

Structure and mechanism of LARGE1 matriglycan polymerase

Soumya Joseph¹, Nicholas J. Schnicker², Zhen Xu², Tiandi Yang¹, Jesse Hopkins³, Maxwell Watkins³, Srinivas Chakravarthy³, Omar Davulcu⁴, Mary E. Anderson¹, David Venzke¹, Kevin P. Campbell^{1*}

¹ Howard Hughes Medical Institute, Senator Paul D. Wellstone Muscular Dystrophy Specialized Research Center, Department of Molecular Physiology and Biophysics and Department of Neurology, Roy J. and Lucille A. Carver College of Medicine, The University of Iowa, Iowa City, Iowa, USA

²Protein and Crystallography Facility, University of Iowa, Iowa City, Iowa, USA

³BioCAT Beamline 18-ID-D, Advanced Photon Source, Argonne National Laboratory, Lemont Illinois, USA

⁴Pacific Northwest National Laboratory, Environmental Molecular Sciences Laboratory, Richland, Washington, USA

*Corresponding author. Email: kevin-campbell@uiowa.edu

Abstract:

Matriglycan is a linear polysaccharide of alternating xylose and glucuronate that binds extracellular matrix proteins and acts as a receptor for Lassa fever virus. LARGE1 synthesizes matriglycan on dystroglycan and mutations in *LARGE1* cause muscular dystrophy with abnormal brain development. However, the mechanism of matriglycan polymerization by LARGE1 is unknown. Here, we report the cryo-EM structure of LARGE1. We show that LARGE1 functions as a dimer to polymerize matriglycan by alternating activities between the xylose transferase domain on one protomer and the glucuronate transferase domain on the other protomer. Biochemical analyses using a recombinant Golgi form of dystroglycan reveal that LARGE1 polymerizes matriglycan processively. Our results provide mechanistic insights into LARGE1 function and may facilitate novel therapeutic strategies for treating neuromuscular disorders or arenaviral infections.

One-Sentence Summary:

Dimeric LARGE1 processively polymerizes matriglycan on dystroglycan using orthogonal active sites on alternate protomers.

Introduction:

Matriglycan is a linear polysaccharide of alternating xylose and glucuronate that is polymerized on dystroglycan (1), which is a mucin-like protein that is part of the dystrophin-glycoprotein complex (DGC) (2). Matriglycan binds laminin-globular (LG) domain-containing proteins in the extracellular matrix (3,4) to link the cytoskeleton with the basement membrane via the DGC (2, 5-6). The absence or reduction of matriglycan can cause various forms of muscular dystrophy with or without abnormalities in brain development (7, 8). Additionally, matriglycan is a receptor for Lassa fever virus (LASV) and other Old-World mammarenaviruses (9, 10). Matriglycan polymer length appears to be cell-type specific and is physiologically regulated during development and regeneration (11) but can be pathologically decreased by the expression of LASV or the related lymphocytic choriomeningitis virus spike glycoproteins (12).

The bifunctional glycosyltransferases, LARGE1 and LARGE2 transfer alternating xylose and glucuronate monosaccharides repetitively to polymerize matriglycan on human dystroglycan (DAG1) residues threonine-317, 319 and 379 (11, 13, 14), that carry a phosphorylated core M3 O-glycan stem (15). Mutations in *LARGE1* are rare but can cause congenital muscular dystrophies (16) with severe intellectual disability (17). Interestingly, GWAS studies found single nucleotide polymorphisms (SNPs) in *LARGE1* are common (35%) in the West African Yoruba population of Nigeria where LASV is endemic, but absent among the Luo and Masai people of Kenya and Tanzania where LASV is not found, suggesting that the *LARGE1* gene has undergone selection in populations where LASV is endemic (18).

LARGE1 is a single-pass type II transmembrane protein localized to the Golgi apparatus (19). The luminal portion consists of a putative coiled-coil domain and two globular catalytic domains. The tandem, but independent, glycosyltransferase-A (GT-A) domains, adopt a Rossmann-like fold, that typically coordinate a manganese co-factor in each active site using

signature double-aspartate sidechains (DXD) and a C-terminal histidine sidechain (20). Despite knowing the domain architecture, it is unclear how LARGE1 polymerizes matriglycan on dystroglycan. We show that LARGE1 proteins form dimers and that LARGE1 polymerizes matriglycan, of defined length, processively on dystroglycan by using the xylose transfer activity on one protomer with the glucuronate transferase on the other. Determining the structure of LARGE1 and the mechanism of matriglycan polymerization on dystroglycan may form the basis for managing matriglycan-deficient neuromuscular disease and LASV infections.

Results:

LARGE1 is a glycosylated dimer

We performed light scattering analysis using multi-angle light scattering (MALS) and small-angle X-ray scattering (SAXS) to determine the oligomeric state of LARGE proteins and to obtain low-resolution molecular envelopes. *LARGE2*, which is a gene duplication product of *LARGE1*, does not encode the coiled-coil domain (19, 20). We analyzed $\text{LARGE1}\Delta\text{TM}$ and $\text{LARGE2}\Delta\text{TM}$ in parallel to determine whether the coiled-coil domain was responsible for multimerization (Fig. 1A). Light scattering analysis showed that soluble recombinant constructs of LARGE proteins, in which the transmembrane domain was replaced by 3xFLAG (13), form dimers in solution (Fig. 1A–C and Table 1). $\text{LARGE2}\Delta\text{TM}$ also forms a dimer, despite the absence of a coiled-coil domain (Fig 1A and S1), suggesting that the globular catalytic domains of LARGE proteins are sufficient for self-association. The corrected Porod volume from SAXS data are also consistent with LARGE proteins forming dimers. SAXS-derived molecular envelopes show that $\text{LARGE1}\Delta\text{TM}$ dimers are parallel and consist of a globular head and a stem that is presumably the density representing the coiled-coil domain (Fig. 1C). The maximum dimension of $\text{LARGE1}\Delta\text{TM}$ is ~ 180 Å calculated from the pair-distance distribution function (Table 1; D_{\max} and Fig. S2; $P(r)$). The parallel orientation of LARGE dimers is consistent with

their biological topology as type-II single-pass transmembrane proteins in which the globular catalytic heads protrude into the Golgi lumen (Fig. 1D).

Proteins in the secretory pathway are often glycosylated, which increases their molecular weight above the theoretical protein-only molecular weight that is predicted from the primary sequence. Biochemical and protein-conjugate analyses by MALS and mass spectrometry were used to determine the degree of glycosylation. The weight-average molecular weights of LARGE constructs were slightly higher than double the theoretical monomeric protein-only molecular weight (Table 1), suggesting that up to 20% of the observed molecular weight may be attributed to glycan. We observed an oligomeric state closer to an integer of two for LARGE proteins treated with PNGase F suggesting that N-glycan modifications contribute to the non-integral oligomeric state (Fig. 1B, Table 1 and Fig. S3) observed in untreated samples. A decrease in molecular weight across the elution peak for untreated LARGE protein can be partially attributed to heterogeneous glycosylation. Although the size of LARGE proteins decreases when N-glycans are removed, dimerization is unaffected (Fig. 1, Table 1).

Cryo-EM reconstruction of LARGE1

A volume representing globular catalytic domains of LARGE1 Δ TM was reconstructed using single-particle cryo-EM (Fig. 2 and Fig. S5). The cryo-EM reconstruction confirms that LARGE1 Δ TM is a dimer. Each protomer consists of tandem Rossmann folds, which have a core composed of seven β -sheets surrounded by α -helices – a classic $\alpha\beta\alpha$ -sandwich. Additionally, conservation of surface residues on each LARGE protomer primarily maps to the dimer interface (Fig. S4B). Although we reconstructed LARGE1 Δ TM in both C1 and C2 symmetries, the homodimer is not perfectly symmetric. One xylose transferase domain in each dimer remained incompletely resolved, presumably due to a high degree of flexibility relative to the remainder of the protein. Approximately 16% of chain A and 8% of chain B are outside the map density. The

coiled-coil domain and N-terminal tail were not resolved likely because they were averaged to the level of noise during particle alignment due to their dynamics with respect to the globular catalytic head. For similar reasons, the N-glycans on LARGE1 Δ TM were not resolved in the cryo-EM reconstruction.

The xylose transferase active site is typical in that the manganese ion is chelated by the DXD motif between the β 4 and β 5 strands and a C-terminal histidine located between β 9 and α 12. Histidine sidechains appear to coordinate the manganese ion in the glucuronate transferase domain (Fig. S7). The DXD motif in the glucuronate transferase domain are still expected to coordinate the manganese ion during catalysis because the D563N and D565N mutations ablate glucuronate transferase activity (Fig. 3A). The glucuronate transferase domain may have greater plasticity given the structural adjustments required for metal coordination and sugar transfer. This implies that the LARGE1 Δ TM cryo-EM volume represents a resting, inactive conformation possibly due to the absence of substrates. LARGE1 Δ TM coordinates and maps have been deposited with the following PDB and EMDB accession codes: 7UI6, EMD-26540 (C1 symmetry) and 7UI7, EMD-26541 (C2 symmetry).

The dimeric structure of LARGE1 Δ TM reveals that it has two sets of active sites per molecule – a xylose transferase and glucuronate transferase site on each face. Matriglycan is likely polymerized by switching between the active sites on one face (Fig. 2C) rather than between opposite faces, which is less efficient because it would require the non-reducing end of the polymer to navigate around the catalytic head between additions. We suspect the unused transferase domains provide structural support during matriglycan polymerization.

Structural basis of matriglycan polymerization

The active sites on each protomer in the LARGE1 dimer face opposite directions (Fig. 2) suggesting matriglycan is likely polymerized using the xylosyltransferase site on one protomer with the glucuronate transferase site on the other – that is orthogonal active sites on alternate protomers, which may provide a reason for dimerization. Each glycosyltransferase domain is capable of independently transferring sugars (1). We assessed the ability of LARGE1 constructs that were exclusively devoid of either xylose (D242N/D244N; DXD1) or glucuronate (D563N; DXD3A or D565N; DXD3B) transferase activity to synthesize matriglycan on native, endogenous dystroglycan in wild-type and *LARGE1* KO HAP1 cells.

LARGE1 KO HAP1 cells transduced with adenovirus encoding wild-type LARGE1 restored matriglycan on native, endogenous dystroglycan, whereas LARGE1 mutants lacking either the xylose (DXD1) or glucuronate (DXD3A/B) transferase activity were not able to polymerize matriglycan when transduced in isolation (Fig. 3A). However, a combination of the DXD1 and DXD3A/B LARGE1 mutants synthesized matriglycan on dystroglycan to the same extent as cells transduced with wild-type LARGE1, suggesting that either a DXD1-DXD3 heterodimer or a combination of homodimers (DXD1₂ and DXD3₂) can polymerize matriglycan (Fig 3A). The DXD1-DXD3 heterodimer might be capable of polymerizing matriglycan processively because both functions are present in a single entity, whereas a combination of LARGE1 mutant homodimers are obligated to act distributively (Fig. 3B).

Transduction of *LARGE1* KO Hap1 cells with the glucuronate transferase mutant DXD3B (D565N), which retains xylose transfer activity, can produce a faint matriglycan-positive band on Western blots (Fig. 3A; pink arrow). A similar band for the DXD3A (D563N) mutant may be observed when the image contrast is increased (Fig. S9). Therefore, we suspect that the addition of single xylose unit, results in the faint, lower molecular weight band (Fig. 3A;

pink arrow). The efficiency of xylose transfer is greater for DXD3B than DXD3A, which suggests that the mutation of aspartate-563 in the glucuronate transferase domain is communicated to the xylose transferase domain better than the mutation of aspartate-565. Thus, we may be observing interdomain communication between the catalytic domains of LARGE1.

Complementing *LARGE1* KO Hap1 cells with wild-type LARGE1 produced matriglycan that appeared dispersed and uncontrolled in length and did not recapitulate the discrete band observed in untransduced control wild-type Hap1 cells (Fig. 3A). This may be due to overexpression of the LARGE1 protein from the adenovirus. Similarly, complementing wild-type Hap1 cells with either of the glucuronate transferase mutants, DXD3A or DXD3B, produced indiscrete matriglycan-positive bands. Conversely, transduction of wild-type Hap1 cells with adenovirus encoding a LARGE1 xylose transferase mutant (DXD1) decreased the size of core dystroglycan (Fig. 3A; green arrow) compared to untransduced controls. A heterodimer, which has at least one DXD1 mutant protomer, might compete for dystroglycan or interfere with matriglycan polymerization by preventing the first xylose transfer.

Heterodimers consisting of a wild-type and a mutant protomer are known to abrogate protein function in a variety of contexts including signaling of G-protein coupled receptors (21, 22), DNA binding of basic helix-loop-helix transcriptional factors (23), and activation by co-regulators, such as p53 (24). Given that isolated homodimers of LARGE1 mutants (DXD1₂ and DXD3₂) are incapable of polymerizing matriglycan and that native endogenous LARGE1 is not easily detected by Western blot, we attribute the effects of uncontrolled matriglycan polymerization to the dominant-negative phenotype generated by wild-type mutant heterodimers (wt-DXD1 or wt-DXD3), which would suggest a degree of allostery given that at least one set of orthogonal domains are active.

Matriglycan of discrete length is processively polymerized on dystroglycan

Biopolymers such as polynucleotides and polypeptides are synthesized using templates that regulate their length and permutation. However, the polymerization of glycans is a template-free process. Branched glycans are produced by the stochastic interaction of precursors and glycosyltransferase (25), whereas the length of linear polysaccharides are potentially governed by factors such as competing reactions (25), capping of polymers (26–28), substrate availability, allosteric modulators (29–32) or molecular timers or rulers (33–36) that are intrinsic to the enzyme-substrate complex. In the case of matriglycan, processive polymerization is predicated by the tandem arrangement of catalytic domains in LARGE1. Although most polymerases are not purely processive, the discrete distributions of matriglycan-modified dystroglycan as seen in Western blots from various tissues (11) suggests that LARGE polymerases may be processive (37). We sought to determine whether LARGE1 polymerizes matriglycan predominantly through a distributive or processive mechanism using an *in vitro* assay because *in vivo* co-transduction assays produce a mixture of dimers and because interference from endogenous factors might confound results.

Dystroglycan ordinarily encounters LARGE1 in the Golgi apparatus (Fig. 4A). We wanted a construct of dystroglycan for *in vitro* LARGE1 matriglycan polymerase assays that recapitulate the Golgi enzyme-substrate complex. Therefore, we designed a secreted, soluble construct, DAG1₂₈₋₇₄₉^{R311A/R312A}, that retains the dystroglycan globular N-terminal domain (DGN) (38) to mimic its immature, but relevant, configuration in the Golgi (Fig. 4B). We observed a discrete matriglycan-positive band (150-200 kDa) of recombinant dystroglycan (Fig. 4C) due to the activity of endogenous LARGE1. DAG1₂₈₋₇₄₉^{R311A/R312A} without matriglycan (Fig. 4C; supplementary materials) was used as a substrate for *in vitro* matriglycan polymerization assays (Fig. 5).

Processively synthesized polymers accumulate minimal reaction intermediates (37), whereas products synthesized distributively have a Poisson distribution (Fig. 5B and S6). We compared the distribution of matriglycan products on DAG $128\text{-}749^{\text{R311A/R312A}}$ polymerized by LARGE1 ΔTM made by a mixture of the LARGE1 ΔTM mutants, DXD 1_2 and DXD 3_2 , which are obligated to polymerize matriglycan distributively (Fig. 3B and 5). A discrete matriglycan-positive band on Western blot is absent when a combination of LARGE1 ΔTM with either the xylose (DXD $_2$) or glucuronate (DXD 3_2) transferase sites mutated polymerize matriglycan, rather we observe a smear of indiscrete lengths (Fig. 5D), which relates to Poisson distribution of products synthesized by LARGE1 ΔTM on a synthetic substrate (Fig. 5B and S6). In contrast, polymers synthesized on recombinant dystroglycan by wild-type LARGE1 ΔTM produced a discrete band (125-250 kDa), resembling the naturally observed distribution, on a background smear of apparently indiscriminate lengths (Fig. 5C and D). These *in vitro* experiments suggest that a molecular ruler intrinsic to the LARGE1-dystroglycan enzyme-substrate complex controls polymer length, and that LARGE1 primarily polymerizes matriglycan processively.

Interestingly, the faint band observed on Western blots of *LARGE1* KO HAP1 cells transduced with DXD3A/B (Fig. 3A) is also observed in the *in vitro* matriglycan polymerization assay where only UDP-xylose is added as a substrate (Fig. 5C; n = 3). A similarly faint band is observed when protein *O*-mannose kinase (POMK) is inactivated (29).

Because standard, ubiquitous GT-A, Rossman-like domains, each with its own catalytic activity, are uniquely arranged in tandem to polymerize a linear chain of alternating sugars, the protein scaffold of LARGE might be amenable to engineering. Potentially, many polymers of alternating monosaccharides could be synthesized by substituting the individual GT-A domains of LARGE. Their lengths may also be controlled by exploiting the molecular “ruler” created by the LARGE1-dystroglycan enzyme-substrate complex. Engineering linear polymers of

alternating monosaccharides with defined lengths may be broadly applied in the scientific, medical, or industrial fields.

Conclusion:

The bifunctional glycosyltransferase LARGE is a parallel homodimer. LARGE1 polymerizes matriglycan on dystroglycan using orthogonal active sites on alternate protomers. Matriglycan of discrete length is polymerized processively using a molecular ruler that is likely intrinsic to the LARGE1-dystroglycan enzyme-substrate complex. The standard architecture of tandem Rossmann-like GT-A folds (Fig. S4A) in LARGE proteins may lend themselves to enzyme engineering for the synthesis of polymers of alternating sugars of defined length. The structure and mechanism of LARGE elucidated in this study is an essential first step for the modulation of matriglycan polymerization in cases of neuromuscular disorders as well as LASV and other pathogenic mammarenaviruses that bind matriglycan for cell entry.

References:

1. K. Inamori, T. Yoshida-moriguchi, Y. Hara, M. E. Anderson, L. Yu, K. P. Campbell, Dystroglycan function requires xylosyl- and glucuronyltransferase activities of LARGE. *Science*. **335**, 93–96 (2012).
2. J. M. Ervasti, K. P. Campbell, A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin. *J. Cell Biol.* **122**, 809–823 (1993).
3. E. Hohenester, Laminin G-like domains: dystroglycan-specific lectins. *Curr. Opin. Struct. Biol.* **56**, 56–63 (2019).
4. D. C. Briggs, T. Yoshida-Moriguchi, T. Zheng, D. Venzke, M. E. Anderson, A. Strazzulli, M. Moracci, L. Yu, E. Hohenester, K. P. Campbell, Structural basis of laminin binding to the LARGE glycans on dystroglycan. *Nat. Chem. Biol.* **12**, 810–814 (2016).
5. O. Ibraghimov-Beskrovnaya, J. M. Ervasti, C. J. Leveille, C. A. Slaughter, S. W. Sernett, K. P. Campbell, Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature*. **355**, 696 (1992).
6. T. Yoshida-Moriguchi, K. P. Campbell, Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. *Glycobiology*. **25**, 702–713 (2015).
7. C. Godfrey, A. R. Foley, E. Clement, F. Muntoni. Dystroglycanopathies: coming into focus. *Curr. Opin. Genet. Dev.* **21**, 278–285 (2011).
8. K. G. Meilleur, K. Zukosky, L. Medne, P. Fequiere, N. Powell-Hamilton, T. L. Winder, A. Alsaman, A. W. El-Hattab, J. Dastgir, Y. Hu, S. Donkervoort, J. A. Golden, R. Eagle, R. Finkel, M. Scavina, I. C. Hood, L. B. Rorke-Adams, C. G. Bönnemann, Clinical, pathologic, and mutational spectrum of dystroglycanopathy caused by LARGE mutations. *J. Neuropathol. Exp. Neurol.* **73**, 425–441 (2014).
9. M. Katz, J. Weinstein, M. Eilon-Ashkenazy, K. Gehring, H. Cohen-Dvashi, N. Elad, S. J.

Fleishman, R. Diskin, Structure and receptor recognition by the Lassa virus spike complex. *Nature*. **603**, 174–179 (2022).

10. S. Joseph, K. P. Campbell, Lassa Fever Virus Binds Matriglycan—A Polymer of Alternating Xylose and Glucuronate—On α -Dystroglycan. *Viruses*. **13**, 1679 (2021).

11. M. M. Goddeeris, B. Wu, D. Venzke, T. Yoshida-Moriguchi, F. Saito, K. Matsumura, S. A. Moore, K. P. Campbell, LARGE glycans on dystroglycan function as a tunable matrix scaffold to prevent dystrophy. *Nature*. **503**, 136 (2013).

12. J. M. Rojek, K. P. Campbell, M. B. A. Oldstone, S. Kunz, Old World arenavirus infection interferes with the expression of functional α -dystroglycan in the host cell. *Mol. Biol. Cell*. **18**, 4493–4507 (2007).

13. Y. Hara, M. Kanagawa, S. Kunz, T. Yoshida-moriguchi, J. S. Satz, Y. M. Kobayashi, Z. Zhu, S. J. Burden, M. B. A. Oldstone, K. P. Campbell, 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.* **108**, 17426–17431 (2011).

14. K. Inamori, T. Willer, Y. Hara, D. Venzke, M. E. Anderson, N. F. Clarke, P. Guicheney, C. G. Bönnemann, S. A. Moore, K. P. Campbell, Endogenous glucuronyltransferase activity of LARGE or LARGE2 require for functional modification of α -dystroglycan in cells and tissues. *J. Biol. Chem.* **280**, 28138–48, 2014

15. T. Yoshida-Moriguchi, T. Willer, M. E. Anderson, D. Venzke, T. Whyte, F. Muntoni, H. Lee, S. F. Nelson, L. Yu, K. P. Campbell, SGK196 is a glycosylation-specific *O*-mannose kinase required for dystroglycan function. *Science*. **341**, 896–899 (2013).

16. E. Mercuri, S. Messina, C. Bruno, M. Mora, E. Pegoraro, G. P. Comi, A. D’Amico, C. Aiello, R. Biancheri, A. Berardinelli, P. Boffi, D. Cassandrini, A. Laverda, M. Moggio, L. Morandi, I. Moroni, M. Pane, R. Pezzani, A. Pichieccchio, A. Pini, C. Minetti, T. Mongini, E.

Mottarelli, E. Ricci, A. Ruggieri, S. Saredi, C. Scuderi, A. Tessa, A. Toscano, G. Tortorella, C. P. Trevisan, C. Uggetti, G. Vasco, F. M. Santorelli, E. Bertini, Congenital muscular dystrophies with defective glycosylation of dystroglycan: A population study. *Neurology*. **72**, 1802–1809 (2009).

17. E. Clement, E. Mercuri, C. Godfrey, J. Smith, S. Robb, M. Kinali, V. Straub, K. Bushby, A. Manzur, B. Talim, F. Cowan, R. Quinlivan, A. Klein, C. Longman, R. McWilliam, H. Topaloglu, R. Mein, S. Abbs, K. North, A. J. Barkovich, M. Rutherford, F. Muntoni, Brain involvement in muscular dystrophies with defective dystroglycan glycosylation. *Ann. Neurol.* **64**, 573–582 (2008).

18. P. C. Sabeti, P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, E. H. Byrne, Genome-wide detection and characterization of positive selection in human populations. *Nature*. **449**, 913–918 (2007).

19. P. Myriam, S. Eyal, S.-N. Ann-Christin, X. Ya-Gang, H. Fei-Yu, F. Ingegerd, C. John, D. Ian, K.-A. Maria, I. Stephan, D. J. P., The human LARGE gene from 22q12.3-q13.1 is a new, distinct member of the glycosyltransferase gene family. *Proc. Natl. Acad. Sci.* **96**, 598–603 (1999).

20. R. Taujale, A. Venkat, L. C. Huang, Z. Zhou, W. Yeung, K. M. Rasheed, S. Li, A. S. Edison, K. W. Moremen, N. Kannan, Deep evolutionary analysis reveals the design principles of fold a glycosyltransferases. *Elife*. **9**, e54532 (2020).

21. A. U. Gehret, A. Bajaj, F. Naider, M. E. Dumont, Oligomerization of the yeast α -factor receptor: implications for dominant negative effects of mutant receptors. *J. Biol. Chem.* **281**, 20698–20714 (2006).

22. T. Zariñán, M. A. Perez-Solís, G. Maya-Núñez, P. Casas-González, P. M. Conn, J. A. Dias, A. Ulloa-Aguirre, Dominant negative effects of human follicle-stimulating hormone receptor

expression-deficient mutants on wild-type receptor cell surface expression. Rescue of oligomerization-dependent defective receptor expression by using cognate decoys. *Mol. Cell. Endocrinol.* **321**, 112–122 (2010).

23. R. Benezra, R. L. Davis, D. Lockshon, D. L. Turner, H. Weintraub, The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell.* **61**, 49–59 (1990).

24. C. D. Nicholls, K. G. McLure, M. A. Shields, P. W. K. Lee, Biogenesis of p53 involves cotranslational dimerization of monomers and posttranslational dimerization of dimers. Implications on the dominant negative effect. *J. Biol. Chem.* **277**, 12937–12945 (2002).

25. A. Jaiman, M. Thattai, Golgi compartments enable controlled biomolecular assembly using promiscuous enzymes. *Elife.* **9**, e49573 (2020).

26. M. O. Sheikh, D. Venzke, M. E. Anderson, T. Yoshida-Moriguchi, J. N. Glushka, A. V Nairn, M. Galizzi, K. W. Moremen, K. P. Campbell, L. Wells, HNK-1 sulfotransferase modulates α -dystroglycan glycosylation by 3-O-sulfation of glucuronic acid on matriglycan. *Glycobiology.* **30**, 817–829 (2020).

27. C. E. Faller, O. Guvench, Terminal sialic acids on CD44 N-glycans can block hyaluronan binding by forming competing intramolecular contacts with arginine sidechains. *Proteins Struct. Funct. Bioinforma.* **82**, 3079–3089 (2014).

28. A. M. Moody, S. J. North, B. Reinhold, S. J. Van Dyken, M. E. Rogers, M. Panico, A. Dell, H. R. Morris, J. D. Marth, E. L. Reinherz, Sialic acid capping of CD8 β core 1-*O*--glycans controls thymocyte-major histocompatibility complex class I interaction. *J. Biol. Chem.* **278**, 7240–7246 (2003).

29. A. S. Walimbe, H. Okuma, S. Joseph, T. Yang, T. Yonekawa, J. M. Hord, D. Venzke, M. E. Anderson, S. Torelli, A. Manzur, M. Devereaux, M. Cuellar, S. Prouty, S. Ocampo Landa, L. Yu, J. Xiao, J. E. Dixon, F. Muntoni, K. P. Campbell, POMK regulates dystroglycan function

via LARGE1-mediated elongation of matriglycan. *Elife*. **9**, e61388 (2020).

30. Y. Tone, L. C. Pedersen, T. Yamamoto, T. Izumikawa, H. Kitagawa, J. Nishihara, J. I. Tamura, M. Negishi, K. Sugahara, 2-O-phosphorylation of xylose and 6-O-sulfation of galactose in the protein linkage region of glycosaminoglycans influence the glucuronyltransferase-I activity involved in the linkage region synthesis. *J. Biol. Chem.* **283**, 16801–16807 (2008).

31. T. Koike, T. Izumikawa, J. I. Tamura, H. Kitagawa, FAM20B is a kinase that phosphorylates xylose in the glycosaminoglycan-protein linkage region. *Biochem. J.* **421**, 157–162 (2009).

32. S. Nadanaka, S. Zhou, S. Kagiyama, N. Shoji, K. Sugahara, K. Sugihara, M. Asano, H. Kitagawa, EXTL2, a member of the EXTLfamily of tumor suppressors, controls glycosaminoglycan biosynthesis in a xylose kinase-dependent manner. *J. Biol. Chem.* **288**, 9321–9333 (2013).

33. M. R. Levengood, R. A. Splain, L. L. Kiessling, Monitoring Processivity and Length Control of a Carbohydrate Polymerase. *J. Am. Chem. Soc.* **133**, 12758–12766 (2011).

34. D. M. Williams, O. G. Ovchinnikova, A. Koizumi, I. L. Mainprize, M. S. Kimber, T. L. Lowary, C. Whitfield, Single polysaccharide assembly protein that integrates polymerization, termination, and chain-length quality control. *Proc. Natl. Acad. Sci.* **114**, E1215–E1223 (2017).

35. A. R. Hesser, K. Schaefer, W. Lee, S. Walker, Lipoteichoic acid polymer length is determined by competition between free starter units. *Proc. Natl. Acad. Sci.* **117**, 29669–29676 (2020).

36. R. Woodward, W. Yi, L. Li, G. Zhao, H. Eguchi, P. R. Sridhar, H. Guo, J. K. Song, E. Motari, L. Cai, P. Kelleher, X. Liu, W. Han, W. Zhang, Y. Ding, M. Li, P. G. Wang, *In vitro* bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. *Nat. Chem.*

Biol. **6**, 418–423 (2010).

37. L. Yakovlieva, M. T. C. Walvoort, Processivity in Bacterial Glycosyltransferases. *ACS Chem. Biol.* **15**, 3–16 (2020).

38. D. Bozic, F. Sciandra, D. Lamba, A. Brancaccio, The structure of the N-terminal region of murine skeletal muscle α -dystroglycan discloses a modular architecture. *J. Biol. Chem.* **279**, 44812-44816 (2004).

Acknowledgments:

We thank Drs. Sandipan Chowdhury, Erhard Hohenester, Andrew Ward, Lance Wells, and Liping Yu for helpful discussions. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. This project was supported by grants P41 GM103622 and P30 GM138395 from the National Institute of General Medical Sciences of the National Institutes of Health. Use of the Pilatus 3 1M detector was provided by grant 1S10OD018090 from NIGMS. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institute of General Medical Sciences or the National Institute of Health. Extraordinary facility operations were supported in part by the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on the response to COVID-19, with funding provided by the Coronavirus CARES Act. A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. We are grateful to Dr. Jennifer Barr of the Scientific Editing and Research Communication Core at the University of Iowa Carver College of Medicine for critical reading of the manuscript. We are grateful to Amber Mower and Jaeda Harmon for assistance with administrative support.

Funding:

Research reported in this publication was supported by the National Institute of Neurological Disorders And Stroke of the National Institutes of Health under Award Number P50NS053672. The content is solely the responsibility of the authors and does not necessarily represent the

official views of the National Institutes of Health. KPC is an investigator of the Howard Hughes Medical Institute.

Author contributions:

Conceptualization: SJ, KPC

Methodology: SJ, KPC, NJS, ZX, TY, JH, SC, OD, MEA, DV

Investigation: SJ, KPC, NJS, ZX, JH, TY, JH, SC, OD, MEA, DV

Visualization: SJ, NJS, ZX, TY, MEA, DV

Supervision: KPC

Writing—original draft: SJ, KPC

Writing—review & editing: SJ, KPC, NJS, ZX, TY, MEA, DV

Competing interests:

Authors declare that they have no competing interests.

Data and materials availability:

All data are available in the main text or the supplementary materials. LARGE1 coordinates and maps have been deposited with the accession codes 7UI6 and EMD-26540 (C1 symmetry) and 7UI7 and EMD-26541 (C2 symmetry).

Supplementary Materials

Materials and Methods

Figures S1-S13

Tables S1-S3

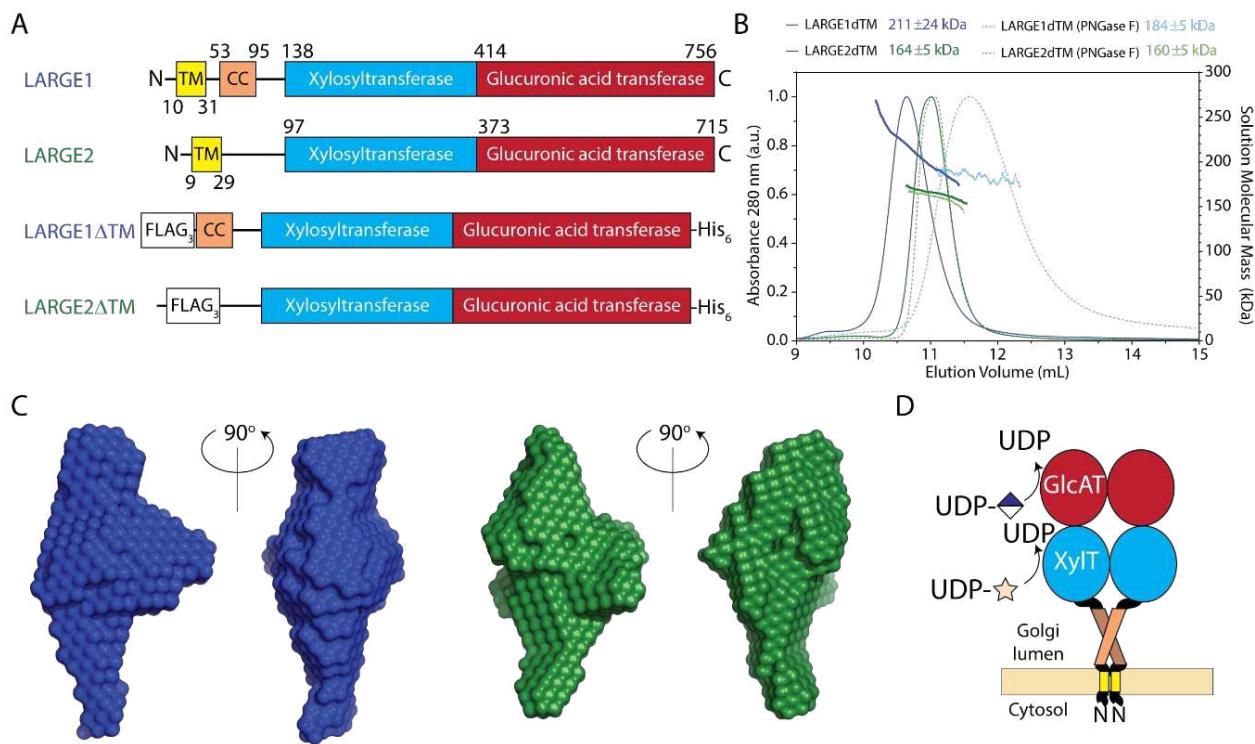


Fig. 1. Light scattering analyses of LARGE1 Δ TM and LARGE2 Δ TM. (A) Schematic of LARGE proteins. LARGE2 lacks a coiled-coil domain (CC) found in LARGE1. Transmembrane domains (TM) were replaced with 3xFLAG tag (FLAG₃) and a C-terminal hexahistidine tag was added (His₆). (B) Size-exclusion chromatogram (SEC) with on-line multi-angle light scattering (MALS) of LARGE1 Δ TM (blue) and LARGE2 Δ TM (green) or treated with PNGase F (dotted blue and light green, respectively). The elution profile was monitored using absorbance at 280 nm (continuous line) over the elution volume (mL). Discrete points represent solution weight-average molecular weight (MW, kDa) calculated using a change in refractive index $dn/dc = 0.185$ for protein and $dn/dc = 0.135$ for glycan. (C) Molecular envelopes of LARGE1 Δ TM (blue) and LARGE2 Δ TM (green) from small-angle X-ray scattering (SAXS). (D) Probable organization of dimeric LARGE1 in the Golgi membrane.

Protein	Radius of gyration (Å)				Molecular Weight (kDa)		
	Guinier	GNOM	D _{max} (Å)	Porod Volume (Å ³)	SAXS (Porod)	MALS	Theoretical
LARGE1ΔTM	43.99±0.08	45.05±0.17	184	264,000	219.5	211±24	
LARGE1ΔTM (PNGase F)	42.83±0.16	44.22±0.29	180	249,000	206.7	184±5	89.7
LARGE2ΔTM	41.56±0.05	42.18±0.08	165	257,000	213.0	164±5	
LARGE2ΔTM (PNGase F)	40.87±0.03	41.66±0.04	159	231,000	191.4	160±5	84.2

Table 1. Biophysical parameters of LARGE proteins derived from multi-angle light scattering (MALS) and small-angle X-ray scattering (SAXS). The radius of gyration was estimated in reciprocal space from Guinier fitting and in real space from GNOM. The maximum dimension (D_{max}) was estimated from the pair-distance distribution function (P(r)). The corrected Porod volume and molecular weight obtained from SAXS are calculated from Porod's law. The solution weight-average molecular weight from MALS can be compared to the protein-only theoretical monomer weight.

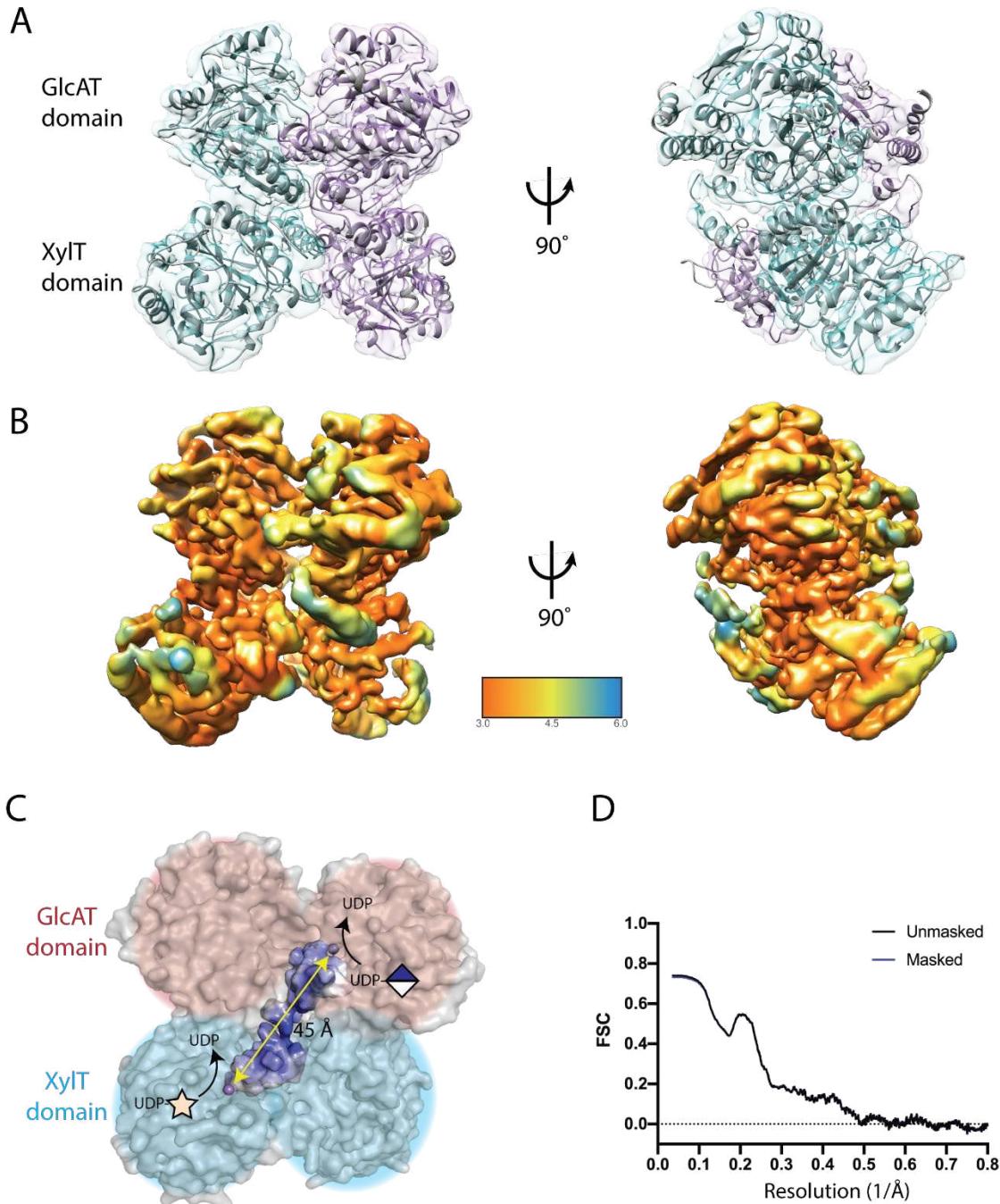
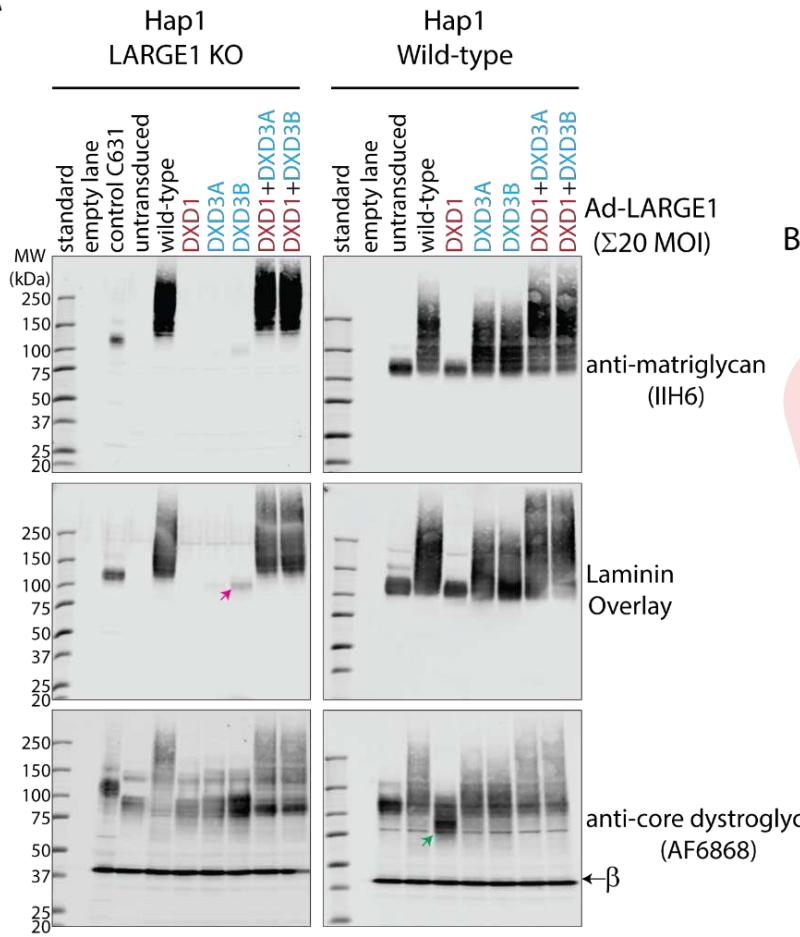


Fig. 2. Cryo-EM structure of LARGE1ΔTM. (A) Two molecules of LARGE1 (grey cartoon) were refined into the 3.7-Å map of LARGE1ΔTM; protomers are displayed as cyan and purple transparent surfaces. Spheres indicate catalytic manganese ions. (B) Reconstructed volume colored by local resolution. (C) Orthogonal active sites each coordinating a manganese ion (purple sphere) on alternate protomers face the same direction. A groove (highlighted surface) connects the xylose (cyan shading) and glucuronate (pink shading) transferase active sites on alternate protomers. The quaternary structure implies that matriglycan is polymerized using orthogonal active sites on alternate protomers of LARGE1 (diagonal yellow arrow). (D) Map-to-model FSC of LARGE1 catalytic domains' fit into the reconstructed volume.

A



B

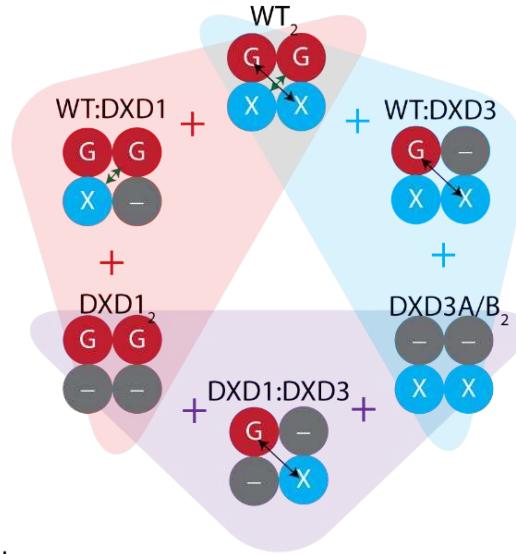
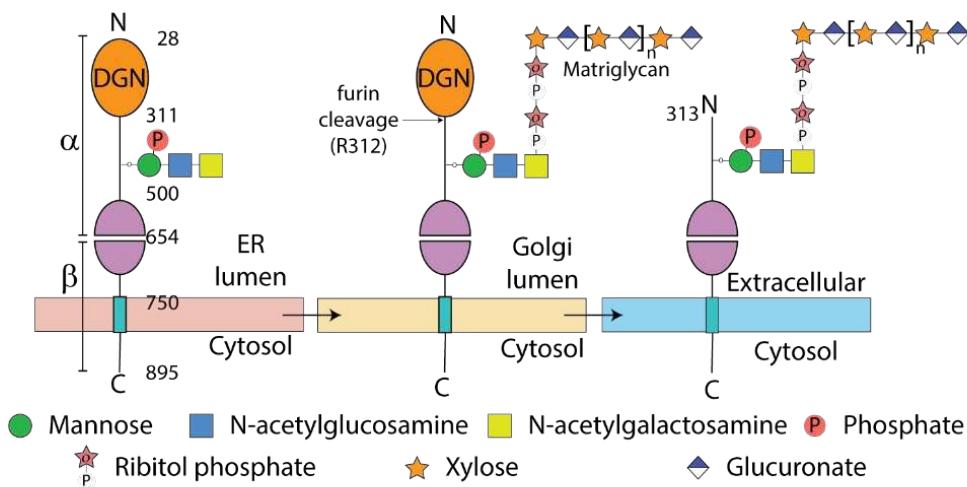
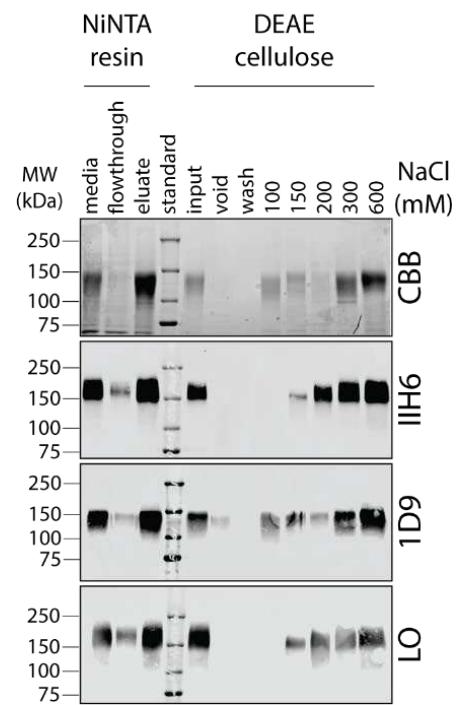


Fig. 3. A combination of LARGE1 active site mutants restore matriglycan on native, endogenous dystroglycan. (A) Western blot of *LARGE1* KO and wild-type Hap1 cell lines transduced with adenovirus encoding LARGE1 (20 MOI): wild-type (WT), xylose transfer mutant (DXD1; D242N/D244N), glucuronate transfer mutant (DXD3A/B; D563N or D565N, respectively) or a combination of both (DXD1+DXD3A/B). Matriglycan is detected using anti-matriglycan antibody (IIH6) or laminin overlay, whereas core dystroglycan protein is detected using AF6868. Pink arrows indicate a faint matriglycan-positive band in cells transduced with DXD3B in *LARGE1* KO Hap1 cells. Green arrows indicate a perturbation in the migration of dystroglycan in wild-type Hap1 cells transduced with DXD1. (B) Adenovirus produce combinations of LARGE1 dimers in Hap1 wild-type or co-transduced *LARGE1* KO cells. Background shading and color-coded plus symbols indicate dimeric combinations present in DXD1 and DXD3 co-transduced *LARGE1* KO Hap1 cells (purple), DXD1 (red) or DXD3 (blue) transduced Hap1 wild-type cells. Double-headed arrows indicate potential processive polymerization of matriglycan by orthogonal active sites on alternate protomers.

A



C



B

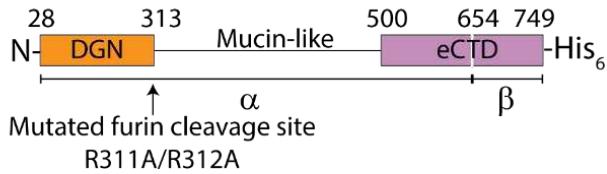


Fig. 4. Construct of recombinant Golgi dystroglycan and purification. (A) Schematic of dystroglycan glycan modifications through the secretory pathway. The signal peptide (residues 1-27), which is cleaved in the ER, is omitted. Dystroglycan globular N-terminal domain (DGN; orange) and the autoproteolytic extracellular C-terminal domain (eCTD; purple) flank the mucin-like domain (black line). Threonine 317, 319, and 379 in human dystroglycan are post-translationally modified in the ER by POMT1/2, POMGNT2, B3GALNT2, and POMK to make phosphorylated core M3 (N-acetylgalactosamine-β3-N-acetylglucosamine-β4-(phosphate-6-)mannose). In the Golgi, FKTN (FCMD) and FKRP extend core M3 by two phosphoribitol units followed by RXYLT1 (TMEM5) and B4GAT1 which add a xylose-glucuronate primer. Mature dystroglycan lacks DGN. (B) Schematic of secreted recombinant Golgi luminal human dystroglycan retains DGN, which is essential for binding to LARGE1 in the Golgi apparatus, because the furin cleavage site was mutated (R311A/R312A). (C) Coomassie brilliant blue (CBB)-stained SDS-PAGE (topmost panel) and Western blot (bottom panels) of recombinant dystroglycan purified from culture media of stably transfected HEK 293 F cells. Protein was eluted using imidazole from IMAC resin and then subject to anion exchange chromatography (DEAE cellulose) to separate dystroglycan carrying matriglycan from unmodified protein. The fraction lacking matriglycan (IIH6) and laminin-binding capacity (LO) were used for matriglycan polymerization assays. The presence of DGN on recombinant dystroglycan was confirmed by Western blotting with 1D9.

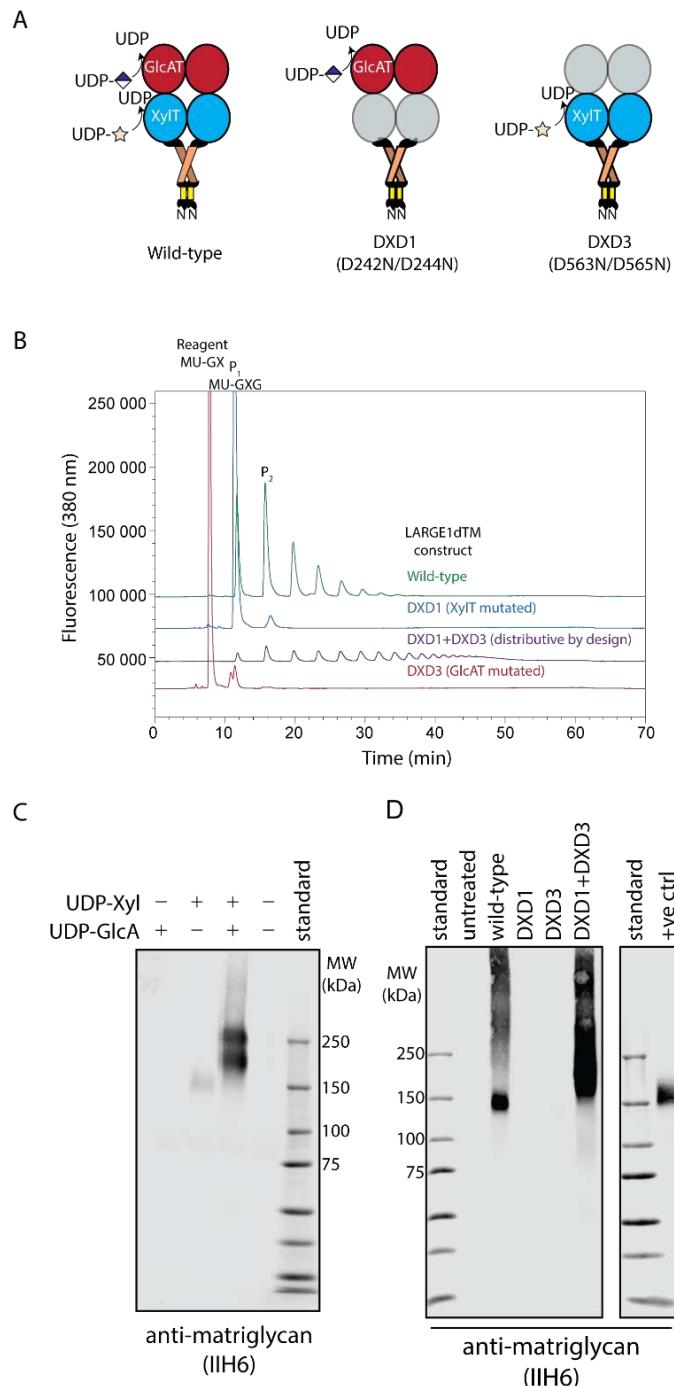


Fig. 5. LARGE1 processively polymerizes matriglycan of discrete length on dystroglycan *in vitro*. (A) Schematic of LARGE1 Δ TM constructs used for determining the mechanism of matriglycan polymerization. Wild-type LARGE1 Δ TM construct has both xylose and glucuronate transfer activities whereas DXD1 and DXD3 retain either glucuronate or xylose transfer activities, respectively. (B) Anion exchange chromatogram of matriglycan polymerized on 4-methylumbelliferyl-glucuronate-xylose by LARGE1 Δ TM constructs. (C) Western blot of matriglycan polymerized on recombinant dystroglycan by LARGE1 Δ TM in the presence of UDP-sugars at pH 7.4 (n=3). (D) Western blot of matriglycan polymerized by LARGE1 Δ TM constructs at pH 6.6 (n=3).