

Published in final edited form as:

Nat Chem Biol. 2022 July 01; 18(7): 713–723. doi:10.1038/s41589-022-01002-w.

Evolutionary balance between foldability and functionality of a glucose transporter

Hyun-Kyu Choi^{#1,2}, Hyunook Kang^{#1}, Chanwoo Lee^{#1,2}, Hyun Gyu Kim^{1,2}, Ben P. Phillips³, Sohyung Park⁴, Charlotte Tumescheit¹, Sang Ah Kim^{1,2}, Hansol Lee^{1,2}, Soung-Hun Roh^{1,2}, Heedeok Hong⁵, Martin Steinegger¹, Wonpil Im⁴, Elizabeth A. Miller^{3,*}, Hee-Jung Choi^{1,*}, Tae-Young Yoon^{1,2,*}

¹School of Biological Sciences, Seoul National University, Seoul 08826, South Korea

²Institute for Molecular Biology and Genetics, Seoul National University, Seoul 08826, South Korea

³Medical Research Council (MRC) Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK

⁴Departments of Biological Sciences and Chemistry, Lehigh University, Bethlehem, PA 18015, USA

⁵Department of Chemistry and Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA

These authors contributed equally to this work.

Abstract

Despite advances in resolving structures of multi-pass membrane proteins, little is known about the native folding pathways of these complex structures. Using single-molecule magnetic tweezers, we here report a folding pathway of purified human glucose transporter 3 (GLUT3) reconstituted within synthetic lipid bilayers. The N-terminal major facilitator superfamily (MFS) fold strictly forms first, serving as structural templates for its C-terminal counterpart, in which polar residues comprising the conduit for glucose molecules present major folding challenges. The ER membrane protein complex facilitates insertion of these hydrophilic transmembrane helices, thrusting GLUT3's microstate sampling toward folded structures. Final assembly between the N- and C-terminal MFS folds depends on specific lipids that ease desolvation of lipid shells surrounding the domain interfaces. Sequence analysis suggests that this asymmetric folding

*Correspondence should be addressed to emiller@mrc-lmb.cam.ac.uk (E.A.M.), choihj@snu.ac.kr (H.-J.C.) or tyoon@snu.ac.kr (T.-Y.Y.).

Author contributions

H.-K.C. and T.-Y.Y. conceived the project. H.-K.C., H.-J.C., E.A.M., and T.-Y.Y. designed the experiments. H.-K.C., C.L. and S.A.K. performed magnetic tweezers experiment. H.K. and H.-J.C. expressed and purified GLUT3 proteins. B.P.P. and E.A.M. expressed and purified EMC. H.G.K. performed flow cytometry experiment. S.P. and W.I. performed molecular dynamics simulation. H.G.K., C.T. and M.S. performed the bioinformatic analysis. S.A.K performed structural analysis in Extended Data Fig. 10. H.L. and S.-H.R. performed TEM imaging. H.-K.C., C.L. and T.-Y.Y. prepared the manuscript, with assistance from H.-J.C., E.A.M. and H.H. and with input from all authors.

Competing interests

The authors declare no competing interests.

propensity across the N- and C-terminal MFS folds prevails for metazoan sugar porters, revealing evolutionary conflicts between foldability and functionality faced by many multi-pass membrane proteins.

Multi-pass membrane proteins are essential gatekeepers of cells, regulating flow of information and material across cell membranes¹. While electron cryo-microscopy is revealing tertiary and quaternary structures of the multi-pass membrane proteins at an unprecedented pace², remarkably little is known about how these complex structures fold following their synthesis in the endoplasmic reticulum (ER) membrane^{3,4}. Despite the formidable complexity of membrane protein biogenesis, it is increasingly evident that some common principles guide this process. Many membrane protein families share a remarkable conservation of their tertiary structures despite huge evolutionary distances across different members⁵⁻⁸. In particular, the process of transmembrane helix (TMH) assembly is facilitated by ER chaperones, although dedicated TMH chaperones are poorly understood and seem to function by preventing aggregation rather than promoting the correct fold^{9,10}. It is thus tempting to hypothesize that the basic information for navigating the folding pathway – likely conserved across each family – is primarily encoded in the amino acid sequence of membrane proteins. Notwithstanding these prevailing models, the folding pathways of multi-pass membrane proteins remain largely elusive.

We recently used single-molecule magnetic tweezers to observe the native folding pathway of *E. coli* GlpG and human beta-2 adrenergic receptor (β 2AR) with a resolution of a few amino acids in lipid bilayers¹¹. We here extended our approach to studying the folding pathway of human glucose transporter 3 (GLUT3) as an archetype of the major facilitator superfamily (MFS) that composes the largest group of solute carrier proteins (Fig. 1a)^{1,12-16}. Like most members of MFS, GLUT3 consists of two MFS folds, each with 6 TMHs (hereafter referred to as N- and C-domains) that pack against each other in a pseudo mirror-symmetry, connected by an intra-cellular helical (ICH) domain consisting of 5 helices (Fig 1a, left and right insets)^{8,16,17}. To conduct transport, the N- and C-domains undergo rocking motions within the bilayer, alternating between conformational states with access to extracellular and cytoplasmic spaces^{18,19}. The interface between N- and C-domains is enriched with polar residues to create a conduit for glucose molecules in otherwise impermeable lipid bilayers (Fig. 1a)¹⁷.

By examining folding of various GLUT3 constructs under disparate lipid bilayer conditions, we not only delineated a complete folding order of GLUT3 domains, but also dissected detailed pathways of forming individual MFS folds. While C-domain folding and domain-domain assembly present major challenges in GLUT3 folding, we found that different types of agents involving an ER chaperone, the ER membrane protein complex (EMC), and specific lipid species facilitate GLUT3 folding in overcoming these energetic barriers. Our bioinformatics analysis further suggests that the marginal foldability of the C-domain may be a common pattern for metazoan sugar transporters, which is likely linked to enhancement or versatility of transporter functions. Thus, our observations reveal a conflict between foldability and functionality that has likely been faced by many membrane proteins throughout evolution^{20,21}.

Results

Single-molecule magnetic tweezers monitoring GLUT3 folding

To employ magnetic tweezers to observe folding of single human GLUT3 proteins, we attached DNA handles to the N- and C-termini of GLUT3 using the SpyTag/SpyCatcher system (Fig. 1a and Extended Data Fig. 1a-c)^{11,22}. After attaching the DNA handles to a magnetic bead and a polymer-coated surface, we introduced bicelle solutions, with varying lipidic compositions, to provide lipid bilayer environments to GLUT3 (Fig. 1a and Extended Data Fig. 1d-i)²³⁻²⁵. While applying a varying level of magnetic force to the bead by moving a pair of neodymium magnets, we recorded the vertical position of the magnetic bead (referred to as the extension) at sampling rates up to 1.2 kHz. The uncertainty in our bead tracking could be reduced to ~1 nm through median filtering at 5 Hz (Extended Data Fig. 2a)¹¹.

We first examined the force-extension curve (FEC) during gradual stretching and relaxation of single GLUT3 molecules (Fig. 1b). Under high mechanical tension above 20 pN, single GLUT3 proteins showed unfolding via discrete steps. This high-force unfolding culminated in a state of fully-stretched, unstructured polypeptides (referred to as U_c). During relaxation, we observed a transition from the theoretical curve for U_c to U_h in the force range from 20 to 10 pN. Since the U_h curve was generated assuming a fully-stretched state with α -helical structures restored for all TMHs, the observed transition indicated gradual coil-to-helix transitions in twelve TMHs of GLUT3. When further relaxing tension to below 5 pN, the FEC continued to follow the U_h curve (Fig. 1b, right inset). This persistent U_h state presumably resulted from weak membrane penetration of TMHs, likely due to lower hydrophobicity of GLUT3's TMHs compared with those of GlpG and β 2AR (Fig. 1b, left inset and Supplementary Fig. 1)²⁶⁻²⁸.

To observe folding of GLUT3, we applied high force of 25 pN to induce the U_c state, subsequently relaxed the force to 5 pN (taking 200 ms) and maintained the tension, in which the U_h state consequently worked as the starting state of our refolding trial at 5 pN (Fig. 1c). As anticipated from the weak propensity to penetrate membranes, single GLUT3s showed limited progression in their folding efforts at 5 pN. Under bicelle conditions that permit complete folding of GlpG and β 2AR (30 mol% of 1,2-Dimyristoyl-sn-glycero-3-phosphorylglycerol (DMPG) and 70 mol% of 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC))¹¹, we found that none of the trials achieved complete refolding (Fig. 1d,e). We observed apparent partial folding comprising about 35% (i.e., 17.2 nm) of the extension difference between the unfolded (U_h) and native (N) states (48.8 nm) (Fig. 1d). Applying hidden Markov modeling (HMM) and Bayesian information criteria (BIC)^{29,30} indicated that these traces with 35% folding progress were best fit assuming four intermediates (I_{f1} to I_{f4}) in addition to the U_h state (Fig. 1d and Extended Data Fig. 2b-d).

We searched for physicochemical conditions that could further enhance the folding progress. Given previous observations that addition of negatively charged lipids facilitates membrane protein folding^{11,25}, we tested a bicelle phase consisting purely of DMPG lipids (Fig. 1e). We observed a remarkable enhancement in folding progress, reaching up to 73 % of full folding (i.e., an extension decrease of 35.6 nm) (Fig. 1e and Extended Data Fig. 2e). The

HMM and BIC analysis revealed that the positions of the first four intermediates, I_{f1} to I_{f4} , remained largely invariant (Fig. 1d,e) and that there were two intermediates (I_{f5} and I_{f6}) in the extension space newly charted by the use of 100 %-DMPG bicelles (Fig. 1e and Extended Data Fig. 2b,e). Notably, the final 27 % of the folding progress, corresponding to an extension decrease of about 13 nm (from I_{f6} to N), remained as an intractable barrier to reaching full folding of single GLUT3 (Fig. 1e).

Mapping the folding order of single GLUT3 domains

To map the observed folding progress to specific domains of GLUT3, we constructed a variant of GLUT3 with two mutations: S265C and A469C (referred to as GLUT3_{CC}) (Fig. 2a, inset and Supplementary Fig. 2). The introduced cysteines formed a disulfide bond that knotted the entire C-domain, rendering it as one fixed unit in our mechanical interrogation (Fig. 2a and Extended Data Fig. 3a-d). When examining the folding traces of GLUT3_{CC} obtained at 5 pN, we observed almost identical extents of extension decreases for bicelle conditions with either 30 or 100 mol% DMPG lipids (Fig. 2b,c and Extended Data Fig. 3e,f). This was in sharp contrast with the observation for wild-type (WT) GLUT3 where use of 100 mol% DMPG doubled the folding progress. The last gap before the native state, which slightly shrank to ~12 nm, persisted for both bicelle compositions. Using HMM and BIC analyses, we found four intermediates as the maximum likelihood estimation for the extension traces obtained for GLUT3_{CC} (Fig. 2b,c and Extended Data Fig. 3g). The positions and transition kinetics of these four intermediates were largely identical to those of the first four intermediates observed for WT GLUT3 (Fig. 2d,e). The folding step sizes for, I_{f1} and I_{f4} , however, became notably different, which was likely due to the presence of the folded C-domain in GLUT3_{CC} (Fig. 2d).

Based on these results, we propose that the first four intermediates (*i.e.*, I_{f1} to I_{f4}) correspond to folding of GLUT3's N-domain. The following two folding intermediates (I_{f5} and I_{f6}), which could be accessed in the DMPG-100 mol% condition for WT GLUT3 but vanished for GLUT3_{CC}, are attributed to C-domain folding. To determine whether there were indeed partial structures formed in individual intermediates, we applied a force jump to 25 pN when WT GLUT3 reflected I_{f6} (the last intermediate before the 13 nm gap). We found a partially folded structure prior to unfolding with a large step size of ~94 nm (Fig. 2f,g), closely matching what would be theoretically expected for unfolding of both N- and C-domains (but separated) (Fig. 2f,g and Supplementary Fig. 3). When we applied the same force jumps to I_{f4} of WT GLUT3 and GLUT3_{CC}, we found a partially folded structure with an unfolding step of 45 nm under 25 pN, an expected value for N-domain unfolding (Fig. 2f,h,i). These results support our assignment of I_{f4} and I_{f6} to completion of N- and C-domain folding respectively. They also suggest that the remaining ICH domains are responsible for the tenacious 13 nm gap as a blockade to reaching the N state.

Dissecting folding steps of the MFS folds

We next attempted to dissect more detailed folding steps within individual N- and C-domains. To this end, we conducted force jump experiments for WT GLUT3 and GLUT3_{CC} multiple times commencing from the native folded state, and collected all extension values reflected before reaching U_c (Fig. 3a). With 50 Hz median filtering applied, the unfolding

extensions displayed clearly peaked distributions, each of which we assigned as a high-force unfolding intermediate (Fig. 3b,c). In addition, as demonstrated for I_{f4} and I_{f6} in Fig. 2g-i, we applied force jump to each intermediate observed during folding trials at 5 pN. This series of experiments permitted establishment of a crucial one-to-one correspondence between the low-force and the high-force intermediates (Fig. 3b,d). For instance, in the case of WT GLUT3, the first four intermediates observed at 5 pN (I_{f1} to I_{f4}) were mapped to the last four unfolding peaks positioned before U_c (Fig. 3b,d). In addition, the positions of these peaks reasonably coincided with those of the last four peaks determined for GLUT3_{CC} (except for one small peak in the middle) (Fig. 3b,c). These results support our assignment of the early intermediates (I_{f1} to I_{f4}) as corresponding to folding of the N-domain, which appears after unfolding of the ICH and C-domains during the high-force unfolding.

To dissect the detailed folding/unfolding order within individual MFS folds, we first focused on the C-domain that showed only two dominant intermediates of I_{f5} and I_{f6} during the 5 pN re-folding process (e.g., Fig. 1e and 3b). In such two-step folding, the folding process should start from either N- or C-terminus of the C-domain. Otherwise, a partially folded structure at I_{f5} would have flanking N- and C-terminal tails, requiring more than one steps to finish C-domain folding and thus incompatible with the observed two-step folding. Furthermore, the unfolding step from I_{f5} to I_{f4} corresponded to two third of the unfolding extension of the entire C-domain, suggesting that the folding step in the reverse direction (i.e., from I_{f4} to I_{f5}) would involve four out of six TMHs of the C-domain (Fig. 3b). Moreover, inspection of the C-domain structure indicates that TMH 7 is flanked by TMHs 11 and 12, a topological constraint that would force folding of TMHs 11 and 12 only after that of TMH 7 (Fig. 3e). The scenario meeting all these requirements is that TMHs 7 to 10 first fold together (I_{f4} -to- I_{f5} transition), with TMHs 11 and 12 making a helical hairpin to complete C-domain folding (I_{f5} -to- I_{f6} transition) (Fig. 3e).

Given the remarkable pseudo-symmetry of the N- and C-domains of GLUT3^{8,17,31,32}, we assumed that a similar pathway guides folding of the N-domain. Indeed, we found a similar TMH topology for the N-domain, with TMHs 5 and 6 embracing TMH 1 (Fig. 3f), which would allocate TMHs 5 and 6 as the last structural unit in N-domain folding. For the partial structure composed of TMHs 1 to 4, TMHs 1 and 2 in turn wrap around TMH 4 while making multiple atomic contacts among them, which likely renders TMHs 1 and 2 tailing TMH 4 in the folding order (Fig. 3f and Extended Data Fig. 4a). To examine this hypothesis, we generated another GLUT3 mutant harboring T45C and K115C (GLUT3_{TM23C}), in which TMHs 2 and 3 were knotted together by the disulfide linkage (Extended Data Fig. 4b,c and Supplementary Fig. 2). Comparing the high-force unfolding pattern of GLUT3_{TM23C} with that of WT GLUT3 affirmed a scenario of four step folding of N-domain (Fig. 3g,h and see Extended Data Fig. 4d,e and 5 for details), in which TMHs 3 and 4 form the first helical hairpin (U_h to I_{f1}), followed by sequential addition of TMHs 2 and 1 to the structure (I_{f1} to I_{f2} and I_{f2} to I_{f3} each) and completed with addition of TMHs 5 and 6 (I_{f3} to I_{f4}) (Fig. 3f).

EMC facilitates insertion of hydrophilic TMHs of GLUT3

Our observations indicated that GLUT3 has a weaker propensity for folding than GlpG and β 2AR and thus requires assistance by, for example, addition of more negatively charged

lipids to the bilayer. We sought to find a more physiological alternative mechanism that might assist GLUT3 folding. We turned our attention to EMC³³, a large multi-protein complex with 9 members in humans³⁴ (Extended Data Fig. 6a,b). EMC is shown to induce effective membrane insertion of tail-anchored proteins and the first TMHs of G-protein-coupled receptors^{35,36}. Specifically, this membrane insertase activity is manifested when TMHs of target proteins exhibit lower levels of hydrophobicity³⁵⁻³⁷.

We purified human EMC and added the complex reconstituted in bicelles to our single-molecule magnetic tweezers assay (Fig. 4a and Extended Data Fig. 6c). We anticipated that EMCs could be delivered to tweezed single GLUT3 proteins because individual bicelles undergo frequent fusion and fission with one another³⁸. Indeed, when adding 500 nM EMCs to the single GLUT3 folding assay, which corresponded to approximately one EMC in each bicelle (Fig 4a, inset), we observed remarkable facilitation of GLUT3 folding under the 30 mol% DMPG condition. Many of single GLUT3 folding traces progressed as far as ~ 34.7 nm, a direct indication of EMC contribution to successful folding of the entire N- and C-domains (Fig. 4b and Extended Data Fig. 6d). This stimulation of folding progression virtually disappeared when we added an unrelated membrane protein reconstituted in bicelles, indicating the specificity of the EMC results (Extended Data Fig. 6e-g). Indeed, when assessing the time required to first reach the extension value of 17.2 nm (corresponding to I_{f4}), this first-passage time was increasingly shortened as a higher EMC concentration was used (Fig. 4c). Using the HMM and BIC analysis, we analyzed patterns in the folding traces and found that the number and positions of the intermediates were essentially preserved in the presence of EMC, an observation recapitulated for C-domain-knotted GLUT3_{CC} (Fig. 4b,d and Extended Data Fig. 6d,h,i). These observations suggest that EMC helps GLUT3 navigate down the folding intermediates encoded in its native amino acid sequence, rather than creating novel folding pathways.

To examine EMC's effects on GLUT3 folding with higher resolution, we deconvoluted the extension distributions to remove some of the broadening effects caused by Brownian noises from the magnetic beads and DNA handles (Fig. 4e and Supplementary Fig. 4). The resulting extension distribution clearly showed populations markedly increased beyond I_{f4} , indicating that EMC indeed helped GLUT3 sample microstates for C-domain folding (Fig. 4e, red vs. black distributions). Comparing the relative populations allowed us to estimate lowering of the free energy landscape by $\sim 4 k_B T$ at I_{f4} in the presence of EMC (Supplementary Fig. 5). In addition, we note a major valley in the extension distribution at around 25 nm, a major setback for GLUT3's efforts in C-domain folding, which was also observed with the 100 mol%-DMPG condition (Fig. 4e, red vs. blue distributions). Remarkably, this valley approximately coincides with the folding steps of TMHs 7 (a broken helix) and 11 that are estimated to confer the highest energetic costs for TMH insertion (Fig. 4e, upper versus lower panels and Supplementary Fig. 1). We found that EMC successfully propelled single GLUT3s through these barriers to reach I_{f6} (Fig. 4e). Thus, our observation suggests that EMC helps TMH insertion for GLUT3 beyond its first TMHs, which becomes most accentuated for TMHs with low scales of hydrophobicity.

To test a physiological relevance of our observations, we adopted the dual-color ratiometric assay that examines membrane protein stabilities in the cellular milieu (Fig. 4f)^{35,36}. All

partial constructs of N-domain, even that missing only TMH 6 (*i.e.*, TMHs 1 to 5), showed negligible expression signals relative to an expression benchmark (Fig. 4g). While the complete N-domain construct showed an increased stability compared to the truncated N-domain constructs (Fig. 4g and Supplementary Fig. 6a), the construct expressing the entire C-domain failed to be stably expressed. We further found that the stability levels of the full-length and the complete N-domain constructs were mitigated in EMC-knock out (KO) cells, suggesting that the biogenesis of GLUT3 was indeed dependent on EMC in the cellular milieu (Extended Data Fig. 7 and Supplementary Fig. 6b,c). Our cellular folding data revealed that the C-domain could not fold by itself but required the N-domain, a result aligned with the hierarchical order in N- and C-domain folding observed in the magnetic tweezer experiments. Lastly, we noticed that even with EMC, stretched ICH domains failed to fold, observed as the persistent 13 nm gap before the *N*state (Fig. 4b,d and Extended Data Fig. 6d).

PE lipids boost domain-domain assembly of GLUT3

We next asked whether we could induce assembly between the N- and C-domains to complete the known tertiary structure of GLUT3 (Fig. 5a). Since neither negatively charged lipids nor EMC could facilitate domain-domain assembly, we propose that there exists a high energy barrier that arises from a molecular mechanism distinct from poor TMH insertion. To gain insights into this late stage of folding, we employed molecular dynamics (MD) simulations, in which GLUT3 was embedded in lipid bilayers with different lipid compositions and simulated for 1.0 μ s using the CHARMM force field³⁹. The MD simulation results suggest that the high content of polar/charged residues on the interface between N- and C-domains induce considerable distortions in the surrounding bilayer structure as well as increased penetration of water molecules (Fig. 5b-d, Extended Data Fig. 8 and Supplementary Fig. 7)⁴⁰. We reasoned that these structurally distorted lipid shells and penetrated water molecules need to be removed to expose the interfaces for domain-domain assembly, analogous to dehydration of water molecules before binding between soluble proteins⁴¹.

We further reasoned that if the membrane shells between the N- and C-domains indeed define a major barrier to domain-domain assembly, the lipid bilayer composition might play a pivotal role in the final step of GLUT3 folding^{42,43}. Because negatively charged lipids were not effective for this purpose (Fig. 1e), we tested 1,2-dimyristoyl-*sn*-glycero-3-phosphoethanolamine (DMPE) lipids at 15 mol% in bicelles. Strikingly, the presence of DMPE lipids not only induced C-domain folding but also facilitated domain-domain assembly, making the extension traces finally cross the 13-nm gap to reach the native folded state (Fig. 5e,f and Extended Data Fig. 9a,b). In our MD simulations, the frequencies by which the polar/charged residues contacted either water or lipid headgroup markedly decreased with the inclusion of PE lipids (Fig. 5b-d and Extended Data Fig. 8). Together, these observations corroborated our notion that PE lipids ease membrane remodeling, an effect more pronounced when polar/charged residues on the domain interfaces are exposed during membrane protein biogenesis.

We next added EMC to see whether there was synergy between the effects of DMPE lipids and EMC (Fig. 5g and Extended Data Fig. 9c,d). Under the DMPG-30 mol% condition, folding probability after waiting for 500 s at 1 pN was as low as 5.4 %, re-confirming that GLUT3 was not competent for folding by itself (Fig. 5h,i and Supplementary Fig. 8). Either addition of EMC alone (7.6 %) or switching to DMPG 100 mol% (13.2 %) marginally increased the folding probability, consistent with our observations that these conditions facilitated C-terminal domain folding, but not the domain-domain assembly (Fig. 5i). Addition of DMPE lipids increased the folding probability to ~30 %, and addition of both DMPE and EMC further increased the probability to 60 % (Fig. 5i), pointing to a strong synergy between EMC and DMPE lipids.

Asymmetric TMH distributions of metazoan sugar transporters

Finally, we examined whether our observations made for human GLUT3 hold for other sugar transporters that exist across all domains of life¹³⁻¹⁵. We investigated the TMH-insertion energy values for 143 transporters in the sugar porter family (Fig. 6a). To this end, we searched for potential TMH regions in these transporters by comparing their sequences with those of the reference transporters that have high-resolution structures and thus exact, known locations of TMH regions⁴⁴. We then estimated the insertion energy values of putative TMHs using the biological hydrophobicity scale (Fig. 6b,c)²⁶.

Remarkably, when we compare the average of the three highest insertion energy values (out of six), the metazoan sugar transporters exhibited a marked asymmetry in which TMHs of the C-domain had higher insertion energy values than those of the N-domain, a pattern that did not hold true for those bacterial transporters (Fig. 6d). In addition, the C-domains (but not the N-domains) of metazoan sugar transporters showed far larger variances among their six constituting TMHs (Fig. 6e and Extended Data Fig. 10a-d), reminiscent of our finding that TMH 7 and TMH 11 in GLUT3's C-domain have particularly high insertion energies. This asymmetric pattern vanished again for the bacterial sugar porters (Fig. 6b,c,e). We also studied the TMH-insertion energy distributions for other clades and found a similar level of asymmetry between the N- and C-domains for plant sugar transporters, but not for fungal proteins (Extended Data Fig. 10e-g).

Given that all transporters in the sugar porter family are assumed to have sugar binding sites in the C-domain¹⁶, we wondered whether sampling of these more hydrophilic TMHs in the C-domains could be coupled to enhancement of transporter function. Indeed, we found that the QLS motif in TMH 7, which plays a crucial role in improving the selectivity of sugar binding⁴⁵, is identified only for a subset of metazoan sugar transporters that are most close to GLUT3 in our phylogenetic analysis (Fig. 6f). The presence of the QLS motif increases the insertion energy of TMH 7 by ~1 kcal/mol, strongly coupled to higher TMH-insertion energy values of the entire C-domains (Fig. 6g,h).

Discussion

Our single-molecule data reveal the complete folding pathways of a human glucose transporter, allowing us to identify critical setbacks along the pathways and understand how cells remedy these obstacles to promote membrane protein biogenesis. At a resolution

of a few amino acids, our data dissect orders for weaving individual MFS folds. Given the high level of structural conservation^{5,8,16}, we expect the folding order described here to be shared by many transporters belonging to the MFS. We further note that the revealed folding pathway is compatible with the C2 pseudo-symmetry inherent to the MFS fold, allowing us to group the first- and last-folding three TMHs as helix triplets, respectively (Extended Data Fig. 10h). This new grouping – based on the resolved folding order – generates fully-packed, stable helix triplets that can be simply adjoined together to form the MFS fold without involving complicated topological entanglement (Extended Data Fig. 10i versus 10j). Symmetry is a prevailing feature in the conserved structural folds of membrane proteins^{8,32}. Our results may represent an example of the general principle that the folding pathways of membrane proteins have evolved to be commensurate with their symmetry properties, a natural requirement to build such structures of high symmetry.

On the domain level, folding of the N-domain strictly precedes that of the C-domain, which likely mirrors or leverages a co-translational folding pathway in cells^{11,46}. Our cell-based folding/stability assay further revealed that the C-domain that forms most of the glucose binding site does not fold by itself, suggesting that the N-domain likely serves as a structural template for C-domain folding. Still, a failure in C-domain folding would put the fate of the entire protein at risk, including that of the successfully folded N-domain, raising the question of why GLUT3 has a connected structure despite such disadvantages for folding. Primordial transporters before gene duplication or fusion – missing in the current MFS^{14,15} – might have formed homo- or heterodimers, in which both foldability and functionality would be managed within a single subunit⁴⁷. In this vein, the domain structure of GLUT3 can be viewed as specialization of each domain in its role, with the C-domain contributing the functional requirement (but becoming less foldable) and the N-domain becoming the primary driver of folding and structural stability.

Our bioinformatics analysis suggests that the metazoan sugar transporters have most proactively taken this evolutionary venture through sequence space to sample more unstable TMHs in their C-domains. The outcome of less hydrophobic TMHs in metazoans may be aligned with improved performance in the transporting function. In addition to the example of acquiring the QLS motif, the metazoan sugar transporters seem to have implemented versatility in transporter functions with widely differing Michaelis constants (K_M) and catalytic rates (k_{cat}). For example, while GLUT3 is mainly expressed in neurons and transports glucose molecules with a high turnover rate ($k_{cat} > 1,000 \text{ s}^{-1}$)⁴⁸, GLUT2 is expressed in beta cells and mainly works as a glucose sensor with its uniquely high K_M ^{49,50}. Finally, we note different families in MFS have different domain structures. It is thus an open question whether our findings – the N-to-C hierarchical folding pathway and the evolutionary development of insertion-energy asymmetry – are generally observed beyond the sugar porter family.

Thus, our data collectively point to evolutionary conflicts between functionality and foldability faced by many of the metazoan sugar transporters. The resulting evolutionary pressure might have driven the ER membranes of these metazoan cells to be equipped with accessory machineries (e.g. EMC) and distinct lipid compositions that work in concert to help such poor-folding multi-pass membrane proteins. Recent studies suggest that EMC,

along with YidC⁵¹, GET1⁵² and TMCO1⁵³, belong to the Oxa1 superfamily that makes a remarkably conserved family of insertases⁵⁴. We indeed found that even with the bicelle membranes that have lower energy barriers for protein insertion than true lipid bilayers, most TMHs of GLUT3 still need to be assisted by EMC for their efficient membrane insertion, corroborating the notion that the membrane insertion steps do present considerable energy barriers during folding of these transporter proteins. While the PE-headgroup lipids are known to affect TMH orientations and thus establishment of a right topology of TMHs⁵⁵, our results suggest a novel role of the PE lipids – and presumably other lipid species with specific geometric curvatures – during a later stage of membrane protein folding. The presence of PE lipids facilitates removal of lipid shells from the domain-domain (or subunit-subunit) interfaces and assembly of higher-order membrane protein structures. This observation is intriguing because it provides a glimpse into how two biogenesis processes in the ER membrane – membrane protein biogenesis and lipid synthesis – are intricately intertwined with one another^{56,57}.

Methods

Expression and purification of the human GLUT3

For single molecule assays, the GLUT3 glycosylation site N43 was deleted by mutating to Threonine (Thr, T). GLUT3 N43T is referred to as wild type (WT) GLUT3 throughout this work unless otherwise specified. To develop C-domain knotted GLUT3, we mutated S265 and A469 to Cysteine (Cys) based on the structure (PDB:4ZWB, PyMol 2.3 was used for analyzing structure of GLUT3). Likewise, for N-domain knotted GLUT3, we mutated T45 and K115 to Cys. GLUT3 was tagged with Spytag on the N-terminus and Spytag-HRV3C-GFP-10xHis on the C-terminus. GLUT3 WT, GLUT3 S265C/A469C, and GLUT3 T45C/K115C were expressed and purified as described previously¹¹ with minor modification. Briefly, GLUT3 was cloned to a modified pFastBac vector and each virus was made using Bac-to-Bac system (Invitrogen). Virus was added when *Spodoptera frugiperda* (Sf9) cells reached a density of approximately 3.0×10^5 cells/ml. Cells were harvested after 48 hrs and stored at -80°C. Lysis was done with hypotonic buffer (20 mM HEPES 8.0, 1 mM EDTA, 1 mM PMSF) and membrane fractions were collected through centrifugation. Solubilization was done with 20 mM HEPES 8.0, 150 mM NaCl, 1% DDM, 0.5% CHS, 1 mM PMSF at 4 degree for 1 hr. Solubilized GLUT3 was removed from the insoluble fraction and was bound to Ni-NTA resin (Qiagen) at 4 degree for 1 hr. Resin was washed with high salt buffer (20 mM HEPES 8.0, 1 M NaCl, 20 mM imidazole, 0.05% DDM, 0.0025% CHS) and low salt buffer (20 mM HEPES 8.0, 150 mM NaCl, 30 mM imidazole, 0.05% DDM, 0.0025% CHS) sequentially. Elution was done using low salt buffer with 300 mM imidazole. GFP-10xHis tag was cleaved with home-made HRC3V protease at 4 degree. Uncut product and cleaved GFP tag were removed using home-made GFP nanobody column. GLUT3 was finally purified with size exclusion chromatography (GE healthcare) equilibrated with 20 mM HEPES 8.0, 150 mM NaCl, 0.03% DDM, 0.015% CHS.

To assess whether two cysteines in the mutant GLUT3 formed disulfide-bonds, we used labeling with BODIPY-L-Cysteine. BODIPY-L-Cysteine (Invitrogen) becomes fluorescent when their inter-BODIPY disulfide bonds are replaced by bonds with cysteine residues

exposed on protein surfaces (Extended Data Fig. 3a). We determined the melting temperature of GLUT3 WT, GLUT3 S265C/A469C and GLUT3 T45C/K115C by measuring increases in the BODIPY fluorescence signals, which indicated melting of tertiary structures of the GLUT3 constructs and exposition of cysteine residues to aqueous buffer spaces to allow for BODIPY labeling. 4 μ g of GLUT3 was reacted with 5 μ M BODIPY-L-Cysteine (Invitrogen) using Rotor-Gene Q Thermocycler (Qiagen). Temperature was increased from 25°C to 95°C in 1 degree increment with an 10-second interval between each step. To identify the S265C/A469C disulfide bond using gel electrophoresis, a previously described protocol was used⁵⁸ with slight modifications. 4 μ g of GLUT3 was reacted with 1 mM tris(2-carboxyethyl) phosphine (TCEP) for 15 min. 5 μ M BODIPY-L-Cysteine was added and reacted for 30 min in the dark, followed by addition of 5x SDS buffer. All reactions were done at room temperature (RT). The extent by which GLUT3 was labeled with BODIPY was assessed via SDS-PAGE gel followed by imaging using ChemiDoc XRS+ (Bio-Rad). Dye signals were quantified using ImageJ.

Expression and purification of the human EMC

The human ER membrane protein complex (EMC) was prepared essentially as described previously⁵⁹, but using n-Dodecyl-beta-Maltoside (DDM, Anatrace) instead of Lauryl Maltose Neopentyl Glycol (LMNG), with a concentration of 0.2% DDM in all early stage wash buffers and 0.02% DDM in all buffers following FLAG elution. Glycerol was added to a final concentration of 10% before snap freezing in liquid nitrogen. Freeze-thawed aliquots were checked by nanoDSF using a Prometheus NT.48 and had melting curves similar to samples that had not been freeze-thawed.

Expression and purification of the human β_2 AR

This β_2 AR construct (residues 25-365) was expressed and purified as described previously¹¹. β_2 AR was expressed in Sf9 (*Spodoptera frugiperda*) cells using the Bac-to-Bac system (Invitrogen). 10 g of cells expressing β_2 AR were lysed with hypotonic buffer (10 mM HEPES (pH 7.5) and 1 mM EDTA) and β_2 AR was extracted from the cell membrane using 50 ml solubilization buffer consisting of 20 mM HEPES (pH 7.5), 150 mM NaCl, 1% (w/v) n-dodecyl- β -D-maltoside (DDM), 0.1% (w/v) cholesterol hemisuccinate (CHS), and protease inhibitors (benzamidine and leupeptin). The solubilized fraction was loaded onto 8 ml Ni-NTA column and the protein was eluted with 3-4 column volumes of elution buffer including 20 mM HEPES (pH 7.5), 150 mM NaCl, 250 mM imidazole, 0.05% DDM, 0.005% CHS after extensive column washing with 10 column volumes of wash buffer (20 mM HEPES (pH 7.5), 0.5 M NaCl, 20 mM imidazole, 0.05% DDM, 0.005% CHS). The eluted protein was loaded onto 3 ml of second affinity column, an M1 FLAG resin. The column was washed with 10 column volumes of wash buffer including 20 mM HEPES (pH 7.5), 500 mM NaCl, 4 mM CaCl₂, 0.05% DDM, and 0.005% CHS and the protein was eluted.

Preparation of DNA handles

The SpyCatcher proteins covalently linked to DNA handles were prepared following a previously established procedure²². In short, an amine group at one end of the 512 bp DNA fragment made by PCR was reacted with SM(PEG)₂ (PEGylated SMCC crosslinker;

ThermoFisher Scientific) using an amine-sulphydryl crosslinker for 30 min at RT. After purification via DNA maxiprep, DNA fragments labeled with either biotin or digoxigenin at the other end were mixed in 1:1 molar ratio. Mixed DNA fragments were then covalently conjugated to purified SpyCatcher/Maltose Binding Protein (MBP) protein through a thiol-maleimide crosslinking reaction overnight at 4°C. To purify SpyCatcher-conjugated DNA only, anion exchange chromatography using a 1 ml Mono Q column (GE healthcare) and amylose affinity chromatography (New England BioLabs) were used to exclude unconjugated SpyCatcher and unconjugated DNA, respectively. The purified SpyCatcher-DNA handles (in 50 mM Tris pH 7.5 and 150 mM NaCl buffer) were then concentrated up to ~ 100 nM and stored in 10 µl aliquots at - 80°C.

Dynamic light scattering (DLS) measurement

To determine sizes of the bicelles, a dynamic light scattering (DLS) apparatus (Otsuka electronics ELSZ-1000) was used. Bicelles ([lipids]:[CHAPSO] = 2.8:1 in molar ratio) with different composition of lipids (70:30:0, 0:100:0 and 55:30:15 of DMPC:DMPG:DMPE in mol%) were measured at 296K. 2 ml of bicelle buffer (1.3% (w/v) bicelles in 50 mM Tris (pH 7.5), 150 mM NaCl) in a glass-clear polystyrene cuvette (Ratiolab) was placed on the sample stage of DLS analyzer instrument. Data were analyzed using the associated software (Otsuka electronics Photal). Mean diameter of bicelle were obtained as follows. First, we calculated the product of diameter and population probability of each data point in Extended Data Fig. 1d. Then, mean diameter is sum of these products.

Grid preparation and TEM imaging of bicelle

Copper grids coated with a continuous carbon film (CF200-CU, Electron Microscopy Sciences) were used for negative stained imaging. The grids were glow-discharged for 60 sec, and 2.5 µL of bicelle solution was adsorbed. Samples were rinsed by water and stained with 1% uranyl formate. Prepared grids were dried for 5 min in room air and then visualized on a JEM-2100Plus 200kV electron microscope (JEOL) equipped with a Rio 9 CCD camera (Gatan).

Graphene oxide-coated EM grids were used for cryo imaging. Holey carbon grids (Quantifoil R1.2/1.3 200 mesh) were glow-discharged and placed into a small water chamber. Methanol:graphene oxide (9:1) solution was floated on the water surface and the spread layer of graphene oxide was loaded on the grids by slowly venting the water⁶⁰. 2.5µl of bicelle solution was dispensed on the prepared grid and vitrified using Vitrobot IV (ThermoFischer). Vitrified specimens were then visualized by 200kV JEM-2100Plus (JEOL) electron microscope with a Rio 9 camera (Gatan) and 200kV Glacios electron microscope with Falcon 4 camera, respectively. Micrographs were captured at the pixel size of 2.1Å/pixel for JEM-2100Plus and 1.1Å/pixel for Glacios using ~40 e⁻/Å² total dose with defocus range of -3.5~4.5µm. All imaging data were collected at the Center for Macromolecular and Cell Imaging at Seoul National University.

Instrumentation of magnetic tweezers

As established and described previously^{61,62}, a magnetic tweezer instrument was custom built on an inverted microscope (Olympus Live Cell Instrument). The vertical position of a

pair of permanent magnets (Neodymium magnets) was controlled using a translational stage (Physik Instrumente) to generate mechanical forces from ~10 fN to 50 pN. Illumination with a super-luminescent diode ($\lambda = 680$ nm, Qphotonics) generated diffraction patterns for magnetic and reference beads (stuck on surface), of which images were recorded at an acquisition rate up to 1.2 kHz using a high-speed CMOS camera (Mikrotron). Diffraction patterns were pre-recorded by moving an objective lens using a piezoelectric nano-positioner (Mad City Labs) with respect to the sample in order to generate calibration tables for individual beads (magnetic and reference beads both). By comparing diffraction patterns of magnetic beads with the corresponding calibration table in real-time, 3D positions of the magnetic bead were tracked. Custom-written LabView programs were written and used for the single-molecule magnetic tweezers experiments (see Data and code availability section).

Single-molecule magnetic tweezers experiments

Samples for single-molecule magnetic tweezers experiment were prepared as described previously⁶³. In brief, WT or mutant GLUT3 proteins reconstituted in 0.02 to 0.04% of DDM were mixed with the SpyCatcher-DNA handles (with DDM added to a final concentration of 0.1% and TCEP added to 2 mM for WT GLUT3) and incubated for 20 to 22 h at 4°C to attach DNA handles at both ends of the GLUT3 proteins. 10:1 to 20:1 molar ratio for GLUT3 protein: SpyCatcher-DNA handles were used. After incubation, the protein-DNA hybrid complexes were diluted to ~ 1 nM final concentration of DNA using 1.3% (w/v) bicelle buffer (50 mM Tris pH 7.5 and 150 mM NaCl; DDM was thus diluted to below half its CMC). The membrane proteins connected with two DNA handles were then stored in 40 μ l aliquots at -80°C. Bicelles with specific mixtures of DMPC, DMPG and/or DMPE lipids (Avanti polar lipids) and 3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO, Sigma-Aldrich) were prepared with 2.8:1 molar ratio (*i.e.*, $Q \equiv [\text{lipids}]/[\text{detergent}] = 2.8:1$).

For single-molecule magnetic tweezers experiments, 4 μ l of 0.01 mg/ml neutravidin (NTV) was added to 40 μ l of the sample and incubated for 5 min at RT. After binding NTV to one end of the DNA handle, the sample was then further diluted to a final concentration of ~ 500 pM. We first injected 0.02% (w/v) streptavidin-coated polystyrene particles (3.11 μ m, Spherotech, *i.e.*, reference bead) into a home-made flow-cell consisting of two cover slips (VWR No 1.5). The bottom cover slip was coated with mPEG and biotin-PEG at 100:3 molar ratio. After 5 min incubation, unbound reference beads were removed by extensive microfluidic buffer exchange. The final sample was injected and incubated for 10 min. After washing with bicelle buffer to remove unbound samples, anti-digoxigenin coated magnetic beads (2.8 μ m diameter, Invitrogen for magnetic beads and Roche for antibody) were injected and incubated for 30 min after 100 times dilution. For EMC studies, EMC reconstituted in bicelle (300 - 600 nM) was additionally injected.

Force-extension curves (FEC) analysis

The FECs for DNA and unstructured polypeptide were fitted with the extensible worm-like chain (eWLC) model that describes behavior of the semi-flexible biopolymers under tension⁶⁴.

$$F = \left(\frac{k_B T}{l_{p,i}} \right) \left[\frac{1}{4(1 - z_i/(n_i l_{c,i}))} - \frac{1}{4} + \frac{z_i}{n_i l_{c,i}} - \frac{F}{K_i} + \sum_{j=1}^7 a_j \left(\frac{z_i}{n_i l_{c,i}} - \frac{F}{K_i} \right)^j \right],$$

where the index i indicates either DNA or unstructured polypeptide (p), $k_B T$ is the thermal energy, l_c and l_p are the contour length and persistence length, respectively ($l_{c,DNA} = 0.338$ nm, $l_{c,p} = 0.36$ nm and $l_{p,DNA} = 38.5$ nm, $l_{p,p} = 0.39$ nm)⁶⁵⁻⁶⁷. K_j is the elastic modulus ($K_p \sim 50$ μ N and $k_{DNA} \sim 500$ pN)^{62,68}, F is the applied force and a_j are polynomial coefficients for the improved approximation. n_i is the total number of constituent monomers of each component such as DNA and polypeptide ($n_{DNA} = 512$ for each handle, $n_{linker,p} = 18$ between the GLUT3 and DNA handle, and $n_{GLUT3,p} = 463$ (i.e., $n_{GLUT3,p} = n_{N-domain,p} + n_{C-domain,p} + n_{ICH-domain,p} = 198 + 207 + 58$) for GLUT3¹⁷).

To describe a rigid-like biopolymer such as helical states (U_h), the Kessler-Rabin (KR) model was used^{61,67},

$$Z_h \approx -\frac{1}{2f} - \frac{\chi}{f \tanh 2\chi} + \frac{L_h}{\tanh 2\chi} - \frac{2\chi^2}{3f} \left(\frac{1}{\tanh f L_h} - \frac{f L_h}{(\sinh f L_h)^2} - 1 \right),$$

where $\chi \equiv \sqrt{f \frac{(n_h l_{c,h})^2}{4l_{p,h}}}$, $f \equiv \frac{F}{k_B T}$ and n_h is the number of amino acids consisting of the transmembrane helix. The persistence length ($l_{p,h}$) is 9.17 nm and the contour length ($l_{c,h}$) along helical axis is averagely 0.16 nm per amino acid.

In the force-ramp and force-jump experiments, observed extension values can be estimated from a linear superposition of extensions of all components in tweezing system. The fully unstructured coil state (U_c) and helical state (U_h) are thus described as follows.

$$z_m = z^{\text{handle}} + z^{\text{GLUT3}} = (z_p + z_{DNA})^{\text{handle}} + z_{p \text{ or } h}^{\text{GLUT3}}$$

where z_m is measured extension, z_p is the extension of the unstructured polypeptide linker between DNA and target protein (linkers from each end of the protein to SpyCatcher), z_{DNA} is the extension of the DNA handle, and $z_{p \text{ or } h}^{\text{GLUT3}}$ is the total molecular extension of GLUT3 with contributions from unstructured and/or helical parts. The z_p and z_{DNA} values are inversely calculated from the eWLC model at given force levels, and $z_{p \text{ or } h}^{\text{GLUT3}}$ are calculated from the eWLC or KR model, respectively. In the case of stretching GLUT3 in its native state (N), $z_{p \text{ or } h}^{\text{GLUT3}}$ is replaced by a d_N value of 3.9 nm, an end-to-end distance determined from the native state structure (PDB:4ZW9)¹⁷.

To analyze relative extension changes during high-force unfolding, we treated N- and C-domains independently because two domains are separated by the ICH-domains. We used the relation that the extension increase observed for an intermediate state ($z_{i,p}$) is proportional to the number of unfolded amino acids (n_i), giving

$\frac{z_{i,p}}{n_{N(C)-domain,p}} = \frac{\Delta n_i}{n_{N(C)-domain,p}}$, where $n_{N(C)-domain,p}$ is the extension increase expected when N- or C-domain is fully unraveled and $n_{N(C)-domain,p}$ is the total number of amino acids in N- or C-domain, respectively. Because the remaining partially folded structures have finite thickness values along the pulling axis (d_i), we used the relation of $z_{m,i} = z_{i,p} + d_i - d_{N(C)-domain}$ for N(C)-domain. $d_{N-domain}$ ($d_{C-domain}$) is the initial thickness of the fully folded N- or C-domain, determined to be 1.3 (0.7) nm. By using first-order approximation, a recurrence relation can be derived as $\Delta n_i \approx \frac{n_{N(C)-domain,p}}{z_{N(C)-domain,p}} (z_{m,i} - d_i + d_{N(C)-domain})$. The intersection between functions in the left-hand side and right-hand side yields the number of amino acids from the reference point where unfolding starts. Fixed rate of 1 pN/s was applied for FEC obtained by force-ramp experiment.

Hidden Markov Model analysis

Hidden Markov Model (HMM) analysis was employed to determine the folding/unfolding intermediate states from the time-resolved low-force extension traces recorded at 1.2 kHz³⁰. The adjustable parameters in our system are the number of states (n), the extension position for i -th intermediate state (\bar{Z}_i), and the transition matrix of rates between states (\vec{k}). The optimal number of states (n) was obtained from Bayesian Information Criterion (BIC): $BIC = q \ln(N) - 2 \ln(L^{\wedge})$ where q is the number of output parameters given by model, N is sample size and L^{\wedge} is the maximum value of the likelihood function. Maximum likelihood estimation was performed using the Baum-Welch algorithm. BIC as a function of the number of states determines the optimal number by finding the point where the BIC slope substantially changes²⁹. The extension traces were median-filtered with 5-Hz window, and the extension position/deviation for each state was estimated from the Gaussian Mixture Model (GMM) in the HMM analysis. The rates (*i.e.*, the transition matrix) were then determined using the optimal parameters for the number of states and extension positions. The rates estimated from HMM were confirmed by checking single exponential fitting of the dwell time distributions. In this process, dwell time data shorter than 50 ms were considered artifact and ignored because we used median-filtered traces (5 Hz or 200 ms). Finally, the resulting traces were verified by the Viterbi algorithm.

Deconvoluted extension probability analysis

To obtain an extension distribution of single GLUT3 protein ($\tilde{P}_p(Z_p; F)$) with Brownian noises of magnetic beads and handles removed, we implemented deconvolution of the measured marginal probability distribution ($\tilde{P}_m(Z_m; F)$) in real space as previously established in optical tweezers studies^{69,70}. Because the magnetic bead in magnetic tweezers is not physically trapped unlike with optical tweezers (*i.e.*, magnetic force is not a fluctuating variable but stably fixed), the marginal probability distribution ($\tilde{P}_m(Z_m; F)$) from Hamiltonian of the bead in the presence of magnetic force could be directly described as $\tilde{P}_m(z_m; F) \approx \frac{1}{4} e^{\beta F z} P_m(z) = \frac{1}{4} P_m(z; F)$ where $P_m(z; F)$ is the measured equilibrium probability of the total bead-handle-protein system with separation z at the constant force \mathcal{F} ; β is $1/k_B T$. By performing deconvolution in real-space, we can derive the following integral.

$$\int dz_p \tilde{P}_{bh}(z_m - z_p; F) \tilde{P}_p(z_p; F) = \tilde{P}_m(z_m; F)$$

Where $\tilde{P}_{bh}(z; F)$ is conjugated probability of handles (PEG polymers (peg), two DNA handles (dh; dh1 defined as DNA handle directing towards magnetic bead, dh2 towards peg) and two polypeptide linkers (ph) between DNA and GLUT3) and magnetic bead. In brief, $\tilde{P}_{bh}(Z; F) \equiv \mathcal{F}^{-1}(\tilde{P}_b(k; F) \tilde{P}_{peg}(k; F) \tilde{P}_{dh1}(k; F) \tilde{P}_{dh2}(k; F) \tilde{P}_{ph}^2(k; F))$ where \mathcal{F}^{-1} indicates inverse Fourier-transformation and k is the wave-vector in Fourier-space. The probability of the magnetic bead, $\tilde{P}_b(k; F)$ is $\frac{f \sinh((f - ik)R_b)}{(f - ik) \sinh(fR_b)}$ where R_b is the radius of the magnetic bead f , βF and i is the complex number. The rest terms in $\tilde{P}_{bh}(Z; F)$ can be described by

$$\tilde{P}_j(k; F) \equiv \frac{\sum_{n,l,l'} \Psi_{l',B,C}^T(L_{c,j}) \left[e^{-E_{n,j}(f-ik)L_{c,j}} \right]_{l',l} \Psi_{l,B,C}^{(0)}}{\sum_{n,l,l'} \Psi_{l',B,C}^T(L_{c,j}) \left[e^{-E_{n,j}L_{c,j}} \right]_{l',l} \Psi_{l,B,C}^{(0)}}$$
 where the index j represents the

components composed of peg, dh1, dh2 and ph. The corresponding total contour length is $L_{c,j}$. $\Psi_{B,C}$ and $E_{n,j}$ are an eigen state and eigen value (total energy), respectively as previously defined and estimated from effective Hamiltonian equation of propagator of biopolymer in Markovian regime⁶⁶. Index B.C in eigen state indicates whether semi-flexible biopolymer is half-constrained (one side of peg and dh1) or unconstrained (dh2, ph).

To avoid any numerical instability and ill-conditioned result, we used suitable fitting functions for all probability distributions (\tilde{P}_{bh} , \tilde{P}_p and \tilde{P}_m). Linear superposition of Gaussians was employed to determine the pure probability of GLUT3 (we used median-filtered traces with 5-Hz window. Because the characteristic time scale of magnetic bead is less than 30 ms, we could implement deconvolution of the behavior of the bead from the measurement).

$$\tilde{P}_\lambda(z_\lambda; F) = \sum_{i=1}^{N_\lambda} w_i^\lambda g(z_\lambda, \mu_i^\lambda, \sigma_i^\lambda)$$

Where λ means bh (handles and bead), p (GLUT3) or m (total system) and

$g(z_\lambda, \mu_i^\lambda, \sigma_i^\lambda)$ is Gaussian distribution $\left((2\pi\sigma_i^{\lambda 2})^{-1/2} e^{-(z_\lambda - \mu_i^\lambda)^2 / 2\sigma_i^{\lambda 2}} \right)$. w_i^λ is weighting

factor in linear combination and N_λ is total number of Gaussian components (for simplicity, $N_{bh} = 1$ was chosen). Then, parameters of the deconvoluted extension

distribution of the single GLUT3 are described as $w_i^p \approx w_i^m$, $\mu_i^p \approx \mu_i^m - \sum_{j=1}^{N_{bh}} w_j^{bh} \mu_j^{bh}$ and

$\sigma_i^{p2} = \sigma_i^{m2} - \sigma_{bh}^{2} - 2cov(z_m, z_{bh})$. For ensemble averages of the deconvoluted probability

distributions, weighted arithmetic mean was used to visualize the average probability

distribution (*i.e.*, $\langle \tilde{P}_p(z_p; F) \rangle \equiv \sum_{m=1}^M a_m \tilde{P}_{p,m}(z_p; F) / \sum_{m=1}^M a_m$ where m is the number of traces, M is the total number of measured traces and a_m is the normalized weighting factor, which depends on sample size in each trace.

First-passage time (FPT) analysis

FPT decreased exponentially as the EMC concentration was increased. This is because EMC's effects on GLUT3 folding can be described by the same framework that applies chemical folding and unfolding of a protein by a denaturant. EMC effects on single GLUT3 folding can thus be described as follows.

$$\tau(C) = \tau(0)e^{-m^{\ddagger}C} + \tau_0$$

Where τ is the first-passage time (s) defined as the shortest dwelling time between U_h and the certain intermediate state, C is the concentration of EMC and τ_0 is the convergent time from experimental accessibility. The kinetic m^{\ddagger} -value is related to the change of accessible surface area for binding, resulting in a transition towards a certain intermediate state. Also, based on the linear extrapolation method, the exponent $m^{\ddagger}C$ could be expressed in terms of free energy as follows⁷¹.

$$m^{\ddagger}C = (\Delta G_C - \Delta G_0)/k_B T$$

Where ΔG_C is the free energy change from U_h to a certain intermediate state in the presence of EMC with concentration C and ΔG_0 is the free energy change from U_h to a certain intermediate state in the absence of EMC.

Constructs for flow cytometry experiments

eGFPs and mCherry are chosen to determine the stability of respective constructs by measuring eGFP fluorescence relative to that of mCherry that served as an expression benchmark. The parent mCherry-P2A-eGFP construct was generated by inserting mCherry (from pmCherry-C1 vector, Clontech), and eGFP (from pEGFP-C1 vector, Clontech) fragments into the pENTR221 vector using Gibson Assembly (New England BioLabs). For mCherry-P2A-eGFP-Protein constructs, GLUT3 N43T and GLUT3 WT (glycosylation site N43 is unmodified) with GSGSGS linker at both ends were amplified by PCR and inserted into the parent mCherry-P2A-eGFP construct using Gibson assembly. The GLUT3 WT construct was generated from an original GLUT3 N43T-containing plasmid using site-directed mutagenesis.

For the truncated GLUT3 WT constructs, coding regions of GLUT3 TMH-1 (residues 2-51), GLUT3 TMH1-2 (residues 2-90), GLUT3 TMH1-3 (residues 2-116), GLUT3 TMH1-4 (residues 2-146), GLUT3 TMH1-5 (residues 2-182) and GLUT3 TMH1-6 (residues 2-208) were amplified via PCR using the original mCherry-P2A-eGFP-GLUT3 N43T and mCherry-P2A-eGFP-GLUT3 WT constructs with the GSGSGS linkers attached at both ends of the inserts. These amplified PCR products were then inserted into the parent mCherry-P2A-eGFP construct by Gibson Assembly. The mCherry-P2A-GLUT3 C-domain (residues 252-496) was generated by the same method.

GLUT3 N43T-P2A-eGFP and mCherry-P2A-GLUT3 N43T constructs were used as controls for FACS compensation. For the GLUT3 N43T-P2A-eGFP construct, vector DNA

segment of the parent mCherry-P2A-eGFP plasmid excluding the mCherry fragment was amplified. GLUT3(N43T; with GSGSGS linker at C-terminal) was amplified with PCR and inserted to the amplified vector DNA segment using Gibson Assembly.

For mCherry-P2A-GLUT3 N43T construct, vector DNA segment of the parent mCherry-P2A-eGFP plasmid excluding the eGFP fragment was amplified. GLUT3(N43T; with GSGSGS linker at N-, and C-terminal) was amplified with PCR, and inserted to the amplified vector DNA segment using Gibson Assembly. All inserts in the pENTR221 vector constructs were transferred to pInducer20 vector by Gateway recombination cloning (Thermo Scientific). pENTR221 and pInducer20 vector plasmids are generous gifts from Dr. Kang (Seoul National University).

Flow cytometry analysis

Flow cytometry experiments were conducted as described previously^{9,35} with modifications. In brief, HEK293T cells (ATCC) were plated in 6-well plates and transiently transfected with appropriate plasmids (0.5 ug/ml) using polyethylenimine (PEI; Sigma) for 24 h. After transfection, cells were treated with 100 ng/ml doxycycline (Sigma) for 24 h. Following doxycycline treatment, cells were trypsinized, then washed in DPBS, and resuspended in 1ml of ice-cold DPBS. Then, cells were passed through a cell strainer (Falcon) and analyzed with the Sony SH800S cell sorter. Live cells were first gated with FSC/BSC, and then gated for mCherry (mCherry-P2A-eGFP-Protein) or eGFP (TRAM2-mCherry-P2A-eGFP). A total of 20,000 cells were analyzed, and data analysis was done using Cell Sorter Software (v2.1.5) and customized MATLAB codes. GFP fluorescence intensity levels were normalized by most probable GFP:mCherry ratio of mCherry-2A-GFP or mCherry-2A-GFP-GLUT3 (Full length).

For the Wild-type and EMC6 knockout Flp-In T-Rex 293 cells³⁶, cells were plated in 6-well plates and transiently transfected with appropriate plasmids (0.4 ug/ml) using TransIT-293 (Mirus) for 24 h. After transfection, cells were treated with 100 ng/ml doxycycline (Sigma) for 18 h. The wild-type and EMC6 knockout Flp-In T-Rex 293 cell lines are generous gifts from Dr. Christianson (Oxford University). The expression of EMC5 and EMC6 were double-checked by western blot (EMC5 & EMC6 antibodies: abcam), using β -tubulin (antibody: Cell Signaling Technology) as a loading control.

Molecular dynamics simulations

The simulation systems were prepared using CHARMM-GUI *Membrane Builder*³⁹ with a crystal structure of GLUT3 (PDB:4ZWC)¹⁷. The structures for the N-terminal (residues 3 to 205) and C-terminal (residues 264 to 470) domains were extracted from the full-length structure. To mimic experiments, the proteins were embedded in mixed bilayers of DMPC and DMPG (molar ratio of 7:3) or DMPC, DMPG, and DMPE (molar ratio of 5.5:3.0:1.5) solvated with bulk water and 150 mM NaCl at $T = 296.15$ K. Because the mixed bilayers with d14:0 tails are close to their phase transition temperature (e.g., $T_m \sim 23.5$ °C for DMPC), we prepared additional mixed bilayers of DMPC, DMPG, or DMPE at a higher temperature, $T = 306.15$ K to examine temperature effects. In addition, to explore the effect of tail saturation, we also prepared mixed bilayers composed of palmitoylloleoyl

(PO) PC, POPG, and POPE at $T = 296.15$ K. The molar ratios of POPC:POPG and POPC:POPG:POPG in bilayers were set to be the same as those for DMPC/DMPG and DMPC/DMPG/DMPE bilayers, respectively. To ensure sufficient number of lipid shells around each domain or the full-length structure, the initial xy -dimensions of bilayers (i.e., the bilayer surface area) were set to be $\sim 130 \times 130 \text{ \AA}^2$ (N- and C-domains) and $\sim 150 \times 150 \text{ \AA}^2$ (full-length GLUT3), respectively. Each system was subjected to 0.5 or 1.0 μs production run following a series of short equilibration runs. All simulations were carried out using OpenMM⁷² with the CHARMM36 force fields^{73,74} and TIP3P water model⁷⁵. The integration time step was set to 4 fs with the SHAKE algorithm⁷⁶ and hydrogen mass repartitioning method⁷⁷ during production runs. Lennard-Jones interactions were switched off over 10-12 \AA by a force-based switching function⁷⁸ and the electrostatic interactions were calculated by the particle-mesh Ewald method⁷⁹ with a mesh size of $\sim 1 \text{ \AA}$. Temperature and pressure (1 bar) were controlled by Langevin dynamics⁸⁰ with a friction coefficient of 1 ps^{-1} and a semi-isotropic Monte Carlo barostat⁸¹ with a pressure coupling frequency of 100 steps in OpenMM simulations. Trajectories were analyzed using CHARMM⁸² and in-house PYTHON scripts, where the interaction frequency between the polar/charged residues and their environments and the number of contacting water to these residues were calculated with 4.5 \AA heavy-atom distance criterion. The snapshots from simulations trajectories were prepared using VMD⁸³.

Determination of helix insertion energy

We compared the helix insertion energies into the endoplasmic reticulum membrane across kingdoms using DGpred²⁶ for a 19 residue window in the center of the helix. Since only a few sugar transporters have a determined structure available in the PDB, we transferred its helix annotation to additional 138 sequences using alignments. All entries from the sugar porter families (TC 2.A.1.1) without any known structures were selected from the TCDB database⁸⁴. We collected the entries which had currently available UniProt IDs. The phylogenetic tree was generated by Dendroscope.

As reference sequences, we used the following sugar transporter sequences with corresponding UniProt ID: Homo sapiens Solute carrier family 2 facilitated glucose transporter member 3 (P11169), Escherichia coli D-xylose-proton symporter (P0AGF4), Plasmodium falciparum Hexose transporter 1 (O97467), Staphylococcus epidermidis Glucose transporter (A0A0H2VG78) and Arabidopsis thaliana Sugar transport protein 10 (Q9LT15). The PDB accession number for structures for these reference proteins are 4ZWC (P11169), 4GBZ (P0AGF4), 6RW3 (O97467), 4LDS (A0A0H2VG78) and 6H7D (Q9LT15).

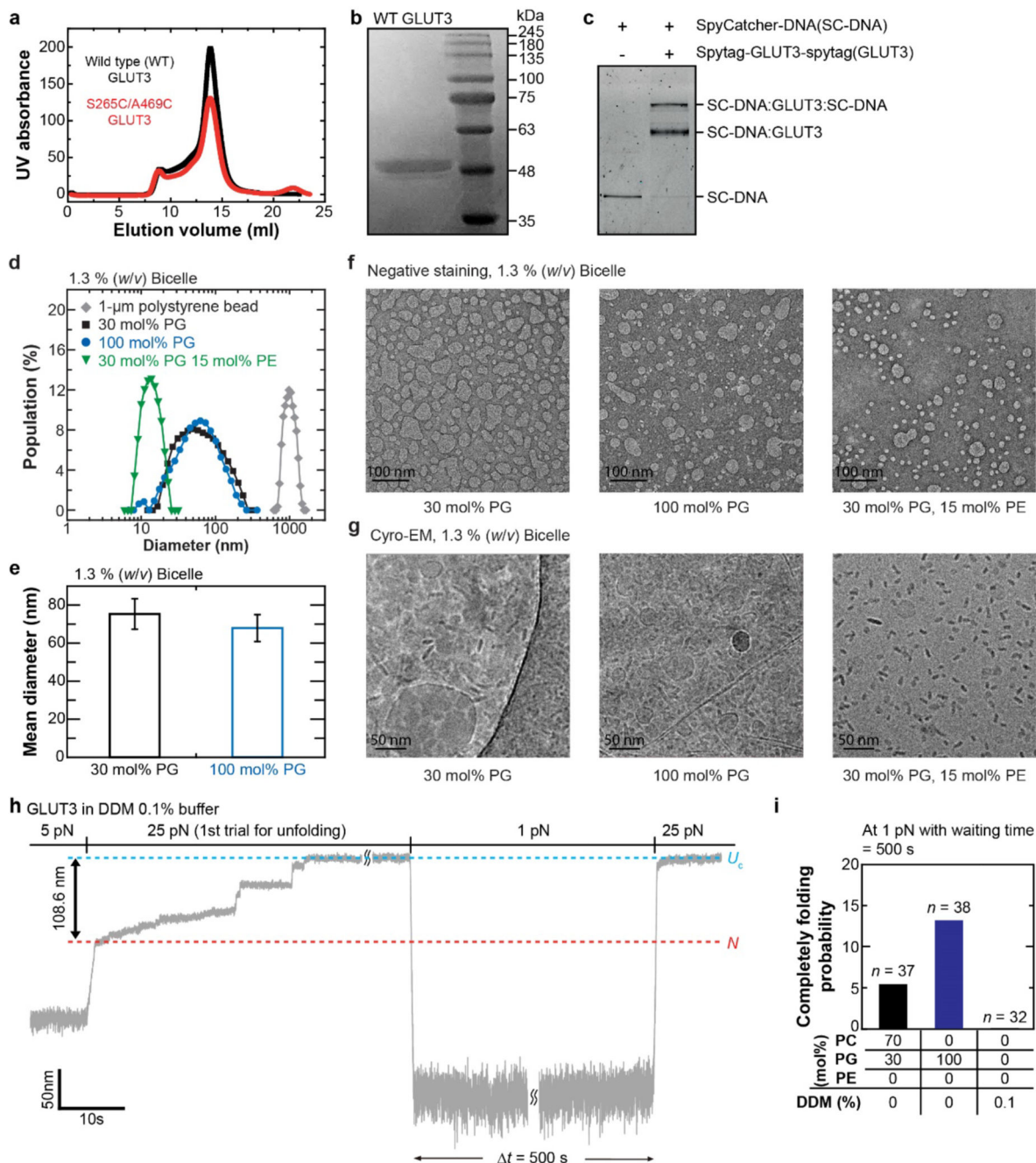
Since the sequences are divergent and therefore difficult to align, we employ two strategies to improve the alignment: (1) multiple references and (2) aligning profile HMMs instead of sequences. We built profile-HMMs for the 143 sugar transporters by aligning them against the UniRef30 database (2020_06)⁸⁵ using hhblits (v3.3.0, parameter -mact 0.1)⁴⁴. Each non-reference profile is aligned to the references using halign (parameter - glob)^{86,87}. The reference with the highest pairwise alignment score was chosen to transfer its annotation. First, the center is inferred from the residue that aligns with the reference center, however, in three cases the helix center did not align and we chose one of the directly adjacent residues.

Second, to counteract misaligned helix positions, we refined the center positions by using the position with the minimal energy for a ± 3 offset, calculated with DGpred. For the resulting coordinates, we extracted the helices and calculated the insertion energy for the 19 residue long helices with DGpred.

Statistics and Reproducibility

Each gel image in Extended Data Fig. 1b-c, 3c, 5b, 5e is a representative gel of a purification protocol that were performed at least 10 independent times with similar results. BODIPY gel assay in Extended Data Fig. 3c was performed 2 independent times with similar results. Gel image in Extended Data Fig. 7a is a representative Western blot and the similar result is also reproducible in the reference, Guna et al. 2018. Each TEM image in Extended Data Fig. 1f-g is a representative micrograph among the reproducible features obtained from more than 20 spots on TEM grids. For repeatability, at least three TEM grids were prepared for each bicelle-forming condition.

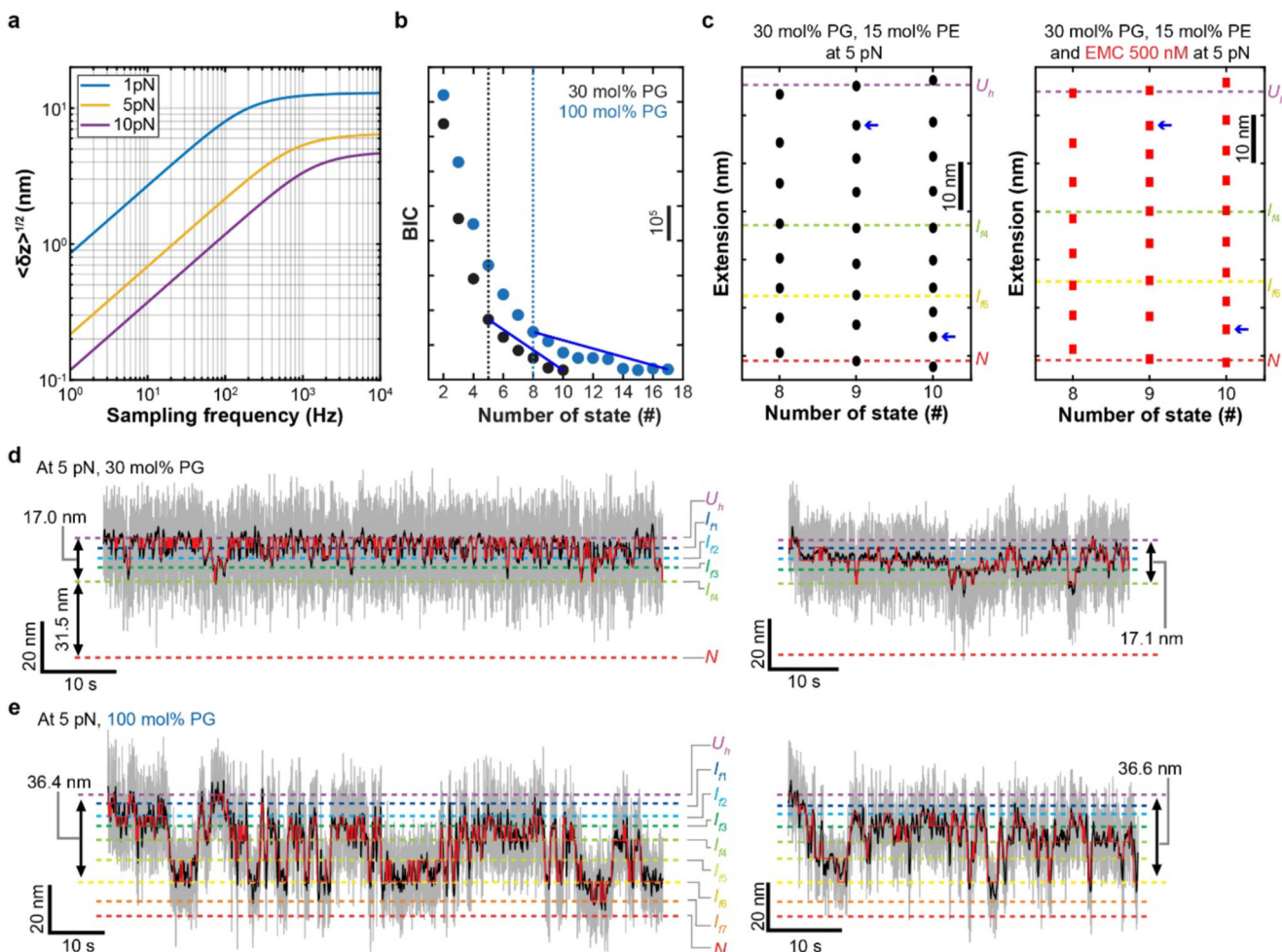
Extended Data



Extended Data Fig. 1. Sample preparation of WT and C-domain-knotted GLUT3 proteins and bicelle membranes

a, Elution profile obtained by size exclusion chromatography (SEC) of WT GLUT3 (black) and S265C/A469C GLUT3 (red). **b**, Purified WT GLUT3 protein analyzed by SDS-PAGE. The major peak position in (a) was used for the gels. The right lane is molecular weight standards. **c**, Representative gel image of SDS-PAGE after SYBR green staining. The left lane shows SpyCatcher-DNA handle only, while the right lane exhibits a mixture of

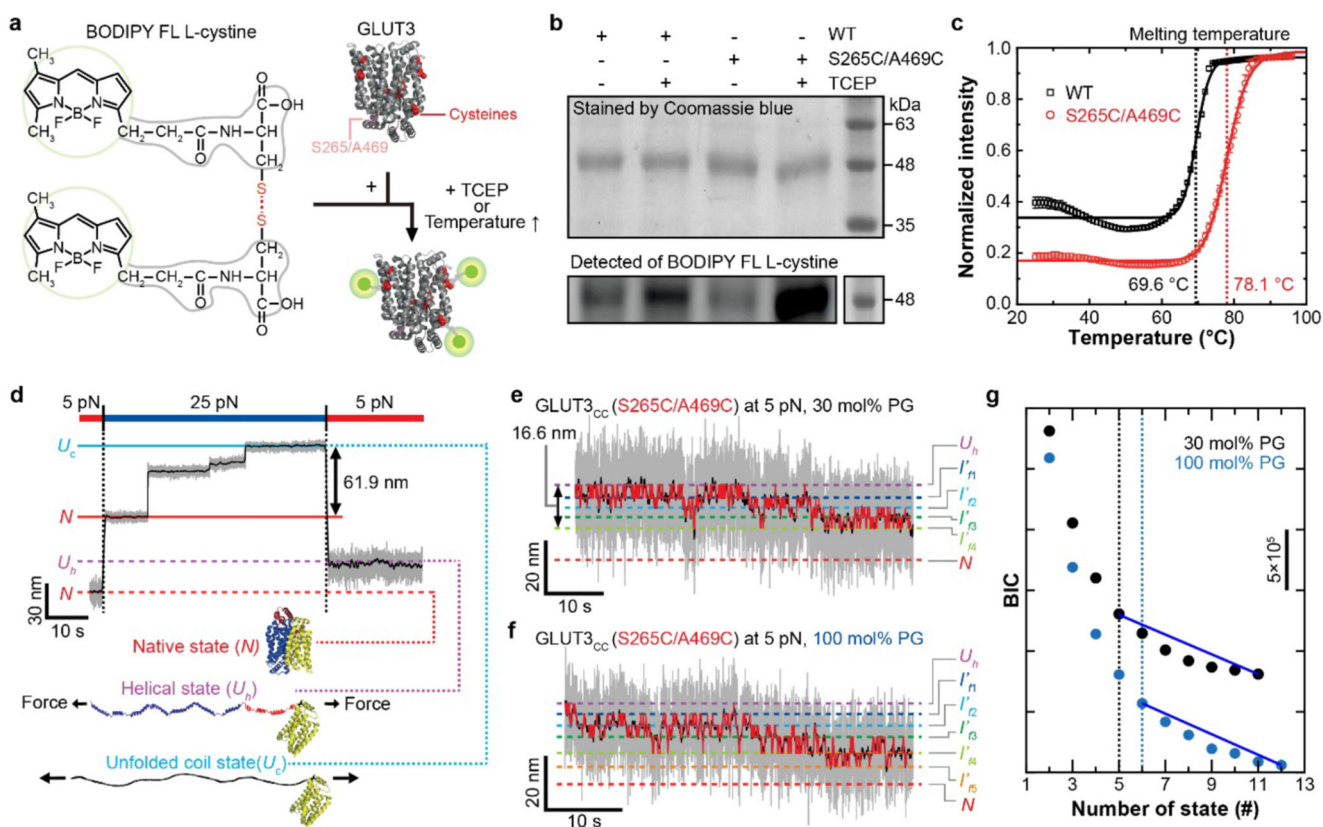
the SpyCatcher-DNA handle and the purified Spytag-GLUT3-spytag. **d**, Bicelle size as measured by dynamic light scattering (DLS) under indicated lipid compositions. The data with gray color is a control sample of 1- μm polystyrene beads. **e**, Mean diameter of the bicelles determined for each lipid composition. Error bars represent SEM ($n = 22$ and 25 for 30 mol% PG and 100 mol% PG, respectively). **f,g**, Electron microscope images of bicelles with different lipid compositions. Two specified electron microscopy methods are used. **h**, Representative trace for the force-jump experiment to determine the refolding probability. Force was first increased to 25 pN to induce full unraveling of the protein and then relaxed to 1 pN for 500 s before checking the folding status through re-application of 25 pN. The folding status was determined by the unfolding steps observed under 25 pN. **i**, Probability for observing the completely folded state for indicated buffer conditions. The refolding probability virtually abolished when GLUT3 was embedded in DDM micelles at 0.1 % (w/v), indicating that the lipid bilayer environments provided by the bicelle membranes are essential for inducing the fully folded state.



Extended Data Fig. 2. Analysis of single-molecule magnetic tweezer data

a, Precision in determination of the vertical position of a bead as a function of the measurement bandwidth. The plot indicates an ~ 1 nm resolution when bead positions are

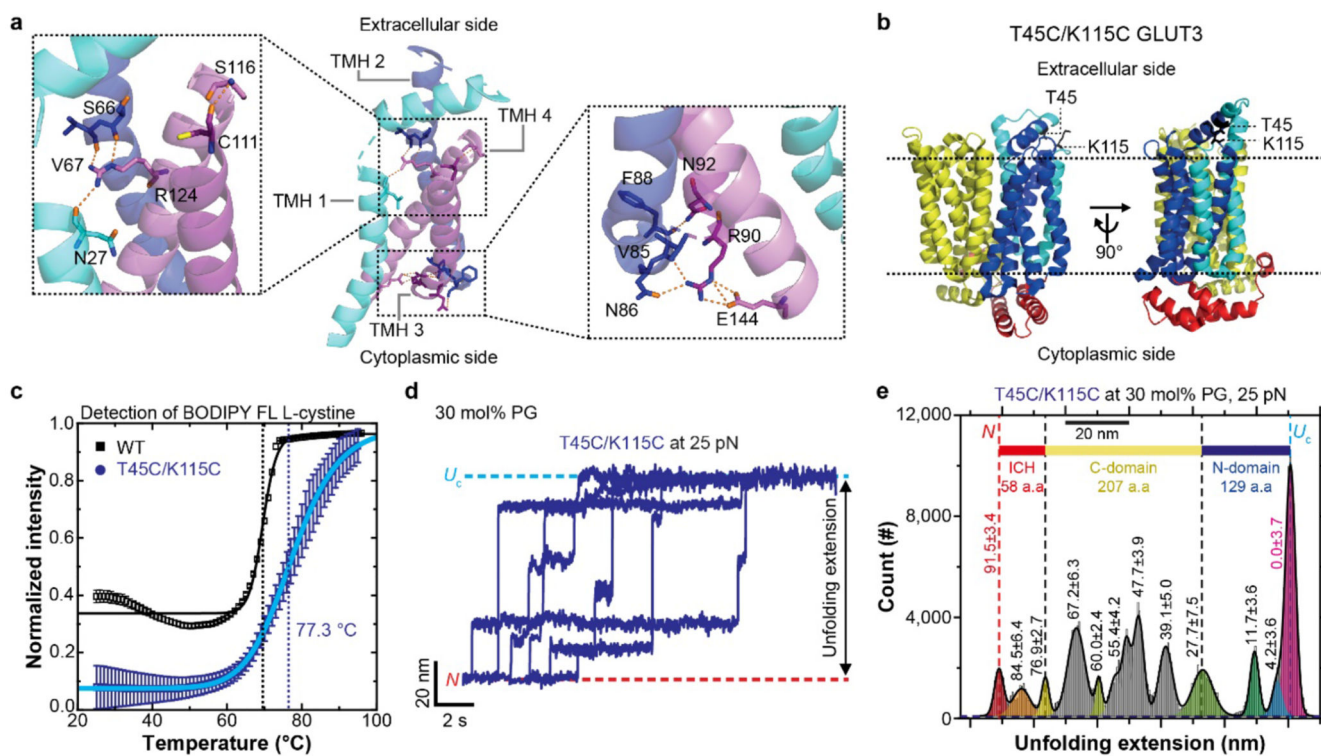
averaged over 50 ms (~20 Hz sampling). In our magnetic tweezer experiments, the bicelle phase used for providing the lipid bilayer environments to the target membrane proteins offers additional low-frequency fluctuations, forcing a longer averaging time of 200 ms to achieve the 1 nm accuracy in our membrane protein folding studies. **b**, Bayesian Information Criteria (BIC) values of WT GLUT3 for each number of states with different bicelle compositions ($n = 16$ and 11 for 30 mol% and 100 mol% PG, respectively). **c**, Position of folding intermediates determined by HMM with for different number of states assumed in HMM analysis. The positions of the key intermediates (U_h , I_{f4} (N-domain folded), I_{f6} (C-domain folded) and N) are essentially preserved when the number of assumed states are changed, which only generates additional intermediates either in the middle of either N-domain folding or domain-domain assembly (blue arrows). Thus, our HMM analysis does not randomly assign the intermediate positions out of noisy data, but rather identifies the intermediates implicated in our time-resolved traces in a robust way. **d, e**, Representative folding traces for WT GLUT3 with 30 mol% PG (**d**) and 100 mol% PG (**e**) at 5 pN. Two replicates are shown for each condition, and the gray and black traces are 1.2-kHz raw data and 5-Hz median-filtered data, respectively. Red traces indicate the transitions between intermediates identified by HMM.



Extended Data Fig. 3. Sample preparation and folding behavior of S265C/A469C GLUT3

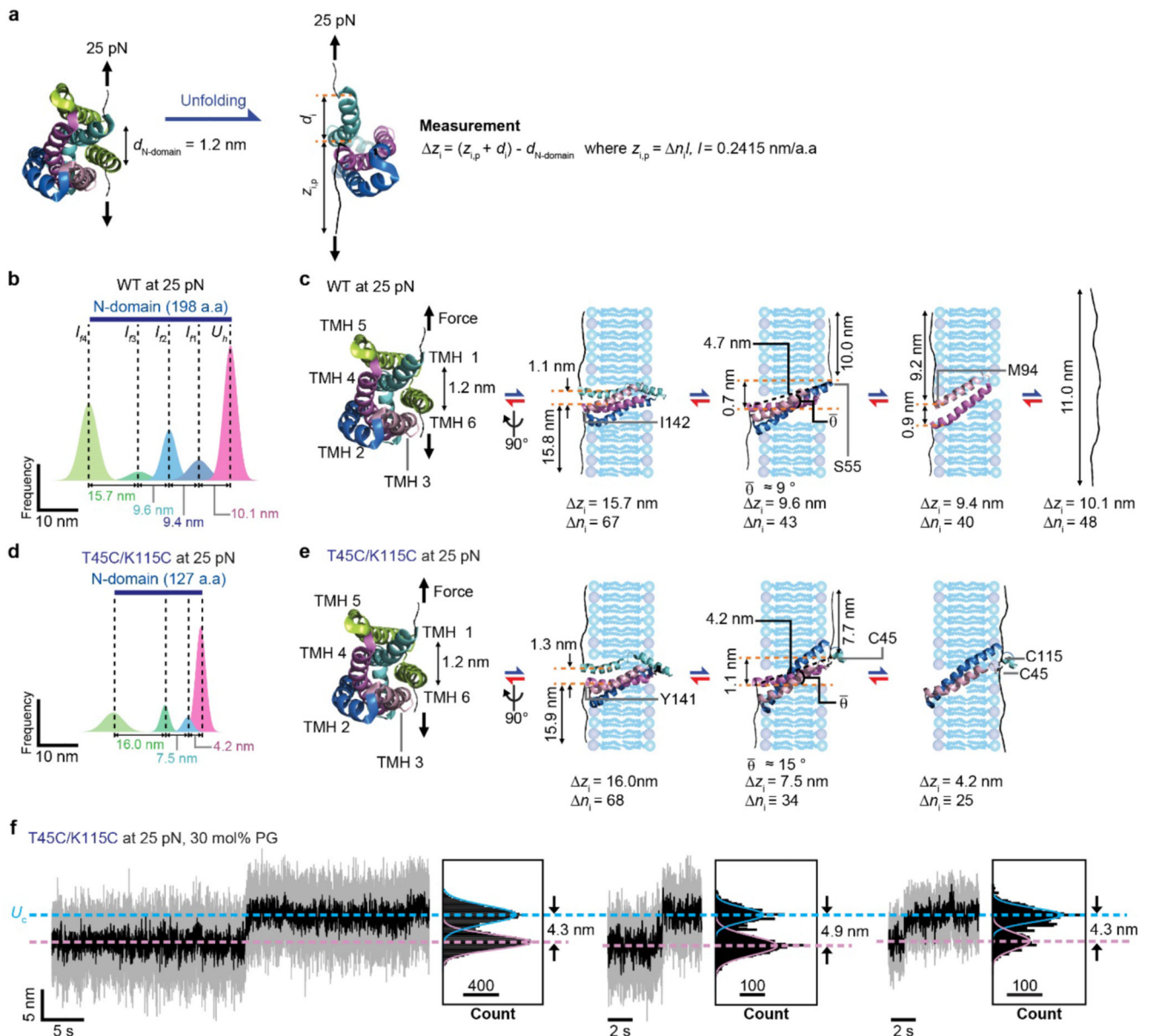
a, Schematic of the assay using BODIPY-L-cystine. The left panel is the chemical structure of two BODIPY FL fluorophores attached to the amino groups of the disulfide-containing amino acid, cysteine. The right panel shows the structure of GLUT3 before and after the

treatment (addition of TCEP or increasing the temperature). Green dots in the right panel are the BODIPY FL fluorophores reacted with cysteines in GLUT3. **b**, Gel analysis for WT and S265C/A469C GLUT3 in the presence of TCEP. Upper gel shows the amount of GLUT3 stained by Coomassie blue. Lower gel shows the amount of BODIPY FL fluorophores reacted with cysteines in GLUT3. The stained positions are same in both gels. **c**, Fluorescence profile of BODIPY FL fluorophore-labeled GLUT3 as temperature increased. Dashed lines indicate the melting temperatures of the WT (black) and S265C/A469C GLUT3 (red). Error bars represent SEM ($n = 4$). **d**, A mechanical cycle for inducing refolding of a single S265C/A469C GLUT3 (GLUT3_{CC}) and corresponding structural states of the protein. **e,f**, Representative folding traces of GLUT3_{CC} with 30 mol% PG (**f**) and 100 mol% PG (**g**) at 5 pN. Definitions of the traces are identical to those of Extended Data Fig. 2d. **g**, BIC values of GLUT3_{CC} for each number of states ($n = 22$ and 12 for 30 mol% and 100 mol% PG, respectively).



Extended Data Fig. 4. Sample preparation and unfolding characteristics of T45C/K115C GLUT3

a, Atomic contacts among TMHs 1, 2, and 4. Inset shows detailed position of interacting residues (blue for amino group, orange for carboxyl group, and yellow for thiol group). **b**, The positions of two mutations, T45C/K115C in GLUT3 (GLUT3_{TM23C}). **c**, An absorbance profile of BODIPY FL fluorophore-labeled GLUT3_{TM23C} as temperature increases. The experiment was done as depicted in Extended Data Fig. 3c. Error bars represent SEM ($n = 4$). **d**, Collection of 50Hz-median filtered unfolding traces initiated from *N* state for GLUT3_{TM23C}. **e**, Distributions of extension values recorded during high-force unfolding of single GLUT3_{TM23C} proteins. Extension values represent mean \pm SD ($n = 19$).

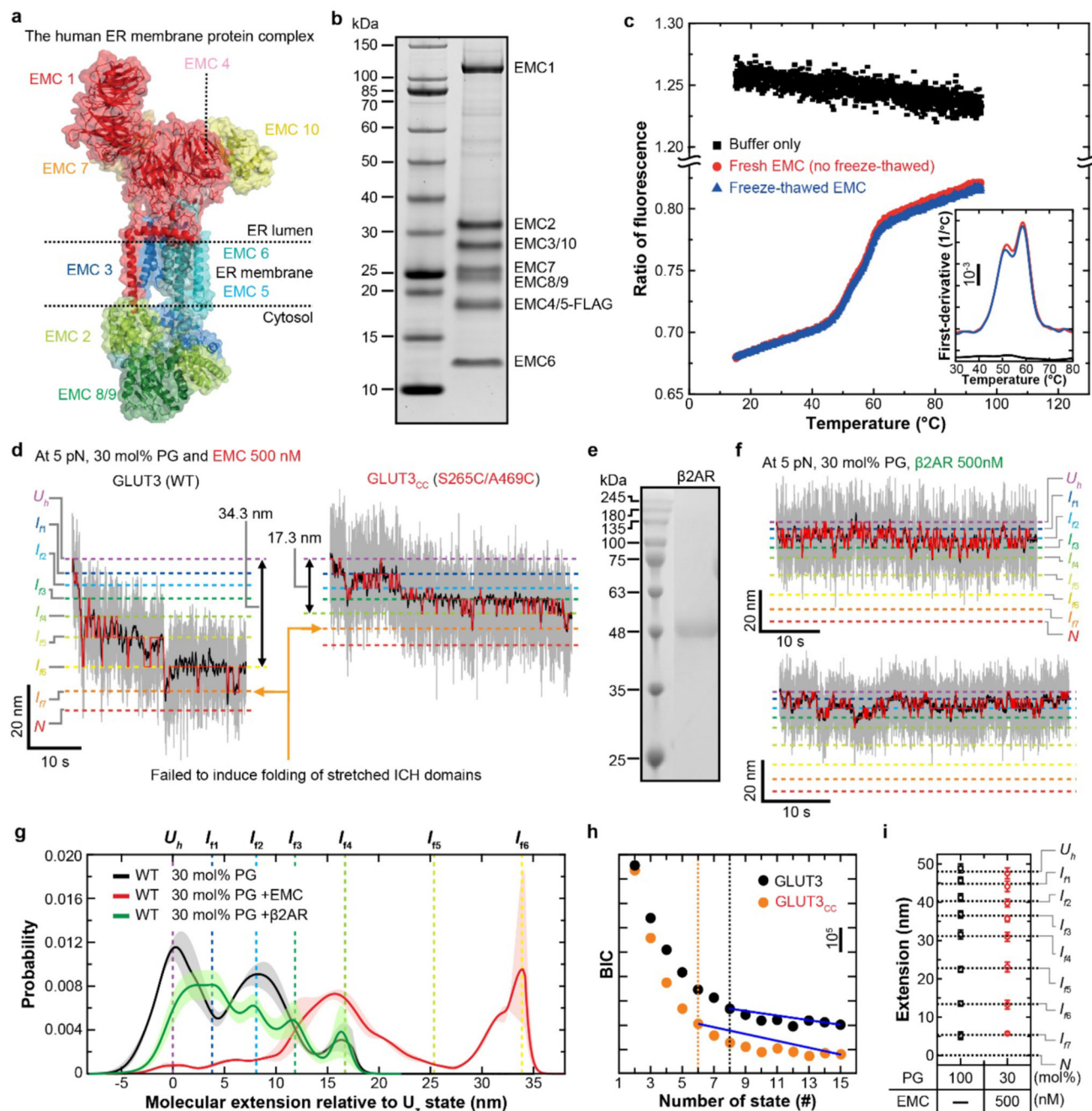


Extended Data Fig. 5. Determination of folding order for N-domain of GLUT3

a, Schematic of pulling geometry for N-domain of GLUT3 at 25 pN. $d_{N\text{-domain}}$ is the distance between two points of force application before unfolding (PDB: 4ZWC). z_i indicates the expected extension increase for GLUT3 for the i th intermediate. $z_{i,p}$ is the extension of the unfolded portion along the membrane for the i th intermediate. d_i denotes the distance between the points of force application for the i th intermediate. n_i is the number of amino acids of the unfolded portion. l is the length of a single amino acid.

b, Unfolding extension distribution for the N-domain part of the WT GLUT3 at 25 pN. **c**, Structural information and folding/unfolding order of the N-domain of WT GLUT3. The distance between two orange dashed lines (perpendicular to the membrane) represents the vertical distance between the two points of application (d_i). This orange dashed line forms

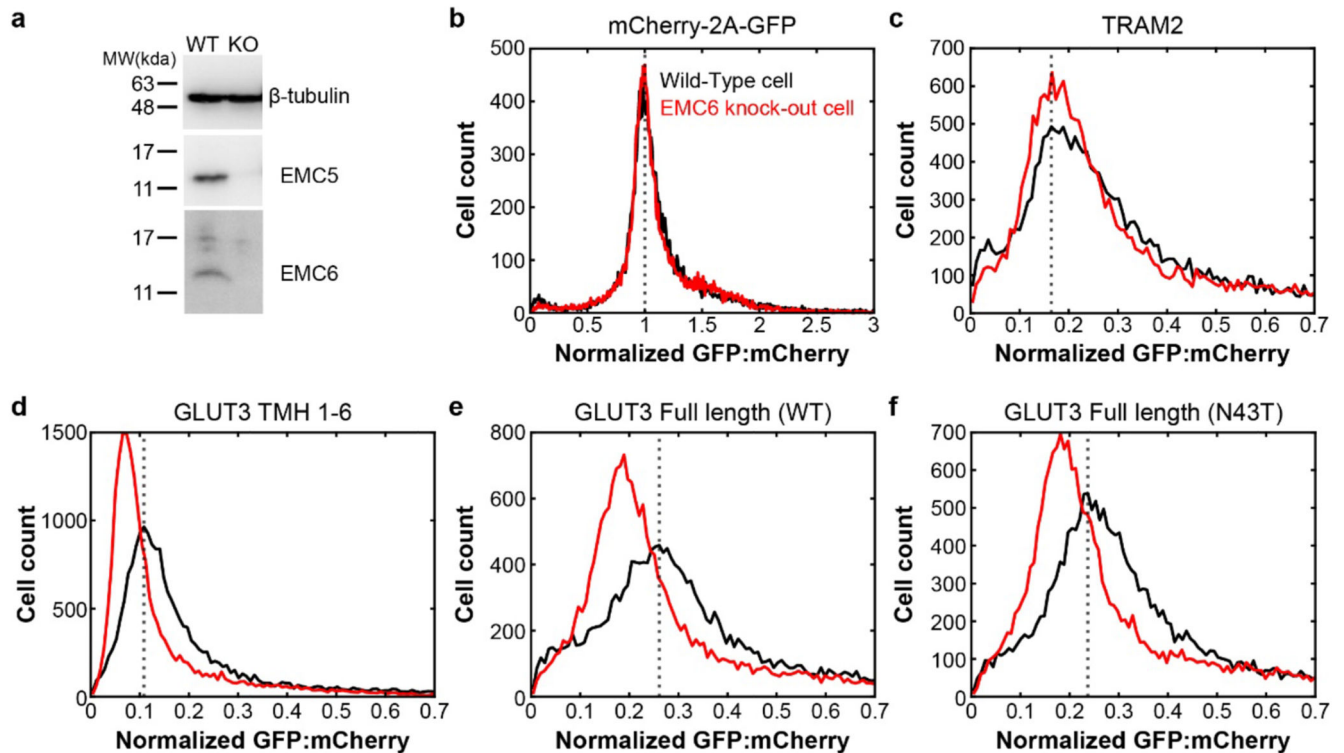
an angle of $\bar{\theta}$ with a black dashed line to the unfolded portion of N-domain in the membrane. **d**, Unfolding extension distribution for the N-domain part of GLUT3_{TM23C} at 25pN. **e**, Structural information and folding/unfolding order of the N-domain of GLUT3_{TM23C}. The description is the same as (c) except for protein construct. The disulfide bond of GLUT3_{TM23C} did not affect the first unfolding step for N-domain that amounted to ~15.7 nm, confirming that TMHs 5 and 6 constitute the first unfolding step of N-domain. The second unfolding step was slightly reduced to 7.5 nm, which was consistent with the length of last helical turn of the long linker region when TMH 1 would be protected by knotting, mapping the second unfolding step to that of TMH 1 and its linker region. The last two unfolding steps before U_c were reduced to a single step of 4.2 nm, which would reflect unfolding of TMH 4 outside the knotted region. **f**, Representative traces showing the final unfolding step of a 4.2 nm extension increase for GLUT3_{TM23C}. Three replicates are shown, and the value indicates the distance between two gaussian peaks.



Extended Data Fig. 6. EMC preparation and folding characteristics of GLUT3 with β2AR

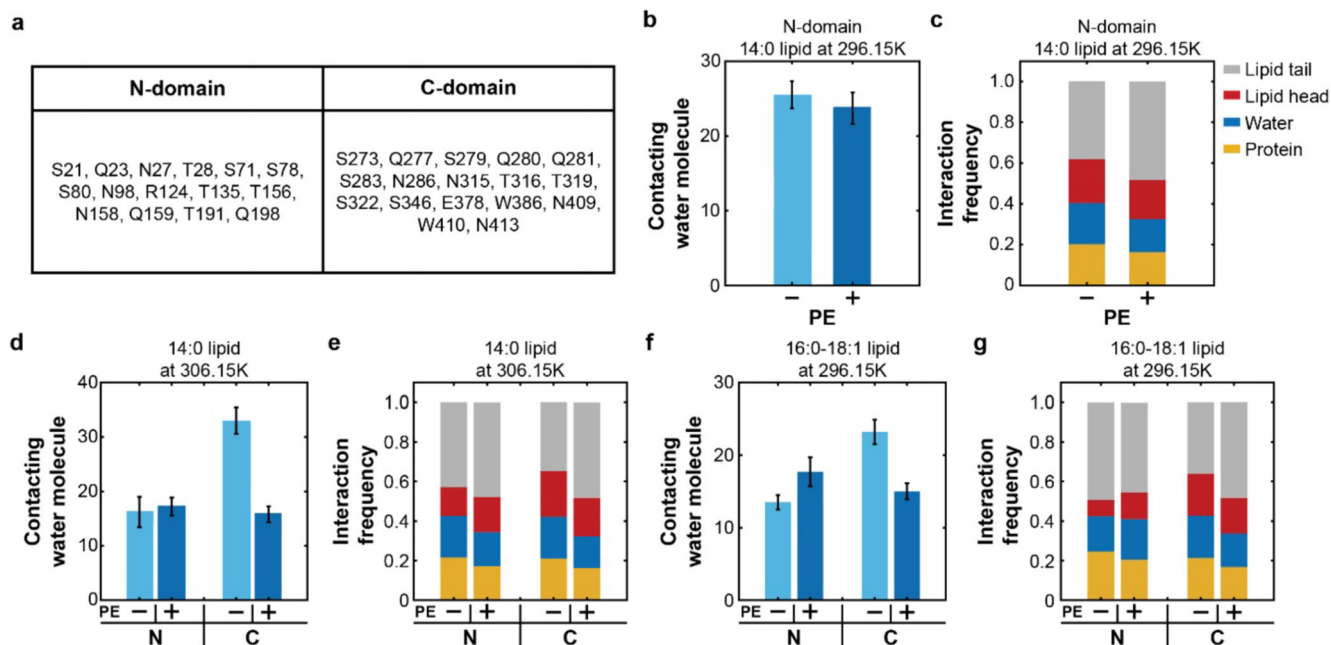
a, Structural model of the human ER membrane protein complex (EMC). **b**, Purified EMC analyzed by SDS-PAGE. The left lane is molecular weight standards and the right lane shows purified EMC. **c**, Fluorescence profile of EMC before and after freeze-thaw as the temperature is increased. Inset displays the profile for first derivative of fluorescence intensity. **d**, Representative folding traces of WT (left) and GLUT3_{CC} (right) in the presence of EMC. The definition of each trace is identical to the traces in Extended Data Fig. 2d. **e**, Purified β2AR analyzed by SDS-PAGE. The left lane is molecular weight standards and the

right lane shows β 2AR. **f**, Representative folding traces for single GLUT3 at 5 pN with 30 mol% PG in the bicelles in the presence of 500 nM β 2AR. Two replicates are shown. **g**, Probability distributions of deconvoluted extension values observed under indicated folding conditions at 5pN ($n = 11$ for the reaction with β 2AR). The black and red distributions are revisited from Fig. 4e. The shaded area means SEM. **h**, BIC values for the indicated number of states ($n = 13$ and 11 traces for WT GLUT3 and S265C/A469C GLUT3, respectively). **i**, Positions of folding/unfolding intermediates identified with HMM are depicted for the indicated conditions. Error bars represent SEM ($n = 22$ and 35 traces for 100 mol% DMPG and 30 mol% DMPG with 500nM EMC, respectively).



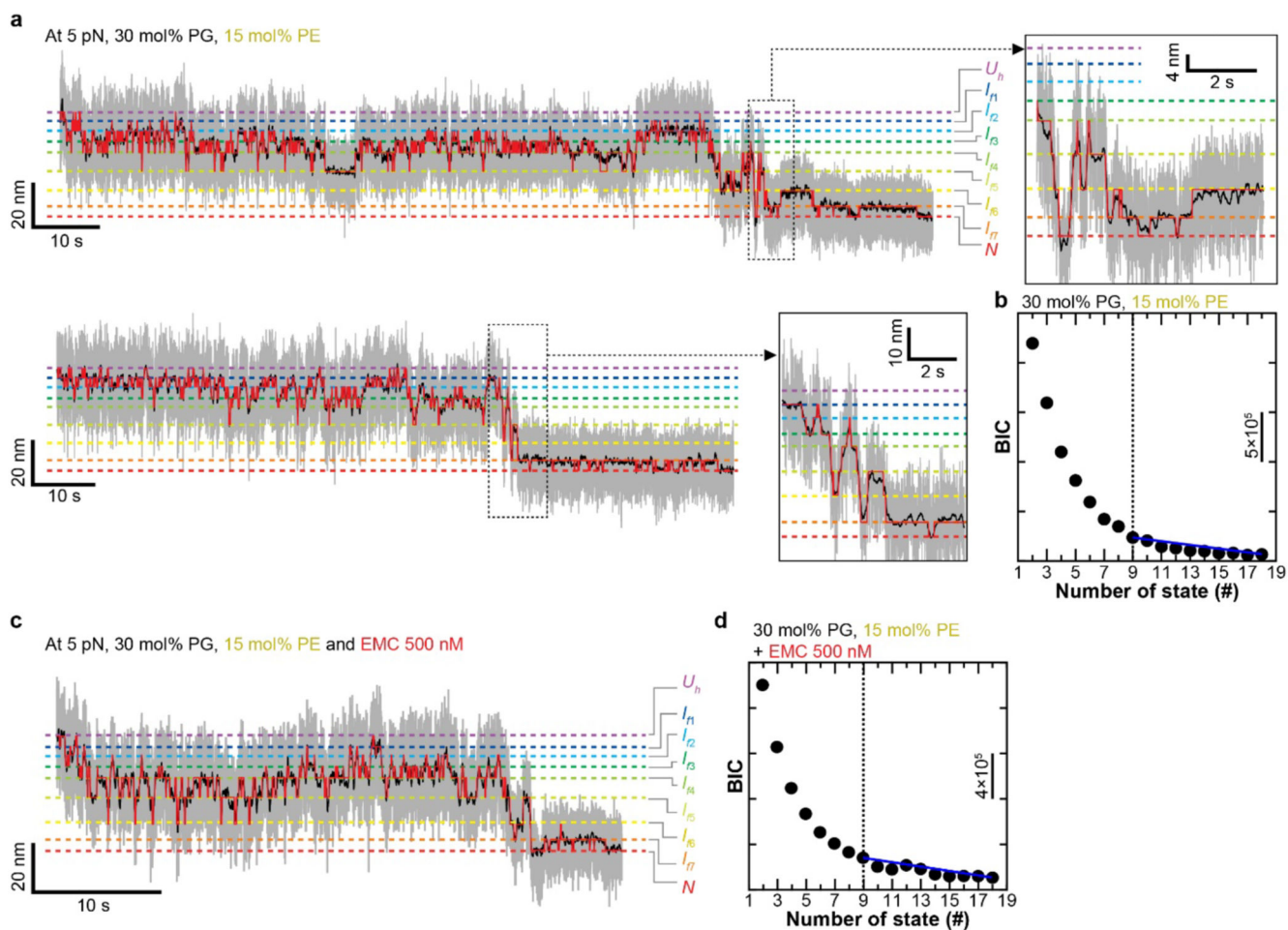
Extended Data Fig. 7. Dual-color ratio-metric assay data for WT and EMC6 KO Flp-In T-Rex 293 cells.

a, Western blot for EMC5 and EMC6 in wild-type (WT) and EMC6 knock-out Flp-In T-Rex 293 cells. **b-f**, Histograms of eGFP:mCherry ratios for the indicated constructs. Each Line with black and red means WT and EMC6 knock-out cell respectively.



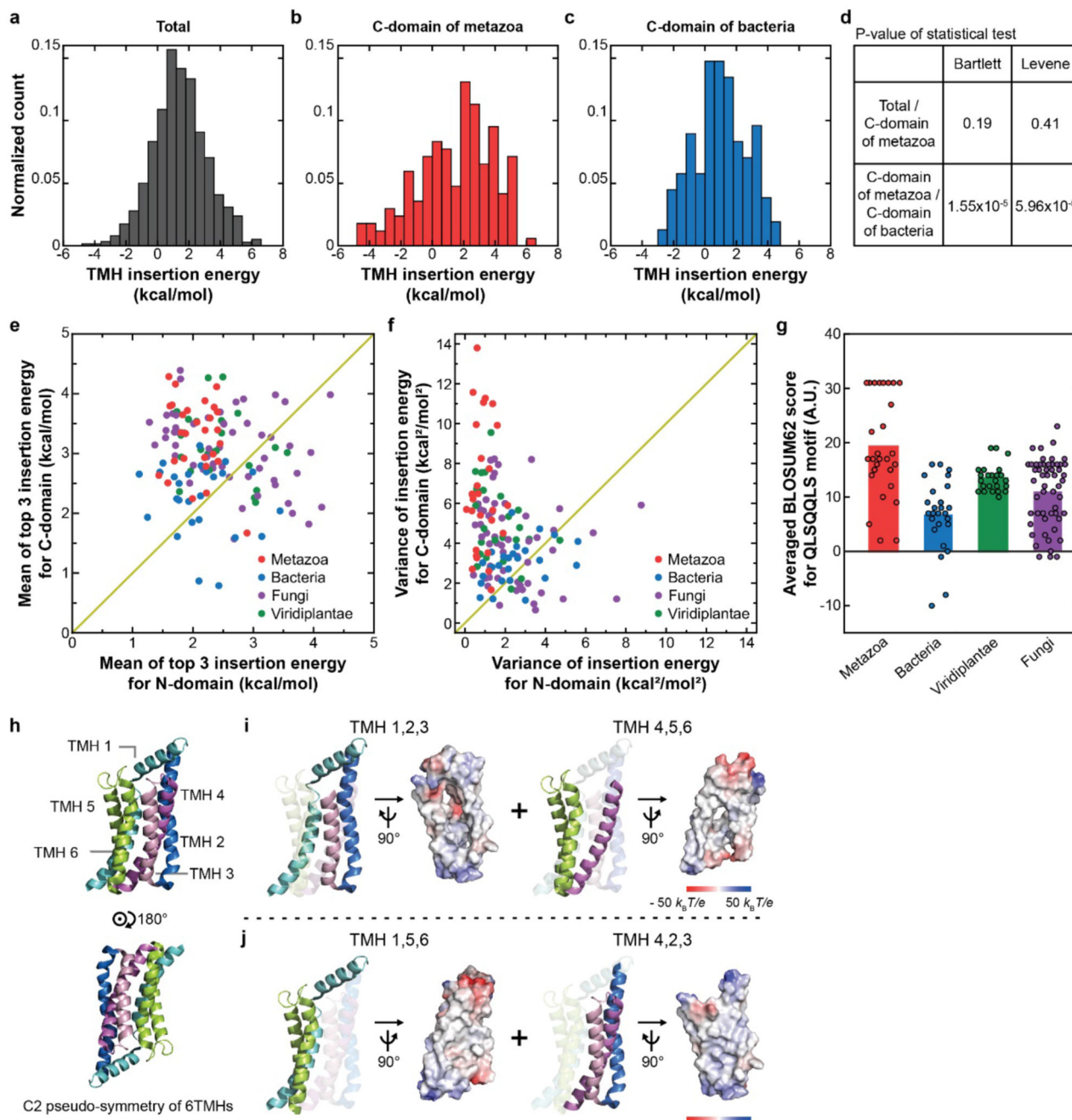
Extended Data Fig. 8. Analysis of MD simulation for GLUT3 in lipid bilayer

a, List of polar/charged residues in TMHs of GLUT3 for the analysis. Residues near the GLUT3 pore entries are not chosen which are likely to be exposed to bulk water. **b**, The average number of contacting water molecule to polar/charged residues in TMHs of N-domain with or without DMPE. Error bars represent SD ($n = 4000$ for each case). **c**, Interaction frequency of polar/charged residues in N-domain interface with or without DMPE. The value in (**b,c**) is the average value from $0.6\mu\text{s}$ to $1.0\mu\text{s}$. **d**, The average number of contacting water molecule to polar/charged residues in TMHs of GLUT3. 'N' and 'C' represent N, C-domain, respectively. Error bars represent SD ($n = 4000$ for each case). **e**, Interaction frequency of polar/charged residues in domain interfaces. The value in (**d,e**) is the average value from $0.6\mu\text{s}$ to $1.0\mu\text{s}$. **f**, The average number of contacting water molecule to polar/charged residues in TMHs of GLUT3. Error bars represent SD ($n = 2000$ for each case). **g**, Interaction frequency of polar/charged residues in domain interfaces. The value in (**f,g**) is the average value from $0.3\mu\text{s}$ to $0.5\mu\text{s}$.



Extended Data Fig. 9. Folding characteristics with PE lipid bicelles

a, Representative folding trace of WT GLUT3 with PE-containing bicelle. Inset shows close-up view of the folding trace. The definition of each trace is identical to the traces in Extended Data Fig. 2d. Two replicates are shown. **b**, BIC values for the indicated number of states with 15 mol% PE bicelle ($n = 11$). The intermediate was largely preserved upon addition of DMPE lipids. **c**, Representative folding trace of WT GLUT3 with PE-containing bicelle in the presence of EMC. **d**, BIC values for the indicated number of states with 15 mol% PE bicelle in the presence of EMC ($n = 10$).



Extended Data Fig. 10. Insertion energy of sugar transporters and symmetrical structure of GLUT3

a, Insertion energy histogram estimated for TMHs of all sugar porters. **b**, Insertion energy histogram estimated for C-domain TMHs of metazoan sugar porters. **c**, Insertion energy histogram estimated for C-domain TMHs of bacteria sugar porters. **d**, P-values from the Bartlett and Levene tests. 2 sets are used for statistical testing. **e**, Scatter plot of mean of top 3 insertion energy for N-domain as x-axis and mean of 3 top insertion energy for C-domain as y-axis for sugar transporters. **f**, Scatter plot of insertion energy variance for

N-domain as x-axis and insertion energy variance for C-domain as y-axis. **g**, Average values of BLOSUM62 score for QLSQQLS motif is calculated for each group. ($n = 26, 28, 54$ and 24 for bacteria, metazoa, fungi and viridiplantae). **h**, Structural view of GLUT3's N-domain with its C2 pseudo-symmetry. **i**, Structural view and electrostatic potential of the helix triplets composed of TMHs 1, 2, 3 and 4, 5, 6 each. **j**, Structural view and electrostatic potential of the helix triplets composed of TMHs 1, 5, 6 and 4, 2, 3 each.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank E. Kweon for helps with preparing illustrations. This work was supported by National Creative Research Initiative Program (NRF-2021R1A3B1071354 to T.-Y.Y.), the Bio Medical Technology Development Program (NRF-2018M3A9E2023523 to T.-Y.Y.) and NRF grants (NRF-2019M3E5D6063903 and NRF-2020R1A2C2003783 to H.-J.C.; NRF-2019R1A6A1A10073437 and NRF-2020M3A9G7103933 to M.S.; NRF-2021M3A9I4021220 and NRF-2020R1A6C101A183 to S.-H.R), all funded by the National Research Foundation of South Korea. This work was also supported by the UK Medical Research Council (MRC_UP_12-1/10 to E.A.M.), the US National Science Foundation (MCB-181069 to W.I.) and National Institute of Health grant (R01GM118684 to H.H.).

Data availability

All data that support the findings of this study are available in the manuscript or Supplementary Figures. Raw data have been deposited in Github (https://github.com/tyoonlab/Nat_Chem_biol_NCHEMB-A210813554). The following PDB IDs were used (4ZWB, 4ZWC, 4ZW9, 4GBZ, 6RW3, 4LDS and 6H7D). Also, The following UniProt IDs were used (P11169, P0AGF4, O97467, A0A0H2VG78 and Q9LT15).

Code availability

A program, written in LabView, to control the magnetic tweezers apparatus has been deposited in Github (https://github.com/tyoonlab/Science_aaw8208) and is available at Zenodo (doi:10.5281/zenodo.3528913). Codes for analyzing the magnetic tweezers and FACS data has been deposited in Github (https://github.com/tyoonlab/Nat_Chem_biol_NCHEMB-A210813554). A code for estimating the helix insertion energy is available at Github under GPL3.0 license (https://github.com/schnamo/TMH_insertion_energy).

References

1. Hediger MA, et al. The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflügers Archiv*. 2004; 447: 465–468. [PubMed: 14624363]
2. Cheng Y. Membrane protein structural biology in the era of single particle cryo-EM. *Current Opinion in Structural Biology*. 2018; 52: 58–63. [PubMed: 30219656]
3. Marinko JT, et al. Folding and misfolding of human membrane proteins in health and disease: from single molecules to cellular proteostasis. *Chemical Reviews*. 2019; 119: 5537–5606. [PubMed: 30608666]

4. Guna A, Hegde RS. Transmembrane domain recognition during membrane protein biogenesis and quality control. *Current Biology*. 2018; 28: R498–R511. [PubMed: 29689233]
5. Quistgaard EM, Löw C, Guettou F, Nordlund P. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nature Reviews Molecular Cell Biology*. 2016; 17: 123. [PubMed: 26758938]
6. Oberai A, Ihm Y, Kim S, Bowie JU. A limited universe of membrane protein families and folds. *Protein Science*. 2006; 15: 1723–1734. [PubMed: 16815920]
7. Almén MS, Nordström KJV, Fredriksson R, Schiöth HB. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology*. 2009; 7: 50–50. [PubMed: 19678920]
8. Shi Y. Common folds and transport mechanisms of secondary active transporters. *Annual Review of Biophysics*. 2013; 42: 51–72.
9. Chitwood PJ, Hegde RS. An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature*. 2020; 584: 630–634. [PubMed: 32814900]
10. Kota J, Ljungdahl PO. Specialized membrane-localized chaperones prevent aggregation of polytopic proteins in the ER. *The Journal of Cell Biology*. 2005; 168: 79–88. [PubMed: 15623581]
11. Choi H-K, et al. Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science*. 2019; 366: 1150–1156. [PubMed: 31780561]
12. Saier MH Jr. Computer-aided analyses of transport protein sequences: gleaned evidence concerning function, structure, biogenesis, and evolution. *Microbiological Reviews*. 1994; 58: 71. [PubMed: 8177172]
13. Saier MH Jr. Families of transmembrane sugar transport proteins: microreview. *Molecular Microbiology*. 2000; 35: 699–710. [PubMed: 10692148]
14. Pao SS, Paulsen IT, Saier MH Jr. Major facilitator superfamily. *Microbiology and Molecular Biology Reviews*. 1998; 62: 1. [PubMed: 9529885]
15. Reddy VS, Shlykov MA, Castillo R, Sun EI, Saier MH Jr. The major facilitator superfamily (MFS) revisited. *The FEBS Journal*. 2012; 279: 2022–2035. [PubMed: 22458847]
16. Yan N. Structural biology of the major facilitator superfamily transporters. *Annual Review of Biophysics*. 2015; 44: 257–283.
17. Deng D, et al. Molecular basis of ligand recognition and transport by glucose transporters. *Nature*. 2015; 526: 391–396. [PubMed: 26176916]
18. Madej MG. Function, structure, and evolution of the Major Facilitator Superfamily: the LacY manifesto. *Advances in Biology*. 2014; 2014
19. Drew D, Boudker O. Shared molecular mechanisms of membrane transporters. *Annual Review of Biochemistry*. 2016; 85: 543–572.
20. Longo L, Lee J, Blaber M. Experimental support for the foldability–function tradeoff hypothesis: Segregation of the folding nucleus and functional regions in fibroblast growth factor-1. *Protein Science*. 2012; 21: 1911–1920. [PubMed: 23047594]
21. Hingorani KS, Gierasch LM. Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge. *Current Opinion in Structural Biology*. 2014; 24: 81–90. [PubMed: 24434632]
22. Min D, Arbing MA, Jefferson RE, Bowie JU. A simple DNA handle attachment method for single molecule mechanical manipulation experiments. *Protein Science*. 2016; 25: 1535–1544. [PubMed: 27222403]
23. Findlay HE, Booth PJ. The biological significance of lipid–protein interactions. *Journal of Physics: Condensed Matter*. 2006; 18 S1281 [PubMed: 21690841]
24. Ujwal R, Bowie JU. Crystallizing membrane proteins using lipidic bicelles. *Methods*. 2011; 55: 337–341. [PubMed: 21982781]
25. Findlay HE, Booth PJ. The folding, stability and function of lactose permease differ in their dependence on bilayer lipid composition. *Scientific Reports*. 2017; 7 13056 [PubMed: 29026149]
26. Hessa T, et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*. 2007; 450: 1026–1030. [PubMed: 18075582]
27. Snider C, Jayasinghe S, Hristova K, White SH. MPEX: A tool for exploring membrane proteins. *Protein Science*. 2009; 18: 2624–2628. [PubMed: 19785006]

28. Moon CP, Fleming KG. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proceedings of the National Academy of Sciences*. 2011; 108: 10174–10177.
29. Lee T-H. Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden Markov models. *The Journal of Physical Chemistry B*. 2009; 113: 11535–11542. [PubMed: 19630372]
30. Zhang Y, Jiao J, Rebane AA. Hidden Markov modeling with detailed balance and its application to single protein folding. *Biophysical Journal*. 2016; 111: 2110–2124. [PubMed: 27851936]
31. Radestock S, Forrest LR. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *Journal of Molecular Biology*. 2011; 407: 698–715. [PubMed: 21315728]
32. Forrest LR. Structural symmetry in membrane proteins. *Annual Review of Biophysics*. 2015; 44: 311–337.
33. Jonikas MC, et al. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science*. 2009; 323: 1693–1697. [PubMed: 19325107]
34. Pleiner T, et al. Structural basis for membrane insertion by the human ER membrane protein complex. *Science*. 2020; 369: 433–436. [PubMed: 32439656]
35. Chitwood PJ, Juskiewicz S, Guna A, Shao S, Hegde RS. EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell*. 2018; 175: 1507–1519. e16 [PubMed: 30415835]
36. Guna A, Volkmar N, Christianson JC, Hegde RS. The ER membrane protein complex is a transmembrane domain insertase. *Science*. 2018; 359: 470–473. [PubMed: 29242231]
37. Kauko A, et al. Repositioning of transmembrane α -helices during membrane protein folding. *Journal of Molecular Biology*. 2010; 397: 190–201. [PubMed: 20109468]
38. Lai G, Renthal R. Integral membrane protein fragment recombination after transfer from nanolipoprotein particles to bicelles. *Biochemistry*. 2013; 52: 9405–9412. [PubMed: 24328096]
39. Jo S, Lim JB, Klauda JB, Im W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophysical Journal*. 2009; 97: 50–58. [PubMed: 19580743]
40. Min D, et al. Unfolding of a ClC chloride transporter retains memory of its evolutionary history. *Nature Chemical Biology*. 2018; 14: 489–496. [PubMed: 29581582]
41. Cheung MS, García AE, Onuchic JN. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proceedings of the National Academy of Sciences*. 2002; 99: 685–690.
42. Bowie JU. Solving the membrane protein folding problem. *Nature*. 2005; 438: 581–589. [PubMed: 16319877]
43. Bogdanov M, Dowhan W. Lipid-dependent generation of dual topology for a membrane protein. *Journal of Biological Chemistry*. 2012; 287: 37939–37948. [PubMed: 22969082]
44. Steinegger M, et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*. 2019; 20: 473. [PubMed: 31521110]
45. Seatter MJ, De La Rue SA, Porter LM, Gould GW. QLS motif in transmembrane helix VII of the glucose transporter family interacts with the C-1 position of D-glucose and is involved in substrate selection at the exofacial binding site. *Biochemistry*. 1998; 37: 1322–1326. [PubMed: 9477959]
46. Cymer F, von Heijne G. Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proceedings of the National Academy of Sciences*. 2013; 110: 14640–14645.
47. Lolkema JS, Dobrowolski A, Slotboom D-J. Evolution of antiparallel two-domain membrane proteins: tracing multiple gene duplication events in the DUF606 family. *Journal of Molecular Biology*. 2008; 378: 596–606. [PubMed: 18384811]
48. Walmsley AR, Barrett MP, Bringaud F, Gould GW. Sugar transporters from bacteria, parasites and mammals: structure–activity relationships. *Trends in Biochemical Sciences*. 1998; 23: 476–481. [PubMed: 9868370]
49. Mueckler M, Thorens B. The SLC2 (GLUT) family of membrane transporters. *Molecular Aspects of Medicine*. 2013; 34: 121–138. [PubMed: 23506862]

50. Thorens B. GLUT2, glucose sensing and glucose homeostasis. *Diabetologia*. 2015; 58: 221–232. [PubMed: 25421524]
51. Serdiuk T, et al. YidC assists the stepwise and stochastic folding of membrane proteins. *Nature Chemical Biology*. 2016; 12: 911–917. [PubMed: 27595331]
52. Wang F, Chan C, Weir NR, Denic V. The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase. *Nature*. 2014; 512: 441–444. [PubMed: 25043001]
53. McGilvray PT, et al. An ER translocon for multi-pass membrane protein biogenesis. *Elife*. 2020; 9 e56889 [PubMed: 32820719]
54. Hegde RS, Keenan RJ. The mechanisms of integral membrane protein biogenesis. *Nature Reviews Molecular Cell Biology*. 2021. 1–18. [PubMed: 33244181]
55. Zhang W, Campbell HA, King SC, Dowhan W. Phospholipids as determinants of membrane protein topology: phosphatidylethanolamine is required for the proper topological organization of the γ -aminobutyric acid permease (GabP) of *Escherichia coli*. *Journal of Biological Chemistry*. 2005; 280: 26032–26038. [PubMed: 15890647]
56. Volmer R, Ron D. Lipid-dependent regulation of the unfolded protein response. *Current Opinion in Cell Biology*. 2015; 33: 67–73. [PubMed: 25543896]
57. Jacquemyn J, Cascalho A, Goodchild RE. The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. *EMBO Reports*. 2017; 18: 1905–1921. [PubMed: 29074503]
58. Rossius, M, Hochgräfe, F, Antelmann, H. *Microbial Proteomics*. Springer; 2018. 261–275.
59. O'Donnell JP, et al. The architecture of EMC reveals a path for membrane protein insertion. *Elife*. 2020; 9 e57887 [PubMed: 32459176]
60. Palovcak E, et al. A simple and robust procedure for preparing graphene-oxide cryo-EM grids. *Journal of Structural Biology*. 2018; 204: 80–84. [PubMed: 30017701]
61. Shon MJ, Kim H, Yoon T-Y. Focused clamping of a single neuronal SNARE complex by complexin under high mechanical tension. *Nature Communications*. 2018; 9: 1–12.
62. Shon MJ, Rah S-H, Yoon T-Y. Submicrometer elasticity of double-stranded DNA revealed by precision force-extension measurements with magnetic tweezers. *Science Advances*. 2019; 5 eaav1697 [PubMed: 31206015]
63. Min D, Jefferson RE, Bowie JU, Yoon T-Y. Mapping the energy landscape for second-stage folding of a single membrane protein. *Nature Chemical Biology*. 2015; 11: 981–987. [PubMed: 26479439]
64. Bouchiat C, et al. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophysical Journal*. 1999; 76: 409–413. [PubMed: 9876152]
65. Oesterhelt F, et al. Unfolding pathways of individual bacteriorhodopsins. *Science*. 2000; 288: 143–146. [PubMed: 10753119]
66. Seol Y, Li J, Nelson PC, Perkins TT, Beternton M. Elasticity of short DNA molecules: theory and experiment for contour lengths of 0.6–7 μm . *Biophysical Journal*. 2007; 93: 4360–4373. [PubMed: 17766363]
67. Sarkar A, Caamano S, Fernandez JM. The mechanical fingerprint of a parallel polyprotein dimer. *Biophysical Journal*. 2007; 92: L36–L38. [PubMed: 17158577]
68. Gebhardt JCM, Bornschlöggl T, Rief M. Full distance-resolved folding energy landscape of one single protein molecule. *Proceedings of the National Academy of Sciences*. 2010; 107: 2013–2018.
69. Hinczewski M, von Hansen Y, Netz RR. Deconvolution of dynamic mechanical networks. *Proceedings of the National Academy of Sciences*. 2010; 107: 21493–21498.
70. Hinczewski M, Gebhardt JCM, Rief M, Thirumalai D. From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proceedings of the National Academy of Sciences*. 2013; 110: 4500–4505.
71. Santoro MM, Bolen D. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl. α -chymotrypsin using different denaturants. *Biochemistry*. 1988; 27: 8063–8068. [PubMed: 3233195]
72. Eastman P, et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*. 2017; 13 e1005659 [PubMed: 28746339]

73. Venable RM, et al. CHARMM all-atom additive force field for sphingomyelin: elucidation of hydrogen bonding and of positive curvature. *Biophysical Journal*. 2014; 107: 134–145. [PubMed: 24988348]
74. Klauda JB, et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The Journal of Physical Chemistry B*. 2010; 114: 7830–7843. [PubMed: 20496934]
75. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. 1983; 79: 926–935.
76. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*. 1977; 23: 327–341.
77. Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of Chemical Theory and Computation*. 2015; 11: 1864–1874. [PubMed: 26574392]
78. Steinbach PJ, Brooks BR. New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal of Computational Chemistry*. 1994; 15: 667–683.
79. Essmann U, et al. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*. 1995; 103: 8577–8593.
80. Chow K-H, Ferguson DM. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Computer Physics Communications*. 1995; 91: 283–289.
81. Åqvist J, Wennerström P, Nervall M, Bjelic S, Brandsdal BO. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters*. 2004; 384: 288–294.
82. Brooks BR, et al. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*. 2009; 30: 1545–1614. [PubMed: 19444816]
83. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996; 14: 33–38. [PubMed: 8744570]
84. Saier MH Jr, Tran CV, Barabote RD. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Research*. 2006; 34: D181–D186. [PubMed: 16381841]
85. Mirdita M, et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Research*. 2016; 45: D170–D176. [PubMed: 27899574]
86. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*. 2012; 9: 173–175.
87. Gabler F, et al. Protein sequence analysis using the MPI bioinformatics toolkit. *Current Protocols in Bioinformatics*. 2020; 72 e108 [PubMed: 33315308]

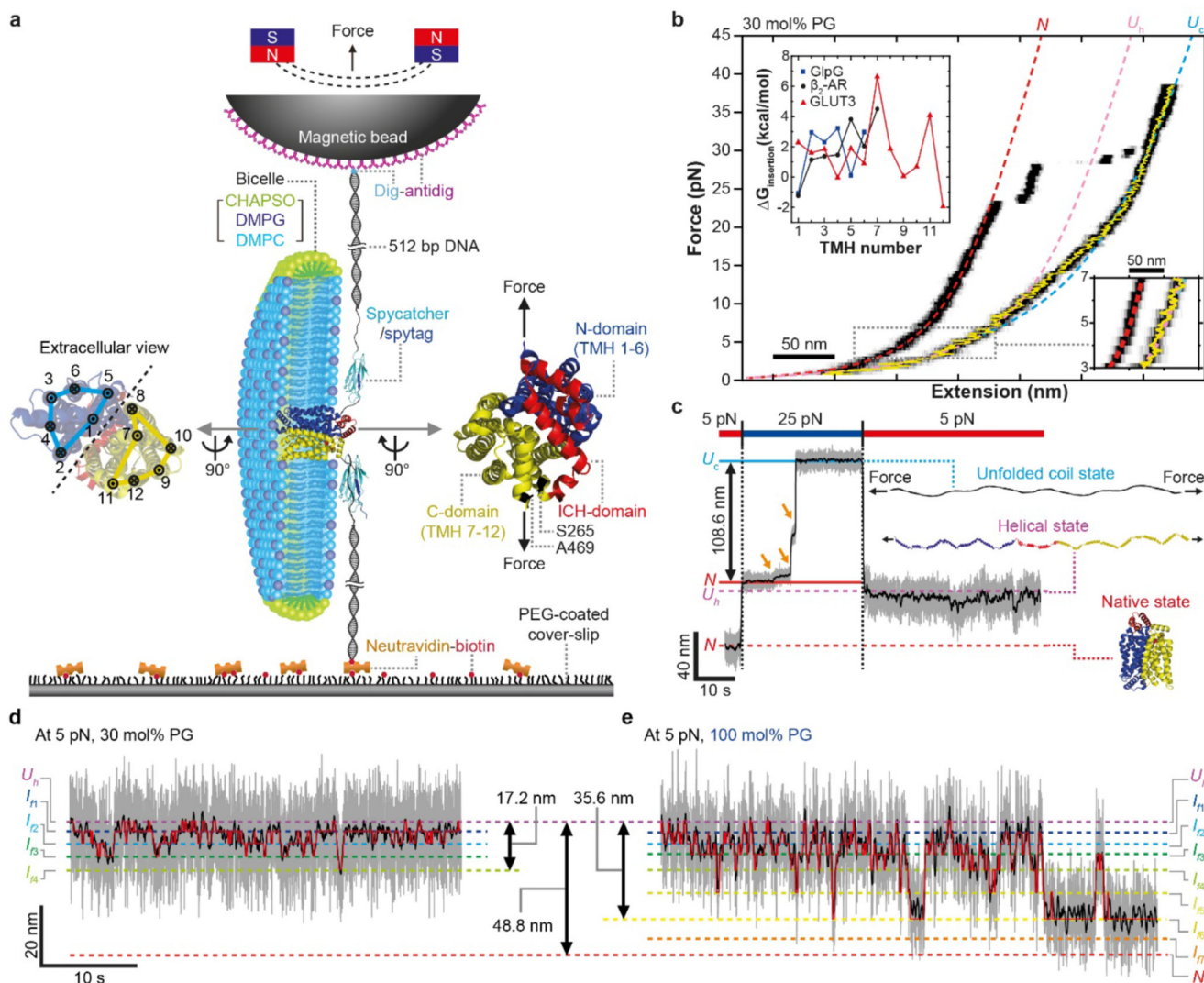


Fig. 1. Single-molecule magnetic tweezers assay for observing folding of single GLUT3 proteins

a, Schematic of magnetic tweezers (MT) experiment for observing folding of a single GLUT3 protein. Extracellular and intracellular view of GLUT3 structures are shown with TMH numbers and pulling positions depicted (left and right inset). **b**, Representative FEC of a single GLUT3 protein is shown as black heat map. The yellow trace shows the mean extension value in the relaxation phase. Theoretically expected FECs for the N , U_h , and U_c states are overlaid as red, pink, and light blue dashed lines, respectively. An observation is markedly different from our previous observations that FECs of GlpG and β_2 AR fell significantly shorter than U_h below 5 pN¹¹. Upper inset shows insertion energy values calculated for individual TMHs for *E. coli* GlpG, human β_2 AR and GLUT3 using the biological hydrophobicity scale from the translocon-ER membrane system²³. Lower inset shows a close-up view of FEC in the range of 3 pN to 7 pN force. **c**, Designed mechanical cycle for inducing refolding of a single GLUT3 protein. The gray and black traces are 1.2-kHz raw data and 5-Hz median-filtered data, respectively. **d**, **e**, Representative time-resolved traces for GLUT3 folding at 5 pN with 30 mol% (**d**) and 100 mol% DMPG (**e**) in bicelles.

The gray and black traces are defined as in (c). Red traces show the transitions between intermediates identified by HMM.

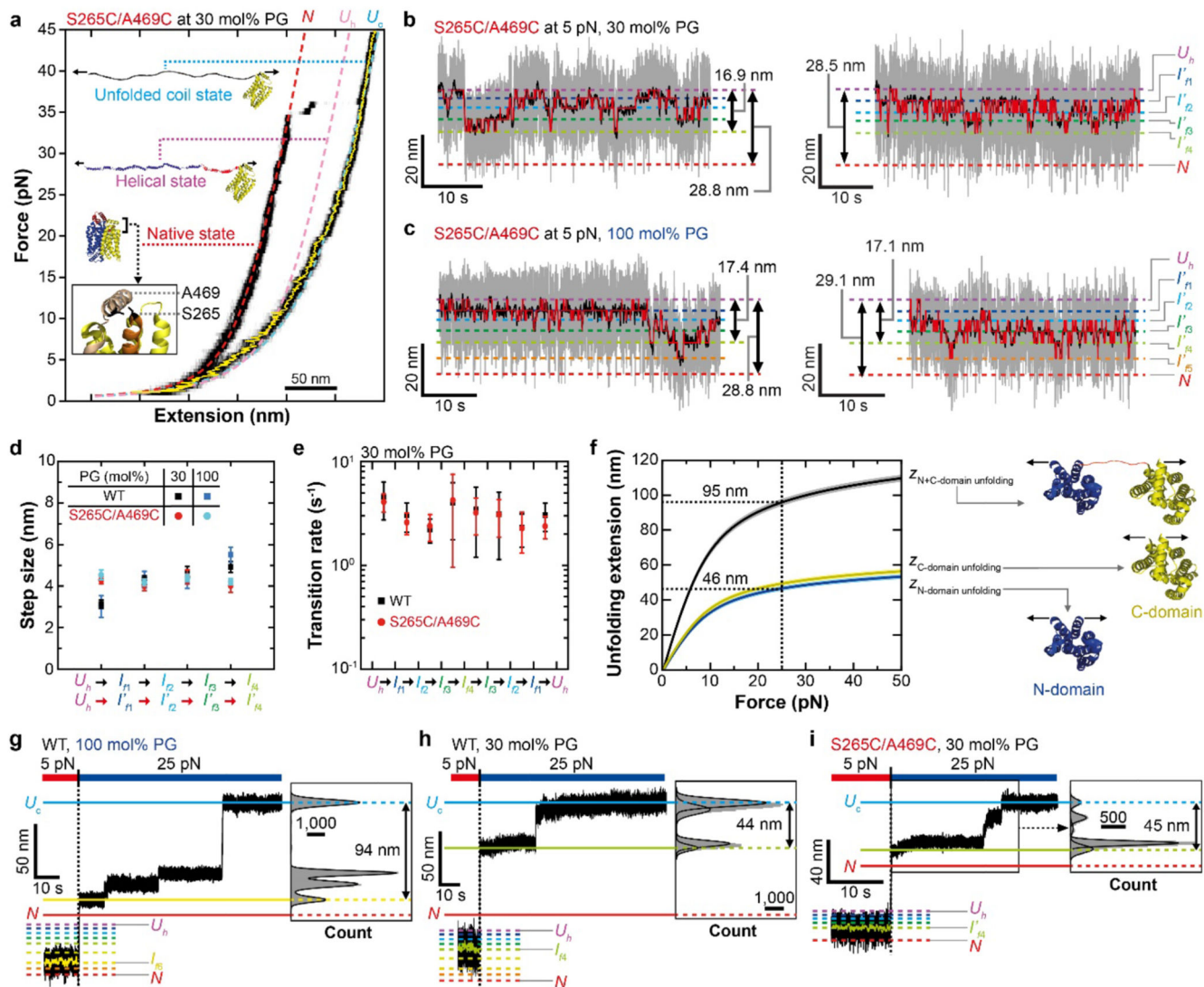


Fig. 2. Identification of the folding order of GLUT3's domains

a, Representative FEC of a single S265C/A469C GLUT3 protein (*i.e.*, GLUT_{CC}) shown as black heat map. Inset shows the position of cysteine mutations on TMHs 7 and 12. Other definitions are as in Fig. 1b. The unfolding step size of GLUT_{CC} under tension above 20 pN was almost halved compared with that WT GLUT3. **b,c**, Folding traces with HMM results for S265C/A469C GLUT3 folding at 5 pN with 30 mol% (**b**) and 100 mol% DMPG (**c**). Two replicates are shown for each condition, and each colored trace is defined as in Fig. 1d. **d**, Step sizes between the neighboring states at 5 pN ($n = 16, 11, 22$ and 12 traces for WT GLUT3 and GLUT_{CC} with 30 and 100 mol% DMPG, respectively). Error bars are SEM. **e**, Transition kinetics between the neighboring states at 5 pN. The number of traces is as in (**d**). Error bars are SEM. **f**, Estimated unfolding step sizes for linked N- and C-domains (black) and isolated N- (blue) and C-domains (yellow). The shaded area means SEM. **g-i**, Representative traces for the force-jump experiments applied to I_{f6} (J), I_{f4} (K), and I'_{f4} (L). The unfolding intermediate withstood the 25 pN tension for more than

hundreds of milliseconds. Unfolded parts were unraveled to unstructured polypeptides and dissociated from the bilayers in the course of high-force unfolding. These additional events incur large free energy costs ranging hundreds of $k_B T$, which presumably accounts for the increased stability of the partially folded structures during high-force unfolding. Insets show the distributions of extensions recorded after force jumps to 25 pN.

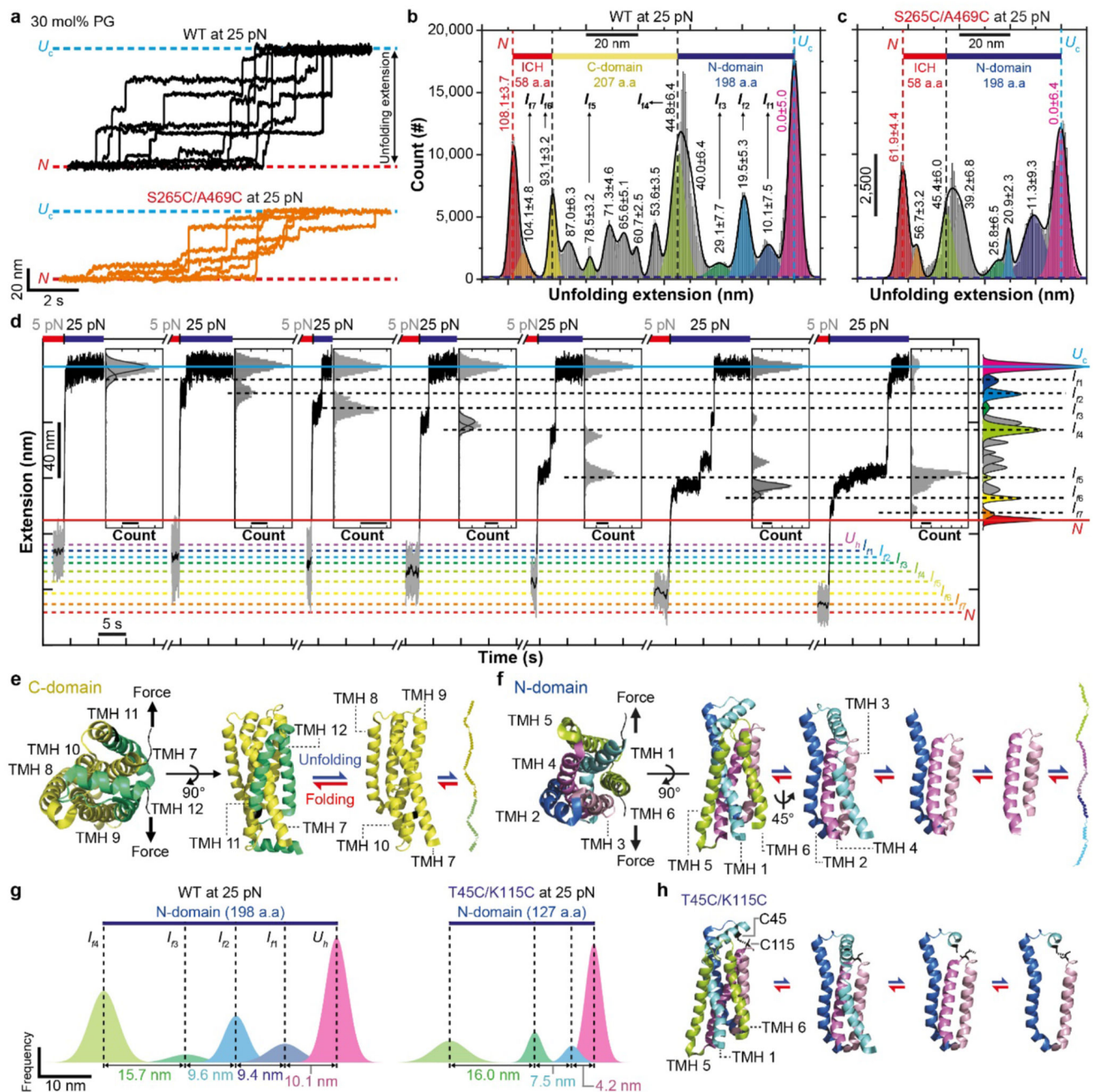


Fig. 3. Dissection of the folding order of individual MFS folds

a, Representative traces of high-force unfolding of single WT (black) and S265C/A469C (orange) GLUT3 proteins. A lipid composition of DMPC:DMPG=70:30 (mol/mol) was used for the bicelles. The traces were recorded at 1.2 kHz and subsequently median-filtered at 50 Hz. **b,c**, Distributions of extension values recorded during high-force unfolding. The peaks indicate the fit centers of multiple Gaussian functions. Relative extension values are measured from the U_c state and represent mean \pm SD. The upper diagrams depict the number of amino acids of corresponding domains to guide mapping onto the structure.

$n = 32$ and 18 high-force unfolding traces for WT (**b**) and S265C/A469C (**c**) GLUT3, respectively. **d**, Representative traces from the force-jump experiments applied to individual low-force folding intermediates. Each inset shows an extension distribution recorded after the force jump to 25 pN (scale bar is 500 count). Dashed lines indicate close alignment of the extension states after the force jumps with one of the unfolding peaks identified in (**b**). **e,f**, Schematics of folding and unfolding of the C- (E) and N-domain (F) of GLUT3. **g**, Intermediate structures identified for high-force unfolding of N-domain for single WT (left) and T45C/K115C (right) GLUT3 proteins. **h**, Schematics of folding and unfolding of the N-domain of T45C/K115C GLUT3.

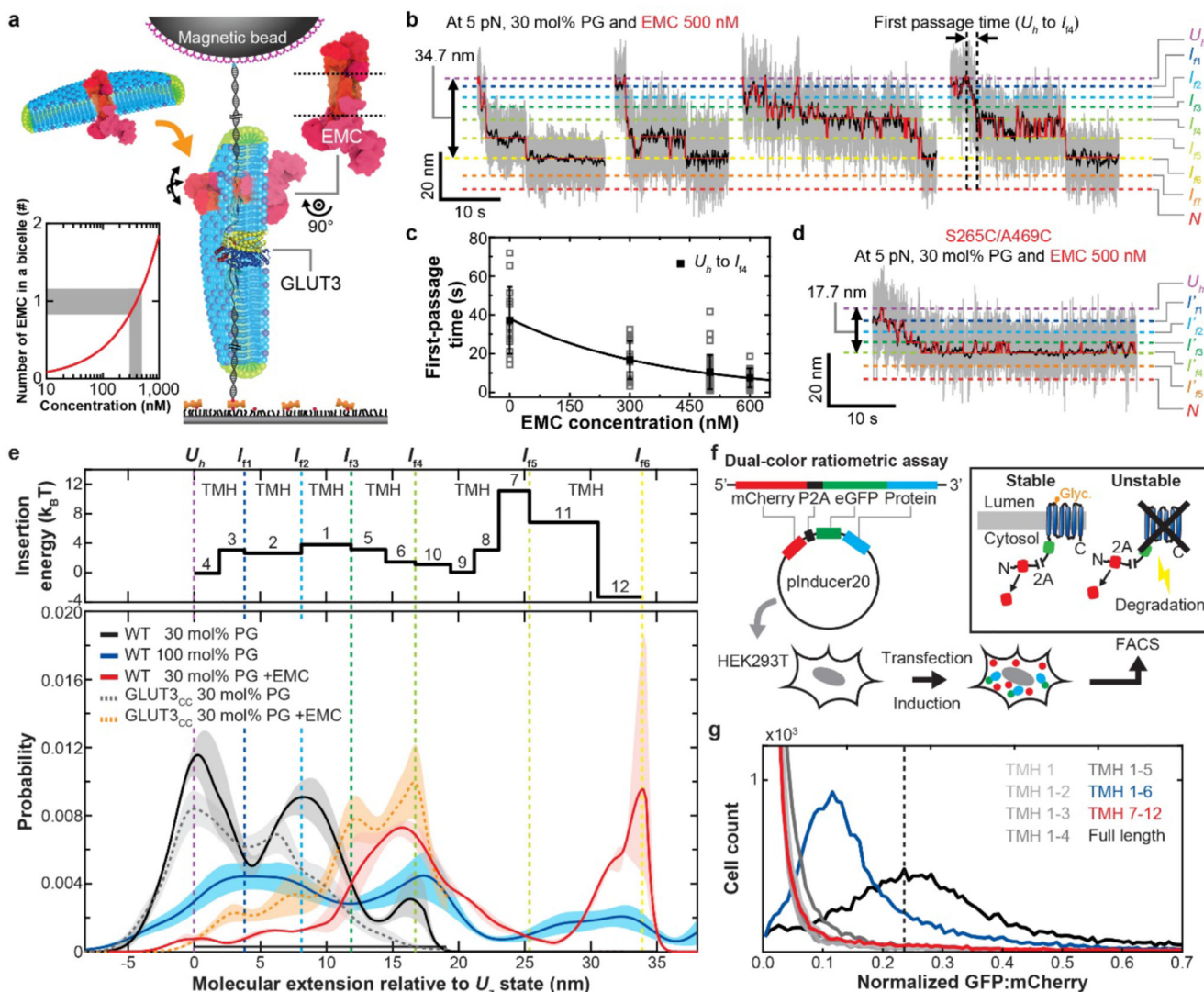


Fig. 4. EMC facilitates folding of single GLUT3 proteins

a, Schematic of an MT experiment that examines the effects of EMC on folding of GLUT3.

Inset shows the average number of EMCs in individual bicelles as a function of EMC concentration. **b**, Representative folding traces of WT GLUT3 obtained with 30 mol% DMPG and 500 nM EMC. Four replicates are shown. **c**, Mean first-passage time for I_{f4} determined for different EMC concentrations. Error bars mean SD ($n = 12, 15, 35$ and 15 traces for the cases with EMC = 0, 300, 500 and 600 nM, respectively). Curve fitting for estimation of free energy (black line, see Methods). **d**, Representative folding traces of GLUT3_{CC} obtained with 30 mol% DMPG and 500 nM EMC. **e**, Probability distributions of deconvoluted extension values observed under indicated folding conditions at 5 pN. The shaded area means SEM. Upper panel shows insertion energy values of individual TMHs aligned along the folding order identified in Fig. 3e,f. The insertion energy values were calculated based on the biological hydrophobicity scale from the translocon-ER membrane system as in upper inset of Fig. 1b. **f**, Schematics of the dual-color ratio-metric assay. **g**, FACS histogram of cell counts vs normalized GFP:mCherry.

Inset shows the average number of EMCs in individual bicelles as a function of EMC concentration. **b**, Representative folding traces of WT GLUT3 obtained with 30 mol% DMPG and 500 nM EMC. Four replicates are shown. **c**, Mean first-passage time for I_{f4} determined for different EMC concentrations. Error bars mean SD ($n = 12, 15, 35$ and 15 traces for the cases with EMC = 0, 300, 500 and 600 nM, respectively). Curve fitting for estimation of free energy (black line, see Methods). **d**, Representative folding traces of GLUT3_{CC} obtained with 30 mol% DMPG and 500 nM EMC. **e**, Probability distributions of deconvoluted extension values observed under indicated folding conditions at 5 pN. The shaded area means SEM. Upper panel shows insertion energy values of individual TMHs aligned along the folding order identified in Fig. 3e,f. The insertion energy values were calculated based on the biological hydrophobicity scale from the translocon-ER membrane system as in upper inset of Fig. 1b. **f**, Schematics of the dual-color ratio-metric assay. **g**, FACS histogram of cell counts vs normalized GFP:mCherry.

assay. FACS, fluorescence activated cell sorting and Glyc, glycosylation. **g**, Histograms of eGFP:mCherry ratios for the indicated constructs. Each histogram is normalized by the most probable eGFP:mCherry ratio determined for a control construct consisting only of eGFP and mCherry proteins separated by the 2A sequence. Full-length GLUT3 showed the highest stability among the GLUT3 constructs.

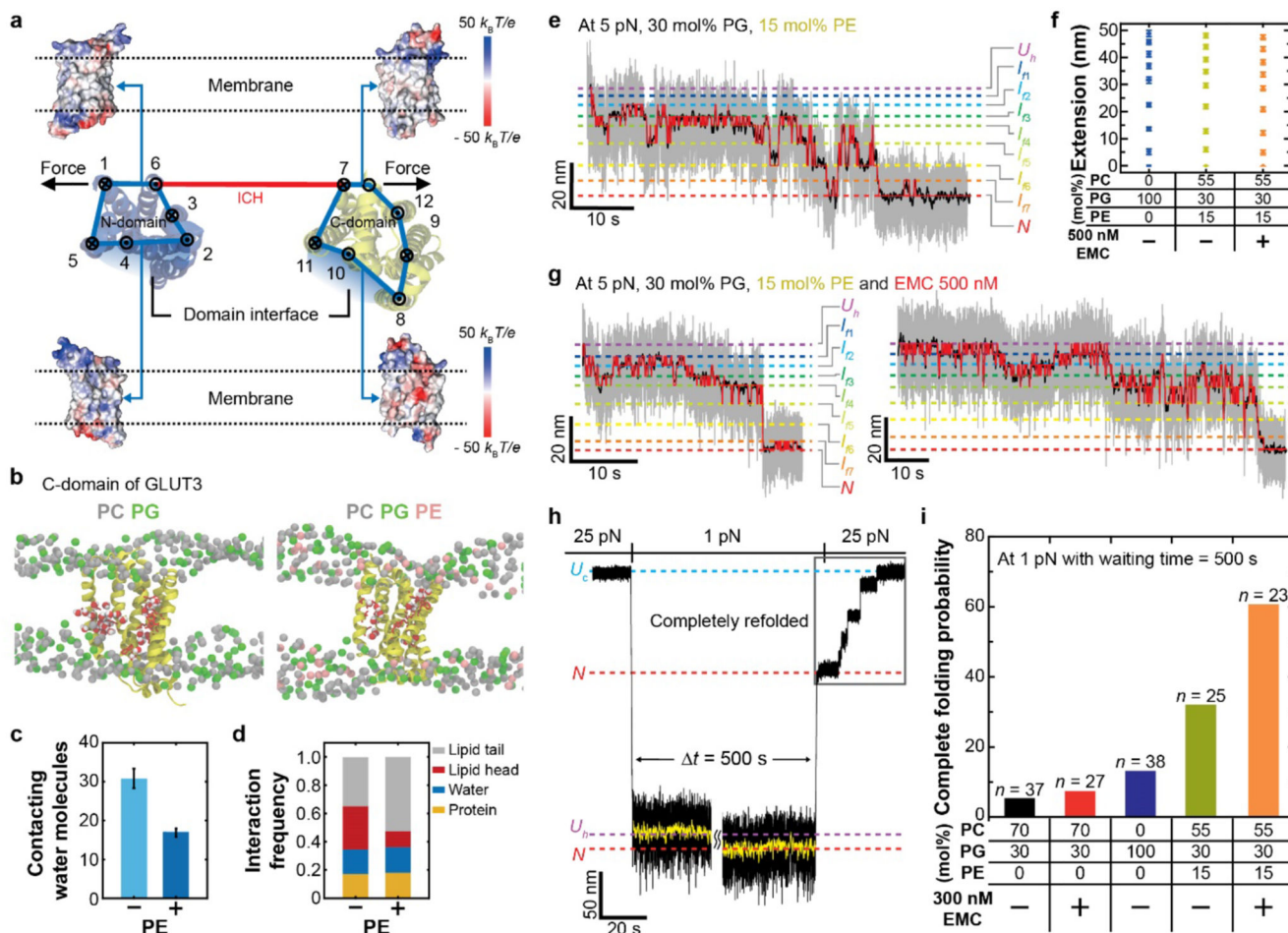


Fig. 5. PE lipids facilitate GLUT3's inter-domain assembly

a, Cartoon of a single GLUT3 protein at I_{f6} before domain-domain assembly. The N- and C-domains are folded and ICH-domains are stretched under mechanical tension. The electrostatic potentials of the outer and inner surfaces of GLUT3 are shown in upper and lower insets, respectively. **b**, Snapshots from MD simulations for isolated GLUT3 C-domains in mixed bilayers of 70 mol% DMPC and 30 mol% DMPG (left) and 55 mol% DMPC, 30 mol% DMPG and 15 mol% DMPE (right) at 296.15 K. DMPC, DMPG and DMPE lipid head groups are depicted as gray, green and pink spheres, respectively, and water molecules are shown as composites of red and white spheres. **c**, The average number of water molecules in contact with all listed residues under the simulation conditions defined in (b). The value is the average value from 0.6 μ s to 1.0 μ s. Error bars represent SD ($n = 4000$ for each case). **d**, Interaction profiles of interface-exposed residues (N315, T316, T319, E378, W386, N413). **e**, Representative time-resolved traces for folding of single GLUT3 proteins at 5 pN with 30 mol% DMPG and 15 mol% DMPE in the bicelles. Two replicates are shown, and the right inset is the close-up folding trajectory. **f**, Positions of the folding intermediates identified by HMM for denoted folding conditions. ($n = 11$, 11 and 10 traces for the cases with 100 mol% PG (blue), 30 mol% PG and 15 mol% PE without (yellow) and with (orange) EMCs, respectively). Error bars represent SEM. **g**,

Representative time-resolved traces for folding of single GLUT3 proteins at 5 pN with 30 mol% DMPG and 15 mol% DMPE in the bicelles in the presence of 500 nM EMC. Two replicates are shown. **h**, Representative traces for determination of the refolding probability with a force-jump experiment. **i**, Probability for observing the complete folding events under indicated conditions. n is the number of trials.

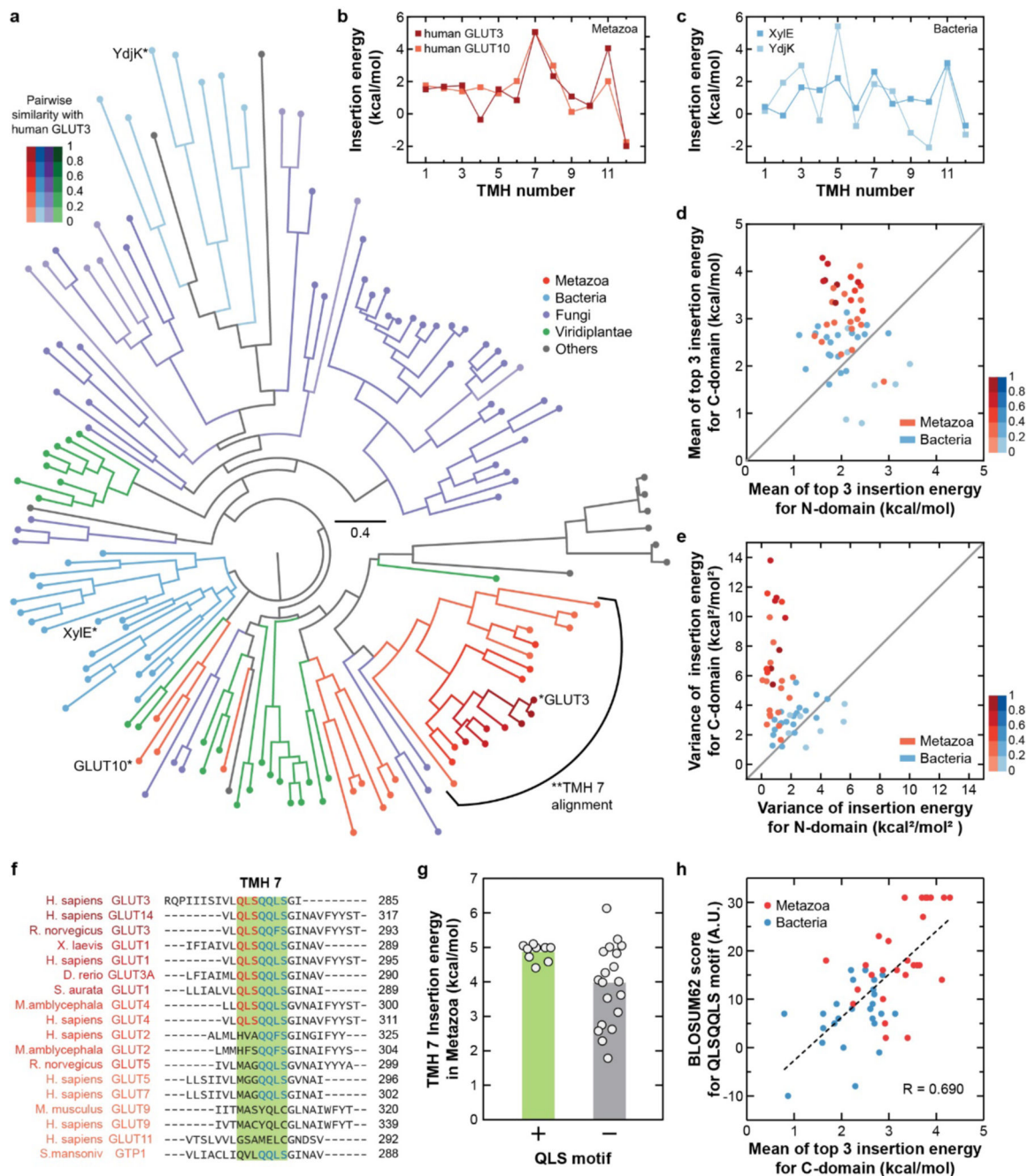


Fig. 6. Phylogenetic analysis reveals an asymmetric folding propensity across the N- and C-domains common for metazoan sugar transporters.

a, Phylogenetic tree of the MFS sugar porter family. Multiple sequence alignment of the sugar porter family was produced in hhsuite⁴⁴. Color code indicates sequence similarity of each protein to human GLUT3 and is applied to branches and nodes in the phylogenetic tree.

b,c, Representative plots for predicted TMH-insertion energy values calculated for human GLUT3 and GLUT10 (**b**), XyleE and YdjK (**c**). The energy was calculated DGpred with 19-amino acid window.

d,e, Scatter plots showing the mean of three highest TMH-insertion

energy values (**d**) and the variance of insertion energy values (**e**) for the N- (x-axis) and C-domain (y-axis) of each metazoan or bacterial transporter in the sugar porter family ($n = 28$ and 26 for metazoa and bacteria, respectively). **f**, Sequence alignment of TMH 7 for a subset of metazoan sugar transporters. The aligned transporters are shown in (**a**). **g**, Mean insertion energy values for TMH 7 calculated for all metazoan sugar transporters. The energy was calculated as in (**b**) ($n = 9$ and 19 for metazoan sugar transporters with and without QLS motif, respectively). **h**, Scatter plot showing the mean of the three highest insertion energy values (x-axis) and the BLOSUM62 score of QLS motif (y-axis) for each sugar porter. R means Pearson correlation coefficient ($n = 28$ and 26 for metazoa and bacteria, respectively).