nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-022-01002-w

Check for updates

Evolutionary balance between foldability and functionality of a glucose transporter

Hyun-Kyu Choi[1,2,6], Hyunook Kang[1,6], Chanwoo Lee[1,2,6], Hyun Gyu Kim[1,2], Ben P. Phillips[3], Soohyung Park[4], Charlotte Tumescheit[1], Sang Ah Kim[1,2], Hansol Lee[1,2], Soung-Hun Roh[1,2], Heedeok Hong[5], Martin Steinegger[1], Wonpil Im[4], Elizabeth A. Miller[3]✉, Hee-Jung Choi[1]✉ and Tae-Young Yoon[1,2]✉

Despite advances in resolving the 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 a structural template for its C-terminal counterpart. We found polar residues comprising the conduit for glucose molecules present major folding challenges. The endoplasmic reticulum 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 the lipid shells surrounding the domain interfaces. Sequence analysis suggests that this asymmetric folding 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 the flow of information and material across cell membranes[1]. Although cryo-EM is revealing the tertiary and quaternary structures of multi-pass membrane proteins at an unprecedented pace[2], 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 remarkable conservation in their tertiary structures despite huge evolutionary distances across different members[5,6]. 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[7,8]. It is thus tempting to hypothesize that the basic information for navigating the folding pathway—probably 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 *Escherichia coli* GlpG and human beta-2 adrenergic receptor (β2AR) with a resolution of a few amino acids in lipid bilayers[9]. Here, we extend our approach to study the folding pathway of GLUT3 as an archetype of the MFS that composes the largest group of solute carrier proteins (Fig. 1a)[1,10,11]. Like most members of the MFS, GLUT3 consists of two MFS folds, each with six TMHs (hereafter referred to as N- and C-domains) that pack against each other in a pseudo mirror-symmetry, connected by an intracellular helical (ICH) domain consisting of five helices (Fig. 1a)[7,11,12]. 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[13,14]. The interface between N- and C-domains is enriched with polar residues to create a conduit for glucose molecules in the otherwise impermeable lipid bilayers (Fig. 1a)[12].

By examining the folding of various GLUT3 constructs under disparate lipid bilayer conditions, we not only delineated a complete folding order for GLUT3 domains, but also dissected detailed pathways forming individual MFS folds. Although 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 probably linked to the enhancement or versatility of transporter functions. Thus, our observations reveal a conflict between foldability and functionality that has probably been faced by many membrane proteins throughout evolution[15].

Results

Single-molecule magnetic tweezers monitoring GLUT3 folding. To employ magnetic tweezers to observe the 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

[1]School of Biological Sciences, Seoul National University, Seoul, South Korea. [2]Institute for Molecular Biology and Genetics, Seoul National University, Seoul, South Korea. [3]Medical Research Council (MRC) Laboratory of Molecular Biology, Cambridge, UK. [4]Departments of Biological Sciences and Chemistry, Lehigh University, Bethlehem, PA, USA. [5]Department of Chemistry and Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, MI, USA. [6]These authors contributed equally: Hyun-Kyu Choi, Hyunook Kang, Chanwoo Lee. ✉e-mail: emiller@mrc-lmb.cam.ac.uk; choihj@snu.ac.kr; tyyoon@snu.ac.kr

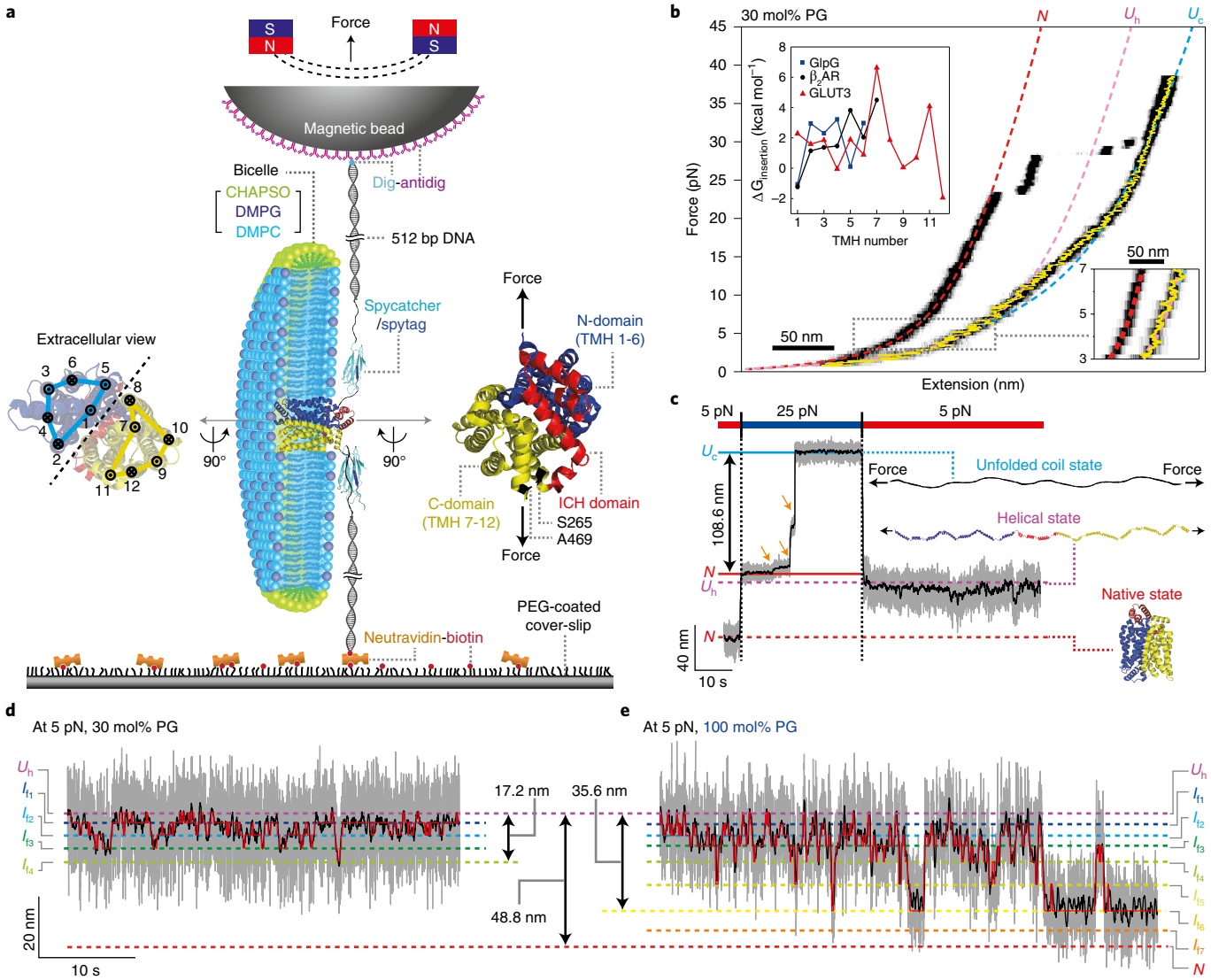


Fig. 1 | Single-molecule magnetic tweezers assay for observing folding of single GLUT3 proteins. a, Schematic of the magnetic tweezers experiment for observing the folding of a single GLUT3 protein. Extracellular and intracellular views of GLUT3 structures are shown with TMH numbers and pulling positions depicted (left and right inset). CHAPSO, 3-([3-cholamidopropyl] dimethylammonio)-2-hydroxy-1-propanesulfonate. **b**, Representative FEC of a single GLUT3 protein shown as a 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. This is markedly different from our previous observations that FECs of GlpG and β2AR fell substantially shorter than U_h below 5 pN (ref. [9]). Upper inset shows insertion energy values calculated for individual TMHs for *E. coli* GlpG, human β_2AR and GLUT3 using the biological hydrophobicity scale from the translocon-ER membrane system[20]. Lower inset shows a close-up view of FEC in the range 3–7 pN force. PG, DMPG. **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 in **c**. Red traces show the transitions between intermediates identified by HMM. Black arrows in (**a**) and (**c**) indicate the direction of exerted force.

and Extended Data Fig. 1a–c)[9,16]. After attaching the DNA handles to a magnetic bead and a polymer-coated surface, we introduced bicelle solutions, of varying lipidic compositions, to provide lipid bilayer environments to GLUT3 (Fig. 1a and Extended Data Fig. 1d–i)[17–19]. 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)[9].

We first examined the force extension curve (FEC) during gradual stretching and relaxation of single GLUT3 (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 over a force range from 20 to 10 pN. Because 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 12 TMHs of GLUT3. With further relaxing of the 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, probably due to the lower hydrophobicity of GLUT3 TMHs

compared with those of GlpG and β2AR (Fig. 1b, left inset and Supplementary Fig. 1)[20–22].

To observe folding of GLUT3, we applied a 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 became 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 β2AR (30 mol% of 1,2-dimyristoyl-*sn*-glycero-3-phosphorylglycerol (DMPG) and 70 mol% of 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC))[9], we found that none of the trials achieved complete refolding (Fig. 1d,e). We observed apparent partial folding comprising about 35% (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)[23,24] 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[9,19], 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 (an extension decrease of 35.6 nm; Fig. 1e and Extended Data Fig. 2e). 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. 2d,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 to the observation for wild-type (WT) GLUT3 in which 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).

On the basis of these results, we propose that the first four intermediates (I_{f1} to I_{f4}) correspond to folding of the GLUT3 N-domain. The following two folding intermediates (I_{f5} and I_{f6}), which could be accessed in the 100 mol% DMPG condition for WT GLUT3 but vanished for $GLUT3_{CC}$, are attributed to C-domain folding. To determine whether partial structures were 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 before 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 allowed the establishment of a crucial one-to-one correspondence between the low-force and 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 refolding process (for example, Figs. 1e and 3b). In such two-step folding, the folding process should start from either the 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 step to finish C-domain folding and thus be incompatible with the observed two-step folding. Furthermore, the unfolding step from I_{f5} to I_{f4} corresponded to two-thirds of the unfolding extension of the entire C-domain, suggesting that the folding step in the reverse direction (from I_{f4} to I_{f5}) would involve four 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 (refs. [6,12,25]), 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 assign 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 between them, which probably leaves 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 with four-step folding of the N-domain (Fig. 3g,h and see Extended Data Figs. 4d,e and 5 for details), in which TMHs 3 and 4

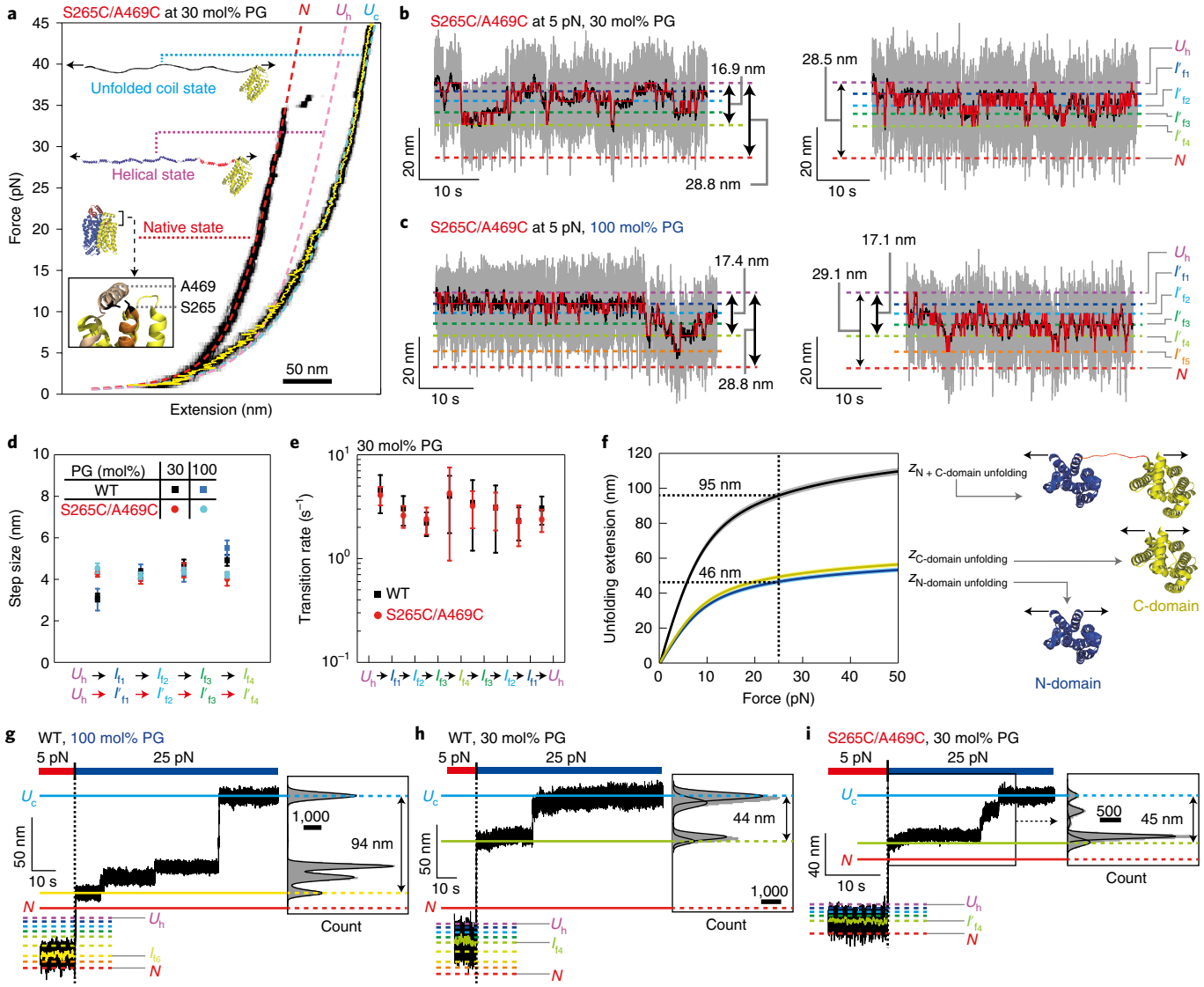


Fig. 2 | Identification of the folding order of GLUT3 domains. a, Representative FEC of a single S265C/A469C GLUT3 protein ($GLUT3_{CC}$) shown as a black heat map. The inset shows the position of cysteine mutations on TMHs 7 and 12. Other definitions are as in Fig. 1b. The unfolding step size of $GLUT3_{CC}$ under tension above 20 pN was almost halved compared with 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 in Fig. 1d. **d**, Step sizes between the neighboring states at 5 pN ($n = 16$, 11, 22 and 12 traces for WT GLUT3 and $GLUT3_{CC}$ with 30 and 100 mol% DMPG, respectively). Error bars show s.e.m. **e**, Transition kinetics between neighboring states at 5 pN. The number of traces is the same as in **d**. Error bars show s.e.m. **f**, Estimated unfolding step sizes for linked N- and C-domains (black) and isolated N- (blue) and C-domains (yellow). The shaded area indicates the s.e.m. **g**–**i**, Representative traces for the force-jump experiments applied to I_{f6} (**g**), I_{f4} (**h**) and I'_{f4} (**i**). The unfolding intermediate withstood the 25 pN tension for hundreds of milliseconds or more. 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 of hundreds of k_BT, 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.

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 by addition of TMHs 5 and 6 (I_{f3} to I_{f4}) (Fig. 3f). This folding order allowed us to group the first- and last-folding three TMHs as helix triplets (Extended Data Fig. 6a). This new grouping generates fully packed, stable helix triplets that can be simply joined together to form the MFS fold without involving complicated topological entanglement (Extended Data Fig. 6b versus 6c).

EMC facilitates insertion of hydrophilic TMHs of GLUT3. Our observations indicated that GLUT3 has a weaker propensity for folding than GlpG and β2AR 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 EMC[26], a large multiprotein complex with nine members in humans[27] (Extended Data Fig. 7a,b). EMC is shown to induce effective membrane insertion of tail-anchored proteins and the first TMHs of G-protein-coupled receptors[28,29]. Specifically, this membrane insertase activity is manifested when TMHs of target proteins exhibit lower levels of hydrophobicity[28–30].

We purified human EMC and added the complex reconstituted in bicelles to our single-molecule magnetic tweezers assay (Fig. 4a and Extended Data Fig. 7c). We anticipated that EMCs could be

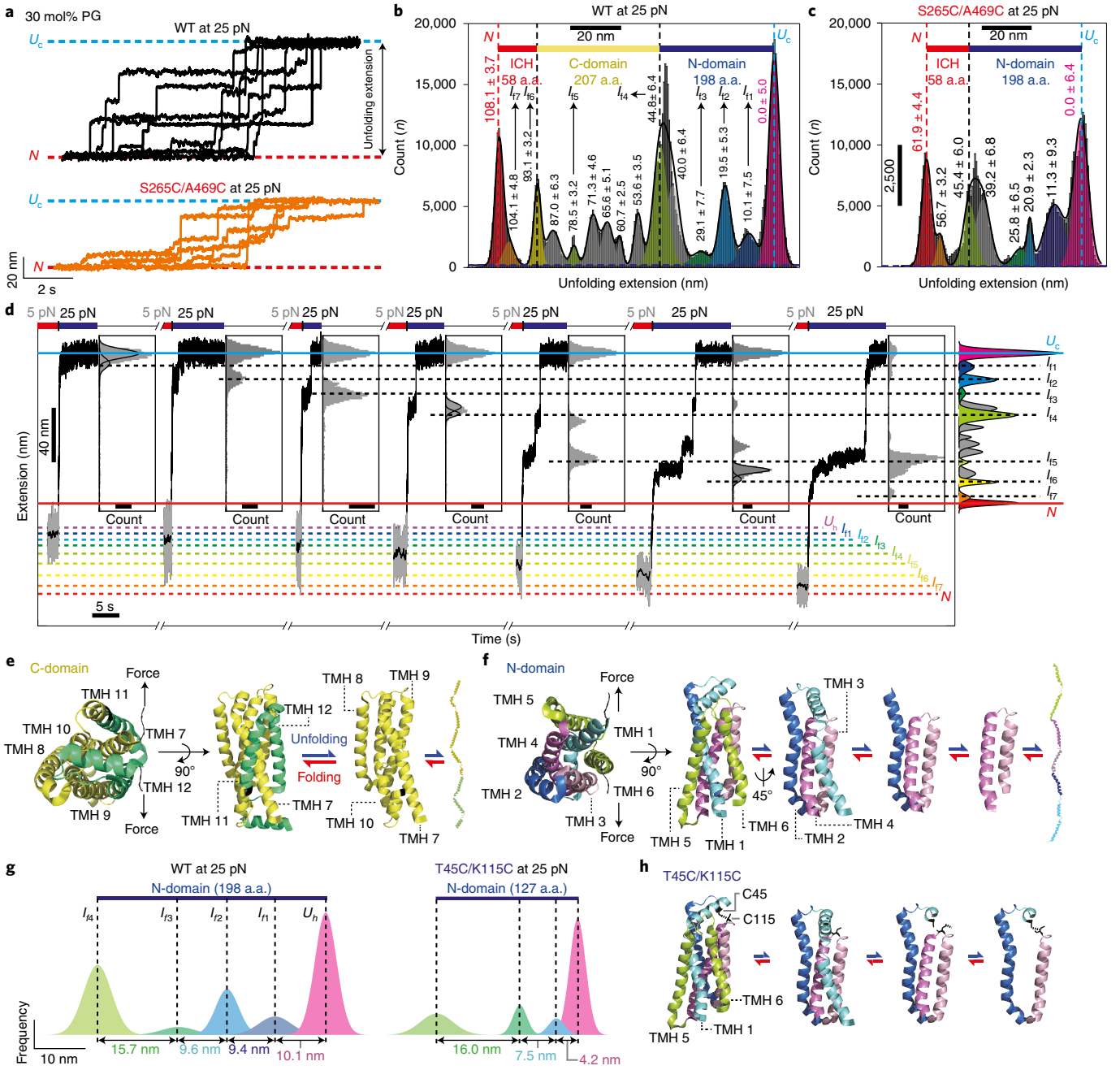


Fig. 3 | Dissection of the folding order of individual MFS folds. a, Representative traces of high-force unfolding of single WT (black) and S265C/A469C (orange) GLUT3 proteins. A lipid composition of DMPC/DMPG = 70:30 (mol/mol) was used for the bicelles. Traces were recorded at 1.2 kHz and subsequently median-filtered at 50 Hz. **b**,**c**, Distributions of extension values recorded during high-force unfolding. Peaks indicate the fit centers of multiple Gaussian functions. Relative extension values are measured from the U_c state and represent mean ± s.d. 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, 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. a.a., amino acids.

delivered to tweezed single GLUT3 proteins because individual bicelles undergo frequent fusion and fission with one another[31]. 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 single GLUT3 folding traces progressed as far as ~34.7 nm, a direct indication of the contribution of EMC to successful folding of the entire N- and C-domains (Fig. 4b and Extended Data Fig. 7d). 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

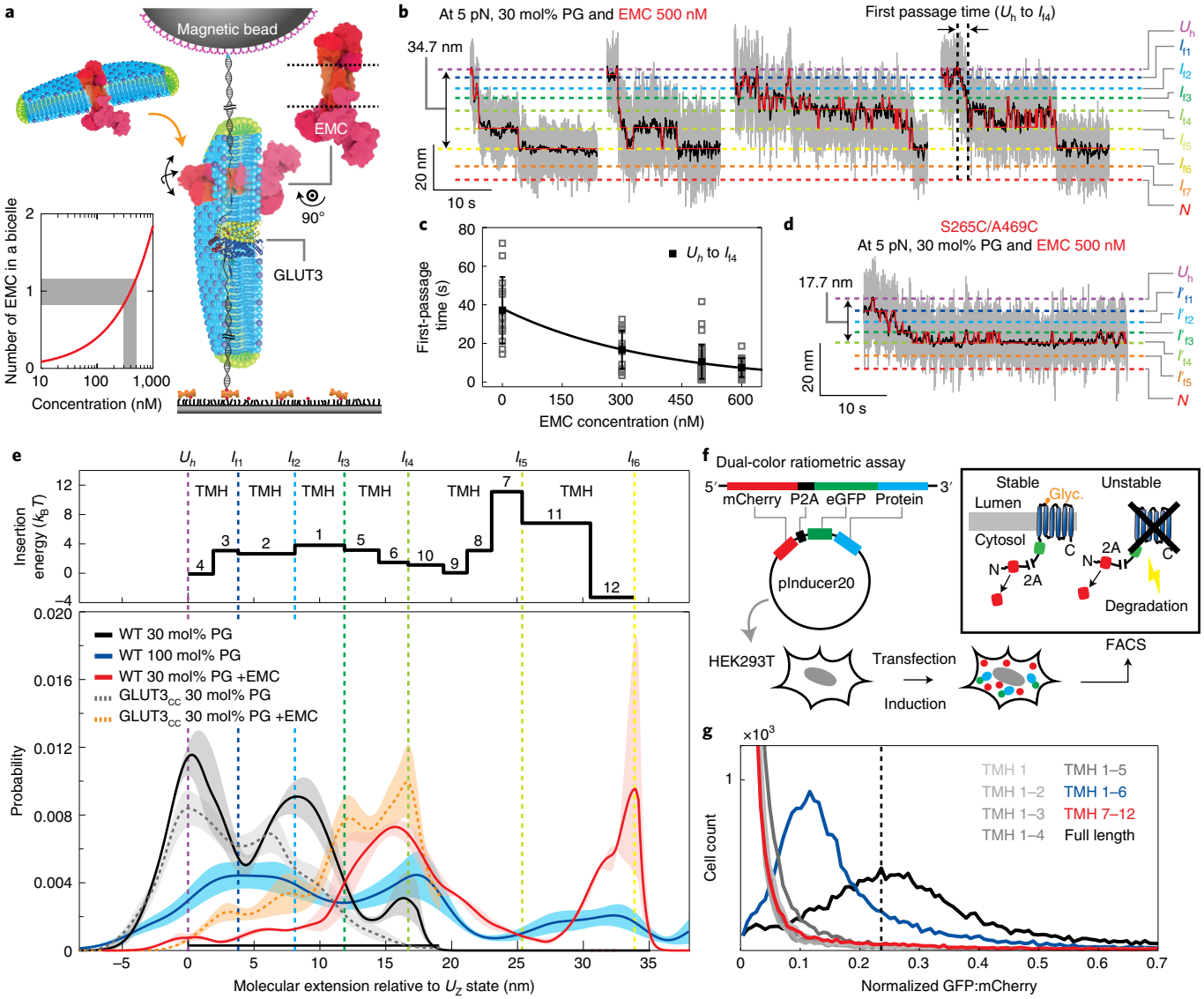


Fig. 4 | EMC facilitates folding of single GLUT3 proteins. a, Schematic of an MT experiment that examines the effects of EMC on folding of GLUT3. Inset shows the average number of EMCs in individual bicelles as a function of EMC concentration. **b**, Representative folding traces of WT GLUT3 obtained 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 indicate s.d. (n = 12, 15, 35 and 15 traces for the cases with EMC = 0, 300, 500 and 600 nM, respectively). **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 the indicated folding conditions at 5 pN. The shaded area indicates the s.e.m. The upper panel shows insertion energy values of individual TMHs aligned along the folding order identified in Fig. 3e,f. Insertion energy values were calculated on the basis of the biological hydrophobicity scale from the translocon-ER membrane system as shown in the upper inset of Fig. 1b. **f**, Schematics of the dual-color ratiometric assay. FACS, fluorescence activated cell sorting; 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. 7e–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 higher EMC concentrations were used (Fig. 4c). Using 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 repeated for C-domain-knotted GLUT3$_{CC}$ (Fig. 4b,d and Extended Data Fig. 7d,h,i). These observations suggest that EMC helps GLUT3 navigate down the folding intermediates encoded in its native amino acid sequence, rather than creating new folding pathways.

To examine the effects of EMC on GLUT3 folding at 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 versus black distributions). Comparing the relative populations allowed us to estimate lowering of the free-energy landscape by ~4 k_BT at I_{f4} in the presence of EMC (Supplementary Fig. 5). In addition, we note a major valley in the extension distribution at

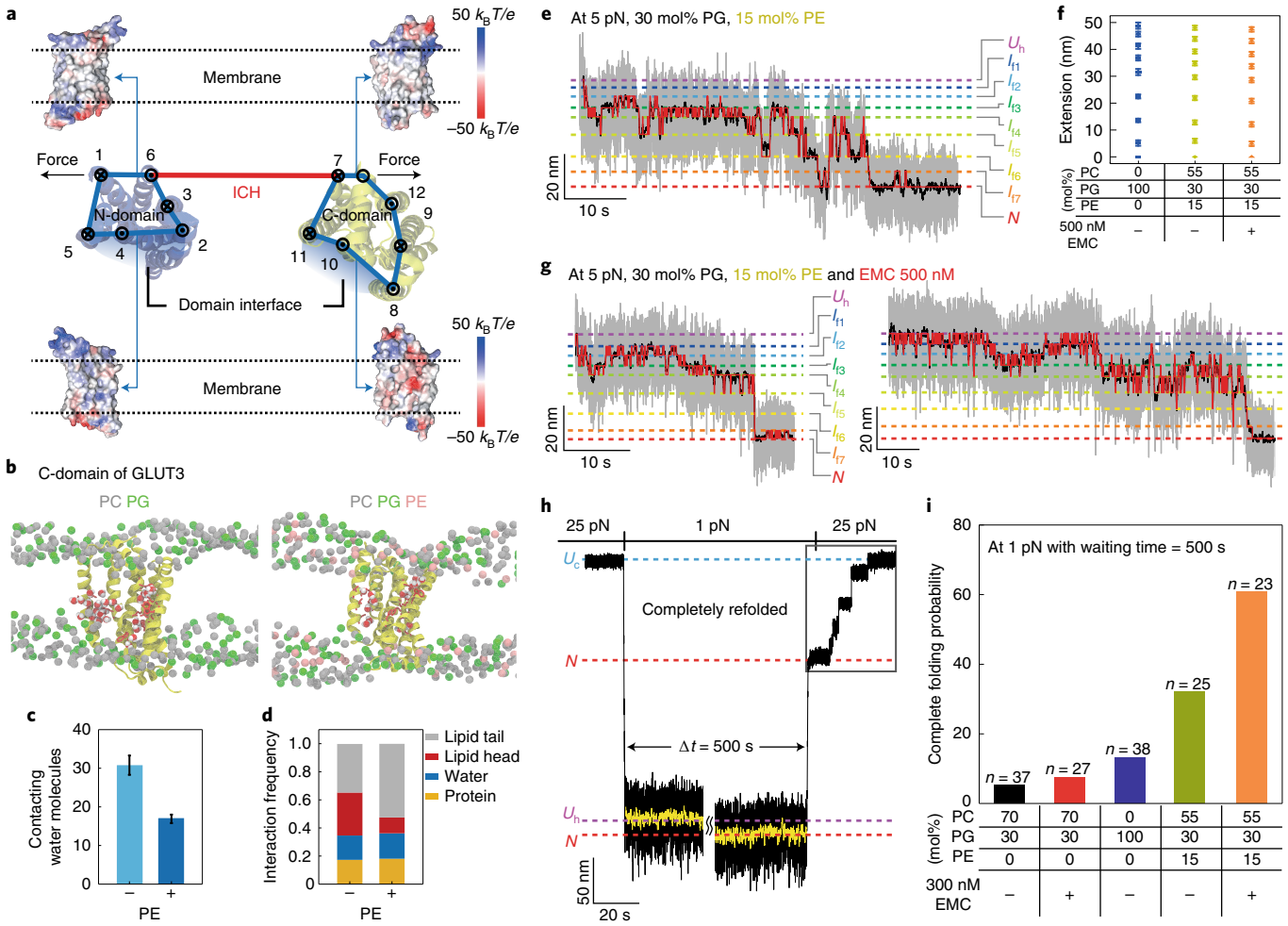


Fig. 5 | PE lipids facilitate GLUT3 inter-domain assembly. a, Cartoon of a single GLUT3 protein at I_{f6} before domain–domain assembly. The N- and C-domains are folded and ICH domains are stretched under mechanical tension. The electrostatic potentials of the outer and inner surfaces of GLUT3 are shown in the 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**, Average number of water molecules in contact with residues under the simulation conditions defined in **b**. The value is the average from 0.6 to 1.0 μs. Error bars represent s.d. (n = 4,000 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. **f**, Positions of the folding intermediates identified by HMM for denoted folding conditions (n = 11, 11 and 10 traces for cases with 100 mol% PG (blue), 30 mol% PG and 15 mol% PE without (yellow) and with (orange) EMCs, respectively). Error bars represent s.e.m. **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 the determination of the refolding probability with a force-jump experiment. **i**, Probability of observing the complete folding events under indicated conditions. n, number of trials.

around 25 nm, a major setback for GLUT3's efforts at C-domain folding, which was also observed with the 100 mol% DMPG condition (Fig. 4e, red versus blue distributions). Remarkably, this valley approximately coincides with the folding steps of TMH 7 (a broken helix) and TMH 11, which 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 the physiological relevance of our observations, we adopted the dual-color ratiometric assay that examines membrane protein stabilities in the cellular milieu (Fig. 4f)[28,29]. All partial constructs of the N-domain, even that missing only TMH 6 (that is, TMHs 1 to 5), showed negligible expression signals relative to an expression benchmark (Fig. 4g). Although the complete N-domain construct showed increased stability compared with 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 complete N-domain constructs were mitigated in EMC-knockout cells, suggesting that the biogenesis of GLUT3 was dependent on EMC in the cellular milieu (Supplementary Figs. 6b,c and 7). 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. 7d).

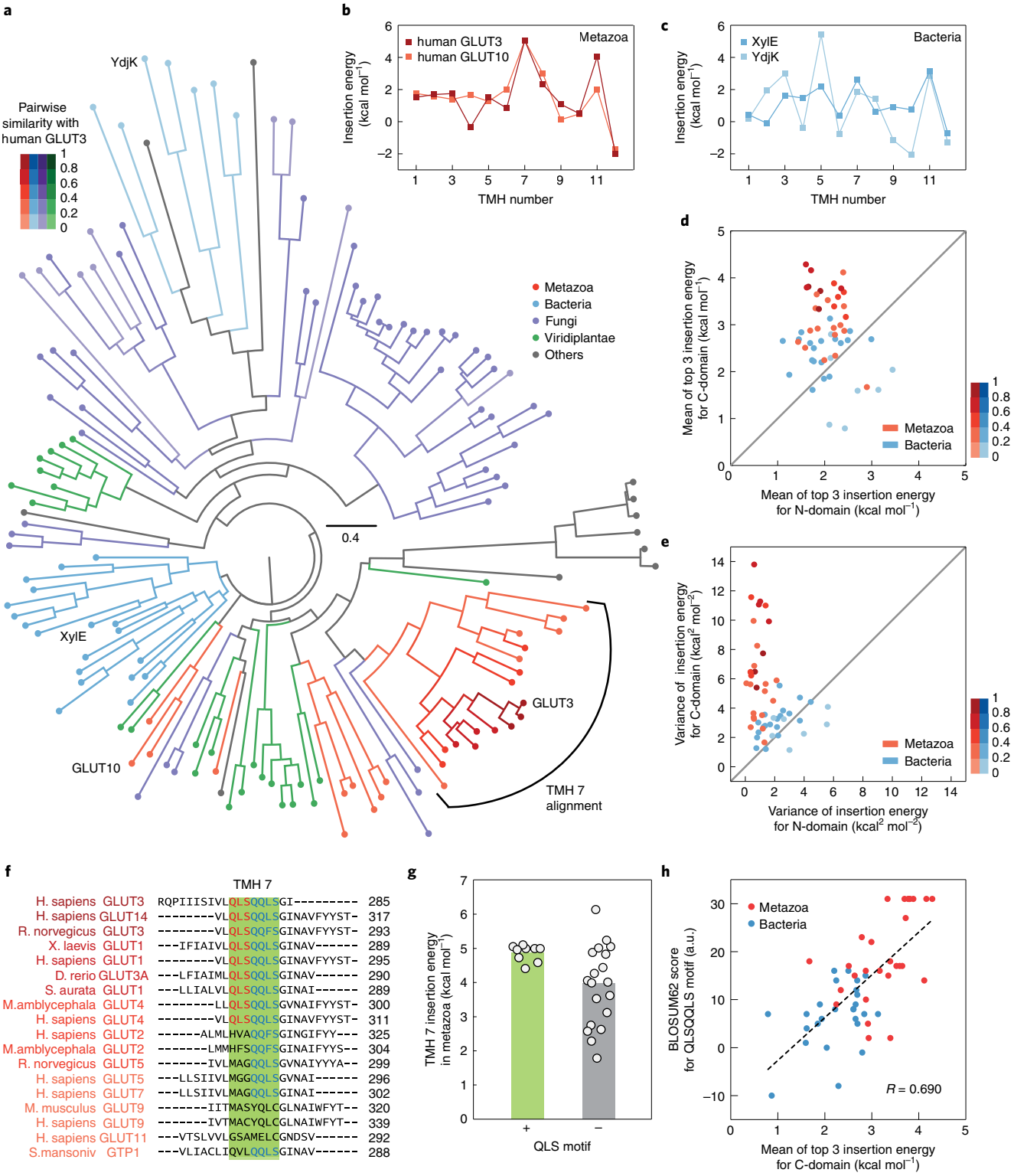


Fig. 6 | Phylogenetic analysis reveals an asymmetric folding propensity across the N- and C-domains common for metazoan sugar transporters. **a**, Phylogenetic tree of the MFS sugar porter family. Multiple sequence alignment of the sugar porter family was produced in hhsuite[36]. The color code indicates the 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**), and XylE and YdjK (**c**). The energy was calculated using DGpred with a 19-amino acid window. **d**,**e**, Scatter plots showing the mean of the three highest TMH insertion energy values (**d**) and the variance in 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 is the Pearson correlation coefficient ($n = 28$ and 26 for metazoa and bacteria, respectively). a.u., arbitrary units.

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). Because 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 µs using the CHARMM force field[32]. The MD simulation results suggest that the high polar/charged residue content on the interface between N- and C-domains induces 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. 8). 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 proteins[33].

We further reasoned that if the membrane shells between the N- and C-domains define a major barrier to domain–domain assembly, the lipid bilayer composition might play a pivotal role in the final step of GLUT3 folding[34,35]. 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 the lipid headgroup markedly decreased with the inclusion of phosphoethanolamine (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). In the 30 mol% DMPE condition, folding probability after 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. 9). Addition of EMC alone (7.6%) or switching to 100 mol% DMPG (13.2%) marginally increased the folding probability, consistent with our observations that these conditions facilitated C-terminal domain folding, but not 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 for human GLUT3 also hold for other sugar transporters that exist across all domains of life[10]. 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 regions[36]. We then estimated the insertion energy values of putative TMHs using the biological hydrophobicity scale (Fig. 6b,c)[20].

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 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 the GLUT3 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-domain[11], 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 binding[37], 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^{-1}, 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[6,11], we expect the folding order described here 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. Symmetry is a prevailing feature in the conserved structural folds of membrane proteins[6,38]. 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.

At 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[9,39]. 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[10]—might have formed homo- or heterodimers, in which both foldability and functionality would be managed within a single subunit[40]. 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 acquirement of the QLS motif, the metazoan sugar transporters seem to have implemented versatility in transporter functions with widely differing Michaelis–Menton constants (K_M) and catalytic rates (k_{cat}). For example, whereas

GLUT3 is mainly expressed in neurons and transports glucose molecules with a high turnover rate ($k_{cat} > 1{,}000\,s^{-1}$)[41], GLUT2 is expressed in beta cells and mainly works as a glucose sensor with its uniquely high K_M (refs. [42,43]). 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 the 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 (for example, 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 YidC[44], GET1 (ref. [45]) and TMCO1 (ref. [46]), belong to the Oxa1 superfamily that forms a remarkably conserved family of insertases[47]. We found that even with 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. Although PE-headgroup lipids are known to affect TMH orientations and establishment of a right topology of TMHs[48], our results suggest a new role for 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[49,50].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-022-01002-w.

Received: 3 August 2021; Accepted: 25 February 2022;
Published online: 28 April 2022

References

1. Hediger, M. A. et al. The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflügers Arch.* **447**, 465–468 (2004).
2. Cheng, Y. Membrane protein structural biology in the era of single particle cryo-EM. *Curr. Opin. Struct. Biol.* **52**, 58–63 (2018).
3. Marinko, J. T. et al. Folding and misfolding of human membrane proteins in health and disease: from single molecules to cellular proteostasis. *Chem. Rev.* **119**, 5537–5606 (2019).
4. Guna, A. & Hegde, R. S. Transmembrane domain recognition during membrane protein biogenesis and quality control. *Curr. Biol.* **28**, R498–R511 (2018).
5. Oberai, A., Ihm, Y., Kim, S. & Bowie, J. U. A limited universe of membrane protein families and folds. *Protein Sci.* **15**, 1723–1734 (2006).
6. Shi, Y. Common folds and transport mechanisms of secondary active transporters. *Annu. Rev. Biophys.* **42**, 51–72 (2013).
7. Chitwood, P. J. & Hegde, R. S. An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature* **584**, 630–634 (2020).
8. Kota, J. & Ljungdahl, P. O. Specialized membrane-localized chaperones prevent aggregation of polytopic proteins in the ER. *J. Cell Biol.* **168**, 79–88 (2005).
9. Choi, H.-K. et al. Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science* **366**, 1150–1156 (2019).
10. Pao, S. S., Paulsen, I. T. & Saier, M. H. Jr. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1–34 (1998).
11. Yan, N. Structural biology of the major facilitator superfamily transporters. *Annu. Rev. Biophys.* **44**, 257–283 (2015).
12. Deng, D. et al. Molecular basis of ligand recognition and transport by glucose transporters. *Nature* **526**, 391–396 (2015).
13. Madej, M. G. Function, structure, and evolution of the major facilitator superfamily: the LacY manifesto. *Adv. Biol.* **2014**, 1–20 (2014).
14. Drew, D. & Boudker, O. Shared molecular mechanisms of membrane transporters. *Annu. Rev. Biochem.* **85**, 543–572 (2016).
15. Hingorani, K. S. & Gierasch, L. M. Comparing protein folding in vitro and in vivo: foldability meets the fitness challenge. *Curr. Opin. Struct. Biol.* **24**, 81–90 (2014).
16. Min, D., Arbing, M. A., Jefferson, R. E. & Bowie, J. U. A simple DNA handle attachment method for single molecule mechanical manipulation experiments. *Protein Sci.* **25**, 1535–1544 (2016).
17. Findlay, H. E. & Booth, P. J. The biological significance of lipid–protein interactions. *J. Phys. Condens. Matter* **18**, S1281–S1291 (2006).
18. Ujwal, R. & Bowie, J. U. Crystallizing membrane proteins using lipidic bicelles. *Methods* **55**, 337–341 (2011).
19. Findlay, H. E. & Booth, P. J. The folding, stability and function of lactose permease differ in their dependence on bilayer lipid composition. *Sci. Rep.* **7**, 13056 (2017).
20. Hessa, T. et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030 (2007).
21. Snider, C., Jayasinghe, S., Hristova, K. & White, S. H. MPEx: a tool for exploring membrane proteins. *Protein Sci.* **18**, 2624–2628 (2009).
22. Moon, C. P. & Fleming, K. G. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl Acad. Sci. USA* **108**, 10174–10177 (2011).
23. Lee, T.-H. Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden Markov models. *J. Phys. Chem. B* **113**, 11535–11542 (2009).
24. Zhang, Y., Jiao, J. & Rebane, A. A. Hidden Markov modeling with detailed balance and its application to single protein folding. *Biophys. J.* **111**, 2110–2124 (2016).
25. Radestock, S. & Forrest, L. R. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *J. Mol. Biol.* **407**, 698–715 (2011).
26. Jonikas, M. C. et al. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009).
27. Pleiner, T. et al. Structural basis for membrane insertion by the human ER membrane protein complex. *Science* **369**, 433–436 (2020).
28. Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S. & Hegde, R. S. EMC is required to initiate accurate membrane protein topogenesis. *Cell* **175**, 1507–1519.e16 (2018).
29. Guna, A., Volkmar, N., Christianson, J. C. & Hegde, R. S. The ER membrane protein complex is a transmembrane domain insertase. *Science* **359**, 470–473 (2018).
30. Kauko, A. et al. Repositioning of transmembrane α-helices during membrane protein folding. *J. Mol. Biol.* **397**, 190–201 (2010).
31. Lai, G. & Renthal, R. Integral membrane protein fragment recombination after transfer from nanolipoprotein particles to bicelles. *Biochemistry* **52**, 9405–9412 (2013).
32. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **97**, 50–58 (2009).
33. Cheung, M. S., García, A. E. & Onuchic, J. N. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl Acad. Sci. USA* **99**, 685–690 (2002).
34. Bowie, J. U. Solving the membrane protein folding problem. *Nature* **438**, 581–589 (2005).
35. Bogdanov, M. & Dowhan, W. Lipid-dependent generation of dual topology for a membrane protein. *J. Biol. Chem.* **287**, 37939–37948 (2012).
36. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinform.* **20**, 473 (2019).
37. Seatter, M. J., De La Rue, S. A., Porter, L. M. & Gould, G. W. 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* **37**, 1322–1326 (1998).
38. Forrest, L. R. Structural symmetry in membrane proteins. *Annu. Rev. Biophys.* **44**, 311–337 (2015).
39. Cymer, F. & von Heijne, G. Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proc. Natl Acad. Sci. USA* **110**, 14640–14645 (2013).
40. Lolkema, J. S., Dobrowolski, A. & Slotboom, D.-J. Evolution of antiparallel two-domain membrane proteins: tracing multiple gene duplication events in the DUF606 family. *J. Mol. Biol.* **378**, 596–606 (2008).
41. Walmsley, A. R., Barrett, M. P., Bringaud, F. & Gould, G. W. Sugar transporters from bacteria, parasites and mammals: structure–activity relationships. *Trends Biochem. Sci.* **23**, 476–481 (1998).

42. Mueckler, M. & Thorens, B. The SLC2 (GLUT) family of membrane transporters. *Mol. Asp. Med.* **34**, 121–138 (2013).
43. Thorens, B. GLUT2, glucose sensing and glucose homeostasis. *Diabetologia* **58**, 221–232 (2015).
44. Serdiuk, T. et al. YidC assists the stepwise and stochastic folding of membrane proteins. *Nat. Chem. Biol.* **12**, 911–917 (2016).
45. Wang, F., Chan, C., Weir, N. R. & Denic, V. The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase. *Nature* **512**, 441–444 (2014).
46. McGilvray, P. T. et al. An ER translocon for multi-pass membrane protein biogenesis. *eLife* **9**, e56889 (2020).
47. Hedge, R. S. & Keenan, R. J. The mechanisms of integral membrane protein biogenesis. *Nat. Rev. Mol. Cell Biol.* **23**, 107–124 (2022).
48. Zhang, W., Campbell, H. A., King, S. C. & 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*. *J. Biol. Chem.* **280**, 26032–26038 (2005).
49. Volmer, R. & Ron, D. Lipid-dependent regulation of the unfolded protein response. *Curr. Opin. Cell Biol.* **33**, 67–73 (2015).
50. Jacquemyn, J., Cascalho, A. & Goodchild, R. E. The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. *EMBO Rep.* **18**, 1905–1921 (2017).

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 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 v.2.3 was used for analysis of the GLUT3 structure). 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 previously[9] 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^{-1}. Cells were harvested after 48 h. Lysis was done with hypotonic buffer (20 mM HEPES pH 8.0, 1 mM EDTA, 1 mM PMSF) and membrane fractions were collected through centrifugation. Solubilization was done with 20 mM HEPES pH 8.0, 150 mM NaCl, 1% *n*-dodecyl-β-D-maltoside (DDM, Anatrace), 0.5% CHS, 1 mM PMSF at 4 °C for 1 h. Solubilized GLUT3 was bound to Ni-NTA resin (Qiagen) at 4 °C for 1 h. Resin was washed with high salt buffer (20 mM HEPES pH 8.0, 1 M NaCl, 20 mM imidazole, 0.05% DDM, 0.0025% CHS) and low salt buffer (20 mM HEPES pH 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 homemade HRC3V protease at 4 °C. Uncut product and cleaved green fluorescent protein (GFP) tag were removed using homemade GFP nanobody column. GLUT3 was finally purified with size exclusion chromatography (GE Healthcare) equilibrated with 20 mM HEPES pH 8.0, 150 mM NaCl, 0.03% DDM, 0.015% CHS.

Expression and purification of the human EMC. Human EMC was prepared essentially as described previously[51], but using DDM instead of lauryl maltose neopentyl glycol. Wash buffers in all early stages included 0.2% DDM, whereas 0.02% DDM was used in all buffers following FLAG elution. Glycerol was added to a final concentration of 10% before snap freezing in liquid nitrogen. Freeze–thaw aliquots were checked by nanoDSF using a Prometheus NT.48 and had melting curves similar to samples that had not been freeze-thawed.

Instrumentation of magnetic tweezers. As established and described previously[52], 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 ~0.1 fN to 50 pN. Illumination with a super-luminescent diode (wavelength = 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 prerecorded by moving an objective lens using a piezoelectric nano-positioner (Mad City Labs) with respect to the sample to generate calibration tables for individual beads. By comparing diffraction patterns of magnetic beads with the corresponding calibration table in real time, the positions of the magnetic bead were tracked and Custom-written LabView programs were used.

Single-molecule magnetic tweezers experiments. Samples for single-molecule magnetic tweezers experiment were prepared as described previously[9]. In brief, GLUT3 proteins were mixed with the SpyCatcher-DNA handles (with DDM added to a final concentration of 0.1% and Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) added to 2 mM for WT GLUT3) and incubated for 20 h at 4 °C. Molar ratios of 10:1 to 20:1 for GLUT3 protein/SpyCatcher-DNA handles were used. After incubation, the protein–DNA 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 diluted to below half its CMC). The membrane proteins connected with two DNA handles were then stored 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 (Sigma-Aldrich) were prepared in a 2.8:1 molar ratio.

For single-molecule magnetic tweezers experiments, 4 µl of 0.01 mg ml^{-1} neutravidin was added to 40 µl of the sample and incubated for 5 min at room temperature. After binding neutravidin to one end of the DNA handle, the sample was further diluted to a final concentration of ~500 pM. We first injected 0.02% (w/v) streptavidin-coated polystyrene particles (3.11 µm, Spherotech; reference bead) into a homemade flow-cell consisting of two cover slips (VWR No 1.5). The bottom cover slip was coated with mPEG and biotin-PEG at a 100:3 molar ratio. After 5 min incubation, unbound reference beads were removed by 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 µm in 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 was additionally injected.

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 tension[53].

$$F=\left(\frac{k_BT}{l_{p,i}}\right)\left[\frac{1}{4\left(1-z_i/(n_il_{c,i})\right)}-\frac{1}{4}+\frac{z_i}{n_il_{c,i}}-\frac{F}{K_i}+\sum_{j=1}^{j\le7}a_j\left(\frac{z_i}{n_il_{c,i}}-\frac{F}{K_i}\right)^j\right]$$

where the index i indicates either DNA or unstructured polypeptide (p), k_BT 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)[54–56]. K_i is the elastic modulus ($K_p\approx50$ µN and $K_{DNA}\approx500$ pN)[52,57], 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 $n_{GLUT3,p}=463$ ($n_{GLUT3,p}=n_{N-domain,p}+n_{C-domain,p}+n_{ICH-domain,p}=198+207+58$)) for GLUT3 (ref. [12]).

To describe a rigid-like biopolymer such as helical states (U_h), the Kessler–Rabin model was used[56],

$$z_h\simeq-\frac{1}{2f}-\frac{\chi}{f\tanh2\chi}+\frac{L_h}{\tanh2\chi}-\frac{2\chi^2}{3f}\left(\frac{1}{\tanh fL_h}-\frac{fL_h}{(\sinh fL_h)^2}-1\right),$$

where $\chi\equiv\sqrt{f\frac{(n_hl_{c,h})^2}{4l_{p,h}}}$, $f\equiv\frac{F}{k_BT}$ and n_h is the number of amino acids consisting of the TMH. The persistence length ($l_{p,h}$) is 9.17 nm and the contour length ($l_{c,h}$) along helical axis is, on average, 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.

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

where z_m is measured extension, z_p is the extension of the unstructured polypeptide linker between DNA and target protein, z_{DNA} is the extension of the DNA handle, and $z_{p\text{ or }h}^{GLUT3}$ 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 $z_{p\text{ or }h}^{GLUT3}$ values are calculated from the eWLC or Kessler–Rabin model, respectively. In the case of stretching GLUT3 in its native state (N), $z_{p\text{ or }h}^{GLUT3}$ is replaced by a d_N value of 3.9 nm, an end-to-end distance determined from the native state structure (PDB 4ZW9)[12].

To analyze relative extension changes during high-force unfolding, we treated N- and C-domains independently because the 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 (Δn_i), giving $\frac{z_{i,p}}{z_{N(C)-domain,p}}=\frac{\Delta n_i}{n_{N(C)-domain,p}}$, where $n_{N(C)-domain,p}$ is the extension increase expected when the N- or C-domain is fully unraveled and $n_{N(C)-domain,p}$ is the total number of amino acids in the 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 the 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 $\Delta n_i\simeq\frac{n_{N(C)-domain,p}}{z_{N(C)-domain,p}}\left(z_{m,i}-d_i+d_{N(C)-domain}\right)$. 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. A fixed rate of 1 pN s^{-1} was applied for FEC obtained by force-ramp experiment.

HMM analysis. HMM analysis was employed to determine the folding/unfolding intermediate states from the time-resolved low force extension traces recorded at 1.2 kHz (ref. [24]). The adjustable parameters in our system are the number of states (n), the extension position for i-th intermediate state ($\overline{zi}$) and the transition matrix of rates between states ($\overleftrightarrow{k}$). The optimal number of states (n) was obtained from BIC: BIC = $q\ln(N)-2\ln(L\hat{})$, where q is the number of output parameters given by model, N is sample size and $\hat{L}$ 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 changes[23]. The extension traces were median-filtered with a 5-Hz window, and the extension position/deviation for each state was estimated from the Gaussian mixture model in the HMM analysis. The rates (that is, 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 <50 ms were considered an 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[58,59]. Because the magnetic bead in magnetic tweezers is not physically trapped, unlike with optical tweezers, the marginal probability distribution ($\tilde{P}_m(z_m;F)$) from the 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 F; β is $1/k_BT$. 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 toward the magnetic bead, dh2 toward peg) and two polypeptide linkers (ph) between DNA and GLUT3) and magnetic bead. In brief, $\tilde{P}_{bh}(z;F) \equiv \mathcal{F}^{-1}\left(\tilde{P}_b(k;F)\tilde{P}_{peg}(k;F)\tilde{P}_{dh1}(k;F)\tilde{P}_{dh2}(k;F)\tilde{P}^2_{ph}(k;F)\right)$, 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 is βF and i is the complex number. The remaining terms in $\tilde{P}_{bh}(z;F)$ can be described by $\tilde{P}_j(k;F) \equiv \frac{\sum_{n,l,l'} \Psi^T_{l',B.C.}(L_{c,j})\left[e^{-E_{n,j}(f-ik)L_{c,j}}\right]_{l',l}\Psi_{l,B.C.}(0)}{\sum_{n,l,l'} \Psi^T_{l',B.C.}(L_{c,j})\left[e^{-E_{n,j}(f)L_{c,j}}\right]_{l',l}\Psi_{l,B.C.}(0)}$, 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 eigenstate and eigenvalue (total energy), respectively, as previously defined and estimated from the effective Hamiltonian equation of propagator of biopolymer in Markovian regime[55]. The index *B.C.* in the eigenstate means boundary condition of biopolymer and indicates whether the semi-flexible biopolymer is half-constrained or unconstrained.

To avoid any numerical instability and ill-conditioned results, we used 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 a 5-Hz window. Because the characteristic time scale of magnetic bead is <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 ($(2\pi\sigma_i^{\lambda^2})^{-1/2}e^{-(z_\lambda-\mu_i^\lambda)^2/2\sigma_i^{\lambda^2}}$). w_i^λ is weighting factor in a linear combination and N_λ is total number of Gaussian components ($N_{bh} = 1$ was chosen). The parameters of the deconvoluted extension distribution of the single GLUT3 are then 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 $\sigma_i^{p^2} = \sigma_i^{m^2} - \sigma_i^{bh^2} - 2\mathrm{cov}(z_m, z_{bh})$.

Constructs for flow cytometry experiments. The parent mCherry-P2A-eGFP construct was generated by inserting mCherry (from pmCherry-C1 vector, Clontech) and enhanced green fluorescent protein (eGFP, from pEGFP-C1 vector, Clontech) fragments into pENTR221 vector using Gibson assembly (New England Biolabs). For mCherry-P2A-eGFP-protein constructs, GLUT3 N43T, GLUT3 WT (glycosylation site N43 is unmodified) and TRAM2 (Sino Biological) with GSGSGS linker at both ends were amplified by polymerase chain reaction and inserted to the parent mCherry-P2A-eGFP construct using Gibson assembly.

For the truncated GLUT3 WT constructs, coding regions of GLUT3 TMH 1 (residues 2–51), GLUT3 TMH 1–2 (residues 2–90), GLUT3 TMH 1–3 (residues 2–116), GLUT3 TMH 1–4 (residues 2–146), GLUT3 TMH 1–5 (residues 2–182) and GLUT3 TMH 1–6 (residues 2–208) were amplified via a polymerase chain reaction 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. mCherry-P2A-GLUT3 C-domain (residues 252–496) was generated by the same method. All inserts in the pENTR221 vector constructs were transferred to pInducer20 vector by Gateway recombination cloning (ThermoFisher Scientific). pENTR221 and pInducer20 vector plasmids are generous gifts from C. Kang (Seoul National University).

Flow cytometry analysis. Flow cytometry experiments were conducted as described previously[8,31] with modifications. In brief, HEK293T cells (ATCC) were plated in six-well plates and transiently transfected with appropriate plasmids (0.5 μg ml^{-1}) using polyethylenimine (PEI; Sigma) for 24 h. After transfection, cells were treated with 100 ng ml^{-1} doxycycline (Sigma) for 24 h. Following doxycycline treatment, cells were trypsinized, then washed in DPBS, and resuspended in 1 ml of ice-cold DPBS. Cells were then passed through a cell strainer (Falcon) and analyzed with the Sony SH800S cell sorter. Live cells were first gated with forward scatter/back scatter (FSC/BSC), and then gated for mCherry. A total of 20,000 cells were analyzed, and data analysis was done using Cell Sorter software (v.2.1.5) and customized MATLAB codes. The WT and EMC6 knockout Flp-In T-Rex 293 cell lines are generous gifts from J. Christianson (Oxford University). The expression of EMC5 and EMC6 were double-checked by western blot (EMC5 and EMC6 antibodies; Abcam), using β-tubulin (antibody; Cell Signaling Technology) as a loading control.

MD simulations. The simulation systems were prepared using CHARMM-GUI Membrane Builder[35] with a crystal structure of GLUT3 (PDB 4ZWC)[15]. The structures for the N-terminal (residues 3–205) and C-terminal (residues 264–470) 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, we prepared additional mixed bilayers of DMPC, DMPG or DMPE at $T = 306.15$ K to examine temperature effects. In addition, to explore the effect of tail saturation, we also prepared mixed bilayers composed of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphorylglycerol (POPG) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) at $T = 296.15$ K. The molar ratios of POPC:POPG and POPC:POPG:POPE 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 the protein structure, the initial *xy* dimensions of bilayers were set to be ~130 × 130 Å^2 (N- and C-domains) and ~150 × 150 Å^2 (full-length GLUT3), respectively. Each system was subjected to a 0.5- or 1.0-μs production run following a series of short equilibration runs. All simulations were carried out using OpenMM with the CHARMM36 force fields and transferable intermolecular potential with 3 points (TIP3P) water model. Trajectories were analyzed using CHARMM and in-house PYTHON scripts. The snapshots from simulations trajectories were prepared using VMD.

Determination of helix insertion energy. We compared the helix insertion energies into the ER membrane across kingdoms using DGpred[23] for a 19-residue window in the center of the helix. Because 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 database[60]. 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), *E. 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 in Data Availability section.

Statistics and reproducibility. The gel in Extended Data Figs. 1b,c, 3c and 5b,e are representatives for the purification protocols that have been performed at least ten independent times with similar results. The BODIPY gel assay in Extended Data Fig. 3c was performed two independent times with similar results. The gel in Supplementary Fig. 7a is a representative western blot. Each transmission electron microscopy (TEM) image in Extended Data Fig. 1f,g is a representative micrograph among the reproducible features obtained from >20 spots on TEM grids. For repeatability, at least three TEM grids were prepared for each bicelle-forming condition.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

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/tyyoonlab/ Nat_Chem_biol_NCHEMB - A210813554). The following PDB IDs were used (4ZWB, 4ZWC, 4ZW9, 4GBZ, 6RW3, 4LDS and 6H7D). The following UniProt IDs were used (P11169, P0AGF4, O97467, A0A0H2VG78 and Q9LT15). Source data are provided with this paper.

Code availability

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

References

51. O'Donnell, J. P. et al. The architecture of EMC reveals a path for membrane protein insertion. *eLife* **9**, e57887 (2020).

52. Shon, M. J., Rah, S.-H. & Yoon, T.-Y. Submicrometer elasticity of double-stranded DNA revealed by precision force-extension measurements with magnetic tweezers. *Sci. Adv.* **5**, eaav1697 (2019).
53. Bouchiat, C. et al. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys. J.* **76**, 409–413 (1999).
54. Oesterhelt, F. et al. Unfolding pathways of individual bacteriorhodopsins. *Science* **288**, 143–146 (2000).
55. Seol, Y., Li, J., Nelson, P. C., Perkins, T. T. & Betterton, M. Elasticity of short DNA molecules: theory and experiment for contour lengths of 0.6–7 μm. *Biophys. J.* **93**, 4360–4373 (2007).
56. Sarkar, A., Caamano, S. & Fernandez, J. M. The mechanical fingerprint of a parallel polyprotein dimer. *Biophys. J.* **92**, L36–L38 (2007).
57. Gebhardt, J. C. M., Bornschlögl, T. & Rief, M. Full distance-resolved folding energy landscape of one single protein molecule. *Proc. Natl Acad. Sci. USA* **107**, 2013–2018 (2010).
58. Hinczewski, M., von Hansen, Y. & Netz, R. R. Deconvolution of dynamic mechanical networks. *Proc. Natl Acad. Sci. USA* **107**, 21493–21498 (2010).
59. Hinczewski, M., Gebhardt, J. C. M., Rief, M. & Thirumalai, D. From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proc. Natl Acad. Sci. USA* **110**, 4500–4505 (2013).
60. Saier, M. H. Jr, Tran, C. V. & Barabote, R. D. TCDB: the transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Res.* **34**, D181–D186 (2006).

Acknowledgements

We thank E. Kweon for help 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 Institutes of Health grant (R01GM118684 to H.H.).

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 the 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 the flow cytometry experiment. S.P. and W.I. performed the molecular dynamics simulation. H.G.K., C.T. and M.S. performed the bioinformatic analysis. H.L. and S.-H.R. performed the 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.

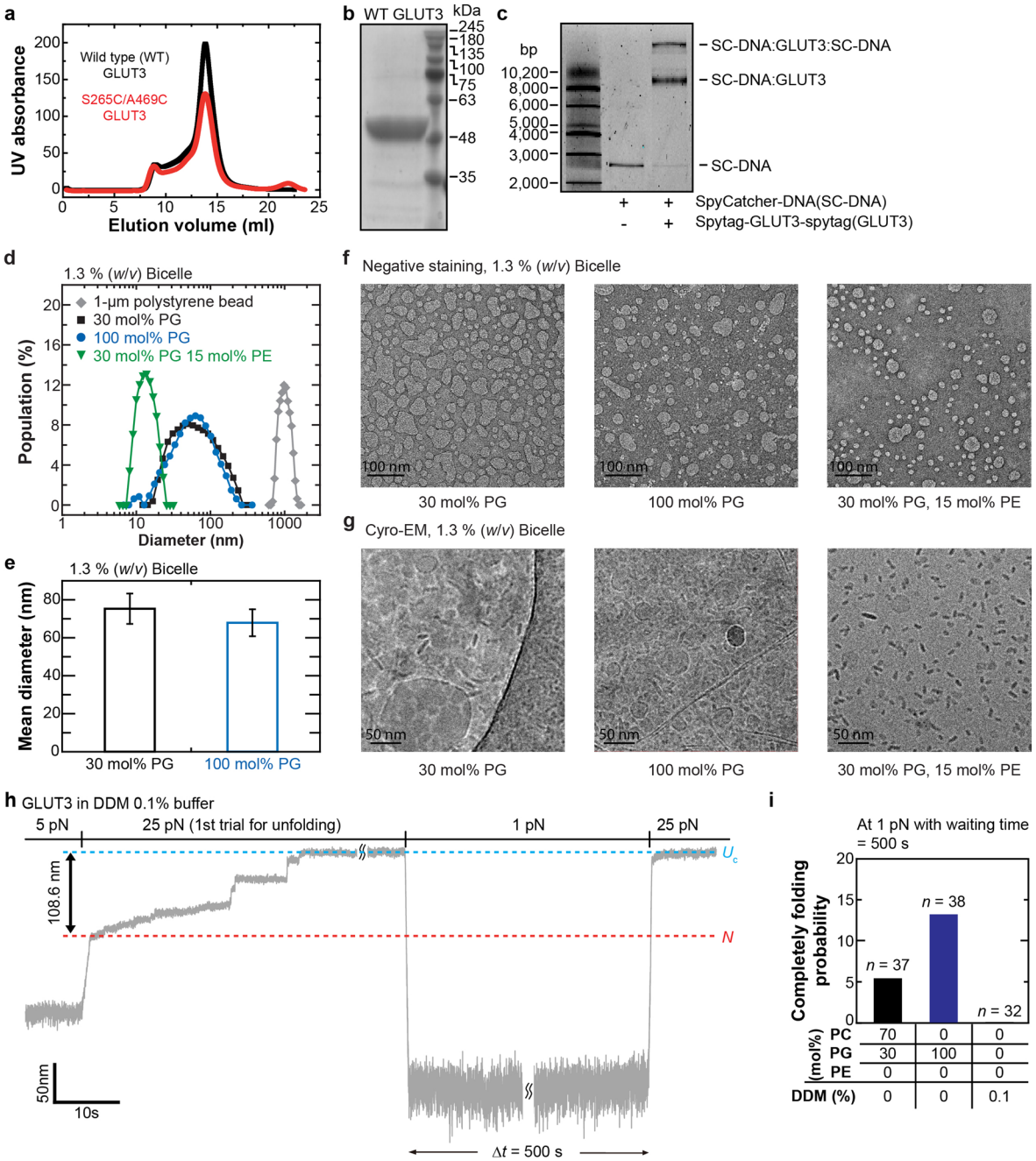
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41589-022-01002-w.

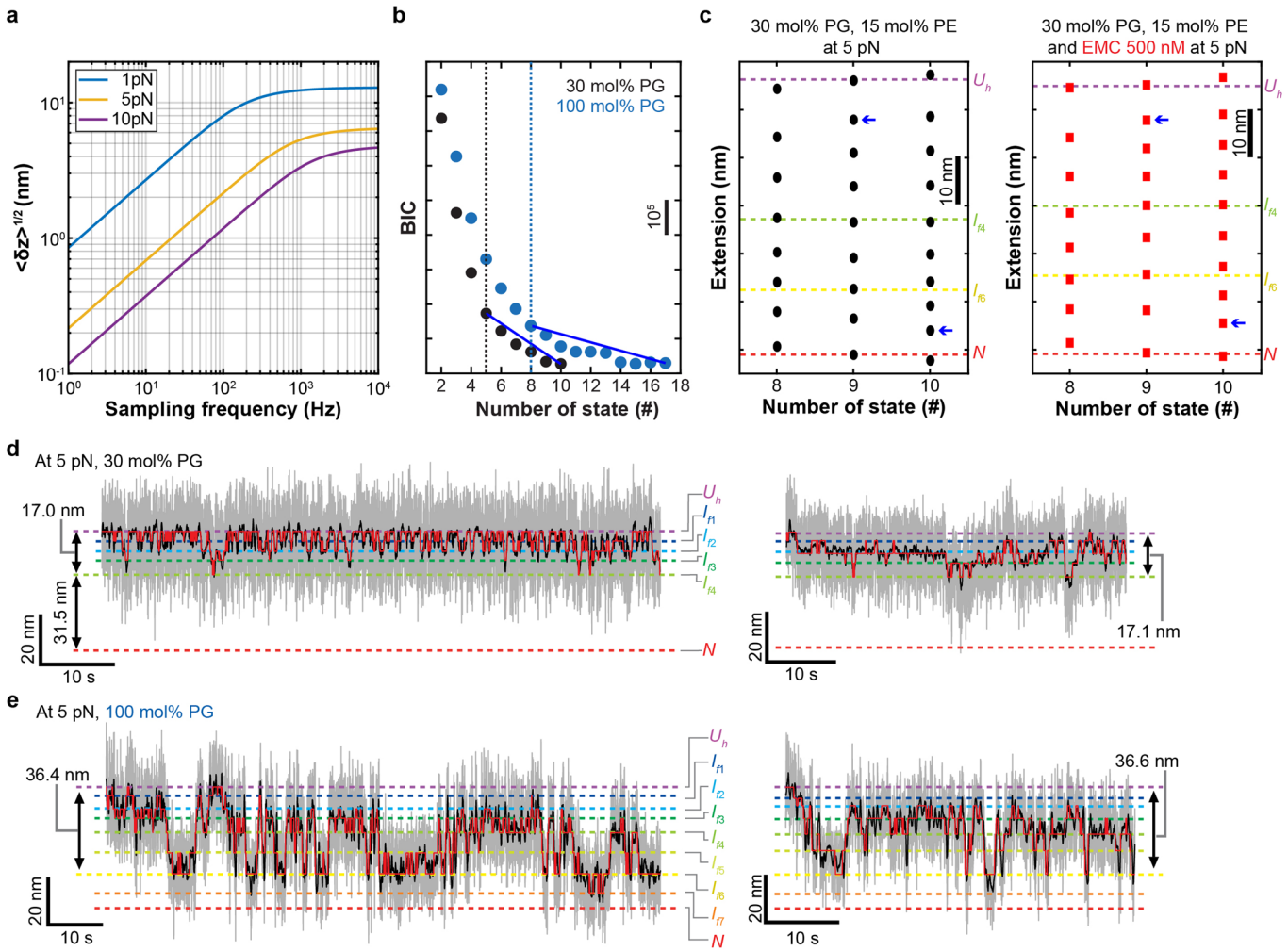
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-022-01002-w.

Correspondence and requests for materials should be addressed to Elizabeth A. Miller, Hee-Jung Choi or Tae-Young Yoon.

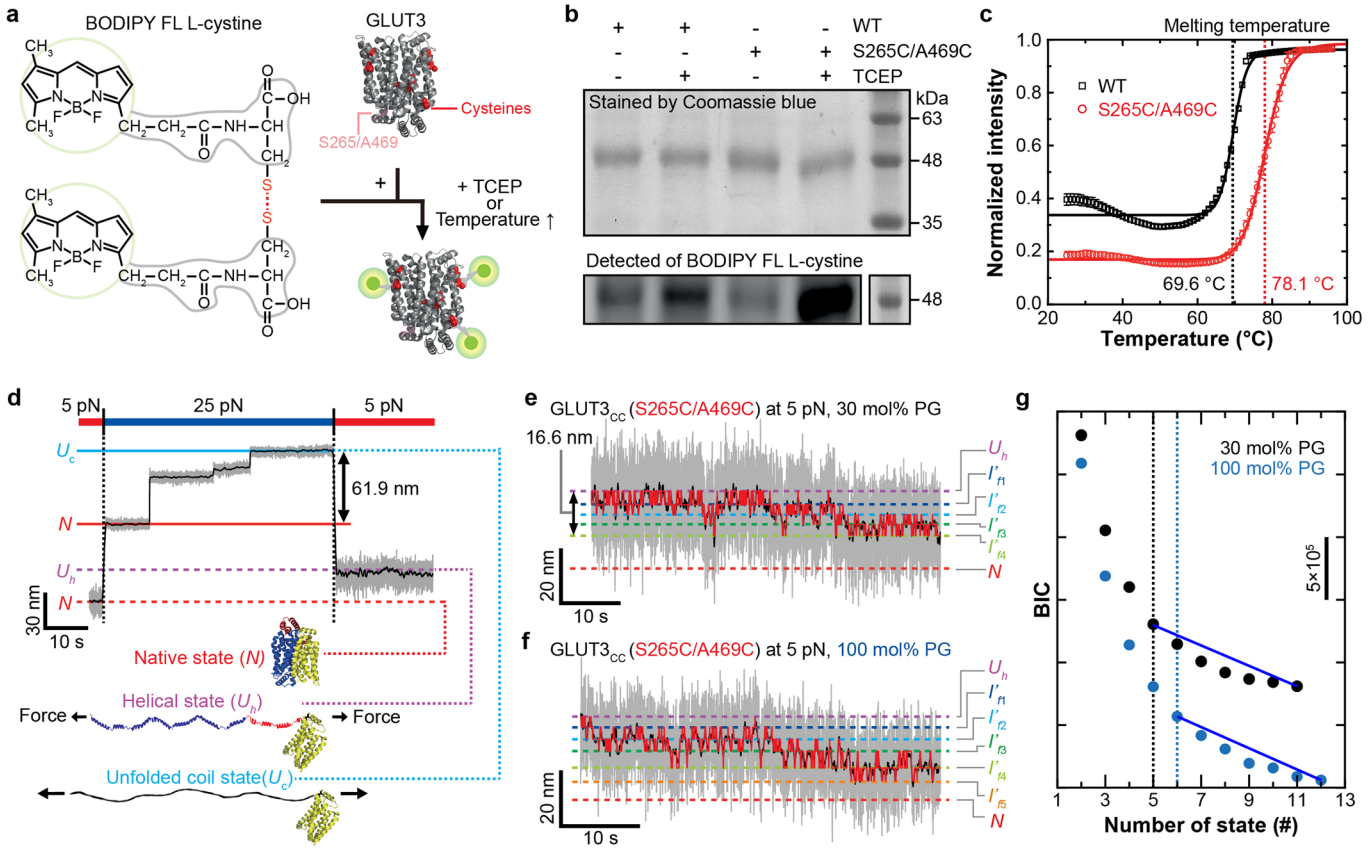
Peer review information *Nature Chemical Biology* thanks Argyris Politis, Horst Vogel and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



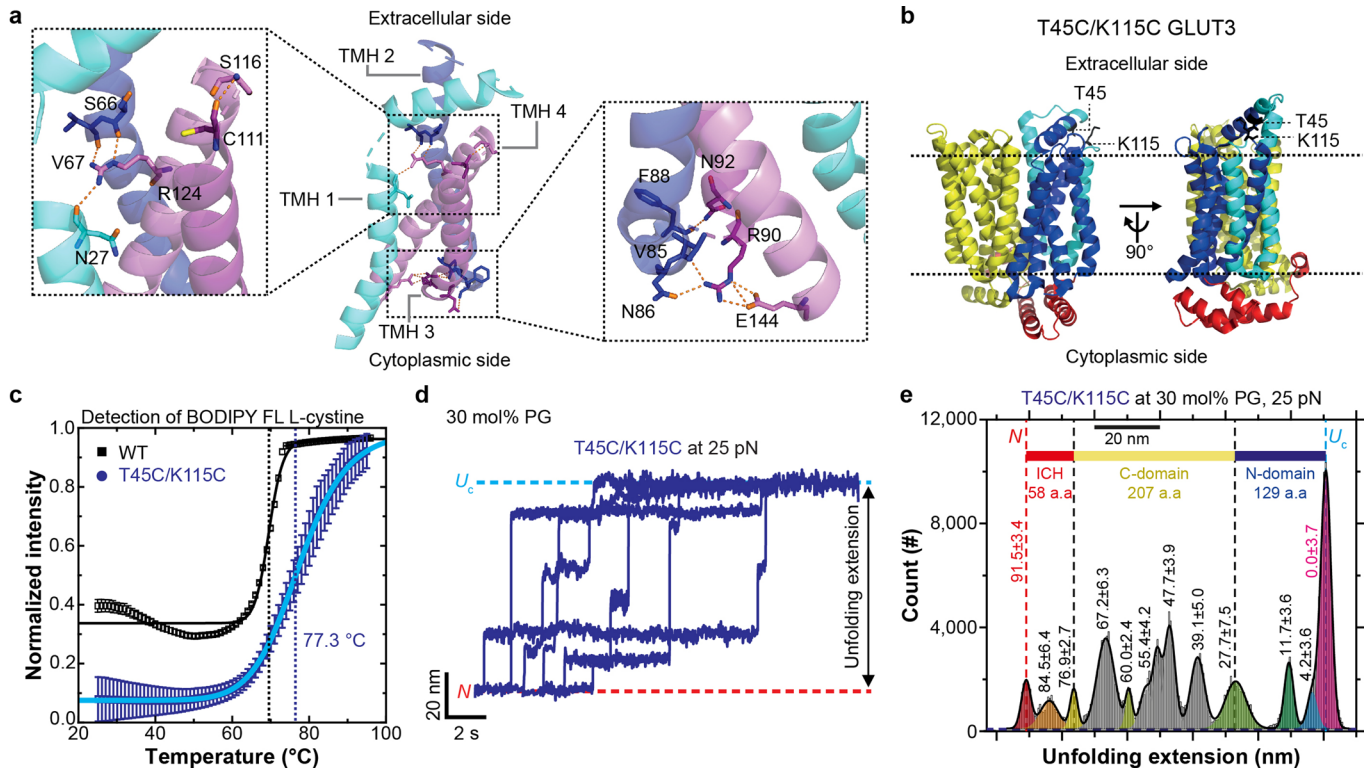
Extended Data Fig. 1 | Sample preparation of WT and C-domain-knotted GLUT3 proteins and bicelle membranes. a, Elution profile obtained by size exclusion chromatography (SEC) of WT GLUT3 (black) and S265C/A469C GLUT3 (red). **b**, Purified WT GLUT3 protein analyzed by SDS-PAGE. The major peak position 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 25pN 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 25pN. **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.



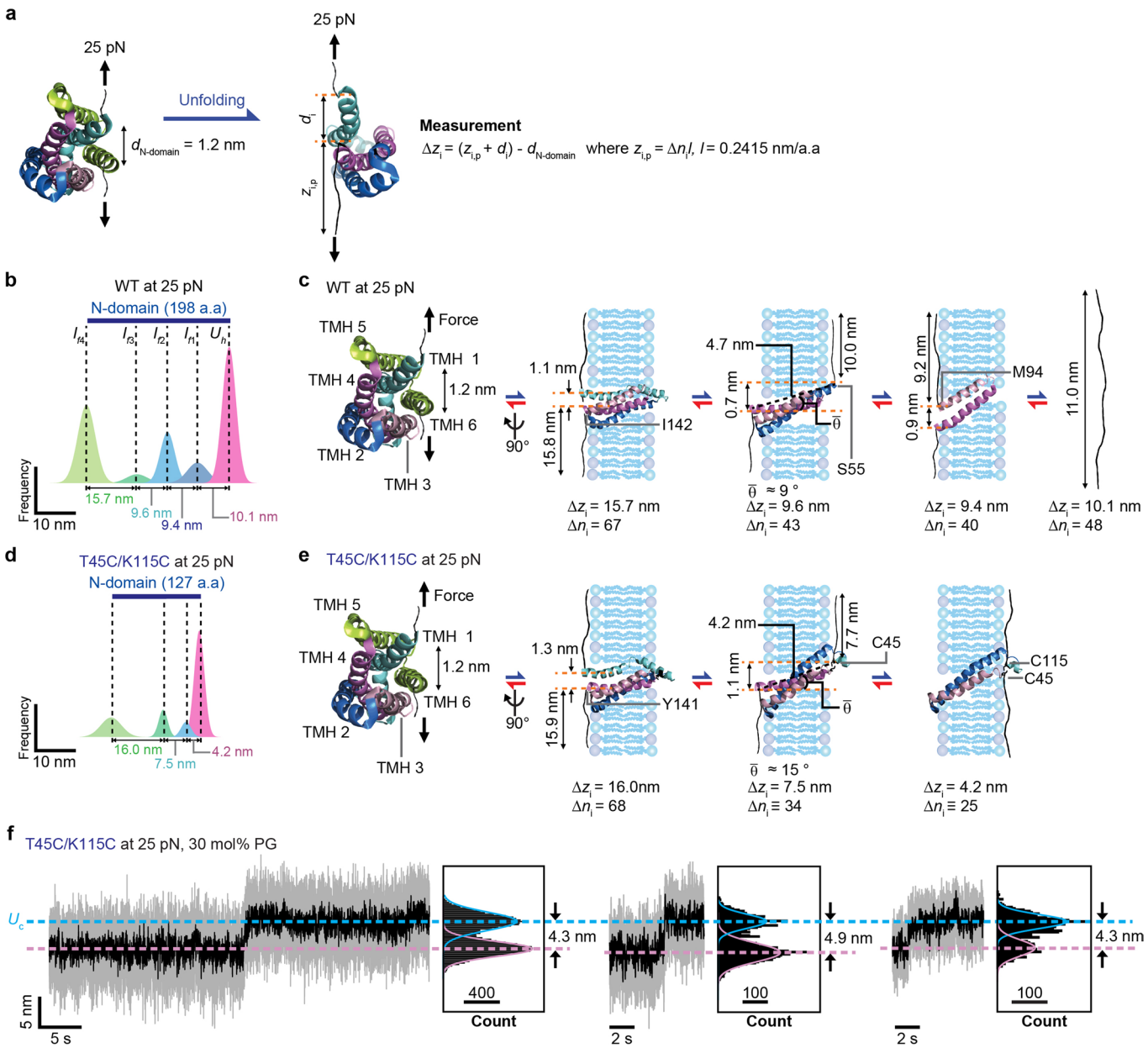
Extended Data Fig. 2 | Analysis of single-molecule magnetic tweezer data. a, Precision in determination of the vertical position of a bead as a function of the measurement bandwidth. The plot indicates an ~1 nm resolution when bead positions are averaged over 50 ms (~20 Hz sampling). In our magnetic 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 5pN. 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.



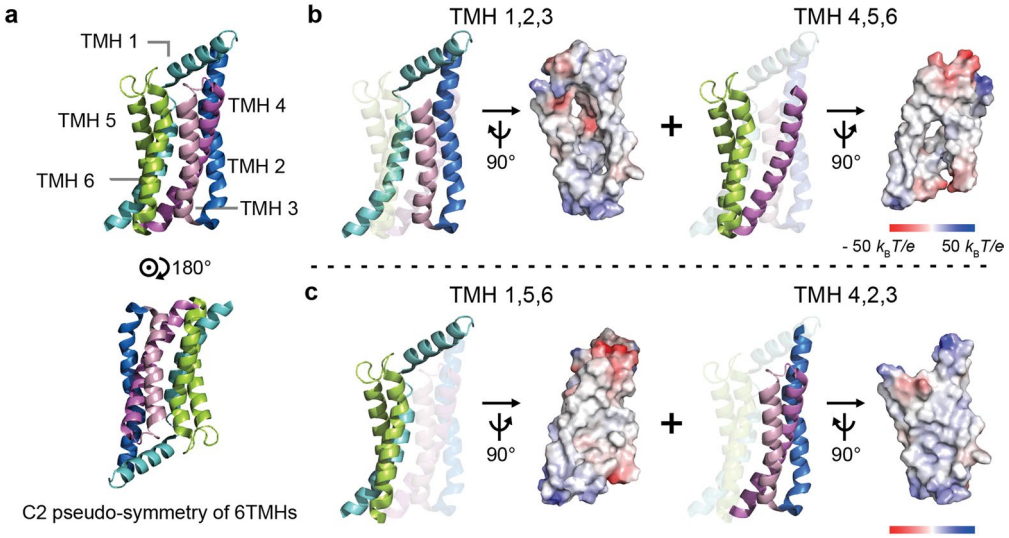
Extended Data Fig. 3 | Sample preparation and folding behavior of S265C/A469C GLUT3. a, Schematic of the assay using BODIPY-L-cystine. The left panel is the chemical structure of two BODIPY FL fluorophores attached to the amino groups of the disulfide-containing amino acid, cysteine. The right panel 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).



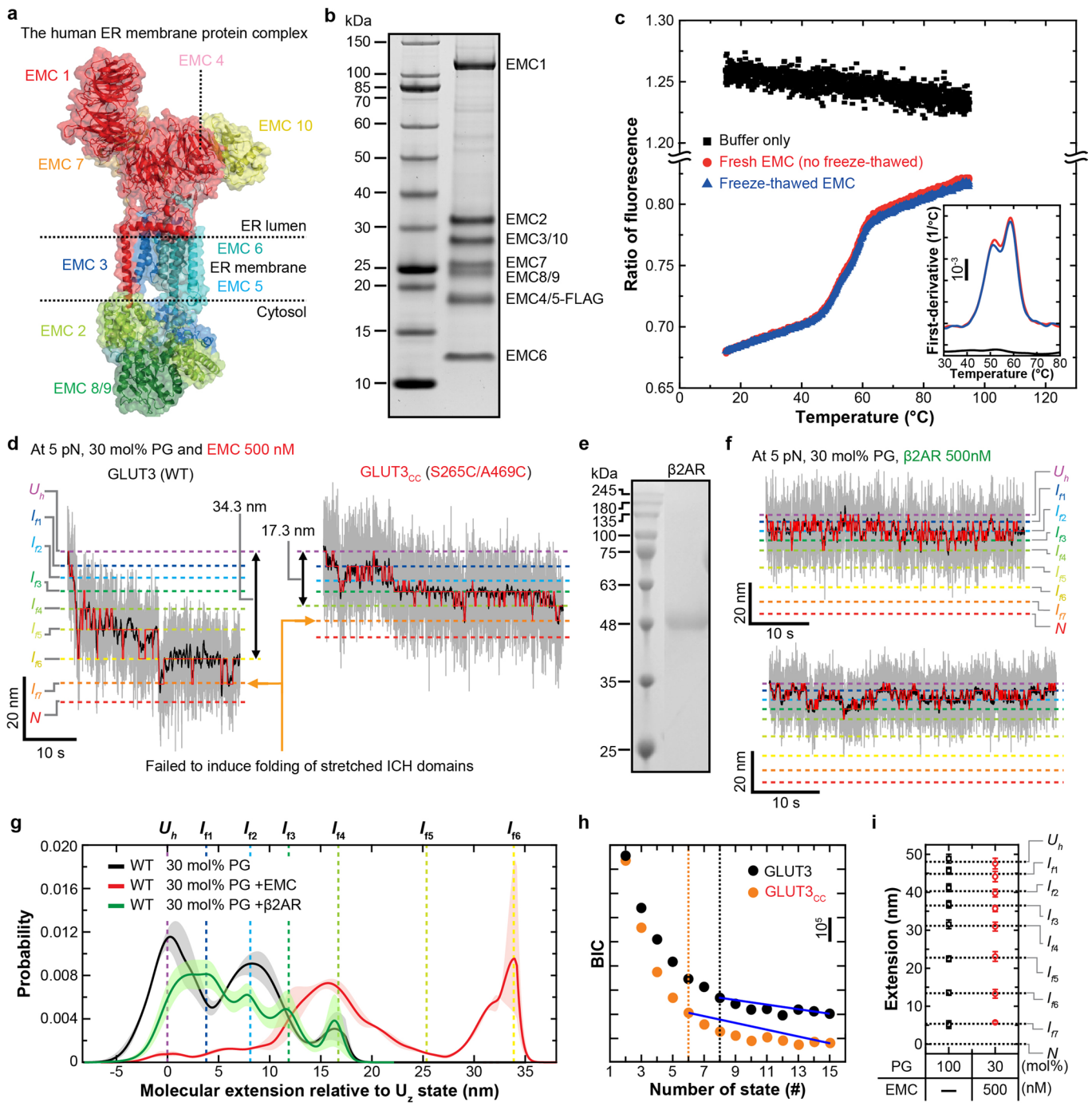
Extended Data Fig. 4 | Sample preparation and unfolding characteristics of T45C/K115C GLUT3. a, Atomic contacts among TMHs 1, 2, and 4. Inset shows detailed position of interacting residues (blue for amino group, orange for carboxyl group, and yellow for thiol group). **b**, The positions of two mutations, 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 ± SD ($n = 19$).



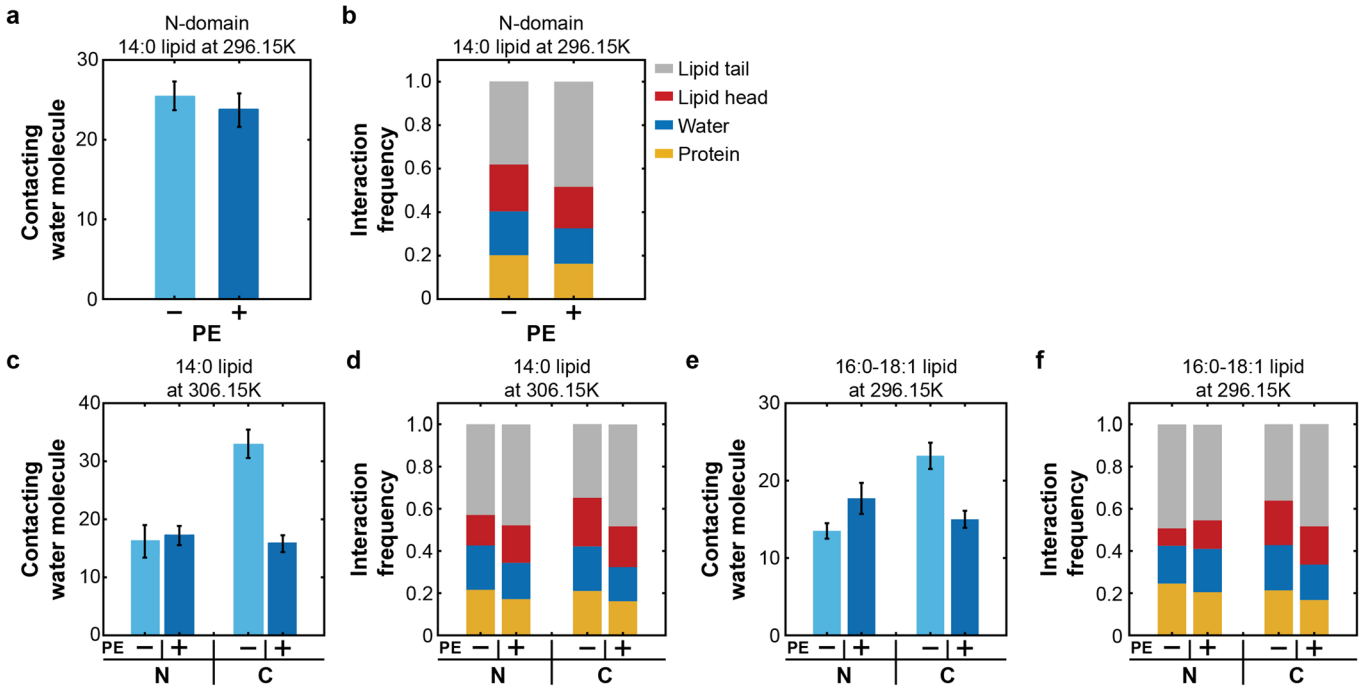
Extended Data Fig. 5 | Determination of folding order for N-domain of GLUT3. a, Schematic of pulling geometry for N-domain of GLUT3 at 25 pN. $d_{N\text{-domain}}$ is the distance between two points of force application before unfolding (PDB: 4ZWC). Δz_i indicates the expected extension increase for GLUT3 for the ith intermediate. $z_{i,p}$ is the extension of the unfolded portion along the membrane for the ith intermediate. d_i denotes the distance between the points of force application for the ith intermediate. Δn_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 25pN. **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 θ 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.



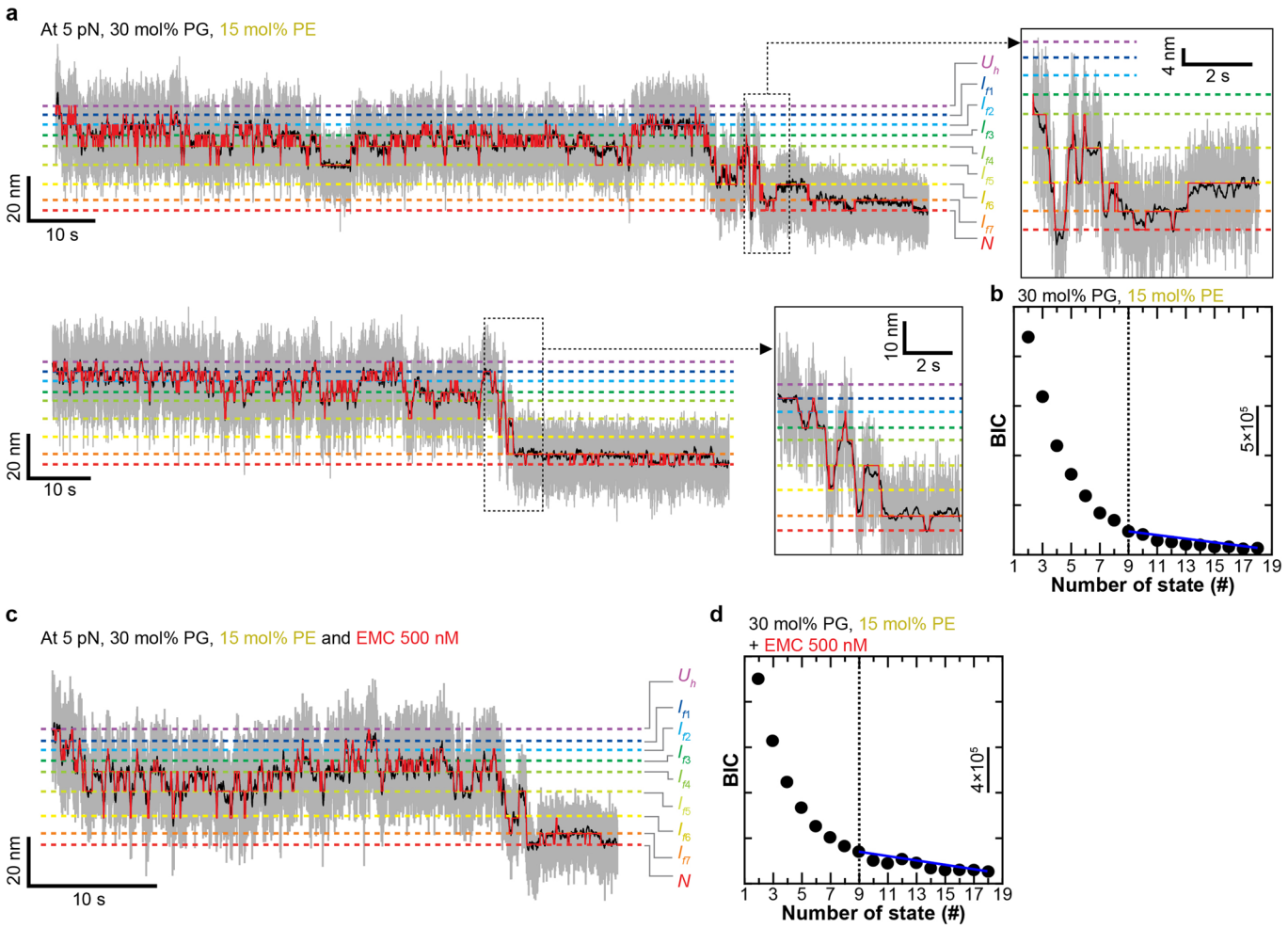
Extended Data Fig. 6 | Symmetrical structure of GLUT3. a, Structural view of GLUT3's N-domain with its C2 pseudo-symmetry. **b**, Structural view and electrostatic potential of the helix triplets composed of TMHs 1, 2, 3 and 4, 5, 6 each. **c**, Structural view and electrostatic potential of the helix triplets composed of TMHs 1, 5, 6 and 4, 2, 3 each.



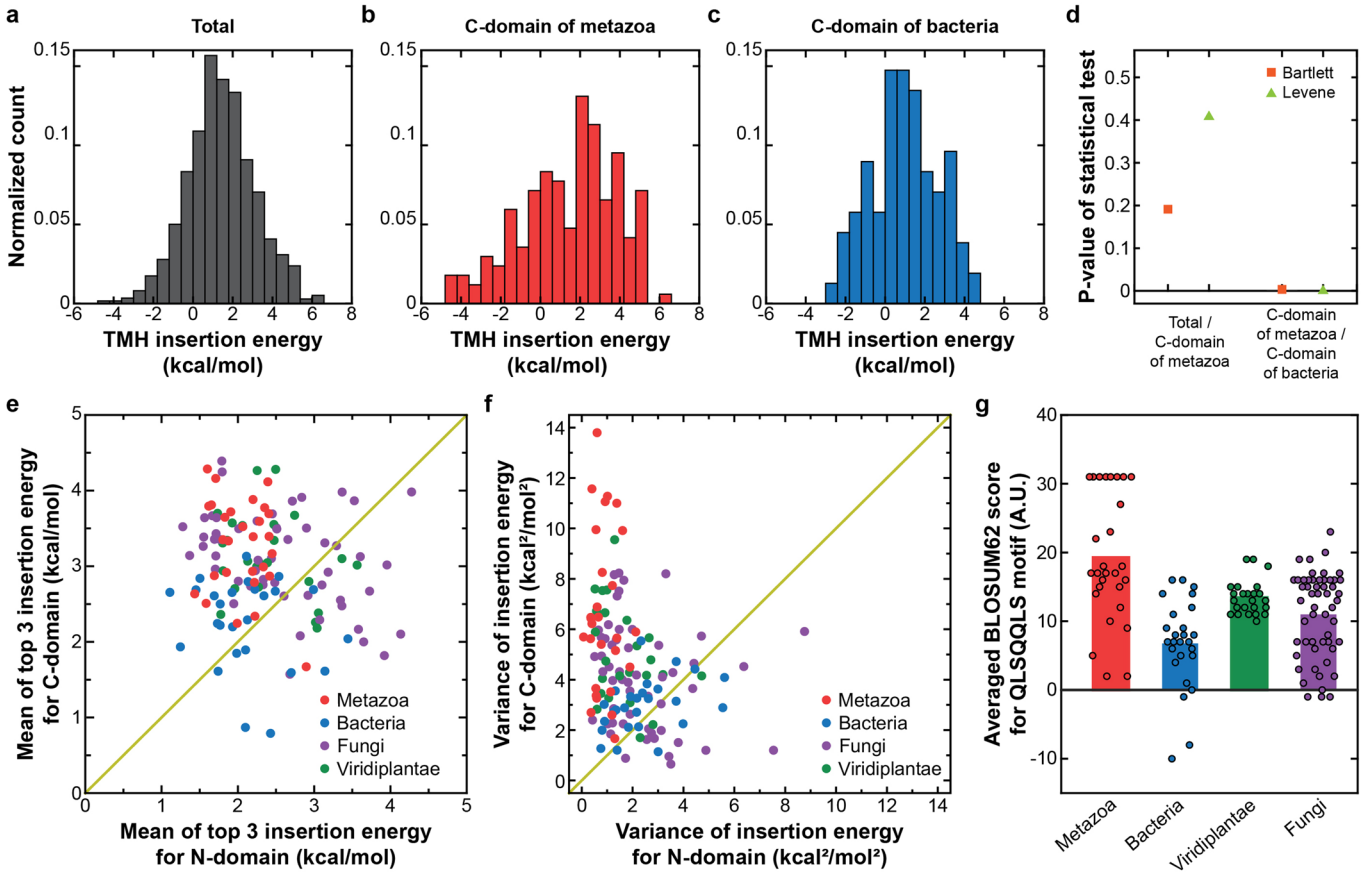
Extended Data Fig. 7 | EMC preparation and folding characteristics of GLUT3 with β2AR. a, Structural model of the human ER membrane protein complex (EMC). **b**, Purified EMC analyzed by SDS-PAGE. The left lane is molecular weight standards and the right lane shows purified EMC. **c**, Fluorescence profile 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 β2AR. **f**, Representative folding traces for single GLUT3 at 5 pN with 30 mol% PG in the bicelles in the presence of 500 nM β2AR. Two replicates are shown. **g**, Probability distributions of deconvoluted extension values observed under indicated folding conditions at 5pN ($n = 11$ for the reaction with β2AR). 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. 8 | Analysis of MD simulation for GLUT3 in lipid bilayer. a, 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). Polar/charged residues in TMHs of GLUT3 for the analysis are as follows. S21, Q23, N27, T28, S71, S78, S80, N98, R124, T135, T156, N158, Q159, T191, Q198, S273, Q277, S279, Q280, Q281, S283, N286, N315, T316, T319, S322, S346, E378, W386, N409, W410, N413. Residues near the GLUT3 pore entries are not chosen which are likely to be exposed to bulk water. **b**, Interaction frequency of polar/charged residues in N-domain interface with or without DMPE. The value in (**a**,**b**) is the average value from 0.6μs to 1.0μs. **c**, 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). **d**, Interaction frequency of polar/charged residues in domain interfaces. The value in (**c**,**d**) is the average value from 0.6μs to 1.0μs. **e**, The average number of contacting water molecule to polar/charged residues in TMHs of GLUT3. Error bars represent SD ($n = 2000$ for each case). **f**, Interaction frequency of polar/charged residues in domain interfaces. The value in (**e**,**f**) is the average value from 0.3μs to 0.5μs.



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. 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).

nature research

Corresponding author(s): Elizabeth A. Miller, Hee-Jung Choi, and Tae-Young Yoon

Last updated by author(s): Feb 9, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Magnetic tweezers : LabView 2015 software (https://github.com/tyyoonlab/Science_aaw8208) FACS : Cell Sorter Software (v2.1.5) Molecular dynamic simulation : OpenMM, VMD Calculation of hydrophobicity scale : DGpred Phylogenetic analysis : Custom pipeline with HHblits and HHalign
Data analysis	Magnetic tweezers : MATLAB FACS : Cell Sorter Software (v2.1.5) Codes for magnetic tweezers and FACS (https://github.com/tyyoonlab/Nat_Chem_biol_NCHEMB-A210813554) Structural analysis : PyMol 2.3 Molecular dynamic simulation : CHARMM, CHARMM-GUI Phylogenetic analysis : Dendroscope Calculation of hydrophobicity scale : DGpred, MPEx 3.3 (https://github.com/schnamo/TMH_insertion_energy) Dynamic light scattering : Otsuka electronics Photal

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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/tyyoonlab/ Nat_Chem_biol_NCHEMB - A210813554).
The followings are the accession codes used in this study.
PDB ID : 4ZWB, 4ZWC, 4ZW9, 4GBZ, 6RW3, 4LDS and 6H7D
UniProt ID : P11169, P0AGF4, O97467, A0A0H2VG78 and Q9LT15

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes for magnetic tweezers experiment was determined by the number of a single proteins whose the extension value under force is satisfied with extensible worm-like chain model described in methods section. Experiments were repeated to ensure reproducibility for at least 3 independent days. The number of analyzed traces for each condition are indicated in figure legends. For flow cytometry experiments, 20,000 cells were analyzed for each sample, after the cells were gated for viability and fluorescent protein expression. This condition is consistent with the previous reports (Chitwood et al. (2018), Chitwood & Hegde (2020)), which worked on similar experiments.
Data exclusions	There is no data exclusion except FACS data. Points with negative intensity were generated by GFP compensation process for the case of cell whose GFP expression level is very low. So, only the points with positive intensity were included in the FACS scatter plot.
Replication	For magnetic tweezers, each time-resolved trace represents an independent measurement of folding/unfolding. Number of time-resolved traces for each plotted data is provided in the figure legends. Flow cytometry experiment was performed 3 independent times with similar results.
Randomization	Randomization is not relevant to this study since there were no experimental groups.
Blinding	Blinding is not relevant to this study since there were no experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-digoxigenin from sheep from Roche (RRID:AB_514496) Anti-MMGT1(Anti-EMC5) from abcam (RRID:AB_11129372)

	Anti-transmembrane protein 93(Anti-EMC6) from abcam (RRID:AB_1925516) Anti-tubulin from Cell Signaling Technology (#6181S)
Validation	Antibodies were validated by the manufacturers according to standard protocols (Western blot, Immunohistochemistry).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Spodoptera frugiperda (Sf9), commercial source : Expression system (RRID:CVCL_0549) HEK293T, commercial source : ATCC (RRID:CVCL_0063) HEK293 TRex EMC5-FLAG, commercial source : Invitrogen (reference : Guna et al., 2018, identifiers MBP04) Flp-In T-Rex 293, commercial source : Invitrogen (reference : Guna et al., 2018, RRID:CVCL_U427)
Authentication	Sf9 and HEK293T cell lines were not authenticated during the study. In the case of HEK293 TRex EMC5-FLAG and Flp-In T-Rex 293, cell line was checked by a presence of Frt site in the chromosome and doxycycline response, which is a characteristic of Flp-In T-Rex cells.
Mycoplasma contamination	Sf9 and HEK293T cell lines were not tested for Mycoplasma contamination.HEK293 TRex EMC5-FLAG and Flp-In T-Rex 293 cell line was checked approximately monthly for mycoplasma contamination and found to be negative.
Commonly misidentified lines (See ICLAC register)	None were used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	HEK293T and Flp-In T-Rex 293 cells were transfected and then induced with doxycycline. After the induction is done, cells were washed and then trypsinized. After the trypsin is inhibited, the cells were then washed with DPBS. Finally, cells were resuspended in ice-cold DPBS and then passed through a cell strainer.
Instrument	Sony SH800S cell sorter
Software	Cell Sorter Software (v2.1.5)
Cell population abundance	Cells were not stained or sorted and all cells were analyzed. Live cells were checked by BSC-A/FSC-A values.
Gating strategy	Cells were gated for live cells with BSC-A/FSC-A values. Additional gating for event rate was performed for the constant results.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.