PLOS ONE

RESEARCH ARTICLE

Structural basis for matriglycan synthesis by the LARGE1 dual glycosyltransferase

Michael Katz, Ron Diskin*

Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel

* ron.diskin@weizmann.ac.il

Citation: Katz M, Diskin R (2022) Structural basis for matriglycan synthesis by the LARGE1 dual glycosyltransferase. PLoS ONE 17(12): e0278713. https://doi.org/10.1371/journal.pone.0278713

Editor: Olga Mayans, University of Liverpool, UNITED KINGDOM

Received: July 10, 2022

Accepted: November 21, 2022

Published: December 13, 2022

Data Availability Statement: Coordinate file, structure factors and EM density maps were deposited and are available under accessions codes: PDB: 7ZVJ, EMDB: EMD-14985, and EMD-14987.

Funding: "The Diskin lab is supported by research grants from the Ernst I Ascher foundation, Ben B. and Joyce E. Eisenberg Foundation, Estate of Emile Mimran, Jeanne and Joseph Nissim Center for Life Sciences Research, Dov and Ziva Rabinovich Endowed Fund for Structural Biology, Donald Rivin, Stanley and Tanya Rossby Endowment Fund, Natan Sharansky, Dr. Barry Sherman Institute for

Abstract

LARGE1 is a bifunctional glycosyltransferase responsible for generating a long linear polysaccharide termed matriglycan that links the cytoskeleton and the extracellular matrix and is required for proper muscle function. This matriglycan polymer is made with an alternating pattern of xylose and glucuronic acid monomers. Mutations in the LARGE1 gene have been shown to cause life-threatening dystroglycanopathies through the inhibition of matriglycan synthesis. Despite its major role in muscle maintenance, the structure of the LARGE1 enzyme and how it assembles in the Golgi are unknown. Here we present the structure of LARGE1, obtained by a combination of X-ray crystallography and single-particle cryo-EM. We found that LARGE1 homo-dimerizes in a configuration that is dictated by its coiled-coil stem domain. The structure shows that this enzyme has two canonical GT-A folds within each of its catalytic domains. In the context of its dimeric structure, the two types of catalytic domains are brought into close proximity from opposing monomers to allow efficient shuttling of the substrates between the two domains. Together, with putative retention of matriglycan by electrostatic interactions, this dimeric organization offers a possible mechanism for the ability of LARGE1 to synthesize long matriglycan chains. The structural information further reveals the mechanisms in which disease-causing mutations disrupt the activity of LARGE1. Collectively, these data shed light on how matriglycan is synthesized alongside the functional significance of glycosyltransferase oligomerization.

Introduction

Like-acetylglucosaminyltransferase 1 (LARGE1) is a bifunctional glycosyltransferase (GTase) [1] that plays a critical role in maintaining proper muscle function [2]. This enzyme is responsible for a post-translational modification on the extracellular matrix receptor, α-dystroglycan (α-DG), which enables it to interact with proteins containing a laminin-G domain [1, 2]. This carbohydrate addition to α-DG is critical for the communication between the F-actin cytoskeleton network and the basal lamina [3, 4]. The polysaccharide which LARGE1 synthesizes, termed matriglycan, is composed of xylose (Xyl) and glucuronic acid (GlcA) subunits in a linear repeating structure of $[\text{-3GlcA-}\beta\text{1,3-Xyl-}\alpha\text{1-}]_n$ [5, 6]. Matriglycan presents multiple binding sites for interacting with laminin G (LG)-like domains, leading to both high avidity and

Medicinal Chemistry, as well as from the Israel Science Foundation (grants No. 3147/19 and 209/20). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript".

Competing interests: The authors have declared that no competing interests exist.

specificity. Binding between the LG-domain and matriglycan is facilitated by the chelation of a Ca^{2+} ion bound to the hydroxyl groups of the GlcA & Xyl subunits [7].

Matriglycan is thought to be a unique modification of α-DG [5] and the initiation of its synthesis depends on a complex network of at least 16 other enzymes [8–11]. Initiation of matriglycan synthesis is dependent on Protein O-Mannose Kinase (POMK) phosphorylation of the mannose of a core M3 O-glycan, GalNAcβ(1,3)-GlcNAcβ(1,4)-Man-O-Thr [10, 12, 13]. Further studies showed that the M3 glycan is elongated with ribitol 5-phosphate (Rbo5P) [14–16], after which the priming enzymes, RXYLT1 [17] and B4GAT1 [18, 19], add GlcAβ(1–4)Xyl, which can then be recognized by LARGE1 [20]. The length of α-DG-conjugated matriglycan chains varies substantially and is highly dependent on the tissue where it is produced [4, 5]. Chain elongation of matriglycan is terminated by 3-O-sulfation on a GlcA subunit by the sulfotransferase, HNK-1ST [21].

Mutations in the genes responsible for the synthesis of these glycan precursors can result in hypoglycosylation of α-DG and disease-causing phenotypes [22]. Improper glycosylation of α-DG is responsible for various forms of congenital muscular dystrophy (CMD), collectively termed dystroglycanopathies, which yield skeletal, brain, and eye abnormalities [23, 24]. Mutations specifically within the *LARGE1* gene can cause severe forms of secondary dystroglycanopathies [25, 26]. Walker-Warburg syndrome (WWS) is considered to be the most serious form of dystroglycanopathies and is characterized by significant brain abnormalities, inability to develop significant motor function, and typically death within infancy [23, 24]. Several point mutations within LARGE1 have been known to result in life-threatening phenotypes [24, 26–28]. *LARGE1* deficiency is also believed to be an underlying cause of epithelium-derived [29] & lung [30] cancers.

LARGE1 is present in the Golgi [31] as a type II membrane protein, composed of a transmembrane/cytosolic region, a predicted coiled-coil/stem (CC), and domains with xylosyltransferase (Xyl-T) and glucuronyltransferase (GlcA-T) catalytic activities, which are separated by a linker [1, 32] (Fig 1a). Most of the currently available structural data for GTases were derived using X-ray crystallography. However, the organization of the GTases solved within a crystal lattice could make it difficult to determine their true native oligomeric states. Structural data on bifunctional GTases, like LARGE1, is scarce [33–36]. While the presence of the CC domain in LARGE1 suggests that it forms higher-order oligomers, it is unknown how LARGE1 would oligomerize. Additionally, how the two domains orchestrate alternating synthesis of this polysaccharide is unclear. GTases have frequently been observed as homo-oligomers and the oligomeric state can be required for the enzymatic function [37–40]. It was previously suggested that the transmembrane/stem region of a GTase is critical for their oligomerization [41, 42] but it is still unclear as to which other factors facilitate oligomer formation.

Here, we investigated the structure of LARGE1 using a combination of X-ray crystallography and single-particle cryo-electron microscopy (cryo-EM). We found that in solution, the catalytic domains of LARGE1 can dimerize in two distinct configurations. We further show that the CC domain promotes dimerization and further selects one of these configurations, which places the active sites of one Xyl-T and one GlcA-T in close proximity that may promote efficient synthesis of matriglycan.

Results

The ectodomain of LARGE1 forms a dimer in solution

To characterize LARGE1, we produced in HEK293F cells two soluble, affinity-tagged, secreted forms of this protein. Both forms include the entire catalytic ectodomain module without the transmembrane part of the protein, but one contains the CC domain, and the other does not

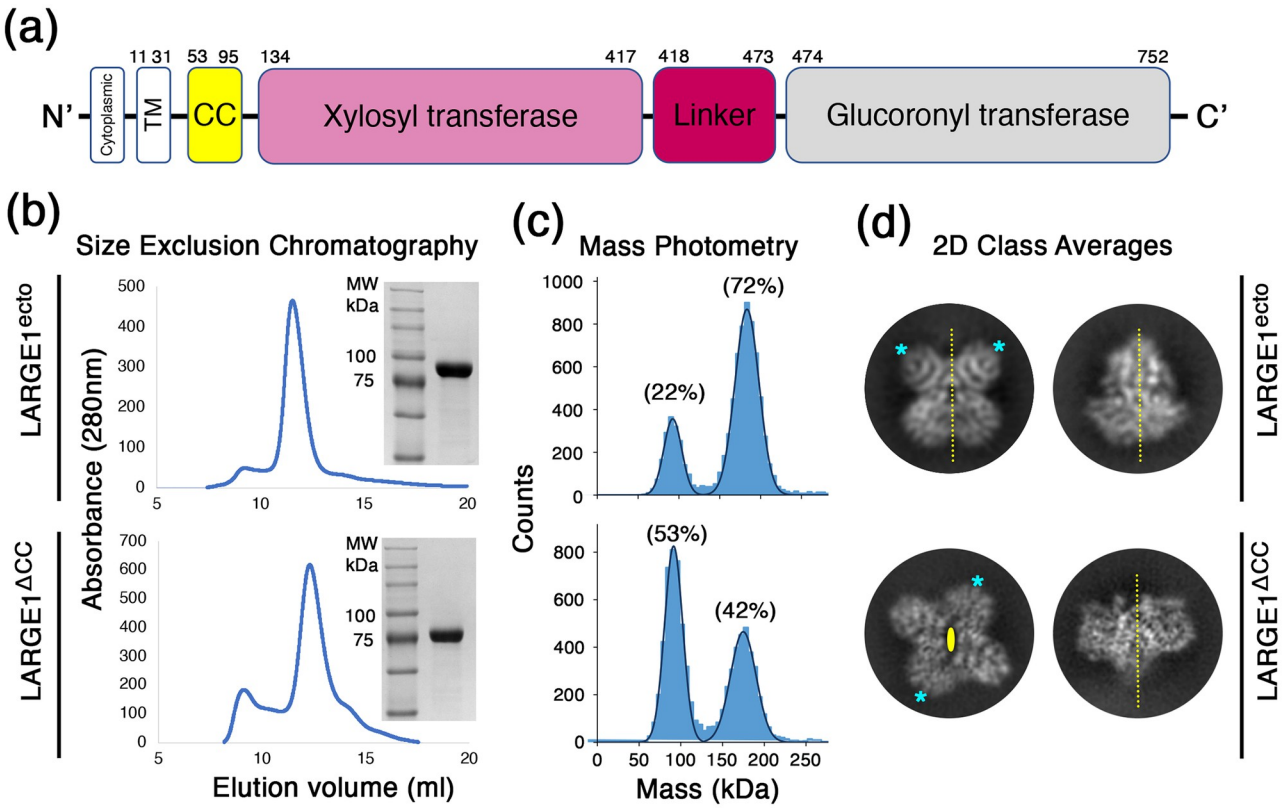

Fig 1. LARGE1 dimerizes in solution. (a) A schematic diagram for the domain organization in the LARGE1 protein. (b) Purification of the recombinant LARGE1 proteins. Chromatograms from SEC for constructs with ($LARGE1^{ecto}$, upper image) or without ($LARGE1^{\Delta CC}$, lower image) the CC domain. Insets show Coomassie-stained SDS-PAGE for the purified proteins. (c) Mass-photometry analysis of the two LARGE1 proteins at a concentration of 30 nM. The fraction of counts is noted for the two main peaks in the data sets that corresponds to the monomeric (~80 kDa) and dimeric (~160 kDa) forms of LARGE1. The rest of the counts (6% and 5%, for with and without CC, respectively) do not belong to these two main peaks. These are representative measurements that were repeated multiple times using different protein batches. (d) 2D class averages of LARGE1 particles from the two forms. For each LARGE1 form, two separate classes are shown, representing two distinct and perpendicular views. The 2-fold symmetry axes are shown as dashed yellow lines (for in-plane axes) or as an oval shape (for out-of-plane axis). Same-type domains are marked with asterisks.

https://doi.org/10.1371/journal.pone.0278713.g001

(will be referred to as $LARGE1^{ecto}$, and $LARGE1^{\Delta CC}$, respectively). Both forms were readily expressed and were easily captured from cells' media using Ni^2+ affinity chromatography. Subsequent purification of these two proteins using size exclusion chromatography (SEC) resulted in nearly-homogenous protein samples (Fig 1b). For both protein samples, the SEC elution profiles were similar, with a single main peak that contained the LARGE1 protein (Fig 1b). Analyzing both protein samples using mass photometry at a concentration of 30 nM (2.5 μg/ml) indicated the existence of a monomer/dimer equilibrium (Fig 1c). At this concentration, $LARGE1^{ecto}$ had a higher propensity to form a dimer compared with $LARGE1^{\Delta CC}$ (Fig 1c). The CC-mediated higher propensity for dimer formation suggests that this region is indeed forming a coiled-coil as was previously predicted [32]. Noteworthy, this analysis indicates that even in the absence of the CC domain, the catalytic module of LARGE1 is sufficient to mediate dimerization.

While both LARGE1 constructs dimerize in solution, single-particle cryo-EM analysis indicated that they do so in two different configurations: 2D class averages indicate that in the presence of the CC domain, the dimer forms with a parallel orientation in which each of the two catalytic domains (Fig 1a) interacts with the same domain of the second protomer (Fig

1d). In contrast, LARGE1$^{\Delta CC}$ forms an anti-parallel dimer in which each of the catalytic domains interacts with the other domain of the second protomer (Fig 1d). Therefore, the ectodomain of LARGE1 has an inherent propensity to dimerize and the presence of the CC domain favors a parallel orientation for dimer formation.

The dimeric structure of LARGE1

Single-particle cryo-EM analysis of the two LARGE1 constructs yielded interpretable density maps (Fig 2). Both dimers exhibit a clear 2-fold symmetry, which was applied during reconstruction. The LARGE1^{ecto} and LARGE1$^{\Delta CC}$ dimers provided density maps extending to gold-standard Fourier shell correlation (FSC) resolution values of 3.9 Å, and 3.7 Å, respectively. While obtaining the single-particle EM data, we also managed to crystallize the LARGE1$^{\Delta CC}$ and to collect X-ray diffraction data, extending to 2.6 Å (Table 1). We, therefore, utilized one-half of the EM density map (Fig 2) as a search model in Phaser [43] to obtain phases for the X-ray diffraction data, following a previously described method [44]. The crystal of LARGE1 belonged to a *C* 2 space group, and we found two copies of LARGE1 in the asymmetric unit.

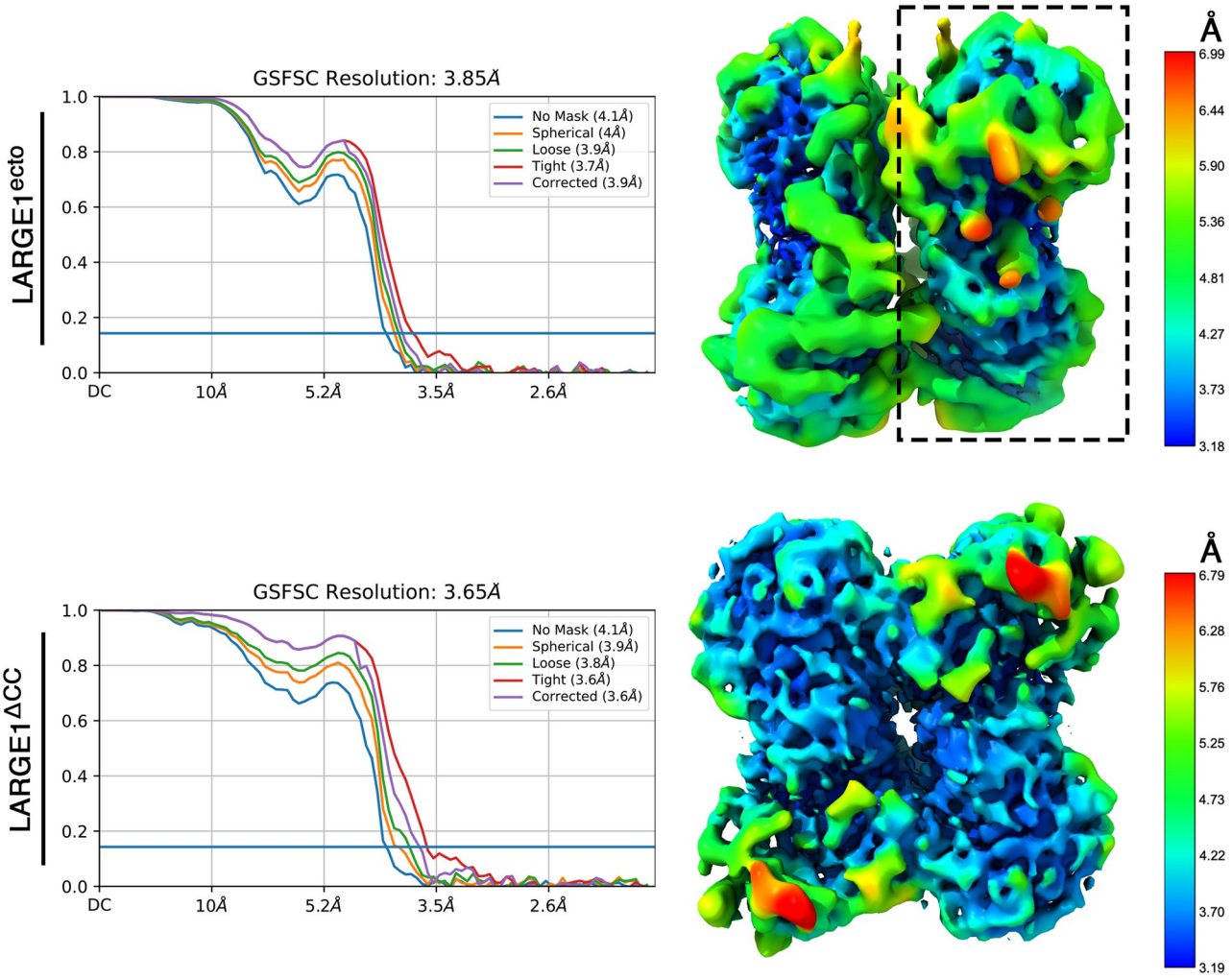

Fig 2. Single particle cryo-EM maps of LARGE1. Reconstructed maps of the LARGE1^{ecto} (upper part) and LARGE1$^{\Delta CC}$ (lower part). The gold standard FSC curves are shown on the left and density maps colored by the local resolution estimates are shown on the right. One half map that was used as a molecular replacement model is highlighted with a dashed rectangle.

https://doi.org/10.1371/journal.pone.0278713.g002

Table 1. Data collection and refinement statistics.

	LARGE1$^{\Delta CC}$
Wavelength	0.9677
Resolution range	45.64–2.61 (2.69–2.61)
Space group	*C* 2
Unit cell	196.53 107.38 100.33 90 120.26 90
Total reflections	140110
Unique reflections	52821
Multiplicity	2.7
Completeness (%)	96.0 (93.0)
Mean I/sigma(I)	4.5 (0.4)
R-meas	18.1 (322.5)
$CC_{1/2}$	98.3 (14.9)
Reflections used in refinement	52411
Reflections used for R-free	1194
R-work	0.215
R-free	0.270
Number of atoms	
macromolecules	9685
ligands	45
solvent	44
RMSD (bonds)	0.008
RMSD (angles)	1.22
Ramachandran favored (%)	97.03
Ramachandran allowed (%)	2.88
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	3.15
Clashscore	11.97
Average B-factor	
macromolecules	83.99
ligands	110.11
solvent	63.87

Statistics for the highest-resolution shell are shown in parentheses.

https://doi.org/10.1371/journal.pone.0278713.t001

We used Phenix-Autobuild [45] to provide an initial model and then manually completed the model using an iterative model building and refinement in Coot [46] and Phenix-Refine [47], respectively.

The crystal structure of LARGE1 consists of two chains that make a dimer in the asymmetric unit (Fig 3a). For both chains (i.e., A and B), electron density was visible and hence allowed us to model residues 134 to 752. The electron density for some loops was either poor or was lacking completely and hence the model lacks residues: 282–286, 362–367, 369–372, 382–391, 626–636, 718–719 for chain A, and 280–286, 361–367, 370–372, 383–391, 713–718 for chain B. The two chains adopt an almost identical configuration with a root mean square deviation (RMSD) value of 1.4 Å for all shared Cα atoms (Fig 3a). Despite the fact that we have crystallized the LARGE1$^{\Delta CC}$ construct, the two chains in the asymmetric unit formed a parallel dimer that resembles the EM map of the LARGE1^{ecto}. Indeed, the crystallographic dimer readily fits into the EM density map of the LARGE1^{ecto} (Fig 3b). Noteworthy, the crystal packing is such that two neighboring LARGE1 dimers have their N-termini facing each other in a way

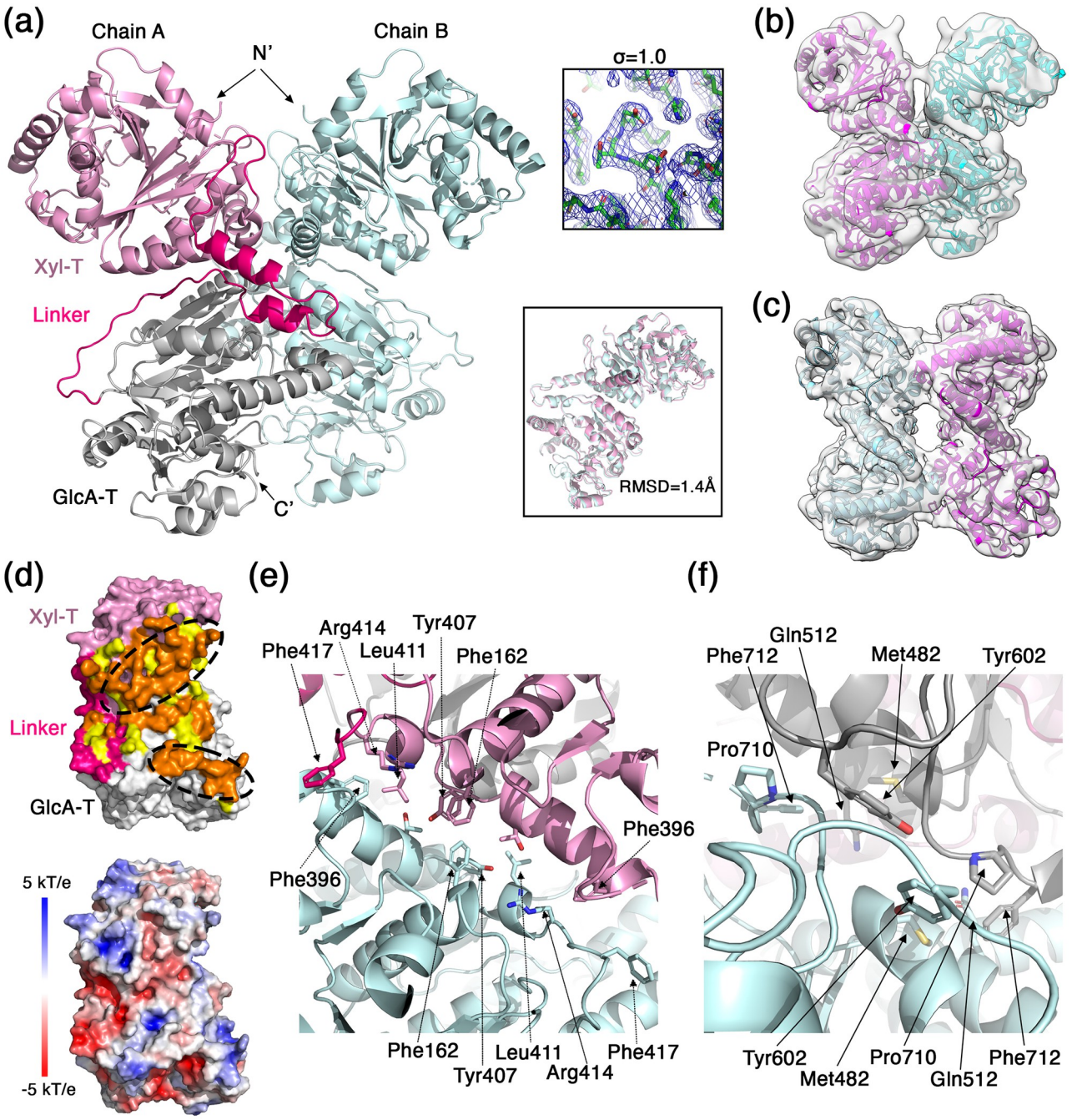

Fig 3. The dimeric structure of LARGE1. (a) Ribbon representation of the two LARGE1 monomers that make the asymmetric unit of the crystal. The Xyl-T, Linker region, and GlcA-T regions are highlighted in pink, hot pink, and gray on one monomer, respectively. The N- and C- termini are noted. Upper inset shows a 2Fo-Fc electron density map of the Xyl-T active site at the indicated sigma level. Lower inset shows a superimposition of the two modeled LARGE1 chains. (b) A fit of the LARGE1 dimer as appeared in the asymmetric unit of the crystal into the EM density map of the LARGE1^{ecto}. (c) An individual fit of two LARGE1 monomers into the EM density map of the LARGE1$^{\Delta CC}$. (d) Analysis of the dimer interface. Upper image shows a LARGE1 monomer in a surface representation. The Xyl-T, Linker region, and GlcA-T regions are highlighted in pink, hot pink, and gray, respectively. Fully buried surface residues due to dimerization are shown in orange (probe size of 1.4 Å). Partially buried surface residues (i.e., making cavities) due to dimerization are shown in yellow (probe size of 2.0 Å). Two patches of buried surfaces on the Xyl-T and GlcA-T domains are marked with dashed oval lines. Lower image shows a LARGE1 monomer in the same orientation as the upper image with surface representation that is color-coded by surface electrostatic potential (blue, 5 kT/e and red, -5 kT/e). (e) Detailed view of the Xyl-T/XylT interface that makes the buried patch. (f) Detailed view of the GlcA-T/GlcA-T interface that makes the buried patch.

https://doi.org/10.1371/journal.pone.0278713.g003

that does not leave enough room for accommodating two CC domains. Hence, LARGE1^{ecto} will not be able to crystallize with a similar crystal packing. In order to fit into the EM map of the LARGE1$^{\Delta CC}$, the chains need to be rotated in respect to one another to form an anti-parallel dimer (Fig 3c). Considering the high similarity of chains, A and B of the crystallographic model (Fig 3a) and since a rotation of the LARGE1 monomers is sufficient to get a good fit into the EM map of the LARGE1$^{\Delta CC}$, it indicates that the relative orientation of the Xyl-T domain in respect to the GlcA-T domain is maintained in all of these states. In addition, the fact that the LARGE1$^{\Delta CC}$ crystallized in the parallel dimeric form as observed for the LARGE1^{ecto} (Fig 3b) implies that it also forms such parallel dimers in solution despite the fact that the EM analysis only revealed anti-parallel dimers (Figs 1d, 2 and 3c). Hence, in equilibrium, and in the absence of the CC domain, the anti-parallel dimeric form is energetically preferred and thus dominates. The parallel dimers do exist, but at a lower abundance that is not apparent in our EM data. Likely due to packing considerations, these low-abundant parallel dimers were the ones that first nucleated and subsequently dictated the organization of the LARGE1 crystals. Noteworthy, the LARGE1 molecules that were incorporated into the solid crystal phase are no longer at equilibrium with the soluble LARGE1 molecules, which according to Le Chatelier's principle, shifts the equilibrium toward the formation of this parallel dimer form.

The selection of the parallel dimeric form by the presence of the CC domain (Figs 1d and 3b) indicates that this may be the preferred physiological dimer *in-vivo*. We hence focus our analysis on this dimeric form. Inspection of the parallel dimer interface shows an overall buried surface area of 2,589 Å^2 (1,350 Å^2 on chain 'A' and 1239 Å^2 on chain 'B') (Fig 3d). However, if we consider additional residues at the interface, which make very narrow cavities (i.e., using a probe with a radius of 2.0 Å for detecting buried surfaces), the buried surface area then becomes 4,443 Å^2. This buried surface area spans the Xyl-T and GlcA-T catalytic domains as well as some residues at the linker region between the domains (Fig 3d). In this dimer form, more surface is buried at the interface of the two Xyl-T domains than between the GlcA-T domains, and in both cases, most of the buried residues make continuous patches (Fig 3d). Examining the surface electrostatic potential shows that the dimer interface is mostly uncharged (Fig 3d). Various residues contribute to the buried, hydrophobic patches on the Xyl-T (Fig 3e) and on the GlcA-T (Fig 3f) domains. Besides the hydrophobic patches, polar residues exist at various locations along the dimeric interface. These residues may participate in water-mediated hydrogen-bond networks that may also contribute for dimerization. Having fairly flat hydrophobic surfaces on both catalytic domains may help to rationalize the ability of LARGE1 to form an anti-parallel dimer when the CC domain is not dictating a parallel configuration.

The Xyl-T and GlcA-T catalytic domains

The Xyl-T domain of LARGE1, which is classified by the Carbohydrate Active Enzymes (CAZy) database (http://www.cazy.org/) as a GT family 8 retaining transferase [48], adopts a canonical GT-A fold. This fold is composed of 7-stranded major & 3-stranded minor β-sheets, encompassed by α-helices (Fig 4a). In its active site, the Xyl-T domain presents a conserved DXD motif (Asp242-X-Asp244) (Fig 4a), which is a motif that mediates binding of nucleoside-bound sugars through a manganese ion and is important for the activity of GTases from many distinct enzyme families [49]. Structures of other xylosyltransferases have been solved. Comparing the Xyl-T domain with the structure of a mouse-derived xyloside α-1,3-xylosyltransferase (XXYLT1) [50], which is an enzyme unrelated to LARGE1, reveals a similar general architecture (Fig 4a). In the active site of the Xyl-T domain, we could not detect a clear density that would indicate the presence of a divalent cation. Nevertheless, the superimposition of

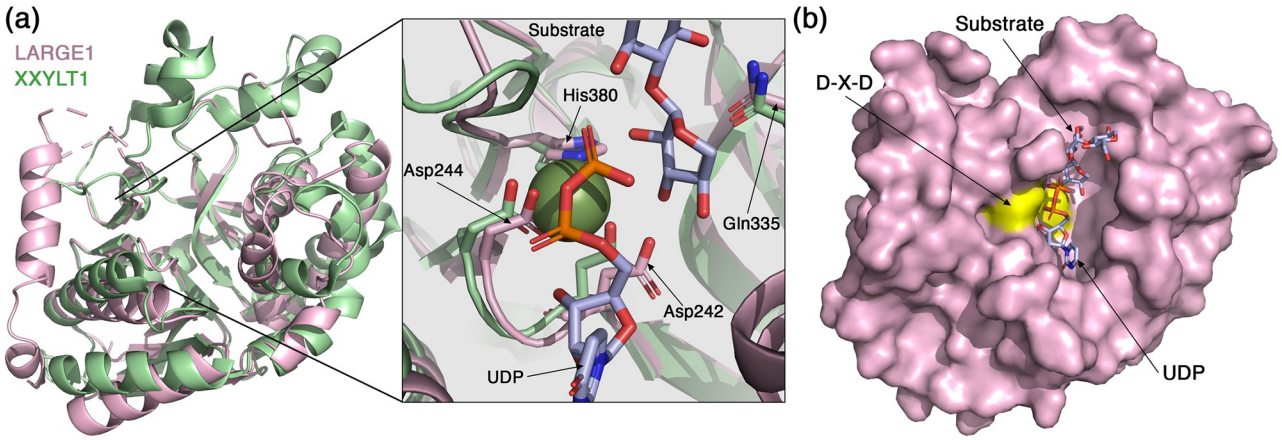

Fig 4. The Xyl-T domain of LARGE1. (a) Superimposition of the Xyl-T domain from LARGE1 (pink) with XXYLT1 (PDB: 4wn2, green). Inset shows a closeup view of the active site. The conserved residues that coordinate the Mn^{2+} are noted, as well as a UDP and a growing substrate that were resolved in the structure of XXYLT1. (b) Surface representation of the LARGE1 Xyl-T domain. The aspartic acids of the DXD motif are colored yellow. The UDP and the growing substrate derived from the structure of XXYLT1 are shown as sticks.

https://doi.org/10.1371/journal.pone.0278713.g004

these two structures places the XXYLT1-derived Mn^{2+} such that it would be coordinated by the DXD-residues Asp242 and Asp244, and by His380 of LARGE1 (Fig 4a). These three residues are conserved in XXYLT1 as well as in other xylosyltransferases [50, 51]. The Mn^{2+} helps to position the UDP-xylose in the active site [50]. Based on the XXYLT1 structure, we can postulate the position of the UDP-sugar donor in the active site of the LARGE1 Xyl-T (Fig 4a and 4b). Gln335 of LARGE1 is also a conserved residue (Fig 4a), and based on the enzymatic mechanism of XXYLT1 [50], it likely binds the substrate and also functions in stabilizing high-energy intermediate states, making it a critical residue for catalysis. Overall, the superimposition with XXYLT1 elucidates how the UDP-xylose and growing substrates would fit into the catalytic site of the LARGE1 Xyl-T (Fig 4b).

The GlcA-T domain of LARGE1 also adopts a canonical GT-A fold and has a conserved DXD motif (Asp563-X-Asp565) (Fig 5a). It is classified by the CAZy database as a GT family 49 inverting transferase [48]. The closest structural homolog of the LARGE1 GlcA-T according to 3D-BLAST [52] is the chondroitin polymerase from K4 *E.coli* (K4CP) (PDB:2z86) [53]. Similar to LARGE1, K4CP is a bifunctional glycosyltransferase that catalyzes the addition of GlcA and *N*-acetylgalactosamine to the chondroitin polymer [53], but unlike GlcA-T, it belongs to GT family 2 according to the CAZy database [48]. In the electron density map, we observed density for a coordinated ion in the active site of the GlcA-T domain (Fig 5b) and we therefore modeled a Mn^{2+} atom, which is an essential metal for the activity of LARGE1 [54]. Compared with K4CP, the Mn^{2+} atoms in the GlcA-T domain are slightly shifted away from the coordinating aspartic acids and histidine residues (Fig 5a, inset). The exact position of Mn^{2+} may change when UDP-GlcA is bound in the active site as this shifted location seems to collide with the UDP as seen in the K4CP structure. The LARGE1 GlcA-T domain has an extra histidine (His708) that seems to participate in the binding of the Mn^{2+}. This histidine residue is not conserved in K4CP which has a proline (Pro389) instead (Fig 5a, inset). As in the Xyl-T domain, the Mn^{2+} in the GlcA-T domain is important for holding the UDP-sugar donor in place. A surface representation of the GlcA-T domain reveals the approximate location of UDP in the active site (Fig 5c). The K4CP-derived UDP fits into a deep pocket, leaving a narrow tunnel for the growing matriglycan substrate to enter (Fig 5c).

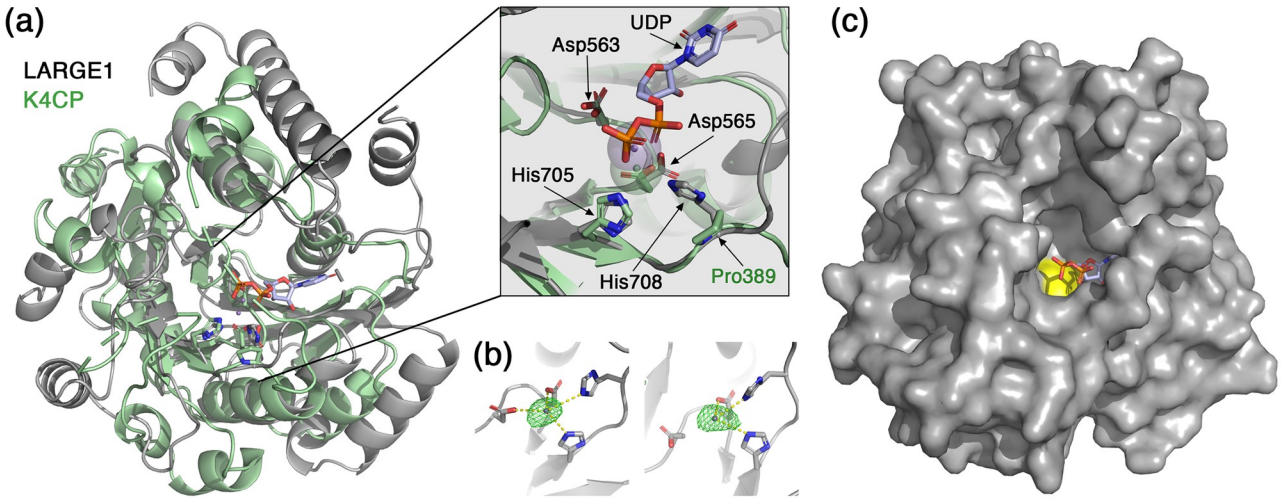

Fig 5. The GlcA-T domain of LARGE1. (a) Superimposition of the GlcA-T domain from LARGE1 (grey) with K4CP (PDB: 2z86, green). Inset shows a closeup view of the active site. Mn^{2+} atoms of LARGE1 (purple) and of K4CP (green) are shown as semi-transparent spheres. A UDP from the K4CP structure is shown in purple. The conserved DXD motif and histidine residues that participate in the coordination of the Mn^{2+} are noted. (b) The GlcA-T active sites of the two LARGE1 monomers (left and right) in the crystallographic asymmetric unit are shown with an Fo-Fc difference map (green mesh, σ = 5) calculated after omitting the Mn^{2+} atoms. (c) Surface representation of the LARGE1 GlcA-T domain. The aspartic acids of the DXD motif are colored yellow. The UDP derived from the structure of K4CP is shown as sticks.

https://doi.org/10.1371/journal.pone.0278713.g005

Matriglycan synthesis by the dimeric LARGE1

The matriglycan polymer that LARGE1 synthesizes is made of alternating Xyl and GlcA monomers. Hence, the growing matriglycan chain needs to enter into the active sites of the Xyl-T and the GlcA-T domains in an alternating manner. Interestingly, the closest Xyl-T and GlcA-T sites in the context of the LARGE1 dimer are located on opposite monomers rather than on a single protomer (Fig 6). These two sites are approximately 40 Å away from each other and make the shortest possible path for a matriglycan non-reducing end to encounter the two different transferases. The alternative path between the two catalytic sites of a single protomer, requires the matriglycan chain to travel around the dimer to reach its far side. Hence the most diffusion-efficient synthesis of matriglycan will make use of the Xyl-T and the GlcA-T domains of separate monomers in the context of the dimer.

Disease-causing mutations in LARGE1

Several alterations in LARGE1 were previously identified as disease-causing mutations. Such mutations include: S331F [27], C443Y [24], W495R [28], and E509K [26]. Having the structural information for LARGE1 allows us to gain some insights for the possible molecular mechanisms that underly the deleterious effect of these mutations. Ser331 is a surface-exposed residue in the Xyl-T catalytic domain (Fig 7a). Introducing phenylalanine in this position does not have an obvious consequence on the structure of LARGE1. Nevertheless, Ser331 is located at the vicinity of some surface-exposed hydrophobic residues like Trp276 and Leu332 (Fig 7a). Potentially, introducing a S331F mutation may form a hydrophobic patch that would drive aggregation of LARGE1. Cys443 is located within the linker between the Xyl-T and GlcA-T domains (Fig 7a). It makes a disulfide bond with a cysteine residue in the GlcA-T domain (Fig 7a). A tyrosine residue instead of a cysteine in position 443 will abrogate the disulfide bridge and will likely also prevent proper folding due to steric clashes with its bulky side chain. Trp495 is located in the GlcA-T domain and is making an integral part of the hydrophobic core (Fig

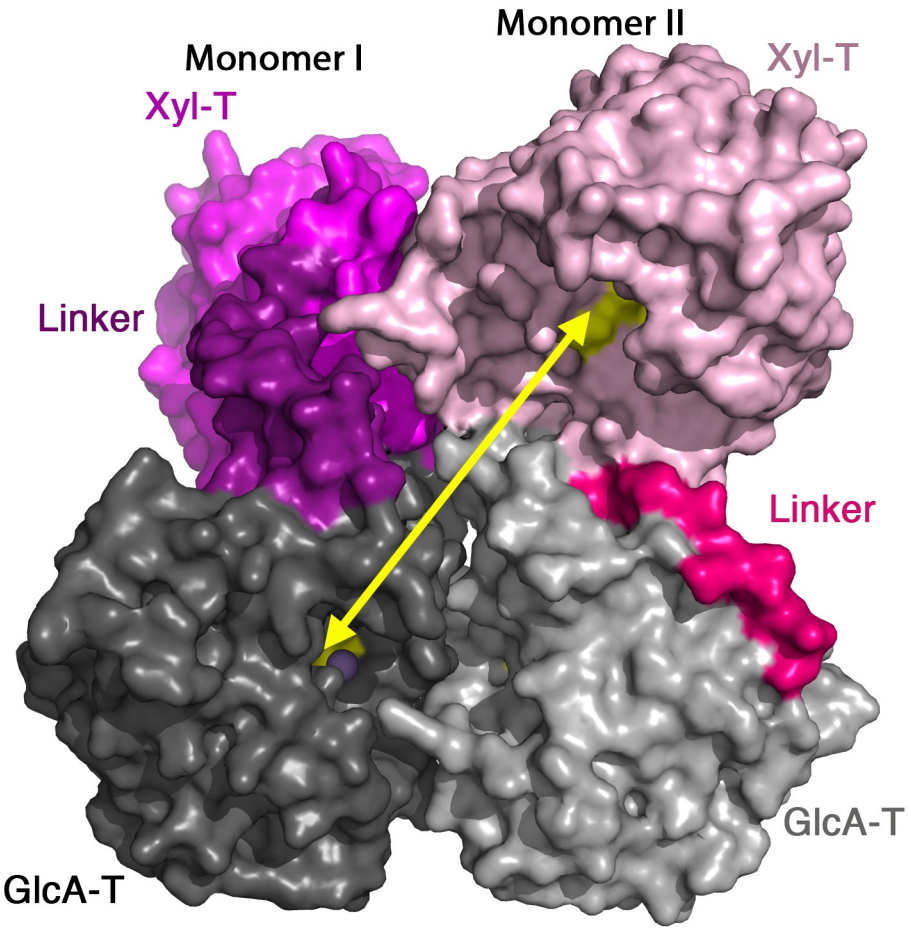

Fig 6. Matriglycan synthesis by the LARGE1 dimer. The closest active sites of the Xyl-T and GlcA-T domains are located of different monomers. The LARGE1 dimer is shown using a surface representation and the catalytic domains are colored as in Fig 3a. The two LARGE1 monomers are highlighted by different tones. A yellow arrow shows the openings of the closest Xyl-T and GlcA-T catalytic sites.

https://doi.org/10.1371/journal.pone.0278713.g006

7a). A positively-charged arginine facing straight into the hydrophobic core of the protein is likely to be highly destabilizing for the folded state of the protein, which will effectively prevent its proper folding. Glu509 is located within the dimer interface of the GlcA-T domains near its 2-fold symmetry axis (Fig 7a). Glu509 is making long-range electrostatic interactions with the nearby Lys536 (Fig 7a), which likely balance the overall surface electrostatic potential at this position. Mutating Glu509 to a lysine will likely cause the accumulation of a local positive surface electrostatic potential, which will make dimer formation to be less favorable.

From the abovementioned structural analysis, it seems like the underlying molecular mechanisms of these four mutations involve the disruption of the global structure of LARGE1 rather than a direct interference at the catalytic sites. Indeed, after individually incorporating each of the four point mutations on the secreted form of LARGE1^{ecto}, we could no longer detect secreted LARGE1 in the cells' media (Fig 7b & 7c). Interestingly, all of these mutated LARGE1 variants do express in the cells as could be seen in the total cell lysates (Fig 7b & 7c). Hence, these proteins are produced but fail to secrete from the cells. This failure in secretion is likely due to the quality control system of the endoplasmic reticulum [55], which supports the notion of a global effect of these mutations on the stability of LARGE1. Interestingly, the clinical

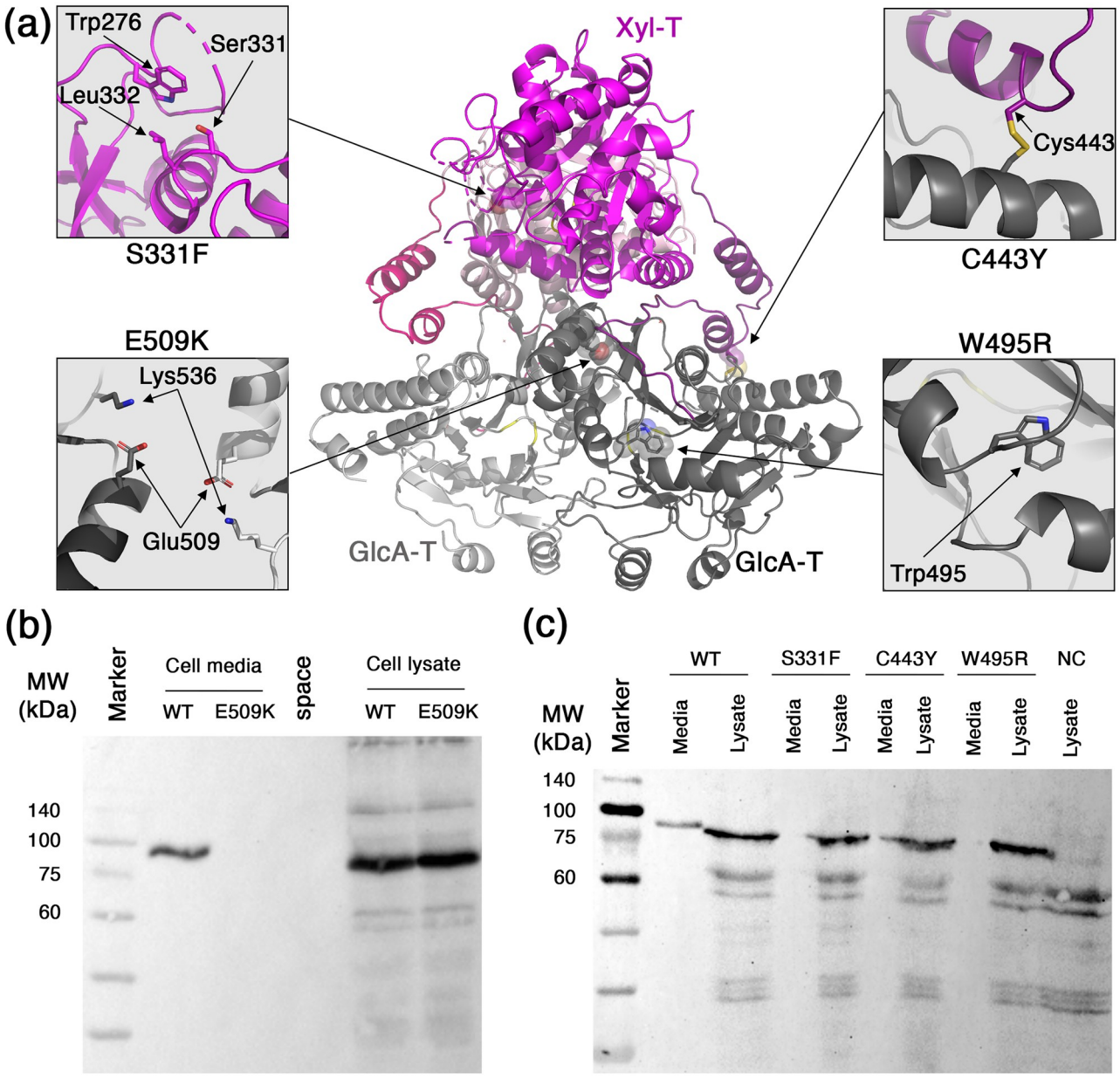

Fig 7. Disease causing mutations in LARGE1. (a) Mapping of four mutations on the structure of LARGE1. Insets show closeup views of the native residues at the mutation sites. (b) Western blot analysis using anti-His antibody showing the expression of soluble LARGE1-His WT or E509K mutant. Samples include the total cell lysates and cells' media as indicated. (c) Similar to 'b', a western blot analysis showing the cells' media and lysates from cell transfected with LARGE1-WT, -S331F, -C443Y, or -W495R. Lysate from non-transfected cells is shown as a negative control (NC).

https://doi.org/10.1371/journal.pone.0278713.g007

manifestations of these mutations are not identical [24, 26–28]. Hence, while the biochemical end result seems to be the same for all of these mutations, subtle changes do exist, which warrant further investigations.

Discussion

LARGE1 has a high propensity to form dimers. Mass-photometry measurements at a LARGE1 concentration of 30 nM indicated the presence of significant amounts of both the monomer

and of the dimer forms in equilibrium (Fig 1c). At higher concentrations, as were used for preparing the EM grids for example (*i.e.*, ~4 μM), the predominant form was already dimeric as no 2D class averages were obtained for a monomeric LARGE1. It is also clear that the CC domain promotes dimerization of LARGE1 (Fig 1c) and further selects a parallel organization for the dimer (Figs 1d, 2 & 3b). Taken together, under physiological conditions, in cells, LARGE1 is likely in the form of a parallel dimer, but future *in-vivo* studies will need to corroborate this notion. Interestingly, this parallel dimer form of LARGE1 includes homo-association of the Xyl-T and the GlcA-T domains (Fig 3a).

Homo-association of other Xyl-Ts and GlcA-Ts, which are not related to LARGE1, has been previously observed [38, 50, 56–59]. Additionally, homo-association has been shown to be important for the enzymatic activity of some GTases. For example, the generation of a monomeric form of human lysyl hydrolase 3 (LH3), which is a multifunctional GTase, results in deterioration of enzymatic activity [60]. Destabilization of the homodimeric interface area by disease causing mutations in O-linked beta-N-acetylglucosamine transferase (OGT) was found to decrease its catalytic activity [61]. Additionally, disruption of homodimerization for the dual glycosyltransferase 1 (dGT1) also lead to a significantly decrease in its GlcNAc-T activity [36]. As postulated below, the association of LARGE1 into a dimer may also be important for its catalytic activity.

The assembly of LARGE1 into a dimer brings into a close proximity the active sites of Xyl-T and GlcA-T from opposite monomers (Fig 6). This proximity between the two active sites increases the probability that the growing matriglycan chain will be sequentially modified by the two domains before the matriglycan polymer dissociates from LARGE1. While preparing this manuscript for submission, a complementary investigation of LARGE1 was deposited in a preprint server [62]. This work also supports the notion that active sites from two opposite chains contribute for the synthesis of matriglycan, which likely represent the most efficient way that LARGE1 can generate matriglycan.

Interestingly, the matriglycan product can potentially consist of hundreds of repeating Xyl-GlcA units [5]. Since HNK-1ST terminates matriglycan synthesis [21], the production of long matriglycan chains suggests that encounters of HNK1-ST with the growing matriglycan are not frequent, and perhaps even restricted, during the elongation process. Delaying capping of matriglycan could potentially be achieved if the growing matriglycan chain will not be completely released from LARGE1 after the addition of each sugar monomer. Due to the high content of GlcA, matriglycan has a negative charge. Examining the surface electrostatic potential of the LARGE1 dimer, we see a positively-charged patch at the interface of the two GlcA-T domains and a significantly larger positively-charged surface formed at the interface of the Xyl-T domains (Fig 8). We therefore speculate that unspecific electrostatic interactions may form between these positively-charged surfaces of LARGE1 and matriglycan, which may delay the release of the polymer during synthesis. By that, the effective local concentration of matriglycan will increase, which will promote synthesis of long matriglycan chains.

Several point mutations within the *LARGE1* gene were linked to pathological conditions. The E509K missense mutation, for instance, results in a significant loss of matriglycan generation, along with α-DG:laminin binding, and thus causes a serious form of α-dystroglycanopathy, termed MDC1D [26, 63]. The E509K LARGE1 mutant was shown to localize in the ER and does not progress to the Golgi [63]. Our observations corroborate this notion as the soluble His-tagged LARGE1^{ecto} that carries the E509K mutation is also produced in cells but is not secreted (Fig 7b), indicating a failure to progress in the secretory pathways. Our structural analysis suggests that the E509K mutation may abrogate dimerization of LARGE1 by localizing positive charge at the symmetry axis of the dimer interface, which will disfavor dimerization

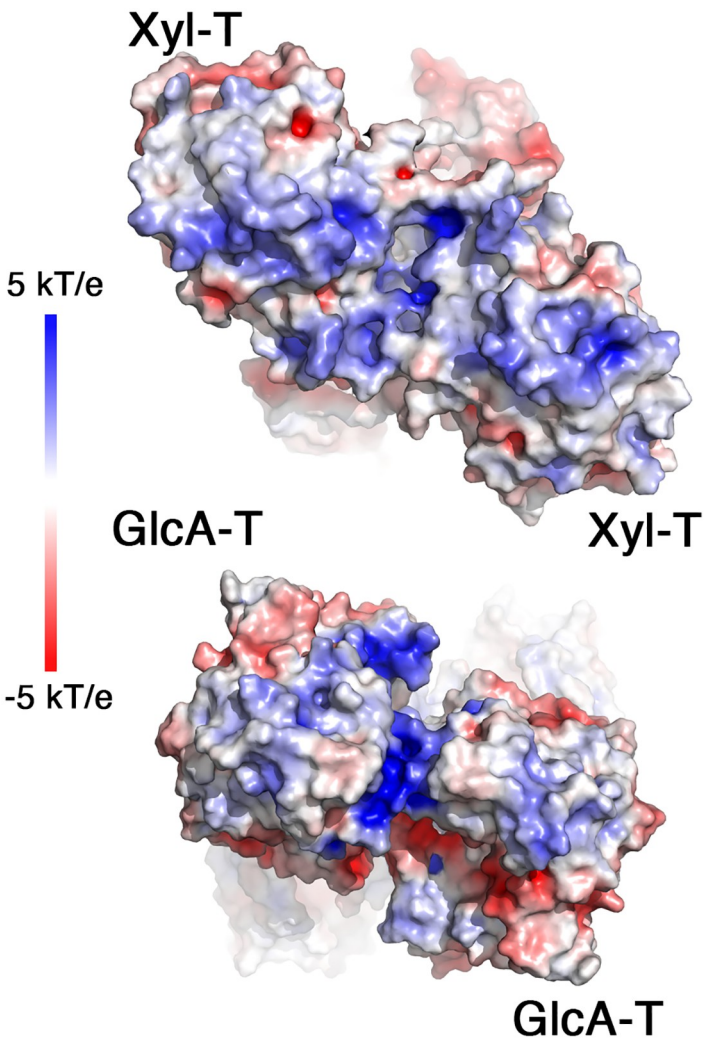

Fig 8. Putative electrostatic interaction of LARGE1 with matriglycan. Surface electrostatic potentials of the LARGE1 dimer from the Xyl-T (top) and from the GlcA-T (bottom) sides of the dimer. Blue to red colors indicate 5 kT/e to -5 kT/e, respectively.

https://doi.org/10.1371/journal.pone.0278713.g008

due to electrostatic repulsion (Fig 7a). Dimerization of LARGE1 seems to be required for progressing through the ER/Golgi network. Interestingly, oligomerization has been shown to be a prerequisite for Golgi translocation for other membrane-bound proteins. The GTases, EXT1 & EXT2 that function as a heterodimer for example, are retained in the ER when expressed alone. Only upon co-expression of both of these GTases can the heterodimer form and reach the Golgi [64, 65]. Other membrane proteins, such as prenylin, are believed to require homodimerization to enter the Golgi [66], and a failure to multimerize retains various other proteins in the ER [67–70]. Taken together, we propose that dimerization of LARGE1 is essential for its proper function in cells.

Overall, we have deciphered the structure of the LARGE1 bi-functional glycosyltransferase. Dictated by the CC/stem domain, a parallel dimer of LARGE1 is formed and is required for proper localization in the cells. Dimerization enables advantageous properties for LARGE1 that likely contribute to its ability to synthesize long chains of matriglycan.

Materials & methods

Enzyme expression & purification

Two variants of LARGE1 were generated, both from the human LARGE1 (xylosyl- and glucuronyltransferase) gene, obtained from the Forchheimer plasmid bank (Weizmann Institute of Science, Clone ID 100000367). Both variants were cloned without the transmembrane and cytosolic tail domains into a modified pHLSEC plasmid. The cloning, expression, and purification for this enzyme have been previously described [71]. Briefly, the LARGE1 gene with its coiled-coil (LARGE1^{ecto}) domain (residues 29–756), with a His tag + GSGG linker at its C-terminal end, was cloned into *pHLsec* using BglII/NotI restriction sites. The same LARGE1 gene without its CC domain (LARGE1$^{\Delta CC}$) (residues 92–756), with a 6x His tag + GSGG linker at its N-terminal end was cloned in the same manner. Both plasmids were then separately transfected into HEK293F cells at a density of approximately 1 x 10^6 million cells/ml. The transfections occurred with 40 kDa poylethylenamine at a ratio of 1:2.5 DNA: PEI (PEI Max, PolySciences). The productions were harvested after 1 week, spun down at 600 xg (the cells were then discarded), and the supernatants was then centrifuged at 15,800 xg to remove cellular debris. The supernatant of each production was then passed through a 0.45 μm Stericup filter (Merck Millipore) and supplemented with 100 μM phenylmethyl sulfonyl fluoride (PMSF) and 0.02% sodium azide. After which, the productions were buffer exchanged to TBS (150 mM NaCl, 20 mM Tris, pH 8.0) using a Pellicon Tangential Flow Filtration system (Merck Millipore) and purified by their 6x His tag by a 5 ml HiTrap IMAC FF Ni^{2+} column (GE Healthcare). Elution from the column was then carried out with 10% imidiazole, and the proteins were concentrated using a 4 ml 30 kDa Amicon (Merck Millipore) and run on a Supderdex 200 10/300 GL column (GE Healthcare) with a running buffer of TBS + 0.2% sodium azide. The aliquots from the main HPLC peak were then pooled, concentrated using another 4 ml 30 kDa Amicon (Merck Millipore), and finally flash frozen and stored at -80˚C.

Cryo-electron microscopy data collection & analysis

LARGE1 variants (0.3 mg/ml) were dispensed onto R 0.6/1 Cu/C Quantifoil grids (EMS) and plunged into liquid ethane using a Vitrobot system (Thermo Fisher/FEI; 4˚C, 100% humidity) and then stored under cryogenic conditions. Plasma treatment for the grids was carried out at 90 seconds, 15 mA, and plunging used 3.5 μl of samples with a blot force of -1 for 3 seconds. Grids were first screened on the Talos Arctica microscope (200 kV; Thermo Scientific) and then transferred to the Titan Krios 3Gi (300 kV; FEI) with a Gatan K3 direct detection camera for data collection. The beam size used was 900 nm and the magnification was 120,000 X, with a defocus range of -1.6 to -0.6 μm and a pixel size of 0.52 Å.

Data processing was carried out using the cryoSPARC v2 software. For both LARGE1 variants, patch motion correction & CTF estimation were first completed. For the LARGE1$^{\Delta CC}$, a total of 7,969 movies were initially obtained and 5,451 movies were then accepted for particle selection using the blob picker. Initially, 1,760,546 particles were chosen. We then extracted the particles using a 400-pixel box, which was Fourier-cropped to a 200-pixel box size with a pixel size of 1.04 Å. After multiple iterations of the 2D classifications, 92,093 particles were finally obtained which belonged to a single class. To improve the GSFCS resolution, we applied a twofold symmetry. The 3D reconstruction had a final GSFSC resolution of 3.65 Å. For LARGE1^{ecto}, the same data processing method was used. A total of 5,250 movies were initially obtained, and 4,896 movies were used for further processing. After patch motion correction & CTF estimation, the blob picker in cryoSPARC v2 selected 223,954 particles. Using twofold

symmetry, 164,336 particles were chosen, all of which related to a single class. A density map with a final GSFSC resolution of 3.86 Å was then obtained.

Crystallization, X-ray data collection, and structure determination

The initial crystallization hit occurred with the SaltRx screen (Hampton Research) using the Mosquito robot (LLP Labtech) in a 96-well plate (LLP). Optimization was carried out with the Dragonfly robot (LLP), where LARGE1- CC crystals were obtained in a 200 nl drop, at a 1:1 ratio (protein:crystallization reagents), containing 1.0 M Na_3PO_4 monobasic monohydrate, K_3PO_4 dibasic / pH 5.0. PEG 200 (20% final concentration) was added to the crystals as a cryo-preservative before storing in liquid N_2 for transport. Data collection was carried out at the European Synchrotron Research Facility (ESRF), with Beamline ID ID30a. A Dectris Eiger 4M detector was used. Diffraction data was successfully obtained using a wavelength of 0.9677 Å to a resolution of 2.6 Å. Data was indexed using XDS [72] and subsequently merged and scaled using Aimless [73]. Half of the EM density map, corresponding to a single LARGE1 monomer, was used for solving the X-ray structure via molecular replacement with Phaser [43], as part of the Phenix suit [47]. An initial model was obtained using the Phenix-Autobuild function [45]. The model was completed by iterative rounds of manual model building using Coot [46] and refinement using Phenix-Refine (NCS & secondary-structure restrains, individual b-factors, TLS, real-space and reciprocal-space coordinate refinement) [47].

Mass photometry

LARGE1 samples were analyzed on a Refeyn One^{MP} (Refeyne) mass photometer. The calibration curve was carried out using jack bean urease (100 μg/ml) (Sigma). LARGE1 samples were diluted in TBS to approximately 30 nM (2.5 μg/ml) and were analyzed at room temperature over 60 second recordings using the Acquire^{MP} software, at 100 frames per second (fps). Analysis occurred using Refeyne's Discover^{MP} software.

Mutagenesis

Mutants were produced with the LARGE1^{ecto} construct in a modified pHLsec plasmid. Point mutations were carried out via the QuikChange Site-Directed Mutagenesis protocol (Aligent). After sequence verification, mutants were amplified (via MaxiPrep) and transfected into HEK293F cells.

Western blot analysis

Cell lysates, collected 4 days-post transfections, were boiled in 4x sample buffer containing 40% glycerol, 240 mM Tris-HCl, pH 6.8, 8% SDS, 0.04% Bromophenol blue, 5% beta-mercaptoethanol. Samples were run on a 10% acrylamide, Tris-glycine gel. Western Blot analysis was then carried out using an anti-His antibody (Qiagen) at a 1:2,000 dilution, in Tris-Buffer Saline + Tween (20 mM Tris, pH 8.0, 150 mM NaCl, 0.5% Tween) containing 1% BSA, incubated overnight at 4˚C. Then, the membrane was incubated with HRP-conjugated anti-mouse secondary antibody (Jackson ImmunoResearch), diluted at 1:10,000, before being developed with EZ-ECl (Biological Industries).

Supporting information

S1 Raw images.
(PDF)

Acknowledgments

We thank Moshe Goldsmith in the Department of Biomolecular Sciences at the Weizmann Institute of Science and Adar Sonn-Segev from Refeyn Ltd. for assisting us to perform the mass photometry measurements.

Accession numbers

Coordinate file, structure factors and EM density maps were deposited and are available under accessions codes: PDB: 7ZVJ, EMDB: EMD-14985, and EMD-14987.

Author Contributions

Conceptualization: Ron Diskin.

Formal analysis: Michael Katz.

Investigation: Michael Katz, Ron Diskin.

Supervision: Ron Diskin.

Writing – original draft: Michael Katz, Ron Diskin.

Writing – review & editing: Michael Katz, Ron Diskin.

References

1. Inamori K, Yoshida-Moriguchi T, Hara Y, Anderson ME, Yu L, Campbell KP. Dystroglycan function requires xylosyl- and glucuronyltransferase activities of LARGE. Science. 2012; 335(6064):93–6. Epub 2012/01/10. https://doi.org/10.1126/science.1214115 PMID: 22223806
2. Kanagawa M, Saito F, Kunz S, Yoshida-Moriguchi T, Barresi R, Kobayashi YM, et al. Molecular recognition by LARGE is essential for expression of functional dystroglycan. Cell. 2004; 117(7):953–64. Epub 2004/06/24. https://doi.org/10.1016/j.cell.2004.06.003 PMID: 15210115.
3. Adams JC, Brancaccio A. The evolution of the dystroglycan complex, a major mediator of muscle integrity. Biol Open. 2015; 4(9):1163–79. Epub 2015/09/01. https://doi.org/10.1242/bio.012468 PMID: 26319583
4. Ibraghimov-Beskrovnaya O, Ervasti JM, Leveille CJ, Slaughter CA, Sernett SW, Campbell KP. Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. Nature. 1992; 355(6362):696–702. Epub 1992/02/20. https://doi.org/10.1038/355696a0 PMID: 1741056.
5. Yoshida-Moriguchi T, Campbell KP. Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. Glycobiology. 2015; 25(7):702–13. Epub 2015/04/18. https://doi.org/10.1093/glycob/cwv021 PMID: 25882296
6. Goddeeris MM, Wu B, Venzke D, Yoshida-Moriguchi T, Saito F, Matsumura K, et al. LARGE glycans on dystroglycan function as a tunable matrix scaffold to prevent dystrophy. Nature. 2013; 503(7474):136–40. Epub 2013/10/18. https://doi.org/10.1038/nature12605 PMID: 24132234
7. Briggs DC, Yoshida-Moriguchi T, Zheng T, Venzke D, Anderson ME, Strazzulli A, et al. Structural basis of laminin binding to the LARGE glycans on dystroglycan. Nat Chem Biol. 2016; 12(10):810–4. Epub 2016/08/16. https://doi.org/10.1038/nchembio.2146 PMID: 27526028
8. Hara Y, Kanagawa M, Kunz S, Yoshida-Moriguchi T, Satz JS, Kobayashi YM, et al. Like-acetylglucosaminyltransferase (LARGE)-dependent modification of dystroglycan at Thr-317/319 is required for laminin binding and arenavirus infection. Proc Natl Acad Sci U S A. 2011; 108(42):17426–31. Epub 2011/10/12. https://doi.org/10.1073/pnas.1114836108 PMID: 21987822
9. Barresi R, Michele DE, Kanagawa M, Harper HA, Dovico SA, Satz JS, et al. LARGE can functionally bypass alpha-dystroglycan glycosylation defects in distinct congenital muscular dystrophies. Nat Med. 2004; 10(7):696–703. Epub 2004/06/09. https://doi.org/10.1038/nm1059 PMID: 15184894.
10. Walimbe AS, Okuma H, Joseph S, Yang T, Yonekawa T, Hord JM, et al. POMK regulates dystroglycan function via LARGE1-mediated elongation of matriglycan. Elife. 2020; 9. Epub 2020/09/26. https://doi.org/10.7554/eLife.61388 PMID: 32975514
11. Johnson K, Bertoli M, Phillips L, Topf A, Van den Bergh P, Vissing J, et al. Detection of variants in dystroglycanopathy-associated genes through the application of targeted whole-exome sequencing

analysis to a large cohort of patients with unexplained limb-girdle muscle weakness. Skelet Muscle. 2018; 8(1):23. Epub 2018/08/01. https://doi.org/10.1186/s13395-018-0170-1 PMID: 30060766

12. Yoshida-Moriguchi T, Yu L, Stalnaker SH, Davis S, Kunz S, Madson M, et al. O-mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding. Science. 2010; 327(5961):88–92. Epub 2010/01/02. https://doi.org/10.1126/science.1180512 PMID: 20044576

13. Sheikh MO, Halmo SM, Wells L. Recent advancements in understanding mammalian O-mannosylation. Glycobiology. 2017; 27(9):806–19. Epub 2017/08/16. https://doi.org/10.1093/glycob/cwx062 PMID: 28810660

14. Praissman JL, Willer T, Sheikh MO, Toi A, Chitayat D, Lin YY, et al. The functional O-mannose glycan on alpha-dystroglycan contains a phospho-ribitol primed for matriglycan addition. Elife. 2016; 5. Epub 2016/05/01. https://doi.org/10.7554/eLife.14473 PMID: 27130732

15. Kanagawa M, Kobayashi K, Tajiri M, Manya H, Kuga A, Yamaguchi Y, et al. Identification of a Post-translational Modification with Ribitol-Phosphate and Its Defect in Muscular Dystrophy. Cell Rep. 2016; 14(9):2209–23. Epub 2016/03/01. https://doi.org/10.1016/j.celrep.2016.02.017 PMID: 26923585.

16. Gerin I, Ury B, Breloy I, Bouchet-Seraphin C, Bolsee J, Halbout M, et al. ISPD produces CDP-ribitol used by FKTN and FKRP to transfer ribitol phosphate onto alpha-dystroglycan. Nat Commun. 2016; 7:11534. Epub 2016/05/20. https://doi.org/10.1038/ncomms11534 PMID: 27194101

17. Manya H, Yamaguchi Y, Kanagawa M, Kobayashi K, Tajiri M, Akasaka-Manya K, et al. The Muscular Dystrophy Gene TMEM5 Encodes a Ribitol beta1,4-Xylosyltransferase Required for the Functional Glycosylation of Dystroglycan. The Journal of biological chemistry. 2016; 291(47):24618–27. Epub 2016/10/14. https://doi.org/10.1074/jbc.M116.751917 PMID: 27733679

18. Willer T, Inamori K, Venzke D, Harvey C, Morgensen G, Hara Y, et al. The glucuronyltransferase B4GAT1 is required for initiation of LARGE-mediated alpha-dystroglycan functional glycosylation. Elife. 2014; 3. Epub 2014/10/04. https://doi.org/10.7554/eLife.03941 PMID: 25279699

19. Praissman JL, Live DH, Wang S, Ramiah A, Chinoy ZS, Boons GJ, et al. B4GAT1 is the priming enzyme for the LARGE-dependent functional glycosylation of alpha-dystroglycan. Elife. 2014; 3. Epub 2014/10/04. https://doi.org/10.7554/eLife.03943 PMID: 25279697

20. Sheikh MO, Capicciotti CJ, Liu L, Praissman J, Ding D, Mead DG, et al. Cell surface glycan engineering reveals that matriglycan alone can recapitulate dystroglycan binding and function. Nat Commun. 2022; 13(1):3617. Epub 2022/06/25. https://doi.org/10.1038/s41467-022-31205-7 PMID: 35750689.

21. Sheikh MO, Venzke D, Anderson ME, Yoshida-Moriguchi T, Glushka JN, Nairn AV, et al. HNK-1 sulfotransferase modulates alpha-dystroglycan glycosylation by 3-O-sulfation of glucuronic acid on matriglycan. Glycobiology. 2020; 30(10):817–29. Epub 2020/03/10. https://doi.org/10.1093/glycob/cwaa024 PMID: 32149355

22. Nickolls AR, Bonnemann CG. The roles of dystroglycan in the nervous system: insights from animal models of muscular dystrophy. Dis Model Mech. 2018; 11(12). Epub 2018/12/24. https://doi.org/10.1242/dmm.035931 PMID: 30578246

23. Taniguchi-Ikeda M, Morioka I, Iijima K, Toda T. Mechanistic aspects of the formation of alpha-dystroglycan and therapeutic research for the treatment of alpha-dystroglycanopathy: A review. Mol Aspects Med. 2016; 51:115–24. Epub 2016/07/18. https://doi.org/10.1016/j.mam.2016.07.003 PMID: 27421908.

24. Meilleur KG, Zukosky K, Medne L, Fequiere P, Powell-Hamilton N, Winder TL, et al. Clinical, pathologic, and mutational spectrum of dystroglycanopathy caused by LARGE mutations. J Neuropathol Exp Neurol. 2014; 73(5):425–41. Epub 2014/04/09. https://doi.org/10.1097/NEN.0000000000000065 PMID: 24709677

25. Endo T. Mammalian O-mannosyl glycans: Biochemistry and glycopathology. Proc Jpn Acad Ser B Phys Biol Sci. 2019; 95(1):39–51. Epub 2019/01/16. https://doi.org/10.2183/pjab.95.004 PMID: 30643095

26. Longman C, Brockington M, Torelli S, Jimenez-Mallebrera C, Kennedy C, Khalil N, et al. Mutations in the human LARGE gene cause MDC1D, a novel form of congenital muscular dystrophy with severe mental retardation and abnormal glycosylation of alpha-dystroglycan. Hum Mol Genet. 2003; 12(21):2853–61. Epub 2003/09/11. https://doi.org/10.1093/hmg/ddg307 PMID: 12966029.

27. Clement E, Mercuri E, Godfrey C, Smith J, Robb S, Kinali M, et al. Brain involvement in muscular dystrophies with defective dystroglycan glycosylation. Ann Neurol. 2008; 64(5):573–82. Epub 2008/12/11. https://doi.org/10.1002/ana.21482 PMID: 19067344.

28. Mercuri E, Messina S, Bruno C, Mora M, Pegoraro E, Comi GP, et al. Congenital muscular dystrophies with defective glycosylation of dystroglycan: a population study. Neurology. 2009; 72(21):1802–9. Epub 2009/03/21. https://doi.org/10.1212/01.wnl.0000346518.68110.60 PMID: 19299310.

29. de Bernabe DB, Inamori K, Yoshida-Moriguchi T, Weydert CJ, Harper HA, Willer T, et al. Loss of alpha-dystroglycan laminin binding in epithelium-derived cancers is caused by silencing of LARGE. The

Journal of biological chemistry. 2009; 284(17):11279–84. Epub 2009/02/27. https://doi.org/10.1074/jbc.C900007200 PMID: 19244252

30. Liu Y, Huang S, Kuang M, Wang H, Xie Q. High LARGE1 Expression May Predict Benefit from Adjuvant Chemotherapy in Resected Non-Small-Cell Lung Cancer. Pharmgenomics Pers Med. 2021; 14:87–99. Epub 2021/01/28. https://doi.org/10.2147/PGPM.S271516 PMID: 33500650

31. Grewal PK, McLaughlan JM, Moore CJ, Browning CA, Hewitt JE. Characterization of the LARGE family of putative glycosyltransferases associated with dystroglycanopathies. Glycobiology. 2005; 15 (10):912–23. Epub 2005/06/17. https://doi.org/10.1093/glycob/cwi094 PMID: 15958417.

32. Peyrard M, Seroussi E, Sandberg-Nordqvist AC, Xie YG, Han FY, Fransson I, et al. The human LARGE gene from 22q12.3-q13.1 is a new, distinct member of the glycosyltransferase gene family. Proc Natl Acad Sci U S A. 1999; 96(2):598–603. Epub 1999/01/20. https://doi.org/10.1073/pnas.96.2.598 PMID: 9892679

33. Clarke BR, Ovchinnikova OG, Sweeney RP, Kamski-Hennekam ER, Gitalis R, Mallette E, et al. A bifunctional O-antigen polymerase structure reveals a new glycosyltransferase family. Nat Chem Biol. 2020; 16(4):450–7. Epub 2020/03/11. https://doi.org/10.1038/s41589-020-0494-0 PMID: 32152541.

34. Brazier-Hicks M, Offen WA, Gershater MC, Revett TJ, Lim EK, Bowles DJ, et al. Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. Proc Natl Acad Sci U S A. 2007; 104(51):20238–43. Epub 2007/12/14. https://doi.org/10.1073/pnas.0706421104 PMID: 18077347

35. Jeong JH, Kim YS, Rojviriya C, Ha SC, Kang BS, Kim YG. Crystal structures of bifunctional penicillin-binding protein 4 from Listeria monocytogenes. Antimicrobial agents and chemotherapy. 2013; 57 (8):3507–12. Epub 2013/05/15. https://doi.org/10.1128/AAC.00144-13 PMID: 23669378

36. Zhang H, Zhou M, Yang T, Haslam SM, Dell A, Wu H. New Helical Binding Domain Mediates a Glycosyltransferase Activity of a Bifunctional Protein. The Journal of biological chemistry. 2016; 291 (42):22106–17. Epub 2016/08/20. https://doi.org/10.1074/jbc.M116.731695 PMID: 27539847

37. Harrus D, Kellokumpu S, Glumoff T. Crystal structures of eukaryote glycosyltransferases reveal biologically relevant enzyme homooligomers. Cellular and molecular life sciences: CMLS. 2018; 75(5):833–48. Epub 2017/09/22. https://doi.org/10.1007/s00018-017-2659-x PMID: 28932871.

38. Pedersen LC, Tsuchida K, Kitagawa H, Sugahara K, Darden TA, Negishi M. Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. The Journal of biological chemistry. 2000; 275(44):34580–5. Epub 2000/08/18. https://doi.org/10.1074/jbc.M007399200 PMID: 10946001.

39. Hashimoto K, Madej T, Bryant SH, Panchenko AR. Functional states of homooligomers: insights from the evolution of glycosyltransferases. J Mol Biol. 2010; 399(1):196–206. Epub 2010/04/13. https://doi.org/10.1016/j.jmb.2010.03.059 PMID: 20381499

40. Kellokumpu S, Hassinen A, Glumoff T. Glycosyltransferase complexes in eukaryotes: long-known, prevalent but still unrecognized. Cellular and molecular life sciences: CMLS. 2016; 73(2):305–25. Epub 2015/10/18. https://doi.org/10.1007/s00018-015-2066-0 PMID: 26474840

41. Nilsson T, Rabouille C, Hui N, Watson R, Warren G. The role of the membrane-spanning domain and stalk region of N-acetylglucosaminyltransferase I in retention, kin recognition and structural maintenance of the Golgi apparatus in HeLa cells. J Cell Sci. 1996; 109 (Pt 7):1975–89. Epub 1996/07/01. https://doi.org/10.1242/jcs.109.7.1975 PMID: 8832420.

42. Sasai K, Ikeda Y, Tsuda T, Ihara H, Korekane H, Shiota K, et al. The critical role of the stem region as a functional domain responsible for the oligomerization and Golgi localization of N-acetylglucosaminyltransferase V. The involvement of a domain homophilic interaction. The Journal of biological chemistry. 2001; 276(1):759–65. Epub 2000/10/12. https://doi.org/10.1074/jbc.M004972200 PMID: 11024048.

43. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007; 40(Pt 4):658–74. https://doi.org/10.1107/S0021889807021206 PMID: 19461840

44. Jackson RN, McCoy AJ, Terwilliger TC, Read RJ, Wiedenheft B. X-ray structure determination using low-resolution electron microscopy maps for molecular replacement. Nat Protoc. 2015; 10(9):1275–84. Epub 2015/08/01. https://doi.org/10.1038/nprot.2015.069 PMID: 26226459

45. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta crystallographica Section D, Biological crystallography. 2008; 64(Pt 1):61–9. Epub 2007/12/21. https://doi.org/10.1107/S090744490705024X PMID: 18094468

46. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta crystallographica Section D, Biological crystallography. 2010; 66(Pt 4):486–501. https://doi.org/10.1107/S0907444910007493 PMID: 20383002

47. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography. 2010; 66(Pt 2):213–21. Epub 2010/02/04. https://doi.org/10.1107/S0907444909052925 PMID: 20124702

48. Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic acids research. 2022; 50(D1):D571–D7. Epub 2021/12/02. https://doi.org/10.1093/nar/gkab1045 PMID: 34850161

49. Wiggins CA, Munro S. Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proc Natl Acad Sci U S A. 1998; 95(14):7945–50. Epub 1998/07/08. https://doi.org/10.1073/pnas.95.14.7945 PMID: 9653120

50. Yu H, Takeuchi M, LeBarron J, Kantharia J, London E, Bakker H, et al. Notch-modifying xylosyltransferase structures support an SNi-like retaining mechanism. Nat Chem Biol. 2015; 11(11):847–54. Epub 2015/09/29. https://doi.org/10.1038/nchembio.1927 PMID: 26414444

51. Culbertson AT, Ehrlich JJ, Choe JY, Honzatko RB, Zabotina OA. Structure of xyloglucan xylosyltransferase 1 reveals simple steric rules that define biological patterns of xyloglucan polymers. Proc Natl Acad Sci U S A. 2018; 115(23):6064–9. Epub 2018/05/23. https://doi.org/10.1073/pnas.1801105115 PMID: 29784804

52. Yang JM, Tung CH. Protein structure database search and evolutionary classification. Nucleic acids research. 2006; 34(13):3646–59. Epub 2006/08/04. https://doi.org/10.1093/nar/gkl395 PMID: 16885238

53. Osawa T, Sugiura N, Shimada H, Hirooka R, Tsuji A, Shirakawa T, et al. Crystal structure of chondroitin polymerase from Escherichia coli K4. Biochem Biophys Res Commun. 2009; 378(1):10–4. Epub 2008/09/06. https://doi.org/10.1016/j.bbrc.2008.08.121 PMID: 18771653.

54. Inamori K, Willer T, Hara Y, Venzke D, Anderson ME, Clarke NF, et al. Endogenous glucuronyltransferase activity of LARGE or LARGE2 required for functional modification of alpha-dystroglycan in cells and tissues. The Journal of biological chemistry. 2014; 289(41):28138–48. Epub 2014/08/21. https://doi.org/10.1074/jbc.M114.597831 PMID: 25138275

55. Ellgaard L, Helenius A. Quality control in the endoplasmic reticulum. Nat Rev Mol Cell Biol. 2003; 4(3):181–91. Epub 2003/03/04. https://doi.org/10.1038/nrm1052 PMID: 12612637.

56. Kakuda S, Shiba T, Ishiguro M, Tagawa H, Oka S, Kajihara Y, et al. Structural basis for acceptor substrate recognition of a human glucuronyltransferase, GlcAT-P, an enzyme critical in the biosynthesis of the carbohydrate epitope HNK-1. The Journal of biological chemistry. 2004; 279(21):22693–703. Epub 2004/03/03. https://doi.org/10.1074/jbc.M400622200 PMID: 14993226.

57. Shiba T, Kakuda S, Ishiguro M, Morita I, Oka S, Kawasaki T, et al. Crystal structure of GlcAT-S, a human glucuronyltransferase, involved in the biosynthesis of the HNK-1 carbohydrate epitope. Proteins. 2006; 65(2):499–508. Epub 2006/08/10. https://doi.org/10.1002/prot.21118 PMID: 16897771.

58. Chou YH, Pogorelko G, Zabotina OA. Xyloglucan xylosyltransferases XXT1, XXT2, and XXT5 and the glucan synthase CSLC4 form Golgi-localized multiprotein complexes. Plant Physiol. 2012; 159(4):1355–66. Epub 2012/06/06. https://doi.org/10.1104/pp.112.199356 PMID: 22665445

59. Egelund J, Petersen BL, Motawia MS, Damager I, Faik A, Olsen CE, et al. Arabidopsis thaliana RGXT1 and RGXT2 encode Golgi-localized (1,3)-alpha-D-xylosyltransferases involved in the synthesis of pectic rhamnogalacturonan-II. Plant Cell. 2006; 18(10):2593–607. Epub 2006/10/24. https://doi.org/10.1105/tpc.105.036566 PMID: 17056709

60. Scietti L, Chiapparino A, De Giorgi F, Fumagalli M, Khoriauli L, Nergadze S, et al. Molecular architecture of the multifunctional collagen lysyl hydroxylase and glycosyltransferase LH3. Nat Commun. 2018; 9(1):3163. Epub 2018/08/10. https://doi.org/10.1038/s41467-018-05631-5 PMID: 30089812

61. Meek RW, Blaza JN, Busmann JA, Alteen MG, Vocadlo DJ, Davies GJ. Cryo-EM structure provides insights into the dimer arrangement of the O-linked beta-N-acetylglucosamine transferase OGT. Nat Commun. 2021; 12(1):6508. Epub 2021/11/13. https://doi.org/10.1038/s41467-021-26796-6 PMID: 34764280

62. Joseph S, Schnicker NJ, Xu Z, Yang T, Hopkins J, Watkins M, et al. Structure and mechanism of LARGE1 matriglycan polymerase. bioRxiv. 2022:2022.05.12.491222. https://doi.org/10.1101/2022.05.12.491222

63. Brockington M, Torelli S, Prandini P, Boito C, Dolatshad NF, Longman C, et al. Localization and functional analysis of the LARGE family of glycosyltransferases: significance for muscular dystrophy. Hum Mol Genet. 2005; 14(5):657–65. Epub 2005/01/22. https://doi.org/10.1093/hmg/ddi062 PMID: 15661757.

64. McCormick C, Duncan G, Goutsos KT, Tufaro F. The putative tumor suppressors EXT1 and EXT2 form a stable complex that accumulates in the Golgi apparatus and catalyzes the synthesis of heparan sulfate. Proc Natl Acad Sci U S A. 2000; 97(2):668–73. Epub 2000/01/19. https://doi.org/10.1073/pnas.97.2.668 PMID: 10639137

65. Senay C, Lind T, Muguruma K, Tone Y, Kitagawa H, Sugahara K, et al. The EXT1/EXT2 tumor suppressors: catalytic activities and role in heparan sulfate biosynthesis. EMBO reports. 2000; 1(3):282–6. Epub 2001/03/21. https://doi.org/10.1093/embo-reports/kvd045 PMID: 11256613

66. Liang Z, Veeraprame H, Bayan N, Li G. The C-terminus of prenylin is important in forming a dimer conformation necessary for endoplasmic-reticulum-to-Golgi transport. Biochem J. 2004; 380(Pt 1):43–9. Epub 2004/02/26. https://doi.org/10.1042/BJ20031788 PMID: 14979871

67. Gao XD, Dean N. Distinct protein domains of the yeast Golgi GDP-mannose transporter mediate oligomer assembly and export from the endoplasmic reticulum. The Journal of biological chemistry. 2000; 275(23):17718–27. Epub 2000/04/05. https://doi.org/10.1074/jbc.M909946199 PMID: 10748175.

68. Salahpour A, Angers S, Mercier JF, Lagace M, Marullo S, Bouvier M. Homodimerization of the beta2-adrenergic receptor as a prerequisite for cell surface targeting. The Journal of biological chemistry. 2004; 279(32):33390–7. Epub 2004/05/25. https://doi.org/10.1074/jbc.M403363200 PMID: 15155738.

69. Dancourt J, Barlowe C. Erv26p-dependent export of alkaline phosphatase from the ER requires lumenal domain recognition. Traffic. 2009; 10(8):1006–18. Epub 2009/06/06. https://doi.org/10.1111/j.1600-0854.2009.00936.x PMID: 19497047

70. Jenne N, Frey K, Brugger B, Wieland FT. Oligomeric state and stoichiometry of p24 proteins in the early secretory pathway. The Journal of biological chemistry. 2002; 277(48):46504–11. Epub 2002/09/19. https://doi.org/10.1074/jbc.M206989200 PMID: 12237308.

71. Katz M, Weinstein J, Eilon-Ashkenazy M, Gehring K, Cohen-Dvashi H, Elad N, et al. Structure and receptor recognition by the Lassa virus spike complex. Nature. 2022; 603(7899):174–+. https://doi.org/10.1038/s41586-022-04429-2 PMID: 35173332

72. Kabsch W. Xds. Acta crystallographica Section D, Biological crystallography. 2010; 66(Pt 2):125–32. https://doi.org/10.1107/S0907444909047337 PMID: 20124692

73. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta crystallographica Section D, Biological crystallography. 2013; 69(Pt 7):1204–14. https://doi.org/10.1107/S0907444913000061 PMID: 23793146