Unfolding and identification of membrane proteins *in situ*

Nicola Galvanetto[a,1,], Zhongjie. Ye[a,1], Arin Marchesi[b], Simone Mortal[a], Sourav Maity[a,c], Alessandro Laio[a,d], Vincent Torre[a,e]*

[a] International School for Advanced Studies (SISSA); Trieste, Italy

[b] Nano Life Science Institute (WPI-NanoLSI); Kanazawa University, Kakuma-machi, Kanazawa, Japan

[c] Moleculaire Biofysica, Zernike Instituut, Rijksuniversiteit Groningen; Groningen, Netherlands

[d] The Abdus Salam International Centre for Theoretical Physics (ICTP); Trieste, Italy

[e] Institute of Materials (IOM-CNR), Area Science Park; Trieste, Italy

[1] These authors contributed equally to this work

[*] Corresponding author. Email: nicola.galvanetto@sissa.it

Keywords: membrane proteins | atomic force microscopy | single-molecule force spectroscopy | protein folding | protein identification

Abstract: Single-molecule force spectroscopy (SMFS) uses the cantilever tip of an AFM to apply a force able to unfold a single protein. The obtained force-distance curve encodes the unfolding pathway, and from its analysis it is possible to characterize the folded domains. SMFS has been mostly used to study the unfolding of purified proteins, in solution or reconstituted in a lipid bilayer. Here, we describe a pipeline for analyzing membrane proteins based on SMFS, that involves the isolation of the plasma membrane of single cells and the harvesting of force-distance curves directly from it. We characterized and identified the embedded membrane proteins combining, within a Bayesian framework, the information of the shape of the obtained curves, with the information from Mass Spectrometry and proteomic databases. The pipeline was tested with purified/reconstituted proteins and applied to five cell types where we classified the unfolding of their most abundant membrane proteins. We validated our pipeline by overexpressing 4 constructs, and this allowed us to gather structural insights of the identified proteins, revealing variable elements in the loop regions. Our results set the basis for the investigation of the unfolding of membrane proteins in situ, and for performing proteomics from a membrane fragment.

Introduction

Mapping and recovering the structure of membrane proteins in their native environment, *i.e.* embedded in a native cellular membrane and at room temperature is a challenging aim: indeed the most successful tool for this purpose is cryo electron microscopy, but it requires frozen samples, and precise single particle measurements can be achieved only with purification.

In fact, much of what we know about the mechanics and the structure at room temperature of cell membranes, of polymers (Carrion-Vazquez et al., 1999; Sarkar et al., 2005) and of membrane proteins (Al-Rekabi and Contera, 2018; Casuso et al., 2012; García-Sáez et al., 2007) comes from atomic force microscopy (AFM) which can operate in liquid environment. Single Molecule Force Spectroscopy (SMSF) uses the tip of an AFM to apply a stretching force to unfold a single protein. SMFS provides a 1D force-distance (F-D) curve which encodes the unfolding pathway so that from the analysis of the sequence of force peaks it is possible to identify the folded domains and their variability (Engel and Gaub, 2008). It has been recently confirmed that from the 1D amino-acid sequence of a protein it is possible to accurately determine its 3D structure (Jumper et al., 2021), therefore it is tempting to continue exploring which information can be recovered from the 1D F-D curves provided by SMFS.

Hitherto SMFS has been mostly used to study the unfolding mechanics of purified proteins, in solution or reconstituted in lipid bilayers. The information that is possible to obtain (e.g. mechanical stability(Sumbul et al., 2018a; Thoma et al., 2015), structural heterogeneity(Hinczewski et al., 2016)) depends on the physical and chemical properties of the cell membrane(Maity et al., 2015; Thoma et al., 2018). However, native membranes are extremely complex, hosting hundreds of different proteins in an environment which is highly heterogeneous even at the single-cell level, and typically different in different cells.

In this manuscript we aim to bridge the gap between all these recent breakthroughs by attempting to identify and recover structural information of membrane proteins embedded in biological samples, i.e. in their native environment and at room temperature.

We describe a complete pipeline, including the experimental methods and the data analysis, which represent a first step forward in this direction. The pipeline allows to identify membrane proteins obtained from SMFS on single cells native membranes. Firstly, we developed a technique to extract a piece of the membrane suitable for SMFS so to obtain millions of F-d curves from native biological membranes. Secondly, we used an unsupervised clustering procedure to detect clusters of similar unfolding curves. Finally, we implemented a Bayesian meta-analysis using information from Mass Spectrometry and protein structure databases that allows to identify a limited list of candidates of the unfolded proteins.

We first characterized the unfolding of purified membrane proteins reconstituted *in vitro* in lipid bilayers, then we focused on the native cell membranes of hippocampal neurons, dorsal root ganglia (DRG) neurons, neuroblastoma cells and the plasma and disc membrane of rod outer segments. Our protocol — even if it cannot compete with the identification accuracy of mass spectroscopy performed on thousands of cells, or with the data quality of a SMFS experiment performed on a single purified protein in an artificial membrane — accomplishes a different task, and in each of these cases it allows gathering biologically meaningful information that cannot be captured otherwise, even at the single cell level.

Results

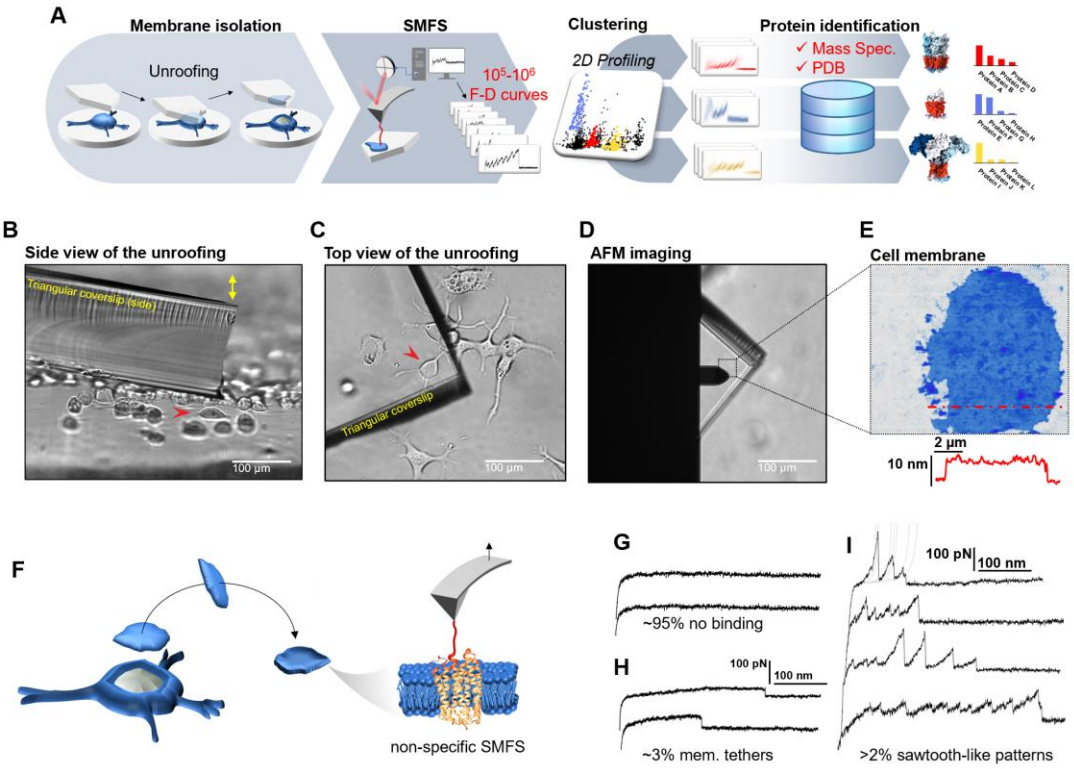


Fig. 1. Experimental method for single-cell membrane isolation and protein unfolding. *(**A**), workflow of the method in four steps: isolation of the apical membrane of single cells; AFM-based protein unfolding of native membrane proteins; identification of the persistent patterns of unfolding and characterization of the population of unfolding curves; Bayesian protein identification thanks to mass spectrometry data, Uniprot and PDB. (**B**), side view and (**C**), top view of the cell culture and the triangular coverslip approaching the target cell (red arrow) to be unroofed. (**D**), positioning of the AFM tip in the region of unroofing. (**E**), AFM topography of the isolated cell membrane with profile. (**F**), cartoon of the process that leads to SMFS on native membranes. Examples of F-D curves of (**G**), no binding events; (**H**), constant viscous force produced by membrane tethers during retraction; (**I**), sawtooth-like patterns, typical sign of the unfolding of a membrane protein.*

Unfolding proteins from isolated cell membranes

The first ingredient of our pipeline is an unroofing method(Galvanetto, 2018) to isolate the apical part of single cell membranes containing membrane proteins with negligible contamination of cytoplasmic proteins (Fig. S1–S3). We sandwiched a single cell between two glass plates, the culture coverslip and another plate mounted on the AFM itself (see Fig. 1B-C, triangular coverslip). The triangular coverslip is coated with polylysine which favors membrane adhesion. A rapid separation of the plates, permits the isolation of the apical membrane of the cell (see Fig. 1D-E). The method is reliable (n=105, ~80% success rate) with cell types grown on coverslips. With non-adherent cells, like freshly isolated rods, membranes were isolated with a lateral flux of medium(Clarke et al., 1975) (see Methods). During the unroofing process, we verified the absence of any adsorption of cytoplasmic material like tubulin, actin, mitochondria or free cytoplasmic proteins from the unroofed membrane patch where only the membrane proteins that are hold by the lipids are present (see Fig. S2–S4).

We verified with the AFM (Fig. 1F) that the isolated membrane patches have a height of 5-8 nm with roughness in the order of 1 nm. Then, we performed conventional SMFS(Oesterhelt et al., 2000) collecting more than two millions of F-D traces from all the samples (hippocampal neurons, DRG neurons, neuroblastoma, rods and rod discs, see below). Among the obtained curves, ~95% shows no binding (Fig. 1G), ~3% shows plateau ascribable to membrane tethers (Fig. 1H), while the remaining ~2% displays the sawtooth-like shape that characterizes the unfolding of proteins(Oesterhelt et al., 2000) (Fig. 1I) with each tooth fitting the worm-like chain (WLC) model with a persistence length of ~0.4nm (Fig. S5)(Li et al., 2002). These ~50,000 traces are dramatically heterogeneous, differing in total length, number of peaks, maximum force, etc. This is not surprising, since cell membranes contain a large number of different proteins, each hosted in different local environment. In the following we describe a procedure which allows, to some extent, disentangling this bundle.

In SMFS it is assumed that the binding between the cantilever and the protein occurs either at the C- or at the N-terminus, and that the protein is fully unfolded by the tip. However, some traces suggest that also other events take place: i) the simultaneous attachment of two or more proteins to the tip(Walder et al., 2017), ii) the incomplete unfolding of the attached protein(Tanuj Sapra et al., 2006), iii) the binding of the AFM tip to a loop of the protein instead of to a terminus end (Fig. S6, C to F).

i) Attachment of multiple proteins (Fig. S6D): the resulting F-D curves will not have a recurrent pattern; if two proteins form a complex, the resulting spectrum is the sum of the two individual spectra: that causes deviations of the measured persistence length (Fig. S7). The simultaneous unfolding of multiple proteins is also characterized by force changes and varying persistence length (Fig. S6D, G and S7).

ii) Incomplete unfolding of the protein (Fig. S6E): if the tip prematurely detaches from the terminus, the F-D curve has similar but shorter pattern compared to a complete unfolding (Fig. S6C). The fraction of curves that prematurely detaches has been reported to be ~23% of the fully unfolded protein(Tanuj Sapra et al., 2006).

iii) Binding of the AFM tip to a loop (Fig. S6F): this case is equivalent to the attachment of multiple proteins. However, if the attachment of the cantilever tip to a loop occurs with some consistency, we will obtain a recurrent pattern with the features described in case i) (deviation of persistence length during intersection, 2 major levels of unfolding force, see Fig. S7).

Therefore, there is no doubt that a large fraction of F-D curves will not represent the proper unfolding of a single protein. But inspired by the successes of single particle Cryo-EM, which can produce atomic-resolution structures, despite selecting less than 20% of the protein images(Yi et al., 2019), we attempted a similar approach on SMFS data, under the hypothesis that 'bad' events are likely to produce *non-recurrent* patterns of unfolding.

The key tool to recognize recurrent patterns is an automated clustering method developed in our group(Ilieva et al., n.d.) based on density peak clustering(Rodriguez and Laio, 2014) (Fig. S11, see Supplementary Methods for a detailed description of the algorithm). In short, this approach detects statistically dense F-D patterns occurring often in the sample (clusters). It is based on a previously established 'similarity distance'(Marsico et al., 2007) in the context of SMFS and, remarkably, it does <u>not require to pre-set neither the shape nor the number of clusters</u>.

Benchmarking the analysis with mixtures of purified proteins reconstituted in lipid bilayers

We first tested the consistency of our pipeline with three known membrane proteins. We reconstituted in lipid bilayers(Shen et al., 2013) highly purified Channelrhodopsin (ChR1)(Nagel et al., 2002), Bacteriorhodopsin (bR)(Oesterhelt et al., 2000) and the cyclic adenosine monophosphate (cAMP) gated channel SthK(Marchesi et al., 2018) (Fig. 2A), and we performed SMFS experiments.

bR, ChR1 and SthK were reconstituted *in vitro* one at a time. Since we did not have control of the protein orientation in the bilayer, we expected 2 main F-D patterns, and therefore two clusters for each sample: one representing the unfolding of the protein from the N-terminus and the other from the C-terminus.
The clustering algorithm identified 2 almost indistinguishable clusters for bR (see Supplementary Methods), and the result is in agreement with the fact that the unfolding of bR from the two termini is almost symmetric as previously reported (Kessler and Gaub, 2006) (Fig. 2D). In the case of ChR1 we obtained 3 clusters where the second cluster — cluster C2 — represents the partial unfolding of ChR1 from the C terminus because the peaks of C2 match perfectly the first three peaks of the full unfolding (cluster C1), while cluster C3 perfectly match the unfolding from the N terminus (Fig. 2B). Sthk generated 2 clusters, cluster S1 and S2 (Fig. 2C), as expected (all the clusters are shown in Fig. S8). For all the three proteins, we observed that: (i) the value of Lc_{max} — i.e. the estimated length of the segment of a.a. stretched in our SMFS experiments — is within 5% equal to the total length of the protein minus the free length of the N-terminus (C-terminus) domain if the cantilever tip attached to the C-terminus (N-terminus) (see Fig. 2B, C and Fig. S8B); and that (ii) the force peaks of the clusters of F-D curves colocalize with the unfolded regions or loops of their structures (see Fig. S8, A and C).
Next, we simultaneously reconstituted two proteins in lipid bilayers (see Methods) obtaining the same clusters associated with the unfolding patterns observed in the single protein reconstitution (Fig. S9). We finally reconstructed all the three proteins together, with different relative abundances: 1:1:1 and 1:7:20 respectively for bR, ChR1 and SthK. We observed that the number of traces per cluster scaled approximately with the abundance of the preparation (Fig. S10).

The histogram of the maximal contour length (Lc max) of all the F-D curves (Figure 2E) encodes the protein content of the sample and allows a rapid detection of the presence and of the abundance of a given protein in sub-femtogram samples, providing also a qualitative indication on the concentration of a given protein. The success of this preliminary *in vitro* investigation prompted us to move to native membranes.

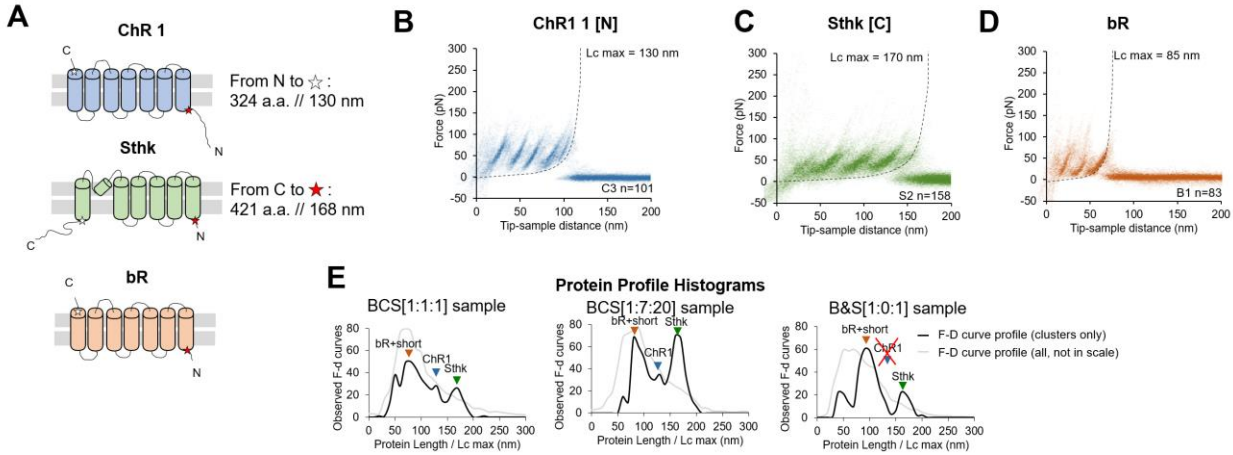


Fig. 2. Unfolding of reconstituted mixtures of membrane proteins. *(**A**): scheme of the structure of the three purified proteins used in the in vitro preliminary step: ChR1, Sthk and bR (cylinders represent alpha helices). (**B**): superimposition of 101 unfolding curves (density plot) of the full unfolding of ChR1 from the N terminus. (**C**): density plot of the full unfolding of Sthk from the C terminus. . (**D**): density plot of the full unfolding of bR. (**E**): protein profile, i.e. the histogram of the maximal contour length (Lc max) of all the F-D curves in the clusters (black line), and of all the F-D curves collected from the sample (grey line, not in scale), from samples with mixtures of bR, ChR1 and Sthk with a relative abundance of 1:1:1, 1:7:20 and 1:0:1 (bR-Sthk only). Arrows indicate where the F-D curves of (**B**), (**C**) and (**D**) accumulate in*

the histogram. The peak at ~90 nm indicated by bR+short also contains the shorter clusters of ChR1 and SthK shown in Fig. S8, that also appear in the mixtures.

Clustering SMFS data from native membrane proteins.

Next, we applied our pipeline to DRG, hippocampal neurons, neuroblastoma, rod outer segments and rod discs membranes (Fig. 3A-E) and we obtained 301,654 curves from the hippocampal neurons, 413,468 from DRG neurons, 623,482 from neuroblastoma, 386,128 from rods and 221,565 from rod discs. From these F-D traces we applied our clustering procedure that identified 15, 10, 11, 8 and 5 statistically dense pattern of unfolding present in the datasets (Fig. 3F-J). These traces show that clean unfolding patterns can be detected even in patches of native membranes.

We identified four major classes of clusters: *Short curves with increasing forces:* clusters DRG12, H5, H8 and R3 shows repeated peaks (ΔLc 10-20 nm, distance between consecutive peaks) of increasing force up to 400 pN. *Long and periodic curves:* R6, H7 and DRG10 display periodic peaks of ~100 pN and with a ΔLc of 30-40 nm whose unfolding patterns are similar to what seen when unfolding LacY(Serdiuk et al., 2016). *Conventional short curves:* the majority of the identified clusters like DRG1, H3, R8 and all clusters from the rod discs have F-D curves with a total length less than 120 nm with constant or descending force peaks. These F-D curves share various features with the opsin family unfolded in purified conditions(Engel and Gaub, 2008), *e.g.* a conserved unfolding peak at the beginning (at contour length < 20 nm) associated to the initiation of the denaturation of the protein. We found also *unconventional* clusters such as DRG7, DRG8 and R7: DRG8 exhibits initial high forces and with variable peaks followed by more periodic low forces; while cluster R7 has a conserved flat plateau at the end of the curve of unknown origin.

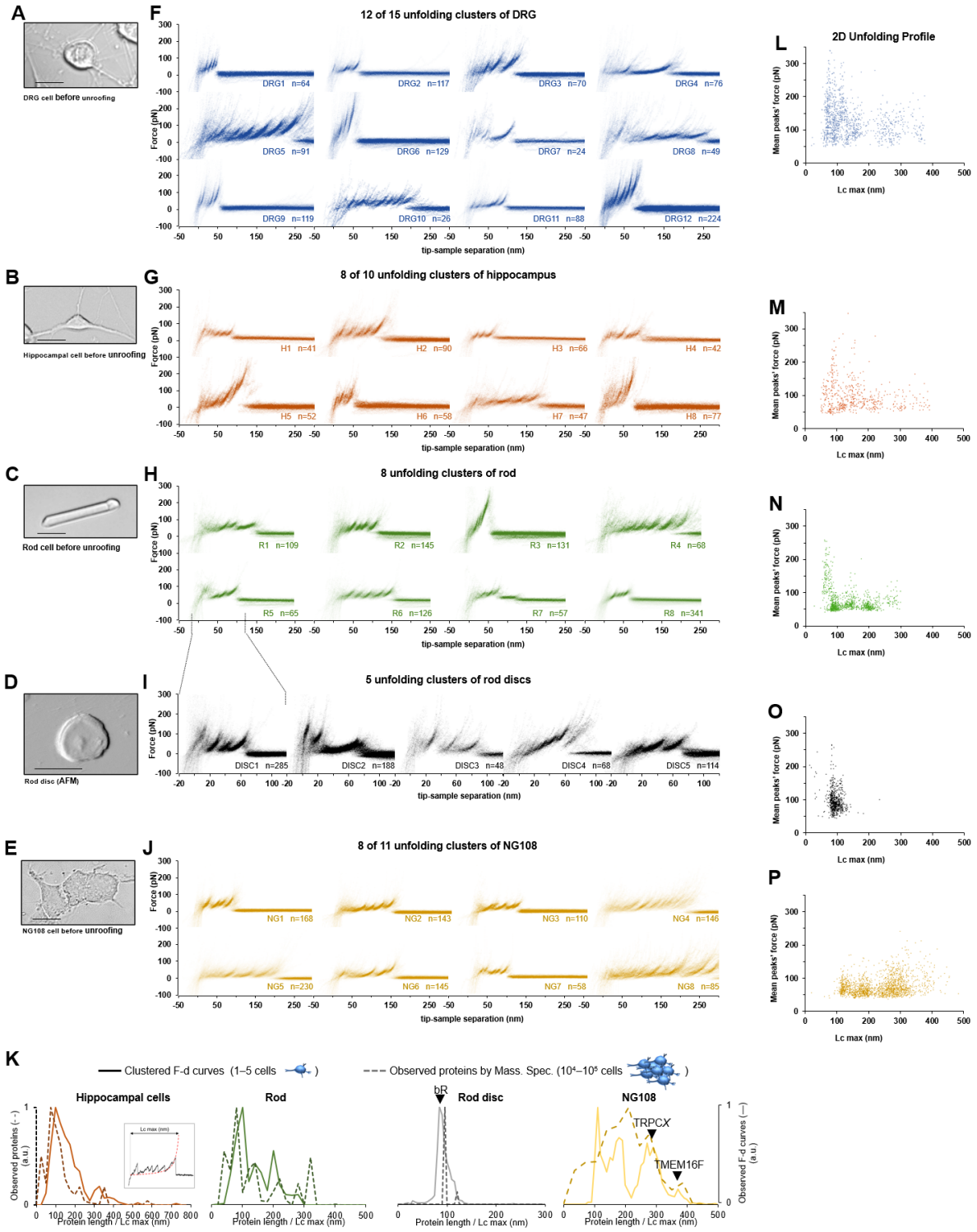


Fig. 3. Unfolding clusters in native cell membranes. *Bright field image of (**A**): dorsal root ganglia neuron; (**B**): hippocampal neuron; (**C**) rod before unroofing (scale bar 15 µm). (**D**): AFM error image of an isolated disc (scale bar 1 µm), (**E**): NG 108 cells. (**F, G, H, I, J**): examples of obtained clusters from the native membranes shown as density plots, that is superposition of the n unfolding curves agglomerated and colored as a heat map. (**K**): comparison of the Protein profiles obtained with mass*

spectrometry vs. the F-D curves obtained with SMFS shows a good correlation. The protein profile from mass spectrometry is the normalized histogram of the total number of proteins found in the sample, considering their abundance. The protein profile from SMFS is the histogram of the maximal contour length (Lc max) of all the F-D curves selected with our clustering procedure. ***(L, M, N, O, P):*** *representation of all the clustered F-D curves in 2D: x-axis is the maximum contour length; y-axis is the average unfolding force (DRG: n = 1255; hippocampus: n = 563; rod: n = 1039; disc: n = 703, NG108: n = 1591).*

A compact representation of clustered F-D curves becomes more important in native samples where the information stored in the SMFS data is more complex (see Fig. 3L-P). The cancerous NG108-15 cells have few short and not particularly stable membrane proteins compared to neural cells which instead have many clusters of proteins under 100 nm that unfold even above 200 pN. On the contrary, NG cells have a higher fraction of long and stable proteins than neurons.

Furthermore, in native membranes we can directly compare the expected protein profile obtained from mass spectrometry (usually obtained from thousands of cells, see Methods) with the F-D curves profile of just few cells (Fig. 3K) which shows a good agreement in particular in the region above 100 nm of protein length. These results suggest that protein length detection of SMFS — restricted on clustered traces according to our protocol — provides a semi-quantitative proxy for mass detection at the single cell membrane level.

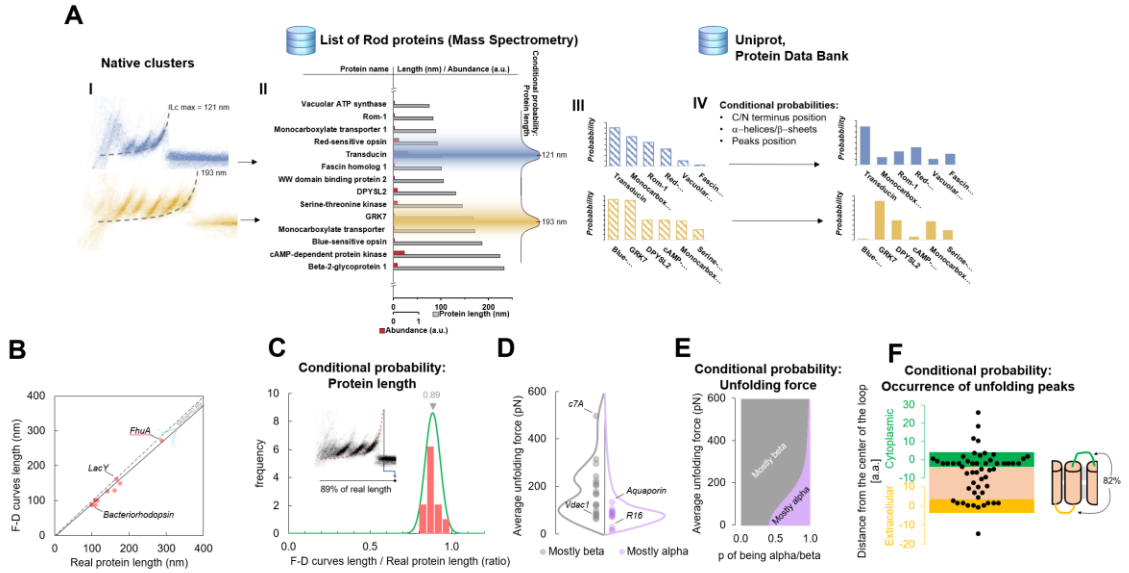


Fig. 4: Bayesian identification and conditional probabilities. ***A****, logical workflow of the Bayesian steps: selection due to total length and abundance (mass spectrometry), refinement with structural and topological information (PDB and Uniprot).* ***B****, Comparison of the real length of the protein vs. the measured maximal contour length of the F-D curves in 14 SMFS experiments on membrane proteins.* ***C****, Conditional probability of the observed maximal length of the clusters obtained from* ***B****.* ***D****, Comparison of the force necessary to unfold beta sheets and alpha helices in 22 SMFS experiments.* ***e****, Conditional probability of the observed unfolding forces obtained from* ***D****.* ***F****, Conditional probability for the occurrence of unfolding peaks extracted from SMFS literature (see methods). Unfolding peaks occurs most likely in the loops (82%) than in transmembrane domains (n_peaks=54, from 11 SMFS experiments of different membrane proteins). The points in the green (yellow) region represents unfolding peaks occurred in a cytoplasmic (extracellular) loop, the point in the pink area occurred in a transmembrane domain. The*

points above the green and below the yellow regions occurred in cytoplasmic and extracellular domains, respectively. The scale is approximate because in rare occasion loops are longer than 10 nm.

Molecular identification of detected clusters

Having identified clusters of F-D curves from native membranes that approximates the overall population of membrane proteins (Fig. 3K), the next question is: *can we identify the membrane proteins whose unfolding corresponds to the identified clusters (Fig. 3)?*

For this purpose, we developed a Bayesian method providing candidate proteins on the basis of the information present in the data from Mass Spectrometry (ProteomeXchange) and in other proteomic databases (Uniprot, PDB) — combined with the empirical results obtained from SMFS literature(Bosshart et al., 2012; Cisneros et al., 2005; Ge et al., 2016; Kawamura et al., 2013; Kedrov et al., 2004; Klyszejko et al., 2008; Maity et al., 2015; Möller et al., 2003; Oesterhelt et al., 2000; Sapra et al., 2009; Serdiuk et al., 2016; Thoma et al., 2017, 2012). The Bayesian identification (Fig. 4A) is based on two steps: firstly, the crossing of information between the clusters with the results of Mass Spectrometry analysis of the sample under investigation (hippocampal neurons, discs, etc.); secondly, a refinement of the preliminary candidates using additional information (structural and topological) present in the PDB and Uniprot databases. The first step relies on the comparison of Lc_{max} of each cluster with the length of proteins identified in the Mass Spectrometry data (Chen et al., 2006; Kwok et al., 2008; Panfoli et al., 2008). By using the protein abundance included in the Mass Spectrometry data as the prior, a first list of candidate proteins is obtained as well as their probabilities (Fig. 4A III). The SMFS-literature contains 14 examples of unfolded membrane proteins allowing a comparison between the Lc_{max} of the measured F-D curves and the real length of the same protein completely stretched (Fig. 4B). Therefore, we extrapolated the first conditional probability of our Bayesian inference, showing that on average, the value of Lc_{max} corresponds to 89% of the real length of the protein (R^2=0.98). When the structure of candidate proteins is known (helices, sheets), these probabilities are further refined considering e.g. the force of unfolding or the position of the loops which are known to constrain the possible unfolding patterns (Fig. 4C-F, see supplementary methods for the formal description).

Like in Mass Spectrometry, this identification is probabilistic (Fig. 5), and it allows to reduce the number of possible candidates from hundreds to 1 to 4 proteins.

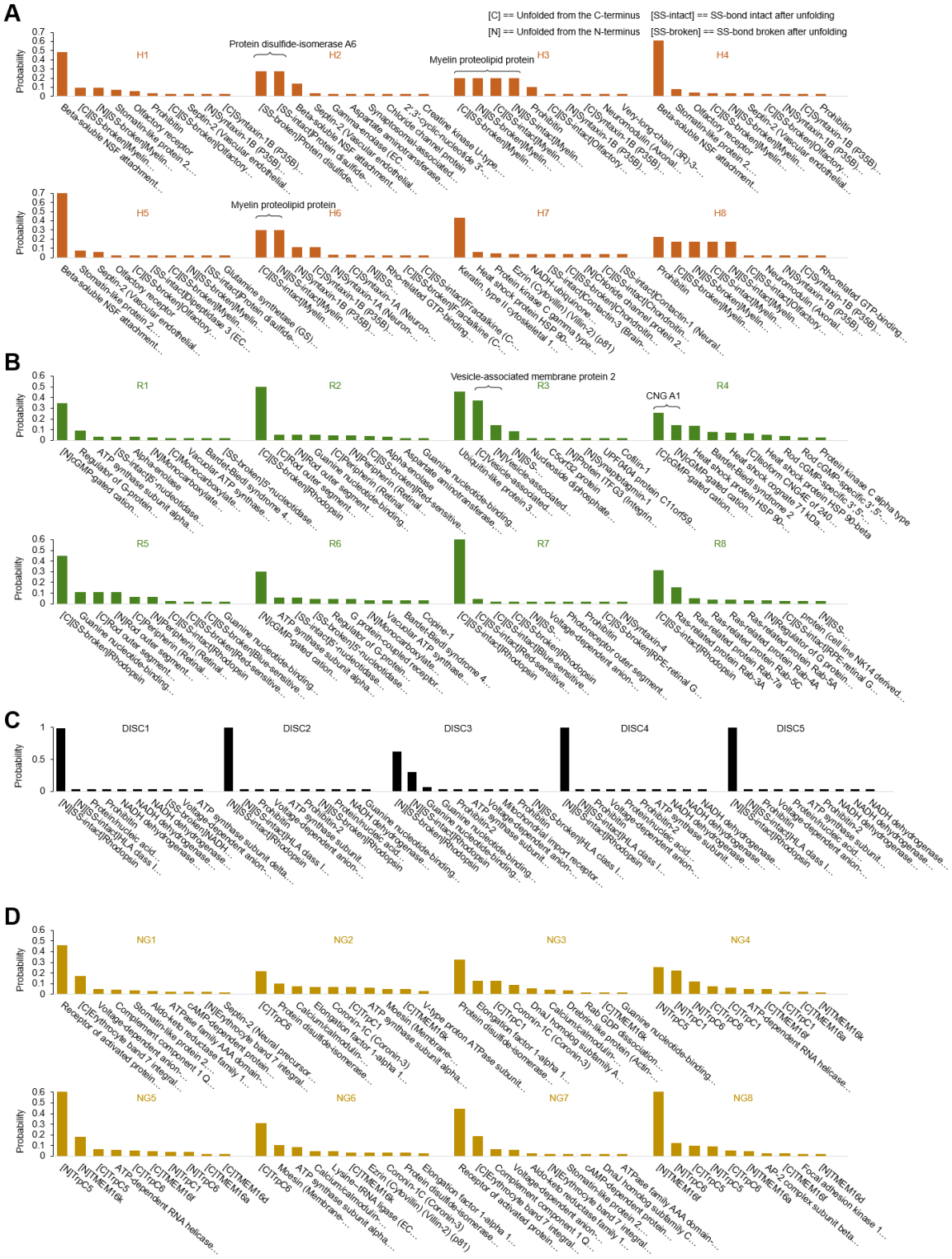


Fig. 5. Bayesian identification of the unfolding clusters*. Most probable candidates for the unfolding clusters found in a, hippocampal neurons; b, rods; c, rod discs; d, NG108-15 cells.*

To prove the reliability of the proposed pipeline protocol, we focused on 4 clusters found in NG108-15 cells where the Bayesian method assigned a high probability to the membrane proteins TMEM16F, TRPC1, TRPC5 and TRPC6. Reverse Transcription Polymerase Chain Reaction and Western Blot analysis (Wu et al., 2007) (Fig. S17B) confirm that these membrane proteins are abundantly expressed in neuroblastoma NG108-15 cells. We overexpressed the construct 6xHis-N2B-*protein*-GFP, where *protein* is the molecular candidate to be validated (Fig. 6A–B, Table S1). The N2B is a chain of 204 aa which can be unfolded with a force less than 10 pN providing a well-known signature(Linke and Grützner, 2008), and GFP is the Green Fluorescent Protein used to identify which NG108-15 cell was successfully transfected, and if the proteins were present in the unroofed membrane patch (Fig. 6C).

We compared the clusters obtained without overexpression (NG8, NG4, NG5 and NG11) with the new clusters obtained from the unfolding of the constructs N2B-TMEM16F-GFP, N2B-TRPC1-GFP, N2B-TRCP5-GFP and N2B-TRPC6-GFP from the transfected cells (Fig. 6D, E). These F-D traces had the well-known signature of the N2B domain consisting in a flat portion with a length of ~85 nm, which was followed by force peaks almost exactly matching those observed in the corresponding density plots (Fig. 6D). This matching was also confirmed by the correspondence of the cumulative peaks observed in the global histogram of the contour length of the NG8, NG4, NG5 and NG11 clusters with those obtained from neuroblastoma cells transfected with the constructs N2B-TMEM16F-GFP, N2B-TRPC1-GFP, N2B-TRCP5-GFP and N2B-TRPC6-GFP translated of ~85 nm because of the presence of the N2B (Fig. 6E).

This pipeline demonstrates that from the obtained clusters (Fig. 3) it is possible to derive the unfolding spectra of membrane proteins in their almost-native conditions without the need of purification. The superposition of the peaks (Fig. 6D–E) confirms the identification and allows to map the unfolding positions along the tertiary structure of the proteins (Fig. 6F).

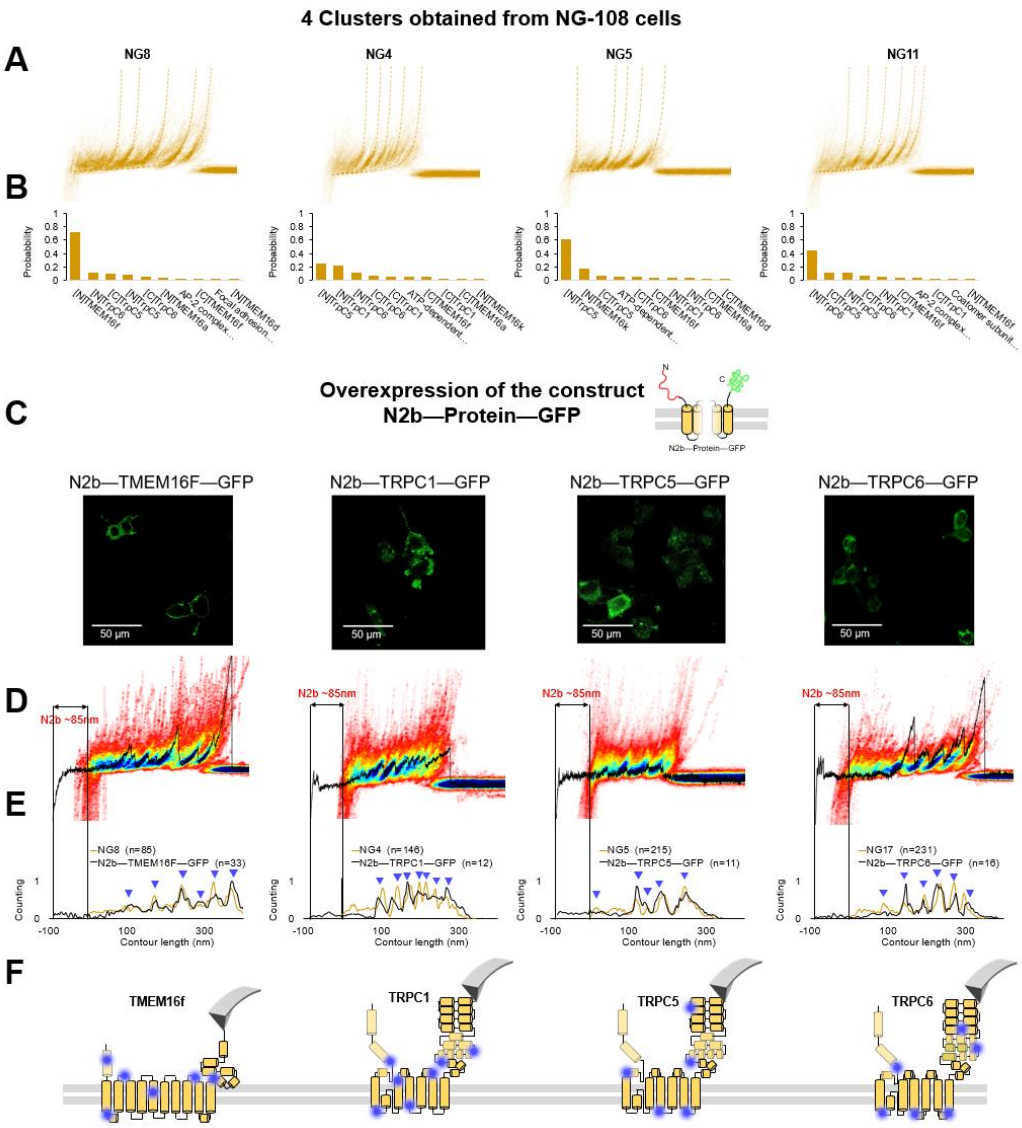


Fig. 6. Validation of the method with a comparison between native clusters and those obtained with overexpression of a construct with the N2B signature at the N-terminal and GFP at the C-terminal. A*: four clusters found in the dataset of F-D traces pulled from NG108-15 cells wild type.* ***B****: the Bayesian identification of the clusters in* ***A*** *with the candidate membrane proteins.* ***C****: confocal images of NG108-15 cells overexpressing the construct N2b—Protein—GFP, as reported by the green fluorescence emitted by GFP.* ***D****: from left to right, color density plots (blue indicating a colocalization among F-D traces of more than 80 %) of the clusters superimposed to a F-D curve bearing the N2B signature (85 nm segment with a flat force below 10 pN) obtained in the sample overexpressing the candidate protein TMEM16F, TRPC1, TRPC5 and TRPC6, respectively.* ***E****: Global histogram of Lc of the clusters in* ***A*** *and the clusters with the N2B signature from cells transfected with the corresponding construct* ***F****: position of the most likely rupture regions of the proteins according to* ***E****. superimposed to the cartoon of the Cryo-EM structure available in the PDB.*

Structural insights from SMFS

The confirmation of the molecular identity of the unfolded proteins allows a better understanding of their molecular structures obtained from Cryo-EM, in particular of the not-well folded regions. SMFS detects the unfolded domains of a protein and can characterize their variability where the more powerful Cryo-EM precisely determines the position of the atoms of the well folded domains. The unfolding of the construct N2B-TRPC6-GFP as well as the endogenous TRPC6 show force peaks at 143, 193 and 230 nm with different properties (Fig. 7). The peak at 143 nm is always well defined, but the force peaks at 193 nm and at 230 nm — occurring at the extracellular loop E1 between S1 and S2 and at the extracellular loop E2 between S3 and S4, respectively (see positions in Fig. 7B) — show variability in the location and strength of the secondary peaks (Fig. 7D and E). The unfolding behavior observed at 193 nm and 230 nm reflects the presence of small structural elements(Yu et al., 2017) which are present in some but not all instances (see inserts in Fig. 7D-E). The peak at 143 nm, in contrast shows a clear and reproducible unfolding behavior (Fig. 7C) typical of a defined and well folded structure(Rico et al., 2013; Takahashi et al., 2018).

In the light of these results, we analyzed the available Cryo-EM structures of TRPC6. There are three published structures (Bai et al., 2020, p. 6; Tang et al., 2018, p. 6): these three Cryo-EM structures are very similar — and almost identical — particularly in the long transmembrane α-helices, but differ in the connecting loops where short α-helices are present only in some of these structures (Fig. 7 C-E, inserts). These short α helices in the loops E1 and E2 are the small structural elements at the origin of the segmentation observed in the F-D traces (Fig. 7 D-E). However, as shown in the structures in yellow, the α helices are not always observed as the single-peak F-D curves suggest. Therefore the most likely conclusion from our SMFS data is that, at room temperature and in their native environment, the different Cryo-EM structures coexist and the connecting loops E1 and E2 have a variable structure.

The observed variability is not restricted only to loop E1 and E2 of TRPC6: also the loop pre-S1 and S1 of TRPC5 shows a similarly complex unfolding behavior as opposed to the unfolding of the loop between S2 and S3. Clusters DRG5 or H3 show an even more exotic behavior. Their unfolding curves show a dual nature: secondary peaks and unfolding humps (Fig. S14).

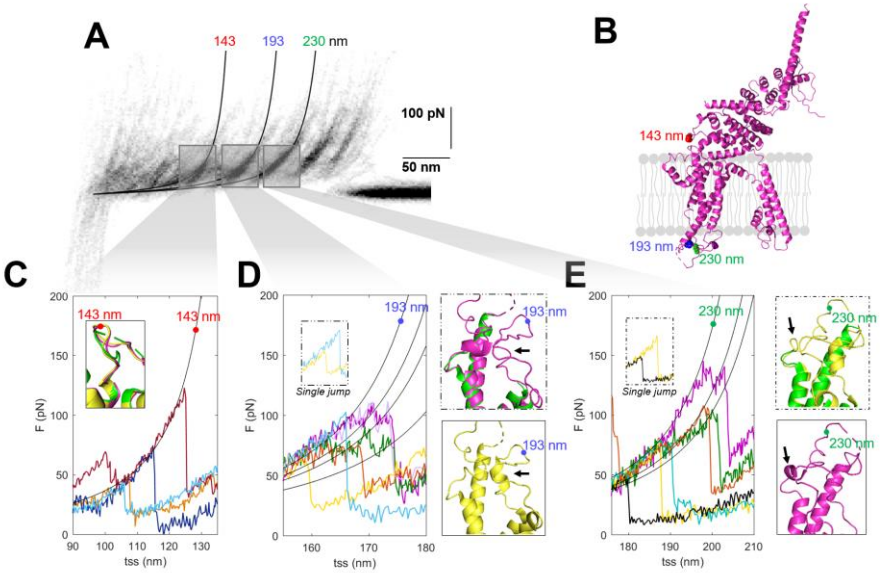


Fig. 7. Structural segmentation in the loops of TRPC6. A*: Density plot of F-D curves of TRPC6 (validated from cluster NG11).* ***B****: 3D structure of TRPC6 (PDB: 6UZA) with highlighted the position corresponding to the unfolding peaks in A.* ***C****: Representative F-D curves showing the single peak at 143 nm. Insert: comparison of the agreement with the three proposed structures of TRPC6 (PDB ID, green:*

5YX9; purple: 6UZA; yellow: 6UZ8). ***D****: Representative F-D curves showing the variable behavior at 193 nm where ~20% of the F-D curves shows a single jump that suggest no structural fragmentation of the loop like in the green/purple structures. Right inserts: comparison of the structure of the loop obtained by Cryo-EM where only the yellow structure show the presence of a 2-turn helix.* ***E****: Representative F-D curves at 230 nm where ~30% of the F-D curves shows a single jump that suggest no structural fragmentation of the loop like in the green/yellow structures. Right inserts: comparison of the structure of the loop obtained by Cryo-EM where only the purple structure show the presence of a 1-turn helix.*

Discussion

Given the high diversity of membrane proteins in cells, the unequivocal identification of any unlabeled protein from single cells is a daunting task. However, the present manuscript shows that by combing SMFS on native membrane patches with information present in databases, a probabilistic identification of groups of proteins is possible. Our approach provides usually less than 4 candidates (Fig. 5) for most of the identified clusters of F-d traces ($p > 95\%$) and we were able to prove the validity of the identification by properly engineered constructs (Fig. 6).
The proposed pipeline offers also the possibility to enrich the structural information derived from Cryo-EM, in their native environment. SMFS cannot distinguish between a β-sheet and an α-helix, but it can determine with a good accuracy the unfolded domains(Engel and Gaub, 2008) of a membrane protein at room temperature (see Fig. 7). Cryo-EM cannot capture the structural heterogeneity of the poorly folded regions, while SMFS can provide important complementary information, for instance in loose connecting loops.
The pipeline, in addition, offers the possibility to obtain the unfolding profile of the membrane proteins from a limited amount of native material (membranes isolated from 1 to 10 cells). We have shown that the unfolding profile of membrane proteins can be used as a fingerprint to characterize the sample under investigation (see Fig. 2–3 and Fig. S12). We foresee that this approach can be extended to characterize, and eventually distinguish, membranes in cells in healthy and sick conditions where other methodology cannot be applied because of sample scarcity.

A major limitation of the pipeline — in its present form — is the possibility to merge in the same cluster the unfolding of proteins with a different molecular identity: indeed, from the Mass Spec data it is clear that different proteins can have the same — or approximately the same — molecular weight, a similar unfolding pattern and a similar total unfolded length Lc. This issue is rather significant for short proteins, i.e., those with values of Lc between 50 nm and 150 nm. In order to overcome this limitation, it will be desirable to couple SMFS with a device (e.g., solid state nanopores(Lee et al., 2018)) able to identify, even qualitatively, specific amino acids during the unfolding process. Another limitation is that a protein, in order to be identified with our approach, must be abundant: the number of distinct membrane proteins that are present, for example in hippocampal neurons is believed to be of order hundreds but with only few of them representing the majority of the expressed proteins. With our approach indeed we identify only 10 clusters.

This work ss a first step for the use of SMFS in native samples, but despite the limitations here reported, we were still able to classify tens of unfolding pathways, identify with high accuracy part of them, and we could also reconcile some contrasting cryo-EM structures for the molecularly identified proteins. We therefore believe that the proposed methodology represents a genuine step in the direction of single-cell membrane proteomics.

Methods:
All experimental procedures were in accordance with the guidelines of the Italian Animal Welfare Act, and their use was approved by the SISSA Ethics Committee board and the National Ministry of Health (Permit Number: 630-III/14) in accordance with the European Union guidelines for animal care (d.1.116/92; 86/609/C.E.).

Cell preparation and culture.
NG108-15.
Mouse neuroblastoma NG108-15 cells were obtained from Sigma-Aldrich. The cells were grown in Dulbecco's Modified Eagle Medium (DMEM, ThermoFisher) plus 10% Fetal bovine serum (FBS, Gibco), 100 U/ml Penicillin and 100 U/ml Streptomycin. The cells were cultured into a humidified incubator (5% CO2, 37 °C).

Hippocampal and DRG neurons.
Hippocampal and DRG neurons were obtained from Wistar rats (P2-P3) as described in ref. (Galvanetto, 2018). In short, the animals were anesthetized with CO_2 and sacrificed by decapitation. The dissociated cells were plated at a concentration of 4×10^4 cells/ml onto glass round coverslips (170 µm in thickness) coated with 0.5 mg/ml poly-D-lysine (Sigma-Aldrich, St. Louis, MO, USA) for 1 h at 37°C and washed 3 times in deionized water. It is fundamental to obtain an optimal adhesion of the cells to prevent detachment in the next step (isolation of the cell membrane). The medium used for hippocampal neurons is in Minimum Essential Medium (MEM) with GlutaMAX supplemented with 10% Fetal Bovine Serum (FBS, all from Invitrogen, Life Technologies, Gaithersburg, MD, USA), 0.6% D-glucose, 15 mM Hepes, 0.1 mg/ml apo-transferrin, 30 µg/ml insulin, 0.1 µg/ml D-biotin, 1 µM vitamin B12 (all from Sigma-Aldrich), and 2.5 µg/ml gentamycin (Invitrogen). The medium used for DRG neurons is Neurobasal medium (Gibco, Invitrogen, Milan, Italy) supplemented with 10% Fetal Bovine Serum (FBS, from Invitrogen, Life Technologies, Gaithersburg, MD, USA).

Rods.
Rod cells were obtained from adult male Xenopus laevis as described in ref. (Mazzolini et al., 2015). Under infrared illumination, the eyes of dark-adapted frogs after anesthesia with MS-222 were surgically extracted. Eyes were preserved in the Ringer solution (110 NaCl, 2.5 KCl, 1 CaCl2, 1.6 MgCl2, 3 Hepes-NaOH, 0.01 EDTA, and 10 glucose in mM; pH 7.8 buffered with NaOH), and hemisected under a dissecting microscope. The extracted retina was maintained in the Ringer solution.

Cell transfection.
NG108-15 cells were transiently transfected with 300 ng of each cDNA expression plasmids (see Supplementary Table 1) by using Lipofectamine 2000 Transfection Reagent (ThermoFisher) according to its handbook. Briefly, the plasmids (Supplementary Table 1) and the Lipo2000 were diluted into Opti-MEM Reduced Serum Medium (Gibco), respectively. 5 mins later, we added the diluted DNA to the diluted Lipo2000 to make the plasmid DNA-lipid complexes. After incubating 30 min, we plated the cells on the 12 mm round coverslips coated with 1x Poly-D-Ornithine (Sigma-Aldrich) in 12 well plate, and in the meanwhile, we added DNA-lipid complexes to the cells. We performed membrane isolation and confocal experiments about 56 hours after transfection.

Isolation of cell membranes.
Single-cell unroofing (for cell types that grow in adhesion).
The apical membrane of Hippocampal neurons, DRGs and NG108-15 cells were isolated with an optimized version of the unroofing method(Galvanetto, 2018). Briefly, additional empty glass coverslips (24 mm in diameter, 170 µm in thickness) were plasma cleaned for 15 seconds and broken in 4 quarters

(with the use of the hands) in order to obtain optically sharp edges, as described in (Galvanetto, 2018). The coverslip quarters were immersed into 0.5 mg/ml poly-D-lysine for 30 minutes, and then they were immersed in deionized water for 10 seconds before use. A petri dish was filled with Ringer solution (3.5 ml) without any glucose, where the glass quarter was placed tilted of 7-15 degrees in the middle of it, supported by a 10 x 10 x 1 mm glass slice and Vaseline. The cover of the petri dish was then fixed on the stage of the AFM-inverted microscope setup (JPK Nanowizard 3 on an Olympus IX71).
The cell culture was then mounted on a 3D printed coverslip holder connected to the head stage of the AFM. The AFM head was put on top of the stage in measurement position. The cell culture was immersed into the solution and a target cell was identified and aligned with the underlying corner of the glass quarter. The cell culture was moved towards the corner of the underlying glass with the motors of the AFM until the target cell was squeezed and it doubled its area. At this point the cell is kept squeezed for 3 minutes, then a loaded spring under the AFM is released to abruptly separate the corner from the cell culture, and break the target cell membrane. The glass quarter with the isolated cell membrane was laid down and fixed on the petri dish. The medium was replaced with Ringer's solution without exposing the cell membrane to the air.

Membrane isolation of non-adherent cells.
Cells that do not grow in adhesion usually do not establish a tight binding with the substrate on top of which they are deposited. For these cells (e.g. rod cells), instead of unroofing, it is more reliable to break the cells with a lateral flux of medium(Clarke et al., 1975).
Isolated and intact rods were obtained by mechanical dissociation of the Xenopus retina in an absorption buffer (150 mM KCl, 25 mM $MgCl_2$, and 10 mM Trizma base; pH 7.5); they were then deposited on cleaved mica as described in ref.(Maity et al., 2017). Incubated rods were maintained for 30–45 minutes over the mica in order to be adsorbed by its negatively charged surface. In the meanwhile, the position of the rods in the field of view of the microscope was annotated. The absorption buffer was substituted by a solution containing (in mM): 150 KCl, 10 Tris-HCl, (pH 7.5) and then a lateral flux of medium was applied to the rods until all the cell bodies were removed.
Experiments on the (almost) protein-free membrane of Oocytes, overexpressing CNG-N2b were made following the protocol of ref. (Maity et al., 2015).

Isolation of rod discs.
Purification techniques with multiple centrifugations are usually required to isolate membrane-only organelles like rod discs or outer membrane vesicles(Thoma et al., 2018). Rod discs were obtained starting from the extracted retina as described in ref. (Maity et al., 2017). Briefly, discs were separated with two series of centrifugations of the sample overlaid on a 15-40% continuous gradient of OptiPrep (Nycomed, Oslo, Norway). 40 µl of the sample were diluted with 40 µl of absorption buffer, and incubated on freshly cleaved mica for 40 minutes. After 40 minutes, the incubation medium was removed and substituted with the solution used in the AFM experiments (150mM KCl, 10mM Tris-HCl, pH 7.5).

<u>Reconstitution of purified membrane proteins.</u>
Sthk channel was purified as described in ref.(Marchesi et al., 2018). bR and ChR1 proteins were bought from Cube-Biotech. The purified protein or they different mixtures (as indicated in the Results section) were brought at a total concentration of 0.5 mg/ml in a buffer containing 20mM HEPES, pH 7.8, 150mM KCl and 0.1% N-dodecyl β–D maltoside (DDM) and aliquoted into 100µL samples. A lipid mixture of 1,2-dioleolyl-*sn*-glycero-3-phosphocholine (DOPC), 1,2-dioleolyl-*sn*-glycero-3-phosphoethanolamine (DOPE), 1,2-dioleolyl-*sn*-glycero-3-phospho-L-serine (DOPS) at 8:1:1 ratio (Avanti Polar Lipids) was added at a lipid-to-protein ratio of 1 (wt/wt). The different ternary mixtures of protein-lipid-detergent were sonicated in an ice-bath sonicator for 2 min and subsequently equilibrated on an orbital shaker (200rmp) for 2h. Detergent was removed by hydrophobic adsorption adding twice ~5 mg of wet bio-beads SM-2 and equilibrating under gentle shaking (200rmp) for 1h at RT and overnight at 4°C,

respectively. Before use samples were diluted 1:3 in adsorption buffer (20mM HEPES, pH 7.8, 150mM KCl)
The reconstituted membrane proteins were adsorbed on freshly cleaved mica for 30 min in a humid chamber. The samples were gently rinsed with imaging buffer (150Mm KCl, 10mM HEPES, Ph=7.8) and subsequently used for AFM imaging and unfolding.

Gene silencing experiment.
The genes of TRPC1/5/6 and TMEM16F were edited using CRISPR/Cas9 and their sgRNAs were cloned into LentiCRISPR-V2 puro vector (Supplementary Table 2). 3000ng of the indicated plasmids were transfected into the NG108-15 cells by using the lipo2000. 52h later after transfection, the cells were used for western blot. In order to select the silencing cell for single-cell unroofing, we treated the cells with culture medium containing 2 µl/ml puromycin(Jiao et al., 2018) after 22h transfection. After about 68h selection of puromycin, the alive cells were subjected to unroofing and the unroofed membrane were used for unfolding experiment.

AFM imaging and Single-Molecule Force Spectroscopy (SMFS).
AFM experiments was performed using an automated AFM (JPK Nanowizard 3) with 50 µm long cantilevers (AppNano HYDRA2R-NGG, nominal spring constant = 0.84 N/m). We calibrated the AFM cantilevers in the experimental medium before each experiment using the equipartition theorem(Butt et al., 1995). The AFM experiments of Hippocampal neurons and DRGs were performed with Ringer's solution (NaCl 145 mM, KCl 3 mM, CaCl2 1.5 mM, MgCl2 1 mM, HEPES 10 mM, pH 7.4); The AFM experiments of NG108-15 cell membranes were performed in the Ca2+-free Ringer solution supplemented with 10mM TCEP (Cat# C4706, Sigma). The TCEP can breakdown the disulfide bonds; Rod membrane and discs experiments were performed with 150mM KCl, 10mM Tris-HCl, pH 7.5. The fishing of the reconstituted membrane proteins (ChR1 and SthK) and bR was performed with imaging buffer (150mM KCl, 10mM HEPES, pH=7.8). All experiments were performed at 24 Celsius.
AFM imaging.
The position of the cells before unroofing was annotated in the monitor of the computer in order to start the AFM imaging where the cells was in contact with the substrate (cell membrane is not visible in bright-field). The membrane obtained with single-cell unroofing (hippocampal neurons, DRG and NG108-15) can easily be found in proximity of the glass corner (~80% success rate). In the case of the rod membrane (non-adherent cells), usually different positions need to be scanned before finding a patch of membrane. Rod discs, and the reconstituted protein patches can be identified only via AFM imaging. We performed imaging both in contact mode (setpoint ~0.4 nN) and intermittent-contact mode (lowest possible), but the intermittent-contact mode if preferable because it does not damage the border of the patches of membrane.

AFM-based SMFS (protein unfolding)
we performed automated SMFS on top of the imaged membranes by setting grid positions for the approaching and retraction cycles of the cantilever. All experiments were performed with a retraction speed of 500 nm/s (for hippocampal neurons, DRG, Rods and Discs), 600nm/s (for NG108-15 cells), 1000nm/s (for bR, ChR1 and SthK). The membrane proteins present in the sample were attached non-specifically to the cantilever tip by applying a constant force of ~1 nN for 1 second between the AFM tip and the cytoplasmic side of the membrane. This method proved to work with different membrane proteins(Tanuj Sapra et al., 2006; Thoma et al., 2017, 2012), and to allow a higher throughput compared to methods that involve a specific attachment between the tip and the protein(Cisneros et al., 2005; Kedrov et al., 2004; Müller et al., 2002; Oesterhelt et al., 2000; Sumbul et al., 2018a, 2018b). Also, in

order to demonstrate the sawtooth patterns truly represent the unfolding of membrane proteins, we performed unfolding experiment on the top of the coverslip coating with the poly-D-lysine.

Western blot.

The transfected NG108-15 cells were harvested in the lysis buffer(150 mM NaCl, 1.0% Triton X-100, 0.1% SDS, 0.5% sodium deoxycholate, 10 mM NaF, 1 mM Na3VO4, 1.0% NP-40, and 40mM Tris-HCl, pH=8.0) supplemented with protease inhibitor(Cat# 469311001, Roche) on ice. The samples were boiled at 100 °C for 30 mins and then the supernatants were loaded in 10% SDS PAGE gel for gel electrophoresis. The proteins on the gels were transferred onto the PVDF membranes (Cat# 88520, ThermoFisher) and these membranes were blocked with 5% nonfat dry milk. The membranes were incubated with the corresponding primary antibodies 1:5000 for α-Tubulin (Cat# T8203, sigma); 1:500 for TRPC1 (Cat# ACC-101, Alomone Labs)/5 (Cat# ACC-020, Alomone Labs)/6 (Cat#ACC-120, Alomone Labs) and 1:200 for TMEM16F (Cat# ACL-016, Alomone Labs) at 4°C overnight with gentle shaking. Secondary antibodies Goat-anti-rabbit HRP (Cat# P0448, Dako) and Goat-anti-mouse (Cat# P0447, Dako) 1:5,000 in 5% BSA were used. Membranes were washed with PBST three times and developed with an ECL-HRP (Cat# WBULS0100, Millipore) system. The gray value ratios of protein bands were quantified using ImageJ software.

Confocal experiments.

NG108-15 whole cells

The transfected cells and non-transfected cells were fixed by using 4% PFA for 60 min at RT. The transfected cells were stained with Hoechst (Life Technologies). The non-transfected cells were permeated by 0.05% TritonX-100 for 4 min and blocked by 10% FBS+5% BSA for 90 min, following by the incubation of corresponding primary antibodies (1: 250) at 4°C overnight. The samples were rinsed with pre-cold 1xPBS 3 times and incubated with the secondary antibody Alexa 594-labeled goat anti-rabbit (1:800, Cat# A11037, Invitrogen) at RT for 90 min. After staining the samples with Hoechst, the samples were tested by Nikon A1R microscope with 60× oil immersion objective (NA 1.40). The results were analyzed with the ImageJ software.

NG108-15 cell membranes

The NG108-15 cells were stained with 200 μM SiR-Actin, 200 μM SiR-Tubulin (Spirochrome) or 25 nM MitoTracker® Red FM (ThermoFisher) for 30 min in a humidified incubator (5% CO2, 37 °C). The stained or transfected cells were unroofted according to the method as mentioned above. Then the NG108-15 cell membranes were tested by Nikon A1R microscope with 20× objective, followed by the imaging scan by AFM. The results were analyzed with the ImageJ software.

Molecular visualization.

We aligned the current protein structures of TRPC superfamily using the PyMOL software, such as human TRPC3, mouse TRPC4, mouse TRPC5 and human TRPC6. Their codes in PDB are 5zbg, 5z96, 6aei and 5yx9, respectively. We zoomed the C-terminal part (from S5 to the C-termini) of these structures and find more details.

Acknowledgments: We thank Alexander Bruce for useful comments on the manuscript. We thank dr. Kosaku Shinoda for support in the emPAI calculation. We thank Prof. Anna Menini who provided the mTMEM16A-GFP and mTMEM16F-GFP plasmids. We thank Prof. Guidalberto Manfioletti who provided the peGFP-N1 plasmid.

References

Al-Rekabi Z, Contera S. 2018. Multifrequency AFM reveals lipid membrane mechanical properties and the effect of cholesterol in modulating viscoelasticity. *PNAS* **115**:2658–2663. doi:10.1073/pnas.1719065115

Bai Y, Yu X, Chen H, Horne D, White R, Wu X, Lee P, Gu Y, Ghimire-Rijal S, Lin DC-H, Huang X. 2020. Structural basis for pharmacological modulation of the TRPC6 channel. *eLife* **9**:e53311. doi:10.7554/eLife.53311

Bosshart PD, Iordanov I, Garzon-Coral C, Demange P, Engel A, Milon A, Müller DJ. 2012. The Transmembrane Protein KpOmpA Anchoring the Outer Membrane of Klebsiella pneumoniae Unfolds and Refolds in Response to Tensile Load. *Structure* **20**:121–127. doi:10.1016/j.str.2011.11.002

Butt H-J, Jaschke M, Ducker W. 1995. Measuring surface forces in aqueous electrolyte solution with the atomic force microscope. *Bioelectrochemistry and Bioenergetics* **38**:191–201. doi:10.1016/0302-4598(95)01800-T

Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, Fernandez JM. 1999. Mechanical and chemical unfolding of a single protein: A comparison. *PNAS* **96**:3694–3699. doi:10.1073/pnas.96.7.3694

Casuso I, Khao J, Chami M, Paul-Gilloteaux P, Husain M, Duneau J-P, Stahlberg H, Sturgis JN, Scheuring S. 2012. Characterization of the motion of membrane proteins using high-speed atomic force microscopy. *Nature Nanotechnology* **7**:525–529. doi:10.1038/nnano.2012.109

Chen P, Li X, Sun Y, Liu Z, Cao R, He Q, Wang M, Xiong J, Xie J, Wang X, Liang S. 2006. Proteomic analysis of rat hippocampal plasma membrane: characterization of potential neuronal-specific plasma membrane proteins. *Journal of Neurochemistry* **98**:1126–1140. doi:10.1111/j.1471-4159.2006.03934.x

Cisneros DA, Oesterhelt D, Müller DJ. 2005. Probing Origins of Molecular Interactions Stabilizing the Membrane Proteins Halorhodopsin and Bacteriorhodopsin. *Structure* **13**:235–242. doi:10.1016/j.str.2004.12.005

Clarke M, Schatten G, Mazia D, Spudich JA. 1975. Visualization of actin fibers associated with the cell membrane in amoebae of Dictyostelium discoideum. *PNAS* **72**:1758–1762.

Engel A, Gaub HE. 2008. Structure and Mechanics of Membrane Proteins. *Annual Review of Biochemistry* **77**:127–148. doi:10.1146/annurev.biochem.77.062706.154450

Galvanetto N. 2018. Single-cell unroofing: probing topology and nanomechanics of native membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1860**:2532–2538. doi:10.1016/j.bbamem.2018.09.019

García-Sáez AJ, Chiantia S, Schwille P. 2007. Effect of Line Tension on the Lateral Organization of Lipid Membranes. *J Biol Chem* **282**:33537–33544. doi:10.1074/jbc.M706162200

Ge L, Villinger S, Mari SA, Giller K, Griesinger C, Becker S, Müller DJ, Zweckstetter M. 2016. Molecular Plasticity of the Human Voltage-Dependent Anion Channel Embedded Into a Membrane. *Structure* **24**:585–594. doi:10.1016/j.str.2016.02.012

Hinczewski M, Hyeon C, Thirumalai D. 2016. Directly measuring single-molecule heterogeneity using force spectroscopy. *PNAS* **113**:E3852–E3861. doi:10.1073/pnas.1518389113

Ilieva NI, Galvanetto N, Allegra M, Brucale M, Laio A. n.d. Automatic classification of single-molecule force spectroscopy traces from heterogeneous samples. *Bioinformatics*. doi:10.1093/bioinformatics/btaa626

Jiao A, Li F, Zhang C, Lv W, Chen B, Zhang J. 2018. Simulated Cholinergic Reinnervation of β (INS-1) Cells: Antidiabetic Utility of Heterotypic Pseudoislets Containing β Cell and Cholinergic Cell. *International Journal of Endocrinology*. doi:https://doi.org/10.1155/2018/1505307

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. doi:10.1038/s41586-021-03819-2

Kawamura S, Gerstung M, Colozo AT, Helenius J, Maeda A, Beerenwinkel N, Park PS-H, Müller DJ. 2013. Kinetic, Energetic, and Mechanical Differences between Dark-State Rhodopsin and Opsin. *Structure* **21**:426–437. doi:10.1016/j.str.2013.01.011

Kedrov A, Ziegler C, Janovjak H, Kühlbrandt W, Müller DJ. 2004. Controlled Unfolding and Refolding of a Single Sodium-proton Antiporter using Atomic Force Microscopy. *Journal of Molecular Biology* **340**:1143–1152. doi:10.1016/j.jmb.2004.05.026

Kessler M, Gaub HE. 2006. Unfolding Barriers in Bacteriorhodopsin Probed from the Cytoplasmic and the Extracellular Side by AFM. *Structure* **14**:521–527. doi:10.1016/j.str.2005.11.023

Klyszejko AL, Shastri S, Mari SA, Grubmüller H, Muller DJ, Glaubitz C. 2008. Folding and Assembly of Proteorhodopsin. *Journal of Molecular Biology* **376**:35–41. doi:10.1016/j.jmb.2007.11.030

Kwok MCM, Holopainen JM, Molday LL, Foster LJ, Molday RS. 2008. Proteomics of Photoreceptor Outer Segments Identifies a Subset of SNARE and Rab Proteins Implicated in Membrane Vesicle Trafficking and Fusion. *Molecular & Cellular Proteomics* **7**:1053–1066. doi:10.1074/mcp.M700571-MCP200

Lee K, Park K-B, Kim H-J, Yu J-S, Chae H, Kim H-M, Kim K-B. 2018. Recent Progress in Solid-State Nanopores. *Advanced Materials* **30**. doi:10.1002/adma.201704680

Li H, Linke WA, Oberhauser AF, Carrion-Vazquez M, Kerkvliet JG, Lu H, Marszalek PE, Fernandez JM. 2002. Reverse engineering of the giant muscle protein titin. *Nature* **418**:998–1002. doi:10.1038/nature00938

Linke WA, Grützner A. 2008. Pulling single molecules of titin by AFM—recent advances and physiological implications. *Pflugers Arch - Eur J Physiol* **456**:101–115. doi:10.1007/s00424-007-0389-x

Maity S, Ilieva N, Laio A, Torre V, Mazzolini M. 2017. New views on phototransduction from atomic force microscopy and single molecule force spectroscopy on native rods. *Scientific Reports* **7**. doi:10.1038/s41598-017-11912-8

Maity S, Mazzolini M, Arcangeletti M, Valbuena A, Fabris P, Lazzarino M, Torre V. 2015. Conformational rearrangements in the transmembrane domain of CNGA1 channels revealed by single-molecule force spectroscopy. *Nature Communications* **6**:7093. doi:10.1038/ncomms8093

Marchesi A, Gao X, Adaixo R, Rheinberger J, Stahlberg H, Nimigean C, Scheuring S. 2018. An iris diaphragm mechanism to gate a cyclic nucleotide-gated ion channel. *Nat Commun* **9**:1–11. doi:10.1038/s41467-018-06414-8

Marsico A, Labudde D, Sapra T, Muller DJ, Schroeder M. 2007. A novel pattern recognition algorithm to classify membrane protein unfolding pathways with high-throughput single-molecule force spectroscopy. *Bioinformatics* **23**:e231–e236. doi:10.1093/bioinformatics/btl293

Mazzolini M, Facchetti G, Andolfi L, Zaccaria RP, Tuccio S, Treu J, Altafini C, Fabrizio EMD, Lazzarino M, Rapp G, Torre V. 2015. The phototransduction machinery in the rod outer segment has a strong efficacy gradient. *PNAS* **112**:E2715–E2724. doi:10.1073/pnas.1423162112

Möller C, Fotiadis D, Suda K, Engel A, Kessler M, Müller DJ. 2003. Determining molecular forces that stabilize human aquaporin-1. *Journal of Structural Biology* **142**:369–378. doi:10.1016/S1047-8477(03)00066-2

Müller DJ, Kessler M, Oesterhelt F, Möller C, Oesterhelt D, Gaub H. 2002. Stability of Bacteriorhodopsin α-Helices and Loops Analyzed by Single-Molecule Force Spectroscopy. *Biophysical Journal* **83**:3578–3588. doi:10.1016/S0006-3495(02)75358-7

Nagel G, Ollig D, Fuhrmann M, Kateriya S, Musti AM, Bamberg E, Hegemann P. 2002. Channelrhodopsin-1: A Light-Gated Proton Channel in Green Algae. *Science* **296**:2395–2398. doi:10.1126/science.1072068

Oesterhelt F, Oesterhelt D, Pfeiffer M, Engel A, Gaub HE, Müller DJ. 2000. Unfolding Pathways of Individual Bacteriorhodopsins. *Science* **288**:143–146. doi:10.1126/science.288.5463.143

Panfoli I, Musante L, Bachi A, Ravera S, Calzia D, Cattaneo A, Bruschi M, Bianchini P, Diaspro A, Morelli A, Pepe IM, Tacchetti C, Candiano G. 2008. Proteomic Analysis of the Retinal Rod Outer Segment Disks. *Journal of Proteome Research* **7**:2654–2669. doi:10.1021/pr7006939

Rico F, Gonzalez L, Casuso I, Puig-Vidal M, Scheuring S. 2013. High-Speed Force Spectroscopy Unfolds Titin at the Velocity of Molecular Dynamics Simulations. *Science* **342**:741–743. doi:10.1126/science.1239764

Rodriguez A, Laio A. 2014. Clustering by fast search and find of density peaks. *Science* **344**:1492–1496. doi:10.1126/science.1242072

Sapra KT, Damaghi M, Köster S, Yildiz Ö, Kühlbrandt W, Muller DJ. 2009. One β Hairpin after the Other: Exploring Mechanical Unfolding Pathways of the Transmembrane β-Barrel Protein OmpG. *Angewandte Chemie International Edition* **48**:8306–8308. doi:10.1002/anie.200904361

Sarkar A, Caamano S, Fernandez JM. 2005. The Elasticity of Individual Titin PEVK Exons Measured by Single Molecule Atomic Force Microscopy. *J Biol Chem* **280**:6261–6264. doi:10.1074/jbc.C400573200

Serdiuk T, Balasubramaniam D, Sugihara J, Mari SA, Kaback HR, Müller DJ. 2016. YidC assists the stepwise and stochastic folding of membrane proteins. *Nat Chem Biol* **12**:911–917. doi:10.1038/nchembio.2169

Shen H-H, Lithgow T, Martin LL. 2013. Reconstitution of Membrane Proteins into Model Membranes: Seeking Better Ways to Retain Protein Activities. *Int J Mol Sci* **14**:1589–1607. doi:10.3390/ijms14011589

Sumbul F, Marchesi A, Rico F. 2018a. History, rare, and multiple events of mechanical unfolding of repeat proteins. *J Chem Phys* **148**:123335. doi:10.1063/1.5013259

Sumbul F, Marchesi A, Takahashi H, Scheuring S, Rico F. 2018b. High-Speed Force Spectroscopy for Single Protein Unfolding In: Lyubchenko YL, editor. Nanoscale Imaging: Methods and Protocols, Methods in Molecular Biology. New York, NY: Springer. pp. 243–264. doi:10.1007/978-1-4939-8591-3_15

Takahashi H, Rico F, Chipot C, Scheuring S. 2018. α-Helix Unwinding as Force Buffer in Spectrins. *ACS Nano* **12**:2719–2727. doi:10.1021/acsnano.7b08973

Tang Q, Guo W, Zheng L, Wu J-X, Liu M, Zhou X, Zhang X, Chen L. 2018. Structure of the receptor-activated human TRPC6 and TRPC3 ion channels. *Cell Research* **28**:746–755. doi:10.1038/s41422-018-0038-2

Tanuj Sapra K, Park PS-H, Filipek S, Engel A, Müller DJ, Palczewski K. 2006. Detecting Molecular Interactions that Stabilize Native Bovine Rhodopsin. *Journal of Molecular Biology* **358**:255–269. doi:10.1016/j.jmb.2006.02.008

Thoma J, Bosshart P, Pfreundschuh M, Müller DJ. 2012. Out but Not In: The Large Transmembrane β-Barrel Protein FhuA Unfolds but Cannot Refold via β-Hairpins. *Structure* **20**:2185–2190. doi:10.1016/j.str.2012.10.006

Thoma J, Burmann BM, Hiller S, Müller DJ. 2015. Impact of holdase chaperones Skp and SurA on the folding of β-barrel outer-membrane proteins. *Nat Struct Mol Biol* **22**:795–802. doi:10.1038/nsmb.3087

Thoma J, Manioglu S, Kalbermatter D, Bosshart PD, Fotiadis D, Müller DJ. 2018. Protein-enriched outer membrane vesicles as a native platform for outer membrane protein studies. *Communications Biology* **1**:23. doi:10.1038/s42003-018-0027-5

Thoma J, Ritzmann N, Wolf D, Mulvihill E, Hiller S, Müller DJ. 2017. Maltoporin LamB Unfolds β Hairpins along Mechanical Stress-Dependent Unfolding Pathways. *Structure* **25**:1139-1144.e2. doi:10.1016/j.str.2017.05.010

Walder R, LeBlanc M-A, Van Patten WJ, Edwards DT, Greenberg JA, Adhikari A, Okoniewski SR, Sullan RMA, Rabuka D, Sousa MC, Perkins TT. 2017. Rapid Characterization of a Mechanically Labile α-Helical Protein Enabled by Efficient Site-Specific Bioconjugation. *Journal of the American Chemical Society* **139**:9867–9875. doi:10.1021/jacs.7b02958

Wu G, Lu Z-H, Obukhov AG, Nowycky MC, Ledeen RW. 2007. Induction of Calcium Influx through TRPC5 Channels by Cross-Linking of GM1 Ganglioside Associated with α5β1 Integrin Initiates Neurite Outgrowth. *J Neurosci* **27**:7447–7458. doi:10.1523/JNEUROSCI.4266-06.2007

Yi X, Verbeke EJ, Chang Y, Dickinson DJ, Taylor DW. 2019. Electron microscopy snapshots of single particles from single cells. *Journal of Biological Chemistry* **294**:1602–1608. doi:10.1074/jbc.RA118.006686

Yu H, Siewny MGW, Edwards DT, Sanders AW, Perkins TT. 2017. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science* **355**:945–950. doi:10.1126/science.aah7124

Supplementary Information for

Unfolding and identification of membrane proteins *in situ*

- Supplementary Methods
- Figs. S1 to S14
- Tables S1 to S2

Automatic classification of SMFS data.

The selection of the F-D curves that represent the unfolding of membrane proteins is usually based on the search for a specific pattern of unfolding in the SMFS data, after a filtering based on the length of the protein under investigation[1,2]. In the case of a native preparation (like ours) that contains a mixture of unknown proteins a) the filtering based on the distance cannot be applied and b) the number of specific patterns to be found is unknown. In order to find recurrent patterns of unfolding in a SMFS dataset we developed an algorithm that consists of five major blocks (Fig. S11 a). In the first block, the parts of the F-D curves not related to the unfolding process are removed, and a coarse filtering aimed at the detection of spurious traces is performed. In the second block, a quality score based on the consistency of the experimental data with the worm-like chain (WLC) model is computed and assigned to each trace. This score is used to select physically meaningful traces for further analysis. In the third block, distances between pairs of traces are computed to assess their similarity. The distances are used in the fourth block for density peak clustering. The fifth and final block consists in the refinement and possibly in the merging of some of these clusters. In what follows we provide a detailed overview of each block.

Block 1: filtering.

The standard F-D curve preprocessing was applied to all the data within 'Fodis'[3]. The zero of the force of the curve was determined averaging the non-contact part (baseline after the final peak) and subtracted to all the points of the curve. The piezo position was transformed in tip-sample-separation considering the contribution of the bending of the tip to the extension of the polymer. Given that the F-D curves are subject to noise (due to thermal fluctuations, coming from the instrument, etc.), we smooth the original signals through interpolation on a grid with width δ_{interp}= 1 nm.

A curve is discarded if it does not contain a:

- detectable contact point (i.e. a transition from negative forces to positive forces in respect to the baseline set at zero force);
- if the points occupy force ranges over 5000 pN;

Some of the F-D curves show deviations from the horizontal zero-force line in the non-contact part (wavy final part due to imperfect detachment of the polymer or other noise from the environment). We detect and discard these traces by computing the standard deviation of the tails from the zero-force line. If it exceeds two times σ_{NOISE} (average standard deviation of the baseline of the batch of curves) the trace is discarded.

Block 2: Quality score.

The quality score is used to refine selection of traces with high information content vs. noisy traces. It is based on the description provided by the worm-like chain (WLC) model, which is the standard model in the analysis of SMFS data[4]. The WLC model implies the equation:

$$F(x) = \frac{k_B T}{l_p}\left(\frac{1}{4}\left(1-\frac{x}{L_c}\right)^{-2} + \frac{x}{L_c} + \frac{1}{4}\right) \quad (1)$$

where F is force, x is extension, k_B is Boltzmann's constant, T is temperature, l_p is persistence length and L_c is contour length. Each unfolding curve in the trace is fitted with the WLC equation and a L_c value, corresponding to the length of the unfolded protein domain is obtained. The L_c values are computed by solving equation (1) for each x and F. An appropriate value for the persistence length l_p for membrane proteins is 0.4 nm as reported in ref. [4]. The WLC model is applicable in the force range 30-500 pN[5].
Once we compute the L_c values, we can build a L_c histogram. Normally, the L_c histogram describing a successful unfolding experiment is characterized by the presence of a few maxima separated by deep minima. We implement these features in the definition of our quality score to distinguish meaningful F-D curves.
An important parameter is the bin width of the L_c histogram. If the bin width is too small the histogram is noisy; if the bin width is too large, peaks corresponding to the unfolding of different domains might be merged. We use bin width 8 nm which is an efficient value for evaluating the goodness of a curve and it allows to consider also curves that deviate from the WLC model (l_p =0.4 nm) but that contain information. Furthermore, the choice of such large bin width is based on visual inspection of the histograms of proteins with known structure. Once the L_c histogram is built, we detect all maxima and minima. A maximum is meaningful if it is generated by more than 5 points and it includes more than 1 % of the force measures of a trace.
For each maximum in the L_c histogram, we compute a score W quantifying the consistency of the peak with the WLC model. A high-quality peak is clearly separated from other peaks of the histogram, therefore it should be surrounded by two minima. We define $f_{left} = \frac{P_{left}}{P_{max}}, f_{right} = \frac{P_{right}}{P_{max}}$ where P_{max} , P_{left} and P_{right} are the probability densities of the maximum, of the left and the right minima. Ideally, $f \sim \frac{1}{2}\left(f_{left} + f_{right}\right)$ should go to 0. We define the peak score as $W = exp(-2f^2)$. According to this definition, if $P_{left} = 1, P_{right} = 2$ and P_{max}=16, W=0.98. Whilst if $P_{left} = 13, P_{right} = 14$, the peak doesn't fit well with the WLC model and W=0.24.
Once a score is computed for each relevant peak in the L_c histogram, that score is assigned to all points in the corresponding trace. This is accomplished in two steps: first, the peak score is assigned to all points in the histogram belonging to that peak. Second, to all points with force values below 30 pN, for which an L_c values cannot be computed due to the model's limitations. To these points, we assign the score of the first successive point with force larger than 30 pN. This criterion applies only to points within 75 nm from the last point assigned to the peak. The peak width value is selected by visual inspection of traces, evaluating the maximum width of their force peaks.
The quality score of a trace, S_w, is the sum of the scores for all points in the trace. The higher the global score, the higher the trace quality. We use the ratio between the quality score and the trace length to select high quality traces. If this ratio is below 0.5, we discard the trace. We assume that if more than half of the trace is inconsistent with the WLC model, it is a low-quality trace and as such we exclude it from the analysis. While if more than half of the trace is in good agreement with the WLC model, it is possibly a meaningful trace.
We point out that the goal of blocks 1–4 is only to find dense recurrent patterns in the SMFS data: in block 5 we reevaluate the F-D curves to form the selections shown in Fig. 3 of the main text.

Block 3: Computing distances
In block 3 we quantify the similarity between the traces in order to find the recurrent pattern of unfolding within the data. To accomplish this goal, we use a modified version of the distance introduced by Marsico

et al.[1]. This distance is defined using the dynamic programming alignment score computed for a pair of traces. For two traces, *a* and *b*, the distance d_{ab} is simply:

$$d_{ab} = 1 - \frac{S_D(N_a, N_b)}{N_{max}} \tag{2}$$

where $S_D(N_a, N_b)$ is the global alignment score, N_a is the length of trace *a*, N_b is the length of trace *b*, and N_{max} is the maximum length between the two. We have modified the match/mismatch scoring function used by Marsico et al as follows:

$$M(i,j) = \begin{cases} 1 - \frac{|F_a(i) - F_b(j)|}{F_{scoring}} & if\ |F_a(i) - F_b(j)| < F_{scoring} \\ -\frac{|F_a(i) - F_b(j)|}{F_{scoring}} & otherwise \end{cases} \tag{3}$$

where $F_a(i)$ and $F_b(j)$ are the forces in points *i* and *j* in traces *a* and *b*, and $F_{scoring} = 4\sigma_{NOISE}$. In the work done by Marsico et al, $F_{scoring}$ is replaced by ΔF_{max}, which is the average of the maximum force values in the two traces. When two widely different traces have high ΔF_{max} their distance will be lower with respect to two traces with low ΔF_{max} but overall higher level of similarity. Namely, the distance magnitude depends on the ΔF_{max} value and traces with high ΔF_{max} have by definition lower distance values. It is important to note that this problem did not occur in Marsico's work since the ΔF_{max} values were uniformly distributed for all traces.
In order to gain computational efficiency, the distance is computed only for traces which differ by no more than 2 peaks in the L_c histograms or by no more than 20 % in their trace length difference.

Block 4: Density peak clustering.
The density peak clustering (DPC) algorithm[6] is used for clustering. This choice is appropriate given that a fraction of traces in the analyzed datasets correspond to statistically isolated events and DPC automatically excludes the outliers. DPC can be summarized in the following steps:

1. We compute the density of data points in the neighborhood of each point using the *k*-nearest neighbor (*k*-NN) density estimator[7]. The density is the ratio between *k* and the volume occupied by the *k* nearest neighbors:

$$\tilde{\rho}_i = \frac{k}{\omega_d r_{k,i}^d} \tag{4}$$

where *d* is the intrinsic dimension (ID) of the dataset[8], ω_d is the volume of the *d*-sphere with unitary radius and $r_{k,i}$ is the distance of point *i* from its *k*-th nearest neighbor. In DPC it is the density rank which is relevant for the final cluster assignation. Therefore, without loss of generality, we compute the density using the following equation:

$$\rho_i = -\log r_{k,i} \tag{5}$$

$\tilde{\rho}_i$ and ρ_i are related by a simple monotonic transformation and thus, have the same rank. By using equation (4) we don't have to compute the intrinsic dimension of the dataset. In order to assign bigger weight to high quality traces, we multiply ρ_i by the score-length ratio of trace *i*.

2. Next, we find the minimum distance between point *i* and any other point with higher density, denoted as δ_i:

$$\delta_i = \min_{j:\rho_j > \rho_i} d_{ij} \tag{6}$$

where d_{ij} is the distance between points *i* and *j*. δ_i is used to identify the local density maxima.

3. We identify the cluster centers as density peaks, e.g. points with high values of both ρ_i and δ_i. For each point we compute the quantity $\gamma_i = \rho_i \delta_i$. Points with high values of γ_i are good cluster center candidates. We sort all points by the value of γ_i in descending order. The first point is a cluster center. The second point is a cluster center unless its distance from the first point is smaller than $r_{cut} = 0.3$ (which represents the distance below which on average two traces are

considered as the same pattern). Regarding the third point, it is a cluster center if it is at a distance smaller than r_{cut} from the preceding two points. Following the same logic, all the points are assessed and all cluster centers are identified.

4. All points that are not cluster centers are assigned to the same cluster of the nearest point with higher density.

Block 5: Refinement and merging

The previous blocks, from 1 to 4, were optimized for finding the centers of dense patterns of unfolding in the SMFS data, but not for finding the borders of the clusters. To solve this issue, i.e. finding the F-D curves that are similar to each pattern of unfolding, we used the conventional definition of similarity (degree of superposition of F-D curves in the Force/tip-sample-separation plane) automated in the Fodis software in the tool 'fingerprint_roi'[3].

In brief, we superimposed each cluster center with its two closest neighbors creating the effective 'area of similarity' (AoS) for each cluster. The AoS is defined as the area generated by all the points of the three curves above 30 pN and before the last peak (see Supplementary Fig. 11 b), each point forming a square of 5 nm x 5 pN. Then, the SMFS curves are preliminary filtered based on their length with their final peak falling between 0.7 x *L* and 1.3 x *L* (with *L* length of the cluster center). Each of the remaining F-D curves is compared with the AoS, and the number of its points that fall within the AoS is annotated: this number constitutes the similarity score. As depicted in Supplementary Fig. 11 c, the plot of the scores in descending order interestingly forms a line with two different slopes. The change of the slope empirically defines a threshold that reflects the limit of similarity for each cluster. If two clusters share more than 40% of the traces above the threshold, they are considered the same cluster, thus merged (all the merges are reported in Supplementary Fig. 11 d).

<u>Formal derivation of the Bayesian identification of the clusters.</u>

Bayesian inference is widely used in modern science[9,10] because it allows to univocally determine the level of uncertainty of a hypothesis[11]. We used the same framework to determine the molecular identity of the unfolding clusters. In the most general terms, we observed the unfolding cluster C_X, and we want to find the probability that the unfolding of a certain protein $Prot_A$ corresponds to the unfolding cluster C_X, i.e. we want to find the posterior probability $P(Prot_A|C_X)$. In the form of the Bayes theorem:

$$P(Prot_A|C_X) = \frac{P(C_X|Prot_A)P(Prot_A)}{P(C_X)} \tag{7}$$

where $P(Prot_A)$ is the prior, i.e. the probability of $Prot_A$ to be in the sample; $P(C_X|Prot_A)$ is the likelihood, i.e. the likelihood to find a cluster with the features of C_X coming from the unfolding of $Prot_A$; and $P(C_X)$ is the normalizing factor. In the case of a classical experiment with a single purified protein, $P(Prot_A|C_X)$ is assumed to be equal to 1, but this is not the case for a native environment where there are $Prot_B$, $Prot_C$, etc.

The observables of an unfolding cluster for which we determined the likelihood functions are the contour length of the last detectable peak $Lc_{max,Cx}$ (~ length of the F-D curve), the peak profile of the cluster in the Lc space $Peaks_{Cx}$ (also called unfolding barriers), and the average unfolding force of the detected peaks $\bar{F}_{Cx}$, but the method is modular and therefore it will incorporate also other observables when available. The equation (7) becomes:

$$P\left(Prot_A \middle| Lc_{max,Cx}, \bar{F}_{Cx}\right) = \frac{P\left(Lc_{max,Cx}|Lc_{Prot_A}\right) P\left(\bar{F}_{Cx}|\bar{F}_{Prot_A}\right) P\left(Peaks_{Cx}|Structure_{Prot_A}\right) P(Prot_A)}{N} \tag{8}$$

where $N = \sum_i (P\left(Lc_{max,Cx}|Lc_{Prot_i}\right) P\left(\bar{F}_{Cx}|\bar{F}_{Prot_i}\right) P\left(Peaks_{Cx}|Structure_{Prot_A}\right) P(Prot_i))$ is the normalizing factor that takes into consideration all the proteins $Prot_i$ present in the sample. In the next paragraphs we will describe the determination of the numerator of equation (8).

Determination of prior $P(Prot_A)$

The most crucial part of the method is the determination of the list of proteins present in the sample, together with all their properties (length, abundance, secondary structure, topology, etc.). To do so we combined the Mass Spectrometry results of the cells under investigation[12–14] with other structural and topological information available in Uniprot and PDB. The crossing of the databases is done thanks to the unique Uniprot identifier. The complete list of proteins of Hippocampal neurons, Rod outer segments and Rod discs with the information necessary for the Bayesian inference are shown in the Supplementary Data of the article. In case the data of the species of interest are not available, cross species proteomic analysis demonstrated that the majority of proteins are conserved in terms of relative abundance[15,16]. $P(Prot_A)$ is the probability of finding $Prot_A$ and not $Prot_B$, $Prot_C$, etc., which corresponds to the normalized relative abundance of $Prot_A$ in the sample—a parameter that is usually calculated in mass spectrometry analysis. Indeed, *in silico* calculation of abundances gives rather trustworthy values:

1. the most accurate option is the emPAI[17];
2. if the emPAI is not available, the second best option is the spectra counting for each peptide (PSM)[18];
3. if the PSM is not available, the sequence coverage can be used as loose estimation[19].

We used the emPAI for hippocampal neurons and rods; for the discs, the emPAI does not give accurate values because of the extreme concentration of Rhodopsin, therefore we used the abundances obtained with other quantitative methods[20].

We demonstrated in Supplementary Fig 1, 2 and 3 that the isolated paches of membrane contain the membrane proteins of the original cells but not the cytoplasmic proteins, therefore we created an additional binary variable $ismembrane$ for each protein, equal to 0 if the protein is not a membrane protein, 1 otherwise. This information is extracted from the annotation in the Uniprot database. The final prior is:

$$P(Prot_A) = abundance_A \times ismembrane_A \tag{9}$$

Modeling the conditional probability $P\left(Lc_{max,Cx} | Lc_{Prot_A}\right)$

The F-D curves encode a reliable structural information, that is the total length of the unfolded protein[21]. We revisited 14 published unfolding clusters of membrane proteins[21–33] plus our own experiments, and that allowed us to create the conditional probability function for the observable $Lc_{max,Cx}$ as shown in Fig. 4 b–c. The distribution of the effective unfolding lengths forms a Gaussian bell centered at $0.89Lc_{Prot_A}$ with a standard deviation of $\sigma_{effective} = 0.05Lc_{Prot_A}$. The final standard deviation is obtained combining $\sigma_{effective}$ with the error due to the unperfect determination of the persistence length $\sigma_p = 0.07Lc_{Prot_A}$, and the error of the $Lc_{max,Cx}$ that is $\sigma_{Lc} = 0.05Lc_{Prot_A}$, so we obtained $\sigma_{Likelihood} = \sqrt{{\sigma_{effective}}^2 + {\sigma_p}^2 + {\sigma_{Lc}}^2} = 0.10Lc_{Prot_A}$. This prior is determined without considering any assumption on the pinning point of the protein (i.e. the position of the amino acid fixed at the level of the membrane that determine $Lc_{max,Cx}$), just considering the experimental data.

Most of these experimental data comes from membrane proteins that do not have a final cytoplasmic/extracellular domain. These data show also that the $0.89Lc_{Prot_A}$ correction of Fig. 4 c originates from a pinning point that can occur anywhere between the last amino acids of the last transmembrane domain and the penultimate loop.[25,30,31] There are only few examples of SMFS of proteins with a large cytoplasmic domain at one terminal[33,34], and these data suggest that the pinning point to the membrane or to another protein can be i) at the terminus end of the protein, ii) everywhere along the domain, or iii) at the end of the transmembrane segments. In other words, the pinning point of a protein that has a final domain cannot be firmly established if not within the domain range. Therefore, for the proteins that have topological annotations, we replaced the Gaussian prior described above with a non-informative prior normalized to 1 and constant between

$(L_{protein} - L_{final\ domain}) < Lc_{max,Cx} < L_{protein}$, zero elsewhere.

We use this non-informative prior only for proteins with $L_{final\ domain} \geq 2\sigma_{Likelihood}$; for proteins with small final domains $L_{final\ domain} < 2\sigma_{Likelihood}$, we used the Gaussian prior centered at $0.89Lc_{Prot_A}$. We chose the threshold of $2\sigma_{Likelihood}$ in order to take into account the uncertainty on the position of the pinning points and the errors σ_p and σ_{Lc} that are still present even in proteins with small final domains.

Determination of the conditional probability $P(Peaks_{Cx}|Structure_{Prot_A})$

Empirical evidence[21–33] and simulations[35] suggests that the occurrence of the peaks is more likely in presence of unstructured regions, i.e. in the intracellular or extracellular loops of the membrane proteins (see Fig. 4f). We used this information to evaluate the probability that the structure of the $Prot_A$ unfolds with the unfolding pattern of cluster C_X (position of the peaks). To do so, we calculated the cross-correlation between $Peaks_{Cx}$ i.e. the global histogram of cluster C_X (of point between 40 pN and 100 pN for normalization porpuses) and $Structure_{Prot_A}$ i.e. the profile of gaussian bells centered in the center of the loops of $Prot_A$ and with a FWHM of 15 nm which is an average estimate of the distance between two loops (see Fig. S13). If no structure is available we cross correlated a flat line. Both $Structure_{Prot_A}$ and $Peaks_{Cx}$ are normalized to have a total area under the profile equal to 1. Then we assigned to $P(Peaks_{Cx}|Structure_{Prot_A})$ the maximal value of the cross correlation with a relative lag of ±15 nm to allow some freedom for fine alignment.

Determination of the conditional probability $P(\bar{F}_{Cx}|\bar{F}_{Prot_A})$

The force necessary to unfold a protein domain depends on the stability of the domain itself. α-helices and β-sheets are unfolded at different force levels as shown in Fig. 4 d. We revised the unfolding forces of 22 proteins[21–33,36] and we used as $P(\bar{F}_{Cx}|\bar{F}_{Prot_A})$ the smoothed trend line of the distribution (Fig. 4 e of the main text).

Supplementary References

(1) Marsico, A.; Labudde, D.; Sapra, T.; Muller, D. J.; Schroeder, M. A Novel Pattern Recognition Algorithm to Classify Membrane Protein Unfolding Pathways with High-Throughput Single-Molecule Force Spectroscopy. *Bioinformatics* **2007**, *23* (2), e231–e236. https://doi.org/10.1093/bioinformatics/btl293.

(2) Spoerri, P. M.; Kato, H. E.; Pfreundschuh, M.; Mari, S. A.; Serdiuk, T.; Thoma, J.; Sapra, K. T.; Zhang, C.; Kobilka, B. K.; Müller, D. J. Structural Properties of the Human Protease-Activated Receptor 1 Changing by a Strong Antagonist. *Structure* **2018**, *26* (6), 829-838.e4. https://doi.org/10.1016/j.str.2018.03.020.

(3) Galvanetto, N. *Fodis: A Software for Single Molecule Force Spectroscopy*; 2018.

(4) Ainavarapu, S. R. K.; Brujić, J.; Huang, H. H.; Wiita, A. P.; Lu, H.; Li, L.; Walther, K. A.; Carrion-Vazquez, M.; Li, H.; Fernandez, J. M. Contour Length and Refolding Rate of a Small Protein Controlled by Engineered Disulfide Bonds. *Biophysical Journal* **2007**, *92* (1), 225–233. https://doi.org/10.1529/biophysj.106.091561.

(5) Petrosyan, R. Improved Approximations for Some Polymer Extension Models. *Rheol Acta* **2016**, 1–6. https://doi.org/10.1007/s00397-016-0977-9.

(6) Rodriguez, A.; Laio, A. Clustering by Fast Search and Find of Density Peaks. *Science* **2014**, *344* (6191), 1492–1496. https://doi.org/10.1126/science.1242072.

(7) Altman, N. S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician* **1992**, *46* (3), 175–185. https://doi.org/10.1080/00031305.1992.10475879.

(8) Facco, E.; d'Errico, M.; Rodriguez, A.; Laio, A. Estimating the Intrinsic Dimension of Datasets by a Minimal Neighborhood Information. *Scientific Reports* **2017**, *7* (1). https://doi.org/10.1038/s41598-017-11873-y.

(9) Wang, Q.; Chen, R.; Cheng, F.; Wei, Q.; Ji, Y.; Yang, H.; Zhong, X.; Tao, R.; Wen, Z.; Sutcliffe, J. S.; Liu, C.; Cook, E. H.; Cox, N. J.; Li, B. A Bayesian Framework That Integrates Multi-Omics Data and Gene Networks Predicts Risk Genes from Schizophrenia GWAS Data. *Nature Neuroscience* **2019**, *22* (5), 691–699. https://doi.org/10.1038/s41593-019-0382-7.

(10) Heenan, P. R.; Perkins, T. T. FEATHER: Automated Analysis of Force Spectroscopy Unbinding and Unfolding Data via a Bayesian Algorithm. *Biophysical Journal* **2018**, *115* (5), 757–762. https://doi.org/10.1016/j.bpj.2018.07.031.

(11) Jaynes, E. T. *Bayesian Methods: General Background*; 1986.

(12) Chen, P.; Li, X.; Sun, Y.; Liu, Z.; Cao, R.; He, Q.; Wang, M.; Xiong, J.; Xie, J.; Wang, X.; Liang, S. Proteomic Analysis of Rat Hippocampal Plasma Membrane: Characterization of Potential Neuronal-Specific Plasma Membrane Proteins. *Journal of Neurochemistry* **2006**, *98* (4), 1126–1140. https://doi.org/10.1111/j.1471-4159.2006.03934.x.

(13) Kwok, M. C. M.; Holopainen, J. M.; Molday, L. L.; Foster, L. J.; Molday, R. S. Proteomics of Photoreceptor Outer Segments Identifies a Subset of SNARE and Rab Proteins Implicated in Membrane Vesicle Trafficking and Fusion. *Molecular & Cellular Proteomics* **2008**, *7* (6), 1053–1066. https://doi.org/10.1074/mcp.M700571-MCP200.

(14) Panfoli, I.; Musante, L.; Bachi, A.; Ravera, S.; Calzia, D.; Cattaneo, A.; Bruschi, M.; Bianchini, P.; Diaspro, A.; Morelli, A.; Pepe, I. M.; Tacchetti, C.; Candiano, G. Proteomic Analysis of the Retinal Rod Outer Segment Disks. *Journal of Proteome Research* **2008**, *7* (7), 2654–2669. https://doi.org/10.1021/pr7006939.

(15) Bayram, H. L.; Claydon, A. J.; Brownridge, P. J.; Hurst, J. L.; Mileham, A.; Stockley, P.; Beynon, R. J.; Hammond, D. E. Cross-Species Proteomics in Analysis of Mammalian Sperm Proteins. *Journal of Proteomics* **2016**, *135*, 38–50. https://doi.org/10.1016/j.jprot.2015.12.027.

(16) Wright, J. C.; Beynon, R. J.; Hubbard, S. J. Cross Species Proteomics. In *Proteome Bioinformatics*; Hubbard, S. J., Jones, A. R., Eds.; Humana Press: Totowa, NJ, 2010; Vol. 604, pp 123–135. https://doi.org/10.1007/978-1-60761-444-9_9.

(17) Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. Exponentially Modified Protein Abundance Index (EmPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Mol Cell Proteomics* **2005**, *4* (9), 1265–1272. https://doi.org/10.1074/mcp.M500061-MCP200.

(18) Liu, H.; Sadygov, R. G.; Yates, J. R. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Anal. Chem.* **2004**, *76* (14), 4193–4201. https://doi.org/10.1021/ac0498563.

(19) Florens, L.; Washburn, M. P.; Raine, J. D.; Anthony, R. M.; Grainger, M.; Haynes, J. D.; Moch, J. K.; Muster, N.; Sacci, J. B.; Tabb, D. L.; Witney, A. A.; Wolters, D.; Wu, Y.; Gardner, M. J.; Holder, A. A.; Sinden, R. E.; Yates, J. R.; Carucci, D. J. A Proteomic View of the *Plasmodium Falciparum* Life Cycle. *Nature* **2002**, *419* (6906), 520. https://doi.org/10.1038/nature01107.

(20) Milo, R.; Phillips, R. *Cell Biology by the Numbers*; 2016.

(21) Oesterhelt, F.; Oesterhelt, D.; Pfeiffer, M.; Engel, A.; Gaub, H. E.; Müller, D. J. Unfolding Pathways of Individual Bacteriorhodopsins. *Science* **2000**, *288* (5463), 143–146. https://doi.org/10.1126/science.288.5463.143.

(22) Kawamura, S.; Gerstung, M.; Colozo, A. T.; Helenius, J.; Maeda, A.; Beerenwinkel, N.; Park, P. S.-H.; Müller, D. J. Kinetic, Energetic, and Mechanical Differences between Dark-State Rhodopsin and Opsin. *Structure* **2013**, *21* (3), 426–437. https://doi.org/10.1016/j.str.2013.01.011.

(23) Cisneros, D. A.; Oesterhelt, D.; Müller, D. J. Probing Origins of Molecular Interactions Stabilizing the Membrane Proteins Halorhodopsin and Bacteriorhodopsin. *Structure* **2005**, *13* (2), 235–242. https://doi.org/10.1016/j.str.2004.12.005.

(24) Sapra, K. T.; Damaghi, M.; Köster, S.; Yildiz, Ö.; Kühlbrandt, W.; Muller, D. J. One β Hairpin after the Other: Exploring Mechanical Unfolding Pathways of the Transmembrane B-Barrel Protein OmpG. *Angewandte Chemie International Edition* **2009**, *48* (44), 8306–8308. https://doi.org/10.1002/anie.200904361.

(25) Möller, C.; Fotiadis, D.; Suda, K.; Engel, A.; Kessler, M.; Müller, D. J. Determining Molecular Forces That Stabilize Human Aquaporin-1. *Journal of Structural Biology* **2003**, *142* (3), 369–378. https://doi.org/10.1016/S1047-8477(03)00066-2.

(26) Kedrov, A.; Ziegler, C.; Janovjak, H.; Kühlbrandt, W.; Müller, D. J. Controlled Unfolding and Refolding of a Single Sodium-Proton Antiporter Using Atomic Force Microscopy. *Journal of Molecular Biology* **2004**, *340* (5), 1143–1152. https://doi.org/10.1016/j.jmb.2004.05.026.

(27) Thoma, J.; Bosshart, P.; Pfreundschuh, M.; Müller, D. J. Out but Not In: The Large Transmembrane β-Barrel Protein FhuA Unfolds but Cannot Refold via β-Hairpins. *Structure* **2012**, *20* (12), 2185–2190. https://doi.org/10.1016/j.str.2012.10.006.

(28) Serdiuk, T.; Balasubramaniam, D.; Sugihara, J.; Mari, S. A.; Kaback, H. R.; Müller, D. J. YidC Assists the Stepwise and Stochastic Folding of Membrane Proteins. *Nat Chem Biol* **2016**, *12* (11), 911–917. https://doi.org/10.1038/nchembio.2169.

(29) Ge, L.; Villinger, S.; Mari, S. A.; Giller, K.; Griesinger, C.; Becker, S.; Müller, D. J.; Zweckstetter, M. Molecular Plasticity of the Human Voltage-Dependent Anion Channel Embedded Into a Membrane. *Structure* **2016**, *24* (4), 585–594. https://doi.org/10.1016/j.str.2016.02.012.

(30) Thoma, J.; Ritzmann, N.; Wolf, D.; Mulvihill, E.; Hiller, S.; Müller, D. J. Maltoporin LamB Unfolds β Hairpins along Mechanical Stress-Dependent Unfolding Pathways. *Structure* **2017**, *25* (7), 1139-1144.e2. https://doi.org/10.1016/j.str.2017.05.010.

(31) Bosshart, P. D.; Iordanov, I.; Garzon-Coral, C.; Demange, P.; Engel, A.; Milon, A.; Müller, D. J. The Transmembrane Protein KpOmpA Anchoring the Outer Membrane of Klebsiella Pneumoniae Unfolds and Refolds in Response to Tensile Load. *Structure* **2012**, *20* (1), 121–127. https://doi.org/10.1016/j.str.2011.11.002.

(32) Klyszejko, A. L.; Shastri, S.; Mari, S. A.; Grubmüller, H.; Muller, D. J.; Glaubitz, C. Folding and Assembly of Proteorhodopsin. *Journal of Molecular Biology* **2008**, *376* (1), 35–41. https://doi.org/10.1016/j.jmb.2007.11.030.

(33) Maity, S.; Mazzolini, M.; Arcangeletti, M.; Valbuena, A.; Fabris, P.; Lazzarino, M.; Torre, V. Conformational Rearrangements in the Transmembrane Domain of CNGA1 Channels Revealed by Single-Molecule Force Spectroscopy. *Nature Communications* **2015**, *6*, 7093. https://doi.org/10.1038/ncomms8093.

(34) Tanuj Sapra, K.; Park, P. S.-H.; Filipek, S.; Engel, A.; Müller, D. J.; Palczewski, K. Detecting Molecular Interactions That Stabilize Native Bovine Rhodopsin. *Journal of Molecular Biology* **2006**, *358* (1), 255–269. https://doi.org/10.1016/j.jmb.2006.02.008.

(35) Yamada, T.; Yamato, T.; Mitaku, S. Forced Unfolding Mechanism of Bacteriorhodopsin as Revealed by Coarse-Grained Molecular Dynamics. *Biophysical Journal* **2016**, *111* (10), 2086–2098. https://doi.org/10.1016/j.bpj.2016.09.051.

(36) Hoffmann, T.; Dougan, L. Single Molecule Force Spectroscopy Using Polyproteins. *Chemical Society Reviews* **2012**, *41* (14), 4781. https://doi.org/10.1039/c2cs35033e.

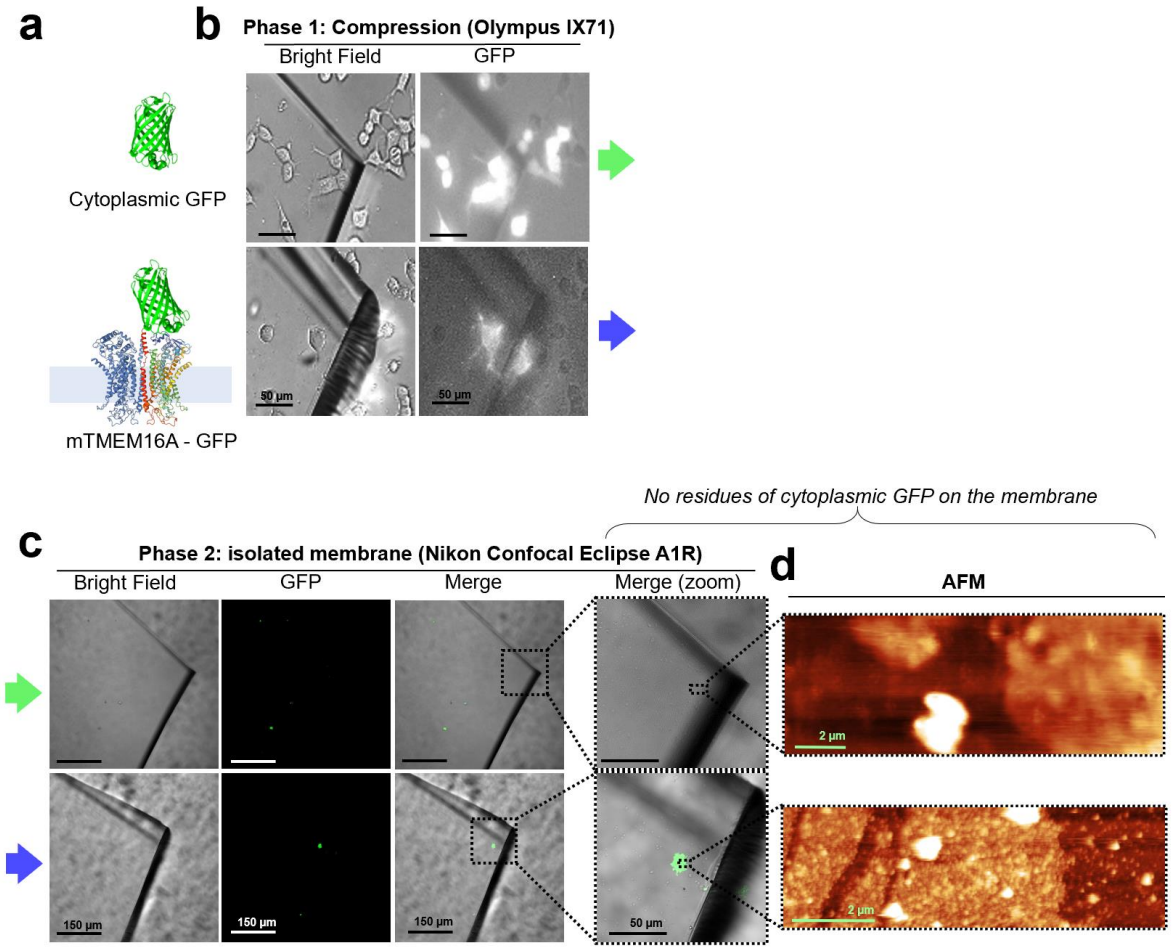


Figure S1
In NG-108 cells, membrane proteins remain into the membrane after unroofing, while cytoplasmic proteins do not.
a, (top) cytoplasmic GFP and (bottom) mTMEM16A-GFP overexpressed in NG-108 cells. **b**, bright field and fluorescence images taken during the compression on the inverted microscope-AFM system. **c**, images taken with a confocal microscope of the coverslip quarter after unroofing; the rough surface under the glass is due to the Vaseline layer used to fix the coverslip quarter. No residues of cytoplasmic GFP are found in the top row. **d**, AFM images of the area in the right most panel of **c** showing the presence of membrane patches in both the samples.

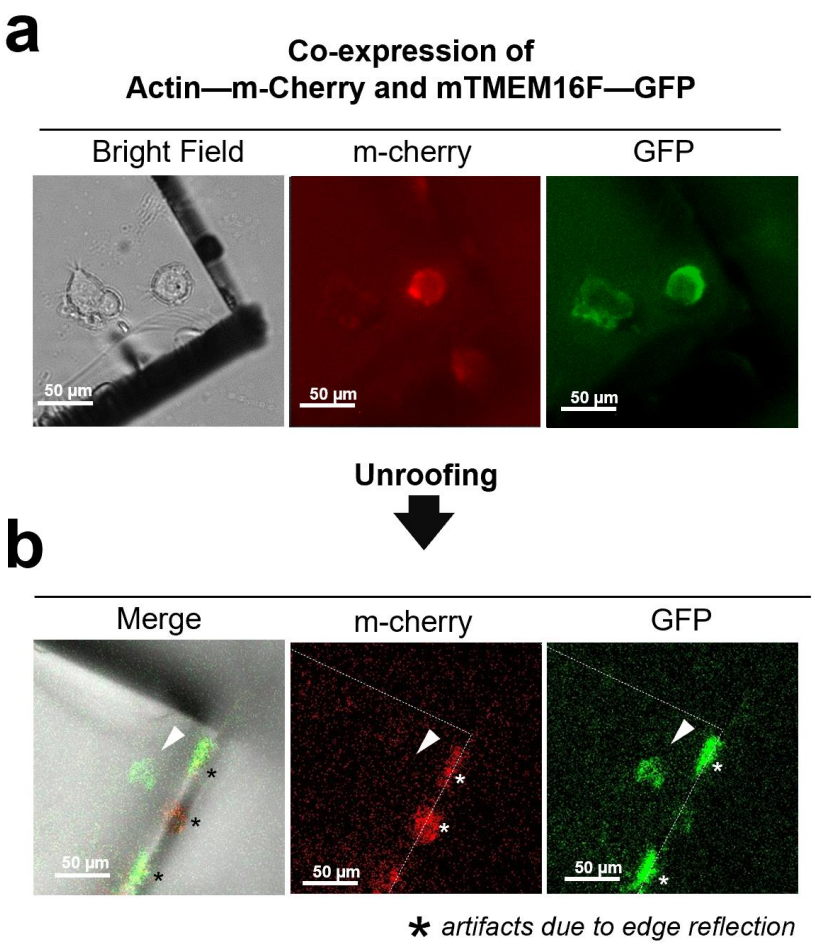


Figure S2
Absence of cytoplasmic proteins after unroofing of NG-108 cells.
a: before unroofing. From left to right: bright field image showing a piece of glass over a neuroblastoma cells transfected with actin labelled with m-Cherry and the membrane protein TMEM16F tagged with GFP; in the center an image of the red fluorescence emitted by the labelled actin; in the right an image of the green fluorescence emitted by the TMEM16F-GFP. **b:** after unroofing only the green signal from TMEM16F is detected (right panel) and no red fluorescence is observed (center panel and the location indicated by the white arrowhead).
***** (asterisks) indicate the fluorescence due to edge reflections.
The green signal is detected exactly where the neuroblastoma cell was located (left panel).

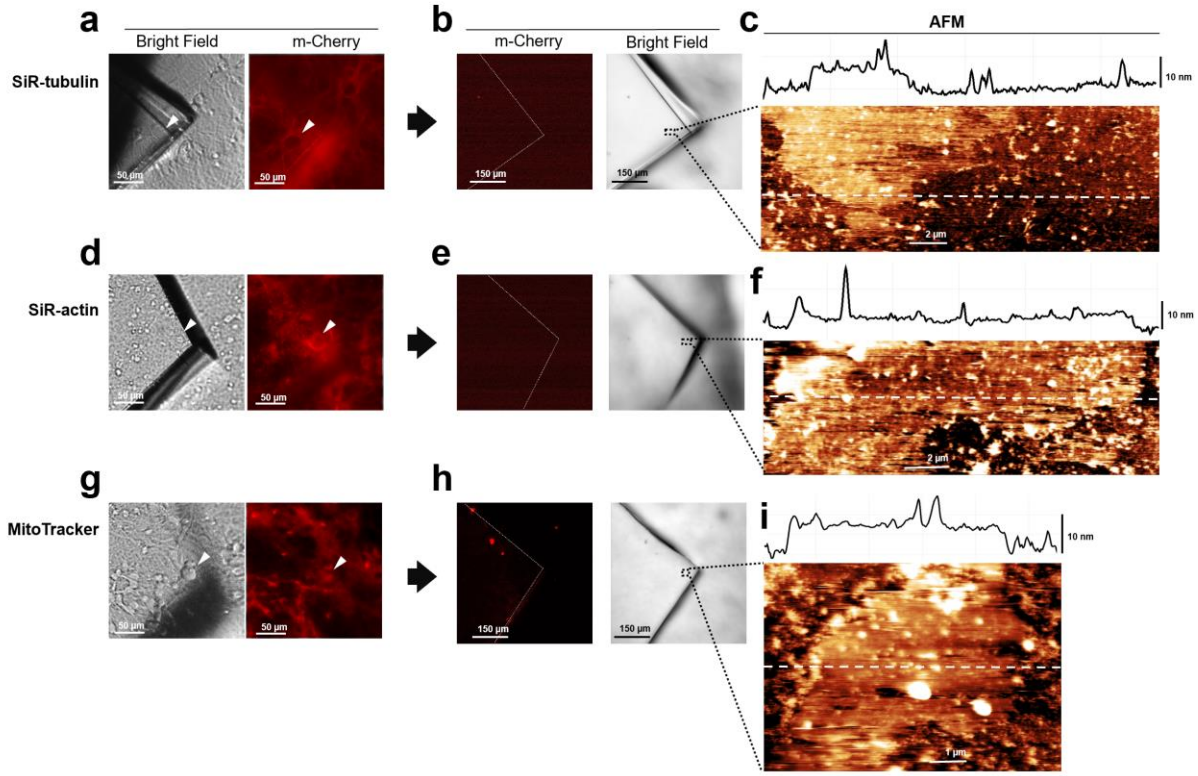


Figure S3
Absence of cytoplasmic proteins in unroofed membranes in hippocampal cells.
a: before unroofing. Bright field image of the piece of glass over a neuroblastoma cell. Red fluorescence emitted by SiR-tubulin**. b:** after unroofing. No red fluorescent signal detected. **c**: AFM imaging of region indicated in **b** showing the presence of a piece of membrane. **d, e** and **f** as in **a, b** and **c** but for Sir-Actin. **g, h** and **i** as in **a, b** and **c** but for Mito Tracker labelling mitochondria.

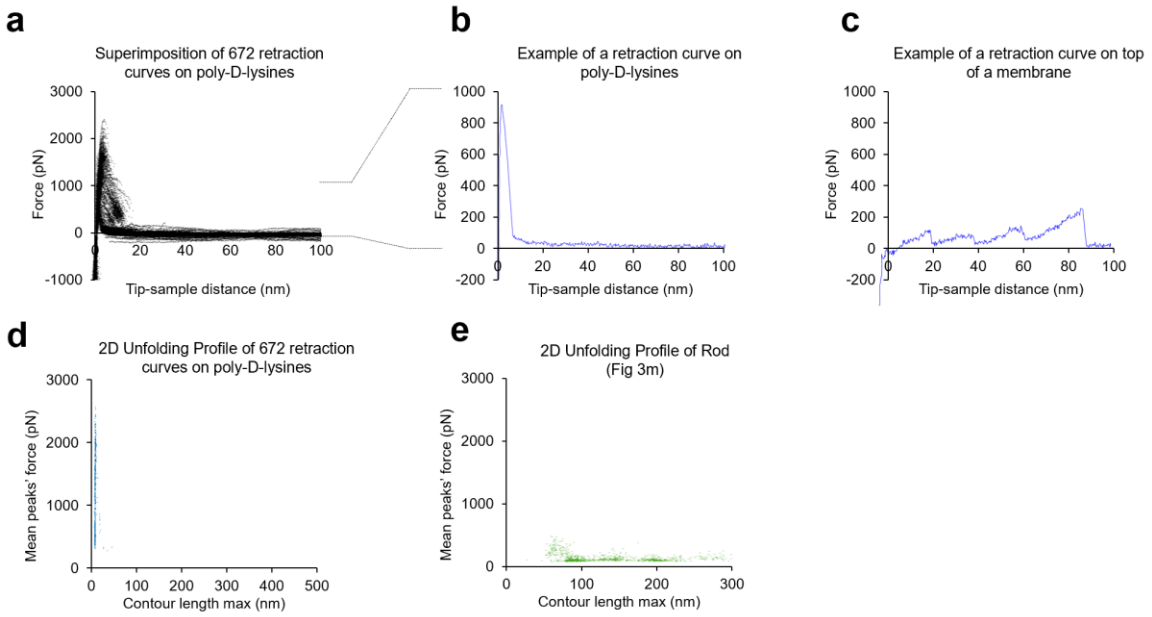


Figure S4
Absence of Poly-D-lysine contamination in the SMFS data from unroofed cell membranes.
a 672 F-d curves obtained from a cover slip just coated with Poly-D-Lysine and in the absence of any cell. **b**: example of one F-d curve from **a**. **c**: one example of a F-d curve obtained in the presence of a biological membrane. **d**: plot of the mean force against maximum contour length of the traces obtained from the experiment shown in **a**; **e**: as in **d** but for F-d curves obtained from rod membranes.

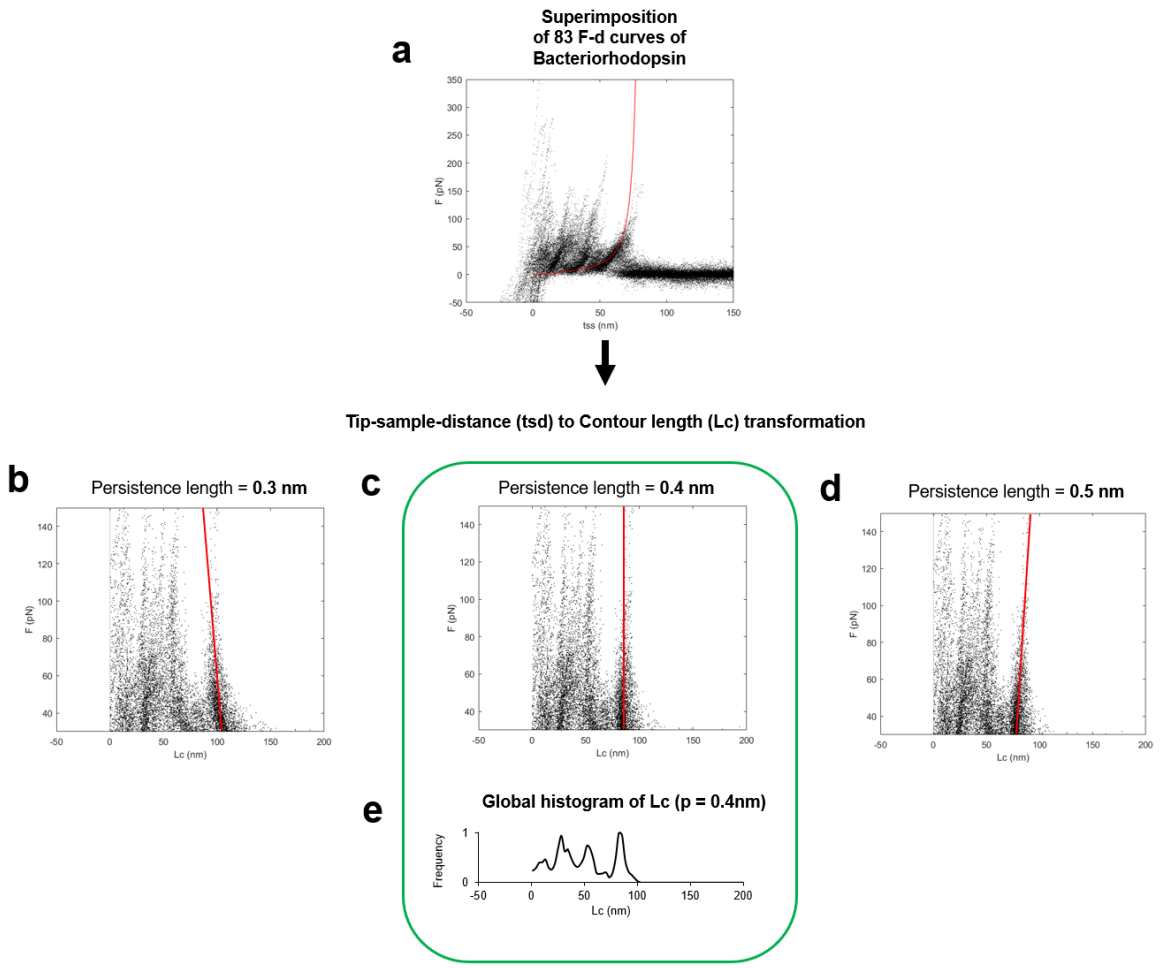


Figure S5
Determination of the optimal persistence length of the detachment peak used for the total contour length calculation (Lc max).
a: superimposition of 83 F-d curves of Bacteriorhodopsin (the red line is the worm like chain curve with persistence length = 0.4 nm and contour length = 85 nm). **b, c, d:** Transformation of the ordinates of the points in **a** to the Contour length space (Lc) with persistence length of 0.3 nm in **b,** 0.4 in **c** and 0.5 in **d**. Solid red lines indicate the best fit of the points belonging to the last unfolding peak: 0.4 nm of persistence length is the correct/optimal value because in the contour length space, the force points representing the unfolding of the polymer with a constant contour length must lay on a vertical line. **d:** Global histogram of contour length of **c**.

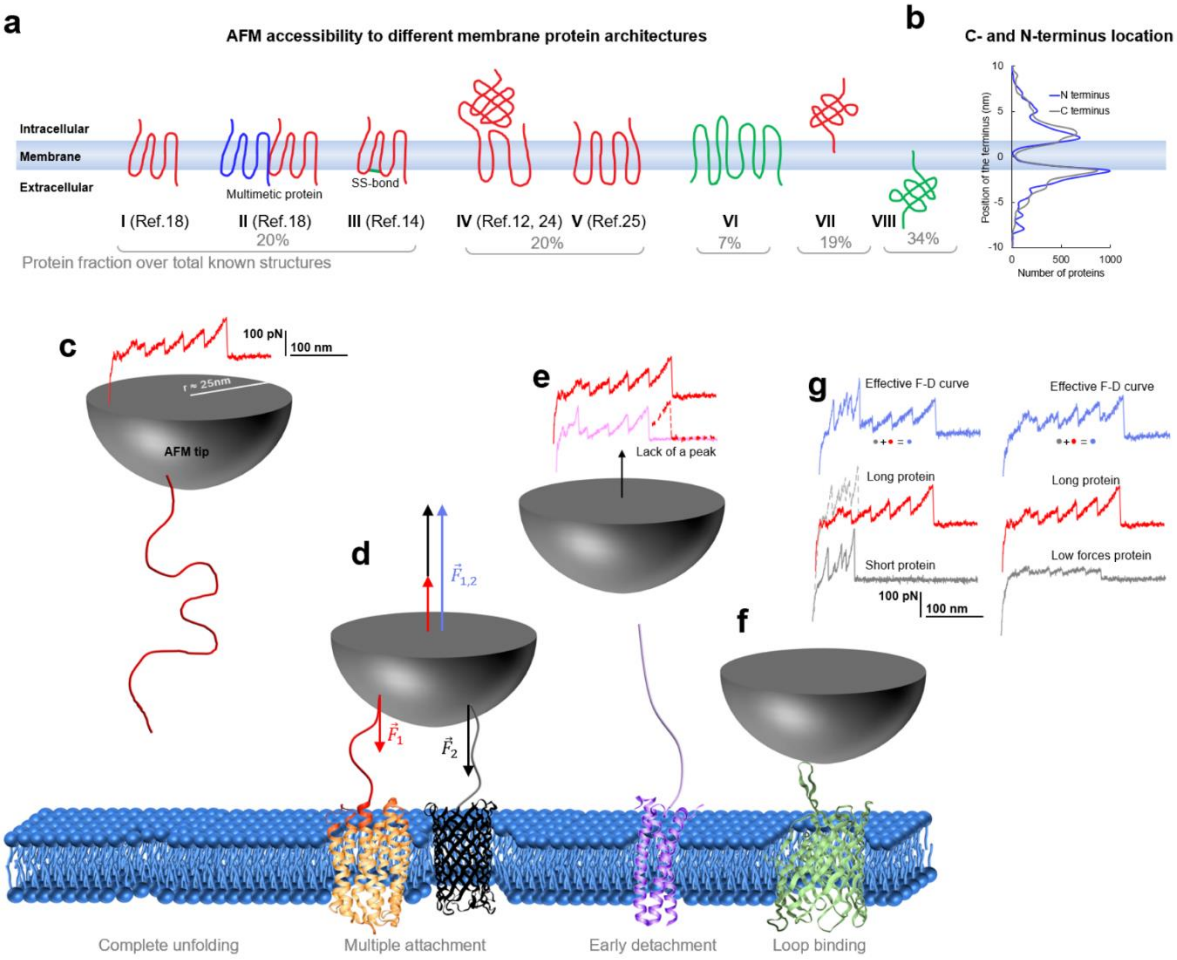


Figure S6
Membrane proteins architectures.
a, eight classes of membrane proteins and their fraction obtained from data present in the PDB-OPM. **b**, distribution of the position of the C and N termini relative to the center of the cell membrane along the axis perpendicular to the membrane. **c**, Cartoon representing the complete unfolding of a membrane protein and its F-d curve, **d**, simultaneous unfolding of two proteins and the resulting involved force which is the arithmetic sum of the one necessary to unfold one protein ant the other. **e**, incomplete unfolding of a protein leads to a similar but shorter unfolding curve, **f**, unfolding from a cytoplasmic loop. **g**, idealized decomposition of F-D curves of a multiple unfolding/ unfolding from a loop. Effective F-d curves are affected by multiple binding: If one of the two protein is short, the unfolding of the longer one can still be recognized although it will differ in the intersection part. If one protein unfolds with lower forces that the other, still the patter of the second one can be recognized, but it will be noisy.

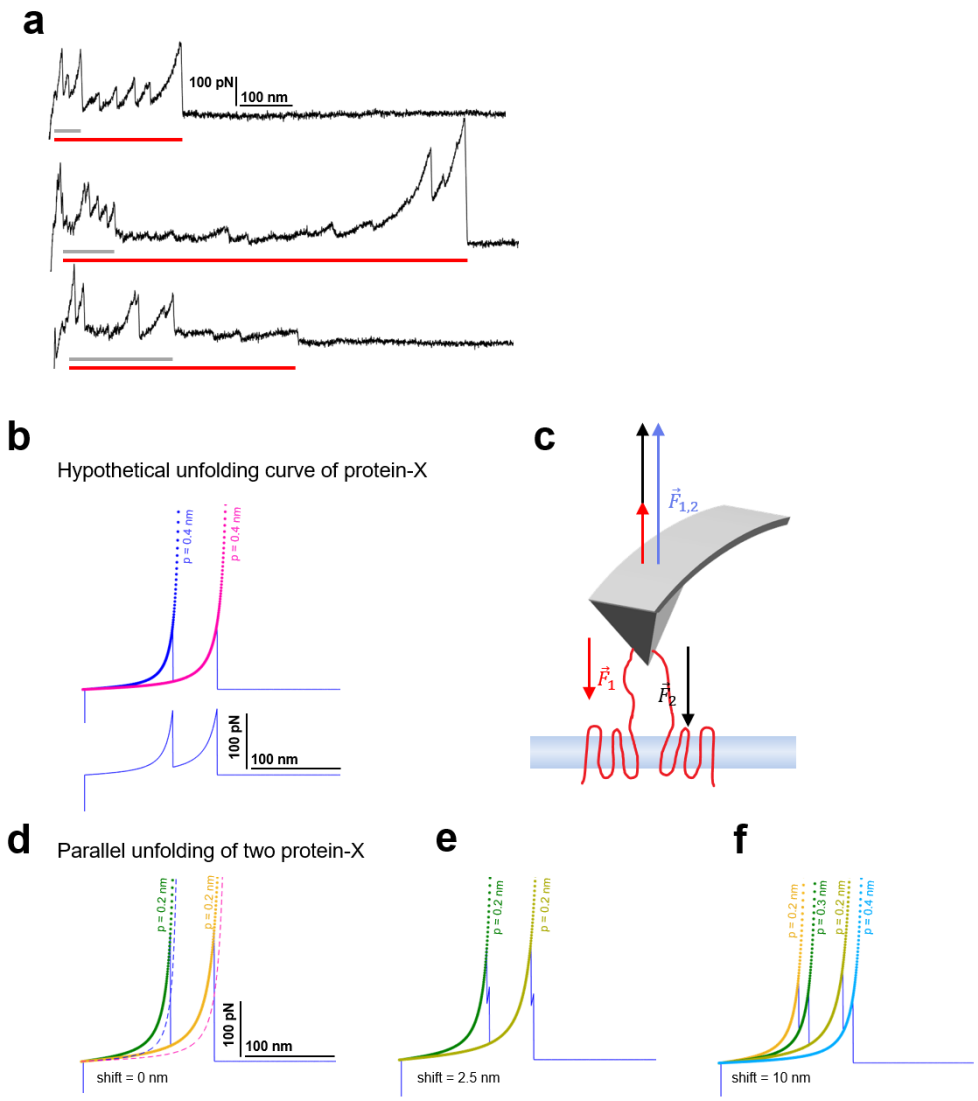


Figure S7
Candidates of multiple unfolding and origin of persistence length deviation.
a, observed F-d curves showing the features of multiple unfolding events (red horizontal bar: long protein, gray horizontal bar: short protein; guessed). **b**, hypothetical unfolding curve of protein-X (peak 1: Lc=100 nm, F=150pN; peak 2: Lc=150 nm, F=150pN) fitted with the WLC model with standard persistence length p=0.4nm. **c**, the force applied by the AFM tip balances the unfolding forces of the two proteins during the retraction. **d**, the effective F-d curve recorded during parallel unfolding of two protein-X corresponds to the sum of a single unfolding curve **b** and it is best fitted with half the persistence length p=0.2nm. **e**, relative shift of 2.5 nm and **f**, 10 nm still result in deviations of the measured persistence length and display the doubling of the peaks.

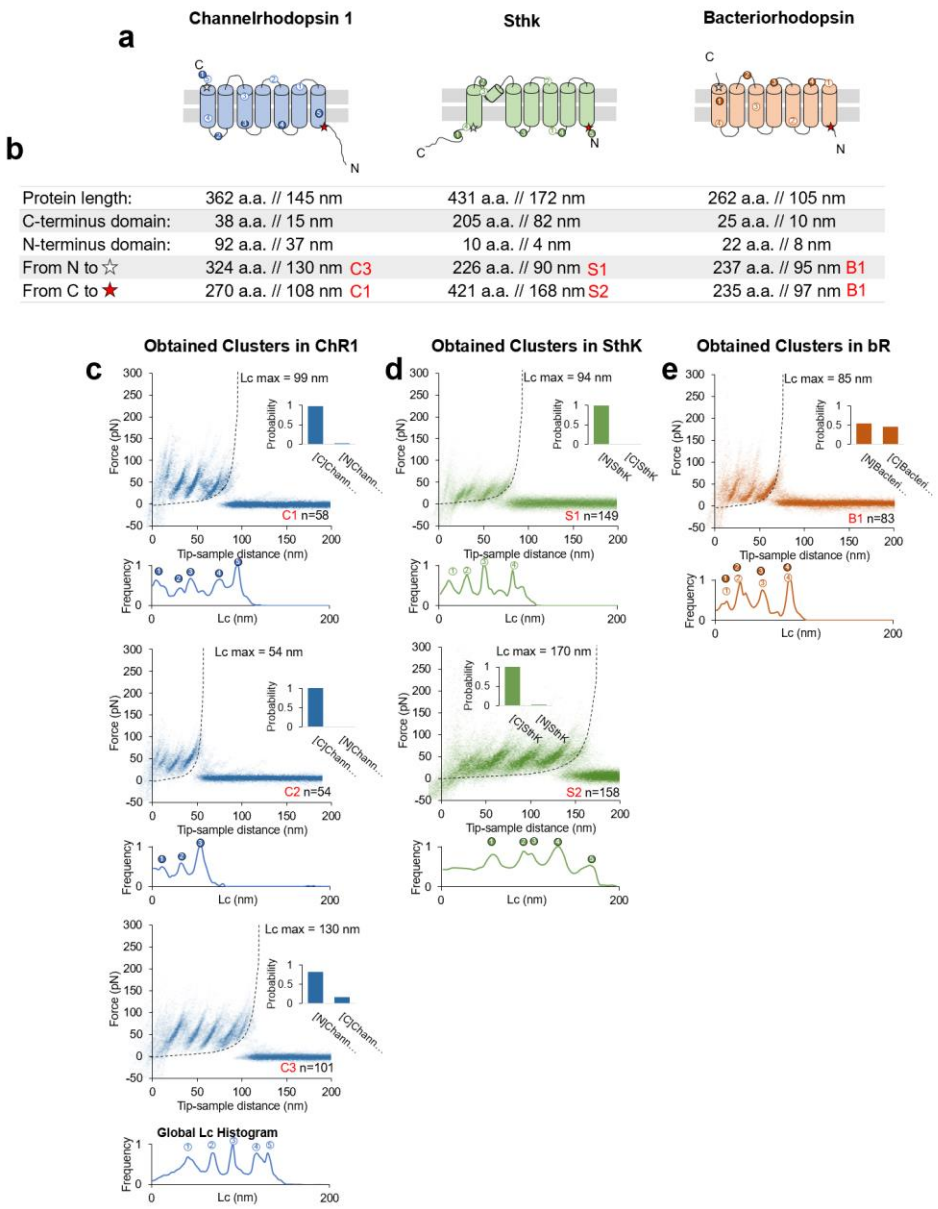


Figure S8

Unfolding of reconstituted membrane proteins.

a: scheme of the structure of the three purified proteins used in the in vitro preliminary step: ChR1, Sthk and bR (cylinders represent alpha helices). **b**: geometrical length of the amino-acid chains constituting the whole protein, the C and N-terminals and of the length of the whole protein minus the C and N-terminal. The codes in red represent the name un the respective unfolding cluster. **c**: clusters obtained from a dataset of n=18868 traces from a sample with ChR1. The histogram bars on the right of each density plot represent the Bayesian assignment according to our procedure (described in Methods and in Fig. 6). [N] and [C] stand for "protein unfolded from N-terminus/C-terminus". Below each density plot the corresponding Global histogram of contour length showing the most likely position for the unfolding peaks to occur: the circles with numbers are then reported in **a** to show the position of the preferential unfolding position in the structure. **d**: clusters obtained from a dataset of n=115155 traces from a sample with SthK. **e**: clusters obtained from a dataset of n=35599 traces from a sample with bR.

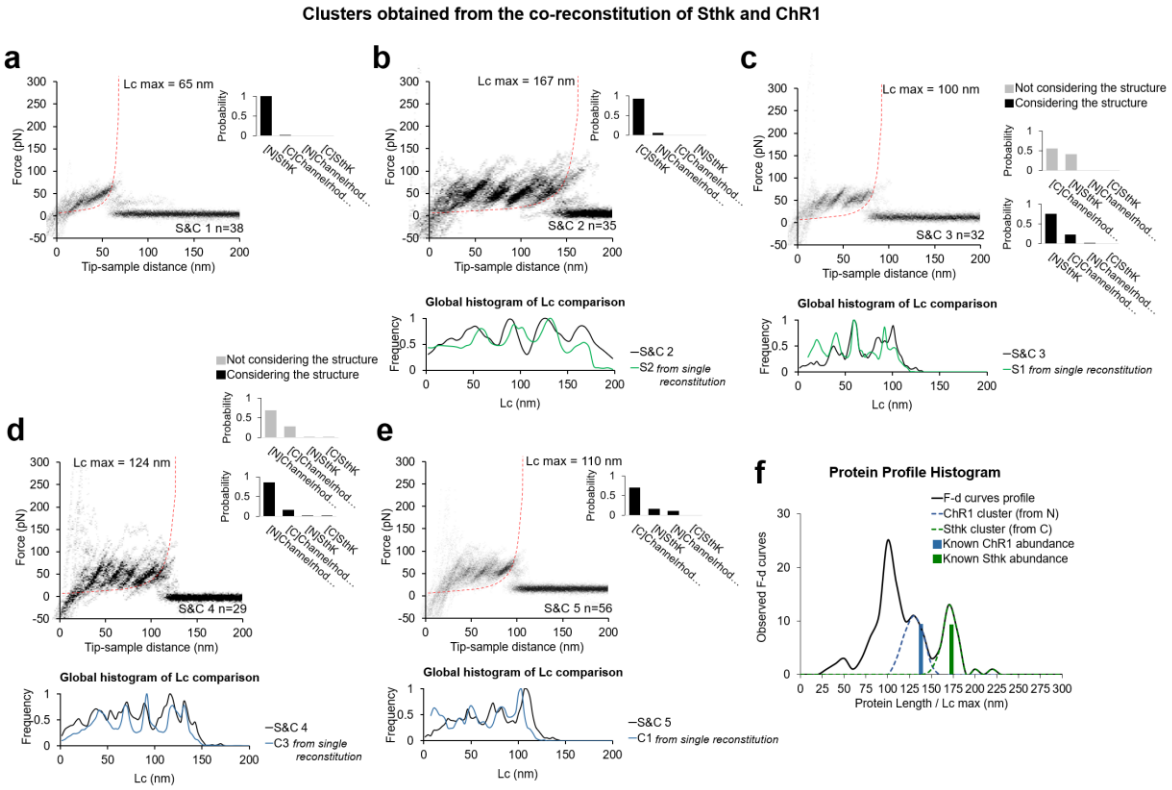


Figure S9

In vitro experiments with co-reconstitution of SthK and ChR1 (relative concentration 1:1) showing the five clusters obtained, and comparison of these new patterns S&Cx to the previous Cx and Sx clusters from single reconstitution.

a–e Clusters (Called S&C 1, …, S&C 5) resulting from the simultaneous reconstitution of ChR1 and SthK in a concentration ratio of 1:1, and Global histogram of the contour length comparison with the best candidate obtained from the single reconstitution (from Fig. 8). Bayesian probability shown on the right of each cluster. For **c** and **d** we report the improvement in the Bayesian identification when the structure of the protein is taken into account (see methods Bayesian estimation). **f:** black line: distribution of the maximum of the contour length (Lc max) of F-d curves of the clusters (except **a**) obtained in the co-reconstituted sample; blue and green bars: real length of the protein in the sample.

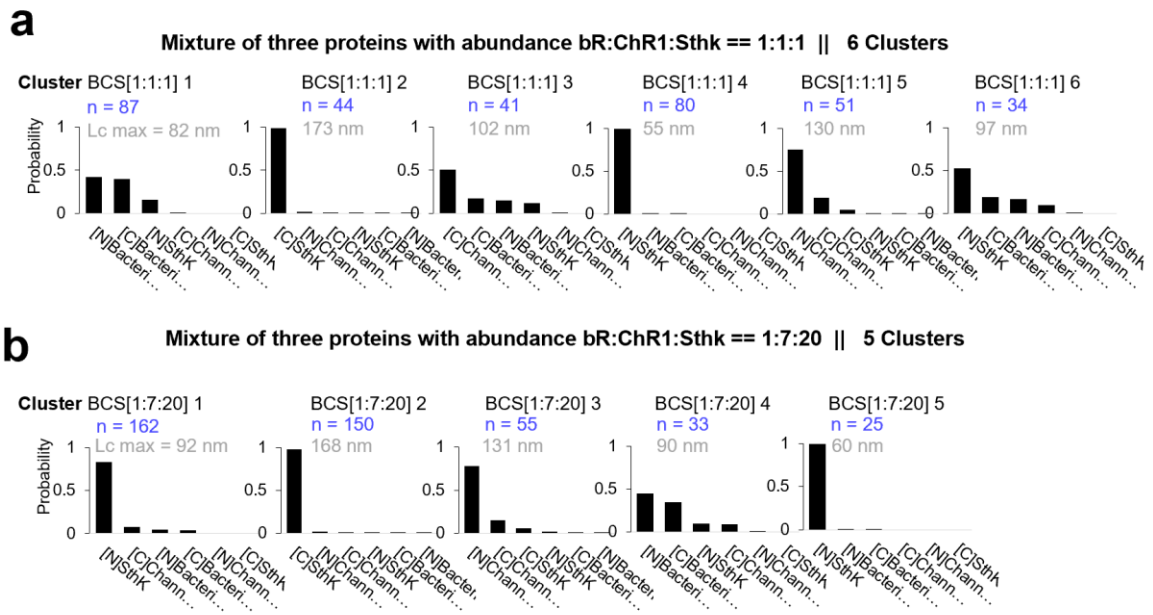
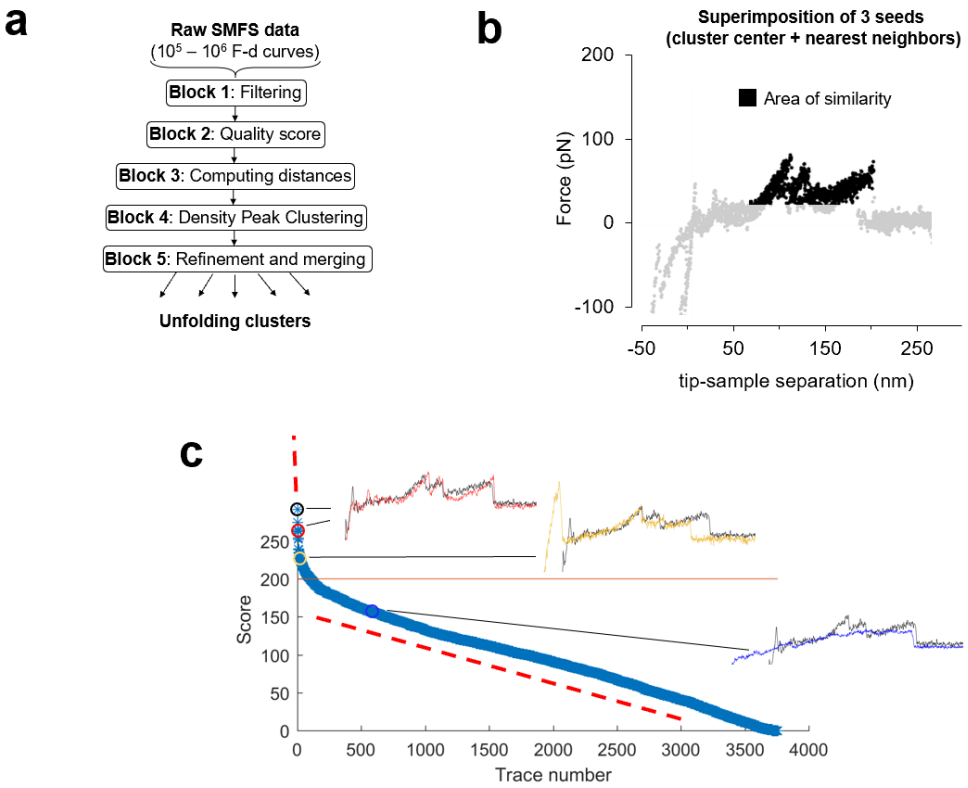


Figure S10
Clusters obtained in *in vitro* experiments with mixtures of bR, ChR1 and Sthk at different relative concentrations [1:1:1] and [1:7:20].
Cluster names: BCS[1:1:1]x and BCS[1:7:20]x.
a: candidates' prediction of the clusters obtained in samples with bR, ChR1 and Sthk and relative abundance of [1:1:1]. (The number in gray indicates the total contour length of the clusters called Lc max, the number in blue the number of traces per cluster). **b:** same as **a** for relative abundance of [1:7:20].



d

DRG	C. merged	Hippo.	C. merged	Rod	C. merged	Disc	C. merged
drg1	1	h1	1	r1	2	disc1	1
drg2	1	h2	1	r2	2	disc2	4
drg3	2	h3	7	r3	2	disc3	1
drg4	1	h4	2	r4	2	disc4	3
drg5	1	h5	1	r5	2	disc5	1
drg6	3	h6	1	r6	1		
drg7	1	h7	1	r7	1		
drg8	1	h8	1	r8	1		
drg9	1	h9	4				
drg10	1	h10	1				
drg11	1						
drg12	1						
drg13	1						
drg14	1						
drg15	1						

Figure S11
Clustering.
a, block scheme of the clustering method. **b**, area of similarity (AoS) for cluster R1 used for block 5 described in Methods. **c**, plot of the scores generated by the AoS on **b** of the F-d curves in the Dataset Rod filtered by length in the descending order. **d**, table showing the number of clusters that was merged to form the final selection of Fig. 3.

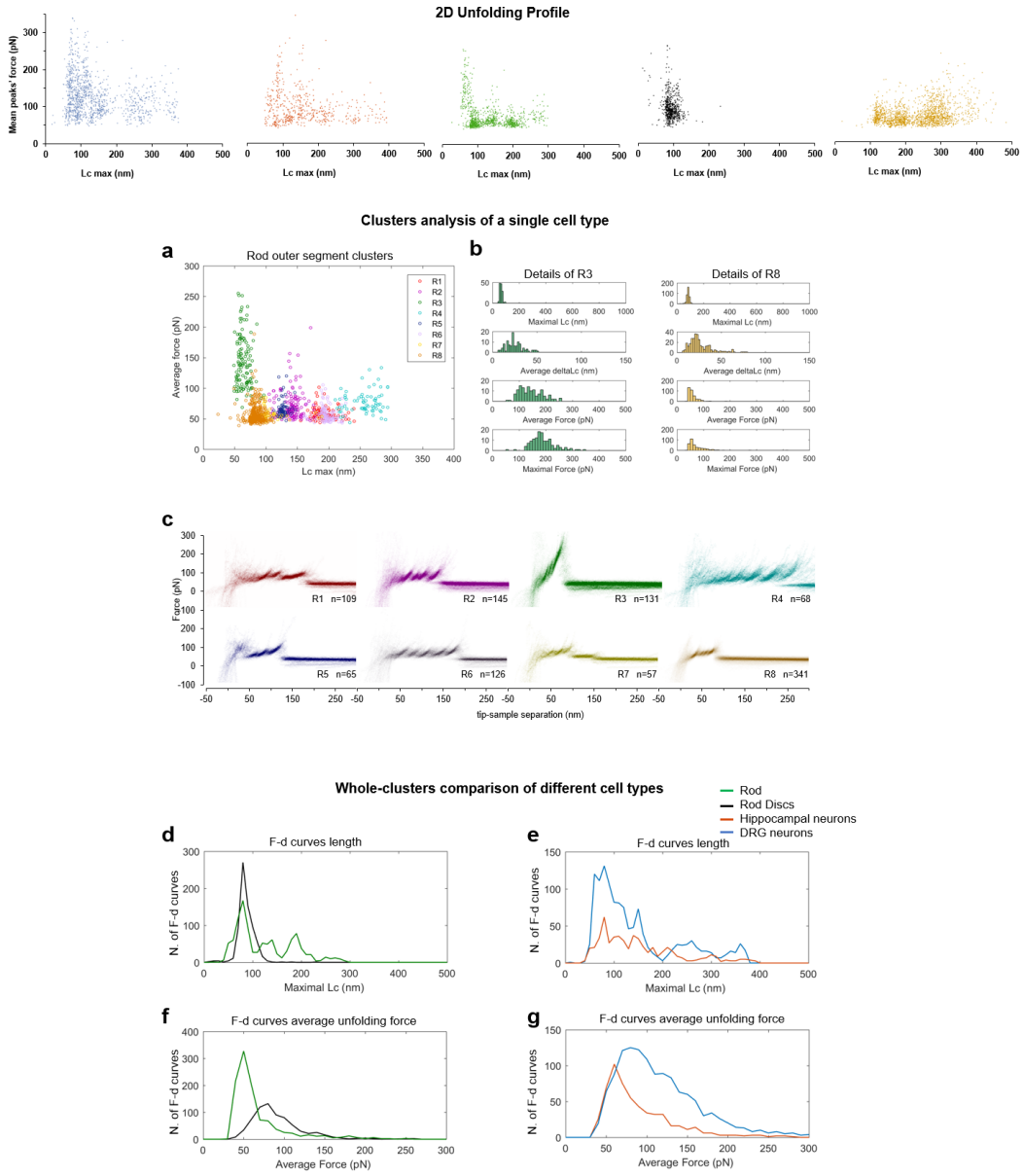


Figure S12

Alternative visualizations of cluster analysis.

Top, two dimensional unfolding profile in the compact representation of all the clustered F-D curves: x-axis is the maximum contour length; y-axis is the average unfolding force (DRG: n = 1255; hippocampus: n = 563; rod: n = 1039; disc: n = 703, NG108: n = 1591)**. a**, plot of all the F-d curves belonging to the clusters of the rod outer segment plotted with different colors in the 'average unfolding force' vs 'maximal contour length' space. **b**, distribution of representative observables for clusters R3 and R8. **c**, density plots of clusters in **a**. Comparison of the maximal contour length profiles (**d** and **e**) and of the average unfolding forces (**f** and **g**) of the four cell types investigated.

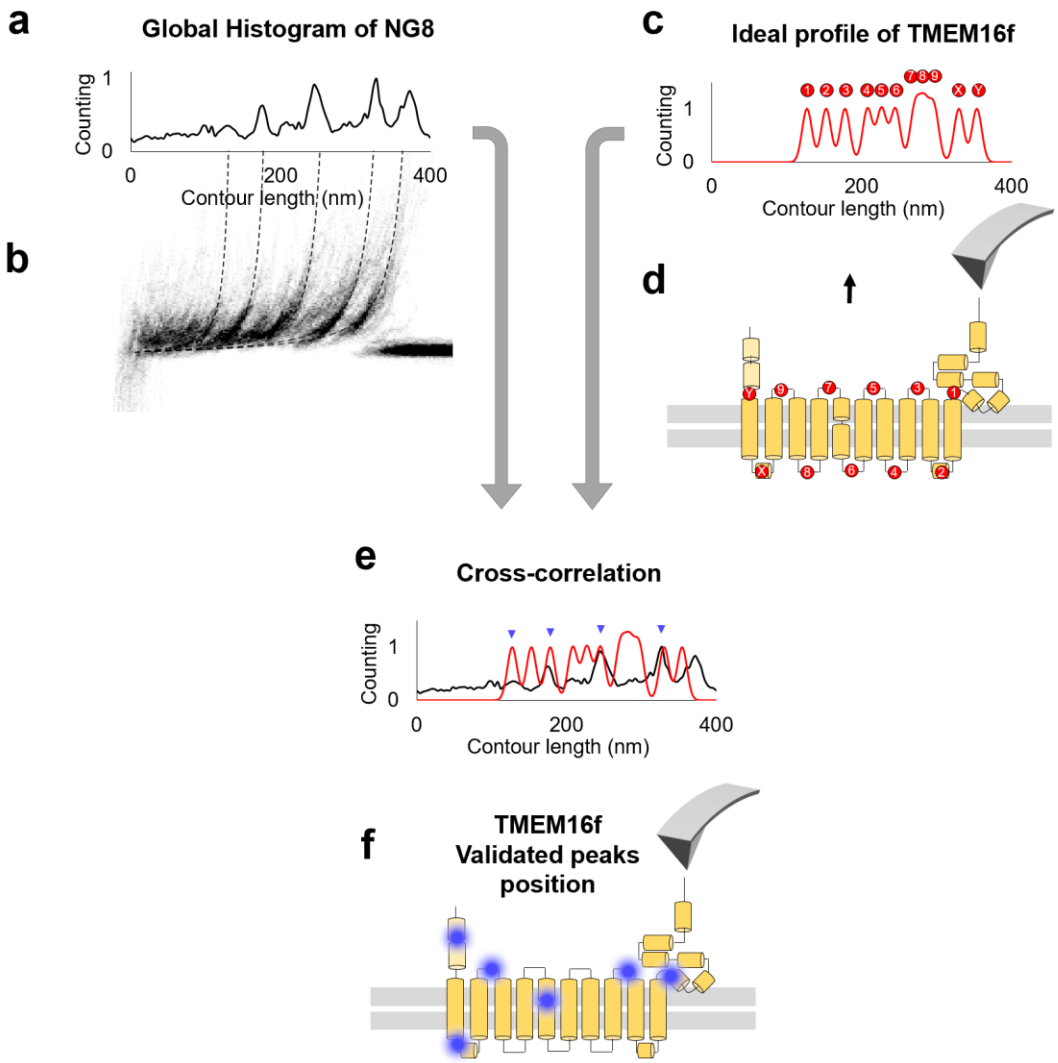


Figure S13

The cross-correlation used to evaluate $P(\mathrm{Peaks}_{\mathrm{Cx}}|\mathrm{Structure}_{\mathrm{Prot_A}})$.

a: Global Histogram of cluster NG8 and the density plot (**b**). The peaks in **a** represent the region where the peaks tend to occur. **c**: ideal profile of TMEM16F built starting from the information that 86% of the peaks happen in the loops (**d**). **e:** superimposition of the real and the idealized profile: the correlation between the two generates a number that is used in the probability estimation. The more peaks correspond to loops, the higher is the cross-correlation value. **f**: Blue dots indicate the location of force peaks in the Cryo-EM structure of the TMEM16F.

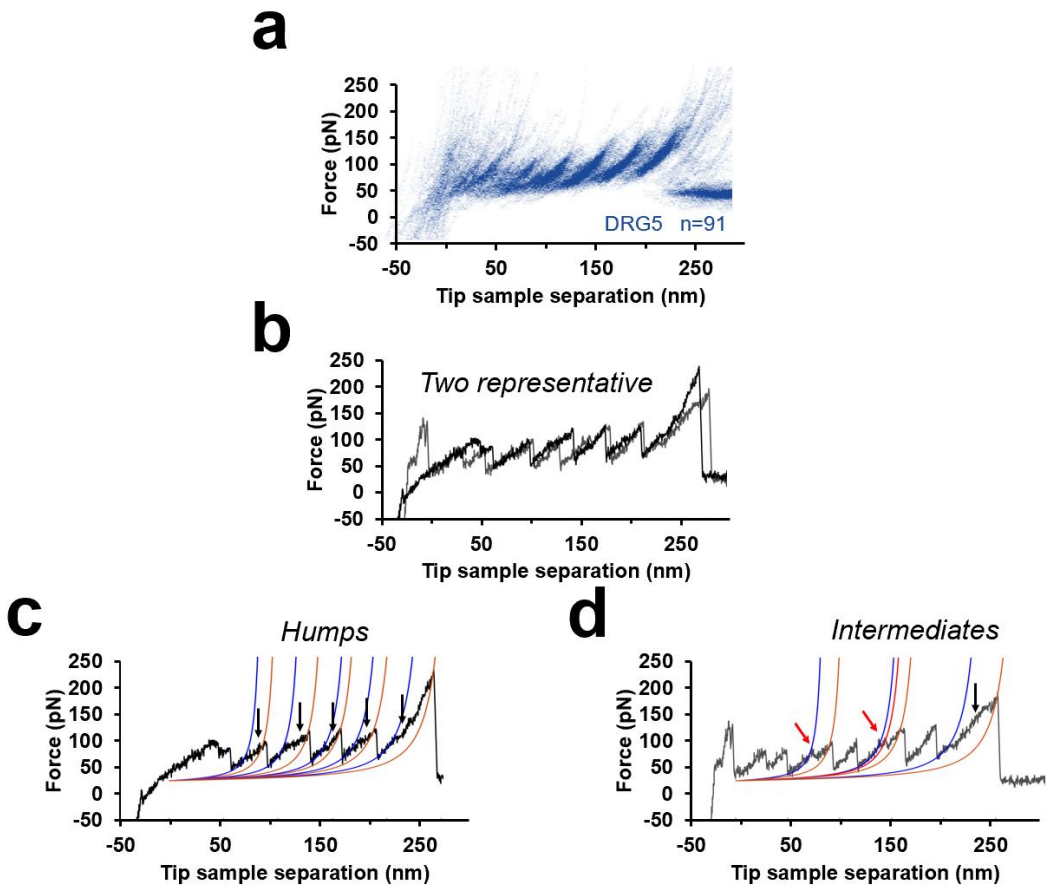


Figure S14
Unfolding intermediates and humps
a, cluster DRG5 from Fig. 3. b, two representative F-D curves from DRG5 (black and gray). c, Black curve in b that shows humps (blue WLC fit identify the beginning of the humps, gold WLC fit identify the end). d, gray curve in b that shows multiple peaks (red arrows) instead of humps.

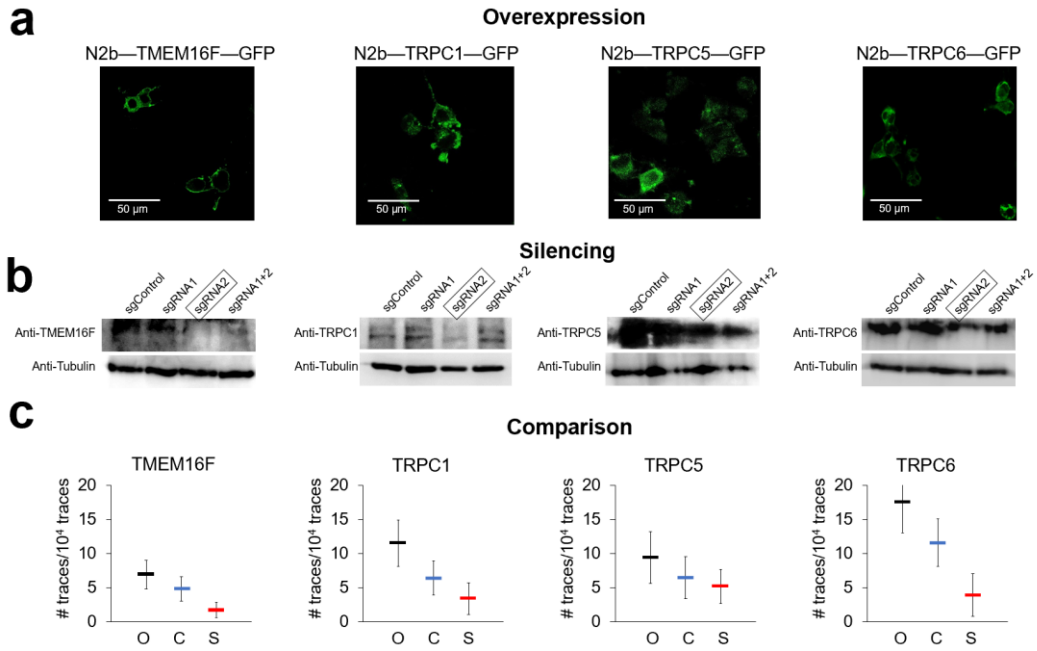


Figure S15
Overexpression and silencing in NG108-15 cells. **a:** fluorescent confocal images of the cells overexpressing the 6xHis-N2B—Protein—GFP construct, where Protein was TMEM16F, TRPC1, TRPC5 and TRPC6. **b**: Western Blots for the four proteins indicating the silencing obtained with different plasmids (the chosen plasmids for the SMFS experiments are the highlighted ones). **c**: For each cluster of F-d traces corresponding to TMEM16F, TRPC1, TRPC5 and TRPC6 we considered the frequency of identified traces over 10000 F-d traces which passed our criteria of good/valid traces associated to the unfolding of a membrane protein.
(C) Control F-d traces were those obtained from not transfected cells. (O) Overexpressed F-d traces were those obtained from cells transfected with the construct 6xHis-N2B-Protein-GFP bearing the N2B signature and those without the N2B signature but classified as corresponding as the unfolding of the Protein under investigation. (S) Silenced F-d traces were those obtained from NG108-15 cells in which the gene coding for the protein under investigation was silenced.
Error bars indicate the standard deviation.

lasmids	Vector	Provider	Company
peGFP-N1		Prof. Guidalberto Manfioletti	
mTMEM16A-GFP	peGFP-N1	Prof. Anna Menini	
mTMEM16F-GFP	peGFP-N1	Prof. Anna Menini	
6xHis-N2B-mTMEM16F-GFP	peGFP-N1		GENEWIZ
6xHis-N2B-mTRPC1-GFP	peGFP-N1		GENEWIZ
6xHis-N2B-mTRPC5-GFP	peGFP-N1		GENEWIZ
6xHis-N2B-mTRPC6-GFP	peGFP-N1		GENEWIZ
mCherry-Lifeact-7		Prof. Michael Davidson	Addgene

Table S1
List of plasmids.
The cDNA of TRPC1/5/6 were from Miao Ling Plasmid Company. All the 6xHis-N2B-Protein-GFP construct is made by Genewiz company.

No	Name	Sequence(20bp)
1	plentiCRISPR V2-TRPC1-	ccgtaagcccacctgtaaga
	sgRNA1(mouse)	
2	plentiCRISPR V2-TRPC1-	acgcttgtagcagaagggct
	sgRNA2(mouse)	
3	plentiCRISPR V2-TRPC5-	attactctacgccatccgca
	sgRNA1(mouse)	
4	plentiCRISPR V2-TRPC5-	ggagtgtgtatccagttcgg
	sgRNA2(mouse)	
5	plentiCRISPR V2-TRPC6-	gcggcagacgattcttcgtg
	sgRNA1(mouse)	
6	plentiCRISPR V2-TRPC6-	taaaggttatgtacggattg
	sgRNA2(mouse)	
7	plentiCRISPR V2-TMEM16F-	agcgagcgttacctcctgta
	sgRNA1(mouse)	
8	plentiCRISPR V2-TMEM16F-	ctctcgggtcaaataccaag
	sgRNA2(mouse)	
9	plentiCRISPR V2-control-sgRNA	tcttgagtttgtaacagctg

Table S2
sgRNA Sequences.
The sgRNA sequences cloned into LentiCRISPR-V2 puro vector were purchased from Miao Ling Plasmid Company.

Data S1. (separate file)
Tables with the protein lists and variables used for the Bayesian identification.