Computational and Structural Biotechnology Journal 21 (2023) 3424–3436

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

New insights into the pathogenicity of TMEM165 variants using structural modeling based on AlphaFold 2 predictions

Dominique Legrand [a], Mélissandre Herbaut [a], Zoé Durin [a], Guillaume Brysbaert [a], Muriel Bardor [a,b], Marc F. Lensink [a], François Foulquier [a,*]

[a] Univ. Lille, CNRS, UMR 8576 – UGSF - Unité de Glycobiologie Structurale et Fonctionnelle, F-59000 Lille, France
[b] Université de Rouen Normandie, Laboratoire GlycoMEV UR 4358, SFR Normandie Végétal FED 4277, Innovation Chimie Carnot, F-76000 Rouen, France

ARTICLE INFO

Article history:
Received 14 April 2023
Received in revised form 15 June 2023
Accepted 15 June 2023
Available online 17 June 2023

Keywords:
TMEM165
CDG
Glycosylation
Manganese
Modeling

ABSTRACT

TMEM165 is a Golgi protein playing a crucial role in Mn^{2+} transport, and whose mutations in patients are known to cause Congenital Disorders of Glycosylation. Some of those mutations affect the highly-conserved consensus motifs E-φ-G-D-[KR]-[TS] characterizing the CaCA2/UPF0016 family, presumably important for the transport of Mn^{2+} which is essential for the function of many Golgi glycosylation enzymes. Others, like the G > R^{304} mutation, are far away from these motifs in the sequence. Until recently, the classical membrane protein topology prediction methods were unable to provide a clear picture of the organization of TMEM165 inside the cell membrane, or to explain in a convincing manner the impact of patient and experimentally-generated mutations on the transporter function of TMEM165. In this study, AlphaFold 2 was used to build a TMEM165 model that was then refined by molecular dynamics simulation with membrane lipids and water. This model provides a realistic picture of the 3D protein scaffold formed from a two-fold repeat of three transmembrane helices/domains where the consensus motifs face each other to form a putative acidic cation-binding site at the cytosolic side of the protein. It sheds new light on the impact of mutations on the transporter function of TMEM165, found in patients and studied experimentally *in vitro,* formerly and within this study. More particularly and very interestingly, this model explains the impact of the G > R^{304} mutation on TMEM165's function. These findings provide great confidence in the predicted TMEM165 model whose structural features are discussed in the study and compared to other structural and functional TMEM165 homologs from the CaCA2/UPF0016 family and the LysE superfamily.

1. Introduction

In 2012, mutations in *TMEM165*, the gene coding for the transmembrane protein 165 (TMEM165), were detected in Congenital Disorders of Glycosylation (CDG) patients presenting a peculiar clinical phenotype, including major skeletal dysplasia, osteoporosis and dwarfism, and a strong defect in the Golgi glycosylation characterized by hypogalactosylation and hyposialylation of total serum N-glycoproteins [1]. Moreover, it was reported that those glycosylation defects also affect O-glycosylated proteins, glycolipids and proteoglycans [2–5]. We showed that whereas some patient mutations affected either the expression or the localization of TMEM165 [6], others clearly impaired the function of TMEM165 in the Golgi, leading to glycosylation defects [7]. Interestingly, our results demonstrated a link between TMEM165 and Mn^{2+} homeostasis, then suggesting that TMEM165 may be a key importer of Mn^{2+} into the Golgi where this cation is a mandatory cofactor for many glycosyltransferases [8]. This hypothesis is highly reinforced by the fact that TMEM165 belongs, according to the IUBMB Transporter Classification Database (TCDB) (https://www.tcdb.org), to the Ca^{2+}:H^{+} Antiporter (CaCA2) family (formerly called the uncharacterized Protein Family 0016 (UPF0016)). Within this family, several homologs from bacteria, plants and eukaryotes were also shown to be involved in Mn^{2+} homeostasis, and possibly also for some of them,

Abbreviations: TMEM165, transmembrane protein 165; CDG, Congenital Disorders of Glycosylation; TMD, Transmembrane helix/domain; LC, Loop in Cytosol; LL, Loop in Lumen; AF2, AlphaFold 2; aa, amino acid

* Corresponding author.
E-mail address: francois.foulquier@univ-lille.fr (F. Foulquier).

https://doi.org/10.1016/j.csbj.2023.06.015

including human TMEM165 and *Saccharomyces cerevisiae* Gdt1p, in Ca^{2+} and H^+ homeostasis [9,10]. Additional pieces of evidence are the sensitivity of TMEM165 to increased cytosolic Mn^{2+} concentrations [8,11] and the measurement of Mn^{2+} transport activity of a truncated form of TMEM165 in *Lactococcus lactis*, as well as that of Gdt1p [10]. In the last years, several studies aimed to characterize the contribution of some specific amino acid (aa) residues to Mn^{2+} sensitivity and/or Mn^{2+} transporter function of TMEM165 allowing normal Golgi glycosylation [7,10], primarily focusing on aa residues of the two consensus motifs E-φ-G-D-[KR]-[TS] highly-conserved among CaCA2 members (^{108}ELGDKT^{113} and ^{248}EWGDRS^{253} in human TMEM165), and on aa residues found mutated in TMEM165-CDG patients. The results confirmed the involvement of specific aa residues within the two motifs, in particular acidic aa residues E^{108}, D^{111}, E^{248} and D^{251}, without however clearly explaining the respective participation of these in the transporter function of TMEM165. Indeed, the classical membrane protein topology prediction methods used till now predicted 6–7 transmembrane helices (TMD) with each consensus motif located at the extremity of TMD 1 and 4 facing both cytosolic and luminal sides of the cell membrane [12–14].

Prediction of protein 3D structure only based on the aa sequence, especially when no homologous structure is available as for TMEM165, has opened a wide field of research since more than 50 years. None of them, however, achieved an atomic resolution. A redesigned version of the neural network-based AlphaFold model, named AlphaFold 2 (AF2), has been validated in the 14th Critical Assessment Techniques for Protein Structure Prediction challenge (CASP14) (https://www.predictioncenter.org/casp14/doc/CASP14_Abstracts.pdf) [15–17]. Using artificial intelligence and machine learning approaches, AF2 was found to accurately predict protein structures even when no homologous structures are available. It has thus been shown to significantly surpass those of previously-described computational methods [15–17]. In the present paper, AF2 has been used to build a TMEM165 model that was then subjected to molecular dynamics (MD) simulation including membrane lipids and water. The model shows that TMEM165 is organized into two repeats of three transmembrane helices/domains (TMD) where the consensus motifs are joined to form a putative acidic cation-binding site. This model is supported by experimental results showing that mutations introducing positive charges at positions 140 and 304 cause a glycosylation defect in cells due to direct electrostatic interactions between the positive charges and several key acidic aa residues of the putative cation-binding site. Very interestingly, these results explain the impact of the previously-unexplained $G > R^{304}$ patient mutation [1,6] on TMEM165 transporter function. The predicted model also gives new insights on the impact of other mutations found in patients and studied experimentally *in vitro* on TMEM165's function. Taken as a whole, our findings provide great confidence in the predicted TMEM165 model whose structural features are discussed and compared to homologs of TMEM165.

2. Material and methods

2.1. Predictive tools

Models of full-size and Δ80-TMEM165, and all other models of TMEM165 homologs, were produced with a locally-installed version of the AlphaFold v2.2 software [15–17]. Five models were generated with pLDDT (predicted local distance difference test) > 70, expected to be modeled with good accuracy [15]. All models shown in the figures were rendered with the UCSF ChimeraX v1.2 software (https://www.cgl.ucsf.edu/chimerax/).

The Orientations of Proteins in Membranes (OPM) database (https://opm.phar.umich.edu/) has been used to calculate the spatial arrangement of the AF2-predicted models with respect to the hydrocarbon core of the lipid bilayer [18,19]. Percentages of identity and homology between protein sequences were calculated using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and Ident and Sim (https://www.bioinformatics.org/sms2/ident_sim.html). The ConSurf server (https://consurf.tau.ac.il/consurf_index.php) has been used for estimating the evolutionary conservation of aa residues in TMEM165 and homologs based on the phylogenetic relations between homologous sequences [20–23]. The AF2-predicted top-ranked PDB files were submitted to the ConSurf server with default parameters using a selection of 150 sequences with % of identity between 95 and 35 and a bayesian evolutionary conservation method.

2.2. Modeling with lipids and water and MD

The AF2-predicted Δ80-TMEM165 model was computationally immersed in an equilibrated 1,2-dipalmitoylphosphatidylcholine (DPPC) bilayer, with all waters removed and the x and y coordinates expanded by a factor of 4. The system was restored to its reference area per lipid (57 $Å^2$) by iteratively shrinking the lipid x and y coordinates by 2 %, then increased to 5 % after 8 iterations. Every iterative step, the system was subjected to 100 steps of steepest descent energy minimization, applying strong position restraints (10^5 kJ/nm^2) on the protein non-hydrogen atoms, and lipids that had their phosphorus atom at a distance closer than 6 Å to any calcium (Ca) atom of the protein were removed [24]. The final system was solvated with roughly 20,000 simple point charge (SPC) water molecules, neutralized by adding Na^+ ions and subjected to 1000 steps of steepest descent energy minimization and 100 ps MD using weak position restraints (10^3 kJ/nm^2). The system was then run for 10 ns free MD, using a time step of 2 fs, by GROMACS version 2018.8, employing the Gromos96 54a7 force field with Berger parameters for the lipid tails. The system was coupled to a temperature bath at 310 K through velocity scaling, with a coupling constant of 0.1 ps. Protein, lipids and solvent were coupled independently. Pressure was maintained at 1 bar using isotropic pressure coupling with a coupling constant of 1 ps [25]. Van der Waals interactions were cut off at a distance of 1.4 nm, and electrostatic interactions were calculated with the particle mesh Ewald method [26]. Equations of motion for the water molecules were solved analytically [27] and all covalent bonds in the system were constrained in the MD simulations [28]. The same protocol was used for modeling TMEM165 within 1-palmitoyl-2-oleoylphosphatidylethanolamine (POPE) and 1,2-dimyristoylphosphatidylcholine (DMPC) lipid bilayers.

2.3. Cell culture

Control and TMEM165 KO HEK293T cells generated as described in [11] were maintained in Dulbecco's Modified Eagle's Medium (Lonza, Basel, Switzerland) supplemented with 10 % fetal bovine serum (FBS) (PAN Biotech, Aidenbach, Germany), at 37 °C with 5 % CO_2 and humidity-saturated atmosphere.

2.4. Antibodies

Mouse monoclonal (IgG) anti-LAMP2 (H4B4) was purchased from Santa Cruz Biotechnology (Dallas, USA) (catalog # sc18822) and diluted at 1:2000. Rabbit polyclonal anti-TMEM165 was purchased from Sigma-Aldrich (Prestige Antibodies, St-Louis, USA) (catalog # HPA038299) and used at 1:100. Mouse monoclonal (IgG1) anti-GM130 was purchased from BD Biosciences (Franklin lakes, USA) (catalog # 610822) and diluted at 1:100. Mouse monoclonal (IgG1) anti-β-Actin was purchased from Sigma-Aldrich (St-Louis, USA) (catalog # A1978) and diluted at 1:10,000. Polyclonal goat anti-mouse immunoglobulins HRP conjugated and polyclonal goat anti-rabbit immunoglobulins HRP conjugated were purchased from DAKO (Glostrup, Denmark) (catalog # 56970 & 7074, respectively)

and diluted at 1:20,000 and 1:10,000, respectively. Conjugated Alexa Fluor Polyclonal goat anti-rabbit (GAR 568) or conjugated Alexa Fluor goat anti-mouse (GAM 488) were purchased from Thermo Fisher Scientific (Waltham, USA) (catalog # A-11011 & A-21121, respectively) and diluted at 1:600.

2.5. Plasmids and transfections

The plasmids used in this study are encoding control and mutant TMEM165 forms (WT-TMEM165, M > R^{140}-TMEM165, G > R^{304}-TMEM165, G > K^{304}-TMEM165, and G > H^{304}-TMEM165). All the plasmids were generated by a tailor-made DNA plasmids method by e-Zyvec (Polyplus, Loos, France).

For transfection, control and TMEM165 KO HEK293T cells were seeded in 6-well plates 24 h before the transfection, for them to reach 70 % of confluency. On the transfection day, cells were washed twice with Opti-MEM culture medium (Gibco, Waltham, USA) and transfected with 1 µg of plasmids using Lipofectamine™ 2000 reagent (Invitrogen, Waltham, USA) in Opti-MEM (Gibco, Waltham, USA) according to the manufacturer's protocol. Four hours and half after the transfection, Opti-MEM was replaced by DMEM containing 10 % of FBS, until collect, 48 h later.

2.6. Western blot analysis

Forty-eight hours post-transfection, cells were placed on ice to be collected. The cells were washed twice with cold Phosphate-buffered saline (PBS Calcium and Magnesium free) (Euromedex, Strasbourg, France) and collected by up and down pipetting. Cells were then centrifuged at 7500 rpm at 4 °C for 10 min. Supernatant was discarded and cells were lysed by up and down pipetting through a pipette tip with RIPA buffer (Tris/HCl 50 mM pH 7.9, EDTA 1 mM, NaCl 120 mM, NP40 0.5 %, NaF 5 mM, Sodium Orthovanadate 1 mM) containing a mixture of protease inhibitors (Roche Holding, Basel, Switzerland). Samples were vortexed 2 times for 10 s and then centrifuged 30 min at 14,000 rpm, 4 °C. Protein concentration in the supernatant (protein lysate) was estimated using the micro BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, USA). For preparation of Western-Blot samples, 10 µg of protein were mixed with water (to reach a volume of 15 µL) and with 5 µL of NuPAGE LDS 4x-concentrated sample buffer (pH 8.4, Invitrogen, Waltham, USA) supplemented with 4 % of β-mercaptoethanol (Sigma-Aldrich, St-Louis, USA). Samples were then denatured for 10 min at 95 °C and separated on NuPAGE 4–12 % Bis-Tris precast polyacrylamide gel in MOPS running buffer (Invitrogen, Waltham, USA) and transferred onto nitrocellulose membrane (0.2 µm pore size) using iBlot2 Dry Blotting System (Thermo Fisher Scientific, Waltham, USA). The membrane was blocked in blocking buffer (5 % non-fatty milk powder (Leclerc, France) resuspended in TBS-Tween 0.05 %) (Euromedex, Strasbourg, France) at room temperature. After 1 h in blocking buffer, the membrane was incubated with a diluted primary antibody in blocking buffer overnight, at 4 °C and under agitation (20 rpm). The next day, the membrane was washed 3 times for 5 min with TBS-Tween. Secondary antibodies were then diluted in the same blocking buffer and incubated at room temperature for 1 h. The membrane was then washed 5 times for 5 min in TBS-Tween. Signal was detected with West Pico plus chemiluminescent substrate (Thermo Fisher Scientific, Waltham, USA) using Camera Fusion (Vilber, Collégien, France) and its software.

2.7. Immunofluorescence and image analysis

Cells were seeded on coverslips (VWR, Radnor, USA) in 6-well plates 24 h before the transfection. Forty-eight hours after the transfection and in non-sterile conditions, the cells were washed three times in Dulbecco's Phosphate-buffered saline (DPBS with calcium and magnesium) (Hyclone, Utah, USA). The coverslips were fixed in 4 % paraformaldehyde for 20 min at room temperature. After three washes with DPBS, the coverslips were placed in a humid atmosphere. Cells were permeabilized with PBS-Triton X-100 0.5 % (Sigma-Aldrich, St-Louis, USA). After 10 min, the coverslips were washed three times with Phosphate-buffered saline (PBS calcium and magnesium free) (Euromedex, Strasbourg, France) and cells were blocked for 1 h in blocking buffer [0.2 % gelatin, 2 % Bovine Serum Albumin (BSA), 2 % FBS in PBS]. Coverslips were incubated with primary antibodies in blocking buffer for another hour. The cells were washed 3 times with PBS and incubated with secondary antibodies conjugated with Alexa-488 and Alexa-568, in darkness, for 1 h. The coverslips were washed 3 times in PBS and cells were incubated for 15 min with 5 µg/mL DAPI (Sigma-Aldrich, St-Louis, USA) in the dark. The cells were washed once again with PBS. The coverslips were individually washed 10 times in Milli-Q water and mounted in 6 µL of MOWIOL 4–88 (Calbiotech, El Cajon, USA) on microscope slides. Fluorescence was detected through an inverted Zeiss LSM780 or LSM700 confocal microscope. Acquisitions were done using the ZEN pro 2.1 software (Zeiss, Oberkochen, Germany) and images were analyzed using Fiji software (https://imagej.net/) and homemade plugin.

3. Results and discussion

3.1. AlphaFold 2 predicts the structure of TMEM165 with high confidence

As previously reviewed, the human TMEM165 primary sequence encompasses 324 aa residues and was initially predicted to contain 7 TMDs by the commonly-used membrane protein topology prediction methods such as TMHMM [12–14]. However, it can be noted that the presence of the first TMD in the mature protein has been questioned, owing to its low prediction score by membrane protein topology prediction methods. Furthermore, both the length and the composition of the *N*-terminal aa residues upstream from the second TMHMM-predicted TMD of TMEM165 are highly variable within the CaCA2 family. For example, the TMEM165 homologs from *S. cerevisiae* (Gdt1p) [8], prokaryotes (*Vibrio cholerae* (MneA)) [29] and *Synechocystis* (SynPAM71/MNX) [30,31] possess a *N*-terminal segment whose length is 31, 11 and 2 aa residues, respectively, instead of the 84 aa residues for the human TMEM165. In contrast, TMEM165 homologs in *Arabidopsis thaliana* thylakoids (PAM71) and chloroplasts (CMT1) have *N*-terminal segments encompassing 150 and 142 aa residues, respectively, with organelle-addressing signal peptides [32].

AF2 prediction conducted on the full-size human TMEM165 sequence (UniProtKB ID Q9HC07) provided five models with average pLDDT (predicted local distance difference test), a per-residue estimate of its confidence on a scale from 0 to 100 [15], ranging from 71.56 to 74.50. Detailed pLDDT values from the best ranked TMEM165 model are illustrated in Fig. 1A. Among the 324 aa residues of the protein, 152 have pLDDT > 90 (up to 97.0), 53 have pLDDT ≥ 80 and < 90 and 3 have pLDDT ≥ 70 and < 80. This indicates that, according to AF2 estimations, a total of 210 aa residues (64.8 %) are expected to be correctly modeled, among which three-quarters with high accuracy. Three regions of the protein have medium/low to very low pLDDT values, preventing sufficient confidence in prediction: the *N*-terminal end of the protein (86 aa residues), aa residues 211–233, and the *C*-terminal tail of the protein (6 aa residues). In particular, the *N*-terminal end exhibits a pLDDT lower than 50, suggesting that this region is either unstructured in physiological conditions or only structured as part of a complex with protein partners. Fig. 1B (up) depicts the full-size TMEM165 sequence

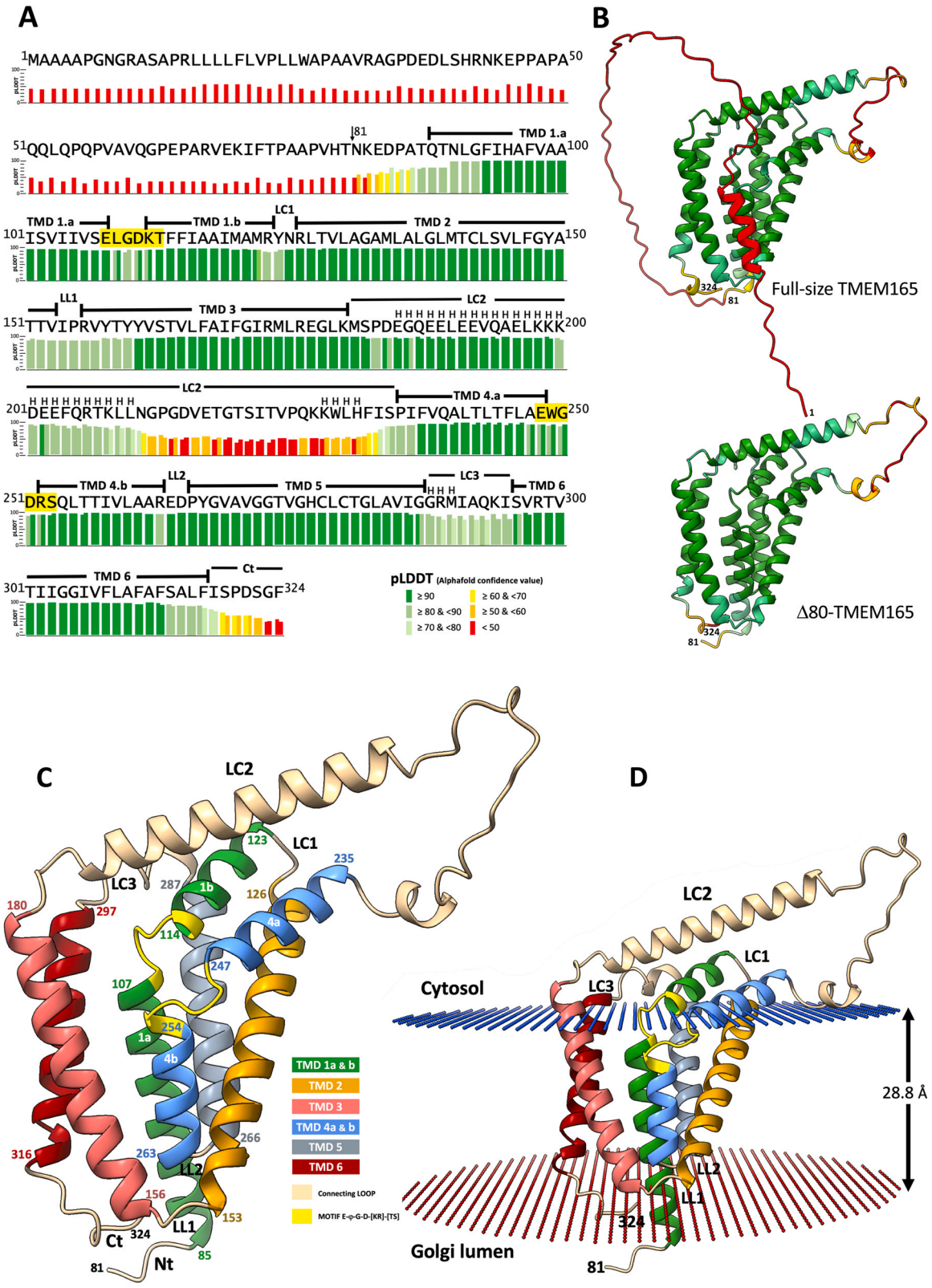

Fig. 1. AF2-predicted structure of human TMEM165. A) Primary sequence of TMEM165 showing the predicted TMDs and loops, together with vertical bars corresponding to the pLDDT values (AF2 confidence value from 0 to 100) for each aa residue. The right-bottom insert indicates the color correspondence with pLDDT values. For each vertical bar, the left and right sides correspond to the pLDDT values from the predicted full-size molecule and Δ80-TMEM165, respectively. H letters indicate the aa residues predicted to form helices in loops LC2 and LC3. Amino acid residues highlighted in yellow belong to the two consensus motifs E-φ-G-D-[KR]-[TS]. B) Full-size (top) and Δ80- (bottom) TMEM165 AF2-predicted models with ribbon diagrams colored according to pLDDT values indicated in the insert. C) Ribbon diagram of the predicted Δ80-TMEM165 structure. The insert indicates the colors used to identify the predicted TMD helices and loops, and the positions of the E-φ-G-D-[KR]-[TS] motifs. Colored numbers indicate the aa residues delimiting the TMD helices. D) Predicted insertion and orientation of the Δ80-TMEM165 model in a flat lipid bilayer. The blue and red disks indicate the surfaces of the hydrophobic layer at the cytosolic and Golgi lumen sides, respectively, as calculated by the PPM 2.0 Web Server of the OPM (Orientation of Proteins in Membranes) database (see Material and methods). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

predicted model showing the *N*-terminal end in a disordered state, except for the presence of a short helix which however was predicted in only two out of the five AF2 models. Owing to the disordered nature of this region, and to avoid possible interference with the rest of the molecule, the first 80 *N*-terminal aa residues were deleted from TMEM165 sequence and a novel AF2 prediction was performed. As shown in Fig. 1A and B (bottom), the obtained top-ranked model was very similar to the previous one between aa residues 81 and 234, with a mean pLDDT of 86.3, and quite similar per-residue confidence values. Hence, the delta-80 *N*-terminal aa residues model (Δ80-TMEM165 model) has been used as a template in the paper for further studies and predictions.

The predicted structural domains of TMEM165 are depicted in Fig. 1 C. As formerly deduced from membrane protein topology prediction methods, the Δ80-TMEM165 scaffold encompasses six transmembrane domains (TMD) formed by a two-fold repeat of three transmembrane helices (TMD 1–3 and TMD 4–6) where TMD 1 and 4 are antiparallel helices bearing at their midpoint a loop containing the consensus motifs E-φ-G-D-[KR]-[TS] (residues 108–113 and 248–253). Very interestingly, whereas the membrane protein topology prediction methods assumed orientation of these two motifs on either side of the Golgi membrane (cytosolic and lumen sides for motifs 1 and 2, respectively) [12,13], our model shows that the motif-containing loops are facing each other and are cytosolically oriented (Fig. 1 C and D). As calculated by the PPM 2.0 Web Server of the OPM (Orientation of Proteins in Membranes) database, the model indeed fits into a flat lipid bilayer with a hydrophobic depth/thickness of 28.8 ± 2.2 Å, a $\Delta G_{transfer}$ of − 49.3 kcal/mol and a tilt angle of 21 ± 1°. In this model, both motif-containing loops are clearly located at the interface of the hydrophobic layer with the cytosol. They are in close proximity to the beginning of the large α-helix of the cytosolic LC2 loop, with both halves of TMD 1 and 4 (TMD 1b and TMD 4a) resting on top of that interface. Elsewhere, whereas cytosolic loops LC1 and LC3 are predicted at the surface of the hydrophobic lipid layer, the short luminal loops LL1 and LL2 are found rather buried in the layer (Fig. 1D). These findings will be supported later in this paper from modeling of TMEM165 immersed in an equilibrated phospholipid bilayer solvated with water molecules. A focus on the predicted structural features of TMEM165 allowing speculation on the effect of protein mutations on its function and its possible mechanism of cation transport will be also made in the following paragraphs.

3.2. Pathogenicity of patient mutations solved in the light of the TMEM165 model

A major outcome of the AF2-predicted model is the presence of a cation-binding site consisting of both conserved consensus motifs E-φ-G-D-[KR]-[TS] (^{108}ELGDKT^{113} and ^{248}EWGDRS^{253} in human TMEM165). Indeed, those motifs form an “acidic cage” potentially responsible for the binding of divalent cations such as Mn^{2+} and/or Ca^{2+} (Fig. 2 A and B). Mn^{2+} and Ca^{2+} are actually known to be coordinated in proteins by oxygen atoms provided by carboxylate moieties of Asp and Glu residues and carbonyls of the polypeptide chain, as well as oxygen atoms of water molecules for Mn^{2+} [33–35]. The TMEM165 model highly supports the hypothesis that those acidic aa residues E^{108}, D^{111}, E^{248} and D^{251} are essential for the function of the protein. As shown in Fig. 2B, they adopt a trigonal bipyramidal geometry that is reported to coordinate Mn^{2+} in Mn^{2+}-dependent proteins [35]. This assumption is corroborated by the findings that *i)* mutation E > G^{108} was found in a TMEM165-CDG patient [36], *ii)* all four acidic aa residues seem crucial for the role of TMEM165 in glycosylation and/or Mn^{2+} sensitivity, as experimentally shown in [7], and *iii)* those aa residues are highly conserved during evolution (Supplementary Fig. 1).

Unexpectedly, another TMEM165-CDG patient mutation greatly enhanced our confidence in the AF2-predicted model. This is the case for mutation G > R^{304} [1], whose impact on protein function remained hitherto unexplained using classical membrane protein topology prediction methods. According to TMHMM prediction, this mutation was indeed located in the last TMD of the protein far away from the conserved consensus motifs [12,14]. The model actually predicts that the guanidium group of R^{304} may readily interact with the side-chain carboxyl group of D^{111}, probably through electrostatic bonding at physiological pH (Fig. 3 A, bottom right model). The occurrence of a pair of hydrogen bonds between two hydrogens of the guanidium group and one oxygen of the carboxyl can even be seen in Fig. 3 A (bottom right model). Hence, interference of R^{304} with the cation-binding site of TMEM165 would explain the molecular mechanisms by which the mutation G > R^{304} impairs its transporter function. In order to go further and support this hypothesis, we specifically expressed different TMEM165 mutated proteins at position 304 in TMEM165 KO HEK293T cells and their function was followed by our LAMP2 glycosylation readout, as previously described [7,8]. We did replace Arg at position 304 by His (G > H^{304})

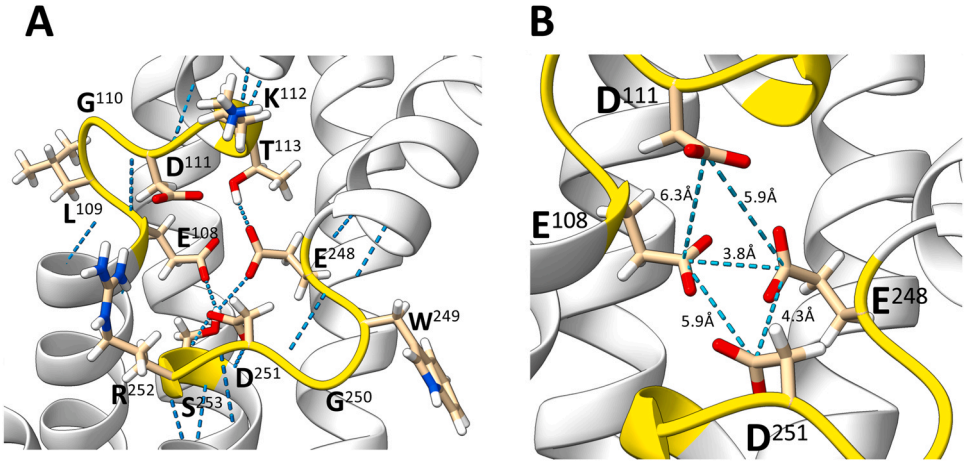

Fig. 2. Putative cation-binding site of TMEM165. A) Partial ribbon diagram of TMEM165 (pale gray) showing the lateral chains of aa residues present within the consensus motifs E-φ-G-D-[KR]-[TS] (colored in yellow). The blue dashed lines show hydrogen bonds exchanged by those aa residues with their neighborhood in the model, as predicted by ChimeraX. B) Partial ribbon diagram of TMEM165 (pale gray) showing the geometry of the “acidic cage” formed by both consensus motifs E-φ-G-D-[KR]-[TS] (colored in yellow). Average distances between the carbonyl groups of aa residues E^{108}, D^{111}, E^{248} and D^{251} are indicated with blue dashed lines and values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

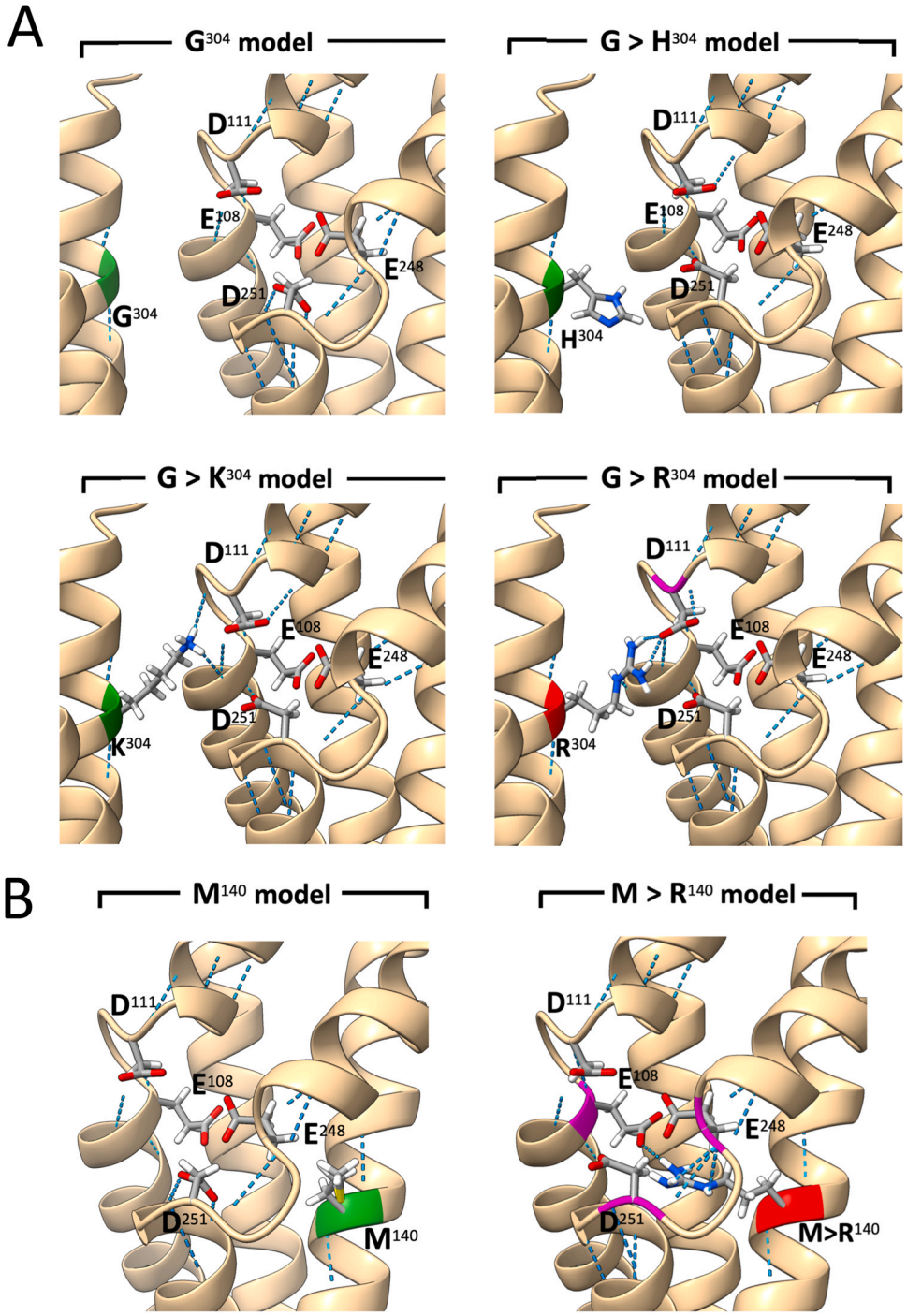

Fig. 3. Partial ribbon diagrams of TMEM165 with mutations at positions 104 and 304. A) Models without (top left) and with mutations G > H^{304} (top right), G > K^{304} (lower left) and G > R^{304} (lower right). B) Models without (left) and with mutation M > R^{104} (right). The models only show the lateral side chains of the acidic aa residues from the conserved motifs and that of the mutated aa residue. The mutated aa residues are colored according to the capacity of corresponding mutants to restore LAMP2 glycosylation when expressed in TMEM165 KO HEK293T cells in the present study: full (green) or no (red) restauration. The acidic aa residues of the motifs possibly impacted by electrostatic interactions and/or hydrogen bonds with the basic charges of mutated aa residues are colored in magenta. The blue dashed lines show the hydrogen bonds exchanged by aa residues of the consensus motifs and mutated aa residues with their neighborhood in the model, as predicted by ChimeraX. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and Lys (G > K^{304}), two basic aa residues with different structural and ionic properties. Very interestingly, our results show that, contrary to the expression of the G > R^{304} mutant but similarly to that of wild type (WT) TMEM165, expression of either G > H^{304} or G > K^{304} mutants in TMEM165 KO HEK293T cells rescued LAMP2 glycosylation (Fig. 4 A). Indeed, LAMP2 protein bands with slower SDS-PAGE migration rates, similar to that of LAMP2 in wild-type cells, can be observed in TMEM165 KO cells expressing either wild type-TMEM165, G > H^{304}-TMEM165 or G > K^{304}-TMEM165. This indicates the neosynthesis of fully-glycosylated LAMP2 forms in those transfected cells and therefore the expression of a functional form of TMEM165. Expression of the TMEM165 variants in transfected cells is shown by Western blot and immunostaining in Supplementary Fig. 2. These results can be explained by the proposed TMEM165 models shown in Fig. 3 A. Indeed, the G > H^{304} model clearly shows that the R group of H^{304} does not interact with D^{111} or any of the aa

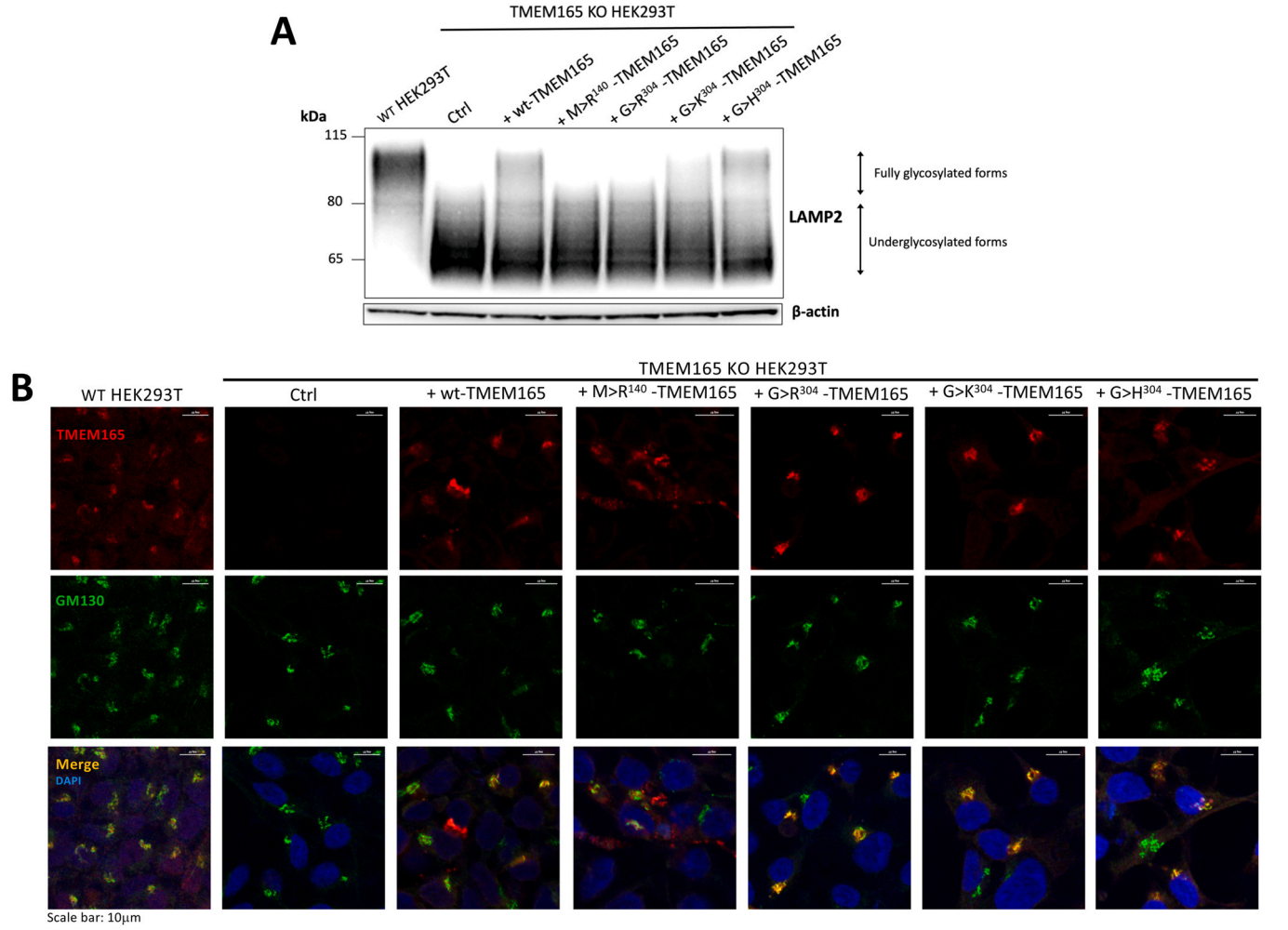

Fig. 4. Expression of TMEM165 mutants in TMEM165 KO HEK293T cells. A) LAMP2 glycosylation profiles of TMEM165 KO HEK293T cells transfected with empty vector (Ctrl), wild-type (WT) and different TMEM165 constructs. Western blot and immunostaining were performed as described in Material and methods. B) Immunofluorescence analysis of the expression and localization of TMEM165 in cells transfected with the mutated forms of TMEM165. Immunofluorescence and image analysis were performed as described in Material and methods. Fluorescence of TMEM165 and GM130 (Golgi marker) is illustrated in red and green, respectively. The fluorescence merge of both proteins is shown with nucleus DAPI staining (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

residues involved in the cation-binding pocket. The same explanation applies to K^{304}. The G > K^{304} TMEM165 model shows that the maximal length of the R group of K^{304} would allow electrostatic interactions with the carboxyl group of D^{111} (distance of 4.3 Å between the nitrogen of the amine group of K^{304} and one oxygen of the carboxylate of D^{111}) but due to the slightly shorter length of its side chain and weaker polarization than R^{304}, it may be hypothesized that its potential to interfere with D^{111} is significantly weaker than in the case of R^{304}. In support to this, differences in the conformation and interactions of both aa residues in a physiological environment has been thoroughly documented [37,38]. Altogether these results not only explain why a mere G > R^{304} mutation abolishes the function of TMEM165 but they also validate the TMEM165 model predicted by AF2.

To further assess our confidence in the predicted TMEM165 model, we investigated the possibility to introduce a mutation in another part of the protein that would be able, like the G > R^{304}, to interfere with the cation-binding site and thus impair TMEM165 function. We hypothesized that introducing a positive charge (Arg) at position 140, which is located on TMD 2 at the opposite of G^{304}, may offer that opportunity. Indeed, as shown in Fig. 3B, the AF2 model predicts that mutation M > R^{140} places the guanidium group of Arg in close proximity to the cation-binding site, most particularly to E^{108}, E^{248} and D^{251} (distances between the guanidium group and carboxylates ranging from 1.7 to 3.5 Å). In contrast, a greater distance separates the R^{140} guanidium group from D^{111} carboxylate (6.2 Å), which makes the interactions between those aa residues less likely. Our choice of M^{140} was reinforced by our previous experiments showing that mutation M > G^{140} did not impair the function of TMEM165 in glycosylation [7], which indicates that the side chain of M^{140} is not essential for the transporter function. Very interestingly, expression of the M > R^{140} mutant in TMEM165 KO HEK293T cells did not rescue LAMP2 glycosylation, similarly to the G > R^{304} mutant (Fig. 4 A). As shown in Fig. 4B, all TMEM165 mutants localized in the Golgi of cells. Nevertheless, it has to be noted that the M > R^{140} mutant also significantly accumulates in punctuated structures, likely endosomal/lysosomal compartments. This suggests that strong interactions between R^{140} and the cation-binding site may not only perturb the function of TMEM165 but also its fate. Taken as a whole, these results greatly support our hypothesis that introducing a basic aa residue in a position that directly interferes with any part of the molecular determinants of the cation-binding site of TMEM165, as it is the case for mutation G > R^{304} found in TMEM165-CDG patients, would impair its function. Altogether, they significantly consolidate our confidence in the TMEM165 model predicted by AF2.

It should also be mentioned that gene mutations other than E > G^{108} and G > R^{304} have been identified in TMEM165-CDG

patients [1,36]. More specifically, in addition to the mutations that activate a cryptic splice donor site (c.792 + 182 G > A) leading to a lack of protein expression, missense mutations c.377 > A and c 0.377 C > T had been found [1,36], leading to proteins with mutations R > H^{126} and R > C^{126}, respectively. Through expression of *C*-terminally RFP-tagged TMEM165 proteins in HeLa cells, we showed that the R > H^{126} and R > C^{126} mutants preferentially localized in endosomes and lysosomes, while the wild-type (wt)TMEM165 preferentially localized in the Golgi [6]. In addition, with regard to R^{126}, whose mutations R > H and R > C in TMEM165-CDG patients preferentially lead the protein towards endosomes and lysosomes [6], it was hypothesized that it belongs to a putative tyrosine-based lysosomal-targeting signal YXXØ in TMEM165 (^{124}YNRL^{127}). This signal sequence is known to be able to interact with heterotetrameric adaptor protein (AP) complexes AP1, AP2, AP3 or AP4, which recruit clathrin to initiate the formation of coated vesicles [39]. According to the TMEM165 model, the ^{124}YNRL^{127} sequence forms the short cytoplasmic loop LC1 which protrudes towards the cytoplasmic space (Fig. 1B and C) but whose degree of accessibility will be commented later in the section of the manuscript dealing with the simulation of TMEM165 in a phospholipid bilayer.

Lastly, previous experimental data not only confirmed the importance of E^{108}, D^{111}, E^{248} and D^{251} in the function of TMEM165, but also showed the importance of aa residues within and at the vicinity of the two consensus motifs, such as T^{113}, F^{114}, S^{253} and Q^{254} [7] (illustrated in both protein sequence and model in Supplementary Fig. 3). Whereas it is hard to speculate about the role of F^{114} and Q^{254} in the binding of the divalent cation by TMEM165 and/or its transport function, since their side chains stay clearly apart from the consensus motifs, it may be hypothesized that both T^{113} and S^{253} play a direct role in this binding, or at least in the stabilization of the binding site. Indeed, as shown in Fig. 2A, the lateral chains of those two aa residues are found in very close positions (about 2 Å) to E^{108}, E^{248} and D^{251}, with which they share several hydrogen bonds. As to the impact of mutations of F^{114} and Q^{254} on the protein, it may be hypothesized that they contribute to destabilize the structure and/or impair the function of the protein. The partial loss of function of TMEM165 mutated at positions 118, 120, 122 and 123 (Supplementary Fig. 3) also suggests that the structural integrity of TMD 1b must be preserved for the transport function.

3.3. TMEM165 and its structural and functional homologs

As previously mentioned, TMEM165 belongs to the CaCA2/UPF0016 family encompassing many homologs from bacteria, plants and eukaryotes (https://www.tcdb.org). Sound pieces of evidence were provided for their function as Mn^{2+} exporters, most particularly as compared to TMEM165 located in the Golgi apparatus of animal cells [8,11], Gdt1p in the Golgi apparatus of yeast [10], PML3/BICAT3 in the Golgi apparatus of plant cells [40], CMT1/PAM71-HL and PAM71 in plant chloroplasts/thylakoids [32], SynPAM71/MNX in cyanobacteria *Synechocystis* [30,31] and MneA in *V. cholerae* [29]. Very interestingly, AF2 predictions performed using the sequences of these homologous proteins, *N*-terminally truncated in most cases to avoid casual interference of the unstructured *N*-terminal tail with the rest of the protein, reveal foldings that are quite similar to human TMEM165, thus confirming their family membership and the consistency of AF2 predictions (Fig. 5). It should be noted that, when

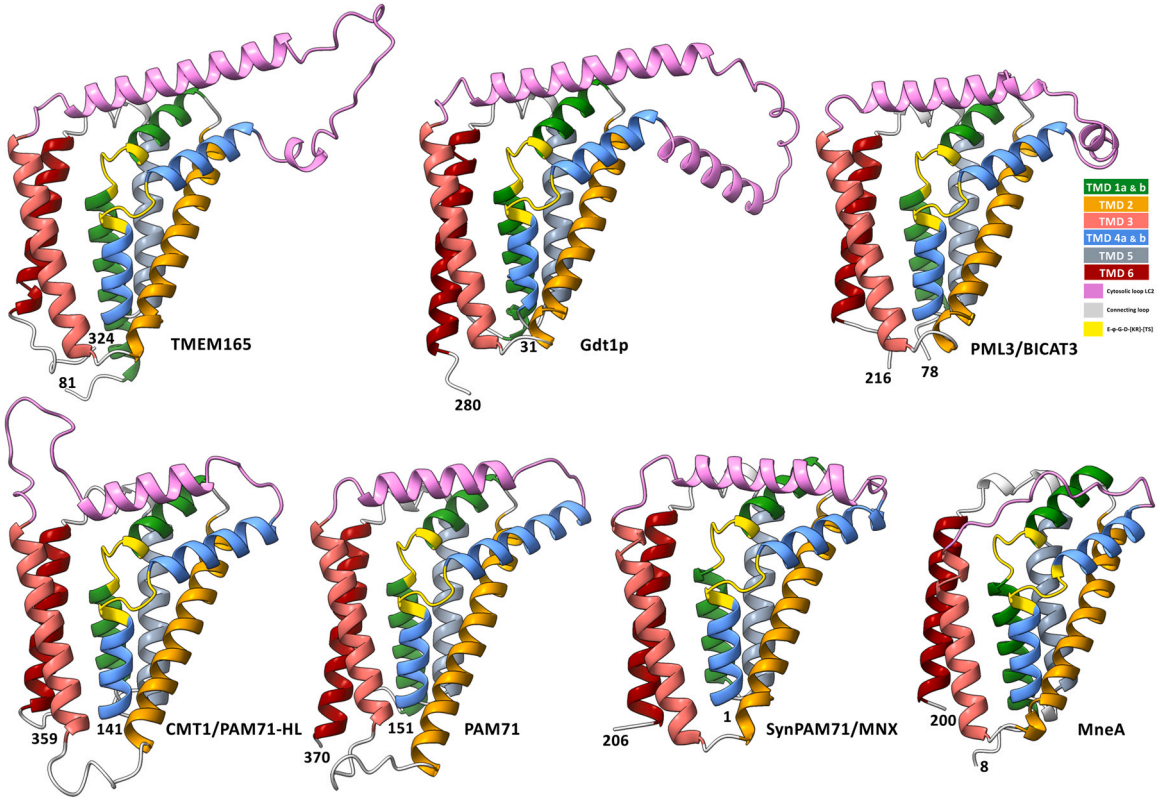

Fig. 5. Compared 3D models of TMEM165 and structural and functional homologs of the CaCA2 family. The figure shows the ribbon diagrams predicted by AF2 of human TMEM165 (UniProtKB ID Q9HC07) (mean pLDDT = 86.3) located in the Golgi apparatus, yeast Golgi Gdt1p (*S. cerevisiae* – UniProtKB ID P38301) (mean pLDDT = 83.8), plant Golgi PML3/BICAT3 (*A. thaliana* - UniProtKB ID Q93Y38) (mean pLDDT = 84.6), chloroplast inner membrane CMT1/PAM71-HL (*A. thaliana* – UniProtKB ID Q9T0H9) (mean pLDDT = 83.9), thylakoid membrane PAM71 (*A. thaliana* – UniProtKB ID Q94AX5) (mean pLDDT = 88.5) and SynPAM71/MNX (*Synechocystis* – UniProtKB ID P52876) (mean pLDDT = 93.9), and plasma membrane MneA (*V. cholerae* – UniprotKB ID A0A0H3AJF5) (mean pLDDT = 88.4). For all proteins, except SynPAM71, *N*-terminally-truncated sequences were used for AF2 predictions, as indicated on the figure. The top-right insert indicates the colors used to identify the homologous TMD helices and loops in the models. The LC2 loop of each protein model is colored in light purple. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

compared to the Δ80-TMEM165 sequence, sequence alignments of the *N*-terminally-truncated protein sequences using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) indicate identity/similarity percentages ranging from 38/55 for Δ77-PML3/BICAT3 to as low as 19/38 for Δ7-MneA. It may also be observed that the equivalent LC2 loop in proteins may highly vary in length, the longest within Gdt1p, the protein belonging to *S. cerevisiae*, and the shortest being MneA belonging to *V. cholerae* (Fig. 5).

Even more interestingly, a previous study expanded the LysE superfamily to new members through the identification of internal repeats and conserved motifs, multiple alignments, phylogenetic trees and average hydropathy, amphipathicity and similarity [41]. Members of this superfamily were formerly identified in prokaryotes as L-Lys and L-Arg efflux transporters (LysE family) but also as exporters of homoserine/threonine (RhtB family) and Cd^{2+} (CadD family) [42]. Tsu and Tseir's study included newly-identified protein families into the LysE superfamily that now encompasses proteins specifically transporting tellurium (TerC family), iron/lead (ILT family), Mn^{2+} (MntP family), Mn^{2+}/Ca^{2+} (CaCA2 family, thus including MneA and TMEM165), Ni^{2+} and Co^{2+} (NiCo family), neutral amino acids (NAAT family), peptidoglycolipids (GAP family) and even electrons (DsbD family) [41]. Most LysE superfamily proteins share similar sizes, around 200 aa residues, with 6 predicted TMDs. Impressively, AF2 predicts for any member of these familes a protein scaffold comparable to that of TMEM165 and its homologs of the CaCA2 family, despite their relatively low identity and similarity percentages with TMEM165, up to 15 % and 30 %, respectively (illustrated for *Thermococcus* sp. SnatA (NAAT/UPF0056 family) and *E. coli* MntP (MNTP family) in Supplementary Fig. 4). Two archaeal membrane electron transporters from the DsbD family, CcdA from *Archaeoglobus fulgidus* and *Thermus thermophilus*, need special attention since their structures had been investigated by NMR [43,44]. Both proteins exhibited a two-fold repeat of 3 TMDs with redox active cysteines located in conserved PCxxP motifs on TMD 1 and 4. This clearly indicates that TMEM165 originates from a large family of proteins, deriving from a single common 3-TMD precursor peptide *via* intragenic gene duplication, as previously postulated for the UPF0016 family members [45], used for the export of not only ions but also electrons and biomolecules.

3.4. Molecular dynamics simulation of TMEM165 in a phospholipid bilayer

To further refine the TMEM165 model in a cell membrane, the AF2-predicted Δ80-TMEM165 model was processed as described in Material and methods. Briefly, the predicted structure was first computationally immersed in a DPPC (1,2-dipalmitoylphosphatidylcholine), POPE (1-palmitoyl-2-oleoylphosphatidylethanolamine) and DMPC (1,2-dimyristoylphosphatidylcholine) bilayer with all water removed, subjected to iterative steps of energy minimization, solvated with SPC (simple point charge) water molecules and subjected again to energy minimization steps. The final systems were then run for 10 ns free molecular dynamics (MD) simulation. The root mean square deviation with respect to the starting structure is shown in Supplementary Fig. 5A. Together with Supplementary Fig. 5C, which shows the average residue root mean square deviation over the last 8 ns of the simulations, we can conclude that all three simulations are sufficiently equilibrated and that any fluctuations are due to the unstructured region linking the LC2 helix and TMD 4a (aa residues 210–225 in full-size TMEM165, 130–145 in the Δ80-TMEM165 model). However, as expected, the DPPC bilayer corresponds best to the hydrophobic thickness of TMEM165 and we will focus our analyses henceforth on that system primarily.

Fig. 6 depicts the topology of TMEM165 within the phospholipid bilayer in presence of water molecules after MD simulation, while the corresponding structure prior to MD is shown in Supplementary

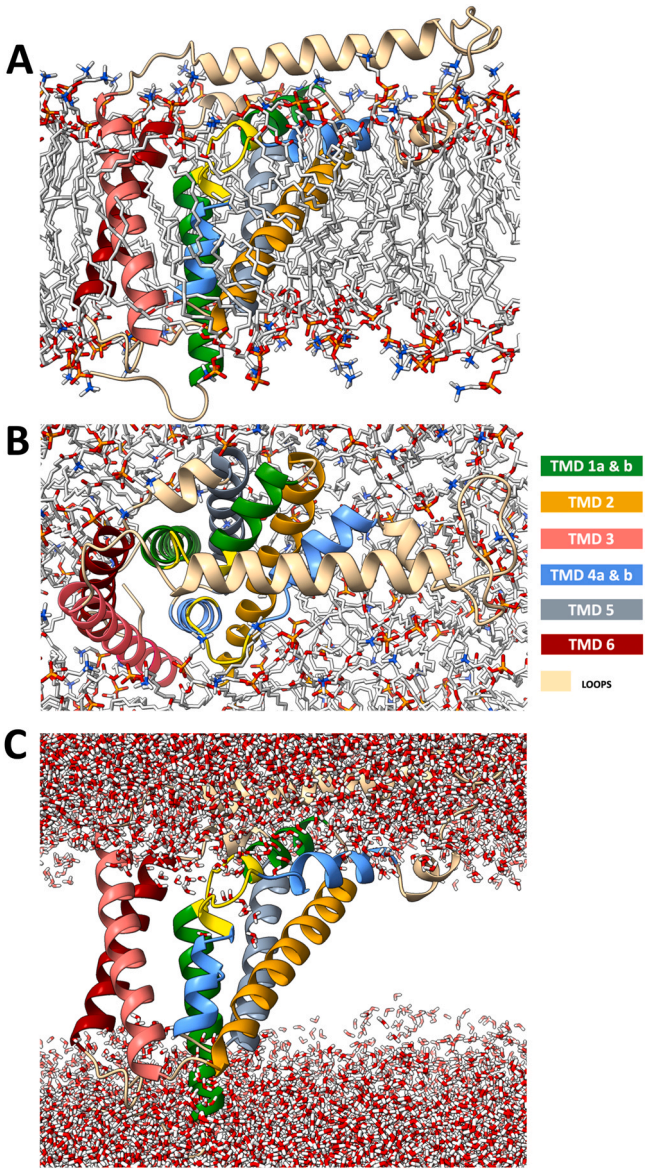

Fig. 6. Topology of Δ80-TMEM165 within a DPPC lipid bilayer in presence of water molecules after MD simulation. Modeling and MD simulation were performed as described in Material and methods. A) Side view of TMEM165 within the phospholipid bilayer, without water molecules. For permitting clearer viewing of the protein ribbon diagram, only the phospholipids surrounding the model at a distance less than or equal to 15 Å are represented. B) Top view of TMEM165 within the phospholipid bilayer. C) Same view as in A) but only the protein model and water molecules are represented. The right-bottom insert indicates the colors used to identify the predicted TMD helices and loops. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 6. TMEM165 is completely embedded within the phospholipid bilayer with the exception of LC2 that protrudes above the bilayer (Fig. 6 A). Furthermore, a cleft is formed between TMD 3 and TMD 6 from one side, and TMD 1a and TMD 4b from the other side, which virtually allows communication across the lipid bilayer (Fig. 6B). However, it is unlikely that, in this predicted configuration, the cleft may be permeable to water and ions. Indeed, although water molecules could be observed along this cleft in the model prior to MD (Supplementary Fig. 6B), these water molecules were excluded from the cleft following MD simulation (Fig. 6 C). In the MD model, water molecules were only present at both cytosolic and luminal ends of the cleft, as well as in the environment of the potential cation-binding site, itself buried at the interface of polar heads and apolar

chains of the cytosolic lipid layer (Fig. 6B). These observations find an explanation by the fact that the cleft is highly hydrophobic in its central part with lots of apolar side chains of aa residues, most of them highly conserved, such as L^{166}, F^{167}, F^{170}, M^{174}, F^{308} and F^{311} on both TMD 3 and 6, and V^{106}, V^{103} and L^{260} on both TMD 1a and 4b (Supplementary Fig. 7A). Conversely, many charged/polar aa residues, such as conserved T^{159}, S^{163} and R^{173}, are present at the cytosolic and luminal ends of the cleft, within the polar end groups of lipids (Supplementary Fig. 7B).

The differences between the two protein models, prior and after MD simulation, mainly relate to the position of the long LC2 loop and the conformation of the second consensus motif (aa residues 248–253) contributing to the potential cation-binding site of TMEM165 (Fig. 7 A and B); this suggests a significant degree of flexibility of the two regions, which is confirmed by the root mean square fluctuation plots of Supplementary Fig. 5B. With regard to LC2 loop, this observation is not surprising because, except for the long helix (aa residues 187–234) which was previously reported as a short coiled-coil helix [6], it belongs to the least-structured region of the predicted Δ80-TMEM165 model (Fig. 1). Interestingly, the MD model clearly shows that the LC2 helix rests at the cytosolic surface of the phospholipid bilayer (Fig. 6 A) and, as this will be commented in the next section, that the charged/polar aa lateral chains of the inner face of the helix are in a position to interact with polar head groups of DPPC. With regard to aa residues 248–253, presumably involved together with aa residues 108–113 in binding cations, a significant conformational change of the inter-helix loop may be observed upon MD simulation. As depicted in Fig. 7B, the major change concerns the *C*-terminal end of TMD 4a that displaces E^{248} from its initial position within the putative cation-binding site, about three-times as far from E^{108}, and about 1.5 times from both D^{111} and D^{251}. Although this has still to be confirmed experimentally, it might be hypothesized that the flexibility of this region could play an important role in the cation capture mechanism by TMEM165 at the cytosolic interface. Another interesting observation is given by MD simulations of TMEM165 in POPE and DMPC lipid bilayer systems, with layer thicknesses higher or lower than DPPC, respectively (Supplementary Fig. 8). Indeed, it can be observed that different positions of the long helix of LC2 and of the unstructured *C*-terminal sequence of LC2, depending on the thickness of the phospholipid bilayer, may have significant impact on the position/conformation of TMD 4a and the second consensus motif. For example, the average distances between the carbonyl groups of E^{108} and D^{248}, calculated over the full MD trajectories, vary from 10.1 Å (SD=2.2 Å) to 14.8 Å (SD=1.3 Å) and 15.3 Å (SD=1.6 Å) in DMPC, DPPC and POPE lipid bilayer systems, respectively. It may thus be hypothesized that LC2, by interacting with the membrane, ions and/or proteins at the cytosolic interface could play a role in the function of TMEM165. This will be further discussed in the last section of the paper.

Lastly, the TMEM165 model predicts that among the ^{124}YNRL^{127} sequence, a putative lysosomal-targeting signal [6], Y^{124} is the most accessible aa residue at the surface of the phospholipid bilayer. Tyrosine is indeed a critical aa residue of the YXXØ signal recognized by the AP complex [6]. With regard to R^{126} whose mutations R > H^{126} and R > C^{126} were found in TMEM165-CDG patients [1], this aa residue is located at the end of TMD 2 and possibly interacts with the phosphorus atoms of phospholipids for stabilizing the LC1 loop (Supplementary Fig. 9). Indeed, it has been reported that Arg possesses a unique ability to form bidentate H-bonds with phosphate groups of lipids [46]. This obviously does not exclude the possibility that R^{126} within the YNRL sequence may interact with the AP complex.

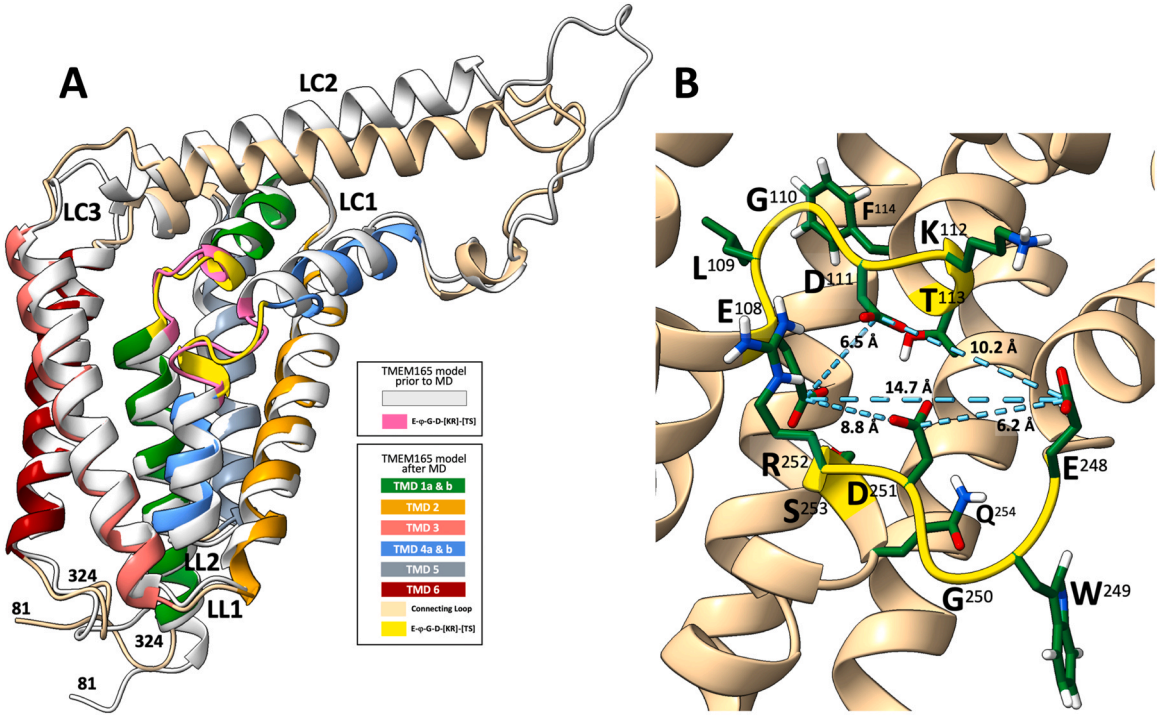

Fig. 7. Comparison of the structures of TMEM165 before and after MD. A) Superimposition of the predicted TMEM165 ribbon diagrams within a DPPC bilayer, prior and after 10 ns MD simulation. The right-side insert indicates the colors used to identify the models, TMD helices and loops. B) Structural changes of the putative cation-binding sites before and after 10 ns MD simulation. The ribbon diagrams show the changes in the geometry of the "acidic cage" formed by both consensus motifs E-φ-G-D-[KR]-[TS] (indicated in yellow with lateral chains in green). Distances between the carbon atoms of carbonyl groups of E^{108}, D^{111}, E^{248} and D^{251} are indicated with blue dashed lines and values. The values correspond to the distances at the end of the MD simulation. The average distances over the full MD trajectory were calculated as: 14.8 Å (standard deviation (SD)=1.3 Å) between E^{108} and E^{248}; 6.5 Å (SD=0.7 Å) between E^{108} and D^{111}; 12.2 Å (SD=1.4 Å) between D^{111} and E^{248}, 8.0 Å (SD=1.1 Å) between E^{108} and D^{251}; 8.4 Å (SD=1.5 Å) between E^{248} and D^{251}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.5. Clues toward the mechanism of cation transport by TMEM165 and its regulation?

Obviously, the *in silico* prediction of TMEM165 3D structure cannot answer all questions regarding the mechanism of cation transport and its regulation. Nevertheless, some clues may be drawn from predictions. The Mn^{2+} exporter MNTP family belonging to the LysE superfamily [47] provides interesting information on the molecular determinants that are essential for Mn^{2+} transport. Indeed, while the predicted 2-fold 3-TMDs protein scaffold of *E. coli* MntP is preserved and may be globally compared to that of TMEM165 (Supplementary Fig. 4B), the acidic aa residues found within the consensus motifs of the CaCA2 family are absent, except for the conservation of a single aa residue in TMD 1 (D^{16} within sequence ^{13}MSMDAFA^{19}) and TMD 4 (D^{118} within sequence ^{115}TSLDAMA^{121}). A glutamate residue (E^{47}) is also observed in TMD 3, just behind D^{16} and D^{118}, but is not conserved in all homologs within the MNTP family (data not shown). Since those two acidic aa residues are in homologous positions to D^{111} and D^{251} of TMEM165, it is expected that, among the acidic tetrad participating to the cation-binding site of TMEM165, D^{111} and D^{251} are the most important for Mn^{2+} binding. This assertion is supported, first, by the experimental data showing that mutation of E^{108} only partially impaired LAMP2 glycosylation [7] and, second, by the flexibility of E^{248} evidenced by MD in this study. Hence, it may be hypothesized that aa residues E^{108} and E^{248} participate in stabilizing the binding of Mn^{2+} and/or allow the binding of other cations such as Ca^{2+}.

As previously mentioned, it could also be postulated that the LC2 loop, whose *N*-terminal end lies directly above the cation-binding site on the model (Fig. 6C), plays an important role in the cation-binding mechanism and/or its regulation. Indeed, on the one hand, MD simulation of TMEM165 in phospholipid bilayers of different thicknesses highlights the potential influence of LC2 on the second consensus motif. On the other hand, as depicted in Supplementary Fig. 10, the solvent-accessible part of the LC2 loop (aa residues 181–227) is highly charged, mostly with negatively-charged aa residues (9 Glu, 3 Asp), with a cluster of Lys and Arg residues in the middle of the long helix, making this loop prone to potential electrostatic interactions with other molecules/proteins that could regulate the function of TMEM165. This may be the case of SPCA1, for example, with which we demonstrated a functional link with TMEM165 and its proximity with TMEM165 in the Golgi of Hailey-Hailey patient fibroblasts [48,49]. Moreover, it should be noted on the TMEM165 model that while the positions of most aa residues of LC2 are not conserved during evolution, three of them (E^{189}, E^{192} and Q^{194}) are highly conserved and actually directly overhang the potential cation-binding site (Supplementary Fig. 10). Even more interestingly, the charged groups of both E^{189} and E^{192} are at distances from that of K^{112}, that are compatible with the formation of salt bridges (4.5 and 5.5 Å, respectively). It is therefore possible that the LC2 loop could interfere with the cation-binding site through interaction with the first consensus motif. Lastly, it would be possible that, owing to its high content in negatively-charged aa residues, the LC2 would bind cytosolic Mn^{2+}/Ca^{2+} and serve as a sensor modulating the transport activity of TMEM165. Both E^{189} and E^{192}, whose distance between their carboxyl group (about 5 Å in the model) is compatible with divalent cation binding, could be one of these binding sites. In strong support to this is the finding that within the NiCo family proteins in prokaryotes [47], which export Ni^{2+} and/or Co^{2+}, the corresponding positions of E^{189} and E^{192} are occupied by highly-conserved His residues. This is the case, for example, of H^{153} and H^{157} in *E. coli* RNCA protein (UniProtKB ID A0A2S8JXG6) [50]. Indeed, His residues are well known to participate to Ni^{2+} and/or Co^{2+} binding [51]. Furthermore, the *N*-terminal end of RNCA LC2, equivalent to aa residues 181–187 in human TMEM165, forms a large loop that is rich in His residues, also very prone to serve as a Ni^{2+}/Co^{2+}-binding region (data not shown).

Lastly, the predicted TMEM165 model may illustrate what could be the "inward open" conformation of TMEM165 with the cation-binding site facing the cytosol, assuming that an "outward open" conformation (facing the Golgi lumen) may also exist. This is generally the case for secondary transporters using an alternating access mechanism: either the "rocker-switch", "rocking-bundle" or "elevator" mechanism [52]. While the "rocker-switch model" requires structurally similar bundles forming the binding site, the two other models require structurally dissimilar bundles, as seems to be the case for TMEM165 and close homologs. Examples of rocking-bundle models from the large monovalent cation proton antiporter (CPA) superfamily [53] are NhaA from *E. coli* [54] and its structural homolog in *Neisseria meningitidis*, the bile acid sodium symporter ASBT*nm* [55]. They both possess two distinct and asymmetric domains delimiting a cavity: a core domain and a dimerization domain, formed by 6 and 4 TMDs in ASBT*nm*, respectively [55]. In the "elevator" mechanism, the substrate is transported by only one of the bundles whereas the other one is fixed, most likely due to dimerization or oligomerization. This third mechanism has been suggested from modeling of both oxidized outward-facing and reduced inward-facing states of *T. thermophilus* CcdA [44]. Although dimerization of TMEM165 has never been neither experimentally demonstrated to date nor convincingly predicted by AlphaFold-multimer [56] (unpublished personal results), similar rocking-bundle or elevator models cannot be excluded, with TMD 1, 2, 4 and 5 acting as the core/transport domain, and TMD 3 and 6 as the dimerization domain. In any case, it is interesting to mention that AF2 structure predictions of several members of the LysE superfamily, including members of CaCA2 in prokaryotes, such as the MneA homolog in *Candidatus Altiarchaeales* archaeon (UniProtKB ID A0A256XC95), reveal models where the metal-binding site faces the extracellular side of the phospholipid bilayer (data not shown). Interestingly, AF2 generates a model for PML3/BICAT3, a close TMEM165 homolog, that could also depict the outward open conformation. Indeed, the fifth-ranked model of Δ77-PML3/BICAT3 (pLDDT = 80.3 *vs* 86.0 for the first-ranked model) shows a protein where TMD 1, 2, 4 and 5 are downwardly tilted causing a translation of the metal-binding site of about 12 Å toward the lumen side of the phospholipid bilayer (Supplementary Fig. 11). Of course, this inward/outward open conformation model remains highly speculative and will require further experimental studies to be confirmed.

4. Conclusion

AF2 and further model refinement using MD with lipids and water have proven to be highly powerful and valuable tools for predicting the 3D structure of TMEM165 within a phospholipid bilayer and, by extension, that of homologous proteins of the CaCA2 family whose structural data hitherto only relied on poor membrane topology predictions. Firstly, our predictive methods provide much clearer insights on the impact of known CDG patient mutations on its transporter function, which for some of them had remained completely unexplained till now on a structural/functional point of view. Secondly, they unveil important structural features that offer interesting research paths allowing understanding the precise mechanism of transport of TMEM165, a key player in the homeostasis of Mn^{2+} within the Golgi, itself crucial for the functioning of Mn^{2+}-dependent enzymes involved in glycosylation.

Funding

This work was supported by grants from the Agence Nationale de la Recherche (ENIGMncA project, ANR-21-CE14–0049–01; DAGENTA, ANR-21-CE20-0038). One part of this work was also funded by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement 774078 (Pharma-Factory).

CRediT authorship contribution statement

DL: Conceptualization, Supervision, Methodology, Visualization, Writing – Original, Writing – review & editing. **MH**: Methodology, Investigation, Writing – review & editing. **ZD**: Methodology, Investigation, Writing – review & editing. **GB**: Software, Resources, Methodology, Writing – review & editing. **MB**: Funding acquisition, Writing – review & editing. **MFL:** Methodology, Investigation, Writing – review & editing. **FF**: Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: FOULQUIER reports financial support was provided by French National Research Agency.

Acknowledgements

We are grateful to the CNRS for supporting the collaboration between the University of Lille, UGSF CNRS UMR 8576, and the University of Rouen Normandie, Glyco-MEV lab. We thank Pr. Vladimir Lupashin and Dr. Leslie Climer for sharing the TMEM165 KO HEK cells. We also acknowledge the support of the PLBS UAR 2014 - US 41, BICeL platform, for the use of the Leica LSM780.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.06.015.

References

[1] Foulquier F, Amyere M, Jaeken J, Zeevaert R, Schollen E, Race V, et al. TMEM165 deficiency causes a congenital disorder of glycosylation. Am J Hum Genet 2012;91:15–26. https://doi.org/10.1016/j.ajhg.2012.05.002
[2] Zeevaert R, de Zegher F, Sturiale L, Garozzo D, Smet M, Moens M, et al. Bone dysplasia as a key feature in three patients with a novel congenital DIsorder of Glycosylation (CDG) type II due to a deep intronic splice mutation in TMEM165. JIMD Rep 2013;8:145–52. https://doi.org/10.1007/8904_2012_172
[3] Bammens R, Mehta N, Race V, Foulquier F, Jaeken J, Tiemeyer M, et al. Abnormal cartilage development and altered N-glycosylation in Tmem165-deficient zebrafish mirrors the phenotypes associated with TMEM165-CDG. Glycobiology 2015;25:669–82. https://doi.org/10.1093/glycob/cwv009
[4] Morelle W, Potelle S, Witters P, Wong S, Climer L, Lupashin V, et al. Galactose Supplementation in Patients With TMEM165-CDG Rescues the Glycosylation Defects. J Clin Endocrinol Metab 2017;102:1375–86. https://doi.org/10.1210/jc.2016-3443
[5] Khan S, Sbeity M, Foulquier F, Barré L, Ouzzine M. TMEM165 a new player in proteoglycan synthesis: loss of TMEM165 impairs elongation of chondroitin- and heparan-sulfate glycosaminoglycan chains of proteoglycans and triggers early chondrocyte differentiation and hypertrophy. Cell Death Dis 2021;13:1–14. https://doi.org/10.1038/s41419-021-04458-1
[6] Rosnoblet C, Legrand D, Demaegd D, Hacine-Gherbi H, de Bettignies G, Bammens R, et al. Impact of disease-causing mutations on TMEM165 subcellular localization, a recently identified protein involved in CDG-II. Hum Mol Genet 2013;22:2914–28. https://doi.org/10.1093/hmg/ddt146
[7] Lebredonchel E, Houdou M, Potelle S, de Bettignies G, Schulz C, Krzewinski Recchi M-A, et al. Dissection of TMEM165 function in Golgi glycosylation and its Mn2+ sensitivity. Biochimie 2019;165:123–30. https://doi.org/10.1016/j.biochi.2019.07.016
[8] Potelle S, Morelle W, Dulary E, Duvet S, Vicogne D, Spriet C, et al. Glycosylation abnormalities in Gdt1p/TMEM165 deficient cells result from a defect in Golgi manganese homeostasis. Hum Mol Genet 2016;25:1489–500. https://doi.org/10.1093/hmg/ddw026
[9] Snyder NA, Stefan CP, Soroudi CT, Kim A, Evangelista C, Cunningham KW. H + and Pi byproducts of glycosylation affect Ca 2+ homeostasis and are retrieved from the Golgi complex by homologs of TMEM165 and XPR1. G3 Genes Genomes Genet 2017;7:3913–24. https://doi.org/10.1534/g3.117.300339
[10] Stribny J, Thines L, Deschamps A, Goffin P, Morsomme P. The human Golgi protein TMEM165 transports calcium and manganese in yeast and bacterial cells. J Biol Chem 2020;295:3865–74. https://doi.org/10.1074/jbc.RA119.012249
[11] Potelle S, Dulary E, Climer L, Duvet S, Morelle W, Vicogne D, et al. Manganese-induced turnover of TMEM165. Biochem J 2017;474:1481–93. https://doi.org/10.1042/BCJ20160910
[12] Dulary E, Potelle S, Legrand D, Foulquier F. TMEM165 deficiencies in congenital disorders of glycosylation type II (CDG-II): clues and evidences for roles of the protein in Golgi functions and ion homeostasis. Tissue Cell 2017;49:150–6. https://doi.org/10.1016/j.tice.2016.06.006
[13] Thines L, Stribny J, Morsomme P. From the uncharacterized protein family 0016 to the GDT1 family: molecular insights into a newly-characterized family of cation secondary transporters. Microb Cell 2020;7:202–14. https://doi.org/10.15698/mic2020.08.725
[14] Foulquier F, Legrand D. Biometals and glycosylation in humans: congenital disorders of glycosylation shed lights into the crucial role of Golgi manganese homeostasis. Biochim Biophys Acta Gen Subj 2020;1864:129674. https://doi.org/10.1016/j.bbagen.2020.129674
[15] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2
[16] Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 2022;50:D439–44. https://doi.org/10.1093/nar/gkab1061
[17] Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, et al. A structural biology community assessment of AlphaFold2 applications. Nat Struct Mol Biol 2022;29:1056–67. https://doi.org/10.1038/s41594-022-00849-w
[18] Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Res 2012;40:D370–6. https://doi.org/10.1093/nar/gkr703
[19] Lomize AL, Todd SC, Pogozheva ID. Spatial arrangement of proteins in planar and curved membranes by PPM 3.0. Protein Sci 2022;31:209–20. https://doi.org/10.1002/pro.4219
[20] Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, et al. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 2003;19:163–4. https://doi.org/10.1093/bioinformatics/19.1.163
[21] Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, et al. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res 2005;33:W299–302. https://doi.org/10.1093/nar/gki370
[22] Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res 2010;38:W529–33. https://doi.org/10.1093/nar/gkq399
[23] Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 2016;44:W344–50. https://doi.org/10.1093/nar/gkw408
[24] Lensink MF, Govaerts C, Ruysschaert J-M. Identification of specific lipid-binding sites in integral membrane proteins. J Biol Chem 2010;285:10519–26. https://doi.org/10.1074/jbc.M109.068890
[25] Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–90. https://doi.org/10.1063/1.448118
[26] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995;103:8577–93. https://doi.org/10.1063/1.470117
[27] Miyamoto S, Kollman PA. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 1992;13:952–62. https://doi.org/10.1002/jcc.540130805
[28] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463–72. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H
[29] Fisher CR, Wyckoff EE, Peng ED, Payne SM. Identification and characterization of a putative manganese export protein in Vibrio cholerae. J Bacteriol 2016;198:2810–7. https://doi.org/10.1128/JB.00215-16
[30] Gandini C, Schmidt SB, Husted S, Schneider A, Leister D. The transporter SynPAM71 is located in the plasma membrane and thylakoids, and mediates manganese tolerance in Synechocystis PCC6803. New Phytol 2017;215:256–68. https://doi.org/10.1111/nph.14526
[31] Brandenburg F, Schoffman H, Kurz S, Krämer U, Keren N, Weber APM, et al. The synechocystis manganese exporter Mnx is essential for manganese homeostasis in cyanobacteria. Plant Physiol 2017;173:1798–810. https://doi.org/10.1104/pp.16.01895
[32] Hoecker N, Honke A, Frey K, Leister D, Schneider A. Homologous proteins of the manganese transporter PAM71 are localized in the Golgi apparatus and endoplasmic reticulum. Plants 2020:9. https://doi.org/10.3390/plants9020239
[33] Permyakov SE, Khokhlova TI, Uversky VN, Permyakov EA. Analysis of Ca2+/Mg2+ selectivity in alpha-lactalbumin and Ca(2+)-binding lysozyme reveals a distinct Mg(2+)-specific site in lysozyme. Proteins 2010;78:2609–24. https://doi.org/10.1002/prot.22776
[34] Dudev T, Lim C. Principles governing Mg, Ca, and Zn binding and selectivity in proteins. Chem Rev 2003;103:773–88. https://doi.org/10.1021/cr020467n
[35] Christianson DW, Cox JD. Catalysis by metal-activated hydroxide in zinc and manganese metalloenzymes. Annu Rev Biochem 1999;68:33–57. https://doi.org/10.1146/annurev.biochem.68.1.33

[36] Schulte Althoff S, Grüneberg M, Reunert J, Park JH, Rust S, Mühlhausen C, et al. TMEM165 deficiency: postnatal changes in glycosylation. JIMD Rep 2015;26:21–9. https://doi.org/10.1007/8904_2015_455
[37] Chakrabarti P. Conformations of arginine and lysine side chains in association with anions. Int J Pept Protein Res 1994;43:284–91. https://doi.org/10.1111/j.1399-3011.1994.tb00392.x
[38] Yamasaki K, Daiho T, Yasuda S, Danko S, Kawabe J-I, Suzuki H. Electrostatic interactions between single arginine and phospholipids modulate physiological properties of sarcoplasmic reticulum Ca2+-ATPase. Sci Rep 2022;12:12200. https://doi.org/10.1038/s41598-022-16091-9
[39] Braulke T, Bonifacino JS. Sorting of lysosomal proteins. Biochim Biophys Acta 2009;1793:605–14. https://doi.org/10.1016/j.bbamcr.2008.10.016
[40] Yang C-H, Wang C, Singh S, Fan N, Liu S, Zhao L, et al. Golgi-localized manganese transporter PML3 regulates Arabidopsis growth through modulating Golgi glycosylation and cell wall biosynthesis. New Phytol 2021. https://doi.org/10.1111/nph.17209
[41] Tsu BV, Saier MH. The LysE superfamily of transport proteins involved in cell physiology and pathogenesis. PLoS One 2015;10:e0137184. https://doi.org/10.1371/journal.pone.0137184
[42] Vrljic M, Sahm H, Eggeling L. A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum. Mol Microbiol 1996;22:815–26. https://doi.org/10.1046/j.1365-2958.1996.01527.x
[43] Williamson JA, Cho S-H, Ye J, Collet J-F, Beckwith JR, Chou JJ. Structure and multistate function of the transmembrane electron transporter CcdA. Nat Struct Mol Biol 2015;22:809–14. https://doi.org/10.1038/nsmb.3099
[44] Zhou Y, Bushweller JH. Solution structure and elevator mechanism of the membrane electron transporter CcdA. Nat Struct Mol Biol 2018;25:163–9. https://doi.org/10.1038/s41594-018-0022-z
[45] Demaegd D, Colinet A-S, Deschamps A, Morsomme P. Molecular evolution of a novel family of putative calcium transporters. PLoS One 2014;9:e100851. https://doi.org/10.1371/journal.pone.0100851
[46] Li L, Vorobyov I, Allen TW. The different interactions of lysine and arginine side chains with lipid membranes. J Phys Chem B 2013;117:11906–20. https://doi.org/10.1021/jp405418y
[47] Zeinert R, Martinez E, Schmitz J, Senn K, Usman B, Anantharaman V, et al. Structure-function analysis of manganese exporter proteins across bacteria. J Biol Chem 2018. https://doi.org/10.1074/jbc.M117.790717
[48] Lebredonchel E, Houdou M, Hoffmann H-H, Kondratska K, Krzewinski M-A, Vicogne D, et al. Investigating the functional link between TMEM165 and SPCA1. Biochem J 2019;476:3281–93. https://doi.org/10.1042/BCJ20190488
[49] Roy A-S, Miskinyte S, Garat A, Hovnanian A, Krzewinski-Recchi M-A, Foulquier F. SPCA1 governs the stability of TMEM165 in Hailey-Hailey disease. Biochimie 2020;174:159–70. https://doi.org/10.1016/j.biochi.2020.04.017
[50] Rodrigue A, Effantin G, Mandrand-Berthelot M-A. Identification of rcnA (yohM), a nickel and cobalt resistance gene in Escherichia coli. J Bacteriol 2005;187:2912–6. https://doi.org/10.1128/JB.187.8.2912-2916.2005
[51] Sun X, Yu G, Xu Q, Li N, Xiao C, Yin X, et al. Putative cobalt- and nickel-binding proteins and motifs in Streptococcus pneumoniae. Metallomics 2013;5:928–35. https://doi.org/10.1039/c3mt00126a
[52] Drew D, Boudker O. Shared molecular mechanisms of membrane transporters. Annu Rev Biochem 2016;85:543–72. https://doi.org/10.1146/annurev-biochem-060815-014520
[53] Brett CL, Donowitz M, Rao R. Evolutionary origins of eukaryotic sodium/proton exchangers. Am J Physiol Cell Physiol 2005;288:C223–39. https://doi.org/10.1152/ajpcell.00360.2004
[54] Lee C, Kang HJ, von Ballmoos C, Newstead S, Uzdavinys P, Dotson DL, et al. A two-domain elevator mechanism for sodium/proton antiport. Nature 2013;501:573–7. https://doi.org/10.1038/nature12484
[55] Hu N-J, Iwata S, Cameron AD, Drew D. Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. Nature 2011;478:408–11. https://doi.org/10.1038/nature10450
[56] Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-multimer. bioRxiv 2022:2021. https://doi.org/10.1101/2021.10.04.463034
[57] Moon CP, Fleming KG. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. Proc Natl Acad Sci USA 2011;108:10174–7. https://doi.org/10.1073/pnas.1103979108