Supplementary Materials for
Design of Biologically Active Binary Protein 2D Materials

Supplemental methods:

- **Microscope calibration and comparison between preformed arrays and arrays made on cells**

List of supplemental figures:

- **Figure S1 | DNA translation and mRNA optimization protocol diagram**
- **Figure S2 | mRNA optimization expression yield**
- **Figure S3 | Designs preliminary SDS-PAGE gels screening**
- **Figure S4 | Circular dichroism analysis of the stabilized components B**
- **Figure S5 | Circular dichroism analysis of the stabilized components A**
- **Figure S6 | Long range order of assembled arrays assessed by TEM**
- **Figure S7 | Designed cyclic pseudo-dihedral building blocks**
- **Figure S8 | Characterization of multi-functional pre-assembled arrays and specific binding to cells expressing Notch1 receptors**
- **Figure S9 | Arrays functionalized with DLL4 recruit Notch in U2OS cells**
- **Figure S10 | Gel source data for the various Western Blots presented in this paper**

List of supplemental tables:

- **Table S1 | mRNA optimization protocol performance**
- **Table S2 | mRNA sequences optimization of component B D_2 homooligomer**
- **Table S3 | Small scale expression SDS screening**
- **Table S4 | Designs and native protein sequences**
- **Table S5 | Sequences of the B component stabilized versions**
- **Table S6 | Designed components pre- and post-stabilization individual solubility**
- **Table S7 | SAXS scattering analysis**
- **Table S8 | B component desymmetrization linkers list**
- **Table S9 | A component desymmetrization linkers list**
- **Table S10 | A-SpyCatcher/spyTag-fDomain and spyTag-DLL4 sequences for arrays-cell receptors binding**

List of supplemental movies:

- **Movie S1. Design strategy PyMOL illustration: dock, design, and propagation**
- **Movie S2. Instantaneous gelation upon components mixing**
- **Movie S3. Clustering of intracellular mScarlet constructs by preformed arrays**
- **Movie S4. 3D rendering of cell incubated with preformed arrays**
- **Movie S5. Stability of receptor clustering assessed by FRAP**
- **Movie S6. Growth of arrays onto cells**

Supplemental methods

Microscope calibration and comparison between preformed arrays and arrays made on cells

To calibrate the TIRF and Spinning disk setup described above in terms of estimated number of GFP and mScarlet molecules, we mixed our previously published GFP-60mer nanocages[1] with an excess of a purified GBP-mscarlet fusion (see Extended Data Figure 10f). Excess of unbound GBP-mscarlet was then removed by size exclusion chromatography on a superose 6 column, and GBP-mScarlet induced a shift in molecular weight of the GFP-60mer (see Extended Data Figure 10g). Near 1:1 binding ratio was confirmed by absorbance measurement at 490 nm and 561 nm. Indeed, absorbance of mScarlet-GBP/GFP-60mer at 570 nm was 0.091 so 907 nM of mScarlet (Extinction coefficient of mscarlet is 100,330 and it does not absorb at 470 nm). On the other hand, absorbance at 490nm was 0.092, so 862 nM of GFP after correction for mScarlet absorbance at 490 nm. This gives a ratio GFP/mScarlet of 0.95. We found that the GFP fluorescence of the mScarlet-GBP/GFP-60mer nanocages was almost identical to that of GFP-60mer nanocages, suggesting that FRET with mScarlet does not lower the fluorescence of GFP (or that it is compensated by the increase of fluorescence due to the "enhancer" nanobody we used[2]).

We then acquired z-stacks of diluted nanocages in the same buffer as the cells' imaging medium, which revealed discrete particles fluorescing on both the GFP and mScarlet channels (see Extended Data Figure 10h). We then z-projected the planes containing particle signal (maximum intensity projection), and automatically detected the particles by 2D Gaussian fitting using the Thunderstorm algorithm[3]. We then assessed the colocalization between GFP and mScarlet-positive particles by considering colocalized particles whose distance between GFP and mScarlet fluorescence centroid is below 200 nm. Non colocalizing particles were discarded, and we then estimated the average fluorescence of one 60-mer by computing the median of the integrated fluorescence intensity from the gaussian fitting (minus the background) for each channel (Extended Data Figure 10h-i). By dividing this median fluorescence by the number of GFP/mCherry per nanocage (i.e. 60), we can estimate the fluorescence of one GFP (respectively one mScarlet) molecule. From this, we can evaluate the approximate number of GFP and mScarlet molecules per diffraction-limited spot on a cell by keeping the exposure and laser power constant between calibration and experiment (see equations below for derivation of the error estimated on these measurements).

Extended Data Figure 10i shows that fluorescence intensity increases linearly with exposure time, suggesting that the instrument (spinning disk in this case) operates in its linear range. This calibration was done for each microscope and to ensure that laser fluctuations were not a variable, calibration datasets were acquired on the same day as an experiment. Care was taken to perform these measurements in areas of the

field of view where illumination was homogenous (about 50% for the spinning disk and about 80% for the TIRF). Note that because of azimuthal illumination, our TIRF instrument does not suffer from shadowing effects, and that for Extended Data Figure 11b, we used 60mer-GFP (not mScarlet-GBP/GFP-60mer) calibration nanocages.

Mathematically, conversions into number of molecules, and their associated error, were performed by building on the elegant work of Picco and colleagues as follows:[4]

I_{GFP} is the integrated intensity of the arrays in the GFP channel (n measurements) and I_{60GFP} is the integrated intensity of the reference 60mer in the same channel (n' measurements). As distribution of dim signals are skewed, estimated average values for I_{GFP}, noted $\widehat{I_{GFP}}$, is computed as median of the distribution. The estimate for the reference 60mer, $\widehat{I_{60GFP}}$, is similarly computed from I_{60GFP}. The respective error associated with these measurements, noted$\delta_{\widehat{I_{GFP}}}$ and $\delta_{\widehat{I_{60GFP}}}$, respectively, are estimated with the Median Absolute Deviation (MAD) corrected for asymptotically normal consistency on the natural logarithm transform of the raw fluorescence values I_{GFP} and I_{60GFP}.

$$\widehat{I_{GFP}} = exp\left(median(ln\,(I_{GFP})\,)\right) = median(I_{GFP})$$

$$\delta_{\widehat{I_{GFP}}} = \widehat{I_{GFP}} \times \frac{1.4826 \times MAD\,(ln\,(I_{GFP})\,)}{\sqrt{n}}$$

$$\widehat{I_{60GFP}} = median(I_{60GFP})$$

$$\delta_{\widehat{I_{60GFP}}} = \widehat{I_{60GFP}} \times \frac{1.4826 \times MAD\,(ln\,(I_{60GFP})\,)}{\sqrt{n'}}$$

The estimate of number of GFP molecule per array was computed as

$$\widehat{n_{GFP}} = \frac{\widehat{I_{GFP}}}{\widehat{I_{60GFP}}} \times 60$$

The uncertainty over this number of molecules, $\delta_{n\,GFP}$, was computed by error propagation as

$$\delta_{\widehat{n_{GFP}}} = \sqrt{\left(\frac{60}{\widehat{I_{60GFP}}} \times \delta_{\widehat{I_{GFP}}}\right)^2 + \left(60 \times \frac{\widehat{I_{GFP}}}{\left(\widehat{I_{60GFP}}\right)^2}\delta_{\widehat{I_{60GFP}}}\right)^2}$$

Similarly, the number of molecules in the mScarlet channel, $n_{mScarlet}$ was estimated from $I_{mScarlet}$, the integrated intensity of the arrays in the mScarlet channel (n measurements) and the intensity of the reference 60mer in the same channel, $I_{60mScarlet}$ (n' measurements).

$$\widehat{I_{mScarlet}} = exp\left(median(ln\,(I_{mScarlet})\,)\right) = median(I_{mScarlet})$$

$$\delta_{I_{m\widehat{Scarlet}}} = I_{\widehat{mScarlet}} \times \frac{1.4826 \times MAD\ (ln\ (I_{mScarlet}))}{\sqrt{n}}$$

$$I_{60\widehat{mScarlet}} = median(I_{60mScarlet})$$

$$\delta_{I_{60\widehat{mScarlet}}} = I_{60\widehat{mScarlet}} \times \frac{1.4826 \times MAD\ (ln\ (I_{60mScarlet}))}{\sqrt{n'}}$$

The estimate of number of mScarlet molecules per array was computed as

$$n_{\widehat{mScarlet}} = \frac{I_{\widehat{mScarlet}}}{I_{\widehat{60mScarlet}}} \times 60$$

The uncertainty over this number of molecules, $\delta_{n\ GFP}$, was computed by error propagation as

$$\delta_{n_{\widehat{mScarlet}}} = \sqrt{\left(\frac{60}{I_{\widehat{60mScarlet}}} \times \delta_{I_{m\widehat{Scarlet}}}\right)^2 + \left(60 \times \frac{I_{\widehat{mScarlet}}}{\left(I_{\widehat{60mScarlet}}\right)^2} \delta_{I_{60\widehat{mScarlet}}}\right)^2}$$

We then estimated the GFP/mscarlet ratio on cells in terms of molecules, $\frac{\widehat{n_{GFP}}}{n_{\widehat{mScarlet}}}$(Extended Data Figure 10k). Its associated error, $\delta_{\frac{\widehat{GFP}}{mScarlet}}$ is computed as:

$$\delta_{\frac{\widehat{GFP}}{mScarlet}} = \sqrt{\left(\frac{\delta_{\widehat{n_{GFP}}}}{n_{\widehat{mScarlet}}}\right)^2 + \left(\frac{\widehat{n_{GFP}}}{n_{\widehat{mScarlet}^2}}\ \delta_{n_{\widehat{mScarlet}}}\right)^2}$$

To compare the lattice order between arrays made on cells and preformed arrays (Fig. 4f), we formed **B(c)GFP**/**A(d)mScarlet** arrays on cells or in vitro, then imaged them and measured array fluorescence by gaussian fitting as above. Preformed arrays were obtained by mixing 5 µM **B(c)GFP** with 5 µM **A(d)mScarlet** in (TBS-0.5 M Imidazole) for 4h at RT, followed by ultracentrifugation (250,000 x *g* ; 30 min) and dilution into PBS for imaging onto the same dishes as the cells. We verified the order of these arrays by EM (Extended Data Figure 8d). Using the notations introduced above, we measured the mScarlet/GFP ratio as

$$\widehat{mScarlet/GFP} = \frac{I_{\widehat{mScarlet}}}{I_{\widehat{GFP}}}$$

Its associated error, $\delta_{\widehat{mScarlet/GFP}}$ is computed as:

$$\delta_{\widehat{mScarlet/GFP}} = \sqrt{\left(\frac{\delta_{n_{\widehat{mScarlet}}}}{I_{\widehat{GFP}}}\right)^2 + \left(\frac{I_{\widehat{mScarlet}}}{I_{\widehat{GFP}^2}}\ \delta_{\widehat{n_{GFP}}}\right)^2}$$

We verified that the mScarlet/GFP fluorescence ratio varies as expected from the structure, and is thus a good proxy of bulk order (Extended Data Figure 10m). To do so, we formed **B(c)GFP**/**A(d)mScarlet** arrays in vitro as above, then incubated them with a 2-fold molar excess of GBP-mScarlet over **B(c)GFP** for 1h at RT, followed by ultracentrifugation (250,000 x *g* ; 30min) and dilution into PBS for imaging onto the same dishes as the cells. As binding of the GBP-mScarlet to GFP does not effectively modify the fluorescence of GFP (see above), the predicted variation of the mScarlet/GFP fluorescence ratio upon saturation of each GFP by GBP-mScarlet is:

$$\left(\widehat{mScarlet}/GFP\right)_{with\ GBP-mScarlet} = 3/2 \times \left(\widehat{mScarlet}/GFP\right)_{without\ GBP-mScarlet}$$

To estimate the **A**/**B** ratio on cells (Extended Data Figure 10n) we incubated cells with **B(c)GFP** and **A(d)mScarlet**. As the distance between GFP and mScarlet within the arrays is $r = 6.09\,nm$, there is significant FRET between the two molecules. The FRET efficiency is given by $E = \frac{1}{1+(r/R_0)^6} = 0.39$ with $R_0 = 56.75$ (https://www.fpbase.org/fret/). To the GFP intensity I_{GFP} is corrected by a factor $\frac{1}{1-E}$ to account for FRET in order to evaluate $\widehat{n_{GFP}}$ as above. As dihedral components have twice more fluorophore than cyclic ones per unit cell, the mean **A**/**B** ratio, noted $\widehat{A/B}$ is computed as follows:

$$\widehat{A/B} = \frac{n_{\widehat{mScarlet}}}{2 \times \widehat{n_{GFP}}}$$

Its associated error, $\delta_{\widehat{A/B}}$ is computed as:

$$\delta_{\widehat{A/B}} = \sqrt{\left(\frac{\delta_{n_{\widehat{mScarlet}}}}{2 \times \widehat{n_{GFP}}}\right)^2 + \left(\frac{n_{\widehat{mScarlet}}}{2 \times \widehat{n_{GFP}}^2}\ \delta_{\widehat{n_{GFP}}}\right)^2}$$

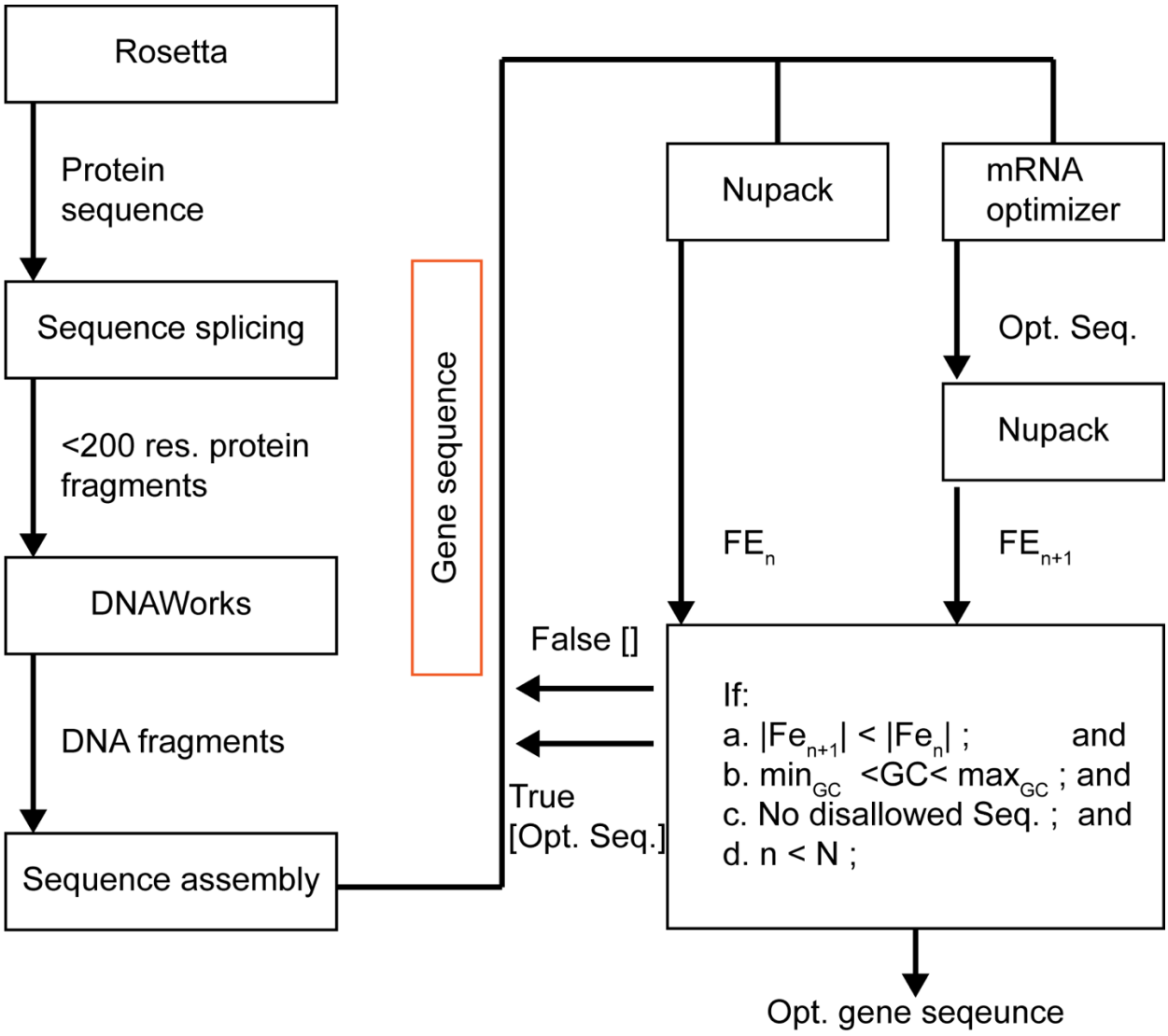


Figure S1. DNA translation and mRNA optimization protocol diagram

DNAworks[5], Nupack[6], and mRNA optimizer[7] are wrapped in a python program to optimize for protein expression in E. Coli. and compatible with some typical requirements (such as GC ratio, repeat, restriction site, ets.) of providers cloning production lines. Once a desired protein sequence is obtained it is parsed to fragments of up to 200 residues (limit of DNAworks) which are passed separately to DNAworks for translation. The DNA sequences are then stitched back into a single fragment and the first *n* nucleotides of each gene, typically 50, are then optimized by the mRNAoptimizer and Nupack iteratively to minimize the mRNA secondary structure ddG. The rationale is minimizing the occurrence of mRNA secondary structures which reduce the yield of protein expression by slowing or blocking the initiation and flow of the mRNA fragment through the ribosomes.[7]

	Pre-optimized ddG	Optimized ddG	ddG	ddG reduction [%]
avg	-13.67	-8.72	4.94	35.68
std	3.58	3.40	3.19	20.49
min	-31.04	-22.84	0.00	0.00
max	-2.68	-0.17	22.69	98.00

Table S1. mRNA optimization protocol performance

44995 sequences were optimized, the table shows the sequence ddG before and after optimization, the difference between the two, and the ratio. An average of 35.68% ddG reduction is shown across all designs, in close agreement with the authors result of 40%, in spite of the additional constraint in our system, e.g. CG ratio limits, repeats, and restriction sites. We acknowledge Zachary R Crook for testing the system and providing the data.

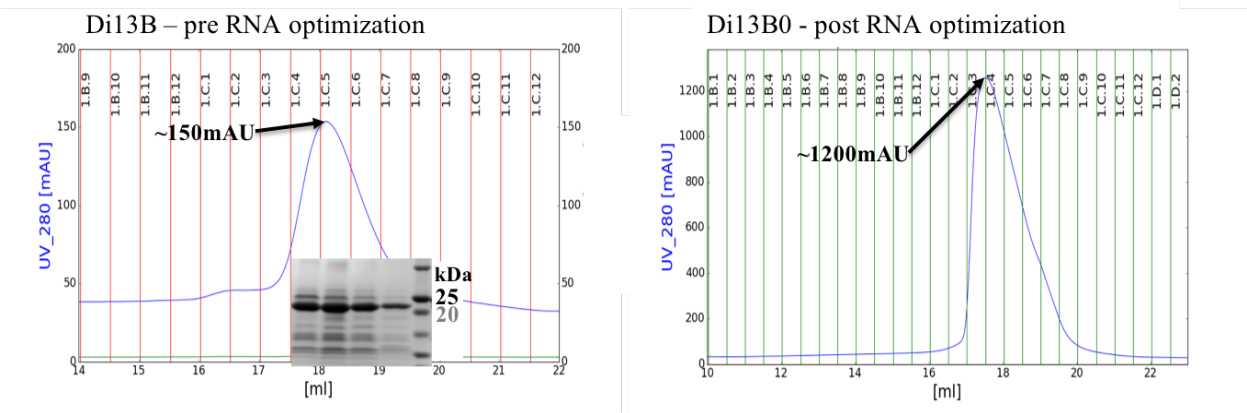


Figure S2. mRNA optimization expression yield

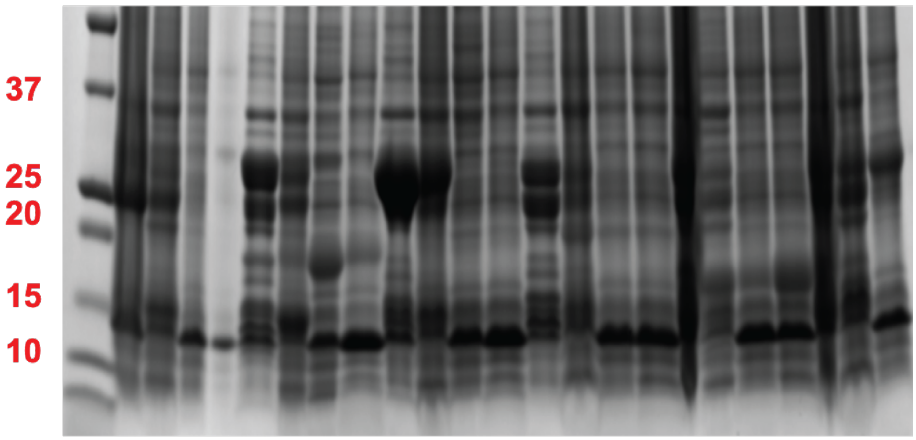
FPLC traces (Superose 6 Increase 10/300 GL) of the **B** component. Both cultures were similarly lysed as described in methods, soluble fractions were separated using centrifugation and further purified using Ni-columns, eluted fractions were concentrated to ~1ml and immediately injected to the FPLC on 1ml loops. We note that both constructs residues sequence is identical and differ only in the DNA sequence of the first 50 nucleotides. SDS-PAGE gel in the inset to the left panel shows the bands of the corresponding fractions at the expected weight. We see an 8-fold yield increase for the mRNA optimized construct vs. the non-mRNA Optimized construct.

Name	Nucleotides sequences	Orig. FE	Opt. FE
B	ATGGGTAGCCTGATAACGCTAGTGGAACTGGAATGGCTAGAG CACCAGCTA	-18.61	-
B0	ATGGGTTCCCTAATCACCCTCGTCGAACTCGAATGGCTCGAA CACCAACTC	-18.61	-7.26

Table S2. mRNA sequences optimization of the B component (D_2 homooligomer)

mRNA optimization was performed on the first 51 bps. We named the optimized construct as B0 for its identical protein sequence. The mRNA optimizer reduced (in absolute value) the mRNA secondary structure expected ddG by 61% to 7.26 kcal/mol.

Design name	Di_1	Di_2	Di_3	Di_4	Di_5	Di_7
Tag position	NB	NA	NA	CB	NB	NA
Comp. A	1zcc	3lub	1i1g	1i1g	1o4v	3mgl
Comp. B	3gvd	2cqz	2c9j	2iz6	4qq8	2ux8
Comp. A	27.38	29.36	17.11	15.85	18.45	15.66
Comp. B	30.73	20.41	25.37	17.84	61.69	31.30



Design name	Di_9	Di_10	Di_11	Di_12	Di_13	Di_14
Tag position	NB	CB	CB	NA	CB	CA
Comp. A	3r2e	3trh	4bop	4e2s	1d2t	1d2t
Comp. B	3no6	2h8n	2vvh	1a17	1tk9	2w3x
Comp. A	14.99	17.61	17.43	31.36	25.15	26.07
Comp. B	28.19	13.62	26.74	18.93	21.38	17.07

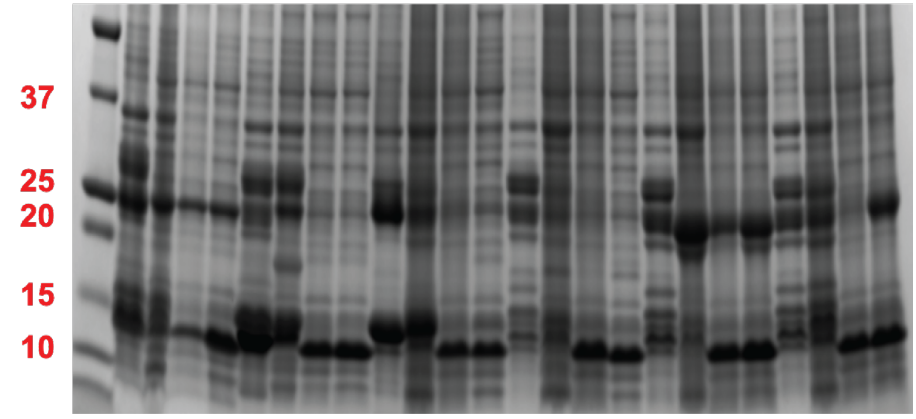


Figure S3. Designs preliminary SDS-PAGE gels screening.

For the initial screening of the 45 designs, bicistronic plasmids were transformed into BL21 Star (DE3) E. coli. cells (Invitrogen) and cultures grown in 4ml LB media in a 96 well plate setup. Protein expression was induced with 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) for 3 hours at 37°C or 15 hours at 22°C, followed by cell lysis in Tris-buffer (TBS; 25 mM Tris, 300 mM NaCl, 1 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, and lysozyme (0.1 mg/ml) using sonication at 20W for 5 min total 'on' time, using cycles of 10s on, 10s off. Soluble and insoluble fractions were separated by centrifugation and protein expression was evaluated by running both fractions on SDS-PAGE. Designs that expressed both proteins, with both present in the insoluble fractions at approximately the correct expected weight, were selected for further TEM screening. For the expression analysis of each design we ran 4 columns in the following configuration (left to right of each 4 columns): insoluble 37^{O}C, Insoluble 22^{O}C, soluble 37^{O}C, soluble 22^{O}C. The table at the top of each gel provides the data for each design: design number ID, location of the 6xHis tag position (N/C on either component **A** or **B**), the PDB 4-digit ID of the design original model, and the expected components weight. Note that in the two columns of the soluble fractions an additional band belongs to the lysozyme used for the lysis process.

Design #	I-37	I-22	S-37	S-22		Design #	I-37	I-22	S-37	S-22
Di_1	1	2	-	-		Di_21	+2	1	-	-
Di_2	+2	2	?2	?2		Di_22	?2	?2	-	1
Di_3	2	2	-	-		Di_23	+1	1	1	1
Di_4	-	-	-	-		Di_24	+2	2	-	-
Di_5	-	-	-	-		Di_25	2	1	1	2
Di_7	-	-	1	-		Di_27	1	?2	-	-
Di_9	2	1	1	1		Di_29	1	?2	1	1
Di_10	?2	1	-	-		Di_30	1	1	-	?1
Di_11	+2	2	-	-		Di_32	?2	?2	?2	?2
Di_12	?2	-	-	-		Di_33	?2	-	-	-
Di_13	2	1	1	1		Di_34	?2	-	-	?2
Di_14	2	?2	-	1		Di_37	?2	1	-	-
Di_15	2	1	1	?2		Di_40	+2	1	1	++2
Di_16	1	2	2	1		Di_41	?2	?2	-	-
Di_17_A	1	1	-	-		Di_42	?2	1	-	1
Di_17_B	1	1	-	-		Di_43	2	2	1	1
Di_18	2	1	1	1		Di_44	?2	?2	?1	1
Di_19	1	2	-	-		Di_45	2	-	-	-
Di_20	1	1	1	1		Total	24	15	4	6

Table S3. Small scale expression SDS screening

For each design we evaluate the number of bands (in a correct weight) found in each expression condition ($22^{O}C/37^{O}C$) in the soluble (S) or insoluble (I) fractions. On average, we found more cases of having two bands when expression was at 37C. Table symbols: "?" sign indicates that the bands are not definitely representing the designed protein, '+' sign indicates very pronounced bands.

Name	Protein sequence
1d2t	LALVATGNDTTTKPDLYYLKNSEAINSLALLPPPPAVGSIAFLNDQAMYEQGRLLRNTE RGKLAAEDANLSSGGVANAFSGAFGSPITEKDAPALHKLLTNMIEDAGDLATRSAKDH YMRIRPFAFYGVSTCNTTEQDKLSKNGSYPSGHTSIGWATALVLAEINPQRQNEILKRG YELGQSRVICGYHWQSDVDAARVVGSAVVATLHTNPAFQQQLQKAKAEFAQHQK
A	MGHHHHHHHGGLALVATGNDTTTKPDLYYLKNSEAINSLALLPPPPAVGSIAFLNDQAM YEQGRLLRNTERGKLAAEDANLSSGGVANAFSGAFGSPITEKDAPALHKLLTNMIEDA GDLATRSAKDHYMRIRPFAFYGVSTCNTTEQDKLSKNGSYPSGHTSIGWATALVLAEIN PQRQNEILKRGYELGQSRVICGYHWQSDVDAARVVGSAVVATLHTNPEFQAQLIKAKI EFKQHQKEL
1tk9	MSLINLVEKEWQEHQKIVQASEILKGQIAKVGELLCECLKKGGKILICGNGGSAADAQH FAAELSGRYKKERKALAGIALTTDTSALSAIGNDYGFEFVFSRQVEALGNEKDVLIGIST SGKSPNVLEALKKAKELNMLCLGLSGKGGGMMNKLCDHNLVVPSDDTARIQEMHILII HTLCQIIDESF
B	MGSLITLVELEWLEHQLIVQLSERLKGQIAKVGELLCECLKKGGKILICGNGGSAADAQ HFAAELSGRYKKERKALAGIALTTDTSALSAIGNDYGFEFVFSRQVEALGNEKDVLIGIS TSGKSPNVLEALKKAKELNMLCLGLSGKGGGMMNKLCDHNLVVPSDDTARIQEMHILI IHTLCQIIDESFLEHHHHHH

Table S4. Designs and native protein sequences

Protein sequence of **A** and **B** components and of the native protein models (1d2t → **A**, 1tk9 → **B**). Color scheme: red - Restriction sites (NcoI/XhoI) and 6xHis purification tags, blue - design mutations.

Name	Protein sequence
Di13B	MGSLITLVELEWLEHQLIVQLSERLKGQIAKVGELLCECLKKGGKILICGNGGSAA DAQHFAAELSGRYKKERKALAGIALTTDTSALSAIGNDYGFEFVFSRQVEALGNEK DVLIGISTSGKSPNVLEALKKAKELNMLCLGLSGKGGGMMNKLCDHNLVVPSDDT ARIQEMHILIIHTLCQIIDESFLEHHHHHH
Di13B1	MGSLITLVELEWLEHQLIVQLSERLKGQIAKVGELLCECLKNGGKILICGNGGSAA DAQHFAAELSGRYKKERKALAGIALTTDTSALSAIGNDYGFEFVFSRQVEALGNEG DVLIGISTSGKSPNVLEALKKAKELNMLCLGLSGKGGGKMNKLCDHNLVVPSDDT ARIQEMHILIIHTLCQIIDEAFLEHHHHHH
Di13B2	MGSLITLVELEWLEHQLIVQLSERLKGQIAKVGELLCECLKNGGKILICGNGGSAA DAQHFAAELSGRYKKERKALAGIALTTDTSALSAIGNDYGFEFVFSRQVEALGNEG DVLIGISTSGKSPNVLEALKKARELNMLCIGLSGKGGGKMNDLCDHNLVVPSDDTA RIQEMHILIIHTLCQIIDEAFLEHHHHHH
Di13B3	MGSLITLVELEWLEHQLIVQLSERLKGQIAKVGELLCRALKNGGKILICGNGGSAA DAQHFAAELSGRYKKERKALAGIALTTDTSALSAIGNDYGFEFVFSRQVEALGNEG DVLIGISTSGKSPNVLEALKKARELGMLCIGLSGKGGGKMNDLCDHCLVVPSDDTA RIQEMHILIIHTLCQIIDEAFLEHHHHHH
Di13B4	MGSLITLVELEWLEHQLIVQLSERLKGQIAKVGELLCRALKNGGKILICGNGGSAA DAQHFAAELSGRYKKERKALAGIALTTDTSALSAIGNDYGFEYVFARQVEALGNEG DVLIGISTSGKSPNVLEALKKARELGMLCIGLSGKGGGKMNDLCDHCLVVPSDDTA RIQEMHILIIHTLCQIIDEAFELHHHHHH

Table S5. Sequences of the B component stabilized versions

As discussed in the main text, stability of the independent components is critical for the usefulness of the binary system. It allows high yield of components production, simple storage, and facile post expression/ preassembly functionalization to adapt the system to different tasks. To improve the protein stability, we used the PROSS server[8]. Because at the time of use the protocol did not include symmetry design we optimized only the monomeric interactions by restricting from design all the residues in proximity to both the intra- and inter-homooligomer interfaces (the first are the interfaces forming the homooligomer, and the second are the arrays forming interfaces). The protocol allows different levels of sequence manipulations, i.e., number of introduced stabilizing mutations. The higher the number of mutations the better is the computationally expected result, however, the higher is the risk to damage the overall protein. While the original **B** component design was aggregating within a day in room temperature, versions **B2** to **B4** were all super stable in room temp and could be stored at over 2mM (see **Extended Data Figure 2** and table **S6**) for periods of months. Following the stabilization process we predominantly use the **B2** version. Color scheme: red - Restriction sites (NcoI/XhoI) and 6xHis purification tags, blue - design mutations, green - stabilization mutations.

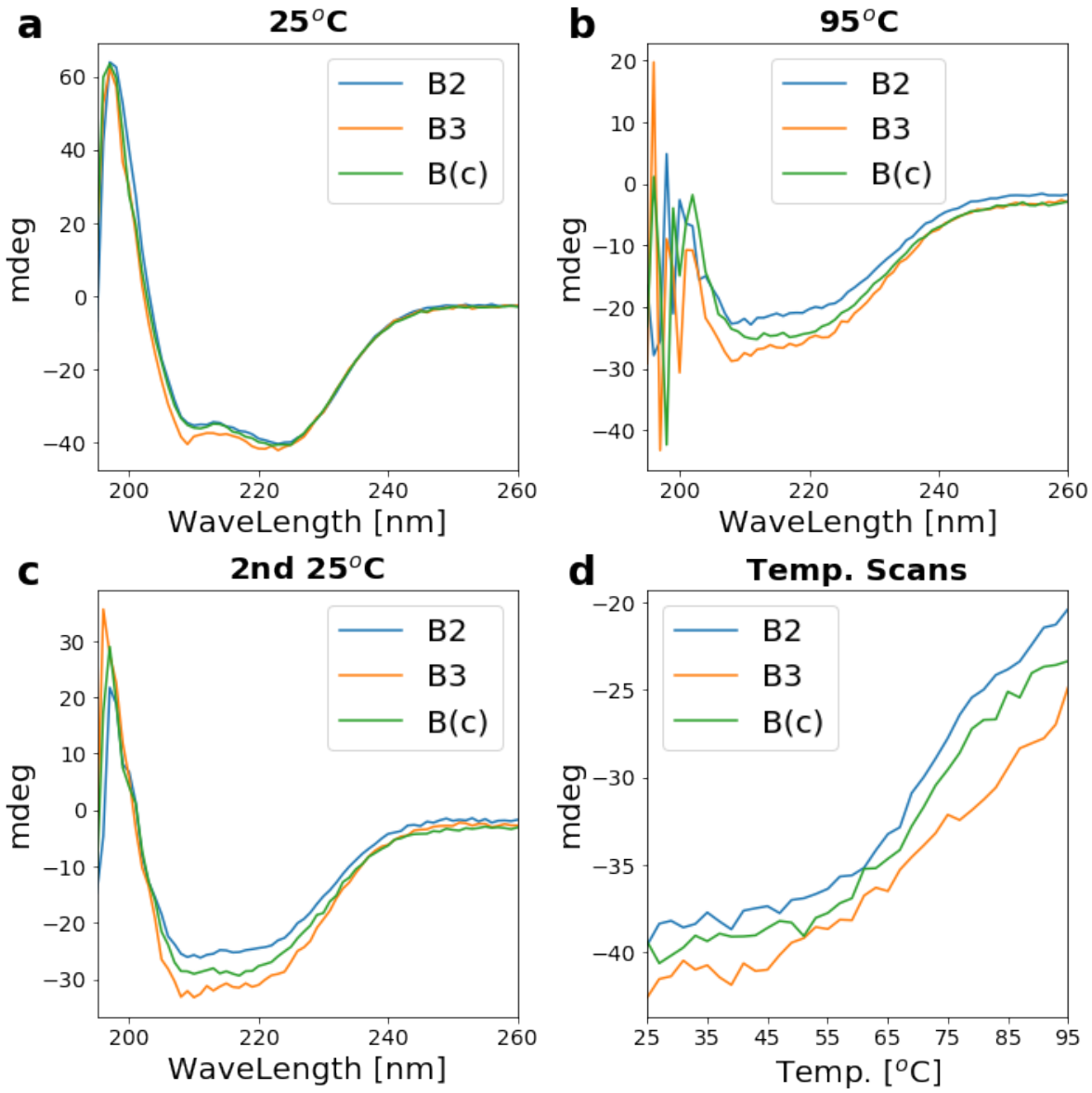


Figure S4. **Circular dichroism analysis of the stabilized components B**

CD spectra at wavelengths scanes (260-195nm) (**a**) at 25°C, (**b**) 95°C , and (**c**) 25°C after cooling are plotted as raw data (millidegrees) for components **B2** at 0.35[mg/ml] (blue), **B3** at 0.30[mg/ml] (orange), and **B(c)** (the cyclic version of **B** discussed later in **Extended Data Figure 8** and **S6**) at 0.29[mg/ml] (green). (**d**) CD Temperature scan from 25°C to 95°C measured at 222nm. Results show that the mutated **B** constructs are stable ambient conditions and could sustain higher temperature before partial unfolding initiates. Experimental method: Far-ultraviolet Circular Dichroism (CD) measurements were carried out with an AVIV spectrometer, model 420. Wavelength scans were measured from 260 to 195 nm at temperatures of 25 and 95 °C. Temperature melts monitored absorption signal at 222 nm in steps of 2 °C/min and 30 s of equilibration time. For wavelength scans and temperature melts a protein solution in PBS buffer (pH 7.4) of concentration 0.2-0.4 mg/ml was used in a 1 mm path-length cuvette.

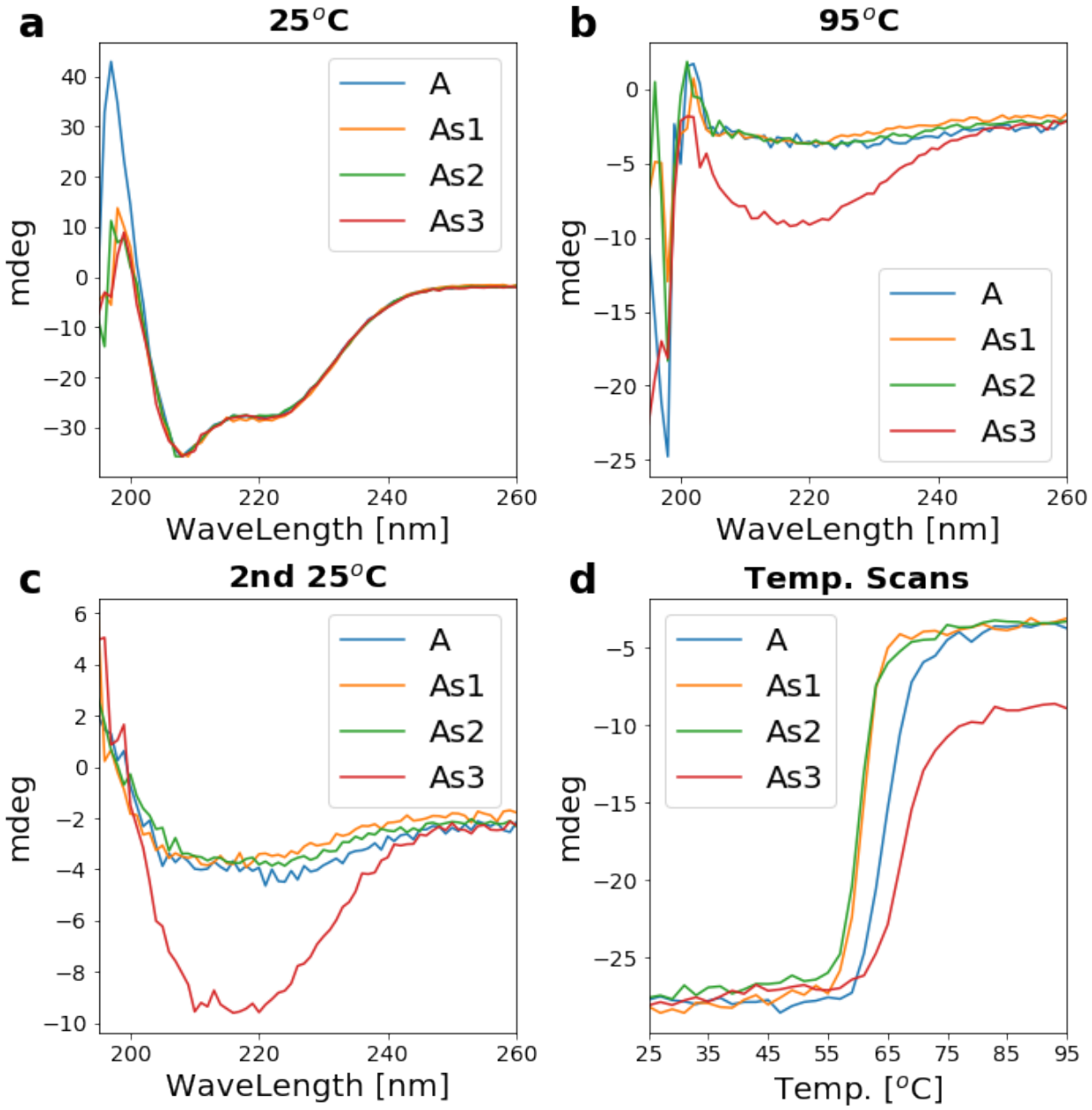


Figure S5. **Circular dichroism analysis of the stabilized components A**

Panels **a-d** are as described in Fig. **S4**. Unlike component **B** required stabilization, component **A** was already stable in ambient conditions and expressing well but we were interested to explore the process ability to further improve components stability at higher temperatures, increased high temperature stability and solubility would have advantages for annealing processes, assembly under various conditions, and storage in non-optimal conditions. As in figure **S4**, **As1** to **As3** are the redesigned constructs with an increasing number of mutations. In this case the protocol did not improve protein stability or thermo stability except the case of construct **As3**. As shown in table **S6** all versions behave approximately similar and exhibit high solubility at room temp., for that reason we kept using the originally designed **A** component unless stated otherwise.

	Di13A	Di13As1	Di13As3	Di13B	Di13B2	Di13Bc_mScarlet
[mg/ml]	46	69	57	X	47	69
[mM]	1.7	2.6	2.1	X	2.2	1.0

Table S6. Designed components pre- and post-stabilization individual solubility

Designs solubility measured using NanoDrop 8000 Spectrophotometer at room temperature. Experimental protocol: following Ni-affinity-chromatography and size-exclusion chromatography (Superose 6 10/300 GL SEC column) an eluted volume of 2ml was collected and concentrated in two step, first to ~400µL and than further to the volumes of 100µL to 200µL. Between each concentration step the collected solutions were centrifuged at 10k for 10min and visually validated for aggregations (or the lack of aggregations). A repeated absorption measurement was performed a week later while solutions were kept on the bench in room temperature. We note that the measured concentrations are close to the nanodrop detection limit (100mg/ml) and interpolation of volume ratios for each tube would bring some of the construct concentrations up to the range of 8 mM/180 mg/ml. As shown in **Extended Data Figure 2**, this system is unique as the individually soluble inter-components distance is much shorter than the obtained inter-component distance upon assembly (the complementary component mediate assembly position each component at higher distance) or solidification.

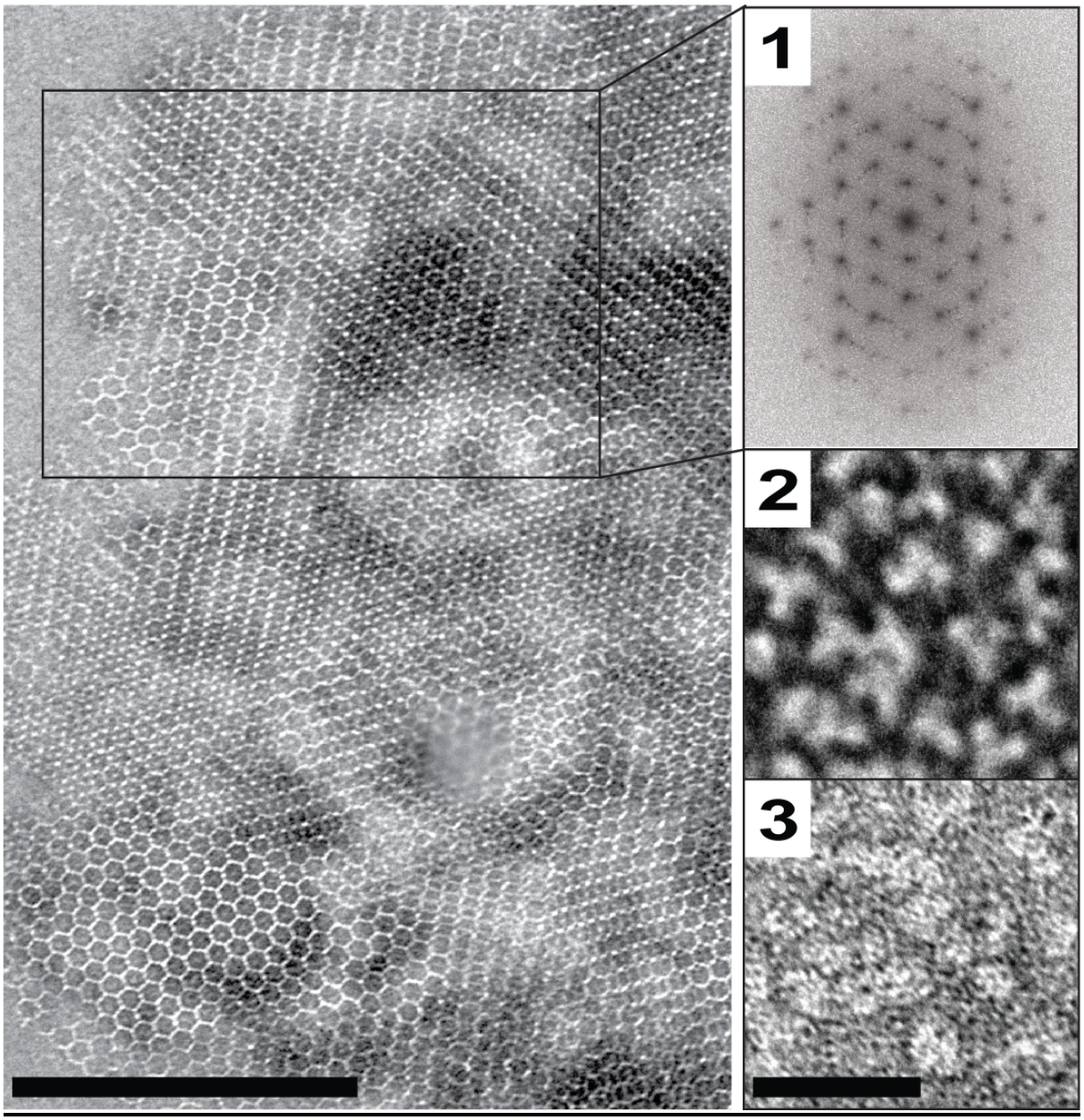


Figure S6. Long range order of assembled arrays assessed by TEM

Negative stain TEM of micron scale arrays overnight assembly from a mixture of 5 μM components concentration in TBS supplemented with 500 mM imidazole. Note that this picture is the same as the one in Fig. 2c, reproduced here for convenience. Insets: 1) FFT of a selected region (black rectangle), 2-3) TEM of only the A and B components, respectively. Note that left panel is identical to Figure 2c, reproduced here for convenience. Scale bars: 500 nm (left panel); 20 nm (2,3).

Peak	Q spacing ($Å^{-1}$)	Bragg Spacing(Å)	Predicted Bragg Spacing from P6 symmetry with spacing 303Å
1	0.024	262	262
2	0.041	153	151.5
3	0.048	131	131
4	0.063	99.7	99.18,87.4
5	0.0825	76.1	76,73

Table S7. SAXS scattering analysis.

Peaks values (at q spacings) derived from the scattering curves of samples prepared by mixing the **A** and **B** components at various equimolar concentrations (0.5, 2, 5 , and 10 μM) as shown in Fig. **2.e** and **Extended Data Figure 4d**. Bragg spacing (middle column) are calculated from the samples q spacing and are in good agreement with Bragg spacing calculated from applying the ASUs to a p6 lattice with 303 Å spacing using CCP4.[9]

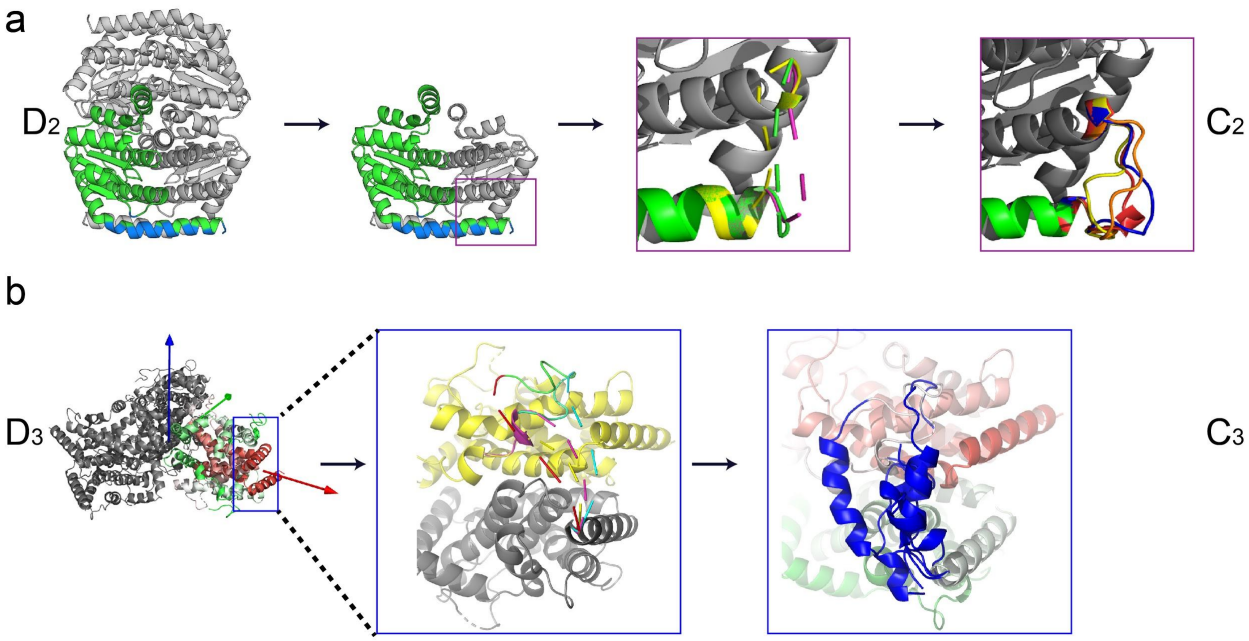


Figure S7. Designed cyclic pseudo-dihedral building blocks

In **Extended Data Figure 1** we described the inherent pros of a dihedral symmetric building blocks for the construction of 2D, planar, assemblies owing to their pair of in-plain rotational symmetry axes. In different scenarios, however, the same pros are found to be a disadvantage. For example, attempting a stepwise assembly over soft substrates such as cell membranes (see the following in **Extended Data Figure 8-11**) where one of the components is initially used as an anchor, results in a failure. We presumed the reason lay in the ligand spatial distribution around the dihedral building blocks, facing both up and down (relative to the plane geometry, see illustration in **Extended Data Figure 8a, b**, where a GFP, in green, is used to bind to the GBP nanobodies displayed on the cell membranes) such that when components bind to the cell membrane, either their orientation is such that the array assembly interfaces are blocked (**Extended Data Figure 8b** leftmost panel, array assembly interfaces are colored in purple and propagation direction is indicated with purple arrows), or the entire component becomes buried within or wrapped by the membrane (**Extended Data Figure 8b** second from the left panel), thereby blocking array assembly. A useful geometry for an anchor unit in such a configuration would be one with vertical inhomogeneous binding sites, a feature inherent to cyclic components (see **Extended Data Figure 8a** right panel for illustration of cyclic components binding sites and **Extended Data Figure 8b** two rightmost panels for the component binding orientation in reference to the lipid substrate). In order to benefit from both geometries and diversify component functionality we chose to redesign the dihedral building blocks to be cyclic pseudo-dihedral ones. For example, every pair of chains in the **B** components (originally a D_2 tetramer) (**a**) are combined to a single chain, resulting in a C_2 dimer "almost", pseudo, identical to the original homooligomer. The C_2pD_2 array interfaces remain unchanged while ligand distribution becomes vertically inhomogeneous, therefore components could now function equally well as both anchor units and planar array building blocks (see figure **4.b** and **Extended Data Figure 8d-f** for array formation on cell membranes and controls)

The computational workflow to alter the building blocks' symmetry from dihedral to cyclic pseudo-dihedral (Dx→Cx) includes a number of steps. We first use pyrosetta[10] to generate the dihedral homooligomer model and choose a pair of monomers such that their C- and N-terminus are as close to each other as possible (a simple case is shown for the **B** components in **(a)** where the C- and N-terminus are adjacent, this is not always the case as shown for component **A** in (**b**)). We then generate a set of blueprints of linkers between a set of positions near the C-terminus of one monomer and a set of positions near the N-terminus of the second components, i.e., we truncate either or both components and suggest linker lengths and secondary structure preferences. We employ Rosetta Remodel[11] to generate fragments that would create ideal linkers (see illustration in (**a**) and (**b**) for different cut sites and fragments generated). We chose to

test a number of linkers with either predicted rigid secondary structure or a flexible one. To generate the full constructs, we cloned the linkers between two different monomers, we chose the best two stable versions of the **A** and **B** components which were generated at the stabilization process (Fig. **S4**, **S5**). Tables **S6** and **S7** show a list of generated linkers. We then expressed the proteins, now referred to as **A(c)** and **B(c)**, verified monomeric weight using SDS-page, homooligomeric weight using SEC-MALS and structural functionality, the ability to form similar hexagonal 2D arrays using negative stain TEM (see **Extended Data Figure 8d, f** for **A+B(c)**, **A(c)+B** and **A(c)+B(c)**). The Final step included genetic fusions of functional groups (GFP or SpyCatcher that could then be peptide fused to spy-tagged ligands) to the cyclic component's N/C-termini or exposed loops. This allows a versatile set of materials to co-assemble in a stepwise fashion, i.e., first cell priming by incubation with one of the components (that need to be cyclic and to carry the binding site), followed by removing the unbound components from the media, and finally introducing the second component (which can be either dihedral or cyclic, functionalized or not) to induce array assembly directly on the cell membrane, allowing a controllable (timely, spatially, uniformly, and receptor or signal specific) assembly process of functionally combinatorial complex structures on cell membranes (see Fig. **4** and **Extended Data Figure 9-11**).

	Construct name	B2 - linker - B4
1	Di13_B2L1B4	IDEAF**GGGSGGS**SLITL
2	Di13_B2L2B4	IDEAF**GGGKDRNG**GSLIT
3	Di13_B2L3B4	IDEAF**TGDAGET**SLITL
4	Di13_B2L4B4	IDEAF**GGETSSKQD**LITLV

Table S8. B component desymmetrization linkers list

Linkers inserted (blue) between the C-terminal of one monomer (green) and the N-terminal of another monomer (red). Note the N-terminal of the second monomer was trimmed in some of the cases. Construct number 2 was the best behaving one and verified under TEM to form the expected hexagonal geometry with the dihedral **A** components with or without the addition of GFP/mcherry labels fused at the C-terminus (see **Extended Data Figure 8d**).

	Construct name	A - linker - As3
1	Di13A_S1L12_As3_n0	KQHQK**FRQQPPPPQQSG**GLALVATGNDATTKPDLYYLKNSEAINSL
2	Di13A_S1L14_As3_n1	KQHQK**DKTPEDSTRSEYKG**GLALVATGNDATTKPDLYYLKNSEAIN
3	Di13A_S8L13_As3_n2	KQHQK**SEPQEVSETQEVP**GNDATTKPDLYYLKNSEAINSLALLPPP
4	Di13A_S18L12_As3_n3	KQHQK**ESTKSWPPTSPA**YYLKNSEAINSLALLPPPPAVGSIAFLND
5	Di13A_S14L10_As3_n4	KQHQK**QQQEERQTDK**KPDLYYLKNSEAINSLALLPPPPAVGSIAFL
6	Di13A_S18L10_As3_n5	KQHQK**DSESSGEPGA**YYLKNSEAINSLALLPPPPAVGSIAFLNDQA
7	Di13A_S14L13_As3_n6	KQHQK**SRDDDKGAKHKPK**KPDLYYLKNSEAINSLALLPPPPAVGSI
8	Di13A_S8L18_As3_n7	KQHQK**SDSKEEEKKKSSDNSSTP**GNDATTKPDLYYLKNSEAINSLA
9	Di13A_S1L18_As3_n8	KQHQK**KPDERSSSKKEEDKKDRG**GLALVATGNDATTKPDLYYLKNS
10	Di13A_S14L11_As3_n9	KQHQK**GSGSGSGSGSG**KPDLYYLKNSEAINSLALLPPPPAVGSIAF
11	Di13A_S8L13_As3_n10	KQHQK**GSGSGSGSGSGSG**GNDATTKPDLYYLKNSEAINSLALLPPP
12	Di13A_S1L14_As3_n11	KQHQK**GSGSGSGSGSGSGS**GLALVATGNDATTKPDLYYLKNSEAIN

Table S9. A component desymmetrization linkers list

Linker inserted (Blue letters) between the C-terminal of one monomer (green) and various truncations of the N-terminal of a monomer version As3 (red). Constructs name nomenclature Di13**A** for the first monomer, SX: X is the number of residues truncated of the second monomer N-terminus, LX: X is the linker length (residues number), and As3 - the stabilized monomer version used as the second monomer. Construct number 3 was best behaving and verified under TEM to form the expected hexagonal geometry both when mixed with dihedral **B** or cyclic **B** components (see **Extended Data Figure 8f**).

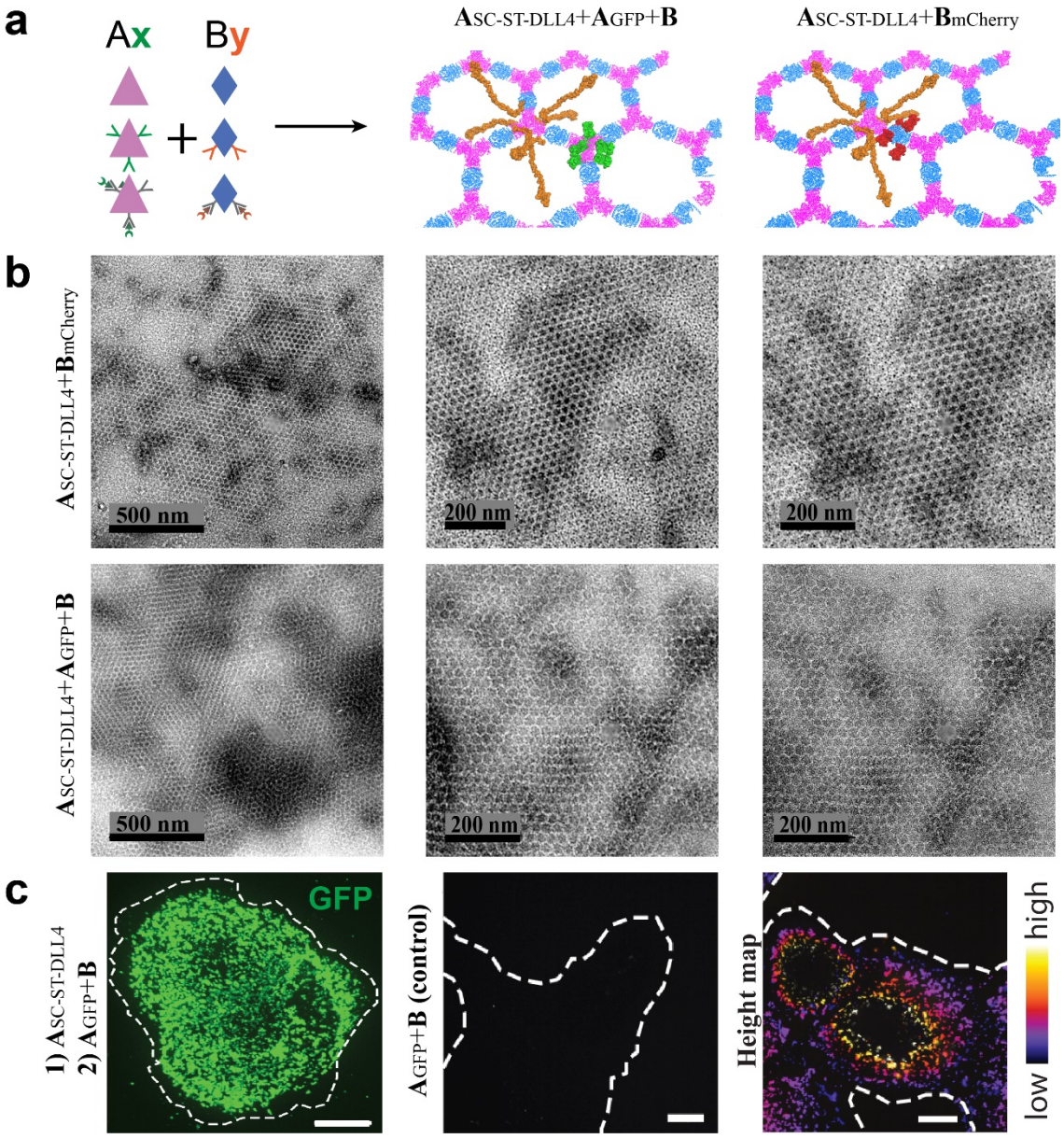


Figure S8. Characterization of multi-functional pre-assembled arrays and specific binding to cells expressing Notch1 receptors

(**a**) Schematics for multi-functional array formation by mixing array components that have been functionalized in various ways, e.g. genetic fusions such as **AGFP** and **BmCherry**, and peptide fusions such as SpyCatcher (SC) - SpyTag (ST) conjugates (e.g. ASC-ST-DLL4). For the formation of arrays, the dihedral versions of **A** and **B** components are mixed in equimolar concentration. For example, to generate **ASC-ST-DLL4 + AGFP + B** arrays, components are mixed in molar ratios of (4:1:5).(**b**) Negative stain TEM of **ASC-ST-DLL4 + BmCherry** (upper panels) and **ASC-ST-DLL4 + AGFP + B** with molar ratios 4:1:5 (lower panel).(**c**) Specific array binding by addition of **AGFP** and **B** following addition of **ASC-ST-DLL4** to cells expressing Notch1 receptors (left panel). No binding is observed in the absence of **ASC-ST-DLL4** (middle panel). Depth-encoded z-stack (right panel). Results for pre-assembled **ASC-ST-DLL4 + BmCherry** (panel **b**) were used for the experiment shown in figure **S9**. Scale bars: (b) indicated in the figure, (c) 10 µm.

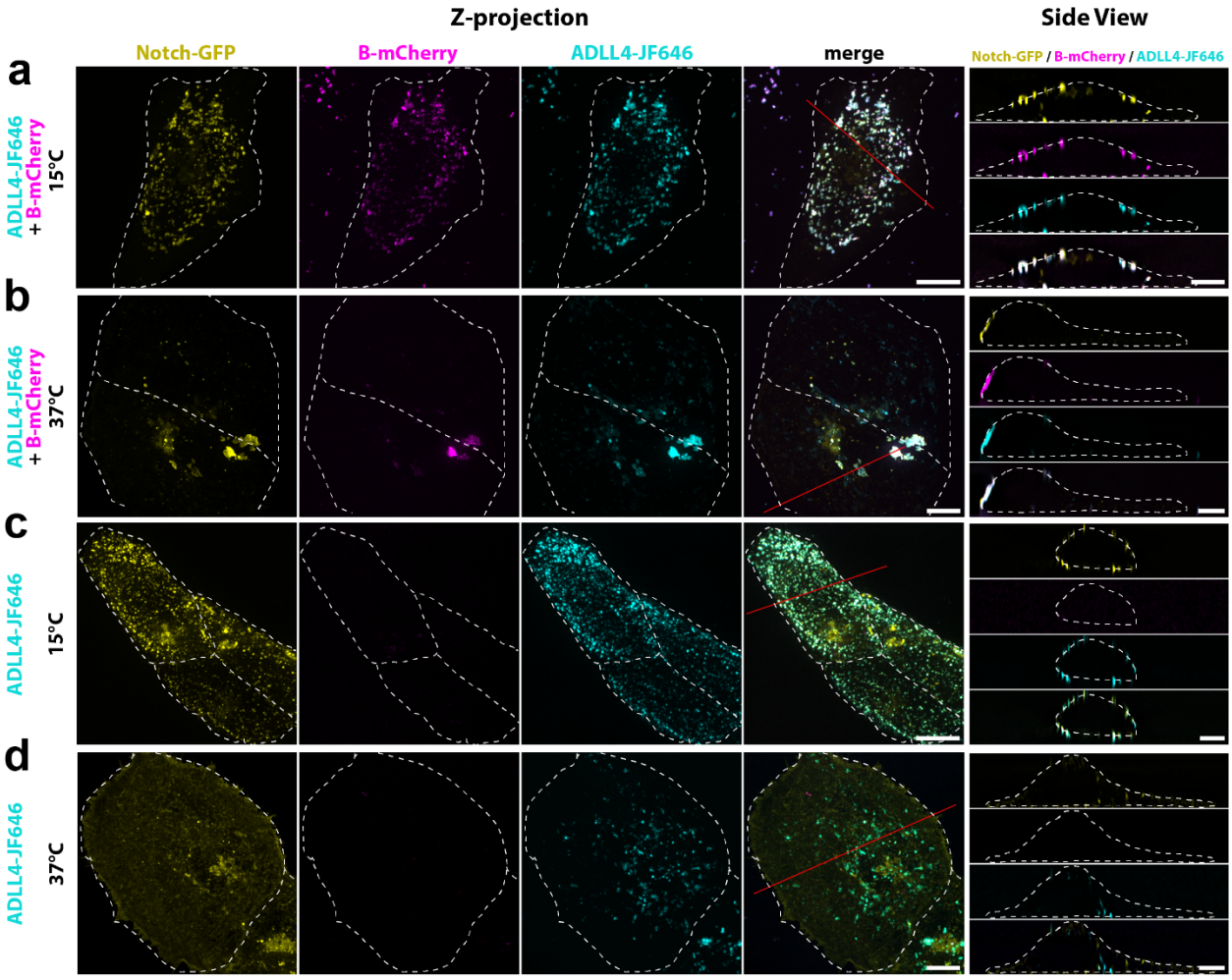


Figure S9. **Arrays functionalized with DLL4 recruit Notch in U2OS cells**

BmCherry/ADLL4-JF646 arrays were formed by mixing 5 µM of each component at 4°C for >18 hr in 25 mM Tris, pH 8.0, 150 mM NaCl, 500 mM imidazole. 0.5 µM arrays (or **ADLL4-JF646** alone) were then incubated with U2OS cells expressing Notch1-EGFP for 15min at the indicated temperature (4°C or 37°C), then washed in PBS and incubated in cell culture medium for 60min at the indicated temperature. Cells were then quickly imaged by spinning disk confocal microscopy at either 37°C (panels b,d) or 15°C (panels a,c). Images correspond to maximum intensity z-projection (left panels) or xz optical slice along the red line after deconvolution (right panels). Dash white lines correspond to cell outlines. (**a, c**) **BmCherry/ADLL4-JF646** arrays or **A ADLL4-JF646** alone binds to cells through a specific DLL4-Notch interaction (high colocalization **ADLL4-JF646** / Notch-GFP**)** and remain on the cell membrane (side view) due to the absence of endocytosis at this restrictive temperature). (**b**) Arrays bound to the cell as in **A** and cluster further into large rafts that remain on the cell membrane upon incubation at 37°C. (**d**) Notch1-EGFP remains on the cell membrane while **ADLL4** dissociates and is internalized. Scale bars: left panels: 10 µm; right panels: 5 µm.

Spy tagged F-domain (**st-fD**)	MAHIVMVDAYKPTKAELASEKPFRDCADVYQAGFNKSGIYTIYINNMPEPKKVFCNMDV NGGGWTVIQHREDGSLDFQRGWKEYKMGFGNPSGEYWLGNEFIFAITSQRQYMLRIELM DWEGNRAYSQYDRFHIGNEKQNYRLYLKGHTGTAGKQSSLILHGADFSTKDADNDNCM CKCALMLTGGWWFDACGPSNLNGMFYTAGQNHGKLNGIKWHYFKGPSYSLRSTTMMI RPLDF
DLL4(N-EFG5)-SpyTag (**st-DLL4**)	MAAASRSASGWALLLLVALWQQRAAGSGVFQLQLQEFINERGVLASGRPCEPGCRTFFR VCLKHFQAVVSPGPCTFGTVSTPVLGTNSFAVRDDSSGGGRNPLQLPFNFTWPGTFSLIIE AWHAPGDDLRPEALPPDALISKIAIQGSLAVGQNWLLDEQTSTLTRLRYSYRVICSDNYY GDNCSRLCKKRNDHFGHYVCQPDGNLSCLPGWTGEYCQQPICLSGCHEQNGYCSKPAEC LCRPGWQGRLCNECIPHNGCRHGTCSTPWQCTCDEGWGGLFCDQDLNYCTHHSPCKNG ATCSNSGQRSYTCTCRPGYTGVDCELELSECDSNPCRNGGSCKDQEDGYHCLCPPGYYG LHCEHSTLSCADSPCFNGGSCRERNQGANYACECPPNFTGSNCEKKGSAHIVMVDAYKP TKGSGHHHHHH
Asc	MGHHHHHHSGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKELAGATMELR DSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGKAT KGDAHIGGSGGSGGNDTTTKPDLYYLKNSEAINSLALLPPPPAVGSIAFLNDQAMYEQGR LLRNTERGKLAAEDANLSSGGVANAFSGAFGSPITEKDAPALHKLLTNMIEDAGDLATRS AKDHYMRIRPFAFYGVSTCNTTEQDKLSKNGSYPSGHTSIGWATALVLAEINPQRQNEIL KRGYELGQSRVICGYHWQSDVDAARVVGSAVVATLHTNPEFQAQLIKAKIEFKQHQK

Table S10. A-SpyCatcher/spyTag-fDomain and spyTag-DLL4 sequences for arrays-cell receptors binding

Asc: A component (black) spy Catcher (orange) and a flexible linker and a His-tag (red).
st-fD: fDomain (black), spytag (orange). **st-DLL4**: DLL4 (black), spyTag (orange).

Supplementary Figure 10 - Uncropped western blots with size marker indications

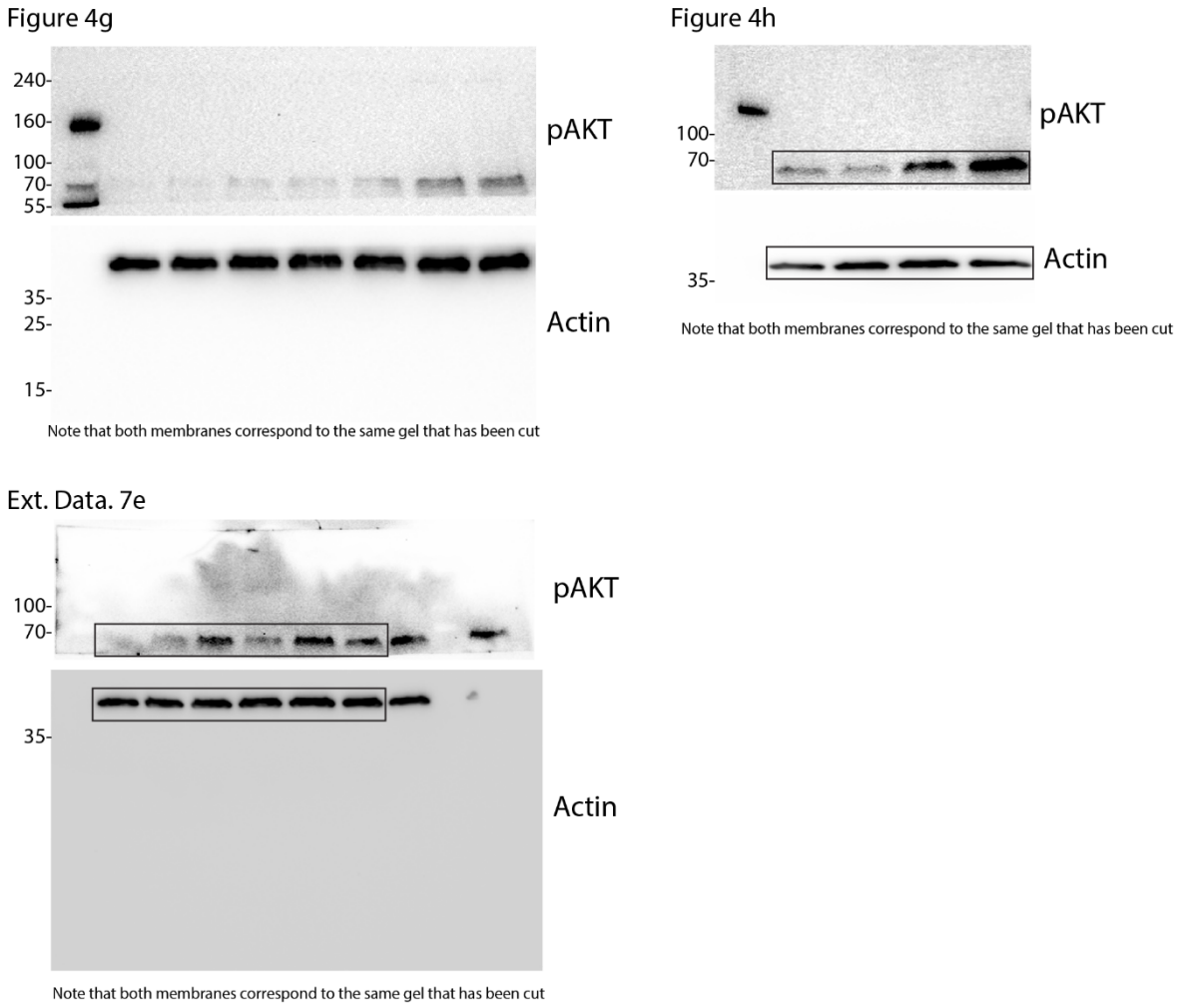


Figure S10. Gel source data for the various Western Blots presented in this paper

Supplementary Movie legends

Supplementary Movie 1. Design strategy PyMOL illustration: dock, design, and propagation

PyMOL illustration demonstrating the docking process, benefits of dihedral components for planar assemblies, and the propagation of ordered structure.

Supplementary Movie 2. Instantaneous gelation upon components mixing

Mixture of 10uL dihedral **A** component at 2mM into 10uL of 1mM B(c)mScarlet (note mixture ratio of 1:2 due to symmetry differences). Upon addition of the second component the mix goes through immediate gelation which clogs the further pipetting.

Supplementary Movie 3. Clustering of intracellular mScarlet constructs by preformed arrays

NIH/3T3 cells expressing GBP-TM-mScarlet were incubated with 10µl/mL of preformed **A**$_{GFP}$+**B** arrays and were imaged immediately by spinning disk confocal microscopy. Upon landing onto the cells, **A**$_{GFP}$+**B** arrays quickly cluster the GBP-TM-mScarlet construct. This movie corresponds to figure **3.b-c**. Scale bar: 6µm.

Supplementary Movie 4. 3D rendering of cell incubated with preformed arrays

NIH/3T3 cells expressing GBP-TM-mScarlet were incubated with 10µl/mL of preformed **A**$_{GFP}$+**B** arrays and were imaged immediately by spinning disk confocal microscopy after 30 minutes. 3D stacks were then processed for 3D reconstruction. This movie corresponds to figure **3.d** and figure **Extended Data Figure 6.d-e**.

Supplementary Movie 5. Stability of receptor clustering assessed by FRAP

GBP-TM-mScarlet expressing NIH/3T3 cells were incubated with the **A**$_{GFP}$+**B** arrays for 1 hour at 37ºC, then the mScarlet signal was bleached and its fluorescence recovery monitored by spinning disk confocal microscopy. Left panel, quantification (see methods). mScarlet signal does not recover, suggesting that arrays cluster stably the GBP-TM-mScarlet construct. This movie corresponds to Figure **Extended Data Figure 6.e-f**. Scale bar: 6µm.

Supplementary Movie 6. Growth of arrays onto cells

NIH/3T3 cells expressing GBP-TM-mScarlet were incubated with 1µM **B**$_{GFP}$, rinsed in PBS, then 0.2µM unlabelled **A** was added and cells were imaged by spinning disk confocal microscopy. Upon addition of **A**, numerous foci positive for extracellular **B**$_{GFP}$ and intracellular mScarlet appear and subsequently fuse with each other. This movie corresponds to Fig. **4a-c**. Scale bar: 12µm.

Supplementary References:

1. Hsia, Y. *et al.* Design of a hyperstable 60-subunit protein icosahedron. *Nature* **535**, 136–139 (2016).
2. Tan, R., Zhu, H., Cao, C. & Chen, O. Multi-component superstructures self-assembled from nanocrystal building blocks. *Nanoscale* **8**, 9944–9961 (2016).
3. Ovesný, M., Křížek, P., Borkovec, J., Svindrych, Z. & Hagen, G. M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinforma. Oxf. Engl.* **30**, 2389–2390 (2014).
4. Picco, A., Mund, M., Ries, J., Nédélec, F. & Kaksonen, M. Visualizing the functional architecture of the endocytic machinery. *eLife* **4**, e04535 (2015).
5. Hoover, D. M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002).
6. Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).
7. Gaspar, P., Moura, G., Santos, M. A. S. & Oliveira, J. L. mRNA secondary structure optimization using a correlated stem–loop prediction. *Nucleic Acids Res.* **41**, e73–e73 (2013).
8. Goldenzweig, A. *et al.* Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **63**, 337–346 (2016).
9. CCP4 Program Suite. http://www.ccp4.ac.uk/html/unique.html.
10. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
11. Huang, P.-S. *et al.* RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLOS ONE* **6**, e24109 (2011).