

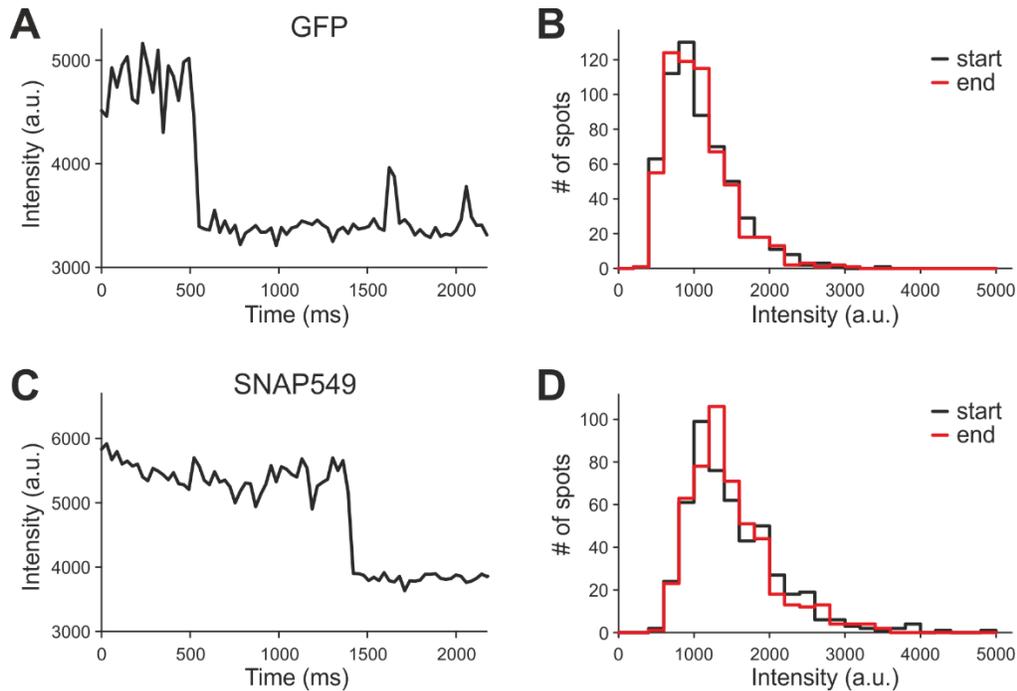
Supplementary information

Designed membrane protein heterodimers and control of their affinity by binding domain and membrane linker properties

Chenyang Lan, Anja Stulz, Nicolas P. F. Barthes, Susan Lauw, Pavel Salavei, Manfred Jung, Heiko Heerklotz, Maximilian H. Ulbrich

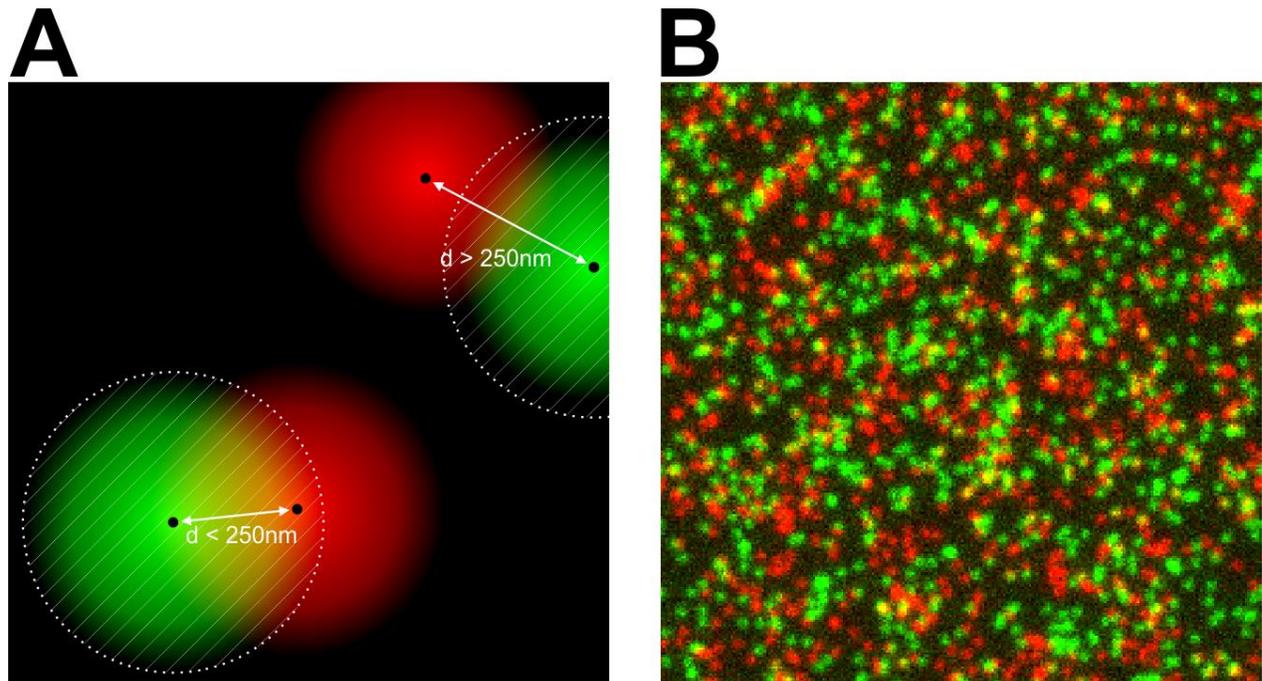
Supplementary Figure 1	2
Intensities along trajectories and intensity histograms	
Supplementary Figure 2	3
Calculation and Monte Carlo simulation for co-incidental overlap	
Supplementary Note 1	4
Amino acid sequences of SYNZIP constructs	
Supplementary Note 2	6
BG-DY549P1 Synthesis	
Supplementary Note 3	7
Model to calculate dimer fraction from yellow fraction	
Supplementary Note 4	8
Model to calculate fluorescent GFP and labeled SNAP-tag fractions	
Supplementary Note 5	9
Isothermal titration calorimetry	
Supplementary Note 6	13
Monte Carlo simulation and calculation of effective distance for a flexible linker	
Supplementary Note 7	15
Analysis using automated spot recognition	
Supplementary Movie 1	
Cell expressing heterodimer-mimicking S549-TM-GFP	
Supplementary Movie 2	
Cell expressing monomer mix HALO-TM-GFP + S549-TM-xGFP	
Supplementary Movie 3	
Cell expressing SYNZIP constructs HALO-SZ1-TM-GFP + S549-SZ2-TM-xGFP	
Scale bars in all movies: 5 μm , insets 1 μm	

Supplementary Figure 1



Supplementary Figure 1: Intensities along trajectories and intensity histograms. (A) A representative intensity course along a trajectory of a green spot in the heterodimer mimic experiment with the S549-TM-GFP construct. Only one bleaching step is visible before the intensity drops to the background level. (B) The histograms of intensities taken from the start (black) and the end (red) of all trajectories (first and last 4 fully illuminated frames, trajectory lengths > 10 frames, background subtracted) are virtually the same ($p < 0.6$, Mann-Whitney U Test). For a dimer or higher order multimer, a shift from higher to lower intensities would be expected. (C) Representative intensity along a red trajectory. (D) Intensity histograms for the red spots. Again, the intensity histograms taken from the start and the end of all trajectories are similar ($p < 0.5$).

Supplementary Figure 2



Supplementary Figure 2: Calculation and Monte Carlo simulation for co-incident overlap. (A) Green and red spots were considered co-localized when their distance was 250 nm or less. With a total spot density of $2.5 \mu\text{m}^{-2}$ ($1.25 \mu\text{m}^{-2}$ for each color), the fraction of the total area that is covered by green spots (shaded area) is $x = \pi \cdot (0.25 \mu\text{m})^2 \cdot 1.25 \mu\text{m}^{-2} = 24.5\%$ (B) A Monte Carlo simulation demonstrates that randomly localized green and red spots yield the calculated fraction of yellow spots.

Supplementary Note 1

Amino acid sequences of SYNZIP constructs

All constructs were assembled in a modular design from the following sequences.

Signal Peptide:

MGVKVLFALICIAVAEAEN

Halo-Tag:

MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTVPVLFHLHGNTSSYVWRNIIPHVAPTHRCIAPDLIGMGK
SDKPDLDGYFFDDHVRFMDAFIEALGLEEVVLIHDWGSALGFHWAKRNPERSVKGI AFMEFIRPIPTWDEWP
EFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAG
EPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIIPAEAAARLAKSLPNCKAVDIGPGLNLLQEDNPDIGS
EIRLWSTLEISG

SNAP-Tag (SNAPf version):

MDKDCEMKRTTLDSPGKLELSGCEQGLHRIIFLGKGTSAADAVEVPAPAAVLGGPEPLMQATAWLNAYFH
QPEAIEEFVVPALHHPVFQQESFTRQVLWKLKLVVVFGEVVISYSHLAALAGNPAATAAVKTALSNGNPVIL
IPCHRVVQGDLDVGGYEGGLAVKEWLLAHEGHRLGKPGLG

SYNZIP1:

FL NLVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEE
Δ1 EVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEE
Δ2 NETLKKKNLHKKDLIAYLEKEIANLRKKIEE
Δ3 NLHKKDLIAYLEKEIANLRKKIEE
Δ4 IAYLEKEIANLRKKIEE
Δ5 IANLRKKIEE

SYNZIP2:

FL ARNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVASHEQ
Δ1 KIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVASHEQ
Δ2 NLQLERDEQNLEKIIANLRDEIARLENEVASHEQ
Δ3 EQNLEKIIANLRDEIARLENEVASHEQ
Δ4 IANLRDEIARLENEVASHEQ
Δ5 IARLENEVASHEQ

Transmembrane domain:

TVAAAVLVLLVIVIIISLIVLVVIWKQK

Linkers between SYNZIP1 and TMD:

8aa: GGSMHSG

26aa: GGKLGSGGSRGSGGSGGGGSGSMHSG

66aa: GGKLSMHSGGSGSTSGTSGSSGSTGSASGGSGGSRGSGGSGGGGSGSGGSSGGSTSSGSTSGL
HSG

123aa: GSKLSMHSGGSGSTSGTSGSSGSTGSASGGSGGSRGSGGSGGGGSGSGGSSGGSTSSGSTSGL
HSGGSGSTSGTSGSSGSTGSASGGSGGSRGSGGSGGGGSGSGGSSGGSTSSGSTSGLHSG

Linkers between SYNZIP2 and TMD:

8aa: GSGSGTSG

26aa: GSGMHGSGSTSGTSGSSGSTGSGTSG

66aa: GSKLSSSGGGSGSTSGTSGSSGSTGSASGGSGGSRGSGGSGGGGSGSGGSSGGSTSSGSTSGP
GTSG

123aa: GSKLSMHSGGSGSTSGTSGSSGSTGSASGGSGGSRGSGGSGGGGSGSGGSSGGSTSSGSTSGL
HSGGSGSTSGTSGSSGSTGSASGGSGGSRGSGGSGGGGSGSGGSSGGSTSSGSTSAGTSG

Supplementary Note 2

BG-DY549P1 Synthesis

HPLC analysis (1 mL/min) and purification (3 mL/min) were performed on an Agilent Technologies 1260 Infinity system using UV detection at 290 nm and – for analysis – a Phenomenex Kinetex® 5u XB-C18 100 Å 250 x 4.6 mm column or – for purification – a Phenomenex Synergi® 10u Hydro-RP 80 Å 250 x 15.0 mm column. Eluent A was water containing 0.05 % trifluoroacetic acid (TFA) and eluent B was acetonitrile containing 0.05 % TFA. Linear gradient conditions were as follows: 0–1 min, A/B (90:10); 1–21 min, linear increase to 100 % of B; 21–23 min, 100 % B; 23–23.3 min: A/B (90:10); 23.3–26 min: A/B (90:10). Characterization was performed through mass spectrometry and mass spectra were recorded on a Thermo Scientific Exactive mass spectrometer using electrospray ionization (ESI) as ion sources.

To a stirred solution of 6-((4-(aminomethyl)benzyl)oxy)-7H-purin-2-amine (1.4 mg, 5.2 µmol, 1.1 eq.) in CH₃CN (250 µL), under inert atmosphere, was added the NHS ester of DY549-P1 (5 mg, 4.8 µmol, 1 eq.) in 250 µL of H₂O, followed by NaHCO₃ (50 µL of a 1 M solution: final concentration of ca. 0.1 M) and the reaction was stirred protected from light for 48 h. After completion (RP-C18 TLC: H₂O/CH₃CN 3:7) the solvent was removed to give a pink oil. The crude BG-DY549P1 was then dissolved in the minimum volume of methanol, and purified using preparative HPLC (see conditions above). The product was isolated as a pink oil (3.3 mg, 55 % yield). C₄₉H₅₇N₈Na₃O₁₅S₄ (1195.25 g/mol). HPLC: t_R = 6.935 min (82 % purity – 18 % of DY549-P1-OH remaining). ESI-HRMS(–): m/z calcd for C₄₉H₅₆N₈Na₃O₁₅S₄: 1193.2447 [M–H][–]; found 1193.2434 [M–H][–]; m/z calcd for C₄₉H₅₇N₈Na₂O₁₅S₄: 1171.2627 [M–Na][–]; found 1171.2611 [M–Na][–]; m/z calcd for C₄₄H₅₄N₃Na₂O₁₅S₄: 1038.2239 [M–2AP–Na][–]; found 1038.2229 [M–2AP–Na][–].

Supplementary Note 3

Model to calculate dimer fraction from yellow fraction

The model takes into account the fraction of fluorescent GFP and SNAP tags and the coincidental overlap of green and red fluorescent membrane proteins. By assuming values for the fraction of fluorescent GFP p_g and SNAP-tags p_r , and ratio of spot area A_s (spot radius r_{spot}) to area A_0 that all spots cover, we established a formula to estimate the values of molecule numbers in monomers $N_{m:S1(GFP)}$ and $N_{m:S2(SNAP)}$ and in heterodimers $N_{d:S1(GFP):S2(SNAP)}$.

The number of GFP- and SNAP-labeled monomers $N_{S1 \rightarrow g}$ and $N_{S2 \rightarrow r}$ are:

$$N_{S1 \rightarrow g} = N_{m:S1(GFP)} \cdot p_g$$

$$N_{S2 \rightarrow r} = N_{m:S2(SNAP)} \cdot p_r$$

Due to non-fluorescent GFP or SNAP-tags, some of dimers with GFP and SNAP-tags can appear as yellow, green or red:

$$N_{d:S1(GFP):S2(SNAP) \rightarrow y} = N_{d:S1(GFP):S2(SNAP)} \cdot p_g \cdot p_r$$

$$N_{d:S1(GFP):S2(SNAP) \rightarrow g} = N_{d:S1(GFP):S2(SNAP)} \cdot p_g \cdot (1 - p_r)$$

$$N_{d:S1(GFP):S2(SNAP) \rightarrow r} = N_{d:S1(GFP):S2(SNAP)} \cdot p_r \cdot (1 - p_g)$$

The numbers of yellow, green, red fluorescent proteins are the sums:

$$N_{y-protein} = N_{d:S1(SNAP):S2(SNAP) \rightarrow y}$$

$$N_{g-protein} = N_{m:S1(GFP) \rightarrow g} + N_{d:S1(GFP):S2(SNAP) \rightarrow g}$$

$$N_{r-protein} = N_{m:S2(SNAP) \rightarrow r} + N_{d:S1(GFP):S2(SNAP) \rightarrow r}$$

Finally, we make an approximation to correct for the random co-localization due to close vicinity of red and green proteins. Then their emission overlaps spatially, resulting in appearance of yellow spots. A green protein will appear as yellow spot when it happens to lie within the 250 nm radius of a red protein. Therefore, the number of green and red spots is smaller than the number of proteins labeled in these colors, and the number of yellow spots is bigger. The chance for a single green spot to not overlap on a particular red receptor is $1 - A_s/A_0$, where $A_s = \pi \cdot r_{spot}^2$ is the area of the red spot (assume the areas of a green and a red spot are identical) and A_0 is the total area the counted spots cover. If one or more green spots overlap with the red spot, the chance for the red spot to be overlapped is then:

$$p_{rand} = 1 - (1 - A_s/A_0)^{N_g}$$

Therefore, the number of yellow spots appearing from green and red overlap is:

$$N_{y-rand} = N_r \cdot (1 - (1 - A_s/A_0)^{N_g})$$

And the final counts of green, red and yellow spots are:

$$N_{g-spot} = N_{g-protein} - N_{y-rand}$$

$$N_{r-spot} = N_{r-protein} - N_{y-rand}$$

$$N_{y-spot} = N_{y-protein} + N_{y-rand}$$

Estimates for the number of designed proteins in monomers $N_{m:S1}$ ($N_{m:S2}$) and in heterodimers $N_{d:S1+S2}$ can be obtained by using them as fit parameters in a least squares fit of N_{g-spot} , N_{r-spot} , and N_{y-spot} to the observed spot counts.

Supplementary Note 4

Model to calculate fluorescent GFP and labeled SNAP-tag fractions

The fractions of fluorescent GFP p_g and fluorescent SNAP-tags p_r can be obtained from the dually labeled S549-PDGFRTM-GFP by counting the numbers N_{g-spot} , N_{r-spot} , and N_{y-spot} of green, red and yellow labeled spots. The coincidental overlap of red and green receptors to form a yellow spot needs to be considered. The formulas from Suppl. Note 3 can be simplified by assuming that $N_{m:S1(GFP)} = N_{m:S2(SNAP)} = 0$ and $N_{d:S1(GFP):S2(SNAP)} = N_0$ because there are only proteins carrying both a GFP and a SNAP-tag.

With the molecule number N_0 , p_g and, p_r we directly get:

$$N_y = N_0 \cdot p_g \cdot p_r$$

$$N_g = N_0 \cdot p_g \cdot (1 - p_r)$$

$$N_r = N_0 \cdot (1 - p_g) \cdot p_r$$

The rest of the equations remains as in Suppl. Note 1:

$$p_{rand} = 1 - (1 - A_s/A_0)^{N_g}$$

$$N_{y-rand} = N_g \cdot (1 - (1 - A_s/A_0)^{N_r})$$

$$N_{g-spot} = N_g - N_{y-rand}$$

$$N_{r-spot} = N_r - N_{y-rand}$$

$$N_{y-spot} = N_y + N_{y-rand}$$

Estimates for the total number of protein N_0 , the fraction of fluorescent GFP p_g , and the fraction of labeled SNAP-tags p_r can be obtained by using them as fit parameters in a least squares fit of N_{g-spot} , N_{r-spot} , and N_{y-spot} to the observed spot counts.

Supplementary Note 5

Isothermal titration calorimetry

The heterodimerization affinities of soluble forms of SYNZIP2 (species A) and SYNZIP1 (species B) were determined via isothermal titration calorimetry (ITC). To increase the solubility of the SYNZIP constructs, the maltose binding protein MBP and a SUMO domain were fused, to obtain SZ1-SUMO-His and MBP-SZ2-His.

Cloning, protein expression and purification

The SZ1-SUMO-His and SZ2-His sequences were synthesized as GeneArt Strings DNA Fragment (ThermoFisher) with codon optimization for *E.coli* K12, and MBP was amplified by PCR from the pMAL-c6X vector. The fragments were fused by PCR and cloned into the pET303/CT-His vector (ThermoFisher).

The proteins were expressed in LB medium by induction at $OD_{600} = 0.7$ with 1 mM IPTG at 30°C for 5 h. After cell lysis by sonication, proteins were purified on a HisTrap FF Crude 5mL column (GE Healthcare). The MBP-SZ2-His construct was digested overnight with TEV protease to remove the His-tag. After digestion, MBP-SZ2 was passed twice through a HisTrap FF Crude 5mL column (GE Healthcare) to remove undigested protein, TEV protease and cut His-tag. Finally, both proteins were purified on a HiLoad 26/600 Superdex 200 column (GE Healthcare) equilibrated with PBS, 10% glycerol.

ITC measurements

ITC experiments were performed with a MicroCal VP-ITC instrument (Malvern Instruments, UK). The sample cell had a volume of 1.4 mL. ITC experiments were conducted by repeating injections of 7 μ L aliquots of 6.5 μ M MBP-SZ2 into the sample cell that contained 0.6 μ M SZ1-SUMO-His. Samples were degassed prior to loading into the calorimeter and all runs were performed at 20 °C with a stirring speed of 307 rpm. A first 1 μ L injection has been discarded from the analysis because it is prone to artifacts.

We found that the SYNZIPs have a tendency to form homodimers at higher concentrations; therefore we developed equations that contain terms for homo- and heteromerization of each species.

During injection of A from the syringe into the cell (initially containing only B), liquid containing both A and B is pushed out from the cell. The ITC-Manual [MicroCal 2004] calculates for the concentrations $[A_{tot}]_i$ and $[B_{tot}]_i$ of species A and B in the cell after the i -th injection, i.e. after injection of a total volume ΔV_i :

$$[B_{tot}]_i = [B_{tot}]_0 \cdot \frac{2V_0 - \Delta V_i}{2V_0 + \Delta V_i}$$
$$[A_{tot}]_i = [A_{tot}]_{syr} \cdot \frac{\Delta V_i}{V_0} \left(1 - \frac{\Delta V_i}{2V_0}\right)$$

where $[B_{tot}]_0$ is the initial concentration of species B in the cell, $[A_{tot}]_{syr}$ is the initial concentration of species A in the syringe, and V_0 is the volume of the cell. To adjust for small deviations between the measured and the real concentrations of either or both species (or e.g. an inactive fraction that does not participate in the reaction), a correction parameter n was introduced by $[B_{tot}]_0 = n \cdot [B_{tot}]_0^{app}$, where $[B_{tot}]_0^{app}$ is the apparent, measured concentration.

With each pair of concentrations $[A_{tot}]_i$ and $[B_{tot}]_i$, we set up an equation system for the monomer, homodimer and heterodimer fractions of A and B that also contains the dissociation constants K_{AA} , K_{BB} , and K_{AB} :

$$[A][A] = K_{AA}[AA]$$

$$[B][B] = K_{BB}[BB]$$

$$[A][B] = K_{AB}[AB]$$

$$[A_{tot}] = [A] + [AB] + 2[AA]$$

$$[B_{tot}] = [B] + [AB] + 2[BB]$$

Next, we calculate the heat contributions of the homodimers AA and BB and the heteromer AB. We also have to take into account the enthalpy that is contained in the syringe and the overflow volume. The heat contributions are calculated as products of the concentration changes, volumes, and molar enthalpies ΔH_X . To account for small deviations of the temperature in the syringe and cell, a species- and concentration-independent term Q_{dil} is introduced.

$$\Delta\Delta V = \Delta V_i - \Delta V_{i-1}$$

$$\Delta[X] = [X]_i - [X]_{i-1}$$

$$[X]_{ov} = \frac{1}{2}([X]_i + [X]_{i-1})$$

$$Q_X = \Delta[X] \cdot V_0 \cdot 2 \Delta H_X - ([X]_{syr} - [X]_{ov}) \cdot \Delta\Delta V \cdot 2 \Delta H_X$$

for $X = AA, BB, AB$

$$Q_{dil} = \Delta\Delta V \cdot q_{dil}$$

$$Q_{obs} = Q_{AA} + Q_{BB} + Q_{AB} + Q_{dil}$$

Since only species A is contained in the syringe, $[BB]_{syr} = [AB]_{syr} = 0$. $[AA]_{syr}$ can be determined from the above equations through a fit to a reference experiment without species B.

The heat per mole of injected species A in the cell is

$$Q_{obs}^m = Q_{obs} / (V_0 \cdot \Delta[A_{tot}]_i)$$

$$\text{with } \Delta[A_{tot}]_i = [A_{tot}]_i - [A_{tot}]_{i-1}$$

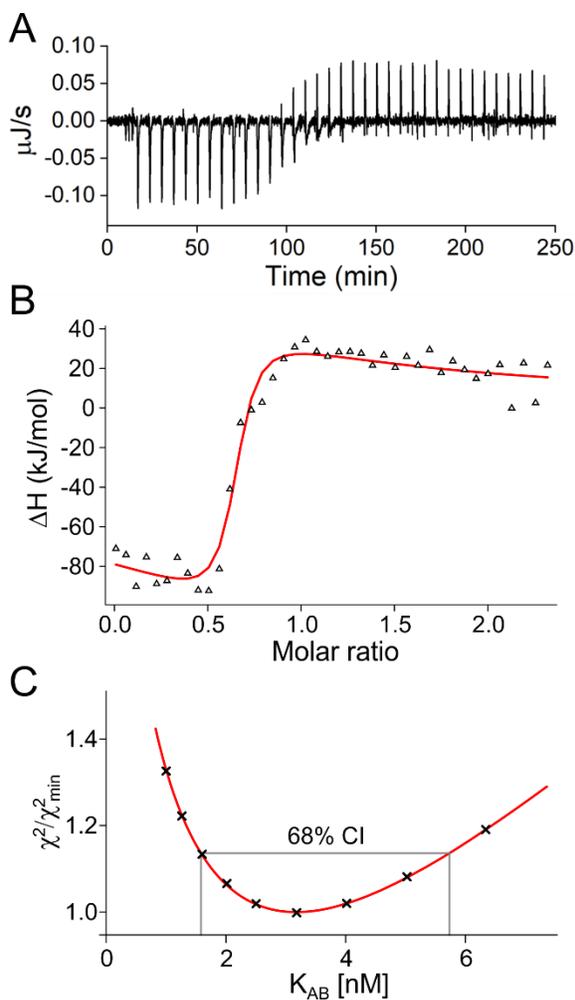
The measured values of Q_{obs}^m should be fitted. The free fit parameters are K_{AA} , K_{BB} , K_{AB} , ΔH_{AA} , ΔH_{BB} , ΔH_{AB} , n , and q_{dil} . In a reference experiment where species A is injected into the cell without B, the values for K_{AA} and ΔH_{AA} can be determined. We assumed $K_{BB} = K_{AA}$ and $\Delta H_{BB} = \Delta H_{AA}$. Then, only K_{AB} , ΔH_{AB} , n , and q_{dil} require fitting in the main experiment.

Suppl. Figure 3A shows a raw ITC thermogram obtained by titrating 6.5 μM MBP-SZ2 into 0.6 μM SZ1-SUMO-His. In Suppl. Figure 3B differential heats are shown as a function of MBP-SZ2/SZ1-SUMO-His molar ratio $[A_{tot}]_i/[B_{tot}]_i^{app}$, together with the fit curve of Q_{obs}^m . The 68% confidence interval was determined by a support plane analysis varying the K_{AB} value as shown in Suppl. Figure 3C [Kemmer 2010]. If several measurements were made, a common fit for all was performed with a single K_{AB} or K_{AA} value, and individual ΔH , n , and Q_{dil} values for each measurement. All results of individual and common fits are summarized in Suppl. Table 1.

We made the assumption that $K_{BB} = K_{AA}$. A change in K_{BB} had a strong impact on the value of K_{AB} (Suppl. Fig. 4A). The impact of variations in ΔH_{BB} is smaller, but visible (Suppl. Fig. 4B). In principle, K_{BB} and ΔH_{BB} can be determined from a reference measurement with species B only. However, the resulting difference in K_{AB} has only a small impact on the conclusions of our work.

Supplementary Figure 3: ITC measurements.

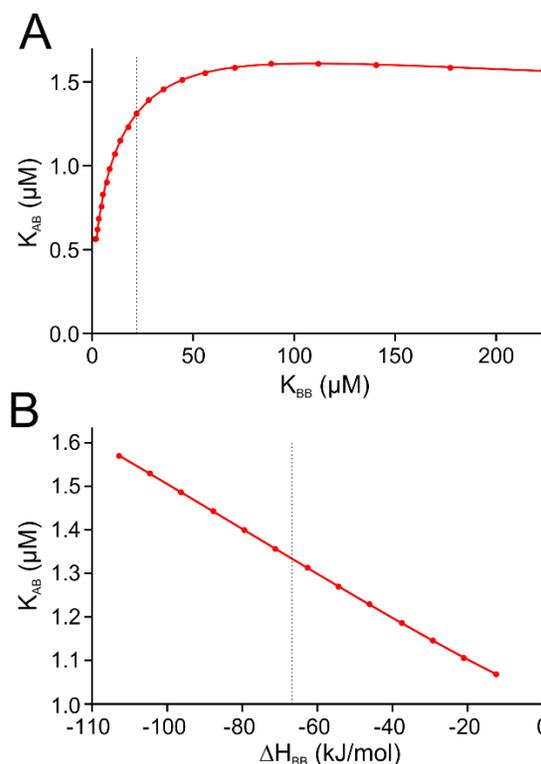
(A) Titration of 6.5 μM MBP-SZ2 into 0.6 μM SZ1-SUMO-His at 20 $^{\circ}\text{C}$. Panel A shows a raw thermogram with arbitrary baseline subtracted. (B) Integrated heats per injection as a function of MBP-SZ2/SZ1-SUMO-His molar ratio (triangles). The red line represents a fit yielding $K_{AB} = 3.2$ nM, $\Delta H_{AB} = -68.4$ kJ/mol, and $n = 0.69$. (C) Confidence assessment for the fit of K_{AB} . K_{AB} was set to a value different from its best fit, and the other free parameters ΔH_{AB} , n , and q_{dil} were re-fitted. χ^2 normalized to its minimal value (when K_{AB} fits best) is plotted against K_{AB} . Only a few values were calculated (crosses) and used for interpolation (red line). Grey line represents the 68% confidence interval, yielding a standard error of 2.1 nM for K_{AB} (full-length sample #1).



Experiment	Fit results				
	$K_{AA/AB}$ (μM)	$\Delta H_{AA/AB}$ (kJ/mol)	Q_{dil} (kJ/mol)	n	68% CI for K (μM)
MBP-SZ2 (6.5 μM) in buffer – homomer AA					
#1	3.5	-71.8	3.0		[1.2,11]
#2	7.5	-109.6	-7.1		[2.2,12]
common fit	5.5	-89.7			[1.9,12]
MBP-SZ2 (6.5 μM) in SZ1-SUMO-His (0.6 μM) – heteromer AB					
#1	0.0032	-68.4	-9.9	0.69	[0.0016,0.0057]
#2	0.0028	-69.8	-3.6	0.63	[0.0018,0.0043]
#3	0.0043	-72.2	-6.7	0.84	[0.0032,0.0059]
common fit	0.0035	-70.2			[0.0023,0.0051]
MBP-SZ2- Δ 1 (30 μM) in buffer – homomer AA					
#1	6.5	-49.3	0.7		[3.5,12.8]
MBP-SZ2- Δ 1 (30 μM) in SZ1- Δ 1-SUMO-His (3 μM) – heteromer AB					
#1	0.39	-78.7	-0.5	0.88	[0.27,0.58]
#2	0.34	-78.0	5.1	1.0	[0.20,0.59]
#3	0.29	-76.2	-0.7	0.92	[0.26,0.33]
common fit	0.34	-77.4			[0.24,0.44]
MBP-SZ2- Δ 2 (25, 50, 100 μM) in buffer – homomer AA					
#1 (25 μM)	24	-77.7	-7.1		[13,41]
#2 (50 μM)	21	-61.2	-3.6		[16,27]
#3 (100 μM)	23	-63.2	-1.7		[19,26]
common fit	22	-66.8			[17,29]
MBP-SZ2- Δ 2 (100 μM) in SZ1- Δ 2-SUMO-His (14 μM) – heteromer AB					
#1	1.31	-57.8	-2.8	0.79	[1.12,1.56]

Supplementary Table 1. Fit values for ITC measurements.

Supplementary Figure 4: Impact of changes in K_{BB} and ΔH_{BB} on K_{AB} . (A) K_{BB} was varied from 2.2 μM to 220 μM and the best fit parameter for K_{AB} was calculated. For small K_{BB} values, K_{AB} decreases significantly. The dotted line represents the value for $K_{BB} = K_{AA}$. As an example, the Δ 2 pair was used, where the impact of changing K_{BB} on the fit of K_{AB} is biggest. (B) ΔH_{BB} was varied from -105 kJ/mol to -15 kJ/mol. The impact on the value of K_{AB} is small. The dotted line represents the value for $\Delta H_{BB} = \Delta H_{AA}$.



[MicroCal 2004] MicroCal LLC, ITC Data Analysis in Origin, Version 7.0 (2004).

[Kemmer 2010] Kemmer G. and Keller S. Nonlinear least-squares data fitting in Excel spreadsheets. *Nat. Protoc.* **5**: 267–281 (2010).

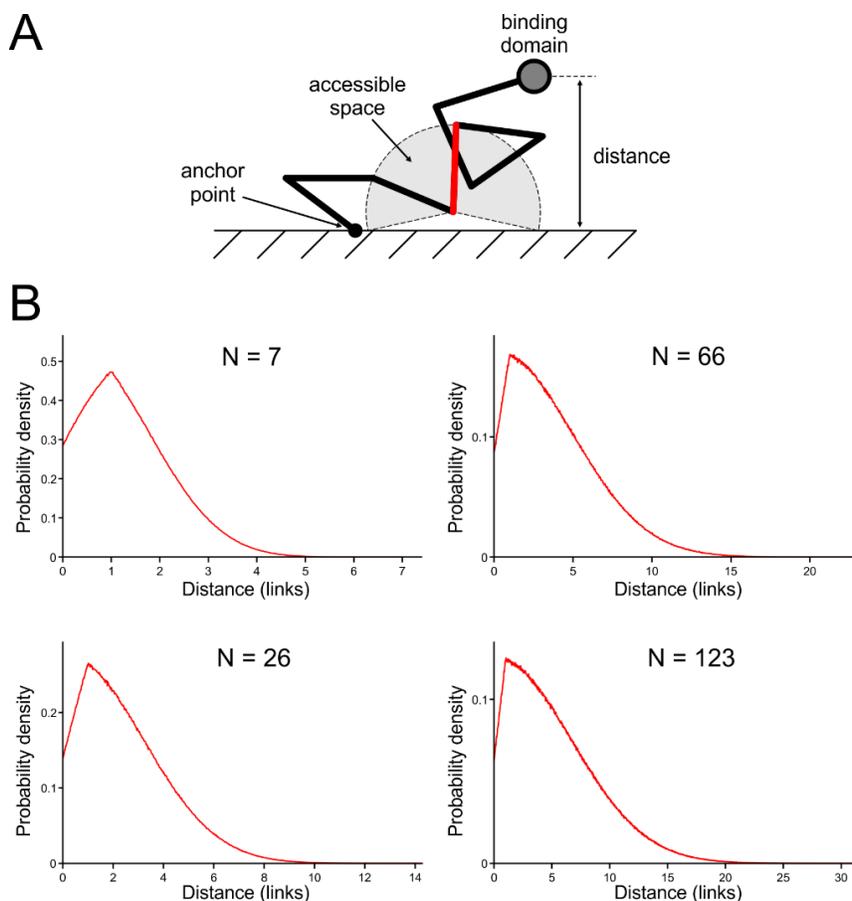
Supplementary Note 6

Monte Carlo simulation and calculation of effective distance for a flexible linker

For the determination of the effective distance of the binding domain from the membrane, we assumed a flexible linker with randomly oriented chain links (Suppl. Fig. 5A). For the number of chain links, we used the number of amino acids.

The probability density distribution of the binding domain's distance from the membrane was simulated for different chain lengths in Monte Carlo simulations of 10^7 random chain configurations, resulting in a characteristic profile (Suppl. Fig. 5B).

From these density distributions, we want to estimate an effective linker length. In our case, since we are interested in the effect of the linker length on the dimerization, we go back to the basic definition of the dissociation constants K_d^{3D} and K_d^{2D} in 3D and 2D.



Supplementary Figure 5: MC Simulation of the flexible chain. (A) The anchor point is tethered to the membrane. Each chain link points to a random direction, but the space occupied by the membrane is not accessible. The binding domain is assumed to be located at the position of the last endpoint of the chain. (B) The probability density of the chain's end point's distance from the membrane in dependence of the linker length N . The distance distribution scales approximately with the square root of N and, except the drop close to the membrane, resembles a half-normal distribution with a half-width of $\sqrt{N/3}$. The x axis unit is given in multiples of a single link's length.

$$c_{m,3D}^2 = K_d^{3D} \cdot c_{d,3D} \quad (1a)$$

$$c_{m,2D}^2 = K_d^{2D} \cdot c_{d,2D} \quad (1b)$$

where $c_{m,3D}$, $c_{d,3D}$, $c_{m,2D}$, and $c_{d,2D}$ are the concentrations of monomers and dimers in 3D and 2D, respectively. The 2D concentrations (i.e. membrane densities) are related to the 3D concentrations through the integrals

$$c_{m,2D} = \int_0^\infty c_{m,3D}(x) dx \quad (2a)$$

$$c_{d,2D} = \int_0^\infty c_{d,3D}(x) dx \quad (2b)$$

In the special case where the concentration is constant from the surface of the membrane up to a distance l and zero beyond that point, the concentration profiles simplify to

$$c_{m,3D}(x) = \begin{cases} c_{m,3D,0} & \text{for } 0 < x \leq l \\ 0 & \text{for } x \leq 0 \text{ or } x > l \end{cases}$$

$$c_{d,3D}(x) = \begin{cases} c_{d,3D,0} & \text{for } 0 < x \leq l \\ 0 & \text{for } x \leq 0 \text{ or } x > l \end{cases}$$

It follows

$$c_{m,2D} = \int_0^\infty c_{m,3D}(x) dx = c_{m,3D,0} \cdot l$$

$$c_{d,2D} = \int_0^\infty c_{d,3D}(x) dx = c_{d,3D,0} \cdot l$$

and finally

$$K_d^{2D} = \frac{c_{m,2D}^2}{c_{d,2D}} = \frac{(c_{m,3D,0} \cdot l)^2}{c_{d,3D,0} \cdot l} = \frac{c_{m,3D,0}^2}{c_{d,3D,0}} \cdot l = K_d^{3D} \cdot l$$

which is equation (1) from the main text. This relation suggests a definition of the effective linker length l_{eff} for the case of a non-homogenous concentration in the vicinity of the membrane as

$$l_{eff} = \frac{K_d^{2D}}{K_d^{3D}}$$

Using the relation (1b) for K_d^{2D} , the integrals (2a) and (2b) and finally (1a) yields

$$l_{eff} = \frac{K_d^{2D}}{K_d^{3D}} = \frac{1}{K_d^{3D}} \frac{c_{m,2D}^2}{c_{d,2D}} = \frac{1}{K_d^{3D}} \frac{\left(\int_0^\infty c_{m,3D}(x) dx\right)^2}{\int_0^\infty c_{d,3D}(x) dx} = \frac{\left(\int_0^\infty c_{m,3D}(x) dx\right)^2}{\int_0^\infty (c_{m,3D}(x))^2 dx}$$

This formula satisfies the expectations for the effective linker length properties: it depends only on the concentration profile, i.e. it is independent of K_d^{3D} , K_d^{2D} , and the total concentration, since a constant factor in $c_{m,3D}(x)$ will cancel out.

For the concentration profiles (Suppl. Fig. 5B) with the chain link numbers of $N = 7, 26, 66,$ and 123 and a length of each link of 0.4 nm (as used in our work), we obtain $l_7 = 1.2$ nm, $l_{26} = 2.2$ nm, $l_{66} = 3.5$ nm, and $l_{123} = 4.7$ nm.

Supplementary Note 7

Analysis using automated spot recognition

For the analysis of green, red and yellow spot numbers in the main text, the spots were selected manually. As an alternative, we also performed automated analysis using the algorithm devised in ref. [Crocker 1996]. However, the algorithm did not perform well at high spot densities that were necessary in our experiments to cover a wide range of membrane densities. In areas where two or more spots were visible by eye, the algorithm only recognized one spot because there was not sufficient spatial separation.

As a consequence of using the automated recognition also for the selection of the spots in the heteromer mimic S549-TM-GFP, the apparent values for fluorescence probabilities p_G and p_R of green and red spots were also smaller ($p_G = 0.42$ and $p_R = 0.48$) than those obtained from the analysis with manual spot selection. The decreased p_G and p_R from the automated analysis compensated for the (compared to manual analysis) smaller recognition efficiency in the analysis of the SYNZIP pairs. Therefore, the results from automated spot selection resembled closely the results obtained from manual spot selection (Suppl. Fig. 6).

Supplementary Figure 6: Analysis after automated spot recognition. (A) Dissociation constants from automated analysis with the linker length according to the extended linker model and (B) according to the coiled linker model.

