Supplementary methods

Protein reconstitution into liposomes

Protein expression and purification was performed as described previously [14, 23]. All lipids were from Avanti Polar Lipids with the exception of ^{3}H-DPPC (^{3}H-1,2-dipalmitoyl phosphatidylcholine), which was from Amersham Pharmacia Biotech. SUV lipid mix = 30 mol% POPC (1-palmitoyl-2-oleoyl-SN-glycero-3-phosphocholine), 15 mol% DOPS (1,2-dioleoyl-SN-glycero-3-phosphoserine), 22,6 mol% POPE (1-hexadecanoyl-2-octadecenoyl-SN-glycero-3-phosphoethanolamine), 5 mol% liver PI (L-α-phosphatidylinositol), 25 mol% cholesterol (from ovine wool), 1.6 mol% Rhodamine-DPPE (1,2-dipalmitoyl-SN-glycero-3-phosphoethanolamine-N-(lissamine rhodamine B sulfonyl)), 0.8 mol% NBD-DPPE (-1,2-dipalmitoyl-SN-glycero-3-phosphoethanolamine-N-(7-nitro-2,1,3-benzoxadiaziole-4-yl)) and trace amounts of ^{3}H-DPPC (1,2-dipalmitoyl-phosphatidylcholine), 3 µmol total lipid. GUV lipid mix = 35 mol% POPC (1-palmitoyl-2-oleoyl-SN-glycero-3-phosphocholine), 15 mol% DOPS (1,2-dioleoyl-SN-glycero-3-phosphoserine), 20 mol% POPE (1-hexadecanoyl-2-octadecenoyl-SN-glycero-3-phosphoethanolamine), 3 mol% liver PI (L-a-phosphatidylinositol), 2 mol% brain PI(4,5)P2 (L-α-phosphatidylinositol-4,5-bisphosphate), 25 mol% cholesterol (ovine wool) and trace amounts of ^{3}H-DPPC, 5 µmol total lipid.

SUVs were formed in the presence of t-SNARE protein (protein-to-lipid ratio 1/1000) or VAMP2 (1/200) and Syt1 (1/800) using the previously described technique of dilution and dialysis followed by a Nycodenz gradient centrifugation [34]. t-SNARE GUVs (protein-to-lipid ratio 1/1000) or v-SNARE GUVs containing VAMP2 (1/200) and Syt1 (1/800) were prepared as described previously [14]. Protein amounts in the reconstituted liposomes were determined using Coomassie blue–stained SDS-PAGE with BSA protein standards and ImageJ Quantitation Software (National Institutes of Health). Lipid recoveries were determined using ^{3}H-DPPC.

Fusion assays

Fusion reactions and data analysis were performed as described [14]. Briefly, unlabeled SUVs or GUVs (14 nmol lipid) and fluorescently labeled SUVs (2.5 nmol

lipid) containing the t-SNARE complex or the v-SNARE VAMP2 and synaptotagmin 1 (as indicated in the figure legends) were incubated in a final volume of 100 µl, containing 25 mM HEPES-KOH, pH 7.4, 135 mM KCl, 100 µM EGTA-KOH, pH 7.4, 0.5 mM $MgCl_2$, and 6 µM complexin in the presence or absence of 0.9 µM Munc18-1 and quickly transferred to a pre-warmed microtiter plate. To block liposome fusion, t-SNARE liposomes were pre-incubated with the cytoplasmic domain of VAMP2 (amino acid residues 1- 94, CD-VAMP) at a concentration of 20 µM for one hour at room temperature before addition to fusion reactions. Samples were measured at 37°C in a Synergy 4 plate reader (BioTek Instruments GmbH) at intervals of 10 seconds for 2 minutes, followed by the addition of 100 µM free Ca^{2+} (prepared as EGTA-buffered 6x stock solution). The NBD fluorescence obtained from control incubations containing v-SNARE liposomes pretreated with BoNT/D were subtracted from individual measurement sets. The fusion-dependent fluorescence is normalized to the maximal fluorescent signal obtained in the presence of 0.4% dodecylmaltoside (Fluka). Fusion kinetics show the average of three independent experiments. Error bars represent SEM.

CryoEM

The reaction partners were mixed as described above and incubated for one minute at 37°C except for the time course assay for fusion (described in Figure 3) where plunging was carried out at 4°C. Samples were then transferred onto freshly glow discharged lacey carbon grids (Plano GmbH, Germany) in a high-humidity chamber (EMBL Heidelberg). Blotting was carried out from the opposite side of the grid with a Whatman 1 filter paper that was extensively washed previously with buffer containing 50 mM EDTA. Immediately after blotting, the samples were plunge frozen in liquid ethane and stored in liquid nitrogen (time taken from sample transfer onto the grid to plunging was approximately one minute). Samples were imaged under low-dose conditions in a FEI Polara microscope (200kV) equipped with a Gatan GIF2002 energy filter.

Large image montages of several grid squares of each sample were collected at 3000x magnification using SerialEM, and spots where a large number of SUVs could be seen in close proximity to GUVs were selected for tomographic acquisition. Tomographic tilt series data was collected at 27,500x magnification (70µm objective

aperture and 50μm C2 aperture, energy filter slit width=30eV) giving an unbinned calibrated pixel size of 4.9Å. Total dose for each series was 60-70 e/Å^2. Data collection conditions were optimized for batch tomography, and 28-45 tomograms were collected for each sample in batch mode using the FEI Tomography 4 software. For each sample described in Figure 2G, three independent biological replicates (independent biochemical experiments, independent electron microscopy experiments) were analyzed to control for sample variability between experiments. Analysis of samples described in Figure 3D-E and in Figure S4C-D is from single experiments.

Image processing and blind membrane morphology assignment

For each biological replicate of each sample described in Figure 2G, low magnification montage images were collected (prior to tomographic acquisition) and renumbered (randomized). In these low-magnification montage images, we counted the number of SUVs in close proximity to a GUV membrane (within ~1-20 nm) and also counted those not in proximity to a GUV. The ratio of docked / undocked vesicles shown in Figure 2G is calculated from these low-magnification montage images.

CryoET tilt series data for all samples described in Figure 2G was similarly randomized and reconstructed in an automated manner using IMOD [31, 32] and RAPTOR [33]. Reconstructed tomograms were inspected and where the automated procedure had failed, were reconstructed again manually. Visualization of reconstructed tomograms was carried out in Amira (Visage Imaging). Details of data analysis of control experiments (Figures 3 and S4) are provided in Table S3.

Docking sites in large-scale cryoET data were identified by visual inspection of tomograms. SUVs that were closer than ~ 20 nm from the GUV membrane were considered as docked, and the adjacent patch of GUV membrane was considered as a docking site. If there was an abrupt increase in positive membrane curvature of the GUV (the target membrane) at the docking site, then that docking site was classified as “protrusion”. If there was no observable change in the curvature of the GUV membrane at the docking site, then the site was classified as a simple “contact” site. When the apposing membranes were pressed together, so that the two membranes were parallel, then the site was classified as an “extended” contact site.

Statistical analysis

Once all docking sites in all tomograms were quantified as described, then the relative proportion of all morphologies was calculated for each sample described in Figure 2G, both for the three independent biological replicates, and the pooled data. These data are shown in Table 1.

Due to the large sample size ($n > 10$) for most entries in table 1, the proportion for each replicate could be estimated very reliably. We focused therefore on computing confidence intervals that assessed the variability between the three independent biological experiments. In order to determine the standard error of a proportion estimate, we employed a James-Stein-shrinkage estimator of the corresponding variances. The estimator shrinks the estimated variance of a proportion towards the median of all variance estimates. This method has been applied successfully in other small sample settings like the identification of differently expressed genes [35].

Specifically, we first probit-transformed the proportions to put them on a z-score scale (using the R-function qnorm), and then applied the shrinkage estimation of the variances using the function var.shrink of the R-package corpcor. 95% confidence intervals were computed in the usual way (mean $\pm 1.96 \times$standard error). Finally, the CI limits were transformed back to the proportion scale (using the R-function pnorm). The estimation of the standard errors was performed separately for the docked-undocked (measured from low-magnification montage images) and the docked proportions (protrusion, contact and extended measured from high-magnification tomography data) in Table 1.

Having obtained standard errors also allowed us to test for differences between proportions using a two-sided z-test of mean/standard error (mean), where standard error(mean) was calculated just as in the Welch-t-test. The p-values from these tests were adjusted using the Benjamini-Hochberg method as implemented in the R-function p.adjust. The rejection of null hypotheses was performed at a 5% false discovery rate level. Results from these comparisons are presented in Table S1-2.

Supplementary references:

34. Weber T, Zemelman BV, McNew JA, Westermann B, Gmachl M, Parlati F, Söllner TH, Rothman JE (1998) SNAREpins: minimal machinery for membrane fusion. *Cell* **92:** 759-772

35. Opgen-Rhein R, Strimmer K (2007) Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Stat Appl Genet Mol Biol* **6:** Article9