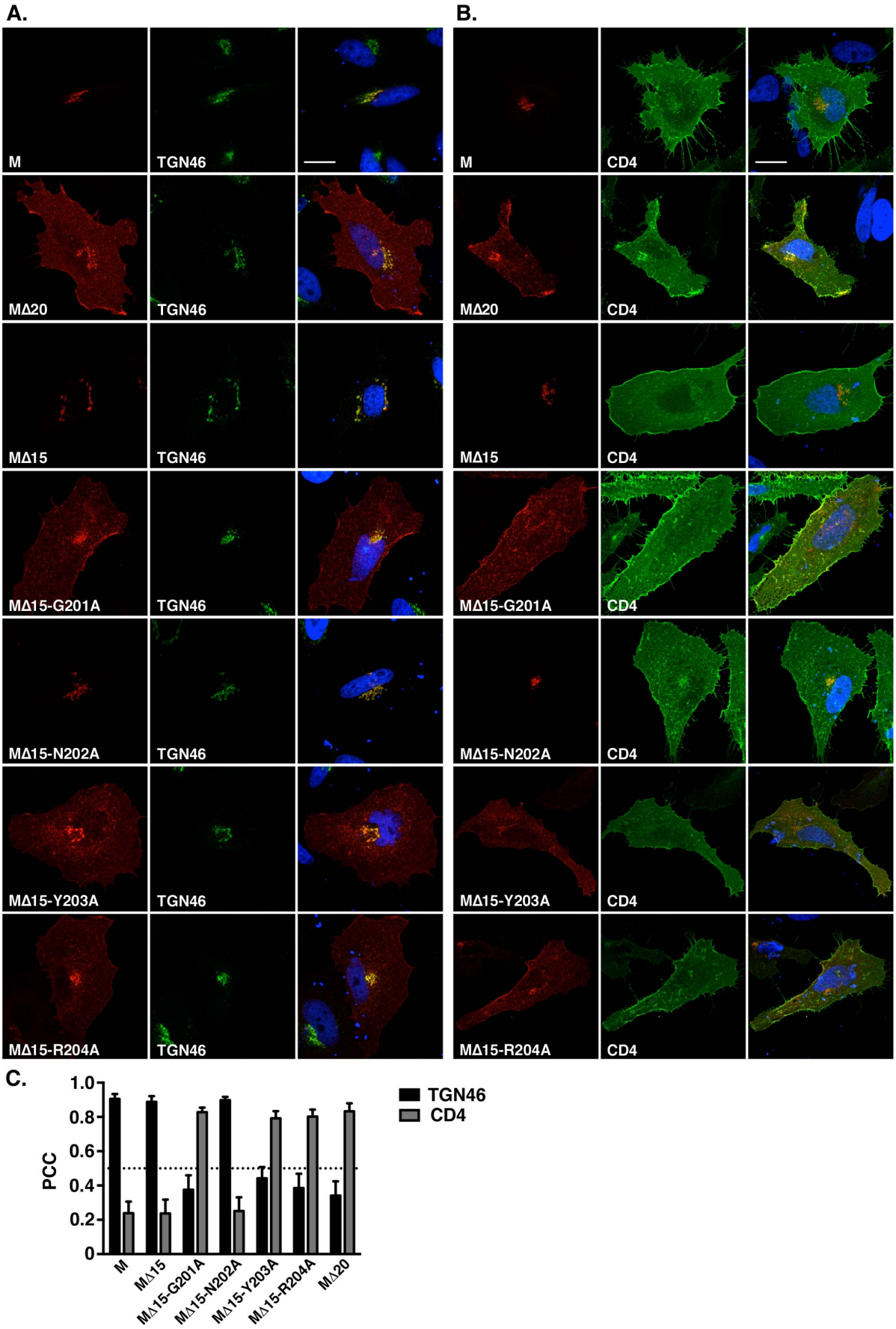

A.
B.
M
TGN46
CD4
MΔ20
MΔ15
MΔ15-G201A
MΔ15-N202A
MΔ15-Y203A
MΔ15-R204A
C.
PCC
1.0
0.8
0.6
0.4
0.2
0
TGN46
CD4
M
MΔ15
MΔ15-G201A
MΔ15-N202A
MΔ15-Y203A
MΔ15-R204A
MΔ20

Figure 5:

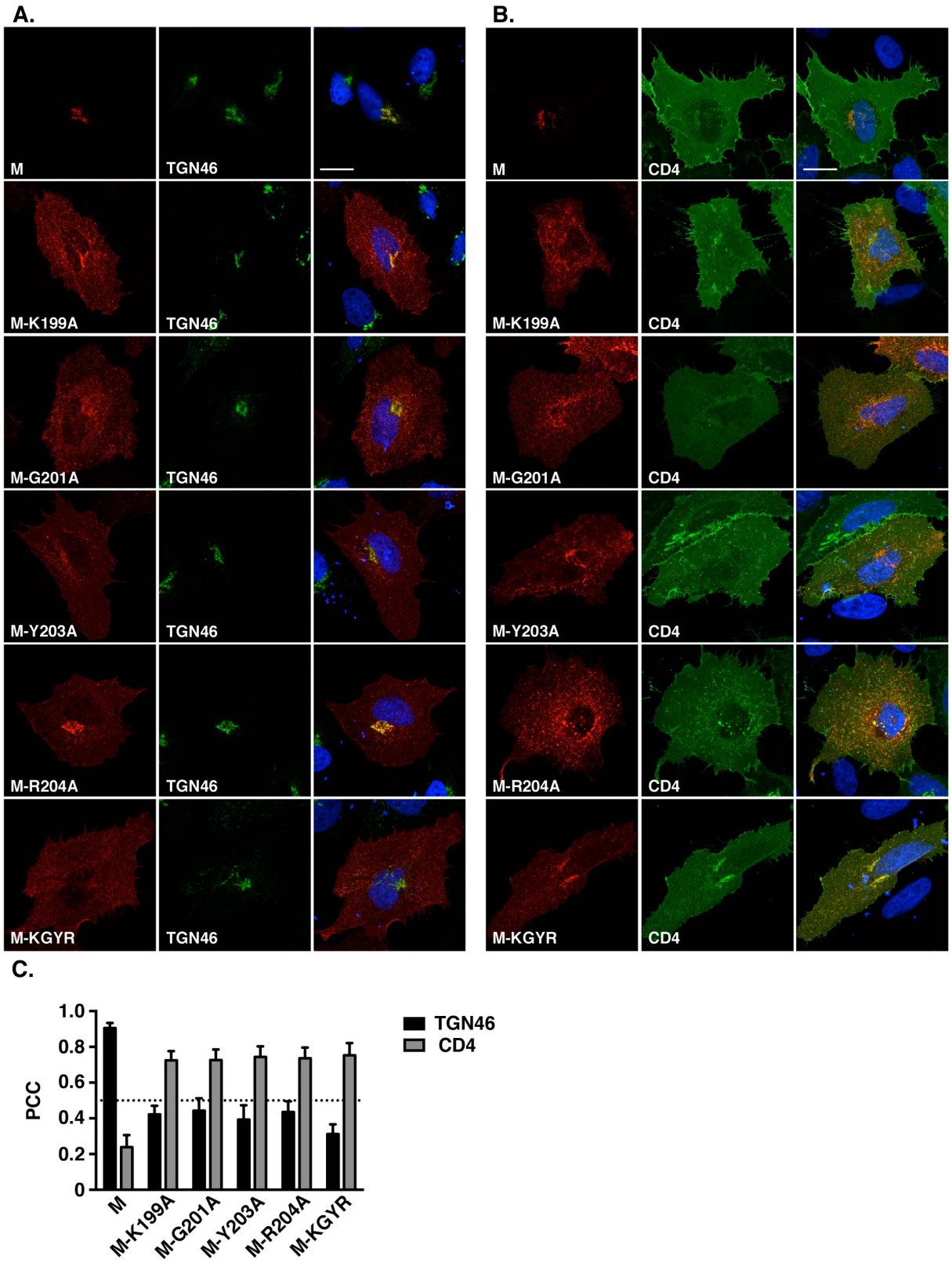

A.
B.
M
TGN46
CD4
M-K199A
M-G201A
M-Y203A
M-R204A
M-KGYR
C.
PCC
1.0
0.8
0.6
0.4
0.2
0
TGN46
CD4
M
M-K199A
M-G201A
M-Y203A
M-R204A
M-KGYR

Figure 6:

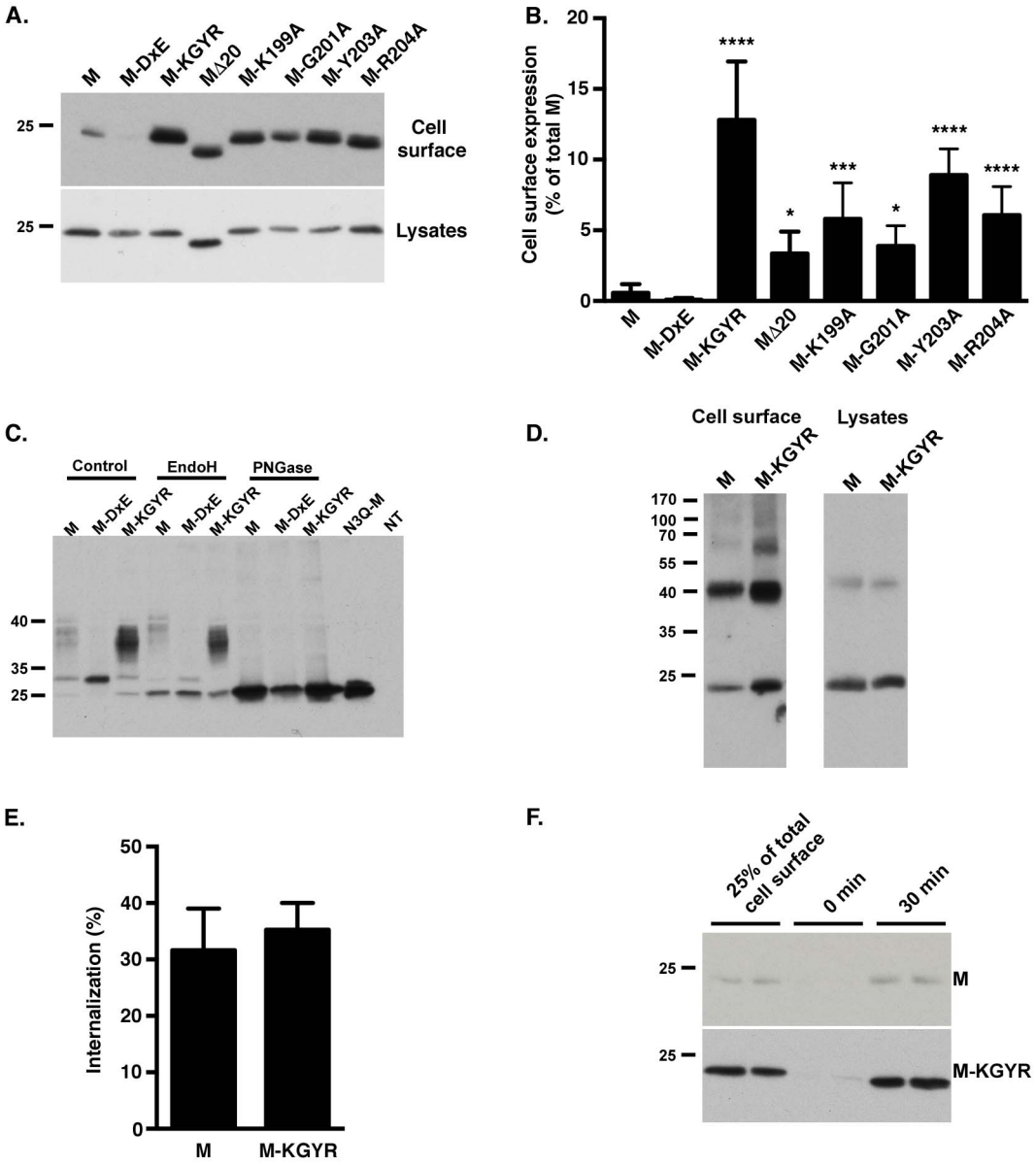

A.
M
M-DxE
M-KGYR
MΔ20
M-K199A
M-G201A
M-Y203A
M-R204A
25
Cell surface
25
Lysates
B.
Cell surface expression (% of total M)
20
15
10
5
0
****
*
***
*
****
****
M
M-DxE
M-KGYR
MΔ20
M-K199A
M-G201A
M-Y203A
M-R204A
C.
Control
EndoH
PNGase
M
M-DxE
M-KGYR
M
M-DxE
M-KGYR
M
M-DxE
M-KGYR
N3Q-M
NT
40
35
25
D.
Cell surface
Lysates
M
M-KGYR
M
M-KGYR
170
100
70
55
40
35
25
E.
Internalization (%)
50
40
30
20
10
0
M
M-KGYR
F.
25% of total cell surface
0 min
30 min
25
M
25
M-KGYR

Figure 7:

A.

```
MERS-M      MSN- MTQLTEAQIIAIIKDWNFAWSLIFLLITIVLQYGYPSRSMTVYVFKMFVLWLLWP
IBV-M       MPNETNCTLDFEQSVQLFKEYNLFITAFLLFLTIILQYGYATRSKVIYTLKMIVLWCFWP
            * *     *   * : ::*::*:  : ::*::**:***** :** .:*.:**:*** :**

MERS-M      SSMALSIFSAVYPIDLASQIISGIVAAVSAMMWISYFVQSIRLFMRTGSWWSFNPETNCL
IBV-M       LNIAVGVISCTYPPNTGGLVAAIILTVFACLSFVGYWIQSIRLFKRCRSWWSFNPESNAV
             .:*:.::*..** : .. : : *::..:.: ::.*::****** *  ********:*.:

MERS-M      LNVPFGGTTVVRP-LVEDSTSVTAVVTNGHLKMAGMHFGACDYDRLPNEVTVAKPNVLIA
IBV-M       GSILLTNGQQCNFAIESVPMVLSPIIKNGVLYCEGQWLAKCEPDHLPKDIFVCTPDRRNI
             .: : .    .  : .    :: ::.** *   *  :. *: *:**::: *..*:

MERS-M      LKMVKRQSYGTNSGVAIYHRYKAGNYRSPPIT-ADIELALLRA-----  219
IBV-M       YRMVQKYTGDQSGNKKRFATF---VYAKQSVDTGELESVATGGSSLYT  225
             :**:: : . ...   :  :    * .  :  .::* .   .
```

B.

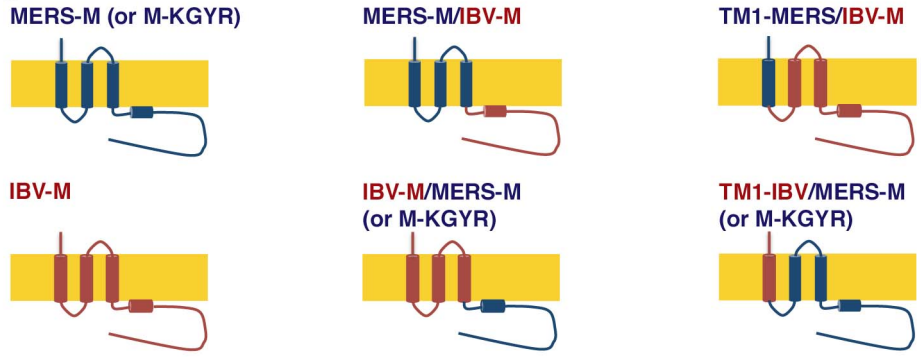
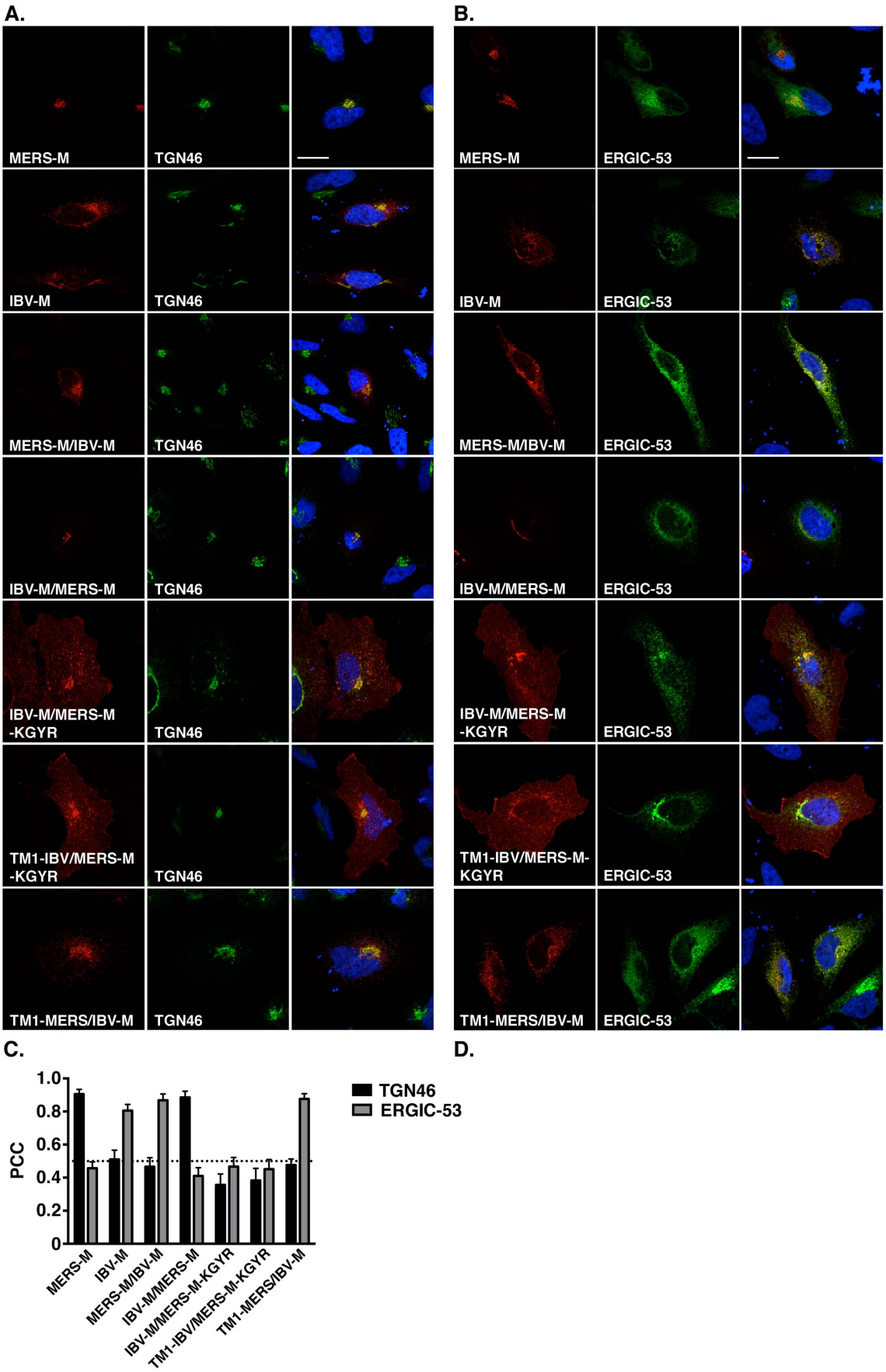


Figure 8:


A.
B.
MERS-M
TGN46
ERGIC-53
IBV-M
MERS-M/IBV-M
IBV-M/MERS-M
IBV-M/MERS-M -KGYR
TM1-IBV/MERS-M -KGYR
TM1-MERS/IBV-M
C.
D.
PCC
1.0
0.8
0.6
0.4
0.2
0
TGN46
ERGIC-53
MERS-M
IBV-M
MERS-M/IBV-M
IBV-M/MERS-M
IBV-M/MERS-M-KGYR
TM1-IBV/MERS-M-KGYR
TM1-MERS/IBV-M