SUPPORTING INFORMATION

Selective crosslinking of interacting proteins using self-labeling tags

Arnaud Gautier[1], Eiji Nakata[2], Gražvydas Lukinavičius, Kui-Thong Tan and Kai Johnsson*

Institute of Chemical Sciences and Engineering; École Polytechnique Fédérale de Lausanne (EPFL); CH-1015 Lausanne, Switzerland.

[1] Present address: Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, United Kingdom,

[2] Present address: Department of Life System, Institute of Technology and Science, Graduate School, The University of Tokushima, Minamijosanjimacho-2, Tokushima 770-8506, Japan.

Corresponding author: kai.johnsson@epfl.ch

This file contains:

Complete Refs 1 and 3

(1) Rual, J. F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitgord, N.; Simon, C.; Boxem, M.; Milstein, S.; Rosenberg, J.; Goldberg, D. S.; Zhang, L. V.; Wong, S. L.; Franklin, G.; Li, S. M.; Albala, J. S.; Lim, J. H.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex, C.; Lamesch, P.; Sikorski, R. S.; Vandenhaute, J.; Zoghbi, H. Y.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, M. E.; Hill, D. E.; Roth, F. P.; Vidal, M. *Nature* **2005**, *437*, 1173-1178.

(3) Ewing, R. M.; Chu, P.; Elisma, F.; Li, H.; Taylor, P.; Climie, S.; McBroom-Cerajewski, L.; Robinson, M. D.; O'Connor, L.; Li, M.; Taylor, R.; Dharsee, M.; Ho, Y.; Heilbut, A.; Moore, L.; Zhang, S.; Ornatsky, O.; Bukhman, Y. V.; Ethier, M.; Sheng, Y.; Vasilescu, J.; Abu-Farha, M.; Lambert, J. P.; Duewel, H. S.; Stewart, I. I.; Kuehl, B.; Hogue, K.; Colwill, K.; Gladwish, K.; Muskat, B.; Kinach, R.; Adams, S. L.; Moran, M. F.; Morin, G. B.; Topaloglou, T.; Figeys, D. *Mol. Syst. Biol.* **2007**, *3*, 89.

Chemical syntheses

An overview of the syntheses performed is given in Schemes S1-3.

General. Chemicals were purchased from Sigma-Aldrich, Fluka, Acros and Advanced Chem Tech, and were used without further purification. BG-NH_2 and BC-NH_2 were obtained from Covalys Biosciences and used without further purification. The syntheses of the intermediate **1** and **3** were described previously[1]. Anhydrous solvents from Sigma-Aldrich were used without further treatment and distillation. ^{1}H nuclear magnetic resonance (NMR) spectra were recorded on a Bruker DPX 400 (400.13 MHz), with chemical shifts (δ) reported in ppm relative to the solvent residual signals of CD_3OD (3.31 ppm), coupling constants reported in Hz. Mass spectra were recorded on a Shimazu Axima CFR*plus* Maldi-TOF MS. High resolution mass spectra (HRMS) were measured on a Micromass Q-TOf Ultima spectrometer with electron spray ionization (ESI). Reversed phase analytical HPLC was run on a Waters 2790 separation module and products were detected at 254 nm and 280 nm using a 2487 dual λ absorption detector. The standard gradient used for the purifications was: starting at 5% acetonitrile for 2 minutes and raising to 100% acetonitrile within 25 minutes. After 5 min at 100% acetonitrile, the column was equilibrated for 2 min with 5% acetonitrile. All HPLC solvents contained TFA at 0.1% v/v. A 3.9 × 300 mm Prep Nova-Pak HR C18 6 μm column from Waters was used to determine the purity of the products. Preparative HPLC was performed on a Waters 600 controller and with a Waters 2487 dual λ absorption detector using a SunFire™ Prep C18 OBD™ 5 μm 19×150 mm column.

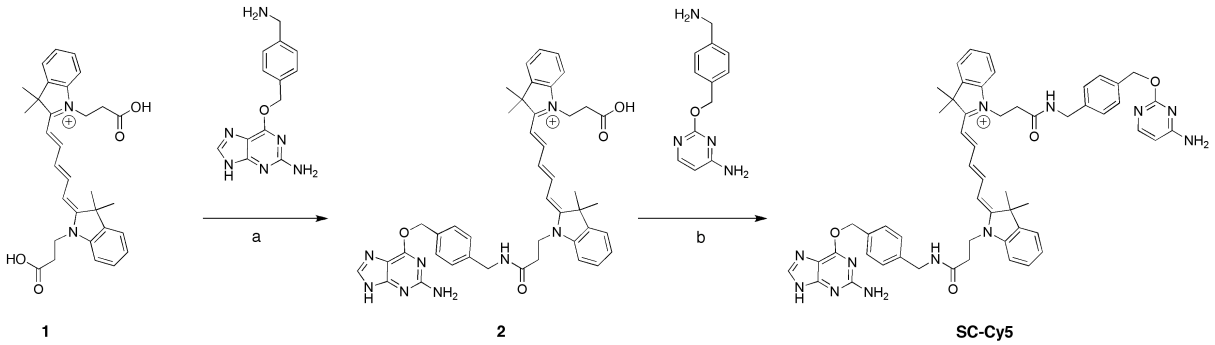
Synthesis of SC-Cy5. To a solution of compound **1** (20 mg, 0.04 mmol), PyBOP (13 mg, 0.025 mmol) and HOBt (3 mg, 0.025 mmol) in DMF (1 mL), DIPEA (3 mg, 0.025 mmol) and BG-NH_2 (7 mg, 0.025 mmol) were added at room temperature and stirred overnight. The solvent was evaporated to dryness *in vacuo*. The crude product was purified by preparative HPLC to give after lyophilization the product **2** as a blue powder in 32% yield. To a solution of compound **2** (5.4 mg, 7.2 μmol), DIC (2.6 mg, 0.025 mmol) and HOBt (3 mg, 0.025 mmol) in DMF (1 mL), BC-NH_2 (2.5 mg, 0.01 mmol) was added at room temperature and stirred overnight. The solvent was evaporated to dryness *in vacuo*. The crude product was purified by preparative HPLC to give after lyophilization the product **SC-Cy5** as a blue powder in 85% yield. ^{1}H-NMR (400MHz, CD_3OD): δ 8.25 (2H, m), 7.85 (2H, m), 7.42-7.22 (12H, m), 7.05

(4H, m), 6.60 (1H, t, J = 12 Hz) 6.35 (2H, m), 6.17 (1H, d, J = 8 Hz), 5.48 (2H, s), 5.25 (2H, s), 4.43 (4H, m), 4.26 (4H, s), 2.74 (4H, m), 1.55 (6H, s), 1.50 (6H, s). ESI-HRMS calcd for $[M+H]^+$ 964.4868, found *m/e* 964.4860.

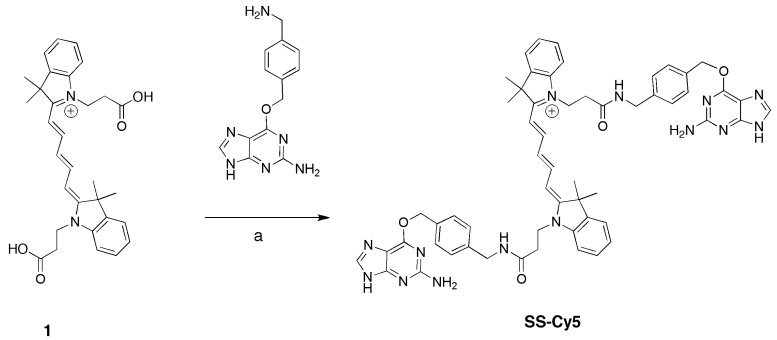
Synthesis of SS-Cy5. To a solution of compound **1** (10 mg, 0.02 mmol), PyBOP (26 mg, 0.05 mmol) and HOBt (3 mg, 0.025 mmol) in DMF (1 mL), DIPEA (7.8 mg, 0.06 mmol) and BG-NH_2 (13.5 mg, 0.05 mmol) were added at room temperature and stirred for 5h. The solvent was evaporated to dryness *in vacuo*. The crude product was purified by preparative HPLC to give after lyophilization the product **SS-Cy5** as a blue powder in 40% yield (8 mg). ^{1}H-NMR (400MHz, CD_3OD): δ 8.21 (2H, t, J = 12 Hz,), 7.75 (2H, s), 7.39-7.20 (12H, m), 6.97 (4H, d, J = 8 Hz), 6.58 (1H, t, J = 12Hz), 6.29 (2H, t, J = 12 Hz), 5.49 (2H, s), 4.42 (4H, m), 4.23 (4H, s), 2.71 (4H, m), 1.45 (12H, s). ESI-HRMS calcd for $[M+H]^{2+}$ 502.7500, found *m/e* 502.7508

Synthesis of compound 4. A solution of compound **3** (0.9 g, 3.87 mmol) in pyridine (8 ml) was heated to reflux. Triethylorthoformate (1.35 ml, 7.76 mmol) was added slowly. The solution was heated for 2h. The mixture was dissolved in CH_2Cl_2 and washed with 0.1 M HCl. After removal of the solvent in vacuo, the residue was dissolved in NH_3.aq (pH 11) and washed with CH_2Cl_2. pH was adjusted to 1-2 and the desired compound was extracted with CH_2Cl_2. The organic layer was dried over Na_2SO_4 and concentrated in vacuo. The compound **4** was obtained as a red powder in 13% yield (236 mg) and was used without further purification. ^{1}H-NMR (400MHz, CD_3OD): δ 8.58 (1H, t, J = 12Hz), 7.54 (2H, d, J = 8Hz), 7.48 (2H, t, J = 8Hz), 7.41 (2H, d, J = 8Hz), 7.34 (2H, d, J = 8Hz), 6.49 (2H, t, J = 12 Hz), 4.46 (4H, t, J =8Hz), 2.87 (4H, t, J = 8Hz), 1.79(12H, s).

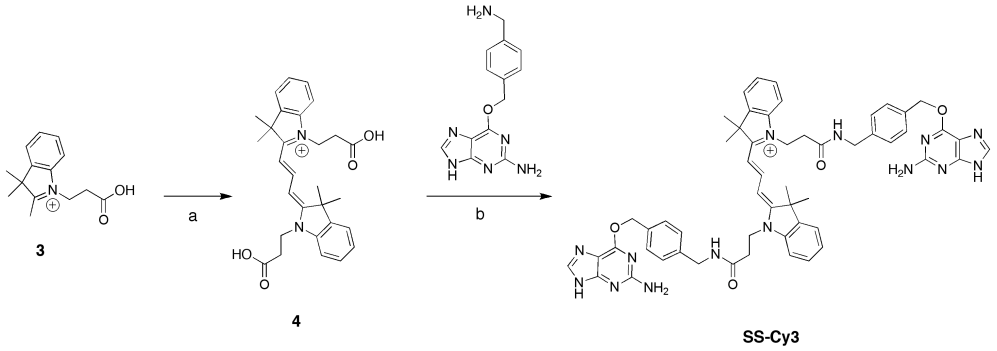
Synthesis of SS-Cy3. A solution of compound **4** (5 mg, 0.01 mmol), BG-NH_2(10 mg, 0.04 mmol), DIEA(3 mg, 0.02 mmol) and PyBOP (10 mg, 0.02 mmol) in dry DMF (1 ml) was stirred for 5h at room temperature. The crude product was purified by preparative HPLC to give after lyophilization the compound **SS-Cy3** as a red powder in 26% yield (2.5 mg). ^{1}H-NMR (400MHz, CD_3OD): δ 8.33 (1H, t, J = 12 Hz), 7.90 (2H, s), 7.44-7.29 (12H, m), 6.95 (4H, d, J = 8 Hz), 6.38 (2H, d, J = 12 Hz), 5.51 (4H, s), 4.44 (4H, m), 4.24 (4H, s), 2.75 (4H, m), 1.45 (12H, s). ESI-HRMS calcd for $[M+H]^{2+}$ 489.2383, found *m/e* 489.2367



Scheme S1: Synthesis of SC-Cy5. (a) PyBOP, HOBt, DIPEA, DMF, room temperature, overnight, 32% yield; (b) DIC, HOBt, DMF, room temperature, overnight, 85% yield.



Scheme S2: Synthesis of SS-Cy5. (a) PyBOP, HOBt, DIPEA, DMF, room temperature, 5 hours, 40% yield.



Scheme S3: Synthesis of SS-Cy3. (a) Triethylorthoformate, pyridine, reflux, 2 hours, 13% yield; (b) PyBOP, DIEA, DMF, room temperature, 5 hours, 26% yield.

Table S1. For each experiment the lysis buffer, the quantity of extract loaded on gel and the concentration of each fusion proteins in nM and pmol/mg are listed.

Experiment	Lysis buffer[a]	Quantity of extract loaded on gel (μg)	Fusion proteins	Conc. of fusion protein in extract (nM)	Conc. of fusion protein in extract (pmol/mg)
Figure 3a/3b	1	50	SNAP-FKBP	130	20
			CLIP-FRB	115	17
Figure 3d/3e	1	40	SNAP-FKBP	350	65
Figure 4a	1	40	SNAP-FKBP	120	22
			CLIP-FRB	100	19
Figure 4c	2	20	SNAP-FKBP	25	10
			CLIP-FRB	20	7.5
			CLIP-ALK5	25	10
Figure 5a	1	50	SNAP-Mdm2	20	3
			CLIP-p53	40	6
Figure 5c	1	65	SNAP-p53	10	1.4
			SNAP-Mdm2	6	2.3
Figure 5d	2	90	SNAP-p53	8	1.5
			SNAP-p53Δ	20	3.3
Figure 5e	2	65	SNAP-p53Δ	8	2
			CLIP-Mdm2	6	1.5
Figure 6a	2	30	SNAP-FKBP-F36M	250	62.5
Figure 7b	2	35	SNAP-MEK1 (lane1)	120	25
			ERK2-CLIP (lane 1)	120	25
			SNAP-MEK1 (lane 2)	150	32
			CLIP-FKBP (lane 2)	150	32
			SNAP-FKBP (lane3)	170	36
			ERK2-CLIP (lane 3)	50	10
			SNAP-MEK1 (lane 4)	70	15
			ERK2-CLIP (lane 4)	50	10
			SNAP-FKBP (lane 4)	130	28

[a] see Materials and Methods for the composition of buffers 1 and 2.

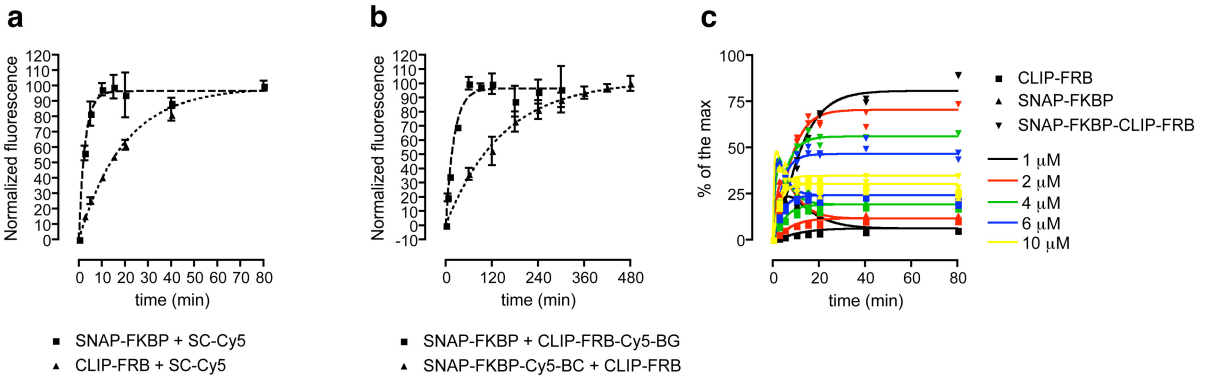


Figure S1. Kinetic study of the inter- and intramolecular crosslinking reactions. (a) Determination of k_1 and k_3 (see Figure 2a). Reaction progress curves of the labeling (represented by the normalized fluorescence) of recombinant SNAP-FKBP (0.2 µM) and CLIP-FRB (0.2 µM) with 2 µM of SC-Cy5. Data represent the mean ± SD of two independent experiments. Data were fitted with a pseudo-first order rate model. (b) Determination of k_2 and k_4 (see Figure 2a). Reaction progress curves of the reactions of purified SNAP-FKBP-Cy5-BC (0.25 µM) with CLIP-FKBP (5 µM), and purified CLIP-FRB-Cy5-BG (0.25 µM) with SNAP-FKBP (5 µM). Data represent the mean ± SD of two independent experiments. Data were fitted with a pseudo-first order rate model. (c) Determination of k'_1, k'_2, k'_3 and k'_4 (see Figure 2b). Recombinant SNAP-FKBP (0.2 µM) and CLIP-FRB (0.2 µM) treated with rapamycin (5 µM) were incubated with different concentration of SC-Cy5 (1, 2, 4, 6 and 10 µM) for different times. The graph shows the percentages (values of two independent experiments) of formed SNAP-FKBP-CLIP-FRB, labeled CLIP-FRB and labeled SNAP-FKBP as a function of time for the different concentration of SC-Cy5. Data were fitted with a kinetic model described in Figure 2b using script given in Annex 1.

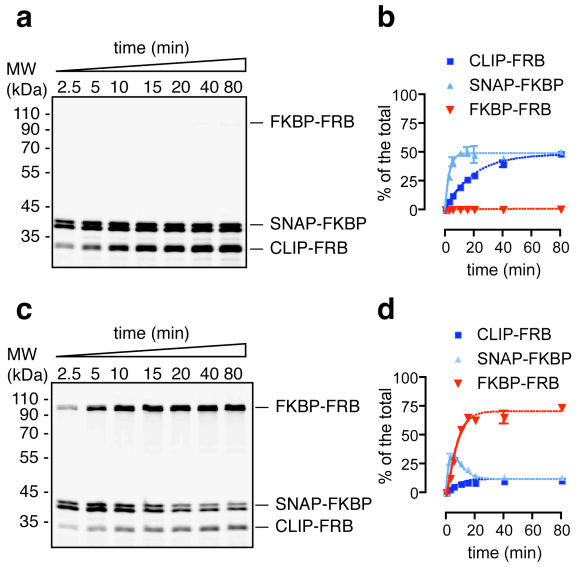


Figure S2. Reactions of recombinant SNAP-FKBP (0.2 µM) and CLIP-FRB (0.2 µM) with 2 µM of SC-Cy5 in presence of 0 µM (a,b) or 5 µM (c,d) of rapamycin. (a, c) Analysis of the crosslinking reaction by SDS-PAGE and in-gel fluorescence scanning (NB: SNAP-FKBP appears as a double band on the gels). (b, d) Reaction progress curves. Data represents mean ± SD of two independent experiments. For clarity, in this figure the names of the crosslinked proteins are given without the name of the tags.

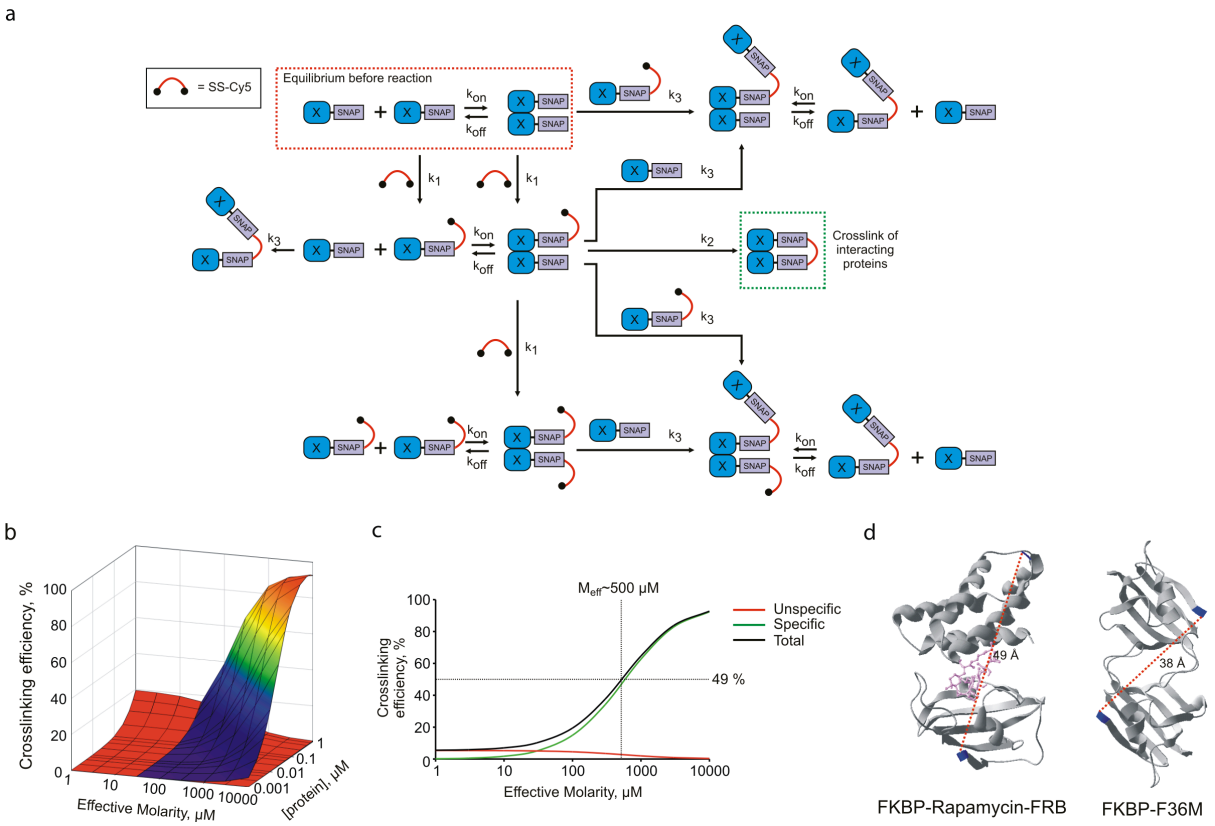


Figure S3. Crosslinking of a weak homodimer; the FKBP-F36M homodimer as an example. (**a**) Kinetic scheme for the crosslinking of a dimeric SNAP-tag fusion protein using SS-Cy5. For simplicity, all the reactions leading to an intermolecular crosslinking are described with the same rate constant k_3; all the reactions leading to labeling of SNAP-tag with SS-Cy5 are described with the same rate constant k_1. (**b**) Calculated crosslinking efficiency after 60 min of reaction as function of the effective molarity ($M_{eff} = k_2/k_3$) and protein concentration while fixing K_D = 30 μM and [SS-Cy5] = 1 μM. The multi-color surface represents the crosslinked product resulting from intramolecular crosslinking (step with rate constant k_2; see script in Annex 2). The red surface represents the product resulting from intermolecular crosslinking (steps with rate constant k_3; see script in Annex 3). The scheme described in (a) was used for the simulation with the following rate constants: k_{on} was fixed to 10 $\mu M^{-1}min^{-1}$; k_{off} was fixed to 300 min^{-1}; k_1 was fixed to 0.036 $\mu M^{-1}min^{-1}$, the rate constant of the reaction of SNAP-FKBP with the crosslinker determined experimentally in cell lysate (data not shown); k_3 was fixed to 0.01 $\mu M^{-1}min^{-1}$, the rate constant of the intermolecular crosslinking with SNAP-FKBP monomer determined experimentally in vitro (rate constant k_4 in Figure 2). The calculation showed that a weak interaction with a K_D of 30 μM can be efficiently crosslinked only when both the protein concentration in the cell extract and M_{eff} are high; the drop of one of these two parameters or of both leads to a drastic decrease of the crosslinking efficiency. (**c**) Two-dimensional cut of (b) at protein concentration of 250 nM (i.e. FKBP-F36M concentration used in experiment). The green line represents the specific signal, the red line represents the unspecific signal and the black line represents the sum of the two. As can be seen from the graph, the observed crosslinking efficiency for SNAP-FKBP-F36M requires an effective molarity of 500 μM. This value was obtained by fitting the dependence of SNAP-FKBP-F36M crosslinking efficiency on protein concentration with the model described in (a) (see script in Annex 4). (**d**) Comparison of structures of FKBP-F36M homodimer (PDB Id: 1EYM) and FKBP-Rapamycin-FRB (PDB Id: 1NSG) heterodimer. The higher effective

molarity value predicted from (c) is consistent with the shorter distance between the N-termini in the homodimer of FKBP-F36M (38 Å) compared to the FKBP-FRB heterodimer (49 Å). The indicated distances are between the N-terminal amino acids.

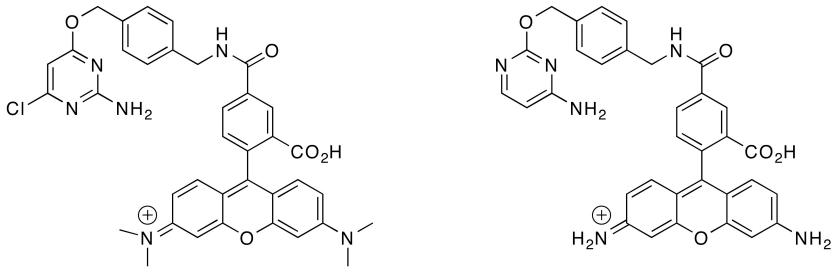


Figure S4. Structures of TMR-star and BC-505.

Supplementary References

(1) Brun, M. A.; Tan, K. T.; Nakata, E.; Hinner, M. J.; Johnsson, K. *J. Am. Chem. Soc.* **2009**, *131*, 5873-5884.

Annex 1. DYNAFIT script used for the determination of the rate constants k'_1, k'_2, k'_3 and k'_4 of the intramolecular crosslinking process (see Figure S1c).

```
[task]
task = fit
data = progress

[mechanism]
P  + S -> PSX      : k1
P  + S -> PSY      : k3
PSX -> PD          : k2
PSY -> PD          : k4
PSX + S -> PSS     : k3
PSY + S -> PSS     : k1

[constants]
k1 = 0.1 ?    ; μM^-1min^-1
k2 = 0.1 ?    ; min^-1
k3 = 0.01 ?   ; μM^-1min^-1
k4 = 1 ?      ; min^-1

[concentrations]
P  = 0.2  ; μM

[progress]
directory      C:\
extension      txt
file      PS1
conc. S = 1
response PD = 0, PSY = 0, PSX = 426208, PSS = 301897
file      PS2
conc. S = 2
response PD = 0, PSY = 0, PSX = 426208, PSS = 301897
file      PS3
conc. S = 4
response PD = 0, PSY = 0, PSX = 426208, PSS = 301897
file      PS4
conc. S = 6
response PD = 0, PSY = 0, PSX = 426208, PSS = 301897
file      PS5
conc. S = 10
response PD = 0, PSY = 0, PSX = 426208, PSS = 301897
file      PSS1
conc. S = 1
response PD = 0, PSY = 0, PSX = 0, PSS = 301897
file      PSS2
conc. S = 2
response PD = 0, PSY = 0, PSX = 0, PSS = 301897
file      PSS3
conc. S = 4
response PD = 0, PSY = 0, PSX = 0, PSS = 301897
file      PSS4
conc. S = 6
response PD = 0, PSY = 0, PSX = 0, PSS = 301897
file      PSS5
conc. S = 10
response PD = 0, PSY = 0, PSX = 0, PSS = 301897
file      PD1
conc. S = 1
response PD = 576244, PSY = 0, PSX = 0, PSS = 0
file      PD2
conc. S = 2
response PD = 576244, PSY = 0, PSX = 0, PSS = 0
file      PD3
conc. S = 4
response PD = 576244, PSY = 0, PSX = 0, PSS = 0
file      PD4
conc. S = 6
response PD = 576244, PSY = 0, PSX = 0, PSS = 0
file      PD5
conc. S = 10
response PD = 576244, PSY = 0, PSX = 0, PSS = 0

[output]
```

Annex 2. DYNAFIT script used for the simulation of the signal generated by the crosslinking of interacting proteins (PD) (see Figure S3a,b).

```
[task]
task = simulate
data = progress

[mechanism]

;----------------------------main crosslinking reactions-----------------------------------

SX + SX <-> SXSX       : kon koff  ; main equilibrium
SX + L -> SXL          : k1        ; labeling of monomer
SXL + SXL <-> SXLSXL   : kon koff  ; equilibrium of labeled monomers
SX + SXL <-> SXSXL     : kon koff  ; equilibrium of labeled and unlabeled monomers
SXSXL -> PD            : k2        ; specific crosslinking
SXSXL + L -> SXLSXL    : k1        ; second step of labeling of dimer
SXSX + L -> SXSXL      : k1        ; first step of labeling of dimer
SXL + SX -> FALSE      : k3        ; Unspecific crosslinking

;----------------------------Dimer-monomer unspecific crosslinking--------------------------

   SXSX + SXL -> SXFALSE    : k3
   SX + FALSE <-> SXFALSE   : kon koff
   SXSXL + SX -> SXFALSE    : k3
   SXSXL + SXL -> SXLFALSE  : k3
   SXL + FALSE <-> SXLFALSE : kon koff
   SXLSXL + SX -> SXLFALSE  : k3

[constants]
kon = 10          ; µM^-1min^-1
koff = 300        ; min^-1
k1  = 0.036       ; µM^-1min^-1
k2  = 0.01        ; min^-1
k3  = 0.01        ; µM^-1min^-1

-------------------------------------------------
[concentrations]
SX = 0.001 ; µM
L = 1  ; µM

[responses]
PD = 200000 ; 2*100/[SX]

[sweep]
k2 = 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100

[progress]
directory C:\data\
extension      txt
mesh       from 0 to 300   step 2.5
file 0_001.txt

[output]
directory C:\output\
-------------------------------------------------

;task done for:
;[SX] = 0.0025, 0.005, 0.0075, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 1 µM

[end]
```

Annex 3. DYNAFIT script used for the simulation of the signal generated by the crosslinking of non-interacting proteins (FALSE) (see Figure S3a,b).

```
[task]
task = simulate
data = progress

[mechanism]

;----------------------------main crosslinking reactions-----------------------------------

SX + SX <-> SXSX       : kon koff  ; main equilibrium
SX + L -> SXL          : k1        ; labeling of monomer
SXL + SXL <-> SXLSXL   : kon koff  ; equilibrium of labeled monomers
SX + SXL <-> SXSXL     : kon koff  ; equilibrium of labeled and unlabeled monomers
SXSXL -> PD            : k2        ; specific crosslinking
SXSXL + L -> SXLSXL    : k1        ; second step of labeling of dimer
SXSX + L -> SXSXL      : k1        ; first step of labeling of dimer
SXL + SX -> FALSE      : k3        ; Unspecific crosslinking

;----------------------------Dimer-monomer unspecific crosslinking-------------------------

   SXSX + SXL -> SXFALSE    : k3
   SX + FALSE <-> SXFALSE   : kon koff
   SXSXL + SX -> SXFALSE    : k3
   SXSXL + SXL -> SXLFALSE  : k3
   SXL + FALSE <-> SXLFALSE : kon koff
   SXLSXL + SX -> SXLFALSE  : k3

[constants]
kon = 10          ; µM-1min-1
koff = 300        ; min-1
k1  = 0.036       ; µM-1min-1
k2  = 0.01        ; min-1
k3  = 0.01        ; µM-1min-1

-------------------------------------------------
[concentrations]
SX = 0.001 ; µM
L = 1  ; µM

[responses]
FALSE =200000 ; 200/[SX]

[sweep]
k2 = 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100

[progress]
directory C:\data\
extension      txt
mesh       from 0 to 300   step 2.5
file 0_001.txt

directory C:\output\
-------------------------------------------------
;task done for:
;SX = 0.0025, 0.005, 0.0075, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 1

[end]
```

Annex 4. DYNAFIT script used for the fitting of the experimentally determined crosslinking efficiencies of FKBP-F36M homodimer (see Figure S3c).

```
[task]
task = fit
data = progress

[mechanism]

;----------------------------main crosslinking reactions-------------------------------------

SX + SX <-> SXSX       : kon koff  ; main equilibrium
SX + L -> SXL          : k1        ; labeling of monomer
SXL + SXL <-> SXLSXL   : kon koff  ; equilibrium of labeled monomers
SX + SXL <-> SXSXL     : kon koff  ; equilibrium of labeled and unlabeled monomers
SXSXL -> PD            : k2        ; specific crosslinking
SXSXL + L -> SXLSXL    : k1        ; second step of labeling of dimer
SXSX + L -> SXSXL      : k1        ; first step of labeling of dimer
SXL + SX -> FALSE      : k3        ; Unspecific crosslinking

;----------------------------Dimer-monomer unspecific crosslinking--------------------------

   SXSX + SXL -> SXFALSE    : k3
   SX + FALSE <-> SXFALSE   : kon koff
   SXSXL + SX -> SXFALSE    : k3
   SXSXL + SXL -> SXLFALSE  : k3
   SXL + FALSE <-> SXLFALSE : kon koff
   SXLSXL + SX -> SXLFALSE  : k3


[constants]
kon  = 10          ; µM^-1min^-1
koff = 300         ; min^-1
k1   = 0.036       ; µM^-1min^-1
k2   = 4 ?         ; min^-1
k3   = 0.01        ; µM^-1min^-1

[concentrations]
L = 1  ; µM

[responses]
PD = 1
FALSE = 1

[progress]
directory      C:\data\
extension      txt
file 1, 2, 3, 4
vary conc. SX = 0.002, 0.02, 0.2, 0.25

[output]
directory C:\output\

[end]
```