



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

이학박사 학위논문

**Study on folding behavior and protein
evolution of glucose transporter using
single-molecule magnetic tweezers**

**(단분자 자기 집게 기술을 이용한 당
수송 막단백질의 접힘 및 진화과정
연구)**

2023년 2월

서울대학교 대학원

생명과학부

이 찬 우

Study on folding behavior and protein evolution of glucose transporter using single-molecule magnetic tweezers

지도 교수 윤 태 영
이 논문을 이학박사 학위논문으로 제출함
2022년 12월

서울대학교 대학원
생명과학부
이 찬 우

이찬우의 이학박사 학위논문을 인준함
2022년 12월

위 원 장 _____ 노성훈 _____ (인)

부위원장 _____ 윤태영 _____ (인)

위 원 _____ 이지오 _____ (인)

위 원 _____ 최명환 _____ (인)

위 원 _____ 최희정 _____ (인)

Abstract

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

Keyword : single-molecule force spectroscopy, magnetic tweezers, membrane protein folding, glucose transporter, protein evolution

Student Number : 2017-21674

Table of Contents

Abstract	i
Chapter 1. Introduction	1
1.1. Folding of helical membrane proteins.....	1
1.2. Biogenesis of membrane proteins	3
1.3. Major facilitator superfamily (MFS)	4
1.4. Single-molecule magnetic tweezers	5
Chapter 2. Methodology.....	7
2.1. Sample Preparation	7
2.1.1. Expression and purification of the human GLUT3	7
2.1.2. Expression and purification of the human ER membrane protein complex....	10
2.1.3. Preparation of DNA handles	12
2.1.4. Preparation of bicelle	13
2.2. Experimental Methods	14
2.2.1. Dynamic light scattering (DLS) measurement	14
2.2.2. single-molecule magnetic tweezers experiments.....	14
2.2.3. Molecular dynamics simulations.....	16
2.2.4. Sequence alignment and determination of helix insertion energy	18
2.3. Data Analysis.....	20
2.3.1. Force-extension curves (FEC) analysis	20
2.3.2. Hidden Markov Model (HMM) analysis.....	22
2.3.3. Deconvoluted extension probability analysis	22
Chapter 3. Results.....	25
3.1. Single-molecule magnetic tweezers monitoring GLUT3 folding	25
3.2. Mapping the folding order of single GLUT3 domains.....	29
3.3. Dissecting folding steps of the MFS folds.....	32
3.4. EMC facilitates insertion of TMHs of GLUT3.....	36
3.5. PE lipids boost domain-domain assembly of GLUT3.....	40
3.6. Asymmetric TMH distributions of metazoan sugar transporters	44
Chapter 4. Conclusion.....	48

Supplementary figure	5 1
Bibliography	6 3
Abstract in Korean (초록)	7 0
감사의 글	7 1

List of figures

Chapter 1. Introduction

Figure 1.1 Co-translational model of membrane protein folding	1
Figure 1.2 Summary of the various interactions that stabilize membrane proteins in lipid bilayers..	2
Figure 1.3 Model for the role of the chaperone in membrane protein biogenesis.....	3
Figure 1.4 The structure of human glucose transporter 3	4
Figure 1.5 Force application as change in the position of the magnet	5

Chapter 2. Methodology

Figure 2.1 GLUT3 construct and Ni-NTA affinity chromatography	7
Figure 2.2 Further purification process of GLUT3.....	8
Figure 2.3 Schematic of the assay using BODIPY-L-cystine	9
Figure 2.4 Analysis of WT and S265C/A469C GLUT3 using BODIPY-L-cystine.....	9
Figure 2.5 Preparation of ER membrane protein complex (EMC)	10
Figure 2.6 Schematic diagram of DNA-SpyCatcher conjugation process.....	12
Figure 2.7 Size distribution of bicelle and representative EM image of bicelle	13
Figure 2.8 Experimental setup of magnetic tweezers instrument	14
Figure 2.9 Representative gel image of SDS-PAGE after SYBR green staining.....	15
Figure 2.10 Snapshot of the protein extent (PDB:2HAC) along the z-axis	16
Figure 2.11 Basic workflow of HHblits	19

Chapter 3. Results

Figure 3.1 Schematic of magnetic tweezers (MT) experiment for observing folding of a single GLUT3 protein	25
Figure 3.2 Representative FEC of a single GLUT3 protein and insertion energy values of TMHs in 3 membrane proteins.....	26
Figure 3.3 Designed mechanical cycle for inducing refolding of a single GLUT3 protein	27
Figure 3.4 Representative time-resolved traces for GLUT3 folding at 5 pN with 30 mol% and 100 mol% DMPG in bicelles	27
Figure 3.5 Representative FEC of a single S265C/A469C GLUT3 protein	29

Figure 3.6 Folding characteristics of GLUT3 _{CC}	2 9
Figure 3.7 The positions and transition kinetics of folding intermediates	3 0
Figure 3.8 Unfolding step size of the intermediates I_{E4} and I_{E6}	3 1
Figure 3.9 Representative traces of high-force unfolding of single GLUT3 proteins.....	3 2
Figure 3.10 Distributions of extension values recorded during high-force unfolding	3 2
Figure 3.11 Representative traces from the force-jump experiments applied to individual low-force folding intermediates.....	3 3
Figure 3.12 Schematics of folding and unfolding of the C- and N-domain of GLUT3	3 4
Figure 3.13 Folding and unfolding behavior of the N-domain of GLUT3 _{TM23C}	3 5
Figure 3.14 Schematic of an MT experiment that examines the effects of EMC on folding of GLUT3.	3 6
Figure 3.15 Folding behavior of WT GLUT3 with EMC.....	3 7
Figure 3.16 Folding characteristics of WT GLUT3 and GLUT3 _{CC} in the presence of EMC.....	3 8
Figure 3.17 Probability distributions of deconvoluted extension values observed under indicated folding conditions at 5 pN.....	3 9
Figure 3.18 Cartoon of a single GLUT3 protein at I_{E6} before domain-domain assembly	4 0
Figure 3.19 Analysis of MD simulation for C-domain of GLUT3 in lipid bilayer	4 0
Figure 3.20 Complete folding of GLUT3 with the aid of PE lipids	4 1
Figure 3.21 Representative folding traces of GLUT3 both with EMC and PE lipids.....	4 2
Figure 3.22 Probability for observing the complete folding events under indicated conditions....	4 3
Figure 3.23 Phylogenetic tree of the MFS sugar porter family.....	4 4
Figure 3.24 Representative plots for predicted TMH-insertion energy	4 5
Figure 3.25 Scatter plots showing the calculated TMH-insertion energy	4 5
Figure 3.26 Sequence alignment of TMH 7 for a subset of metazoan sugar transporters.....	4 6
Figure 3.27 Analysis for insertion energy of TMH 7.....	4 7

Supplementary information

Supplementary figure 1 Precision in determination of the vertical position of a bead as a function of the measurement bandwidth.....	5 1
Supplementary figure 2 Folding characteristic of GLUT3 with 30 mol% PG and 100 mol% PG in bicelle.....	5 2
Supplementary figure 3 Topological model of GLUT3.....	5 3
Supplementary figure 4 Pulling geometry of a single GLUT3.....	5 4
Supplementary figure 5 Sample preparation and unfolding characteristics of T45C/K115C GLUT3.	

.....	5 5
Supplementary figure 6 Determination of folding order for N-domain of GLUT3	5 6
Supplementary figure 7 folding characteristics of GLUT3 with unrelated membrane protein.....	5 7
Supplementary figure 8 Deconvoluted extension probability analysis	5 8
Supplementary figure 9 Analysis of MD simulation for GLUT3 in various lipid bilayers.....	5 9
Supplementary figure 10 Folding characteristics with PE lipid bicelles.....	6 0
Supplementary figure 11 Force-jump experiment for determination of complete folding probability	6 1
Supplementary figure 12 Analysis for insertion energy of sugar transporters.	6 2

Chapter 1. Introduction

1.1. Folding of helical membrane proteins

Cell membranes exist to distinguish between inside and outside of the cell, and in the case of eukaryote, membranes also exist inside the cell for functional distinction (1). There are various proteins in these biological spaces, which can be divided into membrane proteins and soluble proteins depending on whether they are in the membrane or not. These proteins have various functions in our body. Many proteins have a unique function through their sequences and the resulting tertiary structure (2). Especially, helical membrane proteins are essential gatekeepers of cells, regulating flow of information and material across cell membranes (1). Complex tertiary structures, coupled with intricate conformational changes, enable helical membrane proteins to perform their functions (3, 4).

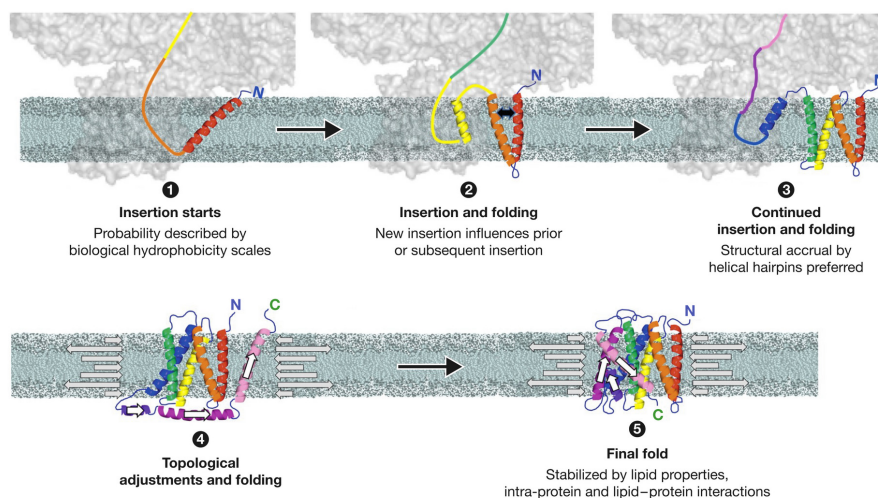


Figure 1.1 | [modified from Corin, Karolina, and James U. Bowie. EMBO reports 23.3 (2022): e53025.] Co-translational model of membrane protein folding

In order to form a structure in which the membrane protein can function, a process of insertion into a membrane and folding is required (5). However, structural

formation of membrane protein is not simple at all. Unlike soluble protein surrounded by a homogenous water environment, membrane protein is in the lipid bilayer, which is composed of polar head group and nonpolar carbon chain. Due to the complexity of this environment, folding of membrane protein is difficult to study, and its driving force is not easily defined (6).

To explain more about the driving force for membrane protein folding, lipid bilayer is the first consideration. The carbon chains in lipids are densely packed at the center, and the density is reduced at the interface close to the head group. These lipids interact and collide with each other to form a lateral pressure (7). The high lateral pressure is formed in the center of the bilayer and the lateral pressure decreases at the interface.

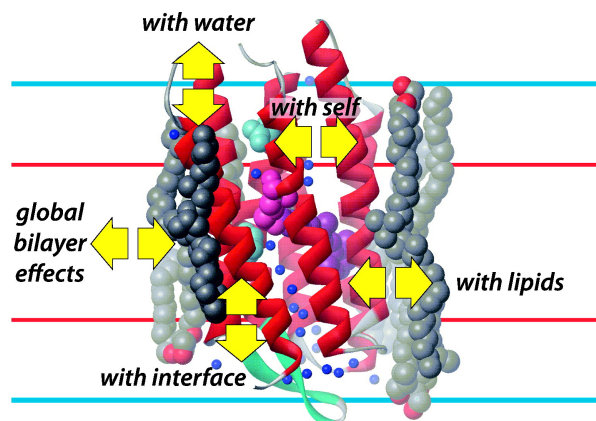


Figure 1.2 | [from Cymer, Florian, Gunnar Von Heijne, and Stephen H. White. *Journal of molecular biology* 427.5 (2015): 999-1022.] Summary of the various interactions that stabilize membrane proteins in lipid bilayers. Blue lines are interface boundaries, and red lines represent boundaries of the lipid hydrocarbon core.

Membrane proteins in these environments interact with a wide variety of molecules (5). First of all, membrane proteins interact with the surrounding lipids, and this interaction is diversified due to the amphiphilic nature of the lipids. There is also polar interaction at the interface of the lipid bilayer. Linker region of membrane proteins mainly interact with the interface. The interaction of membrane

protein itself should also be considered. In the case of helical membrane protein, intra- and inter- hydrogen bond exist, and there are other interactions depending on proteins (8). Also, for most helical membrane proteins, polar residues often serve the function (9). In this case, there is also an interaction with the water molecule.

1.2. Biogenesis of membrane proteins

While electron cryo-microscopy is revealing tertiary and quaternary structures of membrane proteins at an unprecedented pace (10), remarkably little remains known about how these complex structures fold following their synthesis in the endoplasmic reticulum (ER) membrane (11-13). Extensive quality control surveillance in the ER could still result in a low probability of successful folding of multi-pass membrane proteins: as low as 30% depending on the structural complexities of a given protein (12, 14). This success rate further plummets when germ line or somatic mutations are introduced (15, 16).

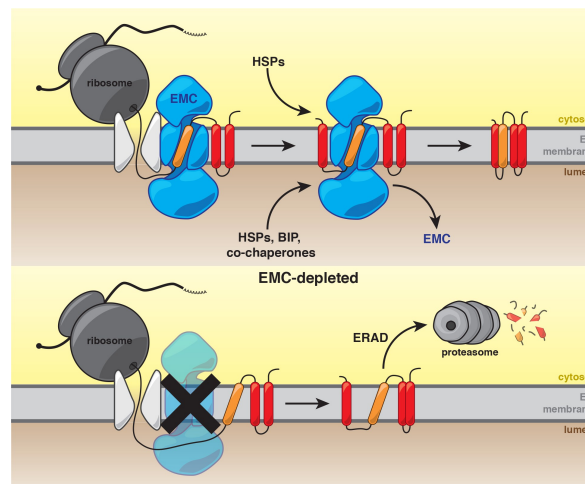


Figure 1.3 | [from Shurtleff, Matthew J., et al. *Elife* 7 (2018): e37018.] Model for the role of the chaperone in membrane protein biogenesis.

Indeed, most of the known deleterious mutations of multi-pass membrane proteins are thought to affect folding and membrane trafficking rather than biochemical function (11, 17). Cells invest significant resources to maintain

homeostasis of multi-pass membrane proteins, and failure to do so is thought to be a main cause of aging and diseased states of cells (18).

Despite the formidable complexity of membrane protein biogenesis, it is increasingly evident that some common principles guide this process. The process of TMH assembly is probably facilitated by ER chaperones, although dedicated TMH chaperones are poorly understood, and seem to function by preventing aggregation rather than promoting the correct fold (13, 19). It is thus tempting to hypothesize that the basic information for navigating the folding pathway – likely conserved across each family – is primarily encoded in the amino acid sequence of membrane proteins. Notwithstanding these prevailing models, the folding pathways of multi-pass membrane proteins remain largely elusive.

1.3. Major facilitator superfamily (MFS)

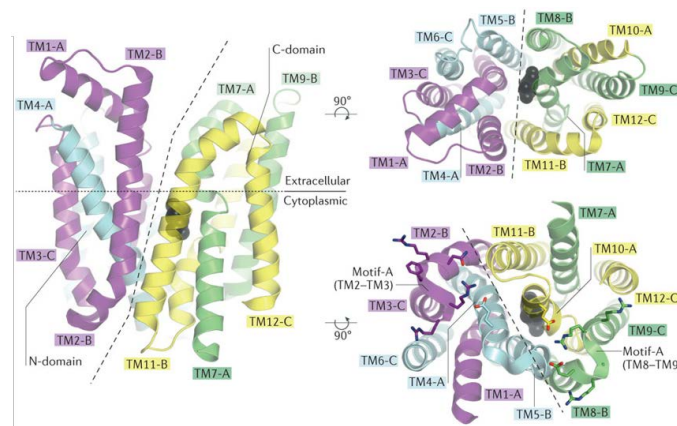


Figure 1.4 | [from Quistgaard, Esben M., et al. *Nature Reviews Molecular Cell Biology* 17.2 (2016): 123-132.] The structure of human glucose transporter 3

Many membrane protein families share a remarkable conservation of their tertiary structures despite huge evolutionary distances and low sequence homology across different members (3, 19-21). In particular, packing of individual transmembrane helices (TMHs) relative to one another helps drive the final conserved fold (22). The members of Major facilitator superfamily (MFS) are the transporter proteins with

such symmetric property (Figure 1.4). MFS transporters are composed of two structurally similar domains (dashes mark the domain barrier in the figure), each consisting of two inverted triple-helix repeats. Also, these proteins are essential for transportation of various substrate across the membrane. Especially, for the case of human glucose transporter 3 (GLUT3), the N- and C-domains undergo rocking motions within the bilayer to conduct transport, alternating between conformational states with access to extracellular and cytoplasmic spaces (23, 24). The interface between N- and C-domains is enriched with polar residues to create a conduit for glucose molecules in otherwise impermeable lipid bilayers (25).

1.4. Single-molecule magnetic tweezers

The magnetic tweezers are the technique that can apply a mechanical force on a single protein(26, 27). First, a string made of DNA is attached to both ends of the protein. One end is fixed to the bottom surface and the other end is conjugated with a magnetic bead. The magnetic bead moves in the direction of the magnet by the generated magnetic force. A tension is exerted on the string of DNA and the force is also transmitted to the target protein.

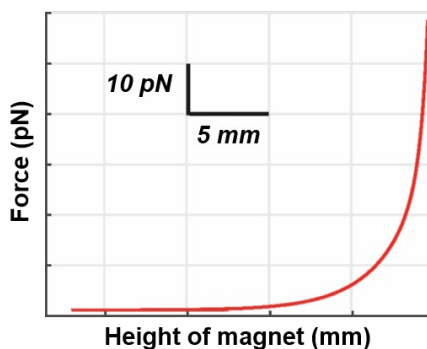


Figure 1.5 | Force application as change in the position of the magnet

The range of applied force using the magnetic tweezers is ~10 fN to 50 pN. The force

of pN scale is suitable for examining biological subject because the biological environment is governed by thermal energy. Also, the change in magnetic position in mm scale can manipulate the small change of force in pN scale (Figure 1.5).

Chapter 2. Methodology

2.1. Sample Preparation

2.1.1. Expression and purification of the human GLUT3

For single molecule assays, the GLUT3 glycosylation site N43 was deleted by mutating to Threonine (Thr, T). GLUT3 N43T is referred to as wild type (WT) GLUT3 throughout this work unless otherwise specified. To develop C-domain knotted GLUT3, S265 and A469 were mutated to cysteine (Cys) based on the structure (PDB:4ZWB). Likewise, for N-domain knotted GLUT3, T45 and K115 were mutated to Cys. GLUT3 was tagged with Spytag on the N-terminus and Spytag-HRV3C-GFP-10xHis on the C-terminus.

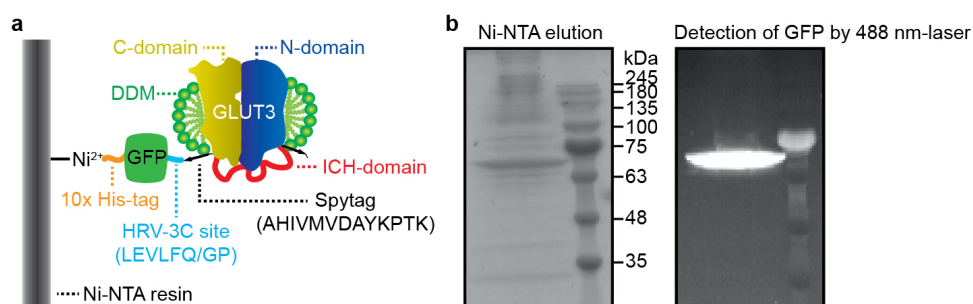


Figure 2.1 | GLUT3 construct and Ni-NTA affinity chromatography. **a**, illustration of GLUT3 construct for single molecule assay. **b**, Elution after his-tag affinity analyzed by SDS-PAGE. The band is detected by Coomassie blue staining for the left gel and by 488nm laser (for tagged GFP) for the right gel.

Preparation of GLUT3 and BODIPY-L-Cysteine assay was done by Dr. Hyunook Kang in Prof. Hee-Jung Choi group. GLUT3 was cloned to a modified pFastBac vector and each virus was made using Bac-to-Bac system (Invitrogen). Virus was added when *Spodoptera frugiperda* (Sf9) cells reached a density of approximately 3.0×10^5 cells/ml. Cells were harvested after 48 hrs and stored at -

80°C. Lysis was done with hypotonic buffer (20 mM HEPES 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% DDM, 0.5% CHS, 1 mM PMSF at 4 degree for 1 hr. Solubilized GLUT3 was removed from the insoluble fraction and was bound to Ni-NTA resin (Qiagen) at 4 degree for 1 hr. Resin was washed with high salt buffer (20 mM HEPES 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.

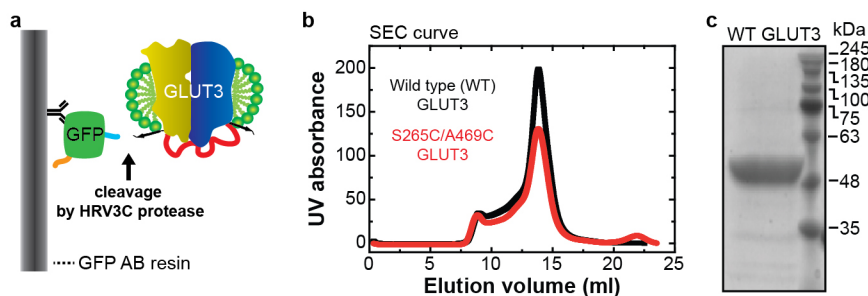


Figure 2.2 | Further purification process of GLUT3. **a**, GFP and 10x His-tag were removed by GFP nanobody column after cleavage reaction. **b**, Elution profile obtained by size exclusion chromatography (SEC) of each construct. **c**, Purified WT GLUT3 protein analyzed by SDS-PAGE. The major peak position in (**b**) was used for the gels.

GFP-10xHis tag was cleaved with home-made HRV3C protease at 4 degree. Uncut product and cleaved GFP tag were removed using home-made GFP nanobody column. GLUT3 was finally purified with size exclusion chromatography (GE healthcare) equilibrated with 20 mM HEPES pH 8.0, 150 mM NaCl, 0.03% DDM, 0.015% CHS.

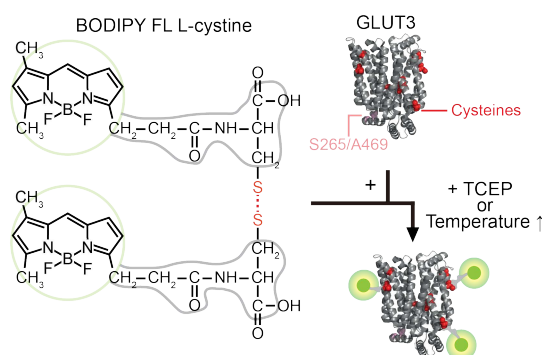


Figure 2.3 | Schematic of the assay using BODIPY-L-cystine. The left panel is the chemical structure of two BODIPY FL fluorophores attached to the amino groups of the disulfide-containing amino acid, cysteine. The right panel shows the structure of GLUT3 before and after the treatment (addition of TCEP or increasing the temperature). Green dots in the right panel are the BODIPY FL fluorophores reacted with cysteines in GLUT3.

BODIPY-L-Cysteine assay was used for assessing whether two cysteines in the mutant GLUT3 formed disulfide-bonds. BODIPY-L-Cysteine (Invitrogen) becomes fluorescent when their inter-BODIPY disulfide bonds are replaced by bonds with cysteine residues exposed on protein surfaces. The melting temperature of GLUT3 proteins was determined by measuring increases in the BODIPY fluorescence signals, which indicated melting of tertiary structures of the GLUT3 constructs and exposition of cysteine residues to aqueous buffer spaces to allow for BODIPY labeling.

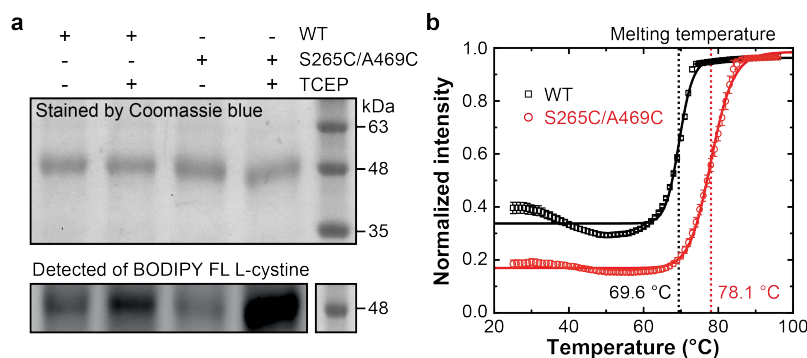


Figure 2.4 | Analysis of WT and S265C/A469C GLUT3 using BODIPY-L-cystine. a, Gel analysis for both constructs in the presence of TCEP. Upper gel shows the amount of GLUT3

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. **b**, 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).

4 μ g of GLUT3 was reacted with 5 μ M BODIPY-L-Cysteine (Invitrogen) using Rotor-Gene Q Thermocycler (Qiagen). Temperature was increased from 25°C to 95°C in 1 degree increment with an 10-second interval between each step. To identify the S265C/A469C disulfide bond, 4 μ g of GLUT3 was reacted with 1 mM tris(2-carboxyethyl) phosphine (TCEP) for 15 min. 5 μ M BODIPY-L-Cysteine was added and reacted for 30 min in the dark, followed by addition of 5x SDS buffer. All reactions were done at room temperature (RT). The extent by which GLUT3 was labeled with BODIPY was assessed via SDS-PAGE gel followed by imaging using ChemiDoc XRS+ (Bio-Rad). Dye signals were quantified using ImageJ.

2.1.2. Expression and purification of the human ER membrane protein complex

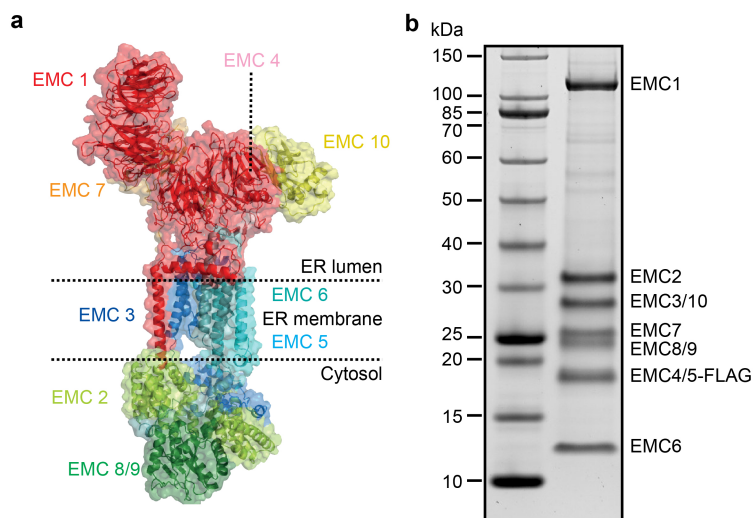


Figure 2.5 | Preparation of ER membrane protein complex (EMC). **a**, Structural information of the human ER membrane protein complex (EMC). **b**, Purified EMC analyzed by SDS-PAGE.

Preparation of EMC was done by Dr. Ben P. Phillips in Prof. Elizabeth A. Miller group. The human EMC was prepared essentially as described previously, but in n-Dodecyl-beta-Maltoside Detergent (DDM, Anatrace) rather than Lauryl Maltose Neopentyl Glycol (LMNG). 40 g pellets of suspension-adapted HEK293 T-REx Flp-In cells overexpressing EMC5-FLAG were solubilized in 40 ml of 2x solubilization buffer consisting of 100 mM HEPES (pH 7.4), 400 mM NaCl, 4 mM MgAc² and 1.6% deoxy big CHAP (DBC, Merck) for 1 h with gentle shaking on ice. All subsequent steps apart from FLAG elution were conducted on ice or at 4°C. Solubilized cells were clarified for 20 min at 21,000g at 4°C and incubated for 1 hour with 1 ml bed volume of anti-FLAG M2 affinity gel (Sigma) which had been pre-equilibrated in DBC wash buffer (0.3% DBC, 50 mM HEPES pH 7.4, 200 mM NaCl, 2 mM MgAc²). FLAG resin was collected by centrifugation (5 min at 1500g) and washed twice with 4 ml DBC wash buffer. Resin was transferred to a 10 ml gravity flow column and washed with 3 × 8 ml of 0.2% DDM wash buffer (0.2% DDM, 50 mM HEPES pH 7.4, 200 mM NaCl) allowing 10 min between washes to permit detergent exchange. Two FLAG elutions were conducted for 25 min each at room temperature by incubating the resin in 2 ml of FLAG elution buffer (0.02% DDM, 50 mM HEPES pH 7.4, 200 mM NaCl, 0.25 mg/ml 3X FLAG peptide, Sigma) with gentle end-over-end mixing. Combined FLAG elutions were diluted with 4 ml low-salt dilution buffer (0.02% DDM, 50 mM HEPES pH 7.4) and were bound to 150 µl bed volume of fast-flow SP sepharose (Cytiva), which had been pre-equilibrated with 2 ml ion exchange buffer A (0.02% DDM, 50 mM HEPES pH 7.4, 50 mM NaCl). SP sepharose was washed with 3 × 1 ml of ion exchange buffer A before 3 × 150 µl rounds of elution in ion exchange buffer B (0.02% DDM, 50 mM HEPES pH 7.4, 400 mM NaCl). The first elution was run over the resin twice and the resulting fractions were checked by A280. The peak fractions were then centrifuged for 15 min at 35,000g to remove insoluble aggregates. Glycerol was added to a final

concentration of 10% before snap freezing in liquid nitrogen for later use. Freeze-thawed aliquots were checked by nanoDSF using a Prometheus NT.48 and had melting curves similar to samples that had not been freeze-thawed.

2.1.3. Preparation of DNA handles

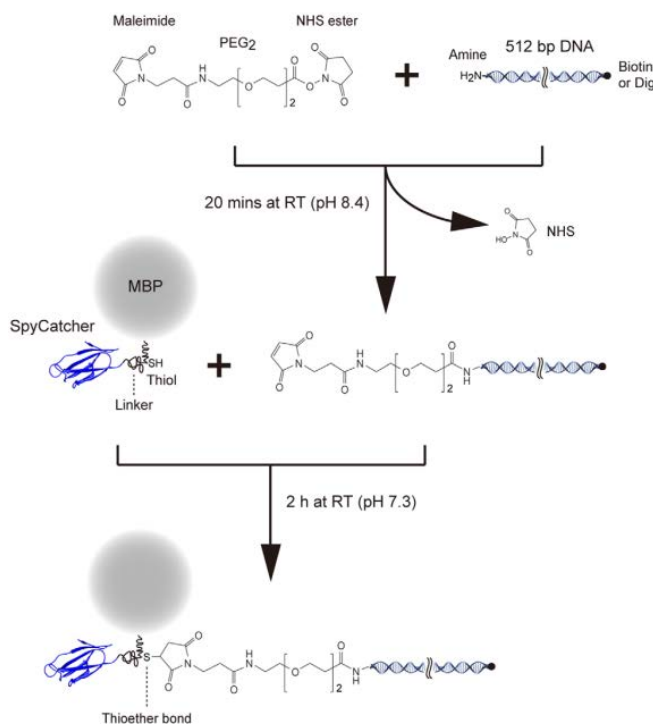


Figure 2.6 | [from Min, Duyoung, et al. *Protein Science* 25.8 (2016): 1535-1544.]
Schematic diagram of DNA-SpyCatcher conjugation process

The SpyCatcher proteins covalently linked to DNA handles were prepared in the following way. An amine group at one end of the 512 bp DNA fragment made by PCR was reacted with SM(PEG)₂ (PEGylated SMCC crosslinker; ThermoFisher Scientific) using an amine-sulphydryl crosslinker for 30 min at RT. After purification via DNA maxiprep, DNA fragments labeled with either biotin or digoxigenin at the other end were mixed in 1:1 molar ratio. Mixed DNA fragments were then covalently conjugated to purified SpyCatcher/Maltose Binding Protein (MBP) protein through

a thiol-maleimide crosslinking reaction overnight at 4°C. To purify SpyCatcher-conjugated DNA only, anion exchange chromatography using a 1 ml Mono Q column (GE healthcare) and amylose affinity chromatography (New England BioLabs) were used to exclude unconjugated SpyCatcher and unconjugated DNA, respectively. The purified SpyCatcher-DNA handles (in 50 mM Tris pH 7.5 and 150 mM NaCl buffer) were then concentrated up to ~ 100 nM and stored in 10 µl aliquots at - 80°C.

2.1.4. Preparation of bicelle

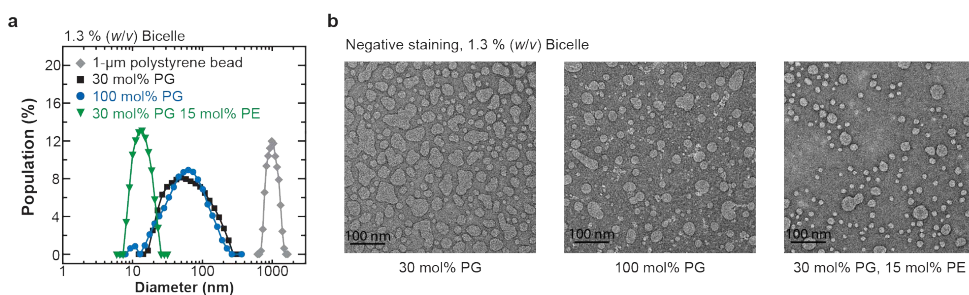


Figure 2.7 | Size distribution of bicelle and representative EM image of bicelle

All the lipids used for preparing the bicelle were purchased from Avanti polar lipids. 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC), 1,2-Dimyristoyl-sn-glycero-3-phosphorylglycerol (DMPG) and 1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine (DMPE) and 3-([3-Cholamidopropyl]dimethylammonio)-2-hydroxy-1-propane sulfonate (CHAPSO, Sigma-Aldrich) were used. Specific mixtures of DMPC, DMPG and/or DMPE lipids and CHAPSO were prepared with 2.8:1 molar ratio (*i.e.*, $Q \equiv [\text{lipids}]/[\text{detergent}] = 2.8:1$). This mixture of powder was dissolved in the buffer (50 mM Tris (pH 7.5), 150 mM NaCl) with concentration of 7.8% (w/v). The solution was fully dissolved after vortexing and incubation at 60°C for 2hrs. Then, freeze-thaw cycle is repeated until the solution is clear. The bicelle solution was stored in 150 µl aliquots at -80°C.

2.2. Experimental Methods

2.2.1. Dynamic light scattering (DLS) measurement

To determine sizes of the bicelles, a dynamic light scattering (DLS) apparatus (Otsuka electronics ELSZ-1000) was used. Bicelles ([lipids]:[CHAPSO] = 2.8:1 in molar ratio) with different composition of lipids (70:30:0, 0:100:0 and 55:30:15 of DMPC:DMPG:DMPE in mol%) were measured at 296K. 2 ml of bicelle buffer (1.3% (w/v) bicelles in 50 mM Tris (pH 7.5), 150 mM NaCl) in a glass-clear polystyrene cuvette was placed on the sample stage of DLS analyzer instrument. Data were analyzed using the associated software (Otsuka electronics Photal).

2.2.2. single-molecule magnetic tweezers experiments

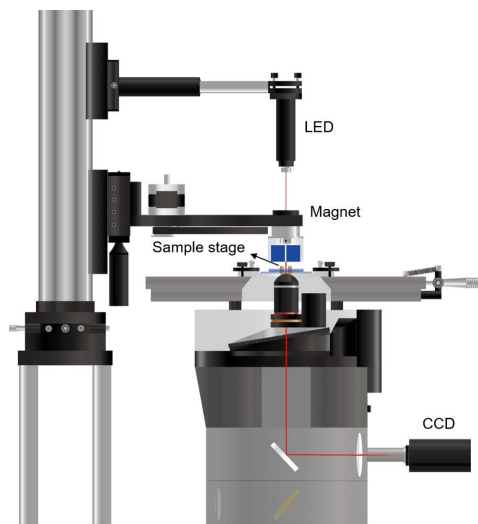


Figure 2.8 | Experimental setup of magnetic tweezers instrument

A magnetic tweezers instrument was custom built on an inverted microscope (28). The vertical position of a pair of permanent magnets (Neodymium magnets) was controlled using a translational stage (Physik Instrumente) to generate mechanical forces. Illumination with a super-luminescent diode ($\lambda = 680 \text{ nm}$, Qphotonics)

generated diffraction patterns for magnetic and reference beads (stuck on surface), of which images were recorded at an acquisition rate up to 1.2 kHz using a high-speed CMOS camera (Mikrotron). Diffraction patterns were pre-recorded by moving an objective lens using a piezoelectric nano-positioner (Mad City Labs) with respect to the sample in order to generate calibration tables for individual beads (magnetic and reference beads both). By comparing diffraction patterns of magnetic beads with the corresponding calibration table in real-time, 3D positions of the magnetic bead were tracked. Custom-written LabView programs were written and used for the single-molecule magnetic tweezers experiments.

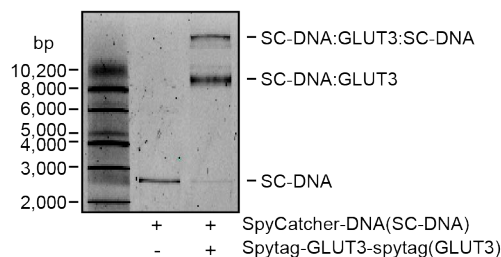


Figure 2.9 | Representative gel image of SDS-PAGE after SYBR green staining. The left lane shows SpyCatcher-DNA handle only, while the right lane exhibits a mixture of the SpyCatcher-DNA handle and the purified Spytag-GLUT3-spytag.

Samples for single-molecule magnetic tweezers experiment were prepared in the following way. WT or mutant GLUT3 proteins reconstituted in 0.02 to 0.04% of DDM were mixed with the SpyCatcher-DNA handles (with DDM added to a final concentration of 0.1% and TCEP added to 2 mM for WT GLUT3) and incubated for 20 to 22 h at 4°C to attach DNA handles at both ends of the GLUT3 proteins. 10:1 to 20:1 molar ratio for GLUT3 protein: SpyCatcher-DNA handles were used. After incubation, the protein-DNA hybrid complexes were diluted to ~1 nM final concentration of DNA using 1.3% (w/v) bicelle buffer (50 mM Tris pH 7.5 and 150 mM NaCl; DDM was thus diluted to below half its CMC). The membrane proteins connected with two DNA handles were then stored in 40 µl aliquots at -80°C.

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

2.2.3. Molecular dynamics simulations

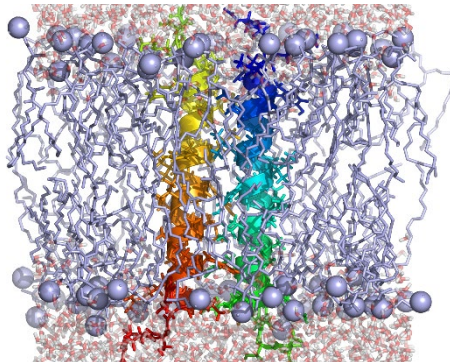


Figure 2.10 | [modified from Jo, Sunhwan, Taehoon Kim, and Wonpil Im., *PLoS one* 2.9 (2007): e880.] Snapshot of the protein extent (PDB:2HAC) along the z-axis

MD simulation was done by Dr. Soohyung Park in Prof. Wonpil Im group. The simulation systems were prepared using CHARMM-GUI Membrane Builder (29-32) with a crystal structure of GLUT3 (PDB:4ZWC). The structures for the N-terminal (residues 3 to 205) and C-terminal (residues 264 to 470) domains were extracted

from the full-length structure. To mimic experiments, the proteins were embedded in mixed bilayers of DMPC and DMPG (molar ratio of 7:3) or DMPC, DMPG, and DMPE (molar ratio of 5.5:3.0:1.5) solvated with bulk water and 150 mM NaCl at $T = 296.15$ K. Because the mixed bilayers with d14:0 tails are close to their phase transition temperature (e.g., $T_m \sim 23.5$ °C for DMPC), additional mixed bilayers of DMPC, DMPG, or DMPE were prepared at a higher temperature, $T = 306.15$ K to examine temperature effects. In addition, to explore the effect of tail saturation, mixed bilayers composed of palmitoyloleoyl (PO) PC, POPG, and POPE were also prepared at $T = 296.15$ K. The molar ratios of POPC:POPG and POPC:POPG:POPG in bilayers were set to be the same as those for DMPC/DMPG and DMPC/DMPG/DMPE bilayers, respectively. To ensure sufficient number of lipid shells around each domain or the full-length structure, the initial xy-dimensions of bilayers (i.e., the bilayer surface area) were set to be $\sim 130 \times 130$ Å² (N- and C-domains) and $\sim 150 \times 150$ Å² (full-length GLUT3), respectively. Each system was subjected to 0.5 or 1.0 μ s production run following a series of short equilibration runs. All simulations were carried out using OpenMM (33) with the CHARMM36 force fields (34, 35) and TIP3P water model (36, 37). The integration time step was set to 4 fs with the SHAKE algorithm (38) and hydrogen mass repartitioning method (39-41) during production runs. Lennard-Jones interactions were switched off over 10-12 Å by a force-based switching function (42) and the electrostatic interactions were calculated by the particle-mesh Ewald method (43) with a mesh size of ~ 1 Å. Temperature and pressure (1 bar) were controlled by Langevin dynamics (44) with a friction coefficient of 1 ps⁻¹ and a semi-isotropic Monte Carlo barostat (45) with a pressure coupling frequency of 100 steps in OpenMM simulations (46). Trajectories were analyzed using CHARMM (47) and in-house PYTHON scripts, where the interaction frequency between the polar/charged residues and their environments and the number of contacting water to these residues were calculated with 4.5 Å heavy-

atom distance criterion. The snapshots from simulations trajectories were prepared using VMD (48).

2.2.4. Sequence alignment and determination of helix insertion energy

Sequence alignment of sugar transporters was done by Dr. Charlotte Tumescheit in Prof. Martin Steinegger group. the helix insertion energies were compared into the endoplasmic reticulum (ER) membrane across kingdoms using DGpred (49) for a 19 residue window in the center of the helix. Since only a few sugar transporters have a determined structure available in the PDB, its helix annotation was transferred 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 (50). The entries were collected from currently available UniProt IDs.

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

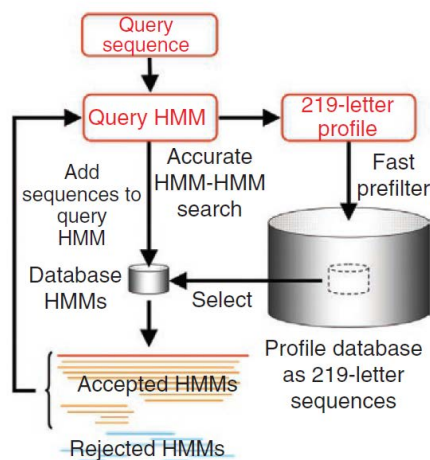


Figure 2.11 | [from Remmert, Michael, et al. *Nature methods* 9.2 (2012): 173-175.]
Basic workflow of HHblits

Since the sequences are divergent and therefore difficult to align, two strategies were employed to improve the alignment: (1) multiple references and (2) aligning profile HMMs instead of sequences. profile-HMMs were built for the 143 sugar transporters by aligning them against the UniRef30 database (2020_06) (51) using hhblits (v3.3.0, parameter -mact 0.1) (52). Each non-reference profile is aligned to the references using hhalgn (parameter - glob) (53, 54). The reference with the highest pairwise alignment score was chosen to transfer its annotation. First, the center is inferred from the residue that aligns with the reference center, however, in three cases the helix center did not align and one of the directly adjacent residues was chosen. Second, to counteract misaligned helix positions, the center positions were refined by using the position with the minimal energy for a ± 3 offset, calculated with DGpred. For the resulting coordinates, the insertion energy of extracted helices was obtained for the 19 residue long helices with DGpred.

2.3. Data Analysis

2.3.1. Force-extension curves (FEC) analysis

The FECs for DNA and unstructured polypeptide were fitted with the extensible worm-like chain (eWLC) model that describes behavior of the semi-flexible biopolymers under tension (55).

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

where the index i indicates either DNA or unstructured polypeptide (p), $k_B T$ is the thermal energy, l_c and l_p are the contour length and persistence length, respectively ($l_{c,DNA} = 0.338$ nm, $l_{c,p} = 0.36$ nm and $l_{p,DNA} = 38.5$ nm, $l_{p,p} = 0.39$ nm)(56-58). K_i is the elastic modulus ($K_p \sim 50$ μ N and $K_{DNA} \sim 500$ pN) (59, 60), F is the applied force and a_j are polynomial coefficients for the improved approximation. n_i is the total number of constituent monomers of each component such as DNA and polypeptide ($n_{DNA} = 512$ for each handle, $n_{linker,p} = 18$ between the GLUT3 and DNA handle, and $n_{GLUT3,p} = 463$ (*i.e.*, $n_{GLUT3,p} = n_{N-domain,p} + n_{C-domain,p} + n_{ICH-domain,p} = 198 + 207 + 58$)) for GLUT3 (25). To describe a rigid-like biopolymer such as helical states (U_h), the Kessler-Rabin (KR) model was used (61, 62),

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

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

In the force-ramp and force-jump experiments, observed extension values can be

estimated from a linear superposition of extensions of all components in tweezing system. The fully unstructured coil state (U_c) and helical state (U_h) are thus described as follows.

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

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

To analyze relative extension changes during high-force unfolding, N- and C-domains were treated independently because two domains are separated by the ICH-domains. The extension increase observed for an intermediate state ($z_{i,p}$) is proportional to the number of unfolded amino acids (Δn_i), giving $\frac{z_{i,p}}{z_{\text{N(C)}-\text{domain},p}} = \frac{\Delta n_i}{n_{\text{N(C)}-\text{domain},p}}$, where $n_{\text{N(C)}-\text{domain},p}$ is the extension increase expected when N- or C-domain is fully unraveled and $n_{\text{N(C)}-\text{domain},p}$ is the total number of amino acids in N- or C-domain, respectively. Because the remaining partially folded structures have finite thickness values along the pulling axis (d_i), the relation of $z_{m,i} = z_{i,p} + d_i - d_{\text{N(C)}-\text{domain}}$ is applicable for N(C)-domain. $d_{\text{N-domain}}(d_{\text{C-domain}})$ 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 $\Delta n_i \simeq$

$\frac{n_{N(C)-\text{domain},p}}{z_{N(C)-\text{domain},p}}(z_{m,i} - d_i + d_{N(C)-\text{domain}})$. The intersection between functions in the left-hand side and right-hand side yields the number of amino acids from the reference point where unfolding starts.

2.3.2. Hidden Markov Model (HMM) analysis

Hidden Markov Model (HMM) analysis was employed to determine the folding/unfolding intermediate states from the time-resolved low-force extension traces recorded at 1.2 kHz (63). The adjustable parameters in MT experiment are the number of states (n), the extension position for i -th intermediate state (\bar{z}_i), and the transition matrix of rates between states (\vec{k}). The optimal number of states (n) was obtained from Bayesian Information Criterion (BIC): $\text{BIC} = q \ln(N) - 2 \ln(L)$ where q is the number of output parameters given by model, N is sample size and 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 (64). The extension traces were median-filtered with 5-Hz window, and the extension position/deviation for each state was estimated from the Gaussian Mixture Model (GMM) in the HMM analysis. The rates (*i.e.*, the transition matrix) were then determined using the optimal parameters for the number of states and extension positions. The rates estimated from HMM were confirmed by checking single exponential fitting of the dwell time distributions. In this process, dwell time data shorter than 50 ms were considered artifact and ignored because median-filtered traces (5 Hz or 200 ms) were used for analysis. Finally, the resulting traces were verified by the Viterbi algorithm.

2.3.3. Deconvoluted extension probability analysis

The deconvoluted extension probability analysis was done in collaboration with Dr.

Hyun-Kyu Choi. To obtain an extension distribution of single GLUT3 protein ($\tilde{P}_p(z_p; F)$) with Brownian noises of magnetic beads and handles removed, established method in optical tweezers studies were utilized (65, 66). Because the magnetic bead in magnetic tweezers is not physically trapped unlike with optical tweezers (*i.e.*, magnetic force is not a fluctuating variable but stably fixed), the marginal probability distribution ($\tilde{P}_m(z_m; F)$) from Hamiltonian of the bead in the presence of magnetic force could be directly described as $\tilde{P}_m(z_m; F) \approx \frac{1}{4} e^{\beta F z} P_m(z) = \frac{1}{4} P_m(z; F)$ where $P_m(z; F)$ is the measured equilibrium probability of the total bead-handle-protein system with separation z at the constant force F ; β is $1/k_B T$. By performing deconvolution in real-space, the following integral is derived.

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

Where $\tilde{P}_{bh}(z; F)$ is conjugated probability of handles (PEG polymers (peg), two DNA handles (dh; dh1 defined as DNA handle directing towards magnetic bead, dh2 towards peg) and two polypeptide linkers (ph) between DNA and GLUT3) and magnetic bead. In brief,

$$\tilde{P}_{bh}(z; F) \equiv \mathcal{F}^{-1} \left(\tilde{P}_b(k; F) \tilde{P}_{peg}(k; F) \tilde{P}_{dh1}(k; F) \tilde{P}_{dh2}(k; F) \tilde{P}_{ph}^2(k; F) \right)$$

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 rest terms in $\tilde{P}_{bh}(z; F)$ can be described by the following equation.

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

Where the index j represents the components composed of peg, dh1, dh2 and ph.

The corresponding total contour length is $L_{c,j}$. $\Psi_{B.C}$ and $E_{n,j}$ are an eigen state and eigen value (total energy), respectively as previously defined and estimated from effective Hamiltonian equation of propagator of biopolymer in Markovian regime (57). Index B.C in eigen state indicates whether semi-flexible biopolymer is half-constrained (one side of peg and dh1) or unconstrained (dh2, ph).

To avoid any numerical instability and ill-conditioned result, suitable fitting functions were substituted 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 (Median-filtered traces with 5-Hz window were used. Because the characteristic time scale of magnetic bead is less than 30 ms, the behavior of the bead could be deconvoluted 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 linear combination and N_λ is total number of Gaussian components (for simplicity, $N_{bh} = 1$ was chosen). Then, parameters of the deconvoluted extension distribution of the single GLUT3 are described as $w_i^p \approx w_i^m$, $\mu_i^p \approx \mu_i^m - \sum_{j=1}^{N_{bh}} w_j^{bh} \mu_j^{bh}$ and $\sigma_i^{p^2} = \sigma_i^{m^2} - \sigma_i^{bh^2} - 2cov(z_m, z_{bh})$. For ensemble averages of the deconvoluted probability distributions, weighted arithmetic mean was used to visualize the average probability distribution (*i.e.*, $\langle \tilde{P}_p(z_p; F) \rangle \equiv \sum_{m=1}^M a_m \tilde{P}_{p,m}(z_p; F) / \sum_{m=1}^M a_m$ where m is the number of traces, M is the total number of measured traces and a_m is the normalized weighting factor, which depends on sample size in each trace.

Chapter 3. Results

3.1. Single-molecule magnetic tweezers monitoring GLUT3 folding

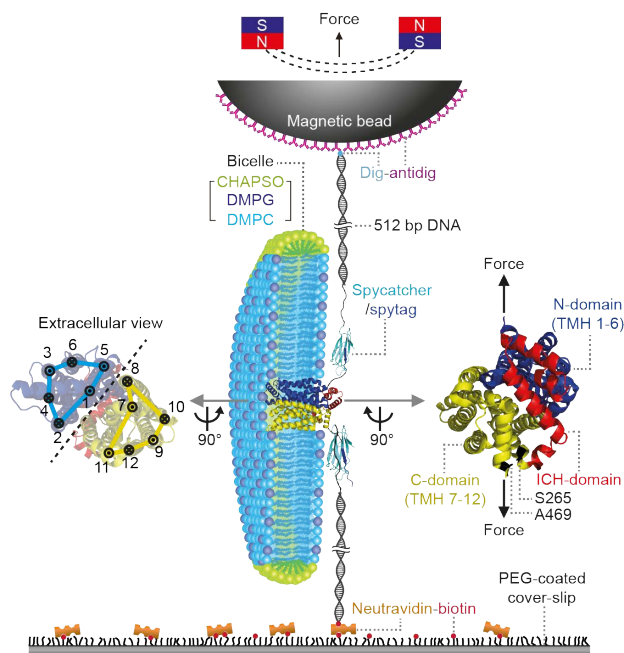


Figure 3.1 |Schematic of magnetic tweezers (MT) experiment for observing folding of a single GLUT3 protein. Extracellular and intracellular view of GLUT3 structures are shown with TMH numbers and pulling positions depicted (left and right inset).

Employing magnetic tweezers to observe folding of single human GLUT3 proteins, I attached DNA handles to the N- and C-termini of GLUT3 using the SpyTag/SpyCatcher system (Figure 3.1) (28, 67). After attaching the DNA handles to a magnetic bead and a polymer-coated surface, bicelle solutions were introduced, with varying lipidic compositions, to provide lipid bilayer environments to GLUT3 (Figure 3.1) (68-70). While applying a varying level of magnetic force to the bead by moving a pair of neodymium magnets, the vertical position of the magnetic bead was recorded (referred to as the extension) at sampling rates up to 1.2 kHz. The

uncertainty in bead tracking could be reduced to ~ 1 nm through median filtering at 5 Hz (Supplementary figure 1) (28).

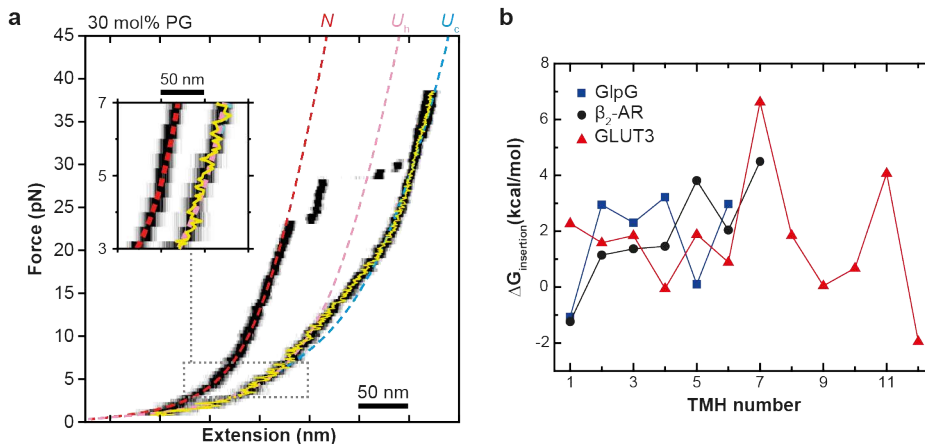


Figure 3.2 | Representative FEC of a single GLUT3 protein and insertion energy values of TMHs in 3 membrane proteins. a, FEC is shown as black heat map. The yellow trace shows the mean extension value in the relaxation phase. Theoretically expected FECs for the N , U_h , and U_c states are overlaid as red, pink, and light blue dashed lines, respectively. **b**, Insertion energy values calculated for individual TMHs for *E. coli* GlpG, human β_2 AR and GLUT3 using the biological hydrophobicity scale from the translocon-ER membrane system (49).

First, the force-extension curve (FEC) was examined during gradual stretching and relaxation of single GLUT3 (Figure 3.2a). 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, a transition was detected from the theoretical curve for U_c to U_h in the force range from 20 to 10 pN. Since the U_h curve was generated assuming a fully-stretched state with α -helical structures restored for all TMHs, the observed transition indicated gradual coil-to-helix transitions in twelve TMHs of GLUT3. When further relaxing tension to below 5 pN, the FEC continued to follow the U_h curve (Figure 3.2a, inset). This observation is markedly different from previous observations that FECs of GlpG and β_2 AR fell significantly shorter than U_h below 5 pN (28). This persistent U_h state presumably

hidden Markov modeling (HMM) and Bayesian information criteria (BIC) (63, 64) 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 (Figure 3.4a and Supplementary figure 2).

Given previous observations that addition of negatively charged lipids facilitates membrane protein folding(28, 70), a bicelle phase consisting purely of DMPG lipids was tested to further enhance the folding progress (Figure 3.4b). A remarkable enhancement in folding progress was observed, reaching up to 73 % of full folding (*i.e.*, an extension decrease of 35.6 nm) (Figure 3.4b and Supplementary figure 2). The HMM and BIC analysis revealed that the positions of the first four intermediates, I_{f1} to I_{f4} , remained largely invariant (Figure 3.4) and that there were two intermediates (I_{f5} and I_{f6}) in the extension space newly charted by the use of 100 %-DMPG bicelles (Figure 3.4b and Supplementary figure 2). 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.

3.2. Mapping the folding order of single GLUT3 domains

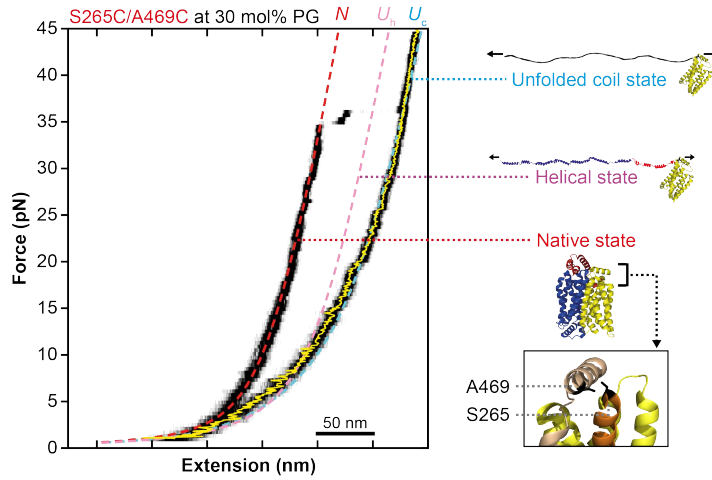


Figure 3.5 | Representative FEC of a single S265C/A469C GLUT3 protein. FEC is shown as black heat map. Figures in the right show the position of cysteine mutations on TMHs 7 and 12. Other definitions are as in Figure 3.1.

To map the observed folding pathway to specific domains of GLUT3, Dr. Choi and I constructed a variant of GLUT3 with two mutations: S265C and A469C (referred to as GLUT3_{CC}) in collaboration with Dr. Kang (Figure 3.5 and Supplementary figure 3). The introduced cysteines formed a disulfide bond that knotted the entire C-domain, rendering it as one fixed unit in the mechanical interrogation (Figure 3.5).

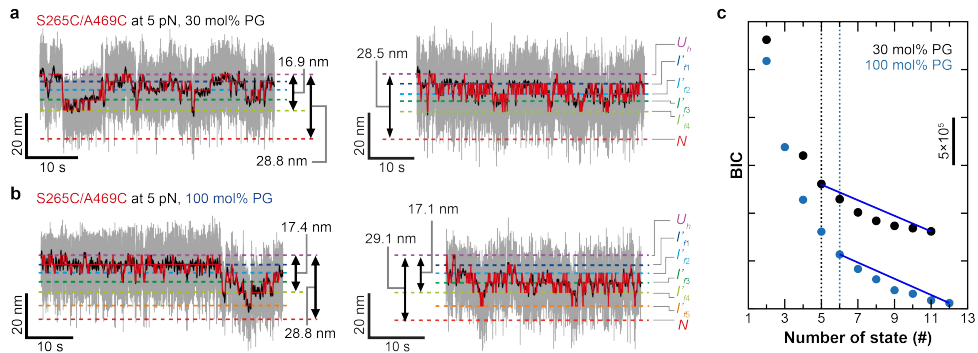


Figure 3.6 | Folding characteristics of GLUT3_{CC}. a,b, Folding traces with HMM results for GLUT3_{CC} folding at 5 pN with 30 mol% (b) and 100 mol% DMPG (c). Two replicates are shown for each condition, and each colored trace is defined as in Figure 3.4. c, BIC values

of GLUT3_{CC} for each number of states ($n = 22$ and 12 for 30 mol% and 100 mol% PG, respectively).

When examining the folding traces of GLUT3_{CC} obtained at 5 pN, extents of extension were almost identical for bicelle conditions with either 30 or 100 mol% DMPG lipids (Figure 3.6a,b). This was in sharp contrast with the observation for wild-type (WT) GLUT3 where use of 100 mol% DMPG doubled the folding progress. The last gap before the native state, which slightly shrank to ~ 12 nm, persisted for both bicelle compositions. Using HMM and BIC analyses, four intermediates was detected as the maximum likelihood estimation for the extension traces obtained for GLUT3_{CC} (Figure 3.6).

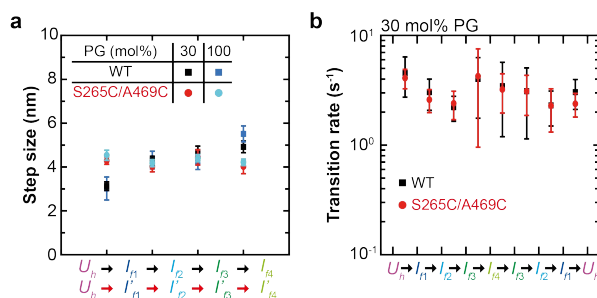


Figure 3.7 | The positions and transition kinetics of folding intermediates. **a**, Step sizes between the neighboring states at 5 pN ($n = 16$, 11 , 22 and 12 traces for WT GLUT3 and GLUT3_{cc} with 30 and 100 mol% DMPG, respectively). Error bars are SEM. **b**, Transition kinetics between the neighboring states at 5 pN. The number of traces is as in **(a)**. Error bars are SEM.

The positions and transition kinetics of these four intermediates were largely identical to those of the first four intermediates observed for WT GLUT3 (Figure 3.7). 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} (Figure 3.7a).

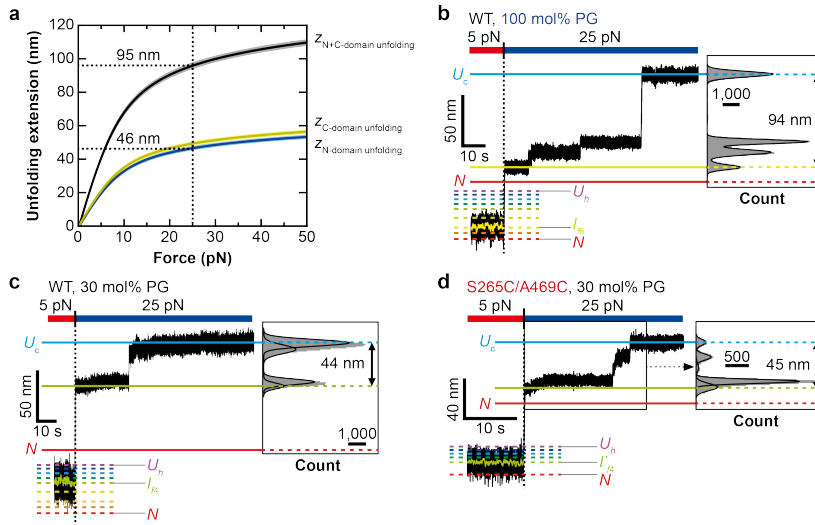


Figure 3.8 | Unfolding step size of the intermediates I_{f4} and I_{f6} . Estimated unfolding step sizes for linked N- and C-domains (black) and isolated N- (blue) and C-domains (yellow). The shaded area means SEM. **b-d**, Representative traces for the force-jump experiments applied to I_{f6} (**b**), I_{f4} (**c**), and I'_{f4} (**d**). The unfolding intermediate withstood the 25 pN tension for more than hundreds of milliseconds. Insets show the distributions of extensions recorded after force jumps to 25 pN.

Based on these results, I propose that the first four intermediates (*i.e.*, I_{f1} to I_{f4}) correspond to folding of GLUT3's N-domain. The following two folding intermediates (I_{f5} and I_{f6}), which could be accessed in the DMPG-100 mol% condition for WT GLUT3 but vanished for GLUT3_{CC}, are attributed to C-domain folding. To determine whether there were indeed partial structures formed in individual intermediates, a force jump to 25 pN was applied when WT GLUT3 reflected I_{f6} (the last intermediate before the 13 nm gap). A partially folded structure was detected and it showed unfolding with a large step size of ~94 nm, closely matching what would be theoretically expected for unfolding of both N- and C-domains (but separated) (Figure 3.8 and Supplementary figure 4). When the same force jumps were applied to I_{f4} of WT GLUT3 and GLUT3_{CC}, a partially folded structure with an unfolding step of 45 nm was detected under 25 pN, an expected value for N-domain unfolding (Figure 3.8a,c,d). These results support the

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.

3.3. Dissecting folding steps of the MFS folds

I next attempted to dissect more detailed folding steps within individual N- and C-domains.

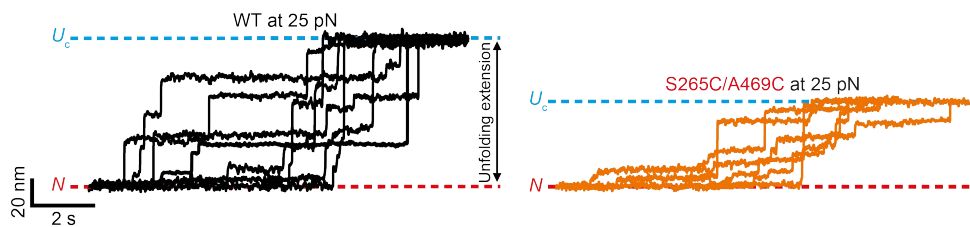


Figure 3.9 | Representative traces of high-force unfolding of single GLUT3 proteins. Black and orange lines represent WT GLUT3 and GLUT3_{cc}, respectively. A lipid composition of DMPC:DMPG=70:30 (mol/mol) was used for the bicelles. The traces were recorded at 1.2 kHz and subsequently median-filtered at 50 Hz.

To this end, I 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 (Figure 3.9).

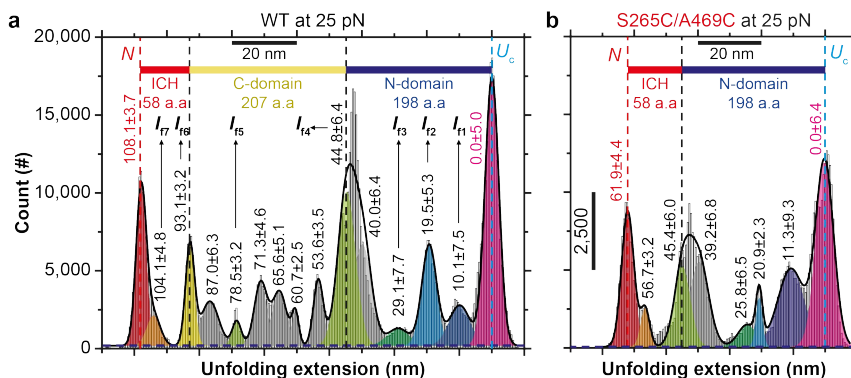


Figure 3.10 | Distributions of extension values recorded during high-force unfolding.

The peaks indicate the fit centers of multiple Gaussian functions. Relative extension values are measured from the U_c state and represent mean \pm SD. The upper diagrams depict the number of amino acids of corresponding domains to guide mapping onto the structure. $n = 32$ and 18 high-force unfolding traces for WT (a) and S265C/A469C(GLUT3_{cc}) (b) GLUT3, respectively.

The unfolding extensions displayed clearly peaked distributions (with 50 Hz median filtering applied), each of which was assigned as a high-force unfolding intermediate (Figure 3.10). In addition, as demonstrated for I_{f4} and I_{f6} in Figure 3.8, force jump experiment at each intermediate was done and unfolding traces were observed during folding trials at 5 pN. This series of experiments permitted establishment of a crucial one-to-one correspondence between the low-force and the high-force intermediates (Figure 3.11).

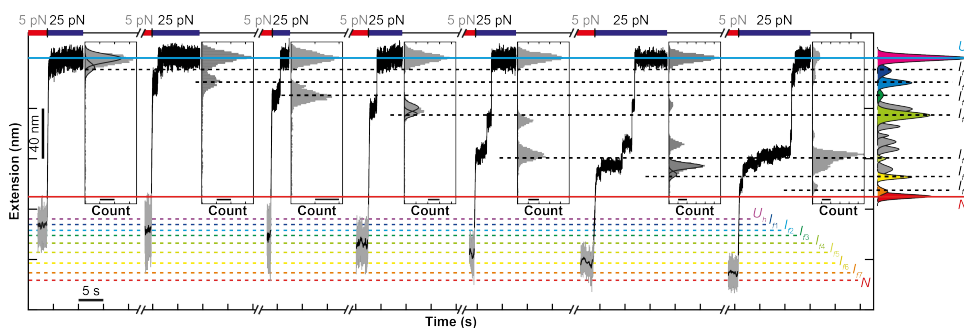


Figure 3.11 | Representative traces from the force-jump experiments applied to individual low-force folding intermediates. Each inset shows an extension distribution recorded after the force jump to 25 pN (scale bar is 500 count). Dashed lines indicate close alignment of the extension states after the force jumps with one of the unfolding peaks identified in Figure 3.10a.

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 (Figure 3.11). 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) (Figure 3.10). These results support the 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, I first focused on the C-domain that showed only two dominant intermediates of I_{f5} and I_{f6} during the 5 pN re-folding process (e.g., Figure 3.4b and Figure 3.10a). In such two-step folding, the folding process should start from either N- or C-terminus of the C-domain. Otherwise, a partially folded structure at I_{f5} would have flanking N- and C-terminal tails, requiring more than one steps to finish C-domain folding and thus incompatible with the observed two-step folding. Furthermore, the unfolding step from I_{f5} to I_{f4} corresponded to two third of the unfolding extension of the entire C-domain, suggesting that the folding step in the reverse direction (*i.e.*, from I_{f4} to I_{f5}) would involve four out of six TMHs of the C-domain (Figure 3.10a). 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 (Figure 3.12a). 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) (Figure 3.12a).

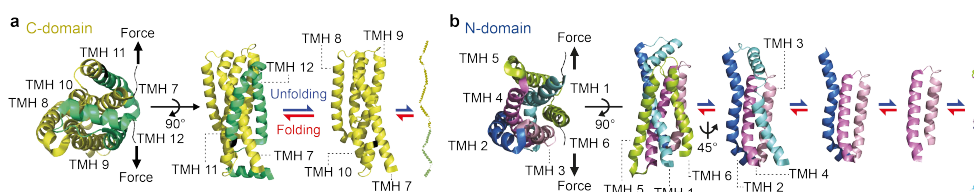


Figure 3.12 | Schematics of folding and unfolding of the C- and N-domain of GLUT3.

Given the remarkable pseudo-symmetry of the N- and C-domains of GLUT3 (21, 25, 72), I assumed that a similar pathway guides folding of the N-domain. Indeed, TMH topology of the N-domain is similar to that of the C-domain, with TMHs 5 and 6 embracing TMH 1 (Figure 3.12b), which would allocate TMHs 5 and 6 as the last structural unit in N-domain folding. For the partial structure

composed of TMHs 1 to 4, TMHs 1 and 2 in turn wrap around TMH 4 while making multiple atomic contacts among them, which likely renders TMHs 1 and 2 tailing TMH 4 in the folding order (Figure 3.12b and Supplementary figure 5a). To examine this hypothesis more directly, Dr. Kang generated another GLUT3 mutant harboring T45C and K115C (GLUT3_{TM23C}), in which TMHs 2 and 3 were knotted together by the disulfide linkage (Supplementary figure 5).

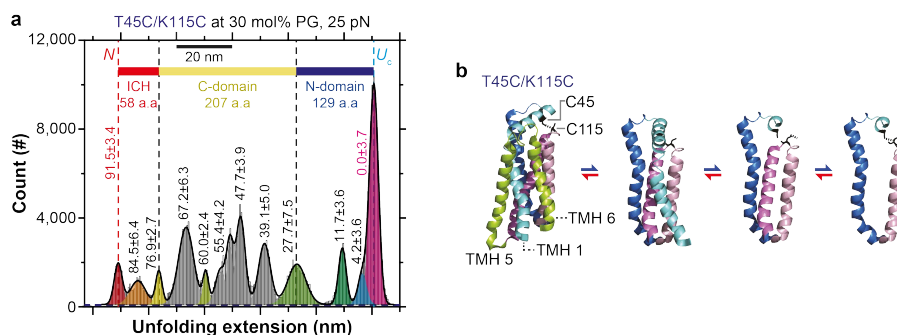


Figure 3.13 | Folding and unfolding behavior of the N-domain of GLUT3_{TM23C}. **a**, Distribution of extension values recorded during high-force unfolding of single GLUT3_{TM23C} proteins. Extension values represent mean \pm SD ($n = 19$). **b**, Schematics of folding and unfolding of the N-domain of GLUT3_{TM23C}.

When examining the high-force unfolding pattern of GLUT3_{TM23C} with that of WT GLUT3, the disulfide bond 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 (Figure 3.13 and Supplementary figure 6a-e). The second unfolding step was slightly reduced to 7.5 nm, which was remarkably consistent with the fact that the last helical turn of the long linker region after TMH 1 would be protected by knotting, mapping the second unfolding step to that of TMH 1 and its linker region (Figure 3.13 and Supplementary figure 6a-e). Furthermore, it was found that the last two unfolding steps before U_c were substantially reduced to a single step of 4.2 nm, which would reflect unfolding of TMH 4 outside the knotted region (Figure 3.13 and Supplementary figure 6f). Together, a scenario consistently

supported by multiple lines of data is four-step folding of N-domain, in which TMHs 3 and 4 form the first helical hairpin (U_z to I_{f1}), followed by sequential addition of TMHs 2 and 1 to the structure (I_{f1} to I_{f2} and I_{f2} to I_{f3} each) and completed with addition of TMHs 5 and 6 (I_{f3} to I_{f4}) (Figure 3.12b).

3.4. EMC facilitates insertion of TMHs of GLUT3

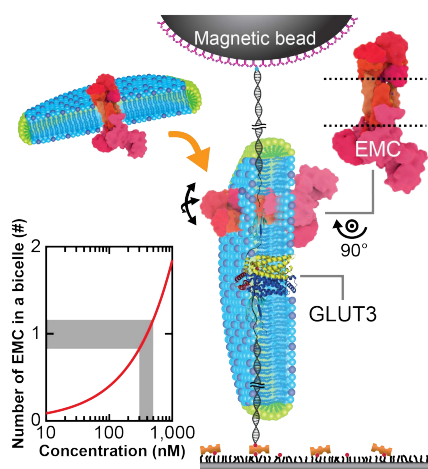


Figure 3.14 | Schematic of an MT experiment that examines the effects of EMC on folding of GLUT3. Inset shows the average number of EMCs in individual bicelles as a function of EMC concentration.

These observations indicated that GLUT3 has a weaker propensity for folding than GlpG and β 2AR and thus requires a more physiological, alternative mechanism that might assist folding. So, I turned the attention to EMC (73), a large multi-protein complex with 9 members in humans (Figure 2.5) (74). EMC is shown to induce effective membrane insertion of tail-anchored proteins and the first TMHs of G-protein-coupled receptors (75, 76). Specifically, this membrane insertase activity is manifested when TMHs of target proteins exhibit lower levels of hydrophobicity (75-77).

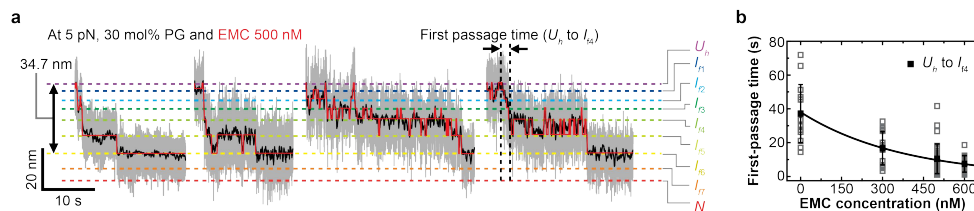


Figure 3.15 | Folding behavior of WT GLUT3 with EMC. **a**, Representative folding traces of WT GLUT3 obtained with 30 mol% DMPG and 500 nM EMC. Four replicates are shown. **b**, Mean first-passage time for I_{f4} determined for different EMC concentrations. Error bars mean SD ($n = 12, 15, 35$ and 15 traces for the cases with EMC = $0, 300, 500$ and 600 nM, respectively).

In collaboration with Prof. Elizabeth A. Miller group, Dr. Phillips purified human EMC and I added the complex reconstituted in bicelles to single-molecule magnetic tweezers assay (

Figure 3.14). We anticipated that EMCs could be delivered to tweezed single GLUT3 proteins because individual bicelles undergo frequent fusion and fission with one another (78). Indeed, when adding 500 nM EMCs to the single GLUT3 folding assay, which corresponded to approximately one EMC in each bicelle (

Figure 3.14, inset), remarkable facilitation of GLUT3 folding was observed under the 30 mol% DMPG condition. Many of single GLUT3 folding traces progressed as far as ~ 34.7 nm, a direct indication of EMC contribution to successful folding of the entire N- and C-domains (Figure 3.15a). This stimulation of folding progression virtually disappeared when I added an unrelated membrane protein reconstituted in bicelles, indicating the specificity of the EMC (Supplementary figure 7). Indeed, when assessing the time required to first reach the extension value of 17.2 nm (corresponding to I_{f4}), this first-passage time was increasingly shortened as a higher EMC concentration was used (Figure 3.15b).

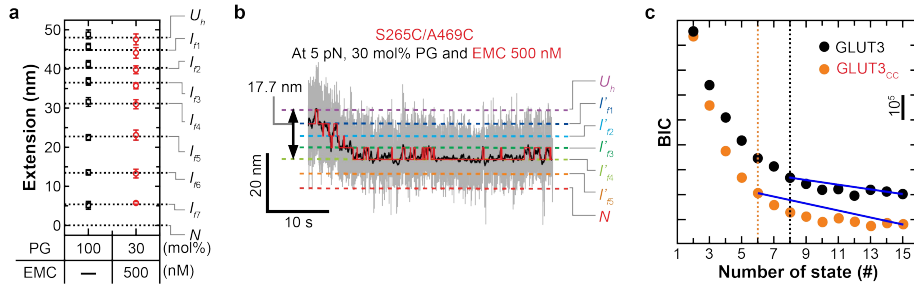


Figure 3.16 | Folding characteristics of WT GLUT3 and GLUT3_{CC} in the presence of EMC. **a**, Positions of folding/unfolding intermediates identified with HMM are depicted for the indicated conditions. Error bars represent SEM ($n = 22$ and 35 traces for 100 mol% DMPG and 30 mol% DMPG with 500 nM EMC, respectively). **b**, Representative folding traces of GLUT3_{CC} obtained with 30 mol% DMPG and 500 nM EMC. **c**, BIC values for the indicated number of states ($n = 13$ and 11 traces for WT GLUT3 and S265C/A469C GLUT3, respectively).

Using the HMM and BIC analysis, Dr. Choi and I analyzed patterns in the folding traces and found that the number and positions of the intermediates were essentially preserved in the presence of EMC, an observation recapitulated for C-domain-knotted GLUT3_{CC} (Figure 3.16). These observations suggest that EMC helps GLUT3 navigate down the folding intermediates encoded in its native amino acid sequence, rather than creating novel folding pathways.

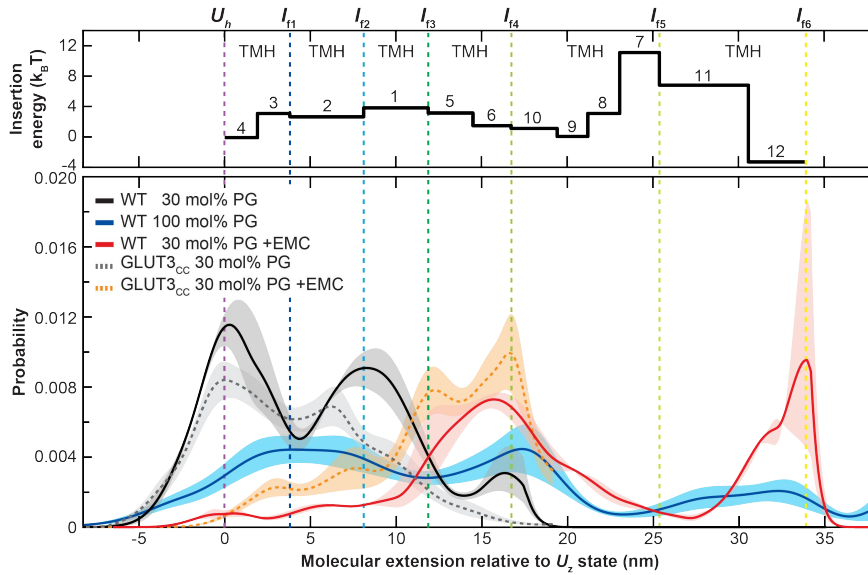


Figure 3.17 | Probability distributions of deconvoluted extension values observed under indicated folding conditions at 5 pN. The shaded area means SEM. Upper panel shows insertion energy values of individual TMHs aligned along the folding order identified in Figure 3.12. The insertion energy values were calculated based on the biological hydrophobicity scale from the translocon-ER membrane system as in Figure 3.2b.

To examine EMC's effects on GLUT3 folding with higher resolution, Dr. Choi and I deconvoluted the extension distributions to remove some of the broadening effects caused by Brownian noises from the magnetic beads and DNA handles (Figure 3.17 and Supplementary figure 8). The resulting extension distribution clearly showed populations markedly increased beyond I_{f4} , indicating that EMC indeed helped GLUT3 sample microstates for C-domain folding (Figure 3.17, red vs. black distributions). In addition, a major valley was appeared in the extension distribution at around 25 nm, a major setback for GLUT3's efforts in C-domain folding, which was also observed with the 100 mol%-DMPG condition (Figure 3.17, red vs. blue distributions). Remarkably, this valley approximately coincides with the folding steps of TMHs 7 (a broken helix) and 11 that are estimated to confer the highest energetic costs for TMH insertion (Figure 3.17). It means that EMC successfully propelled single GLUT3s through these barriers to reach I_{f6} (Figure 3.17). Thus, this observation suggests that EMC helps TMH insertion for GLUT3 beyond its first TMHs, which becomes most accentuated for TMHs with low scales of hydrophobicity. One more notable observation is surging of a peak at 17.2 nm in the presence of EMC, akin to that at 34 nm, suggesting an increased stability of the N- and C-domains once they were completely folded (Figure 3.17, red vs. black distributions). Lastly, stretched ICH domains failed to fold even with EMC, observed as the persistent 13 nm gap before the *N* state (Figure 3.15a and Figure 3.16b).

3.5. PE lipids boost domain-domain assembly of GLUT3

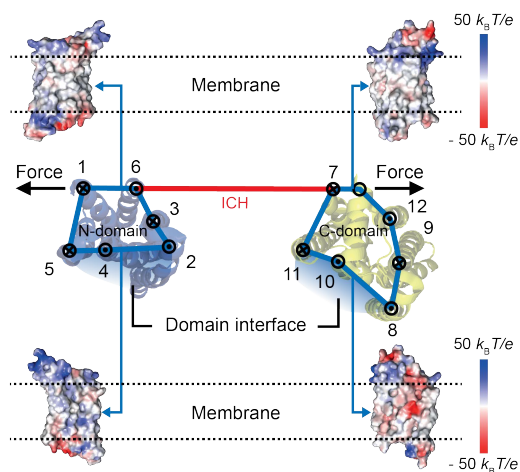


Figure 3.18 | Cartoon of a single GLUT3 protein at I_{16} before domain-domain assembly. The N- and C-domains are folded and ICH-domains are stretched under mechanical tension. The electrostatic potentials of the outer and inner surfaces of GLUT3 are shown in upper and lower insets, respectively.

I next asked whether it is feasible to induce assembly between the N- and C-domains to complete the known tertiary structure of GLUT3 (Figure 3.18). Since neither negatively charged lipids nor EMC could facilitate domain-domain assembly, I propose that there exists a high energy barrier that arises from a molecular mechanism distinct from poor TMH insertion.

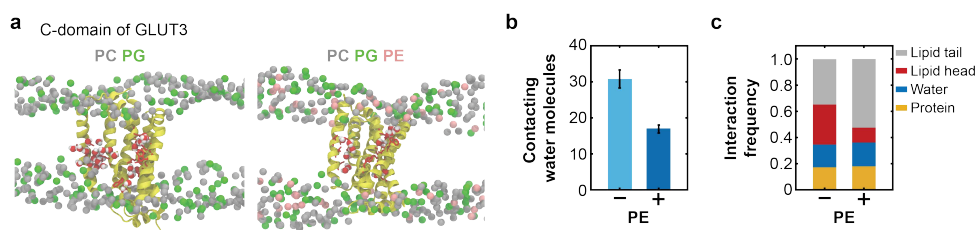


Figure 3.19 | Analysis of MD simulation for C-domain of GLUT3 in lipid bilayer. a, Snapshots from MD simulations for isolated GLUT3 C-domains in mixed bilayers of 70 mol% DMPC and 30 mol% DMPG (left) and 55 mol% DMPC, 30 mol% DMPG and 15 mol% DMPE (right) at 296.15 K. DMPC, DMPG and DMPE lipid head groups are depicted as gray, green and pink spheres, respectively, and water molecules are shown as composites of red

and white spheres. **b**, The average number of water molecules in contact with residues at the domain interface. Error bars represent SD ($n = 4000$ for each case). **c**, Interaction profiles of interface-exposed residues (N315, T316, T319, E378, W386, N413).

To gain insights into this late stage of folding, molecular dynamics (MD) simulations were employed in collaboration with Dr Soohyung Park in Prof. Wonpil Im group. GLUT3 was embedded in lipid bilayers with different lipid compositions and simulated for 1.0 μ s using the CHARMM force field (30). The MD simulation results suggest that the high content of polar/charged residues on the interface between N- and C-domains induce considerable distortions in the surrounding bilayer structure as well as increased penetration of water molecules (Figure 3.19 and Supplementary figure 9). 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 (79).

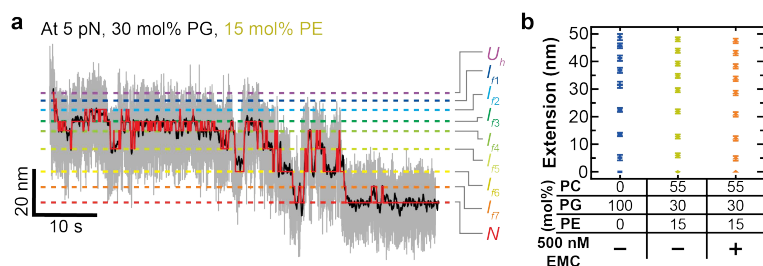


Figure 3.20 | Complete folding of GLUT3 with the aid of PE lipids. **a**, Representative time-resolved traces for folding of single GLUT3 proteins at 5 pN with 30 mol% DMPG and 15 mol% DMPE in the bicelles. **b**, Positions of the folding intermediates identified by HMM for denoted folding conditions. ($n = 11$, 11 and 10 traces for the cases with 100 mol% PG (blue), 30 mol% PG and 15 mol% PE without (yellow) and with (orange) EMCs, respectively). Error bars represent SEM.

We further reasoned that if the membrane shells between the N- and C-domains indeed define a major barrier to domain-domain assembly, the lipid bilayer composition might play a pivotal role in the final step of GLUT3 folding (80). Because negatively charged lipids were not effective for this purpose (Figure 3.4b),

DMPE lipids were added 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 (Figure 3.20a and Supplementary figure 10a). The HMM and BIC analyses applied to individual folding traces indicated that the intermediate structure was largely preserved upon addition of DMPE lipids (Figure 3.20b and Supplementary figure 10b). In MD simulation results, the frequencies by which the polar/charged residues contacted either water or lipid headgroup markedly decreased with the inclusion of PE lipids (Figure 3.19 and Supplementary figure 9). Together, these observations corroborated the 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.

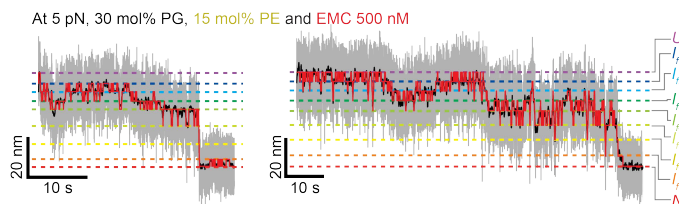


Figure 3.21 | Representative folding traces of GLUT3 both with EMC and PE lipids. Experiment was done at 5 pN with 30 mol% DMPG and 15 mol% DMPE in the bicelles in the presence of 500 nM EMC. Two replicates are shown.

When examining individual extension traces obtained under 5 pN, only a small proportion of the traces successfully consolidated the *N* state. Thus, EMC was added to see whether there was synergy between the effects of DMPE lipids and EMC (Figure 3.21 and Supplementary figure 10c,d).

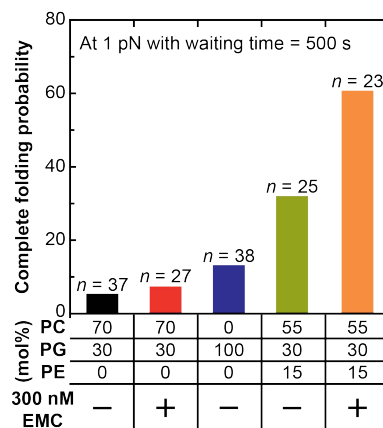


Figure 3.22 | Probability for observing the complete folding events under indicated conditions. *n* is the number of trials. See Supplementary figure 11 in detail for mechanical cycle to check the complete folding probability of GLUT3.

To ensure a higher statistical confidence with a fewer number of trials, a lower tension of 1 pN was used to induce GLUT3 folding (Supplementary figure 11). After waiting for 500 s, the force was increased to 25 pN to determine whether there was complete folding to the native state (Supplementary figure 11). Under the DMPG-30 mol% condition, folding probability was as low as 5.4 %, re-confirming that GLUT3 is not competent for folding by itself (Figure 3.22). Either addition of EMC alone (7.6 %) or switching to DMPG 100 mol% (13.2 %) marginally increased the folding probability, consistent with the observations that these conditions facilitated C-terminal domain folding, but not the domain-domain assembly (Figure 3.22). Addition of DMPE lipids increased the folding probability to ~30 %, and addition of both DMPE and EMC further increased the probability to 60 % (Figure 3.22). These data point to a strong synergy between DMPE lipids and EMC, completing the GLUT3 folding under physiologically relevant conditions in single-molecule magnetic tweezers assay.

3.6. Asymmetric TMH distributions of metazoan sugar transporters

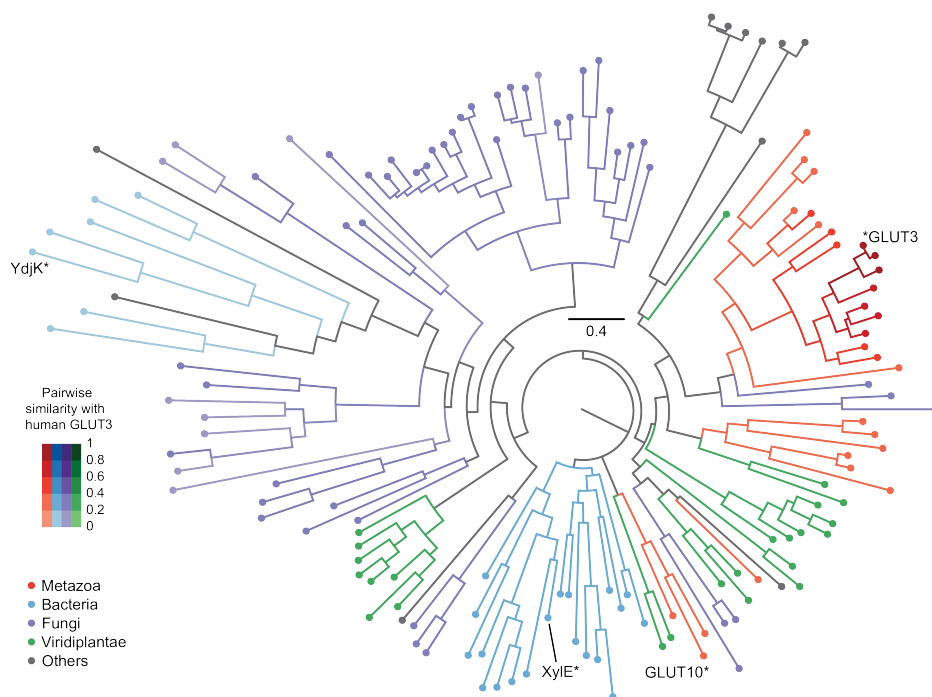


Figure 3.23 | Phylogenetic tree of the MFS sugar porter family. Multiple sequence alignment of the sugar porter family was produced in hhsuite (52). Color code indicates sequence similarity of each protein to human GLUT3 and is applied to branches and nodes in the phylogenetic tree.

Finally, I examined whether the observations made for human GLUT3 hold for other sugar transporters that exist across all domains of life and share a conserved structure of MFS fold (1, 81). In collaboration with Dr. Charlotte Tumescheit in Martin Steinegger group, I investigated the TMH-insertion energy values for 143 transporters in the sugar porter family (Figure 3.23).

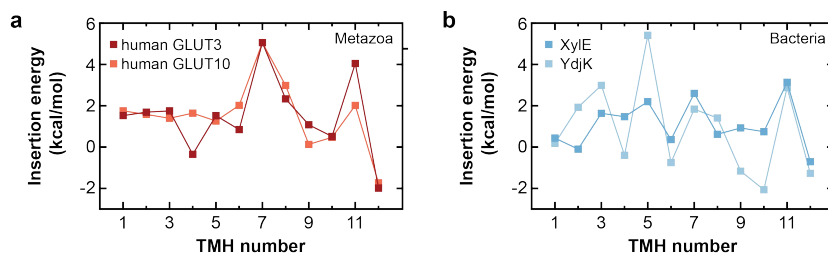


Figure 3.24 | Representative plots for predicted TMH-insertion energy. The values are calculated for human GLUT3 and GLUT10 (b), XylE and YdjK (c). The energy was calculated DGPred with 19-amino acid window.

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 (52). Then, the insertion energy values of putative TMHs were calculated using the biological hydrophobicity scale (Figure 3.24) (49).

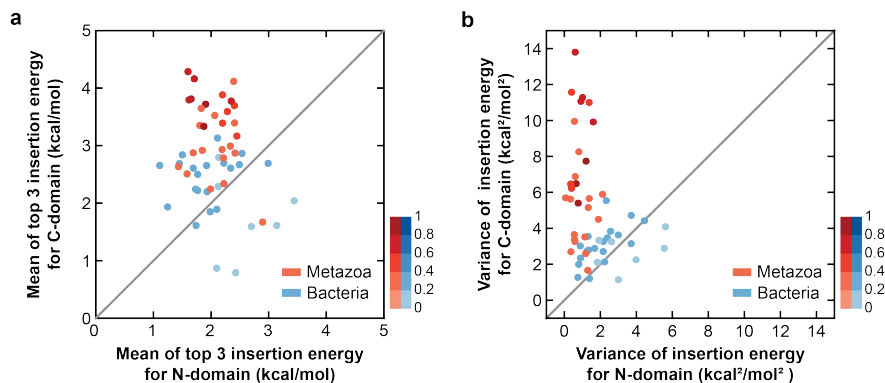


Figure 3.25 | Scatter plots showing the calculated TMH-insertion energy. The mean of three highest TMH-insertion energy values for (a) and the variance of insertion energy values for (b). Plots are shown by the N- (x-axis) and C-domain (y-axis) of each metazoan or bacterial transporter in the sugar porter family ($n = 28$ and 26 for metazoa and bacteria, respectively).

Remarkably, comparing the average of the three highest insertion energy values (out of six), the metazoan sugar transporters exhibited a marked asymmetry in which TMHs of the C-domain had higher insertion energy values than those of the

N-domain, a pattern that did not hold true for those bacterial transporters (Figure 3.25a). In addition, the C-domains (but not the N-domains) of metazoan sugar transporters showed far larger variances among their six constituting TMHs (Figure 3.25b and Supplementary figure 12a-d), reminiscent of finding that TMH 7 and TMH 11 in GLUT3's C-domain have particularly high insertion energies. This asymmetric pattern vanished again for the bacterial sugar porters (Figure 3.24b and Figure 3.25). TMH-insertion energy distributions for other clades showed a similar level of asymmetry between the N- and C-domains for plant sugar transporters, but not for fungal proteins (Supplementary figure 12e,f).

		TMH 7	
H. sapiens	GLUT3	RQPIIIISIVLQLSQQLSGI-----	285
H. sapiens	GLUT14	-----VLQLSQQLSGINAVFYST-	317
R. norvegicus	GLUT3	-----VLQLSQQFSGINAVFYST-	293
X. laevis	GLUT1	---IFIAIVLQLSQQLSGINAV-----	289
H. sapiens	GLUT1	-----VLQLSQQLSGINAVFYST-	295
D. rerio	GLUT3A	---LFIAIMLQLSQQLSGINAV-----	290
S. aurata	GLUT1	---LLIALVLQLSQQLSGINAI-----	289
M. amblycephala	GLUT4	-----LLQLSQQLSGVNAIFYYST-	300
H. sapiens	GLUT4	-----VLQLSQQLSGINAVFYST-	311
H. sapiens	GLUT2	-----ALMLHVAQQFSGINGIFY--	325
M. amblycephala	GLUT2	-----LMMHFSQQFSGINAIFYYS--	304
R. norvegicus	GLUT5	-----IVLMAGQQLSGVNAIYYA--	299
H. sapiens	GLUT5	---LLSIIVLMGGQQLSGVNAI-----	296
H. sapiens	GLUT7	---LLSIIVLMAGQQLSGINAI-----	302
M. musculus	GLUT9	-----IITMASYQLCGLNAIWFT--	320
H. sapiens	GLUT9	-----IVTMACYQLCGLNAIWFT--	339
H. sapiens	GLUT11	---VTSLVVLGSAMELCGNDSV-----	292
S. mansoni	GTP1	--VLIACLIVQLSQQLSGINAV-----	288

Figure 3.26 | Sequence alignment of TMH 7 for a subset of metazoan sugar transporters. The color scale of the name for aligned transporters is same for pairwise similarity in Figure 3.23.

Given that all transporters in the sugar porter family are assumed to have sugar binding sites in the C-domain (82), I wondered whether sampling of these more hydrophilic TMHs in the C-domains could be coupled to enhancement of transporter function. Indeed, QLS motif in TMH 7 plays a crucial role in improving the selectivity of sugar binding (83) and it is identified only for a subset of metazoan sugar transporters that are most close to GLUT3 in the phylogenetic analysis (Figure 3.26). The presence of the QLS motif increases the insertion energy of TMH 7 by ~1 kcal/mol, strongly coupled to higher TMH-insertion energy values of the entire C-

domains (Figure 3.27).

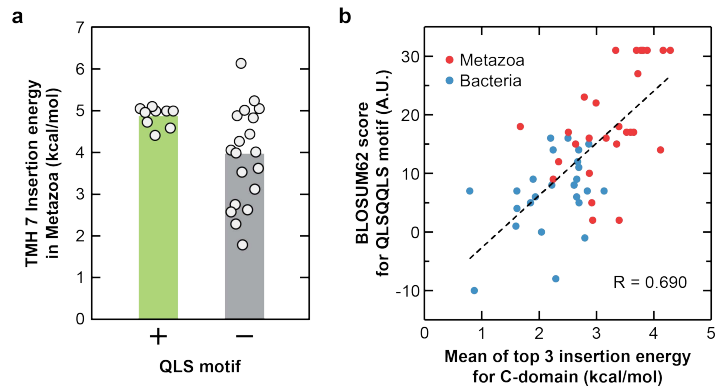


Figure 3.27 | Analysis for insertion energy of TMH 7. **a**, Mean insertion energy values for TMH 7 calculated for all metazoan sugar transporters. The energy was calculated as in Figure 3.24 ($n = 9$ and 19 for metazoan sugar transporters with and without QLS motif, respectively). **b**, Scatter plot showing the mean of the three highest insertion energy values (x-axis) and the BLOSUM62 score of QLS motif (y-axis) for each sugar porter. R means Pearson correlation coefficient. ($n = 28$ and 26 for metazoa and bacteria, respectively).

Chapter 4. Conclusion

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, single-molecule data dissect orders for weaving individual MFS folds. Given the high level of structural conservation (21, 82), I expect the folding order described here be shared by many transporters belonging to the MFS. It is further noted 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 (21, 84). The results may represent an example of the general principle that the folding pathways of membrane proteins have evolved to be commensurate with their symmetry properties, a natural requirement to build such structures of high symmetry.

On the domain level, folding of the N-domain strictly precedes that of the C-domain, which likely mirrors or leverages a co-translational folding pathway in cells (28, 85). This suggests that the N-domain likely serves as a structural template for C-domain folding, but 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 (81) – might have formed homo- or hetero-dimers, in which both foldability and functionality would be managed within a single subunit (86). 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.

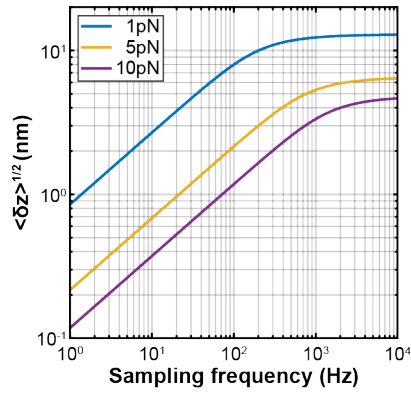
The bioinformatics analysis suggests that the metazoan sugar transporters have most proactively taken this evolutionary venture through sequence space to

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 constants (K_M) and catalytic rates (k_{cat}). For example, while GLUT3 is mainly expressed in neurons and transports glucose molecules with a high turnover rate ($k_{cat} > 1,000 \text{ s}^{-1}$) (87), GLUT2 is expressed in beta cells and mainly works as a glucose sensor with its uniquely high K_M (88). Finally, it is known that different families in MFS have different domain structures. Thus, it is an open question whether the findings – the N-to-C hierarchical folding pathway and the evolutionary development of insertion-energy asymmetry – are generally observed beyond the sugar porter family.

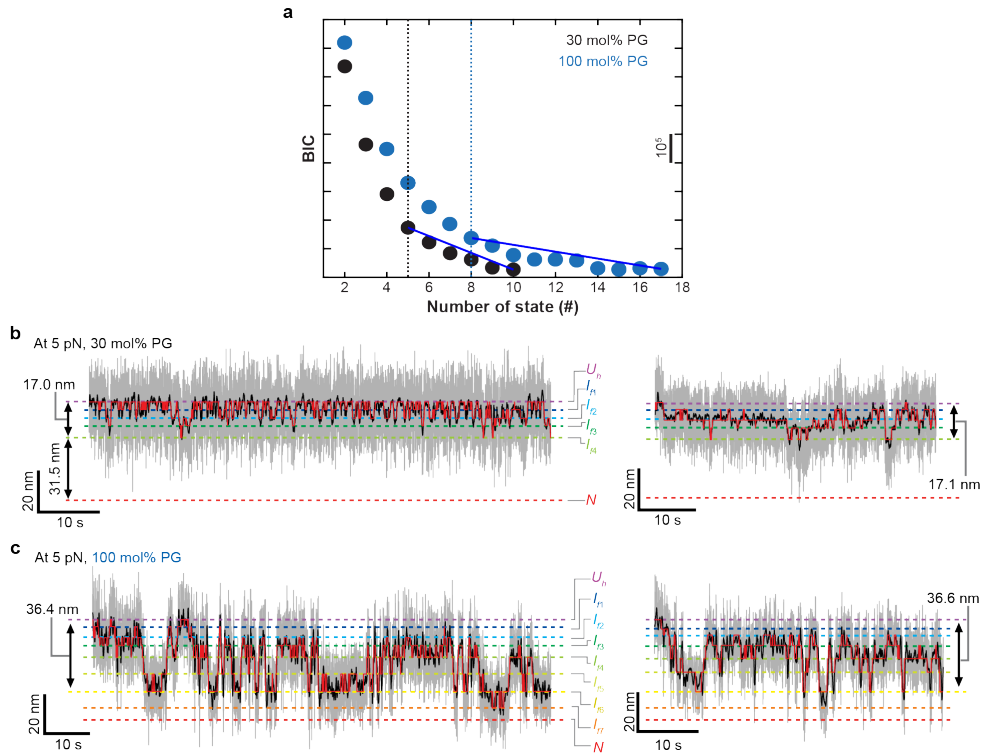
Thus, all the data collectively point to evolutionary conflicts between functionality and foldability faced by many of the metazoan sugar transporters. The resulting evolutionary pressure might have driven the ER membranes of these metazoan cells to be equipped with accessory machineries (e.g. EMC) and distinct lipid compositions that work in concert to help such poor-folding multi-pass membrane proteins. Recent studies suggest that EMC, along with YidC (89), GET1 (90) and TMCO1 (91), belong to the Oxa1 superfamily that makes a remarkably conserved family of insertases (92). Indeed, I found that even with the bicelle membranes that have lower energy barriers for protein insertion than true lipid bilayers, most TMHs of GLUT3 still need to be assisted by EMC for their efficient membrane insertion, corroborating the notion that the membrane insertion steps do present considerable energy barriers during folding of these transporter proteins. While the PE-headgroup lipids are known to affect TMH orientations and thus establishment of a right topology of TMHs (93), the results suggest a novel role of the PE lipids – and presumably other lipid species with specific geometric curvatures

– during a later stage of membrane protein folding. The presence of PE lipids facilitates removal of lipid shells from the domain-domain (or subunit-subunit) interfaces and assembly of higher-order membrane protein structures. This observation is intriguing because it provides a glimpse into how two biogenesis processes in the ER membrane – membrane protein biogenesis and lipid synthesis – are intricately intertwined with one another (94, 95).

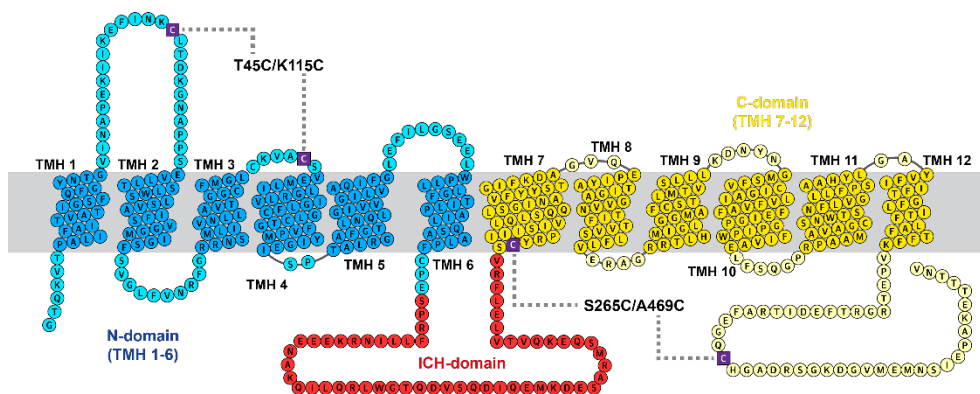
Supplementary figure



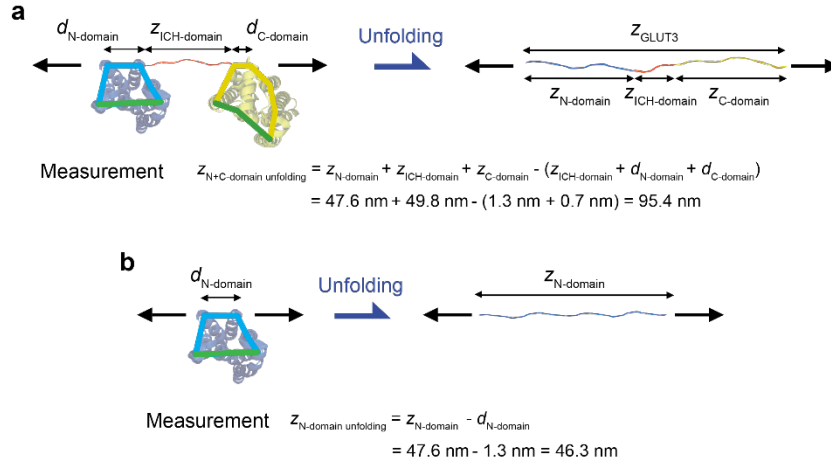
Supplementary figure 1 | Precision in determination of the vertical position of a bead as a function of the measurement bandwidth. The plot indicates an ~1 nm resolution when bead positions are averaged over 50 ms (~20 Hz sampling). In magnetic tweezer experiments, the bicelle phase used for providing 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 membrane protein folding studies.



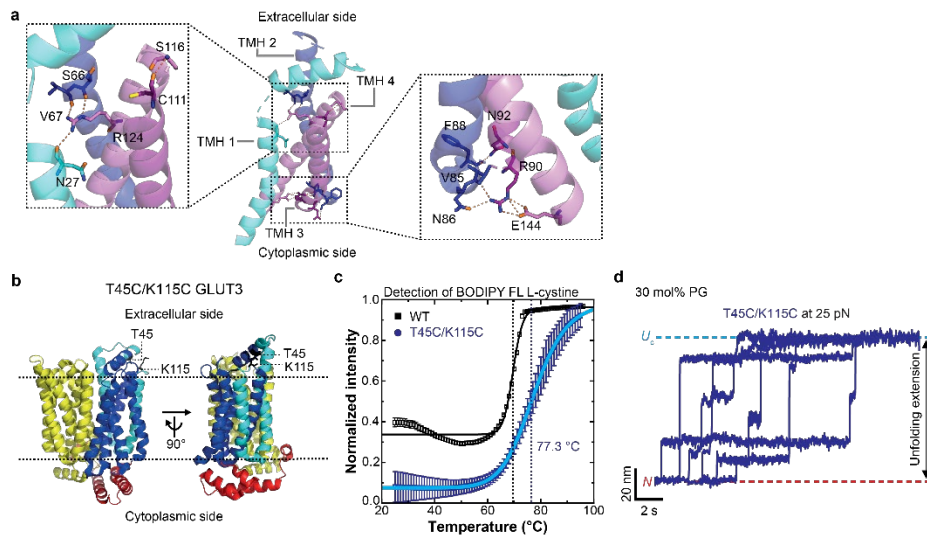
Supplementary figure 2 | Folding characteristic of GLUT3 with 30 mol% PG and 100 mol% PG in bicelle. **a**, Bayesian Information Criteria (BIC) values of WT GLUT3 for each number of states with different bicelle compositions ($n = 16$ and 11 for 30 mol% and 100 mol% PG, respectively). **b**, **c**, Representative folding traces for WT GLUT3 with 30 mol% PG and 70 mol% PC (**b**) and 100 mol% PG (**c**) at 5 pN. Two replicates are shown for each condition, and the gray and black traces are 1.2-kHz raw data and 5-Hz median-filtered data, respectively. Red traces indicate the transitions between intermediates identified by HMM.



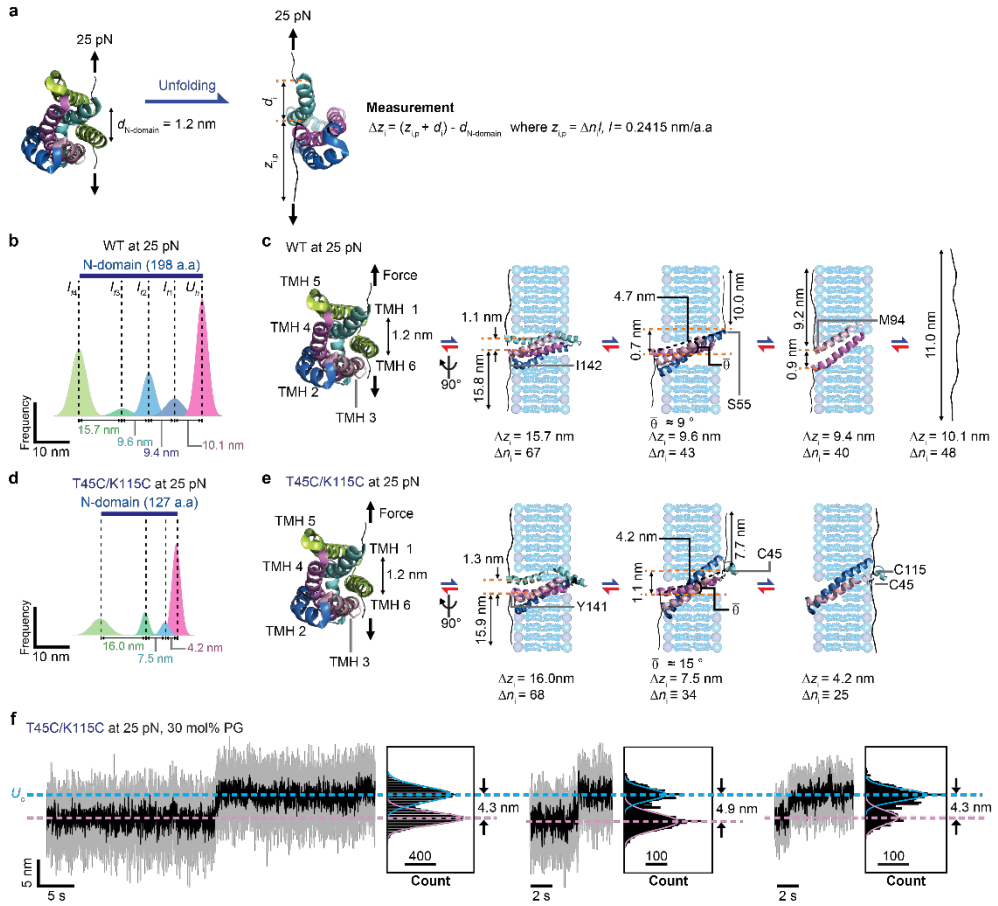
Supplementary figure 3 | Topological model of GLUT3. Each domain is distinguished by different colors. Two pairs of cysteine mutations are indicated.



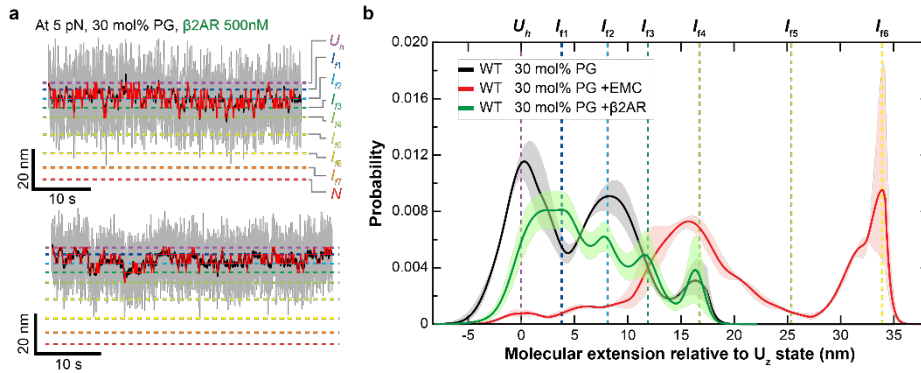
Supplementary figure 4 | Pulling geometry of a single GLUT3. **a**, Schematic of pulling geometry of a single GLUT3 for two cases, where only ICH-domain is unfolded or GLUT3 is fully unfolded at 25 pN (blue for N-domain, yellow for C-domain, and red for ICH-domain). $z_{N+C\text{-domain unfolding}}$ indicates the total extension of a single GLUT3 under 25 pN tension. $d_{N\text{-domain}}$ and $d_{C\text{-domain}}$ are measured by protein structure (PDB: 4ZWC). $z_{N\text{-domain}}$ and $z_{C\text{-domain}}$ are extension values of the unfolded N- and C-domains calculated by eWLC model. The structure is viewed from the cytoplasmic side. Blue (N-domain) and yellow (C-domain) lines indicate the surface of the proteins surrounded by lipid membrane. Green line represents the interface between N- and C-domains. **b**, Schematic of pulling geometry of a single GLUT3 N-domain before and after unfolding at 25pN. $z_{N\text{-domain unfolding}}$ is an expected extension increase for unfolding of GLUT3's N-domain.



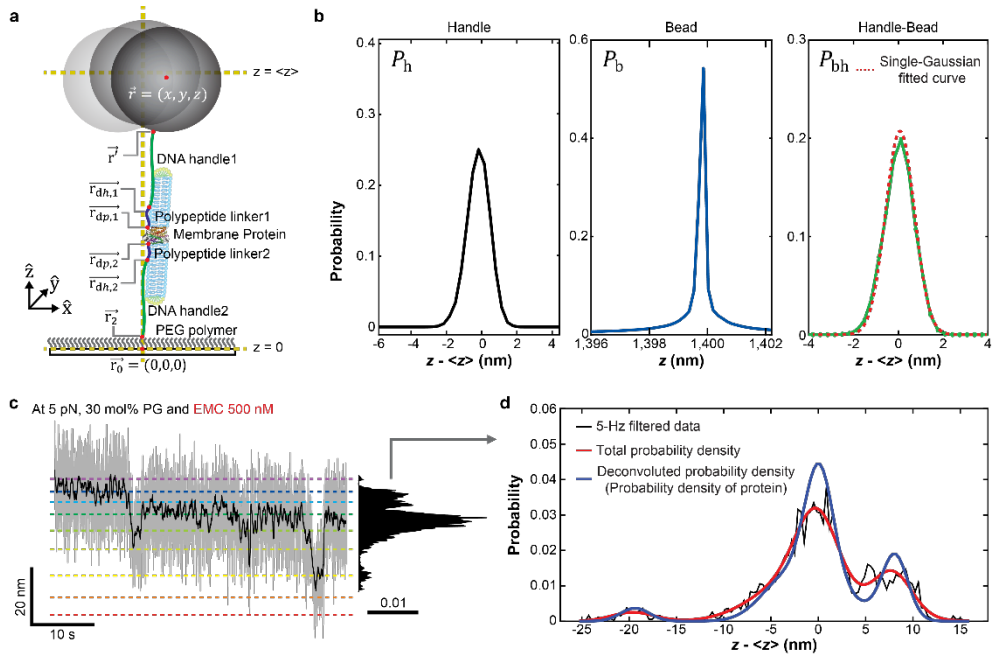
Supplementary figure 5 | Sample preparation and unfolding characteristics of T45C/K115C GLUT3. **a**, Atomic contacts among TMHs 1, 2, and 4. Inset shows detailed position of interacting residues (blue for amino group, orange for carboxyl group, and yellow for thiol group). **b**, The positions of two mutations, T45C/K115C in GLUT3 (GLUT3_{TM23C}). **c**, An absorbance profile of BODIPY FL fluorophore-labeled GLUT3_{TM23C} as temperature increases. Error bars represent SEM ($n = 4$). **d**, Collection of 50Hz-median filtered unfolding traces initiated from *N* state for GLUT3_{TM23C}.



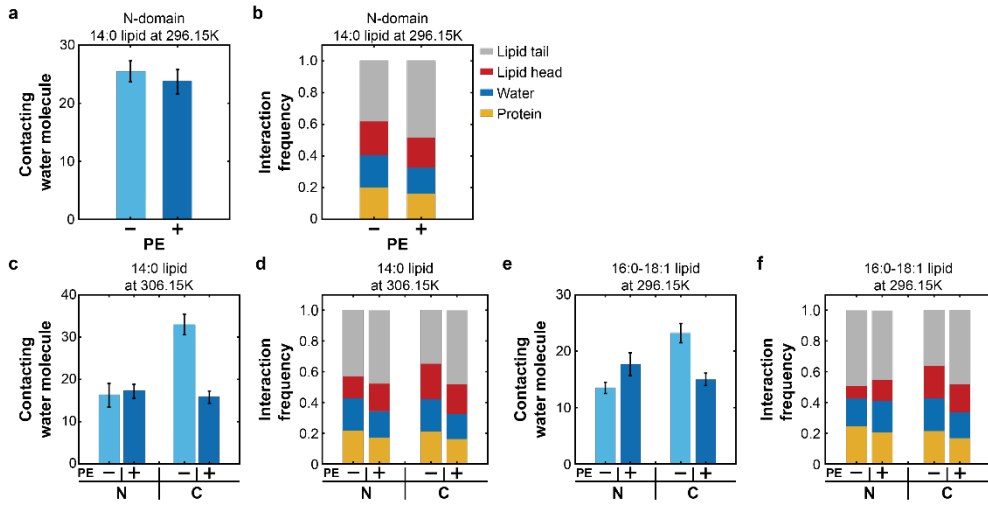
Supplementary figure 6 | Determination of folding order for N-domain of GLUT3. a, Schematic of pulling geometry for N-domain of GLUT3 at 25 pN. $d_{N\text{-domain}}$ is the distance between two points of force application before unfolding (PDB: 4ZWC). Δz_i indicates the expected extension increase for GLUT3 for the i th intermediate. $z_{i,p}$ is the extension of the unfolded portion along the membrane for the i th intermediate. d_i denotes the distance between the points of force application for the i th intermediate. 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 $\bar{\theta}$ with a black dashed line to the unfolded portion of N-domain in the membrane. **d,** Unfolding extension distribution for the N-domain part of GLUT3_{TM23C} at 25pN. **e,** Structural information and folding/unfolding order of the N-domain of GLUT3_{TM23C}. The description is the same as (c) except for protein construct. **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.



Supplementary figure 7 | folding characteristics of GLUT3 with unrelated membrane protein. **a**, Representative folding traces for single GLUT3 at 5 pN with 30 mol% PG in the bicelles in the presence of 500 nM $\beta 2AR$. Two replicates are shown. **b**, Probability distributions of deconvoluted extension values observed under indicated folding conditions at 5pN ($n = 11$ for the reaction with $\beta 2AR$). The shaded area means SEM.

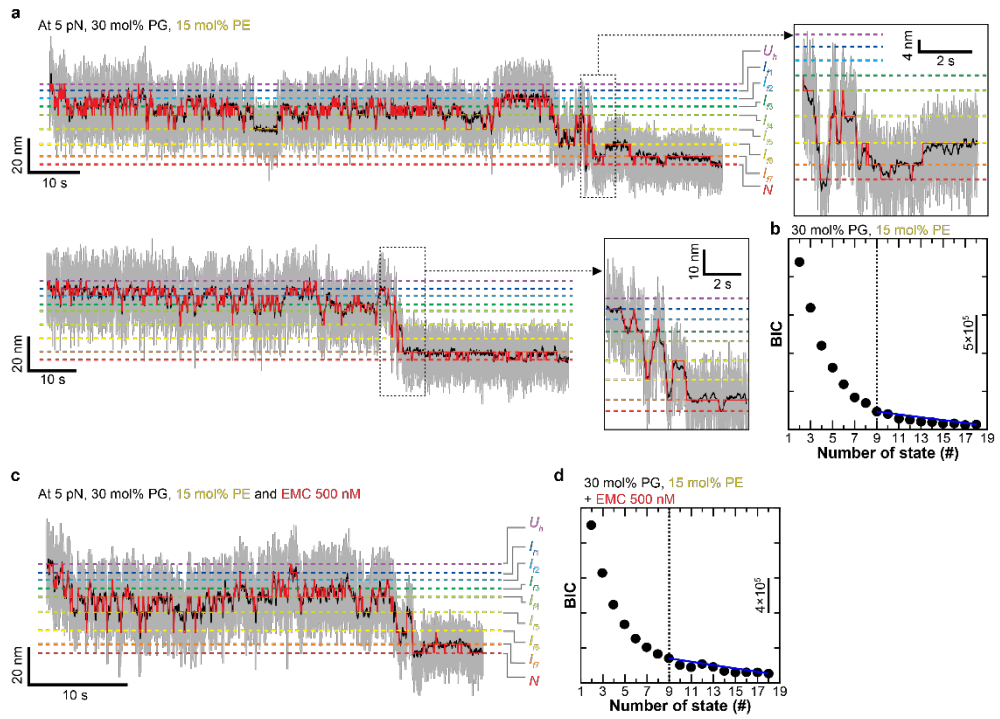


Supplementary figure 8 | Deconvoluted extension probability analysis. **a**, A cartoon depicting components of MT experiment in three-dimensional space for deconvoluted extension probability analysis. The positions of the components are marked with red dots. **b**, First two graphs are probability distributions of handle and bead. Third graph depicts a composed probability distribution of handle and bead with suitable fitting function. **c**, Representative folding trace of GLUT3 with folding distribution. **d**, Set of probability distribution of data in (c). Red graph is the fitting function of 5-Hz filtered data (black graph) and blue graph is the probability distribution generated by deconvoluted extension probability analysis.

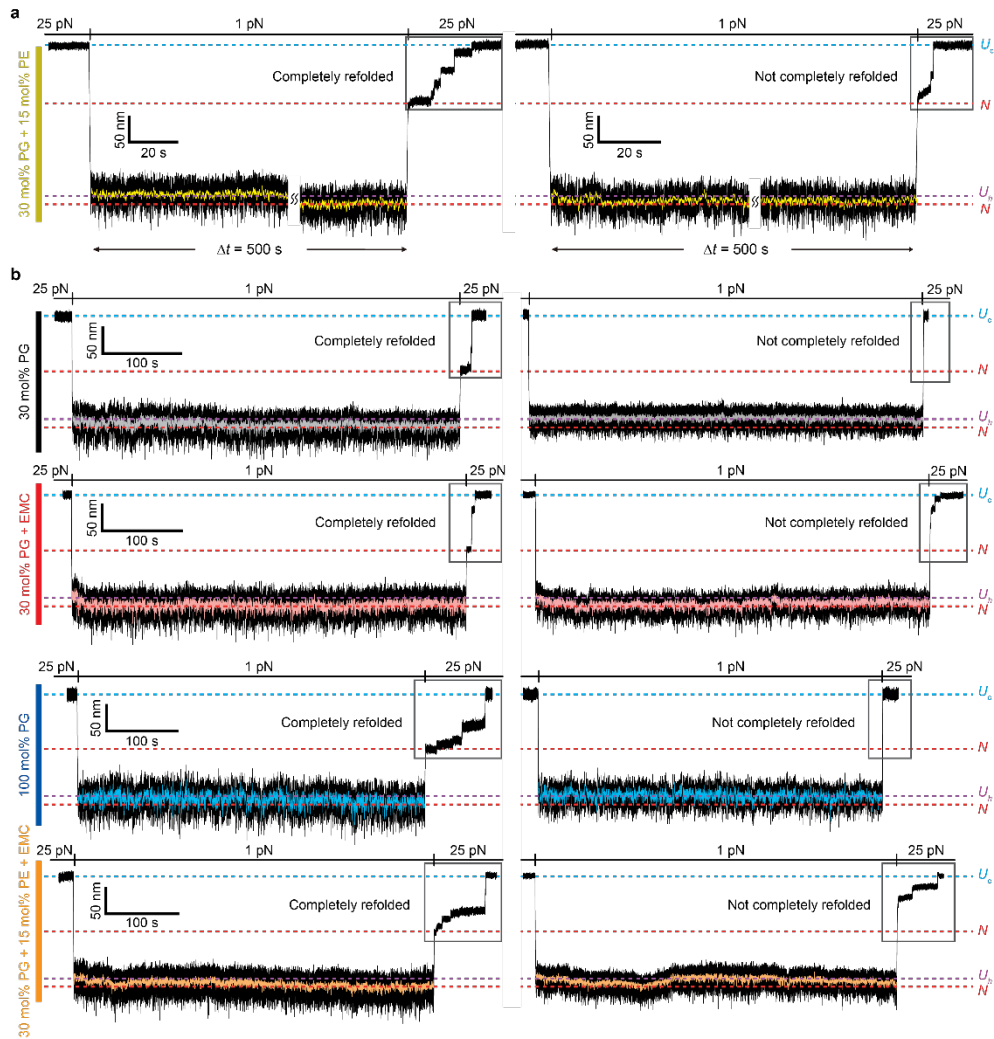


Supplementary figure 9 | Analysis of MD simulation for GLUT3 in various lipid bilayers.

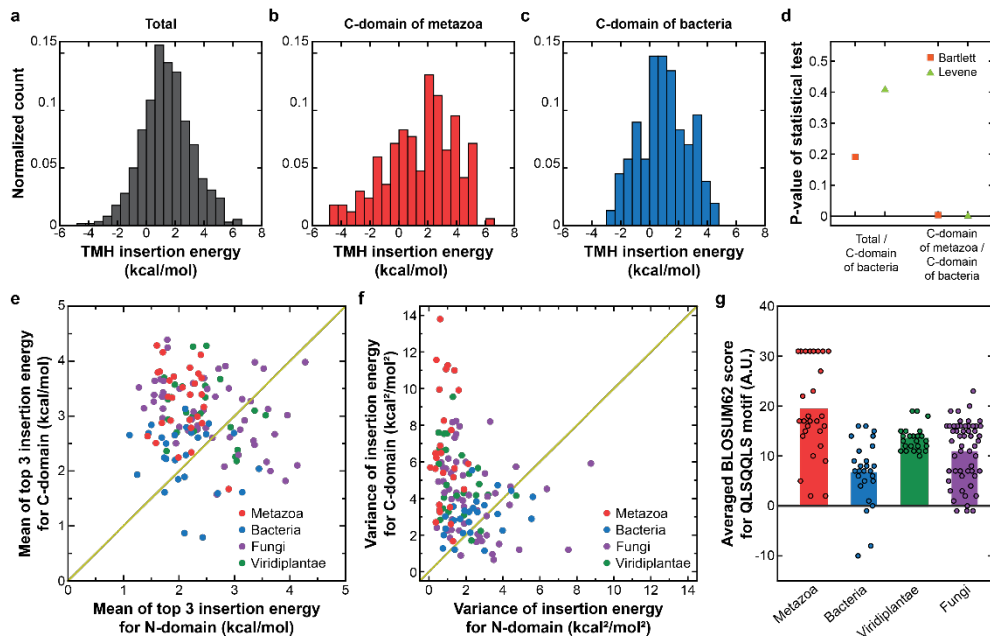
a, The average number of contacting water molecule to polar/charged residues in TMHs of N-domain with or without DMPE. Error bars represent SD ($n = 4000$ for each case). Polar/charged residues in TMHs of GLUT3 for the analysis are as follows. S21, Q23, N27, T28, S71, S78, S80, N98, R124, T135, T156, 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 $0.6\mu\text{s}$ to $1.0\mu\text{s}$. **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 $0.6\mu\text{s}$ to $1.0\mu\text{s}$. **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 $0.3\mu\text{s}$ to $0.5\mu\text{s}$.



Supplementary figure 10 | 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 Supplementary figure 2. Two replicates are shown. **b**, BIC values for the indicated number of states with 15 mol% PE bicelle ($n = 11$). **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 ($n = 10$).



Supplementary figure 11 | Force-jump experiment for determination of complete folding probability. **a**, Representative traces for determination of the complete folding probability using 15 mol% PE bicelle for higher statistical confidence. Force was increased to 25 pN after a fixed period of time (500s) at 1 pN. Then, high force of 25 pN was applied to determine whether there was complete folding to the native state under the respective conditions. (insets). **b**, Traces from the force-jump experiment using four different conditions to determine the refolding probability.



Supplementary figure 12 | Analysis for 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).

Bibliography

1. M. A. Hediger *et al.*, The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins. *Pflügers Archiv* **447**, 465-468 (2004).
2. C. A. Orengo, A. E. Todd, J. M. Thornton, From protein structure to function. *Current opinion in structural biology* **9**, 374-382 (1999).
3. E. M. Quistgaard, C. Löw, F. Guettou, P. Nordlund, Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nature Reviews Molecular Cell Biology* **17**, 123 (2016).
4. K. P. Locher, Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nature structural & molecular biology* **23**, 487 (2016).
5. F. Cymer, G. Von Heijne, S. H. White, Mechanisms of integral membrane protein insertion and folding. *Journal of molecular biology* **427**, 999-1022 (2015).
6. K. Corin, J. U. Bowie, How physical forces drive the process of helical membrane protein folding. *EMBO reports* **23**, e53025 (2022).
7. D. Marsh, Lateral pressure in membranes. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **1286**, 183-223 (1996).
8. J. U. Bowie, Membrane protein folding: how important are hydrogen bonds? *Current opinion in structural biology* **21**, 42-49 (2011).
9. R. Guo *et al.*, Structural cavities are critical to balancing stability and activity of a membrane-integral enzyme. *Proceedings of the National Academy of Sciences* **117**, 22146-22156 (2020).
10. Y. Cheng, Membrane protein structural biology in the era of single particle cryo-EM. *Current opinion in structural biology* **52**, 58-63 (2018).
11. J. T. Marinko *et al.*, Folding and misfolding of human membrane proteins in health and disease: from single molecules to cellular proteostasis. *Chemical reviews* **119**, 5537-5606 (2019).
12. A. Guna, R. S. Hegde, Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. *Current Biology* **28**, R498-R511 (2018).
13. P. J. Chitwood, R. S. Hegde, An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature* **584**, 630-634 (2020).
14. L. Ellgaard, A. Helenius, Quality control in the endoplasmic reticulum. *Nature reviews Molecular cell biology* **4**, 181-191 (2003).
15. R. R. Kopito, Biosynthesis and degradation of CFTR. *Physiological reviews* **79**, S167-S173 (1999).

16. J. P. Schleich *et al.*, Conformational stability and pathogenic misfolding of the integral membrane protein PMP22. *Journal of the American Chemical Society* **137**, 8758-8768 (2015).
17. S. J. Bose *et al.*, Towards next generation therapies for cystic fibrosis: Folding, function and pharmacology of CFTR. *Journal of Cystic Fibrosis* **19**, S25-S32 (2020).
18. W. E. Balch, R. I. Morimoto, A. Dillin, J. W. Kelly, Adapting proteostasis for disease intervention. *science* **319**, 916-919 (2008).
19. A. Oberai, Y. Ihm, S. Kim, J. U. Bowie, A limited universe of membrane protein families and folds. *Protein science* **15**, 1723-1734 (2006).
20. M. S. Almén, K. J. V. Nordström, R. Fredriksson, H. B. Schiöth, Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol* **7**, 50-50 (2009).
21. Y. Shi, Common folds and transport mechanisms of secondary active transporters. *Annual review of biophysics* **42**, 51-72 (2013).
22. A. Krogh, B. Larsson, G. Von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* **305**, 567-580 (2001).
23. M. G. Madej, Function, structure, and evolution of the Major Facilitator Superfamily: the LacY manifesto. *Advances in Biology* **2014**, (2014).
24. D. Drew, O. Boudker, Shared molecular mechanisms of membrane transporters. *Annual review of biochemistry* **85**, 543-572 (2016).
25. D. Deng *et al.*, Molecular basis of ligand recognition and transport by glucose transporters. *Nature* **526**, 391-396 (2015).
26. J. Lipfert, X. Hao, N. H. Dekker, Quantitative modeling and optimization of magnetic tweezers. *Biophysical journal* **96**, 5040-5049 (2009).
27. K. C. Neuman, A. Nagy, Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature methods* **5**, 491-505 (2008).
28. H.-K. Choi *et al.*, Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science* **366**, 1150-1156 (2019).
29. S. Jo, T. Kim, W. Im, Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLOS ONE* **2**, e880 (2007).
30. S. Jo, J. B. Lim, J. B. Klauda, W. Im, CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophysical journal* **97**, 50-58 (2009).
31. E. L. Wu *et al.*, CHARMM-GUI Membrane Builder toward realistic

- biological membrane simulations. *Journal of Computational Chemistry* **35**, 1997-2004 (2014).
32. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859-1865 (2008).
 33. P. Eastman *et al.*, OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **13**, e1005659 (2017).
 34. R. M. Venable *et al.*, CHARMM all-atom additive force field for sphingomyelin: elucidation of hydrogen bonding and of positive curvature. *Biophysical journal* **107**, 134-145 (2014).
 35. J. B. Klauda *et al.*, Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* **114**, 7830-7843 (2010).
 36. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **79**, 926-935 (1983).
 37. A. D. MacKerell *et al.*, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **102**, 3586-3616 (1998).
 38. J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341 (1977).
 39. K. A. Feenstra, B. Hess, H. J. C. Berendsen, Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry* **20**, 786-798 (1999).
 40. C. W. Hopkins, S. Le Grand, R. C. Walker, A. E. Roitberg, Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **11**, 1864-1874 (2015).
 41. Y. Gao *et al.*, CHARMM-GUI Supports Hydrogen Mass Repartitioning and Different Protonation States of Phosphates in Lipopolysaccharides. *Journal of Chemical Information and Modeling* **61**, 831-839 (2021).
 42. P. J. Steinbach, B. R. Brooks, New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal of Computational Chemistry* **15**, 667-683 (1994).
 43. U. Essmann *et al.*, A smooth particle mesh Ewald method. *The Journal of chemical physics* **103**, 8577-8593 (1995).
 44. K.-H. Chow, D. M. Ferguson, Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Computer Physics*

Communications **91**, 283-289 (1995).

45. J. Åqvist, P. Wennerström, M. Nervall, S. Bjelic, B. O. Brandsdal, Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters* **384**, 288-294 (2004).
46. J. Lee *et al.*, CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation* **12**, 405-413 (2016).
47. B. R. Brooks *et al.*, CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **30**, 1545-1614 (2009).
48. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33-38 (1996).
49. T. Hessa *et al.*, Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-1030 (2007).
50. M. H. Saier, Jr, C. V. Tran, R. D. Barabote, TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Research* **34**, D181-D186 (2006).
51. M. Mirdita *et al.*, Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Research* **45**, D170-D176 (2016).
52. M. Steinegger *et al.*, HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).
53. M. Remmert, A. Biegert, A. Hauser, J. Söding, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods* **9**, 173-175 (2012).
54. F. Gabler *et al.*, Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics* **72**, e108 (2020).
55. C. Bouchiat *et al.*, Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements. *Biophysical Journal* **76**, 409-413 (1999).
56. F. Oesterhelt *et al.*, Unfolding pathways of individual bacteriorhodopsins. *Science* **288**, 143-146 (2000).
57. Y. Seol, J. Li, P. C. Nelson, T. T. Perkins, M. Betterton, Elasticity of short DNA molecules: theory and experiment for contour lengths of 0.6–7 μm . *Biophysical journal* **93**, 4360-4373 (2007).
58. A. Sarkar, S. Caamano, J. M. Fernandez, The mechanical fingerprint of a parallel polyprotein dimer. *Biophysical journal* **92**, L36-L38 (2007).
59. M. J. Shon, S.-H. Rah, T.-Y. Yoon, Submicrometer elasticity of double-stranded DNA revealed by precision force-extension measurements with

magnetic tweezers. *Science advances* **5**, eaav1697 (2019).

60. J. C. M. Gebhardt, T. Bornschlöggl, M. Rief, Full distance-resolved folding energy landscape of one single protein molecule. *Proceedings of the National Academy of Sciences* **107**, 2013-2018 (2010).
61. D. Min, R. E. Jefferson, J. U. Bowie, T.-Y. Yoon, Mapping the energy landscape for second-stage folding of a single membrane protein. *Nature chemical biology* **11**, 981-987 (2015).
62. D. A. Kessler, Y. Rabin, Distribution functions for filaments under tension. *The Journal of chemical physics* **121**, 1155-1164 (2004).
63. Y. Zhang, J. Jiao, A. A. Rebane, Hidden Markov modeling with detailed balance and its application to single protein folding. *Biophysical journal* **111**, 2110-2124 (2016).
64. T.-H. Lee, Extracting kinetics information from single-molecule fluorescence resonance energy transfer data using hidden Markov models. *The Journal of Physical Chemistry B* **113**, 11535-11542 (2009).
65. M. Hinczewski, Y. von Hansen, R. R. Netz, Deconvolution of dynamic mechanical networks. *Proceedings of the National Academy of Sciences* **107**, 21493-21498 (2010).
66. M. Hinczewski, J. C. M. Gebhardt, M. Rief, D. Thirumalai, From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proceedings of the National Academy of Sciences* **110**, 4500-4505 (2013).
67. D. Min, M. A. Arbing, R. E. Jefferson, J. U. Bowie, A simple DNA handle attachment method for single molecule mechanical manipulation experiments. *Protein Science* **25**, 1535-1544 (2016).
68. H. E. Findlay, P. J. Booth, The biological significance of lipid-protein interactions. *Journal of Physics: Condensed Matter* **18**, S1281 (2006).
69. R. Ujwal, J. U. Bowie, Crystallizing membrane proteins using lipidic bicelles. *Methods* **55**, 337-341 (2011).
70. H. E. Findlay, P. J. Booth, The folding, stability and function of lactose permease differ in their dependence on bilayer lipid composition. *Scientific Reports* **7**, 13056 (2017).
71. C. Snider, S. Jayasinghe, K. Hristova, S. H. White, MPEx: A tool for exploring membrane proteins. *Protein Science* **18**, 2624-2628 (2009).
72. S. Radestock, L. R. Forrest, The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *Journal of molecular biology* **407**, 698-715 (2011).
73. M. C. Jonikas *et al.*, Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *science* **323**, 1693-1697 (2009).

74. T. Pleiner *et al.*, Structural basis for membrane insertion by the human ER membrane protein complex. *Science* **369**, 433-436 (2020).
75. P. J. Chitwood, S. Juskiewicz, A. Guna, S. Shao, R. S. Hegde, EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell* **175**, 1507-1519.e1516 (2018).
76. A. Guna, N. Volkmar, J. C. Christianson, R. S. Hegde, The ER membrane protein complex is a transmembrane domain insertase. *Science* **359**, 470-473 (2018).
77. A. Kauko *et al.*, Repositioning of transmembrane α -helices during membrane protein folding. *Journal of molecular biology* **397**, 190-201 (2010).
78. G. Lai, R. Renthal, Integral membrane protein fragment recombination after transfer from nanolipoprotein particles to bicelles. *Biochemistry* **52**, 9405-9412 (2013).
79. M. S. Cheung, A. E. García, J. N. Onuchic, Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proceedings of the National Academy of Sciences* **99**, 685-690 (2002).
80. J. U. Bowie, Solving the membrane protein folding problem. *Nature* **438**, 581-589 (2005).
81. S. S. Pao, I. T. Paulsen, M. H. Saier Jr, Major facilitator superfamily. *Microbiology and molecular biology reviews* **62**, 1 (1998).
82. N. Yan, Structural biology of the major facilitator superfamily transporters. *Annual review of biophysics* **44**, 257-283 (2015).
83. M. J. Seatter, S. A. De La Rue, L. M. Porter, G. W. Gould, 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).
84. L. R. Forrest, Structural symmetry in membrane proteins. *Annual review of biophysics* **44**, 311-337 (2015).
85. F. Cymer, G. von Heijne, Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proceedings of the National Academy of Sciences* **110**, 14640-14645 (2013).
86. J. S. Lolkema, A. Dobrowolski, D.-J. Slotboom, Evolution of antiparallel two-domain membrane proteins: tracing multiple gene duplication events in the DUF606 family. *Journal of molecular biology* **378**, 596-606 (2008).
87. A. R. Walmsley, M. P. Barrett, F. Bringaud, G. W. Gould, Sugar transporters from bacteria, parasites and mammals: structure-activity relationships. *Trends in biochemical sciences* **23**, 476-481 (1998).
88. B. Thorens, GLUT2, glucose sensing and glucose homeostasis.

Diabetologia **58**, 221-232 (2015).

89. T. Serdiuk *et al.*, YidC assists the stepwise and stochastic folding of membrane proteins. *Nature chemical biology* **12**, 911-917 (2016).
90. F. Wang, C. Chan, N. R. Weir, V. Denic, The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase. *Nature* **512**, 441-444 (2014).
91. P. T. McGilvray *et al.*, An ER translocon for multi-pass membrane protein biogenesis. *Elife* **9**, e56889 (2020).
92. R. S. Hegde, R. J. Keenan, The mechanisms of integral membrane protein biogenesis. *Nature Reviews Molecular Cell Biology*, 1-18 (2021).
93. W. Zhang, H. A. Campbell, S. C. King, W. Dowhan, Phospholipids as determinants of membrane protein topology: phosphatidylethanolamine is required for the proper topological organization of the γ -aminobutyric acid permease (GabP) of *Escherichia coli*. *Journal of Biological Chemistry* **280**, 26032-26038 (2005).
94. R. Volmer, D. Ron, Lipid-dependent regulation of the unfolded protein response. *Current opinion in cell biology* **33**, 67-73 (2015).
95. J. Jacquemyn, A. Cascalho, R. E. Goodchild, The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. *EMBO reports* **18**, 1905-1921 (2017).

Abstract in Korean (초록)

세포 내에서 일어나는 여러 현상은 막으로 구분되어 있으며, 이러한 막에 존재하는 막 단백질들은 물질을 수송하거나 정보를 전달하는 등 매우 중요한 역할을 담당한다. 이러한 막 단백질이 기능을 하기 위해서는 올바른 접힘 과정을 통해 구조가 형성되어야 한다. 그런데 완성된 구조에 대한 연구는 많이 진행되고 있지만 이러한 구조의 형성 과정에 대한 연구는 부족하다. 막 단백질의 구조 형성과정은 단백질이 인지질 이중막에서 전사되는 과정과 매우 밀접하게 관련이 있다. 이러한 세포내 환경을 모사하고 막 단백질의 폴립, 접힘 과정을 살펴보기 위해 자기 집게 기술 (single-molecule magnetic tweezers)을 이용하였다. 이 기술을 통해 막 단백질에 pN 단위의 미세한 힘을 가할 수 있다. 목표로 한 막 단백질은 복잡한 당 수송 단백질인 GLUT3로 해당 막 단백질의 접힘 과정을 최초로 규명하였다. 특히 소포체막 단백질 복합체(EMC)와 특이구조를 지닌 지질분자를 이용하여 생리화학적 환경에서 막 단백질의 구조 형성 과정을 완전히 밝혀냈다. 이를 통해 막 단백질이 다른 단백질, 그리고 인지질과 같은 주변 환경의 도움을 통해 기능할 수 있는 구조를 형성함을 알아냈다. 이러한 결과를 확장하고자 생물정보학을 이용하여 다양한 당 수송 단백질의 서열을 분석하였다. 당 수송 단백질은 6개의 transmembrane helix로 이루어진 domain 2개가 존재하는데 이러한 사실과 생물정보학 분석을 종합한 결과, 각각의 domain이 단백질의 구조형성 능력과 새로운 수송 기능을 담당하며 균형을 맞추어 진화해 왔음을 발견하였다.

주요어 : 단분자 힘 분광학, 자기 집게, 막단백질 접힘, 당 수송 단백질, 단백질 진화

학번 : 2017-21674

감사의 글

막연히 멀게만 느껴지던 학위 과정을 어느새 마무리하게 되었습니다. 학위 과정을 통해 배운 연구의 시작은 수많은 실패였습니다. 제 스스로가 미숙하고 연구라는 것이 아무도 모르는 새로운 사실을 밝혀내는 과정이기 때문에 실패가 많은 것은 당연하였습니다. 그러나 여기서 좌절하지 않고 그 실패 속에서 나아갈 점을 찾아 실패를 작은 성공으로 바꾸는 일을 하였습니다. 이러한 변화가 쌓여 능동적으로 연구를 할 수 있었고 지금의 저에 도달할 수 있었습니다. 하지만 이렇게 꾸준히 연구를 할 수 있었던 것은 혼자가 아니라 주변의 여러 사람들이 있었기 때문입니다. 연구자로서 그리고 심적으로도 부족한 것이 많은 저였지만 주변 사람들의 도움 덕에 길고 고된 학위 과정을 헤쳐 나갈 수 있었습니다.

우선 학위논문 심사를 진행해주시고 진로에 관해 많은 조언을 주신 심사위원분들께 감사의 말씀을 전합니다. 그리고 학위 과정동안 물심양면으로 아낌없이 지원해주시고 연구가 무엇인지 알려주신 지도교수님, 윤태영교수님께 진심으로 감사하다는 말씀을 드리고 싶습니다. 다음으로 제가 속해 있던 막단백질 단분자 연구실에서 같이 연구를 진행해온 친구들과 선후배, 그리고 박사님께 감사를 드리고자 합니다. 아무것도 모르던 대학원생 초기에 실험 방법에서부터 많은 것을 알려준 KAIST 선배 박사님들께 감사하다는 말씀을 드립니다. 그리고 비슷한 시기에 연구실에 들어와 동고동락한 창원이와 지승이에게도 정말로 고맙다는 말을 전합니다. 막히는 일이 있을 때 편하게 의견을 나눌 친구가 연구실에 있어 든든했습니다. 마지막으로 제 학위 과정이 잘 마무리될 수 있도록 많은 도움을 준 최현규 박사에게 감사의 말을 전합니다.