Title
Unfolding of a ClC chloride transporter retains memory of its evolutionary history.

Journal
Nature chemical biology, 14(5)

ISSN
1552-4450

Authors
Min, Duyoung
Jefferson, Robert E
Qi, Yifei
et al.

Publication Date
2018-05-01

DOI
10.1038/s41589-018-0025-4

Peer reviewed

Published in final edited form as:

Nat Chem Biol. 2018 May ; 14(5): 489–496. doi:10.1038/s41589-018-0025-4.

Unfolding of a ClC chloride transporter retains memory of its evolutionary history

Duyoung Min[1,2], **Robert E. Jefferson**[1,2], **Yifei Qi**[3], **Jing Yang Wang**[1,2], **Mark A. Arbing**[1,2], **Wonpil Im**[4], and **James U. Bowie**[1,2,*]

[1]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

[2]UCLA-DOE and Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA

[3]College of Chemistry and Molecular Engineering, East China Normal University, Shanghai, China

[4]Departments of Biological Sciences and Bioengineering, Lehigh University, Bethlehem PA, USA

Abstract

ClC chloride channels and transporters are important for chloride homeostasis in species from bacteria to human. Mutations in ClC proteins cause genetically inherited diseases, some of which are likely to have folding defects. The ClC proteins present a challenging and unusual biological folding problem because they are large membrane proteins possessing a complex architecture with many re-entrant helices that go only part way through membrane and loop back out. Here we were able to examine the unfolding of the *E. coli* ClC transporter, ClC-ec1, using single-molecule forced unfolding methods. We find that the protein can be separated into two stable halves that unfold independently. The independence of the two domains is consistent with an evolutionary model in which the two halves arose from independent folding subunits that later fused together. Maintaining smaller folding domains of lesser complexity within large membrane proteins may be an advantageous strategy to avoid misfolding traps.

The ClC family includes a large number of both passive channels and active transporters. The Cl^-/H^+ antiporters are ion transporting machines found in both prokaryotic and eukaryotic cells that pump chloride ions into cells using a proton gradient[1,2]. The vital functions of the ClC transporters include the maintenance of membrane potential, volume

*To whom correspondence should be addressed: bowie@mbi.ucla.edu.

Author Contributions

D.M., R.E.J. and J.U.B. conceived and designed the experiments. Y.Q. and W.I. designed and performed MD simulations. D.M., R.E.J., J.Y.W. and M.A.A. performed plasmid cloning and protein purification. D.M. performed DNA handle conjugation and single-molecule forced unfolding experiments. R.E.J. performed the domain isolation, SEC, CD, and vesicle swelling experiments. D.M., R.E.J. and J.U.B. analyzed the experimental data. Y.Q., W.I., D.M. and J.U.B. analyzed the MD simulations. D.M., R.E.J., Y.Q., W.I. and J.U.B. wrote the paper.

Competing Financial Interests

The authors declare no competing financial interests.

homeostasis, epithelial secretion of fluids, control of intravesicular pH, and muscle contraction[3–10]. Mutations in ClC channels and transporters cause numerous genetic diseases including Bartter syndrome, Dent's disease, CNS disorders, epilepsy, heart failure, Thomsen disease, and Becker disease[9]. Disease-causing mutations can result in aberrant protein trafficking or enhanced degradation, suggesting defects in folding[11,12].

Like all characterized ClC family members, the *E. coli* ClC antiporter, ClC-ec1, is a dimer with a single transport pathway per subunit. Mutations in the dimer interface yield monomers that are fully functional[13,14]. Each ClC-ec1 subunit has internal pseudo-symmetry, making the protein divisible into two domains with inverted topology[13,15–18] (Fig. 1a,b).

The inverted topology of the N- and C-terminal halves of the protein suggest that the ClC monomer likely evolved from an ancient gene duplication[19–21], yet it is hard to see how the two halves of the protein could have ever folded separately. Both domains adopt a complex topology that includes typical transmembrane helices, long tilted helices, and so-called re-entrant helices that penetrate only part way and then loop back out[19,22] (Fig. 1a,b). The re-entrant helices expose backbone hydrogen bonds that presumably must be accommodated in the membrane both during the folding process and in the final folded protein (Fig. 1c, yellow loops). Moreover the domain interfaces bear many polar and charged side chains that would be exposed to the hydrocarbon core of the bilayer (see Fig. 1c and electrostatic potential map in Supplementary Fig. 1). The polar surfaces meet at the interface of the two domains, effectively shielding them from the apolar membrane. Thus, each monomer gives the appearance of a highly interconnected unit that must fold with considerable coupling between domains to accommodate the polar interfaces. As described below, however, the structural appearance belies the true energetic architecture of the ClC fold.

Single-molecule forced unfolding methods are particularly useful for the study of membrane protein folding for several reasons. First, since the protein can be held at infinite dilution, aggregation problems that plague membrane protein folding experiments can be eliminated. Second, because the molecular interactions that dominate folding are different throughout the bilayer, most denaturants will have variable effects in different parts of the protein. Force, however, is relatively catholic with respect to the nature of the stabilizing interactions. Third, unlike chemical denaturants, mechanical unfolding does not impact membrane integrity.

Membrane proteins have been extensively characterized by single-molecule force spectroscopy using atomic force microscopy (AFM)[23–26]. The AFM methods have so far been limited to pulling proteins out of the bilayer, however, which is not a physiologically relevant unfolding process for helical membrane proteins. For most helical membrane proteins, biogenesis occurs in two stages: membrane insertion followed by folding of the inserted protein[27–29]. To more directly probe the second-stage folding, we have employed magnetic tweezer techniques[30–35] to apply force along the plane of the bilayer[36,37]. Like the prior AFM methods, the forces required to unfold the protein can also extract the protein from the bilayer. Nevertheless, in favourable cases it is possible to learn about folding within a bilayer because unfolding to the transition state occurs within the bicelle and refolding at

low forces also occurs within the bicelle (see the original and review articles for discussion[36,38]).

Results

Two step unfolding of ClC-ec1

The experimental set up is shown in Fig. 1d. The monomeric variant of ClC-ec1 is covalently linked to a SpyCatcher protein via SpyTag peptide[39] placed at the N- and C-termini of the ClC-ec1. The SpyCatcher protein is in turn covalently linked to DNA handles[37] (Supplementary Fig. 2 and Online Methods). One DNA handle is fastened to a glass surface via a neutravidin-biotin complex and the other DNA handle is attached to a magnetic bead via a digoxigenin/anti-digoxigenin complex. To provide a native-like membrane environment, the ClC-ec1 transporter is reconstituted into bicelles built from a DMPC lipid layer wrapped in CHAPSO detergent molecules[40] (Fig. 1d). Mechanical tension can be applied with a pair of permanent magnets and the response of the system can be observed by measuring the extension (bead height)[36,41–43].

We performed force-ramp experiments in which the force was gradually increased or decreased (~0.5 pN/s). Some representative results are shown in Fig. 1e. The main unfolding events are generally seen above 25 pN, occurring in two large jumps in extension of ~55 nm each (blue arrows) with clear pauses in between (red arrows). As illustrated in Fig. 2a–c, the two unfolding steps are almost identical in size: the increase in extension for the first step is ~51% of the total unfolding extension (see Fig. 2c for analysis). Analyzing unfolding step sizes with worm-like chain (WLC) model (Fig. 2d), we estimated the persistence length as ~0.44 nm (consistent with the previous reports[23,25,26,36]) and used this parameter for theoretical estimation of domain step sizes. The measured step sizes of ~55 nm roughly coincide with theoretical step sizes expected for the complete unraveling of each domain (Online Methods). These results indicate that the two halves of the protein unfold in separate events, rather than cooperatively as one unit. The unfolding probabilities of the first and second steps are indistinguishable in the broad force range of ~20 pN (Fig. 2e), indicating that the two halves of the protein are likely to unfold in stochastic manner[44]. These results suggest that, surprisingly, half the protein may remain folded while the other half is completely unfolded.

Domains dissociate before unfolding

Prior to the main unfolding events in Fig. 1e, we usually observed smaller pre-transitions with a step size of about ~5 nm (green arrows) which often showed reverse transitions of the same step size (yellow arrows). We speculated that the pre-transition oscillations involved dissociation and re-association of the two domains of ClC-ec1 occurring prior to the stochastic, independent unfolding of the two domains. Indeed the domain shape and pulling geometry would cause inter-domain linker stretching and domain rotation that would be consistent with a ~5 nm extension jump (Fig. 3a and Online Methods).

To test the domain separation hypothesis, we engineered a variant of ClC-ec1 with an inter-domain linker that was roughly twice as long by inserting an additional 16 residues (Online

Methods). If the pre-transition involves separation of the two domains, the size of the pre-transition extension should increase in the long linker construct. The WLC model predicts that the additional 16 residue insertion would lead to an additional ~5 nm increase in extension, thereby doubling the extension seen for the wild-type protein. Indeed, as shown in Fig. 3b and 3c, the long linker construct approximately doubled the initial pre-transition step-size (now ~10 nm). The extension size ratio distribution (mutant/WT) also shows a peak of ~1.9 (Fig. 3d). Interestingly, the long linker mutant did not show re-association events, consistent with a lower mutual effective concentration of the two domains upon stretching of the long linker. These results suggest that the ClC-ec1 dissociates into the two stable folded units before the final unfolding steps.

Independent domain isolation

The results described so far indicate that the two domains of the ClC subunit may remain stably folded on their own. Given the unusual architecture of each domain and the fact that domain separation would lead to the exposure of polar backbone and side chain moieties to the hydrophobic membrane environment (Fig. 1a–c), we sought additional validation of this unexpected finding. We therefore attempted to purify the individual domains to see if either could remain stably folded in isolation. To this end, we introduced a TEV protease site into the linker between domains and cut it with TEV protease (Fig. 4a and Supplementary Fig. 3). The migration of the cleaved protein on a gel filtration column remained unchanged indicating that the domains remained stably associated after cleavage (Supplementary Fig. 4). Moreover, when we applied the protein to a Ni-NTA column in n-dodecyl-β-D-maltopyranoside (DDM), which should bind only the His tag at the end of the C-terminal domain, the untagged N-terminal domain did not pass through. To separate the domains (schematics in Supplementary Fig. 5), we bound the protein to a Ni-NTA column and then washed the column with a series of detergents to find one that was sufficiently denaturing to cause the domains to dissociate. We found that n-dodecylphosphocholine (DPC) was able to elute the N-terminal domain, leaving the C-terminal domain bound to the Ni-NTA column. The column was then re-equilibrated with DDM and the C-terminal domain eluted using imidazole, while the eluted N-terminal domain was exchanged back into DDM by application to a size exclusion column.

Figure 4b shows the elution position of the ClC domains on a size exclusion column in comparison to other proteins. The N-terminal domain fragment (calculated radius of gyration from crystal structure, cRg = 18.6 Å) elutes later than two small folded proteins, *E. coli* diacylglyerol kinase (cRg = 19.4 Å) and the transmembrane domain of *E. coli* GlpG (cRg = 16.4 Å), whereas the C-terminal domain fragment (cRg = 17.3 Å) elutes before even the full length ClC-ec1 protein (cRg = 21.2 Å). These results suggest that the N-terminal domain remains compact and substantially folded, while the C-terminal domain was unfolded by our isolation protocol. This conclusion is supported by the circular dichroism (CD) spectra in Fig. 4c indicating a helical content of the N-terminal domain that is similar to the intact, full-length protein (Supplementary Table 1). Although we did not obtain a folded C-terminal domain with our isolation procedure, we cannot rule out that it would be possible under different conditions. Nevertheless, our bulk experiments confirm the results

from the forced unfolding experiments that a domain with a complex topology and exposed polar loops in the hydrocarbon core region can remain largely folded.

Possible domain stabilization by bilayer distortion

To gain insight into how the isolated domains could be stably folded in a bilayer environment, we turned to molecular dynamics simulations. The N-terminal domain, C-terminal domain and full-length ClC-ec1 were individually embedded in DMPC bilayers and simulated for 2 μs using the CHARMM force field[45]. The root-mean square deviations (RMSDs) of the Cα atoms of the transmembrane residues for both the N-terminal domain and the full-length protein generally remained below 2 Å during the last 1 μs while the C-terminal domain showed larger deviations, but still below 3 Å (Fig. 5a). Thus, all the structures remained stable during the simulation, although the larger excursion of the C-terminal domain is consistent with the experimentally observed lower stability of the isolated C-domain.

To examine how the exposed polar residues at the N-C domain interface and the re-entrant loops in the center of the protein could be stabilized upon domain separation, we analyzed the hydrogen bonds between these residues and their environments including water molecules, lipids, and protein moieties (Fig. 5b). In the full-length ClC-ec1, residues at the interface are already highly solvated by water (top panel), but upon domain separation, the solvation by water increases, as does the hydrogen bonding to lipid head groups (bottom panel) (Online Methods). This is not surprising as polar residue exposure to the hydrocarbon region is not energetically favorable and water and lipid head groups must come into play (Supplementary Video). As seen in a representative snapshot of the N-domain simulation during the last 1 μs (Fig. 5c), the polar domain interface is shielded from the apolar hydrocarbon chains by lipid head groups that bend down into the interface and the additional water that enters into the lipid/domain interface. Quantitative analysis of the water density in the transmembrane region (e.g., $|z| < 5$ Å with the bilayer center at $z = 0$) shows that a water pore is formed at the exposed polar interfaces for both N and C domains (Fig. 5d). The N-domain system shows a larger water pore than the C-domain, which is likely due to stronger electrostatic potential on the exposed interface in the N-domain (Supplementary Fig. 6a,b). Along with the water pore, the thickness of the bilayer is considerably reduced around the exposed N-C interface by bilayer distortion, forming a hydrophilic channel that allows water molecules to fill in (Supplementary Fig. 6c). In summary, the MD simulations show that the isolated N and C domains have the potential to be stable in a membrane, and that the exposed polar residues are stabilized by increased penetration of water molecules and lipid head groups.

To experimentally test the implications from the MD simulations that the isolated domains introduce a membrane defect, we employed an osmotic shock vesicle swelling assay. We first prepared proteoliposomes with full length ClC-ec1 or isolated N-domain. We then subjected the vesicles to an osmotic shock by rapid mixing with a mannitol solution in a stopped-flow instrument, and tracked the changes in vesicle size by light scattering at 430 nm. Initially the vesicles shrink as water exits the vesicle and then slowly re-swell as osmolytes enter the vesicle, thereby lowering the osmotic pressure differential

(Supplemental Fig. 7). Fig. 5e shows swelling rates as a function of protein to lipid ratio. It can be seen that the swelling rates for the N-terminal domain are much faster than for the full length protein, even at lower protein/lipid ratios. Indeed the rate of swelling for vesicles containing the N-terminal domain at a protein/lipid ratio of ~7.7×10^{-5} is almost 70-fold higher than for the full length protein at a similar protein/lipid ratio of ~9.4×10^{-5}. Thus, the experiments are consistent with the MD simulations that indicate transient pore formation by the isolated N-domain, allowing more rapid osmolyte entry into the vesicles. Nevertheless, we cannot yet say if the N-domain is monomeric in the membrane vesicles and more work will be needed to learn about how the structure of the N-domain may adjust in the bilayer.

Stable misfolded states

We also examined whether the protein can be refolded. After forced unfolding, we relaxed the force down to zero pN (Fig. 6a), and after waiting 20 mins, we ramped the force back up to observe the unfolding behavior. Among all the follow-up traces examined (N=96), only 11% showed successful refolding as indicated by restored unfolding behavior comparable to what we observed for the first unfolding trace (*i.e.*, identical folded extension, dissociation, and then two large unfolding events at high force) (Fig. 6a, the top boxed traces). The remaining 89% displayed unfolding behaviors that deviated from the initial traces in various ways (Fig. 6a,b). First, the unfolding forces dramatically dropped to less than 20 pN (Fig. 6a, the bottom boxed traces). Second, for the most of traces, the force-loading curve was slightly shifted from the correctly folded one, implying that some portion of protein was partially or incorrectly folded (Fig. 6b and the upper right inset). Third, the unfolding patterns were not consistent with one another. The traces showed different numbers of unfolding steps, inconsistent locations of intermediates, or gradual unfolding transitions (Fig. 6b and the lower right inset). The misfolded states appeared fairly stable since the unfolding forces went up to tens of pN. Moreover, the misfolded states persisted in spite of the long 20 min waiting time. Some traces demonstrated very clear unfolding transitions at over 15 pN (Fig. 6b), indicating that the nonnative interaction network could be very strong. Some of the misfolded states may involve correct folding of one domain and incorrect folding of the remainder of the protein, but it is impossible to know the nature of the misfolded structures at this point.

In the force-unloading curves, we observed ~10 nm compaction of the structure below 5 pN (Fig. 6a, denoted with the black arrow). The compactions are easily seen in the protein extension curve calculated by subtracting force-loading extension from force-unloading extension (Fig. 6a, inset). As the coil-to-helix transition and reconstitution into a bicelle was found to occur at ~18 pN[36], we suggest that the compaction is due to a protein folding event involving non-native structure (Supplementary Fig. 8). Thus, folding might occur from a relatively compact state, possibly explaining why it is hard to find the native structure.

Discussion

Our results indicate that large membrane proteins like ClC-ec1 can find stable misfolded states that can persist for 20 mins or more. Thus, the folding pathway must be an important factor in the efficient folding of large membrane proteins such as ClC-ec1. No doubt part of

the natural process is correct insertion and initial topology determination which would not be constrained in our experiments. Moreover, chaperones such as YidC may play a role in facilitating folding[46,47]. Another possible mechanism to facilitate the folding process would be to break the folding up into smaller pieces (domains) as naturally occurs in soluble proteins[48,49]. Although the folding process is largely invisible to us, the unfolding pathway we observe is consistent with the possibility of two stable folding units. In particular, at high forces we see that both the N- and C-terminal domains can remain folded after separation on the time scale of seconds. How long might stable domains persist at zero force? The dwell time at zero force should exponentially increase with applied force, inferred from Bell equation, $\tau_0 = \tau_F \exp(Fx/k_B T)$, where τ_0 is the dwell time at zero force, τ_F is the dwell time at a given force, F is the applied force, x is the distance between native and transition states, and $k_B T$ is the thermal energy constant. If we roughly assume that x=1 nm and τ_F=5 s at F=30 pN, we obtain τ_0 = 7396 s (~2 hours) at zero force, suggesting that the domains may have an unfolding half-life in bicelles that is on the order of hours at zero force. Moreover, we are able to obtain a largely folded N-domain that is stable for many days in isolation. Thus, once folded, the domains appear to be found in a stable energy well. If a favorable pathway exists for folding the individual domains, it would provide a natural separation of the folding process. In particular, as the N-terminal domain of ClC-ec1 emerges from the translocon first, it might fold initially and possibly template the folding of the C-terminal domain[50] (Supplementary Fig. 9). Our finding that the folded N-terminal domain leads to membrane defects in DMPC bilayers runs counter to this idea, although the holes seen in pure DMPC bilayers might be sealed by proteins or other lipids found in natural membranes. While our results are suggestive of a two part folding mechanism for ClC-ec1, much work remains to learn whether multi-part folding mechanisms actually occur for ClC-ec1 or other large membrane proteins.

Online Methods

ClC-ec1 plasmid constructs

A DNA fragment encoding a polypeptide containing two sequential SpyTags[39] and a C-terminal 6xHis tag, all separated by glycine-serine linkers, was synthesized (IDT, Coralville, IA), digested with the restriction enzymes NdeI and NotI, and cloned into pET24a (EMB Millipore, Billerica, MA) using the one-step ISO DNA assembly method. The resulting plasmid was named pSpy2His in which the encoded amino acid sequence is MAHIVMVDAYKPTKGGSGGSGGSAHIVMVDAYKPTKGGSHHHHHH (SpyTags are underlined). The gene encoding the *E. coil* ClC monomer (ClC-ec1 with mutations I201W and I422W) was PCR-amplified from a plasmid template[13] (courtesy of Janice L. Robertson at University of Iowa) and cloned into the pSpy2His plasmid to generate the following protein sequence:

MAHIVMVDAYKPTKGGSGGSIRQLLERDKTPLAILFMAAVVGTLVGLAAVAF
DKGVAWLQNQRMGALVHTADNYPLLLTVAFLASAVLAMFGYFLVRKYAPEA
GGSGIPEIEGALEDQRPVRWWRVLVKFFGGLGTLGGGMVLGREGPTVQIGGN
IGRMVLDIFRLKGDEARHTLLATGAAAGLAAAFNAPLAGILFIWEEMRPQFRY
TLISIKAVFIGVIMSTIMYRIFNHEVALIDVGKLSDAPLNTLWLYLLGIIFGIFGPI
FNKWVLGMQDLLHRVHGGNITKWVLMGGAIGGLCGLLGFVAPATSGGGFNL

IPIATAGNFSMGMLVFIFVARVITTLLCFSSGAPGGIFAPMLALGTVLGTAFGMV
AVELFPQYHLAGTFAIAGMGALLAASIRAPLTGIILVLEMTDNYQLWLPMIITG
LGATLLAQFTGGKPLYSAILARTLAKQEAEQLARSKAASGGSGGSGGSAHIV
MVDAYKPTKGGSHHHHHH.

We also prepared a long linker construct by inserting a polypeptide fragment into the middle loop between the N- and C-terminal subunits. The amino acid sequences of the original middle loop and the long linker are as follows: NHEVALIDVGKLSDAP and NHEVALIDVGKLSDGGSGGSGGSENLYFQGAP, respectively. This long linker construct was used in the unfolding studies. The insertion also bears a TEV protease site (ENLYFQG), but we were unable to cleave it using TEV protease, likely due to steric hindrance. To isolate the individual domains, we therefore constructed a second long linker construct that moves the TEV cleavage site closer to the center: NHEVALIDVGKLSDENLY FQGGGSGGSGGSAP. The second long linker construct was used for domain purification and full-length controls in SEC and CD experiments.

ClC-ec1 expression and purification

The expression plasmids described above were transformed into *E. coli* C43 (DE3) competent cells. 2 L of LB broth (BD Difco, Franklin Lakes, NJ) containing 50 μg/ml kanamycin sulfate was inoculated with 10 ml of overnight cell culture and grown at 37°C. Protein expression was induced with 400 μM IPTG at an OD_{600} of ~0.7 after lowering the temperature to 16°C. Cells were harvested 20 h after the induction and resuspended in 2 ml lysis buffer per gram of cells. The lysis buffer composition was 25 mM Tris (pH 7.5), 300 mM NaCl, 10% glycerol, 2 mM TCEP, 1% protease inhibitor cocktail (Sigma-Aldrich, P8849), 1 mM PMSF. The resuspension was lysed by two passages through an Emulsiflex-C3 homogenizer (Avestin Inc, Ottawa, Ontario, Canada) at a pressure of 10,000~15,000 psi. Cell debris was removed by centrifugation at 20,000g for 30 mins at 4°C. The cleared lysate was then further centrifuged at 100,000g for 90 mins at 4°C to obtain a membrane pellet. The pelleted membranes were solubilized in 2 ml extraction buffer per gram of cells (25 mM Tris (pH 7.5), 300 mM NaCl, 2% n-dodecyl-β-D-maltopyranoside (DDM), 2 mM TCEP) with the aid of a Dounce homogenizer followed by the rotation for 45 mins at 4°C. Insoluble material was removed by centrifugation at 100,000g for 30 mins at 4°C. The supernatant was bound to 0.25 ml Ni-NTA Superflow resin (Qiagen, Hilden, Germany) per liter of culture with 20 mM imidazole (pH 7.5) for 1 h rotating at 4°C. The column was washed twice with 5 column volumes (CVs) of wash buffer I (25 mM Tris (pH 7.5), 300 mM NaCl, 10% glycerol, 0.1% DDM, 1 mM TCEP, 20 mM imidazole), and then washed twice with 5 CVs of wash buffer II (25 mM Tris (pH 7.5), 300 mM NaCl, 10% glycerol, 0.1% DDM, 1 mM TCEP, 40 mM imidazole). The ClC-ec1 protein was eluted with elution buffer (25 mM Tris (pH 7.5), 300 mM NaCl, 10% glycerol, 0.1% DDM, 1 mM TCEP, 300 mM imidazole). Protein-containing fractions were pooled and concentrated to ~500 μl with a 50K Amicon centrifugal filter device (EMD Millipore, Darmstadt, Germany). The concentrated protein was further purified by size-exclusion chromatography (SEC) using a Superdex 200 Increase 10/300 GL column (GE Healthcare Life Sciences, Marlborough, MA) equilibrated with high-salt SEC buffer (25 mM Tris (pH 7.5), 300 mM NaCl, 10% glycerol, 0.1% DDM, 1 mM TCEP). Fractions were checked for ClC-ec1 protein by 4–20% SDS-PAGE. Purified

fractions were pooled and concentrated by the 50K Amicon centrifugal filter device. The concentrated samples were aliquoted, flash-frozen in liquid nitrogen, and stored at −80°C.

DGK expression and purification

A His-tagged cysless construct of DGK ($DGK_{cysless}$) was expressed and purified as previously described[51]. Final elutions were made with 0.1% DDM. 0.2 mg of purified $DGK_{cysless}$ was analyzed by size-exclusion chromatography (Superdex 200 increase 10/300 GL, GE Healthcare Life Sciences) in the high-salt SEC buffer. The protein sequence is as follows.

MGHHHHHHELANNTTGFTRIIKAAGYSWKGLRAAWINEAAFRQEGVAVLLA
VVIAAWLDVDAITRVLLISSVMLVMIVEILNSAIEAVVDRIGSEYHELSGRAKD
MGSAAVLIAIIVAVITWAILLWSHFG

GlpG expression and purification

A His-tagged cysless construct of the GlpG transmembrane domain ($GlpG_{cysless}$) was created from the previously described $GlpG_{Cys\text{-}TM\text{-}Cys}$[36] by deleting the terminal cysteines with site-directed mutagenesis. $GlpG_{cysless}$ was purified by Ni-NTA metal affinity and anion exchange chromatography as previously described[36]. The flow-through peak after anion exchange was further purified by size-exclusion chromatography in the high-salt SEC buffer. 0.2 mg of purified $GlpG_{cysless}$ was re-analyzed by size-exclusion chromatography in the high-salt SEC buffer. The protein sequence is as follows.

MGHHHHHHELAALRERAGPVTWVMMIAAVVVFIAMQILGDQEVMLWLAW
PFDPTLKFEFWRYFTHALMHFSLMHILFNLLWWWYLGGAVEKRLGSGKLIVI
TLISALLSGYVQQKFSGPWFGGLSGVYALMGYVWLRGERDPQSGIYLQRGLI
IFALIWIVAGWFDLFGMSMANGAHIAGLAVGLAMAFVDSLNARKRK

TEV expression and purification

The His-tagged MBP-TEV fusion construct with a C-terminal truncation and three solubility-enhancing mutations (MHT238Δ) was expressed in BL21-Gold (DE3) at 37°C for 3 h after induction with 0.5 mM IPTG, and purified as previously described[52]. Purified aliquots were stored in storage buffer (10 mM Tris, pH 7.5, 0.3 mM TCEP, 0.5 mM EDTA, 50% glycerol) at −20°C.

DNA handle attachment to ClC-ec1

Following a previously established procedure[37], 100~200 nM of SpyCatcher-conjugated DNA handle (SpyCatcher-DNA) was prepared through maleimide-thiol crosslinking between a maleimide-modified 512 bp DNA and an engineered SpyCatcher protein possessing an unique cysteine. The other end of the DNA handles was modified with biotin or digoxigenin for the attachments to the glass surface and magnetic beads respectively. Unconjugated proteins and DNA were removed by ion exchange chromatography and an amylose affinity resin that captures the maltose-binding protein tag fused to the SpyCatcher[37]. To attach the SpyCatcher-DNA handles, 10 μl of ~100 nM SpyCatcher-DNA sample was mixed with 1 μl of ~20 μM SpyTagged ClC-ec1 (Supplementary Fig. 2a). After ~22 h incubation at 4°C, the handle attachment progress was checked by 4–20% SDS-PAGE

with DNA staining (Supplementary Fig. 2b). 8 μl of the mixture was diluted to ~800 μl with 50 mM Tris (pH 7.5), 150 mM NaCl, 2 mM TCEP, 1.3% bicelle, and then stored at −80°C in 10 μl aliquots. Bicelles were formulated by mixing 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (DMPC; Avanti Polar Lipids, 850345P) and 3-((3-cholamidopropyl)dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO; Affymetrix, C317) at 2.8:1 molar ratio[36,40]. From the gel analysis, the concentration of ClC-ec1 monomer attached to the two distinct DNA handles was estimated as ~150 pM for the final stored sample.

Single-molecule forced unfolding experiment

The single-molecule forced unfolding experiments were performed on a magnetic tweezer apparatus which was custom-built and calibrated as previously described[33,36,37,41–43]. Extension values at the calibrated forces were acquired at a 60 Hz frame rate as also previously described[33,36,37,41–43]. A ~10 μl glass chamber for the ClC-ec1 monomer was prepared by following an established procedure applicable to membrane proteins[36,37]. The buffer condition during the single-molecule pulling cycles was 50 mM Tris (pH 7.5), 150 mM KCl, 1.3% bicelle (DMPC/CHAPSO at 2.8:1 molar ratio). The temperature was maintained at 22°C. In the gradual force ramps, the pair of magnets was translationally moved to the sample chamber at a speed of 0.1 mm/s (average ~0.5 pN/s) in the force range of 1~50 pN. Force-extension curves in Fig. 6 were median-filtered with a 10-point window size.

Analysis of domain unfolding

To map the intermediate pause location in the two step unfolding (Fig. 2), we determined the fractional extension of the pause relative to the total extension upon complete unfolding. We defined the normalized fractional extension as $E_{\mathrm{norm}}(t) \equiv (E_{\mathrm{obs}}(t) - E_{\mathrm{D}}(t))/(E_{\mathrm{U}}(t) - E_{\mathrm{D}}(t))$ where $E_{\mathrm{obs}}(t)$ is the observed extension, and $E_{\mathrm{D}}(t)$ and $E_{\mathrm{U}}(t)$ are the dissociated and unfolded states, obtained from linear fits of the extension *vs* time plots (see Fig. 2b). We overlapped $E_{\mathrm{norm}}(t)$ traces obtained from a number of ClC-ec1 unfolding events and then constructed a count histogram pooled from the paused state (Fig. 2c). The histogram peak was 0.51 (±0.01, SEM), indicating that the two extension jumps are almost of the same size as expected for the same sized domains. The unfolding at forces over 25 pN fits well to the Marko-Siggia (MS) formula of worm-like chain (WLC) model (see Fig. 2d; the formula is below). The persistence length was measured as 0.44 nm (±0.06 nm, SEM, 95% CI) by fitting to the observed unfolding step size with force, with a contour length for each half domain of 209 residues × 0.38 nm/residue (Fig. 2d). The expected unfolding size for each domain (ΔE_{HD}) was estimated from the formula $\Delta E_{\mathrm{HD}} = d_{\mathrm{HD}}(F) - d_1$(or d_2), where $d_{\mathrm{HD}}(F)$ is the end-to-end distance of unraveled random coil from the WLC model and d_1(or d_2) is the distance between the pulling end point and middle linker end point for N-terminal domain (or C-terminal domain) from monomeric ClC-ec1 structural data (PDB 3NMO)[13]. The d_1 and d_2 parameters account for the extension caused by the rotation of each subunit (see below and Fig. 3a). The $d_{\mathrm{HD}}(F)$ is calculated from the MS formula,

$$F = \frac{k_{\mathrm{B}}T}{L_{\mathrm{p}}}\left[\frac{1}{4(1 - d_{\mathrm{HD}}/L_{\mathrm{c}})^2} + \frac{d_{\mathrm{HD}}}{L_{\mathrm{c}}} - \frac{1}{4}\right],$$

where F is the applied force, $k_{\mathrm{B}}T$ is the energy scaling

factor equal to 4.11 pN·nm, L_p is the persistence length of polypeptide (estimated as 0.44 nm in Fig. 2d), and L_c is the contour length of each domain (calculated as 209 residues × 0.38 nm/residue). The measured step sizes of ~55 nm roughly coincide with theoretically estimated values of each domain unfolding of 54.0–58.7 nm for N-domain (52.0–56.7 nm for C-domain) at 30–45 pN.

Analysis of domain separation

To measure the pre-transition size, we performed first-degree polynomial fitting to the extension *vs* time traces for the folded and dissociated states since real-time traces in a limited high-force region under our magnetic tweezer settings are approximately linear (see the fits in Fig. 3b and Supplementary Fig. 10 for simulations). We then calculated the pre-transition size as $E_D(t)-E_N(t)$ at the transition time T where $E_D(t)$ is the fit function for the dissociated state and $E_N(t)$ is the fit function for the folded state (Fig. 3c). The scatter plot in Fig. 3c shows 4.9 (±1.1, SD) nm for the wild-type. The theoretical pre-transition size can be estimated as $\Delta E = d_{ML}(F) + d_1 + d_2 - d_0$ where d_0 as the initial distance between the pulling end points, $d_{ML}(F)$ as the end-to-end distance of the stretched middle linker after domain separation, and d_1 and d_2 as the distance changes due to subunit rotation after separation. The formula reflects middle linker stretching and each domain rotation during domain separation. The geometry of the transition is illustrated in Fig. 3a. $d_{ML}(F) \gg d_1 + d_2 - d_0$ since $d_1 + d_2 - d_0$ = ~0 nm, so the pre-transition step size $\Delta E \sim d_{ML}(F)$. The $d_{ML}(F)$ is calculated from the MS formula of the WLC model (see above). The contour length of the middle linker was calculated as 16 residues × 0.38 nm/residue. The observed extension increase of ~5 nm upon domain separation roughly coincides with the theoretical estimation of 4.1–4.9 nm at 20–45 pN. As for the long linker mutant, we expect that the pre-transition size ΔE should be doubled because the doubly lengthen middle linker doubles $d_{ML}(F)$ (see the section of ClC-ec1 plasmid constructs for the actual linker sequences). The ratio of pre-transition sizes is estimated as $E_{ratio} = \Delta E_{long\ linker}/\Delta E_{WT} \sim 2d_{ML(WT)}/d_{ML(WT)} = 2$. As anticipated, we measured 10.1 (±2.0, SD) nm for the pre-transition for the mutant (Fig. 3c) and 1.9 (±0.1, SEM) for the size ratio (Fig. 3d).

Isolation of N- and C-terminal domains

To cleave the long linker construct described above, 440 ul of 0.54 mg/ml purified ClC-ec1 in the high-salt SEC buffer was cleaved with 20 ul of 2.5 mg/ml TEV in storage buffer for 16 h at room temperature (Supplementary Fig. 3). The TEV protease was removed from the cleaved protein by SEC in the high-salt SEC buffer (the ClC-ec1 domains remain associated as in Supplementary Fig. 4). Fractions containing complexed fragments of ClC-ec1 were pooled, concentrated and bound to 0.5 ml Ni-NTA resin for 1.5 h at 4°C. The column was washed with 10 ml of DPC wash buffer (25 mM Tris, pH 7.5, 300 mM NaCl, 0.5% n-dodecylphosphocholine (DPC), 1 mM TCEP) over 2 h at 4°C. Resin was incubated in the DPC wash buffer overnight at 4°C, and then washed again with 10 ml of the DPC wash buffer over 2 h before washing with DDM wash buffer (25 mM Tris, pH 7.5, 300 mM NaCl, 0.5% DDM, 1 mM TCEP). The remaining His-tagged C-terminal fragment was eluted with imidazole (25 mM Tris, pH 7.5, 300 mM NaCl, 300 mM imidazole, 10% glycerol, 0.1% DDM, 1 mM TCEP). The DPC wash fractions were pooled, concentrated, and purified by SEC in the high-salt SEC buffer to isolate the N-terminal fragment. The imidazole elution

fractions were pooled, concentrated, and purified by SEC in the high-salt SEC buffer to isolate the C-terminal fragment. Fractions containing purified fragments were pooled, concentrated, and re-analyzed by SEC (Fig. 4b).

Circular dichroism

The full-length ClC-ec1 and isolated cleaved fragments were exchanged by SEC into low-salt DDM buffer (25mM Tris, pH 7.5, 30 mM NaCl, 0.1% DDM, 1 mM TCEP) and concentrated to ~0.3 mg/ml. Measurements were carried out on a Jasco J-715 circular dichroism spectrophotometer with a 1 mm cuvette (Fig. 4c). The spectra are the average of 3 scans collected at 0.5 nm intervals from 260 to 199 nm. A buffer only baseline was subtracted from all protein spectra, and the measured ellipticity was converted to mean residue ellipticity (molar ellipticity). Secondary structural content was calculated using the SELCON3 program within the CDPro package using a reference set of 13 membrane proteins and 43 soluble proteins[53] (Supplementary Table 1).

MD simulations

MD simulations were performed with the CHARMM 36 force field[45]. The simulation systems were prepared using CHARMM-GUI Membrane Builder[54] with initial coordinates of the protein from a crystal structure of ClC-ecl (PDB ID 1OTS)[17]. The structures for the N-terminal (residue ID 17 to 240) and C-terminal (residue ID 236 to 460) domains were extracted from the full-length structure. The proteins were embedded in DMPC bilayers and solvated with TIP3P water molecules. KCl was added to neutralize the systems and maintain a salt concentration of 150 mM. The total number of atoms in the N-terminal domain, C-terminal domain, and full ClC system was 103393, 78608, and 109899, respectively. Each system was equilibrated for 20 ns using NAMD 2.12 on a local computer cluster. The non-bonded interactions were smoothly switched off at 10–12 Å by a force-switch function[55]. The particle mesh Ewald algorithm[56] was used to calculate electrostatic forces. Temperature was maintained using Langevin dynamics with a coupling coefficient of 1 ps^{-1}. The Nosé-Hoover Langevin-piston method[57] was used to maintain constant pressure at 1 bar with a piston period of 50 fs and a piston decay of 25 fs. Following equilibration, all systems were transferred to Anton 2 for production runs[58]. The non-bonded interactions were cut off at 10 Å. The u-series approach[58] was used to calculate long-range electrostatic interactions. Pressure and temperature were controlled at 1 bar and 305.5 K with the Multigrator framework[59] using semi-isotropic MTK barostats and Nosé-Hoover thermostats. Each simulation was performed for 2 μs with a time step of 2 fs. The trajectory analysis and resulting probabilities in Fig. 5b was performed and calculated with CHARMM and VMD. Hydrogen bonds were calculated with a distance cutoff of 2.8 Å between the donor and acceptor atoms and an angle cutoff of 120°.

Vesicle swelling assay

DMPC lipid was dissolved in chloroform at 25 mg/ml in glass tubes and dried under a stream of argon gas. Lipids were further dried under vacuum for at least 2 h. Dried DMPC lipids were solubilized in 4% n-octyl-β-D-glucopyranoside (OG) (Anatrace), 25 mM sodium phosphate [pH 7.5], 30 mM NaCl to 25 mg/ml on ice with short bursts of vortexing and sonication in a bath sonicator. Purified ClC-ec1 protein in DDM was added to fully

solubilized DMPC to final concentrations of 3.70, 7.40, or 14.8 μM for full-length protein or 1.85, 3.70, or 7.40 μM for isolated N-domain. Proteoliposomes were formed by removing the solubilizing detergent by dialysis at 4°C against 4 to 6 exchanges of 500 ml of 25 mM sodium phosphate [pH 7.5] using 25 kDa MWCO Spectra/Por tubing (Spectrum Labs). Proteoliposomes were extruded through a 0.2 μm Nucleopore Track-Etch Membrane (Whatman) in a mini-extruder (Avanti Polar Lipids) 15 times to form monodisperse unilamellar vesicles. The size distribution of the resulting vesicles (197.0 ± 36.9 nm, SD, n=6) was confirmed by dynamic light scattering using a DynaPro Plate Reader II (WYATT).

To quantify incorporation of protein into the vesicles, 75% sucrose was added to the extruded vesicles to a final concentration of 20% and layered on top of 30 ml of 50% sucrose. Aggregates were spun down at 72,000 × g for 16 h at 4°C. Floated vesicles were removed by pipetting off the top turbid layer. Vesicles were diluted 3-fold in 25 mM sodium phosphate (pH 7.5) to lower the sucrose concentration. The diluted samples were then concentrated to 1–2 ml using a 10K Amicon centrifugal filtration device (EMD Millipore) to collect any vesicle material that may have been disrupted during the osmotic shock of diluting the sucrose concentration by 3-fold. The amount of protein in the concentrated sample of floated vesicles was measured by quantifying SDS-PAGE bands stained with coomassie-based Instant Stain (Expedeon) using an ImageLab 5.2.1 (BioRad) and comparing to a standard curve of SpyCatcher protein. Lipid concentrations were measured in triplicate using the Malachite Green Phosphate Detection Kit (R&D Systems) according to manufacturer's instruction. The sodium phosphate buffer was removed by dialysis against two exchanges of water using 25 kDa MWCO Spectra/Por tubing (Spectrum Labs). The lipid phosphates were hydrolyzed according to a previously established protocol[60] prior to the malachite green assay.

Proteoliposome shrinking and swelling after osmotic shock was measured at 32.5°C using a SpectroKinetic Monochromoter Stopped Flow Spectrometer (Applied Photophysics) for the N-domain or an SX20 Stopped Flow Spectrometer (Applied Photophysics) for the full-length ClC-ec1. To prevent bubbles forming over longer measurements of the full-length, the extruded vesicles were degassed under vacuum for 15 mins prior to stopped-flow measurements. Proteoliposomes in 25 mM sodium phosphate [pH 7.5] were exposed to osmotic shock by mixing with an equal volume of 25 mM sodium phosphate [pH 7.5] supplemented with 142 mM D-mannitol to increase the osmolarity. Samples were illuminated at 430 nm and the intensity of 90° scattered light was recorded at a maximum rate of 80 kHz oversampled to 10,000 points per measurement. Kinetic data was averaged from 10 shots and fit with a double exponential (Pro-Data Viewer, Applied Photophysics) for the fast shrink phase and slower swell phase, except in the case of 7.4 μM N-domain and longer measurements of the full-length proteoliposomes where the shrink phase was too fast to fit (Supplementary Fig. 7).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health (R01GM063919 to J. U. Bowie and U54GM087519 to W. Im), the National Science Foundation (MCB-1727508 to W. Im), and the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (NRF-2016R1A6A3A03007871 to D. Min). We thank J. L. Robertson at University of Iowa for sending a plasmid template containing monomeric ClC-ec1 gene and the members of our lab for comments on the manuscript. Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through Grant R01GM116961 from the National Institutes of Health. The Anton 2 machine at PSC was generously made available by D. E. Shaw Research.

References

1. Miller C. Open-state substructure of single chloride channels from Torpedo electroplax. Philos Trans R Soc Lond B Biol Sci. 1982; 299:401–411. [PubMed: 6130538]
2. Accardi A, Miller C. Secondary active transport mediated by a prokaryotic homologue of ClC Cl− channels. Nature. 2004; 427:803–807. [PubMed: 14985752]
3. Pusch M, et al. Mechanisms of block of muscle type CLC chloride channels (Review). Mol Membr Biol. 2002; 19:285–292. [PubMed: 12512775]
4. Gouaux E, Mackinnon R. Principles of selective ion transport in channels and pumps. Science. 2005; 310:1461–1465. [PubMed: 16322449]
5. Chen TY. Structure and function of clc channels. Annu Rev Physiol. 2005; 67:809–839. [PubMed: 15709979]
6. Miller C. ClC chloride channels viewed through a transporter lens. Nature. 2006; 440:484–489. [PubMed: 16554809]
7. Dutzler R. A structural perspective on ClC channel and transporter function. FEBS Lett. 2007; 581:2839–2844. [PubMed: 17452037]
8. Jentsch TJ. CLC chloride channels and transporters: from genes to protein structure, pathology and physiology. Crit Rev Biochem Mol Biol. 2008; 43:3–36. [PubMed: 18307107]
9. Stolting G, Fischer M, Fahlke C. CLC channel function and dysfunction in health and disease. Front Physiol. 2014; 5:378. [PubMed: 25339907]
10. Accardi A. Structure and gating of CLC channels and exchangers. J Physiol. 2015; 593:4129–4138. [PubMed: 26148215]
11. Ludwig M, et al. Functional evaluation of Dent's disease-causing mutations: implications for ClC-5 channel trafficking and internalization. Human Genetics. 2005; 117:228–237. [PubMed: 15895257]
12. Peng YJ, et al. Regulation of CLC-1 chloride channel biosynthesis by FKBP8 and Hsp90 beta. Scientific Reports. 2016; 6:32444–32458. [PubMed: 27580824]
13. Robertson JL, Kolmakova-Partensky L, Miller C. Design, function and structure of a monomeric ClC transporter. Nature. 2010; 468:844–847. [PubMed: 21048711]
14. Chadda R, et al. The dimerization equilibrium of a ClC Cl−/H+ antiporter in lipid bilayers. Elife. 2016; 5:e17438. [PubMed: 27484630]
15. Mindell JA, Maduke M, Miller C, Grigorieff N. Projection structure of a ClC-type chloride channel at 6.5 A resolution. Nature. 2001; 409:219–223. [PubMed: 11196649]
16. Dutzler R, Campbell EB, Cadene M, Chait BT, MacKinnon R. X-ray structure of a ClC chloride channel at 3.0 A reveals the molecular basis of anion selectivity. Nature. 2002; 415:287–294. [PubMed: 11796999]
17. Dutzler R, Campbell EB, MacKinnon R. Gating the selectivity filter in ClC chloride channels. Science. 2003; 300:108–112. [PubMed: 12649487]

18. Park E, Campbell EB, MacKinnon R. Structure of a CLC chloride ion channel by cryo-electron microscopy. Nature. 2017; 541:500–505. [PubMed: 28002411]
19. von Heijne G. Membrane-protein topology. Nat Rev Mol Cell Biol. 2006; 7:909–918. [PubMed: 17139331]
20. Bowie JU. Flip-flopping membrane proteins. Nat Struct Mol Biol. 2006; 13:94–96. [PubMed: 16462808]
21. Forrest LR. Structural Symmetry in Membrane Proteins. Annu Rev Biophys. 2015; 44:311–337. [PubMed: 26098517]
22. Bowie JU. Structural biology. Membrane protein twists and turns. Science. 2013; 339:398–399. [PubMed: 23349275]
23. Oesterhelt F, et al. Unfolding pathways of individual bacteriorhodopsins. Science. 2000; 288:143–146. [PubMed: 10753119]
24. Kedrov A, Janovjak H, Sapra KT, Muller DJ. Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy. Annu Rev Biophys Biomol Struct. 2007; 36:233–260. [PubMed: 17311527]
25. Zocher M, et al. Single-Molecule Force Spectroscopy from Nanodiscs: An Assay to Quantify Folding, Stability, and Interactions of Native Membrane Proteins. Acs Nano. 2012; 6:961–971. [PubMed: 22196235]
26. Serdiuk T, et al. YidC assists the stepwise and stochastic folding of membrane proteins. Nat Chem Biol. 2016; 12:911–917. [PubMed: 27595331]
27. Popot JL, Engelman DM. Helical membrane protein folding, stability, and evolution. Annual Review of Biochemistry. 2000; 69:881–922.
28. Engelman DM, et al. Membrane protein folding: beyond the two stage model. Febs Letters. 2003; 555:122–125. [PubMed: 14630331]
29. Bowie JU. Solving the membrane protein folding problem. Nature. 2005; 438:581–589. [PubMed: 16319877]
30. Kim K, Saleh OA. A high-resolution magnetic tweezer for single-molecule measurements. Nucleic Acids Res. 2009; 37:e136. [PubMed: 19729511]
31. Ding F, et al. Single-molecule mechanical identification and sequencing. Nat Methods. 2012; 9:367–372. [PubMed: 22406857]
32. De Vlaminck I, Dekker C. Recent advances in magnetic tweezers. Annu Rev Biophys. 2012; 41:453–472. [PubMed: 22443989]
33. Min D, et al. Mechanical unzipping and rezipping of a single SNARE complex reveals hysteresis as a force-generating mechanism. Nat Commun. 2013; 4:1705. [PubMed: 23591872]
34. Kemmerich FE, et al. Simultaneous Single-Molecule Force and Fluorescence Sampling of DNA Nanostructure Conformations Using Magnetic Tweezers. Nano Lett. 2016; 16:381–386. [PubMed: 26632021]
35. Berghuis BA, Kober M, van Laar T, Dekker NH. High-throughput, high-force probing of DNA-protein interactions with magnetic tweezers. Methods. 2016; 105:90–98. [PubMed: 27038745]
36. Min D, Jefferson RE, Bowie JU, Yoon TY. Mapping the energy landscape for second-stage folding of a single membrane protein. Nat Chem Biol. 2015; 11:981–987. [PubMed: 26479439]
37. Min D, Arbing MA, Jefferson RE, Bowie JU. A simple DNA handle attachment method for single molecule mechanical manipulation experiments. Protein Sci. 2016; 25:1535–1544. [PubMed: 27222403]
38. Jefferson RE, Min D, Corin K, Wang JY, Bowie JU. Applications of Single-Molecule Methods to Membrane Protein Folding Studies. J Mol Biol. 2018; 430:424–437. [PubMed: 28549924]
39. Zakeri B, et al. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc Natl Acad Sci U S A. 2012; 109:E690–697. [PubMed: 22366317]
40. Faham S, Bowie JU. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. J Mol Biol. 2002; 316:1–6. [PubMed: 11829498]
41. Strick TR, Allemand JF, Bensimon D, Bensimon A, Croquette V. The elasticity of a single supercoiled DNA molecule. Science. 1996; 271:1835–1837. [PubMed: 8596951]

42. Gosse C, Croquette V. Magnetic tweezers: Micromanipulation and force measurement at the molecular level. Biophysical Journal. 2002; 82:3314–3329. [PubMed: 12023254]
43. Ribeck N, Saleh OA. Multiplexed single-molecule measurements with magnetic tweezers. Review of Scientific Instruments. 2008; 79:094301–094306. [PubMed: 19044437]
44. Hanggi P, Talkner P, Borkovec M. Reaction-Rate Theory - 50 Years after Kramers. Reviews of Modern Physics. 1990; 62:251–341.
45. Huang J, MacKerell AD Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem. 2013; 34:2135–2145. [PubMed: 23832629]
46. Zhu L, Kaback HR, Dalbey RE. YidC protein, a molecular chaperone for LacY protein folding via the SecYEG protein machinery. J Biol Chem. 2013; 288:28180–18194. [PubMed: 23928306]
47. Kumazaki K, et al. Crystal structure of Escherichia coli YidC, a membrane protein chaperone and insertase. Sci Rep. 2014; 4:7299. [PubMed: 25466392]
48. Dill KA. Theory for the Folding and Stability of Globular-Proteins. Biochemistry. 1985; 24:1501–1509. [PubMed: 3986190]
49. Shen MY, Davis FP, Sali A. The optimal size of a globular protein domain: A simple sphere-packing model. Chemical Physics Letters. 2005; 405:224–228.
50. Paslawski W, et al. Cooperative folding of a polytopic alpha-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112:7978–7983. [PubMed: 26056273]
51. Jefferson RE, Blois TM, Bowie JU. Membrane Proteins Can Have High Kinetic Stability. Journal of the American Chemical Society. 2013; 135:15183–15190. [PubMed: 24032628]
52. Blommel PG, Fox BG. A combined approach to improving large-scale production of tobacco etch virus protease. Protein Expression and Purification. 2007; 55:53–68. [PubMed: 17543538]
53. Sreerama N, Woody RW. On the analysis of membrane protein circular dichroism spectra. Protein Science. 2004; 13:100–112. [PubMed: 14691226]
54. Wu EL, et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem. 2014; 35:1997–2004. [PubMed: 25130509]
55. Steinbach PJ, Brooks BR. New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. Journal of Computational Chemistry. 1994; 15:667–683.
56. Essmann U, et al. A Smooth Particle Mesh Ewald Method. Journal of Chemical Physics. 1995; 103:8577–8593.
57. Feller SE, Zhang YH, Pastor RW, Brooks BR. Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. Journal of Chemical Physics. 1995; 103:4613–4621.
58. Shaw, DE., et al. Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer; Sc14: International Conference for High Performance Computing, Networking, Storage and Analysis; 201441-53
59. Lippert RA, et al. Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. Journal of Chemical Physics. 2013; 139:164106–164116. [PubMed: 24182003]
60. Hess HH, Derr JE. Assay of inorganic and organic phosphorus in the 0.1–5 nanomole range. Anal Biochem. 1975; 63:607–613. [PubMed: 1122033]

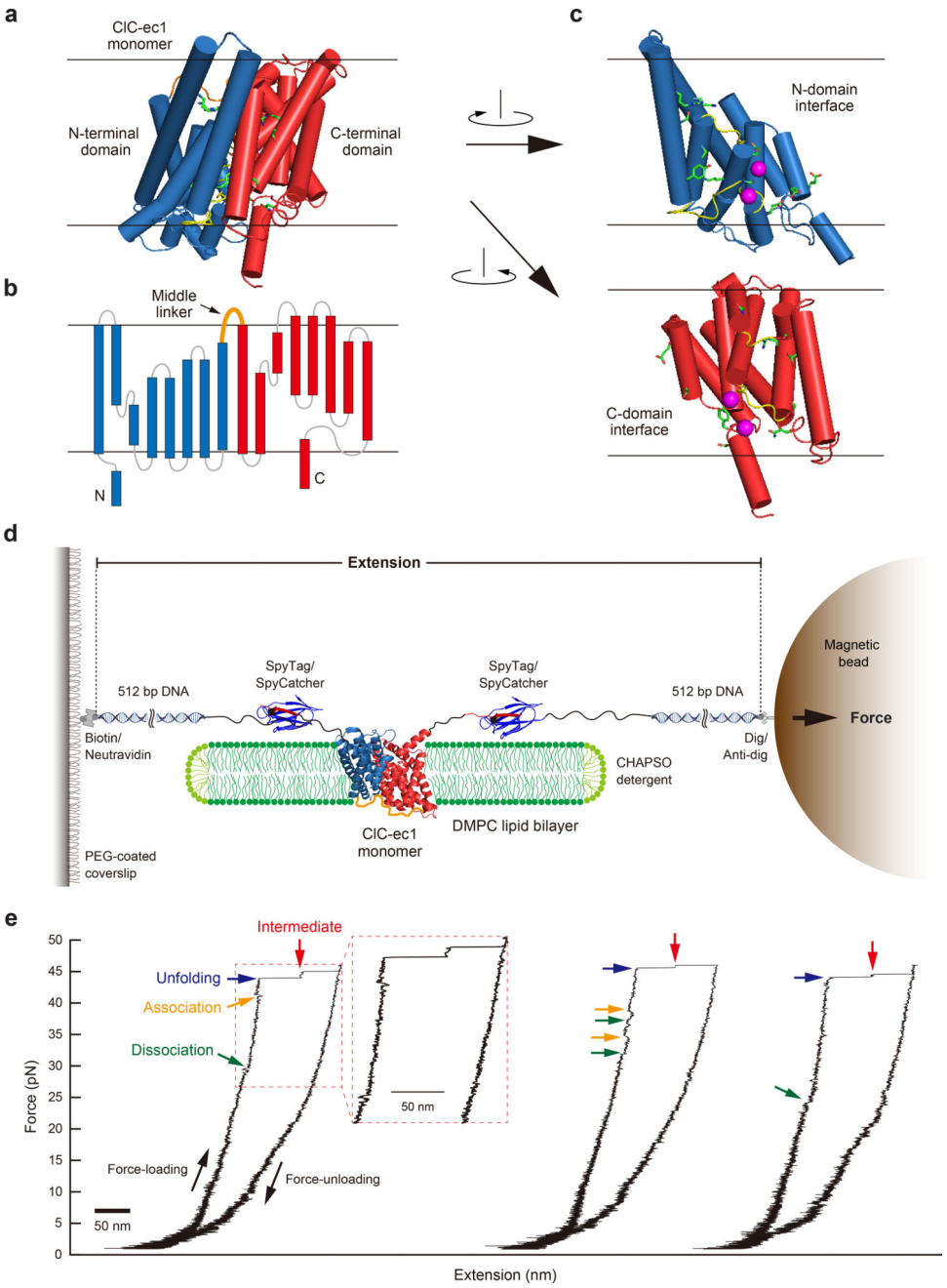

Figure 1. ClC-ec1 structure and single-molecule forced unfolding experiments

(a) Structure of the ClC-ec1 monomer. The N- and C-terminal domains are shown in blue and red, respectively. Re-entrant loops are shown in yellow. Polar residues at the domain interface are shown in stick representation. Carbon, oxygen, nitrogen are colored green, red, blue respectively. **(b)** Topology of the ClC-ec1 monomer. The topology with respect to the membrane is inferred from the known structure[16]. The α-helices are drawn as rectangles. The linker between domains is colored orange. **(c)** Domain interfaces on the N-terminal half (upper panel) and C-terminal half (lower panel). Re-entrant loops in the interface are shown in yellow. Polar residues at the domain interface are shown in stick representation. Carbon,

oxygen, nitrogen are colored green, red, blue respectively. Cl^- ions are colored in magenta. **(d)** Schematic diagram of the single-molecule pulling experiment using magnetic tweezers. A monomeric ClC-ec1 is ensconced in a bicelle disc. The bicelle-incorporated ClC-ec1 is attached to 512 bp DNA handles through covalent SpyTag/SpyCatcher complexes (Supplementary Fig. 2 and Online Methods). The ClC-ec1/DNA/bicelle hybrid construct is tethered to a glass surface and a magnetic bead via biotin-netravidin and digoxigenin-antidigoxigenin linkages. Force direction is denoted by the black arrow. **(e)** Representative force-extension traces from single ClC-ec1 transporters (from total n=112 independent experiments). The initiation of unfolding and the intermediate state between the two jumps are denoted by blue and red arrows, respectively. Apparent dissociations and associations between the N- and C-terminal domains are denoted by green and yellow arrows, respectively. In each curve, the left and right traces correspond to the force-loading and force-unloading phases. A 50 nm extension scale bar is shown.

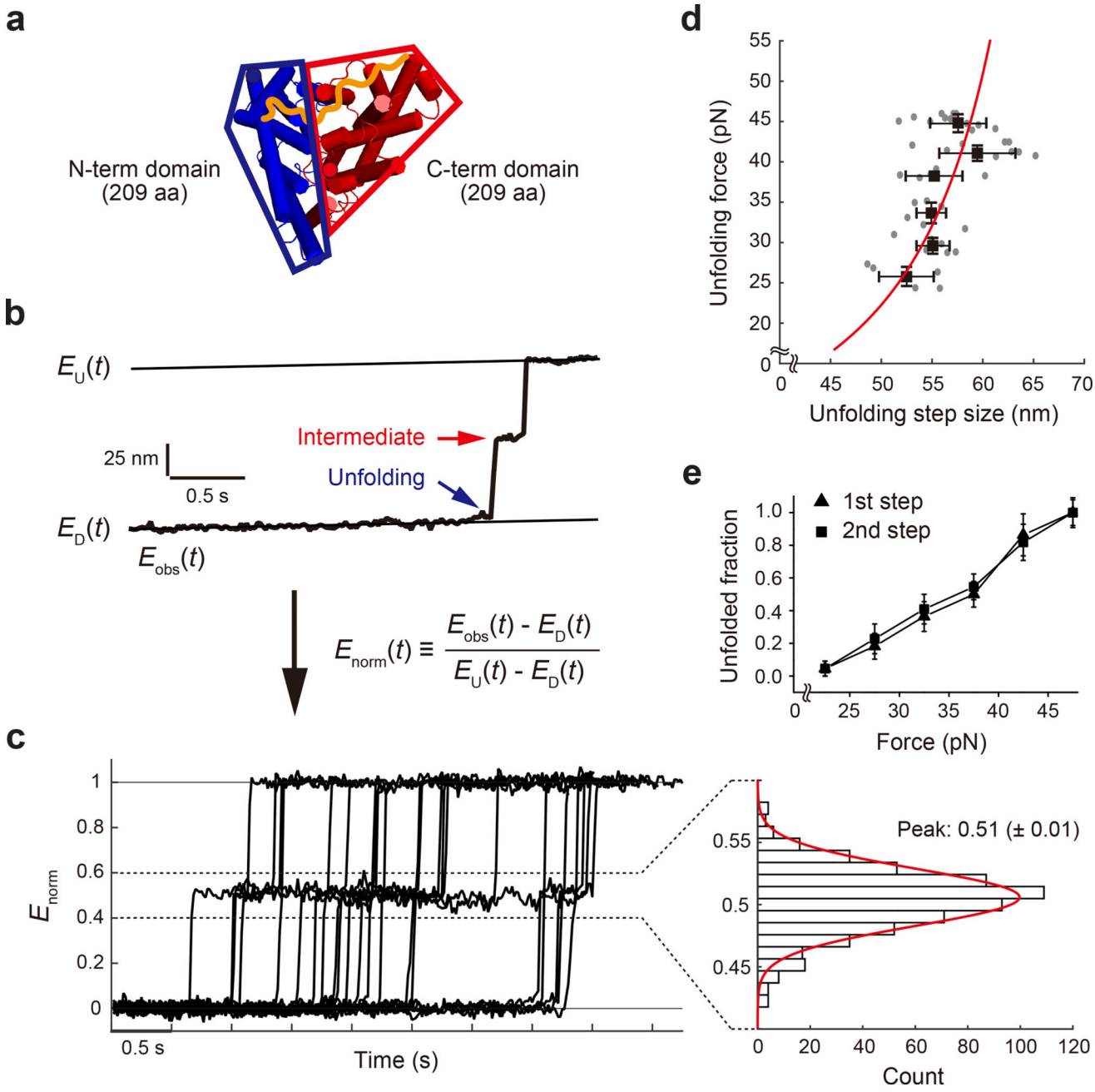

Figure 2. Separate unfolding of N- and C-terminal domains

(a) Structure schematic of the ClC-ec1 structure viewed from the extracellular side. The N- and C-terminal domains have approximately the same number of amino acids. **(b)** Representative real-time extension trace for the ClC-ec1 (from n=22 independent experiments). The black fitting lines are labeled as E_D (t) for the dissociated state (after the pre-transition apparently involving domain separation) and E_U(t) for the unfolded state. The experimental trace is labeled as E_{obs}(t). The normalized extension is defined by E_{norm}(t)≡ (E_{obs}(t)−E_D(t))/(E_U(t)−E_D(t)). Time and extension scale bars are shown. **(c)** Overlapped E_{norm}(t) for n=22 independent traces and count histogram of E_{norm}(t) for the intermediate

pause between the two steps. The count histogram at the right was constructed by polling the values of $E_{norm}(t)$ at the intermediate pause. The peak of Gaussian fitting is 0.51 (±0.01, SEM), indicating that half the protein is unfolded at the pause. **(d)** Unfolding force *vs* unfolding step size. The black dots are the average values (n=5,6,7,4,8,14 for each dot from low to high force) and the error bars represent the standard deviations. A scatter plot of every data point is overlaid (gray dots). The red line is a fit using a worm-like chain model, yielding a persistence length of 0.44 nm (±0.06 nm, SEM, 95% CI) with a contour length of 209 residues times 0.38 nm/residue for each domain (see Online Methods for more details). **(e)** Unfolded fractions as a function of force (from n=22 independent experiments). Each unfolded fraction at a certain force was calculated by integrating the unfolding force distribution up to the force. The plots for the first and second unfolding steps are denoted as triangle and square respectively.

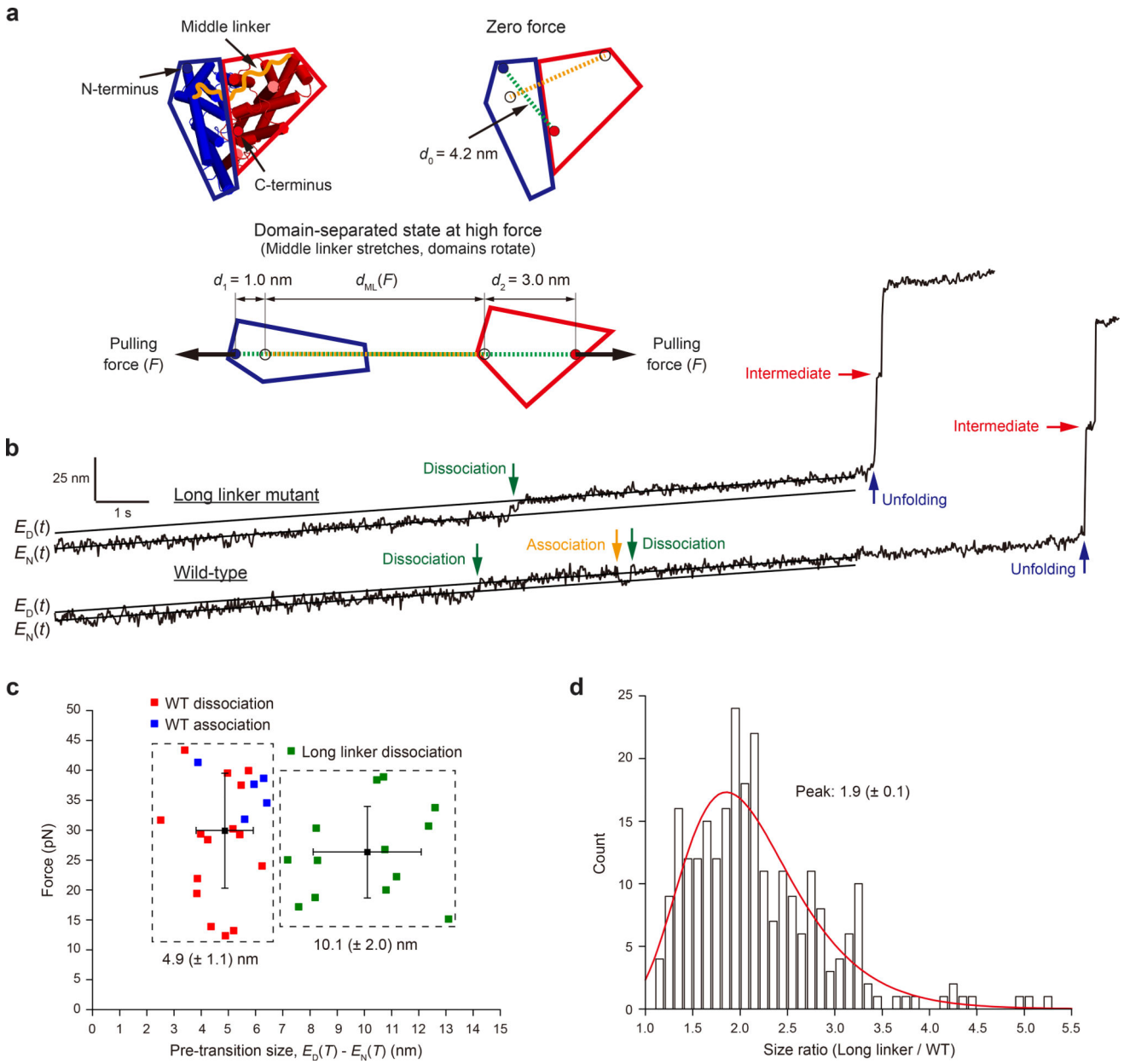

Figure 3. A longer linker between domains increases the size of the pre-transition

(a) Schematic of pulling geometry viewed from the extracellular side and anticipated global structural changes upon domain separation. The middle linker stretches and the two domains rotate during the domain separation event at high force. d_0 is the initial distance between the pulling end points, $d_{ML}(F)$ is the end-to-end distance of the stretched middle linker after domain separation, and d_1 and d_2 are the distance changes due to subunit rotation after separation. $d_{ML}(F) \gg d_1 + d_2 - d_0$ since $d_1 + d_2 - d_0$ = ~0 nm, so the pre-transition step size $\Delta E \sim d_{ML}(F)$. The model predicts that the pre-transition extension for the long linker mutant should be ~2 times the extension of the wild-type protein ($E_{ratio} \sim 2$, see Online Methods for more details). **(b)** Extension traces as a function of time for the wild-type and long linker constructs. The native and dissociated state traces were fit to a line (see Online Methods for

more details). The fitted lines for the native and dissociated state extensions are defined as $E_N(t)$ and $E_D(t)$. The pre-transition size is calculated by $E_D(T)-E_N(T)$ at the time T when the transition occurs. Extension and time scale bars are shown. **(c)** Scatter plot of pre-transition sizes for the wild-type and long linker mutant. The pre-transition size is 4.9 (±1.1, SD) nm at 29.9 (±9.7, SD) pN for the wild-type (n=20 from 13 independent experiments) and 10.1 (±2.0, SD) nm at 26.3 (±7.8, SD) pN for the mutant (n=13 independent experiments). The measure of centers is defined as the mean. For the wild-type protein we observed 13 force ramps from the folded protein and all showed dissociation events. The total number of observed dissociation and association events was 15 and 5, respectively. **(d)** Count histogram of pre-transition size ratio (long linker/wild-type). The size ratios were calculated for all possible combinations between the long-linker mutant and wild-type (n=260). The peak of the histogram is estimated as 1.9 (±0.1, SEM) by fitting with lognormal distribution.

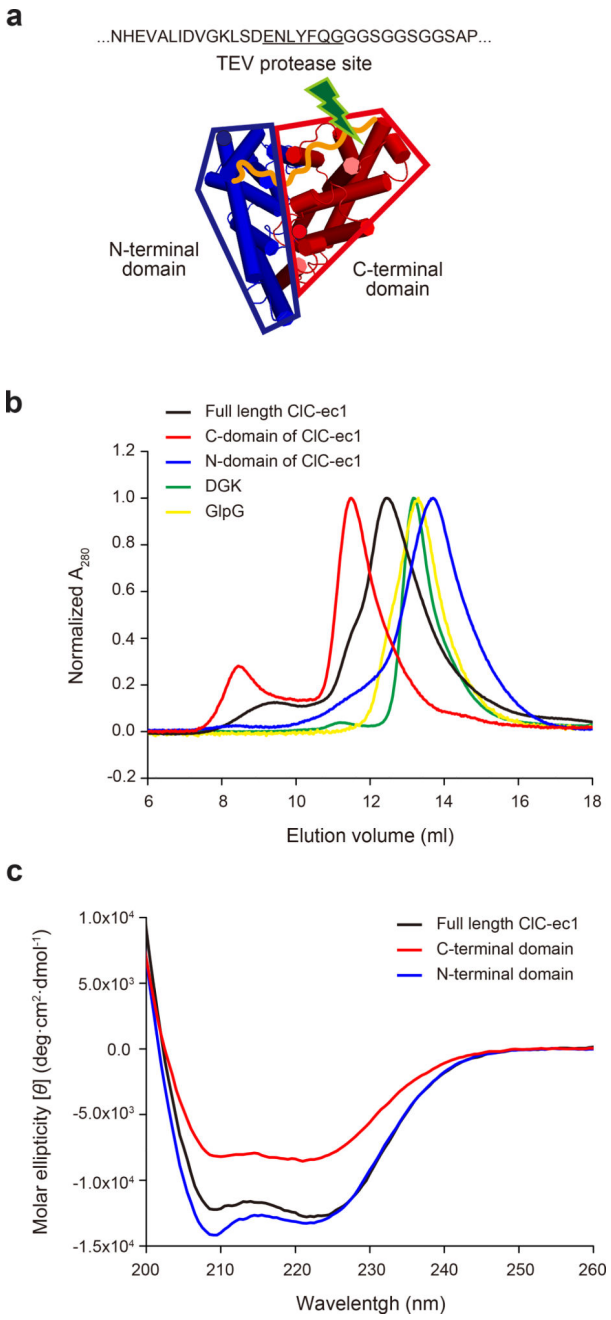

Figure 4. Characterization of the isolated N- and C-terminal domains

(a) TEV protease site inserted into the linker between domains. **(b)** Size-exclusion chromatography elution profiles for full-length ClC-ec1 (black), the C-domain of ClC-ec1 (red), the N-domain of ClC-ec1 (blue), *E. coli* DGK (green), and *E. coli* GlpG (yellow). **(c)** Circular dichroism spectra for full-length ClC-ec1 (black), the C-domain of ClC-ec1 (red), and the N-domain of ClC-ec1 (blue). Supplementary Table 1 provides the estimated fraction of secondary structure for each construct.

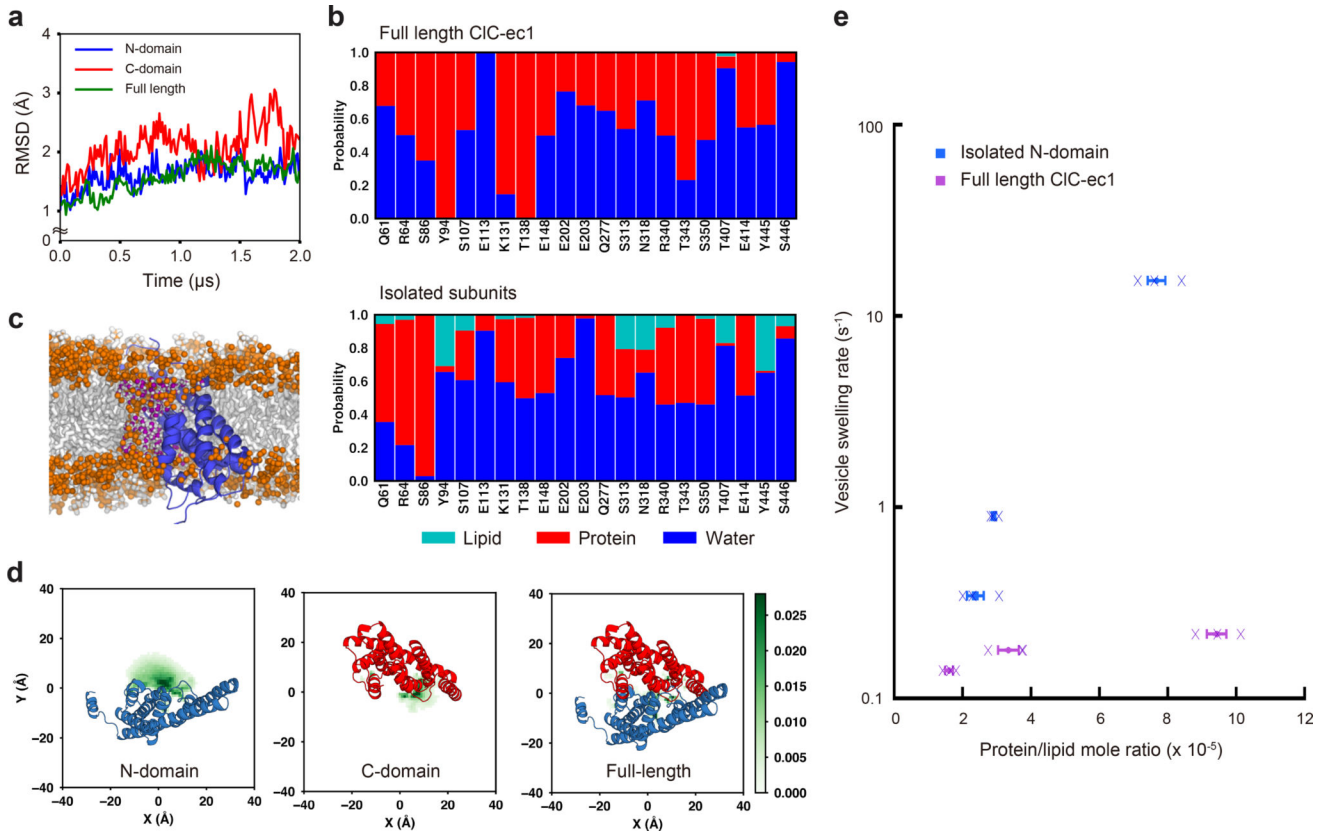

Figure 5. MD simulations and vesicle swelling assay

(a) Cα RMSD profile of the transmembrane region with respect to the crystal structure as a function of simulation time. **(b)** Probability of hydrogen bond partners of the N-C interface residues in full-length ClC-ec1 (top panel) and isolated subunits (bottom panel). The hydrogen bond partners in each MD snapshot were averaged on the whole trajectory to obtain the probability (Online Methods). **(c)** A snapshot from the N-terminal domain simulation. Lipid phosphate atoms are shown as orange spheres, chloride ions as green spheres, and oxygen atoms in water molecules as purple sticks. **(d)** Density of water molecules (water molecules per $Å^3$) in the membrane region. A water molecule was counted if its position along the *Z*-axis was within 5 Å from the bilayer center. **(e)** Vesicle swelling rates after osmotic shock as a function of protein/lipid ratio. The points represent the average of triplicate experiments. The × symbols indicate the actual data points that were averaged. Error bars on rate (very small) represent the fitting errors of rate measurements on proteoliposomes containing full-length ClC-ec1 or the isolated N-domain (average of n=10 independent experiments). The errors in the protein/lipid ratio are the propagated errors from the standard curve fitting error of SDS-PAGE band intensity measurements and the standard deviation of triplicate lipid phosphate quantification assays (n=3 independent experiments).

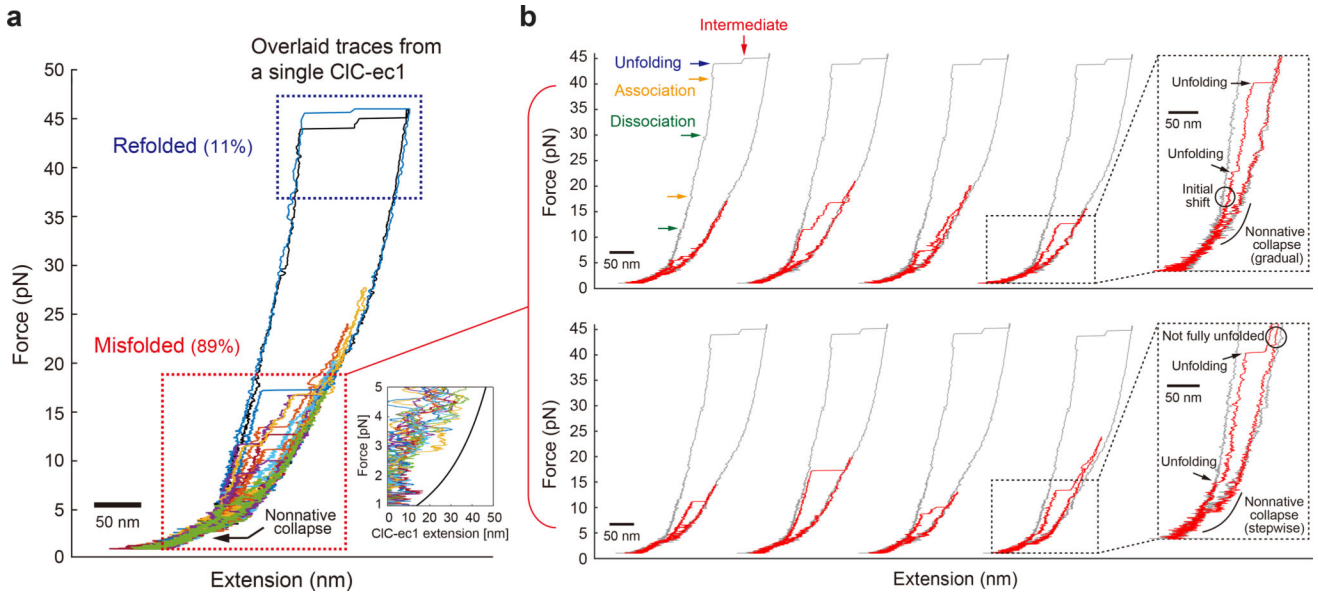

Figure 6. Stable off-pathway misfolded states

(a) Overlaid force-extension curves from a single ClC-ec1 (n=13 independent experiments). Every trace is shown in a different color. The black trace is the initial unfolding event. The top blue dashed box highlights a representative unfolding trace for a refolded protein (11% of total n=96 independent experiments). The bottom red dashed box highlights traces from misfolded proteins (89% of total n=96 independent experiments). The bottom inset shows the ClC-ec1 extension (subtracting extension of DNA handles and polypeptide linkers) which is determined by subtracting the force-loading trace from the force-unloading trace. The black line in the inset is the expected line for unfolded ClC-ec1 given the native secondary structure (see Supplementary Fig. 8 for more details). **(b)** Force-extension curves of misfolded states (red). For comparison, a trace for the refolded protein is also shown as gray. The rightmost insets zoom in on the misfolded state transitions.