Supplementary Information

Barcoded DNA origami structures for multiplexed optimization and enrichment of DNA-based protein-binding cavities

Ali Aghebat Rafat[+], Sandra Sagredo[+], Melissa Thalhammer, and Friedrich C. Simmel*

Physics Department E14 and ZNN, Technical University Munich, Garching, Germany

Key words: DNA nanotechnology, aptamer, multivalent binding, scaffolding, *in vitro* selection

This PDF file includes:

1. Supplementary Methods

DNA-Protein structural modelling

The ternary complex of thrombin and the two aptamers inside the origami cavity was modelled using the 3D molecular visualization software UCSF Chimera[1]. The crystal structure of the ternary complex, in which thrombin is bound to HD22 and a modified TBA1 aptamer (TerΔT12) (PDB ID :5EW2) and the protein bound to TBA1 (PDB ID: 1HUT) were used as a template[2,3]. B-form helices were generated in model.it®Server[4]. Coordinates for modelling the position of the spacers in the cavity were taken from the PDB model of the origami design, which was generated from the corresponding cadnano file[5]. The aptamers were positioned at the ends of the thymidine linkers and rotated to localize the protein between the two helices. The linkage between the aptamers and the unpaired thymidines at the ends of the spacers is expected to be flexible. Subsequently, the torsions of the eight thymidines of each aptamer were modeled and their energy minimized. In case A, in which the spacer stem was a 26 bp duplex, we were not able to position the protein without steric hindrance. In order to create the optimal spacing for the protein, the TBA1 spacer had to be tilted by ≈20°. In case B, where the spacer stem is 20 bp long, the optimal position of the protein sandwiched between the aptamers could be modelled between the spacers without steric clashes.

2 x 2 Array formation and barcodes

Fractal assembly enables the creation of finite DNA origami arrays made of unique and independent structures that fold in separate pots[6]. Since the structures are folded separately, each of them can be equipped with different features that are uniquely addressable.

In order to be able to distinguish between different single structures in AFM imaging, we established a barcoding system, which uses dumbbell hairpin modified staples[7] to generate AFM height contrast between the structure and the barcode patterns. We considered creating higher order fractal structures (4 x 4, 8 x 8 array) with less complexity in barcoding (but where structures could be also identified via their position within the arrays), or to use only single structures or 2 x 2 arrays with a more complicated barcoding scheme. Due to the fact that the yield of fractally assembled structures drops dramatically for higher order assemblies, whereas in our hands the yield of 2 x 2 arrays was reasonably good, we

decided to stick to 2 x 2 arrays and to apply a barcoding scheme which not only enabled us to identify each single structure within a 2 x 2 array, but also to distinguish between different 2 x 2 arrays. Thus our barcoding approach represents a balance between barcoding complexity and assembly yield.

In Supplementary Figure 1a, we show the connection rules for 2 x 2 array formation and the barcoding scheme for 2 x 2 array number 1. Three different types of staple modifications are used at the edges of the tiles: truncation (brown), extension (green) and double hairpin passivation (black), which enable the structures to connect and create 2 x 2 arrays. For truncation and extension 8 of 11 staples at the edge of an DNA origami structure were modified. Colors represent the type of modification for the corresponding staple coordinates.

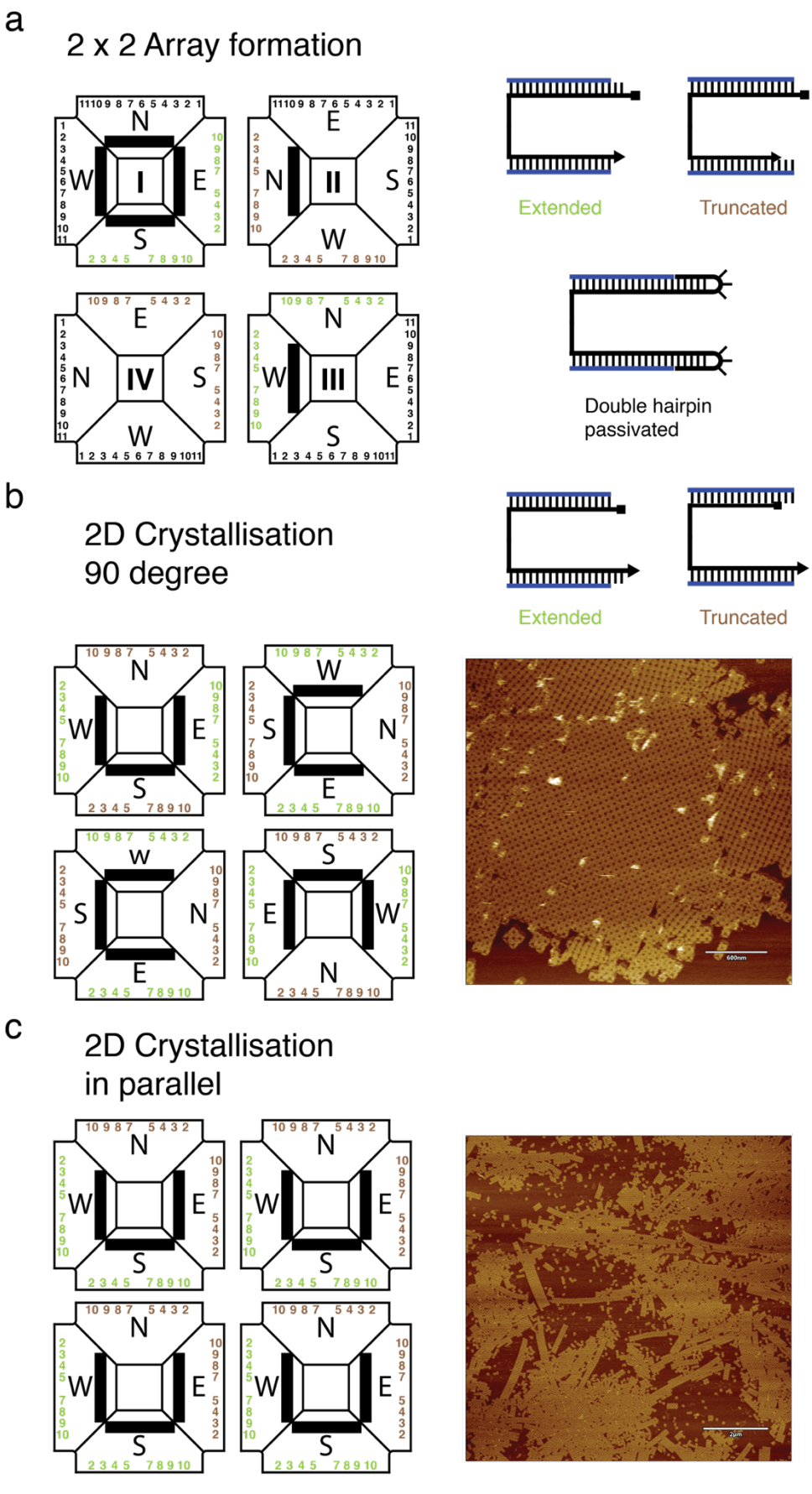
Within a 2 x 2 array, the four origami structures are numbered clockwise using Greek numbers from I to IV. We implemented barcoding patterns that resemble digital numbers. In our experiments we only needed a maximum of eight 2 x 2 arrays in a one-pot experiment, but in principle, our barcoding scheme could be used to label several tens of distinct 2 x 2 arrays in one experiment.

Our barcodes were designed to follow a sufficiently asymmetric pattern in order to be able to distinguish arrays and single structures even when they landed on the mica substrate upside down or had defects. In consideration of these points, we used a zero-shape barcode on structure number I and no barcode at all on structure number IV - this creates the necessary asymmetry for AFM imaging when using only one-digit numbers. Also when using two-digit numbers, enough asymmetry can be implemented to allow to distinguish different structures. Two-digit numbers with too high a degree of symmetry – such as 11, 22, 33 ….99 or 89 and 98 and etc. – cannot be used unless additional asymmetry is introduced elsewhere on the origami tiles.

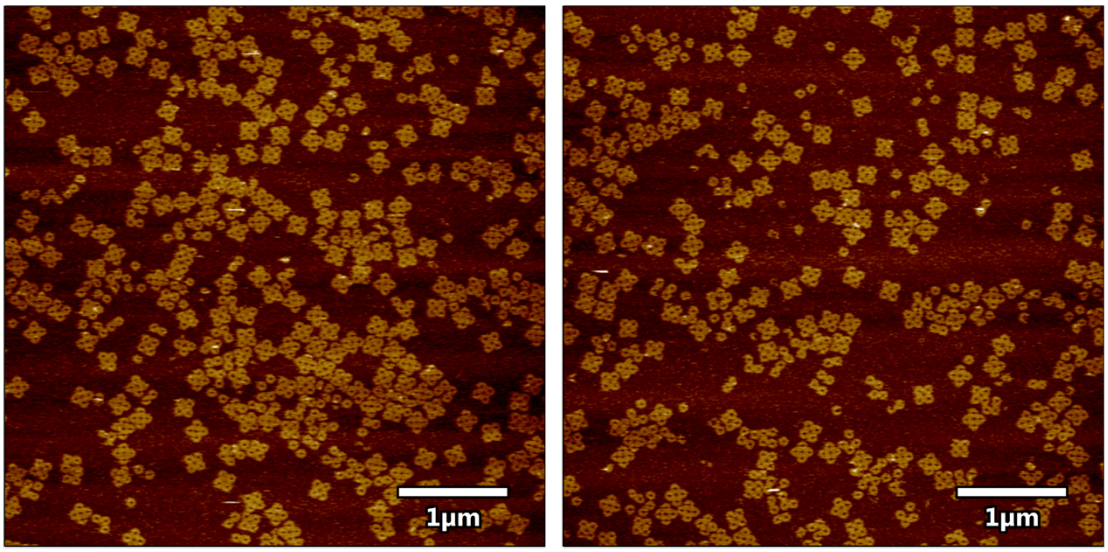
Here we used each trapezoid for one part of a digit to make sure that the patterns can be resolved even when the imaging quality and resolution is not very good. However, in principle one could also create a complete digit on each of the four trapezoids that would enable more barcoding possibilities, but identification by AFM imaging would then be more challenging.

In our experiments, we used the barcoding system to uniquely address each single structure within a 2 x 2 array and to specify which modifications were made in that structure (for example, the configuration of aptamers used or different flexibilities) when it was folded in a separate pot.

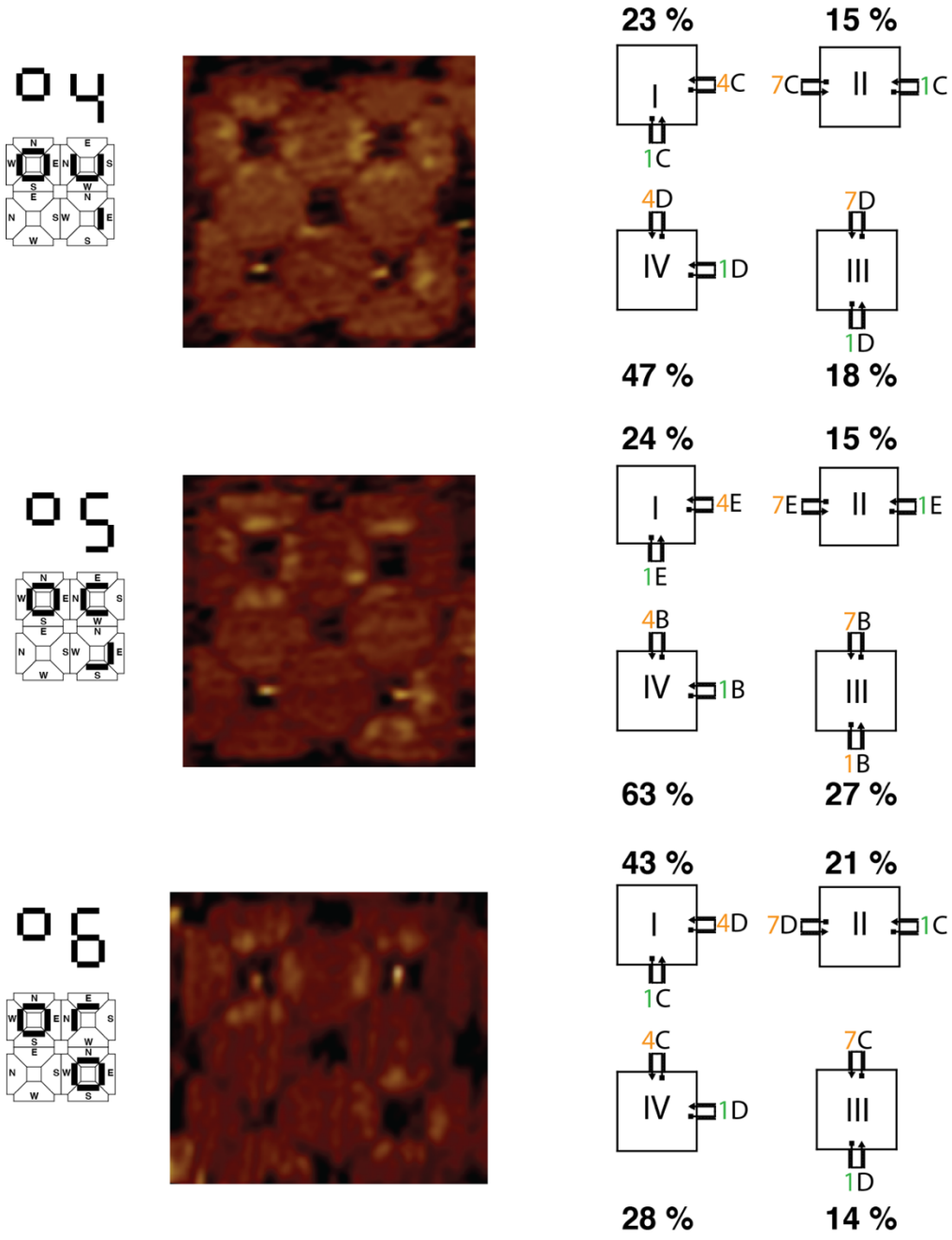
2. Supplementary Figures



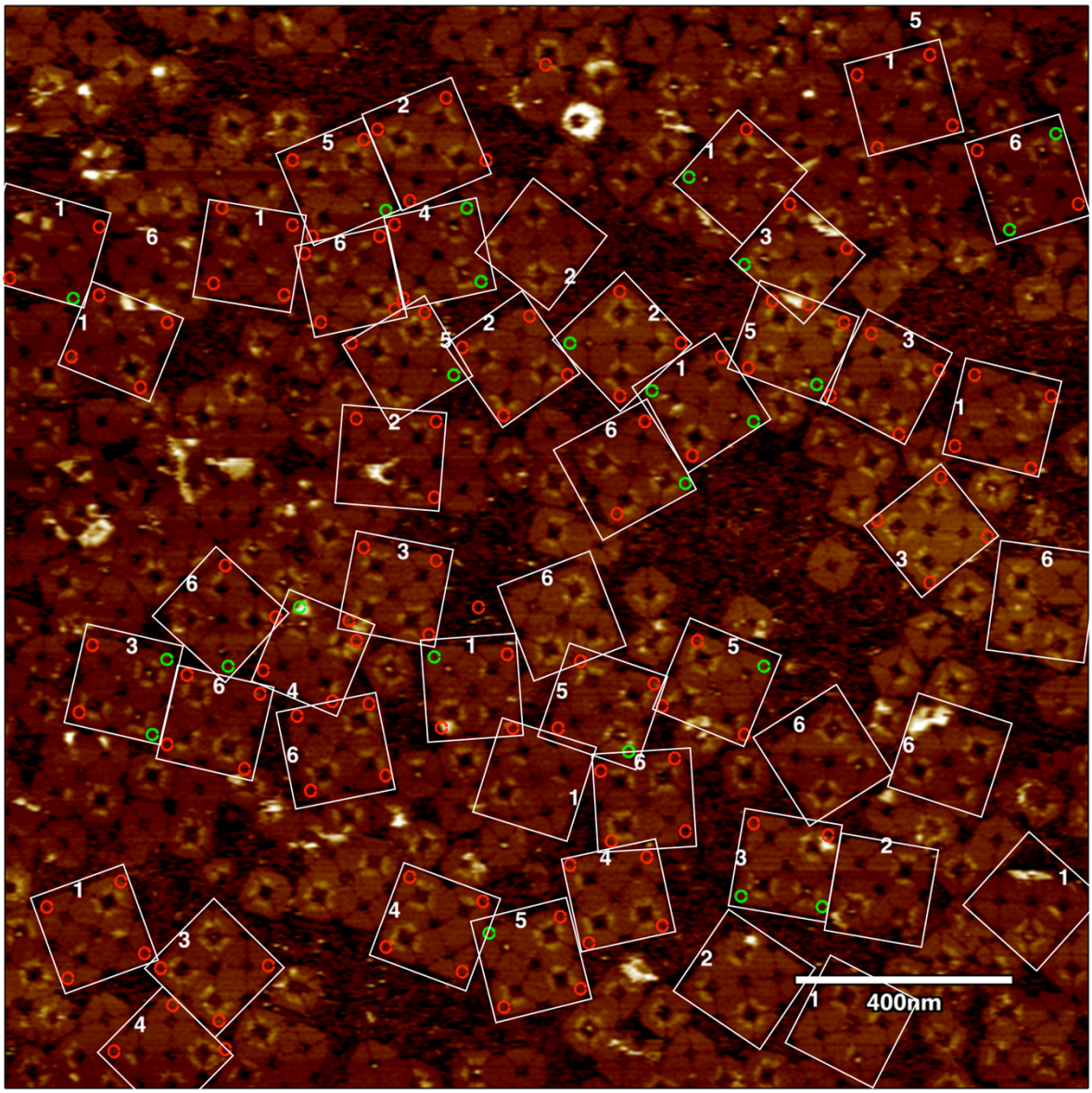
Supplementary Figure 1. a, 2 x 2 array scheme which shows the connection rules between different sides. 1 to 11 are the staple coordinates which are at the edge of the DNA origami structure and their modification enable bounded and unbounded arrays. Color codes in addition to the staple coordinates show on each side of the square (**N**orth, **S**outh, **E**ast, **W**est) which staples are used and what kind of modification (truncation, extension and double hairpin passivation) they have. **b**, Scheme and AFM image of the 2D crystallization with 90° rotation (**N-E**, **W-S**). Color codes and staple coordinates show the type of staple modification and the staples which are used in an edge respectively. **c,** Scheme and AFM image of the 2D crystallization by using parallel sides (**N-S**, **E-W**). As can be seen in the AFM image, elongated crystals were formed using this type of connection rule, potentially due to the build-up of internal strain - a similar effect has been reported by other groups[6]. Color codes and staple coordinates show the type of staple modification and the staples which are used in an edge respectively.



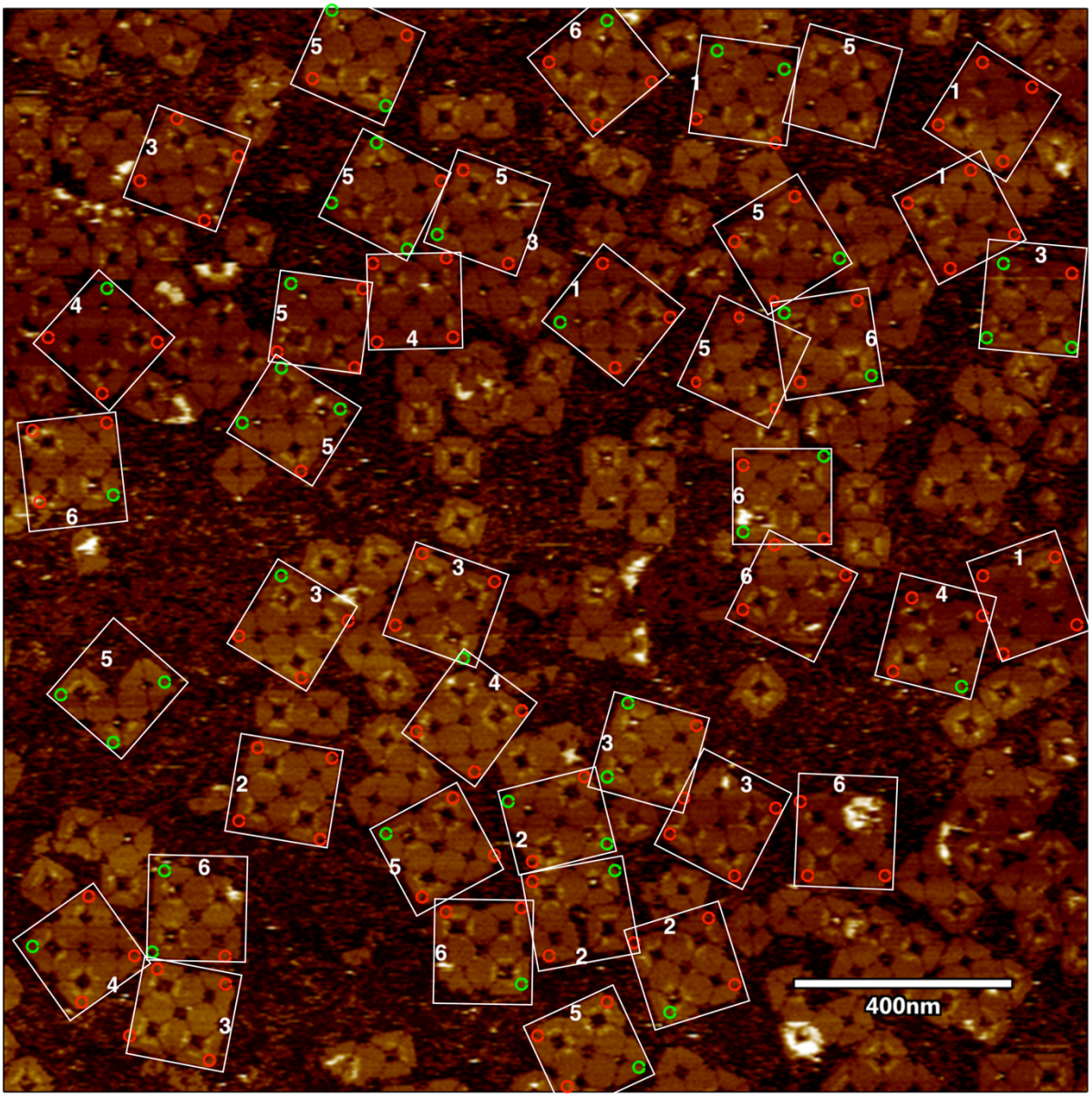
Supplementary Figure 2. AFM images representing 2 x 2 arrays. The yield of the formation of 2 x 2 arrays derived from these images is ≈70%.



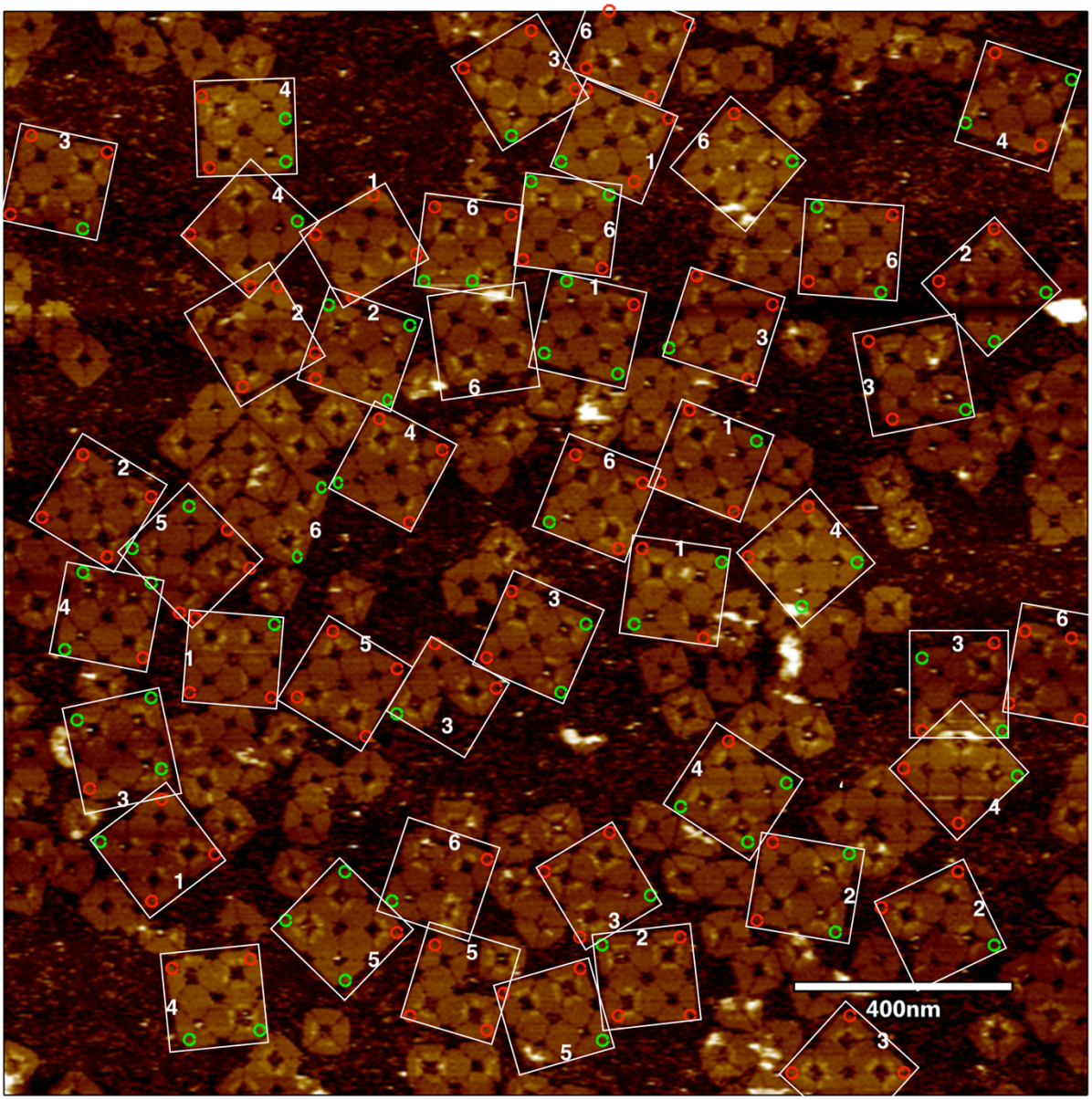
Supplementary Figure 3. DNA arrays used to study thrombin binding while modifying spacer length and flexibility. Left, schematic illustration of the configurations used in 2x2 array barcoded with number 4, 5 and 6. Center, representative AFM image of 2x2 arrays with thrombin bound in some cavities. Right, Binding yield in % obtained from over ≈ 200 tiles for each configuration measured in a single AFM experiment (one sample of a set of barcoded cavities was imaged at multiple locations - for an overview of experimental variability see Supplementary Figure 10). Green denotes HD22, always in position 1, and orange denotes TBA1 aptamer in position 4 (90°) or 7 (180°). The letter indicates the spacer length and flexibility. See Figure 1 and 2 in the main text for more details.



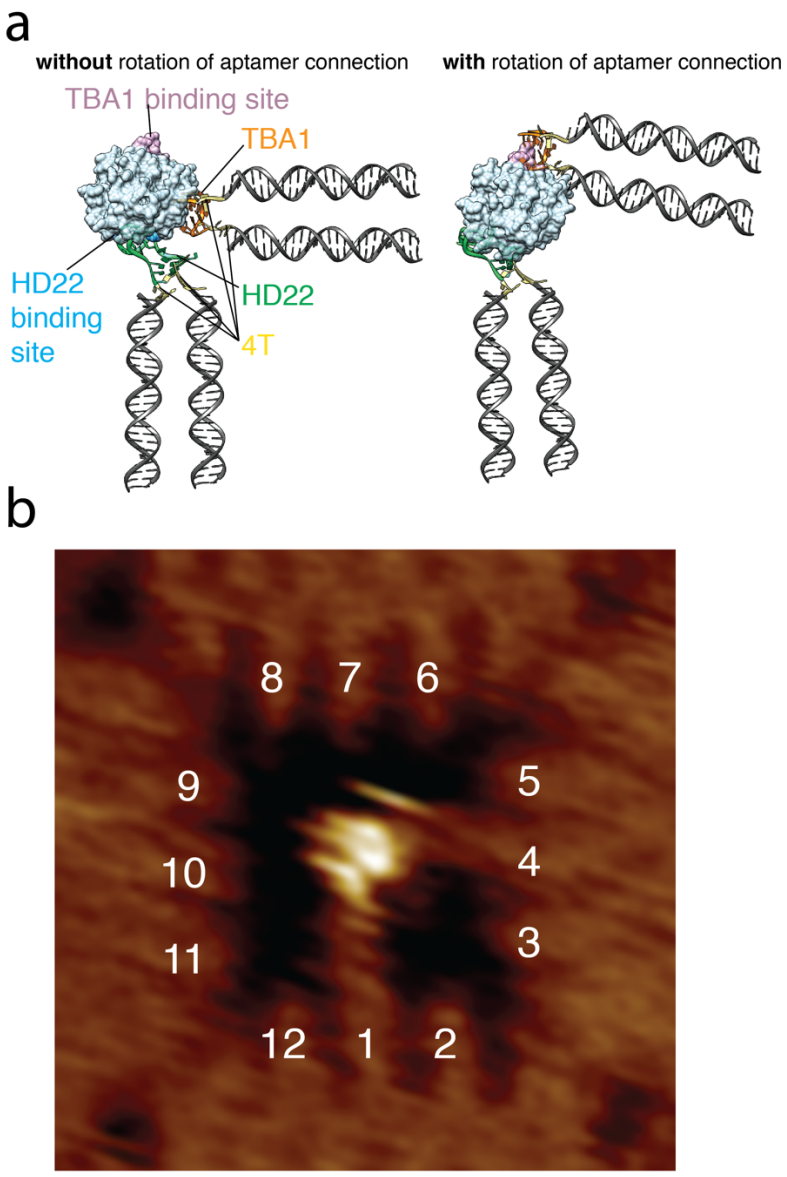
Supplementary Figure 4. AFM quantification of thrombin binding to 2x2 arrays. Example number 1. AFM images of 2x2 arrays with barcodes from 1 to 6 incubated with alpha-thrombin. White squares indicate the 2x2 array, and the numbers indicate the barcode number. Green circles indicate origami with protein bound in the central cavity; red circles indicate empty nanostructures. Single origami structures or ambiguous cases were not considered in our statistics. High-resolution images were analyzed to determine the binding yield.



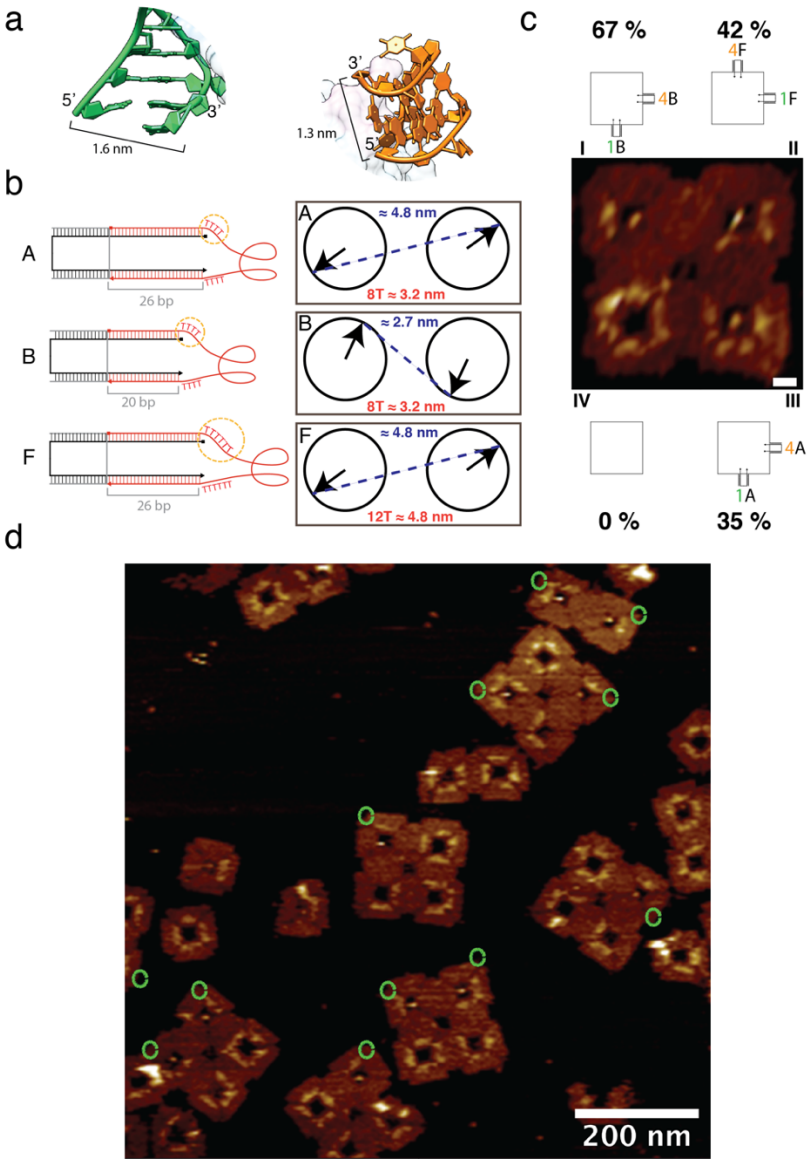
Supplementary Figure 5. AFM quantification of thrombin binding to 2x2 arrays. Example number 2. AFM images of 2x2 arrays with barcodes from 1 to 6 incubated with alpha-thrombin. White squares indicate the 2x2 array, and the numbers indicate the barcode number. Green circles indicate origami with protein bound in the central cavity; red circles indicate empty nanostructures. Single origami structures or ambiguous cases were not considered in our statistics. High-resolution images were analyzed to determine the binding yield.



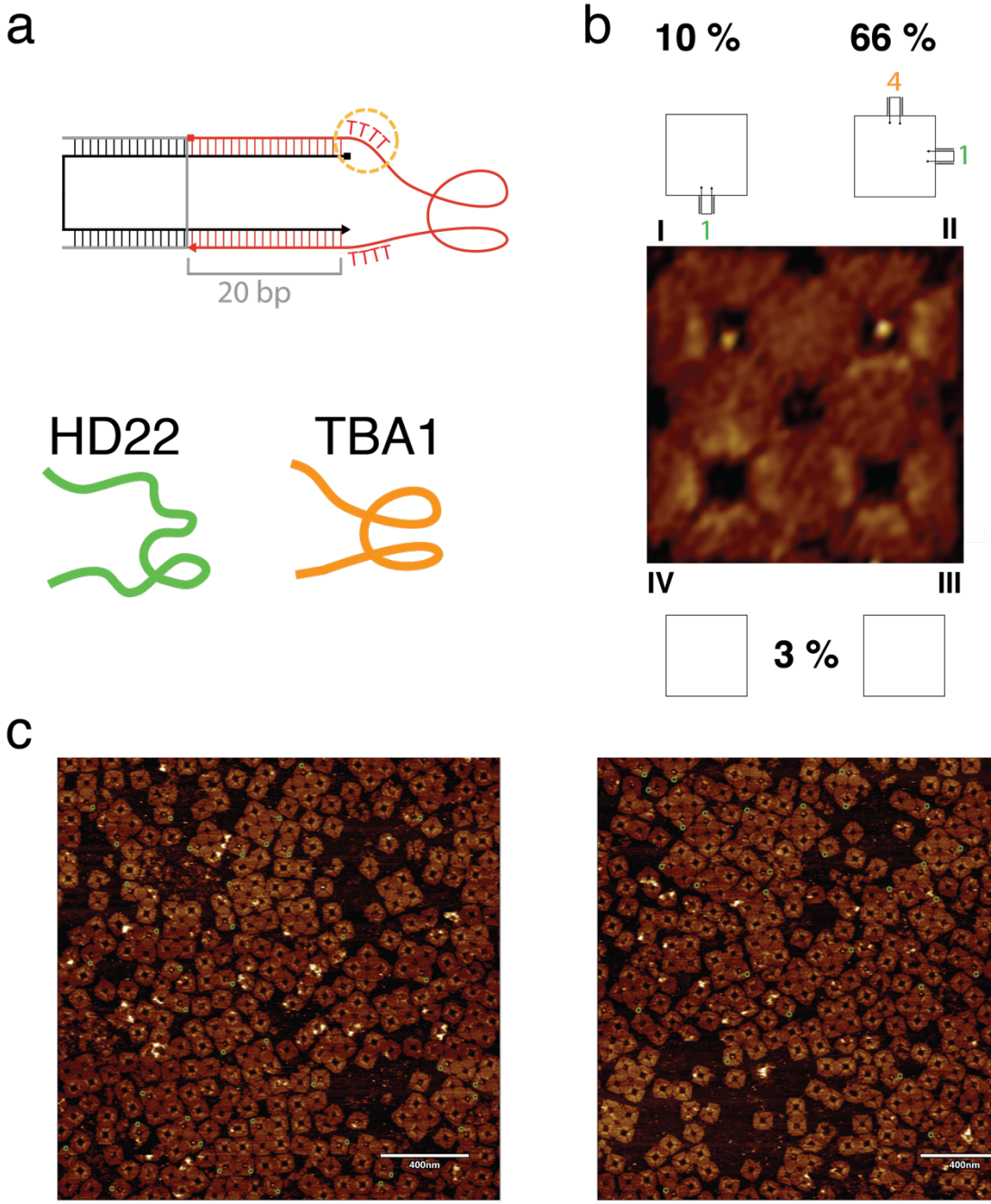
Supplementary Figure 6. AFM quantification of thrombin binding to 2x2 arrays. Example number 3. AFM images of 2 x 2s arrays with barcodes from 1 to 6 incubated with alpha-thrombin. White squares indicate the 2 x 2 arrays, and the numbers indicate the barcode number. Green circles indicate origami with protein bound in the central cavity; red circles indicate empty nanostructures. Single origami structures or ambiguous cases were not considered in the statistics. High-resolution images were analyzed to determine the binding yield.



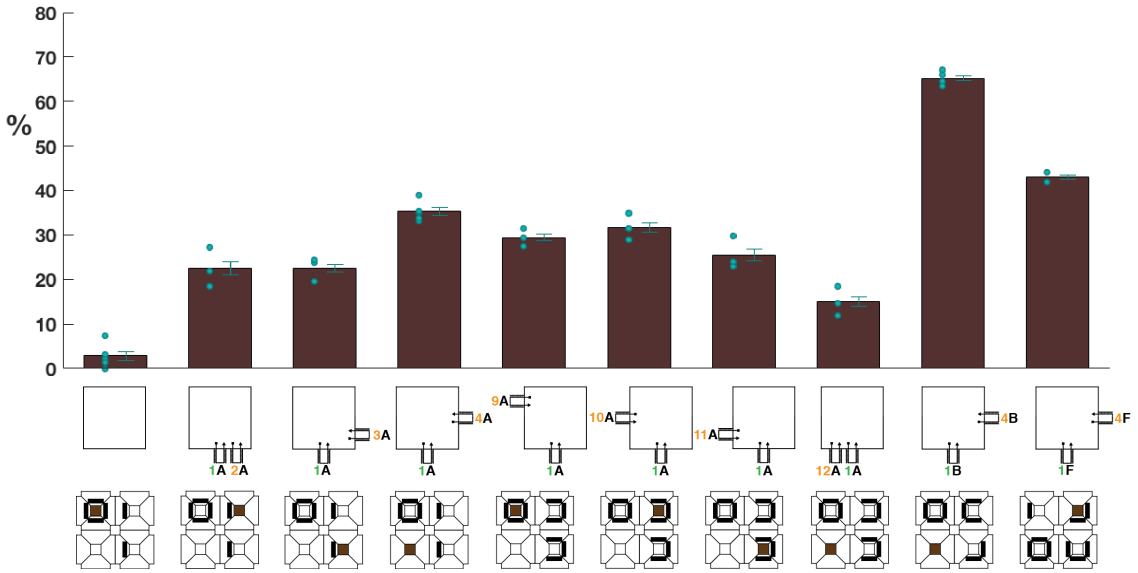
Supplementary Figure 7. Scheme of the interaction of α-thrombin with spacer A (26 bp) with an angle corrected in a way that allows both HD22 and TBA1 aptamers to interact with the protein at their binding sites. **a,** Model of human α-thrombin bound to DNA aptamers in the central cavity of the structure. Thrombin bound to its two aptamers: HD22 (in green) and TBA1 (in orange), which are linked via four consecutive thymidines (in khaki) to the stem in order to allow the aptamer to rotate. The end-to-end distances of the aptamer attachment sites are not large enough to accommodate the protein inside (left image). We assume that the spacers rotate slightly inside the cavity to allow the site-specific interaction of the aptamers with their binding sites (in dark blue and light purple) (right image). Molecular models were rendered and modelled using Chimera and pdb file 5EW2 and 1HUT. **b,** AFM zoom image that shows the binding of thrombin to aptamer with this spacer.



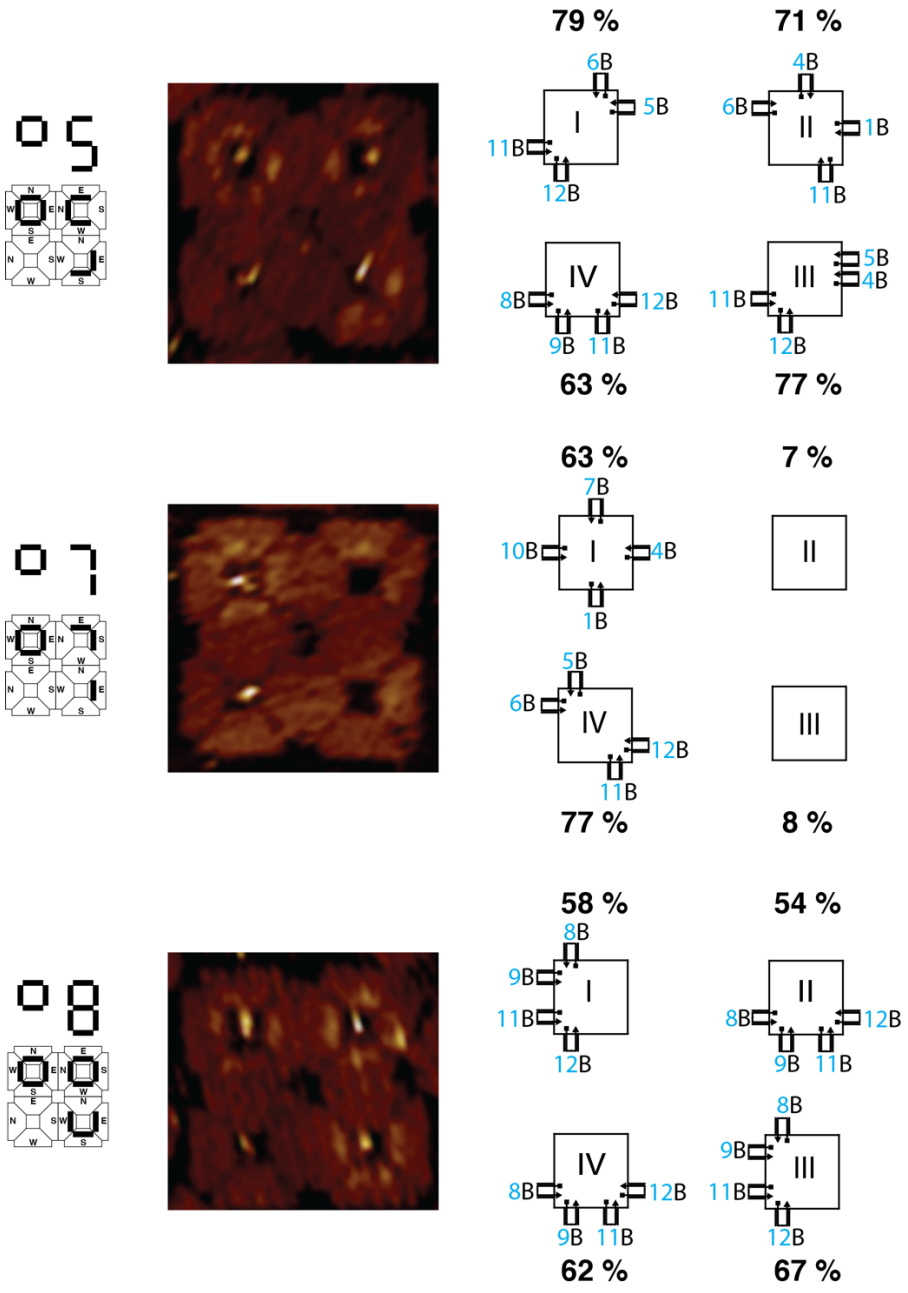
Supplementary Figure 8. Comparison of the binding fraction for nanostructures with 4 or 6 thymidines as linker between the stem and the aptamer. **a,** Distances between the 5' and 3' ends of the aptamers. 3D models rendered in Chimera using as model PDB IDs: 5EW1[3] and 1HUT[2]. HD22 (green) and TBA1 (orange). **b,** the length of the spacer stem varied from 20 bp to 26 bp. The single-stranded linker between the stem and the aptamer consists of four (case A and B) or six (case F) thymidines. The boxes show the DNA orientation in the double helix and the distance between them. This distance is larger in the case of a 26 bp spacer. **c,** Percentages denote the binding yield in each configuration and the negative control for the nanostructure number **IV**. Schemes and AFM image of the three configurations of aptamers are shown. Structure number **IV** has no aptamer as a negative control. HD22 is always in position 1 and TBA1 in position 4. **d,** AFM quantification of thrombin binding to 2 x 2 arrays. Green circles indicate origami structures with protein bound in the central cavity.



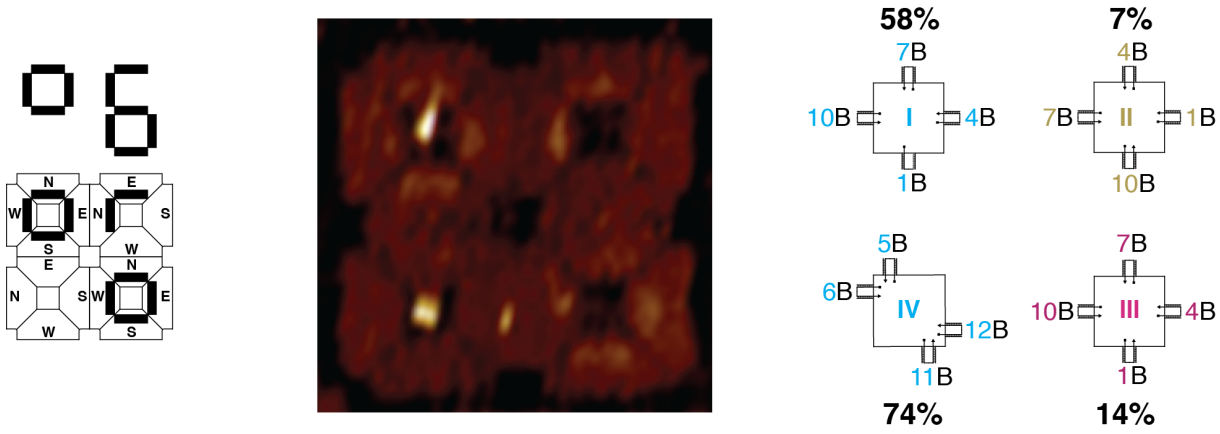
Supplementary Figure 9. Comparison of the binding fraction for nanostructures with only HD22 or both HD22 and TBA1 aptamer. **a**, the length of the spacer is 20 bp. The single-stranded linker between the stem and the aptamer is composed of four thymidines. **b**, Percentages denote the binding yield in the two configurations and for the nanostructures number **III** and **IV** negative controls. Schemes and AFM image of the two configurations of aptamers are shown. HD22 is always in position 1 and the TBA1 position 4. **c,** AFM quantification of thrombin binding to 2 x 2 arrays. Green circles indicate origami structures with a protein bound in the central cavity.



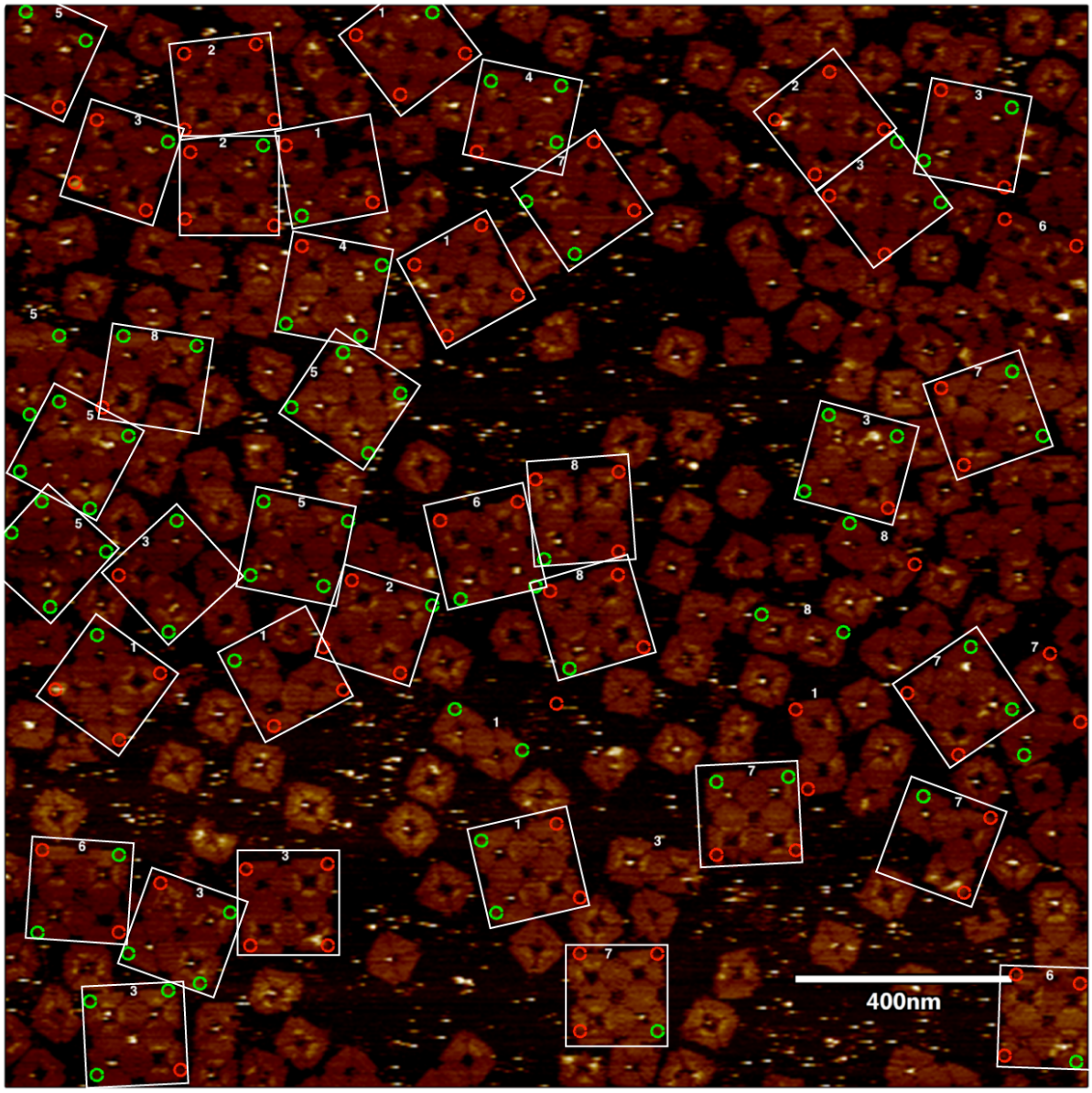
Supplementary Figure 10. Comparison of binding yields for ten different thrombin aptamer configurations and spacer designs measured in independent AFM sessions (different sessions performed on different days). Bar graphs represent the mean ± SEM for each case obtained from n=2-6 experiments, which are all below 3%. Green dots represent individual points (yield percentage obtained from independent AFM sessions). The schematic configuration of aptamers and the illustration of the position of the different configurations in 2 x 2 arrays (brown) plus their respective barcoded 2 x 2 are shown. SEM = standard error of the mean.



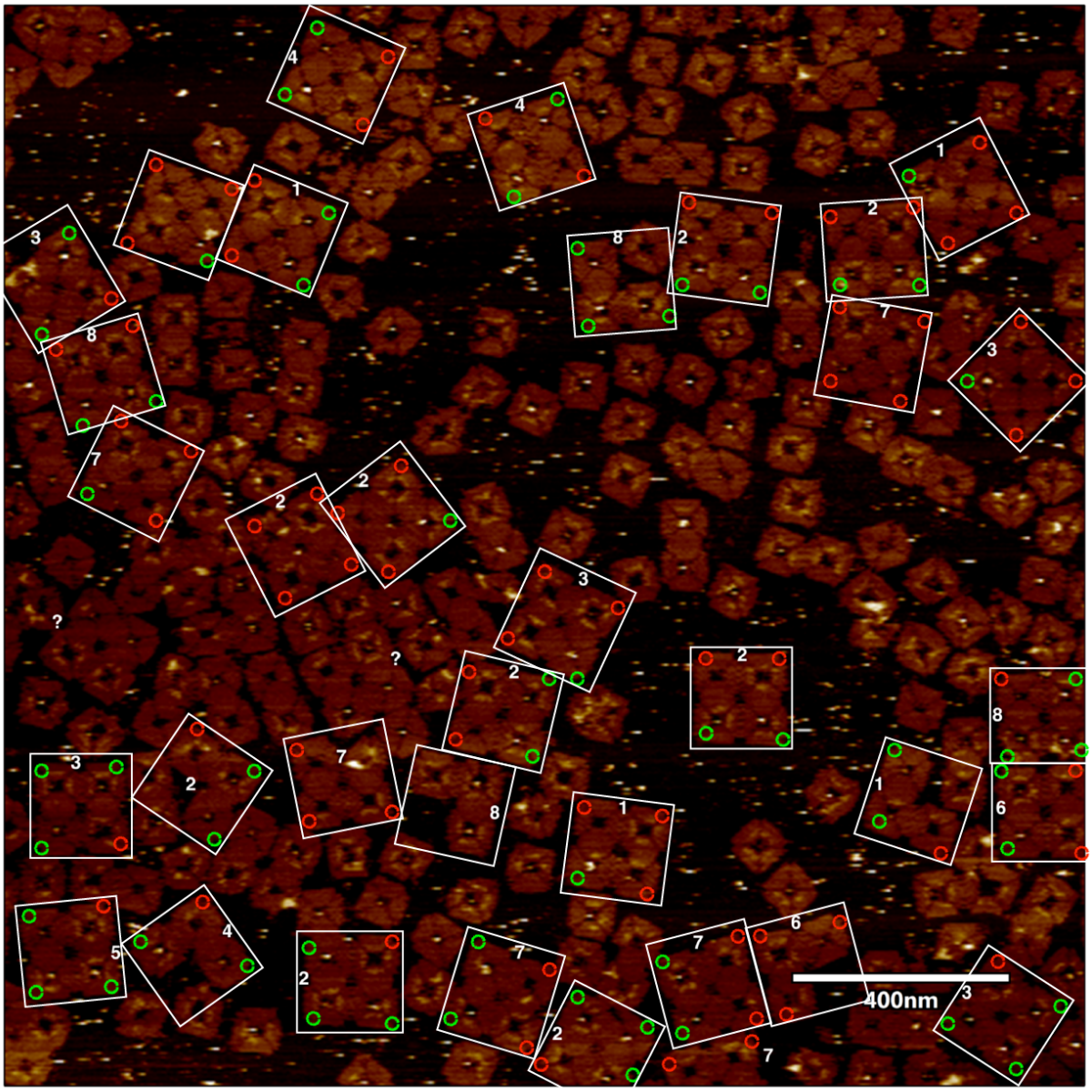
Supplementary Figure 11. DNA arrays 5, 7 and 8, which were used to study streptavidin binding to SSA1 aptamers with different aptamer configurations. Left, schematic illustration of the configurations used in 2 x 2 arrays barcoded with number 5, 7 and 8. Center, representative AFM images of 2 x 2 arrays with streptavidin bound in some of the cavities. Right, Percentages denote the binding yield determined for each configuration. The configuration of the aptamers inside each cavity is represented as a number, that indicates the position of the aptamer, and a letter, that indicates the spacer length and flexibility (case B, 20 bp spacer and 4 thymidines as an aptamer linker). The number in blue indicates that all aptamers are the same SAA1 structure (see Figure 3 in the main text for more details). Array 5 has the same configuration as 2 and 4 but with different staples.



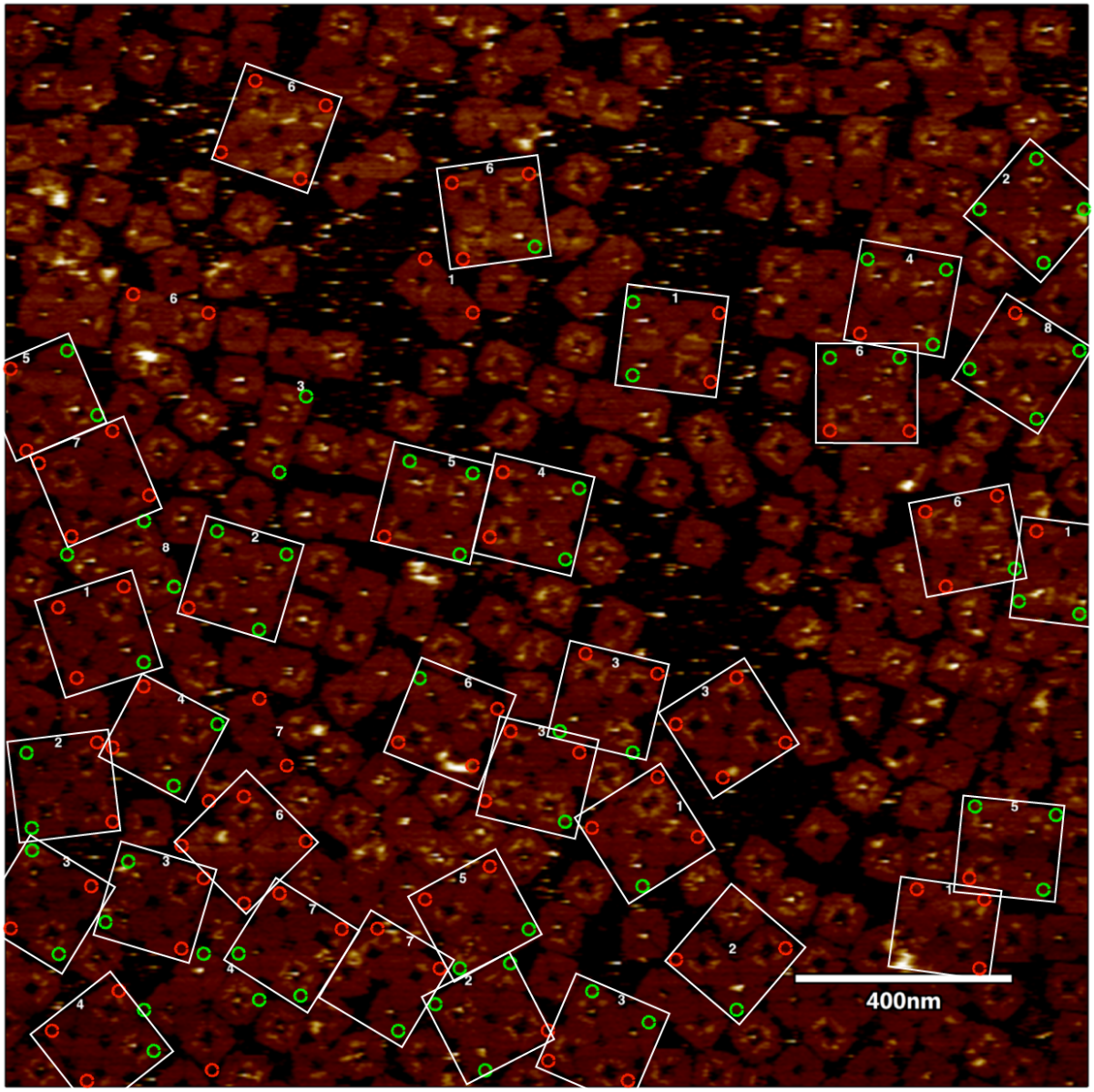
Supplementary Figure 12. DNA array used to compare streptavidin binding to SAA1, SAA2 and SAA3 aptamers. Left, schematic illustration of the configurations used in 2 x 2s barcoded with number 6. Center, representative AFM image of a 2 x 2 array with streptavidin bound in two of the cavities. Right, Percentages denote the binding yield in each configuration. The configuration of aptamers inside each cavity is represented as a number, which indicates the position of the aptamer, and a letter, which indicates the spacer length and flexibility. Blue codes for SAA1 aptamer, khaki for SAA2 and pink for SAA3. See Figure 3 and 4 in the main text for more details.



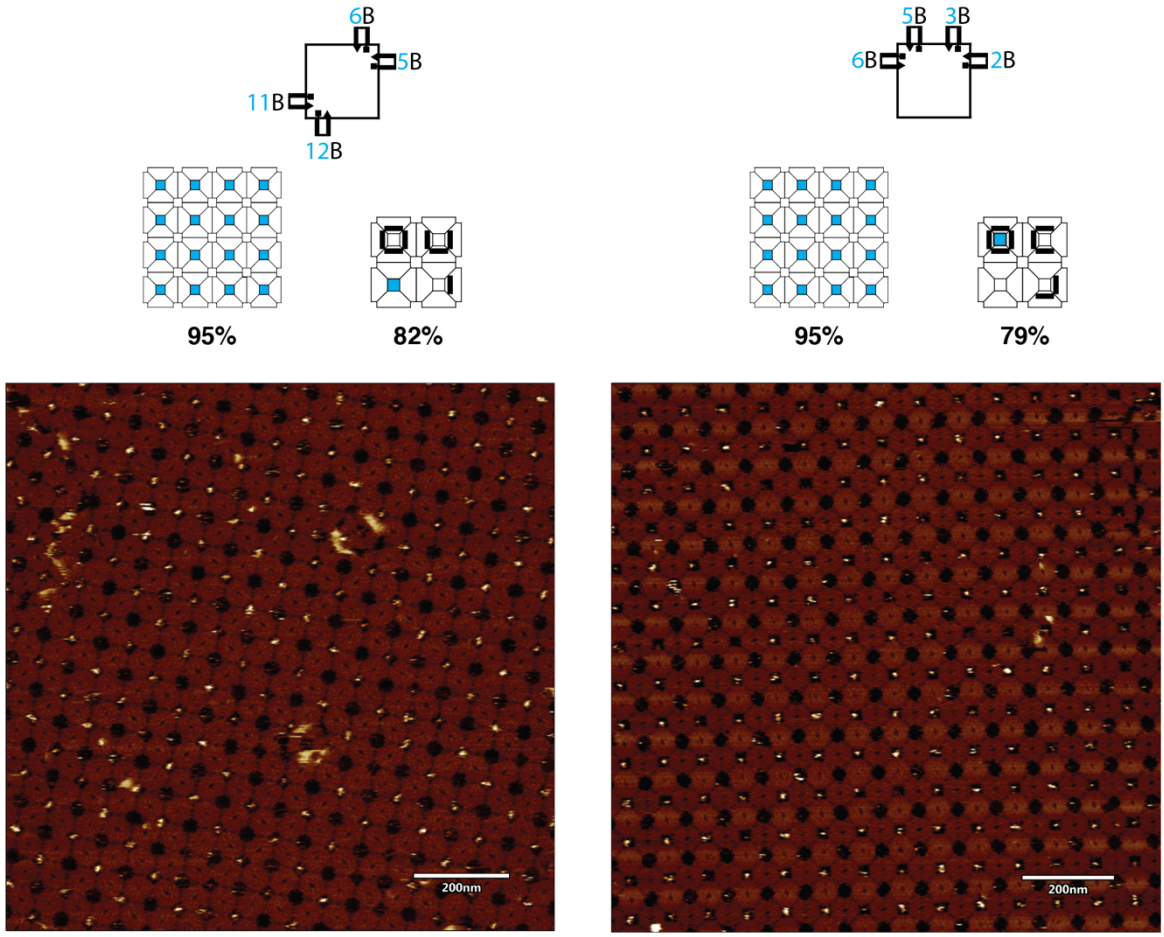
Supplementary Figure 13. AFM quantification of streptavidin binding to 2x2 arrays with barcodes from 1 to 8. Example figure 1. White squares indicate the 2x2 array, and the numbers indicate the barcode number. Green circles indicate origami structures with a protein bound in the central cavity; red circles indicate empty nanostructures. Single origami structures or ambiguous cases were not considered in statistics. High-resolution images were analyzed to determine the binding yield.



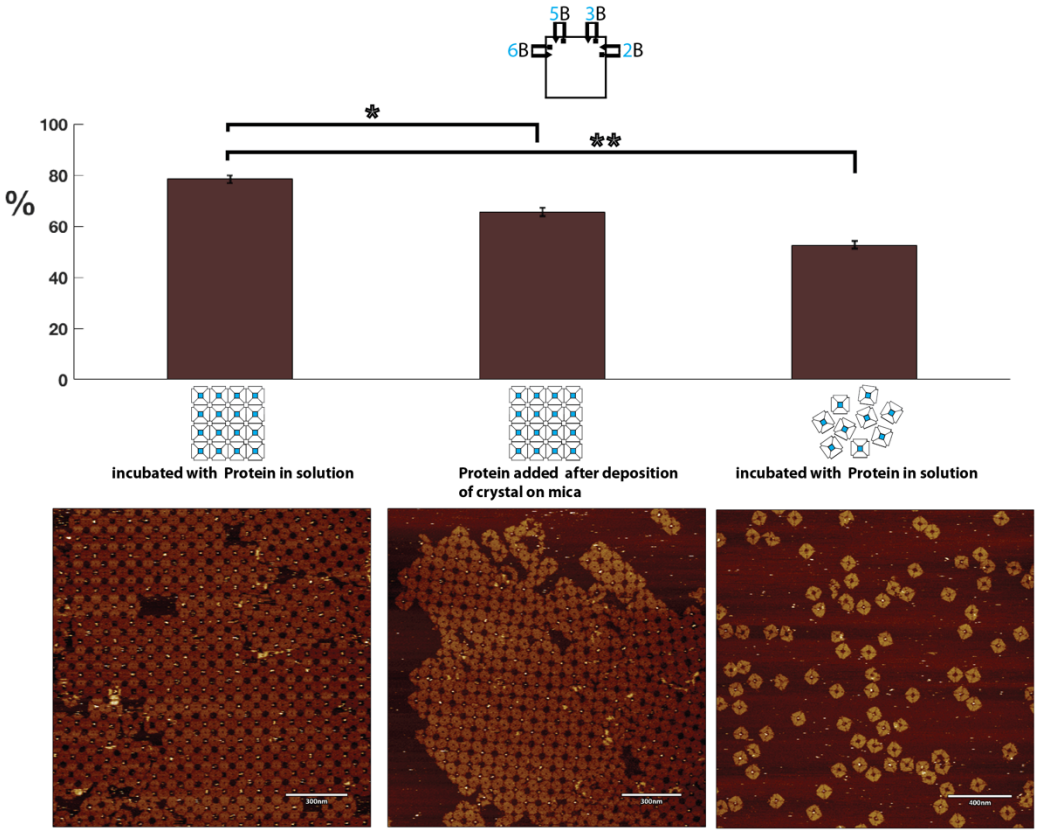
Supplementary Figure 14. AFM quantification of streptavidin binding to 2x2 arrays with barcodes from 1 to 8. Example figure 2. White squares indicate the 2x2 arrays, and the numbers indicate the barcode number. Green circles indicate origami structures with protein bound in the central cavity; red circles indicate empty nanostructures. Question marks indicate structures, which could not be classified. Single origami structures or ambiguous cases were not considered in our statistics. High-resolution images were analyzed to determine the binding yield.



Supplementary Figure 15. AFM quantification of streptavidin binding to 2x2 arrays with barcodes from 1 to 8. Example figure 3. White squares indicate the 2x2 array, and the numbers indicate the barcode number. Green circles indicate origami with protein bound in the central cavity; red circles indicate empty nanostructures. Single origamis or ambiguous cases were not considered in our statistics. High-resolution images were analyzed to determine the binding yield.



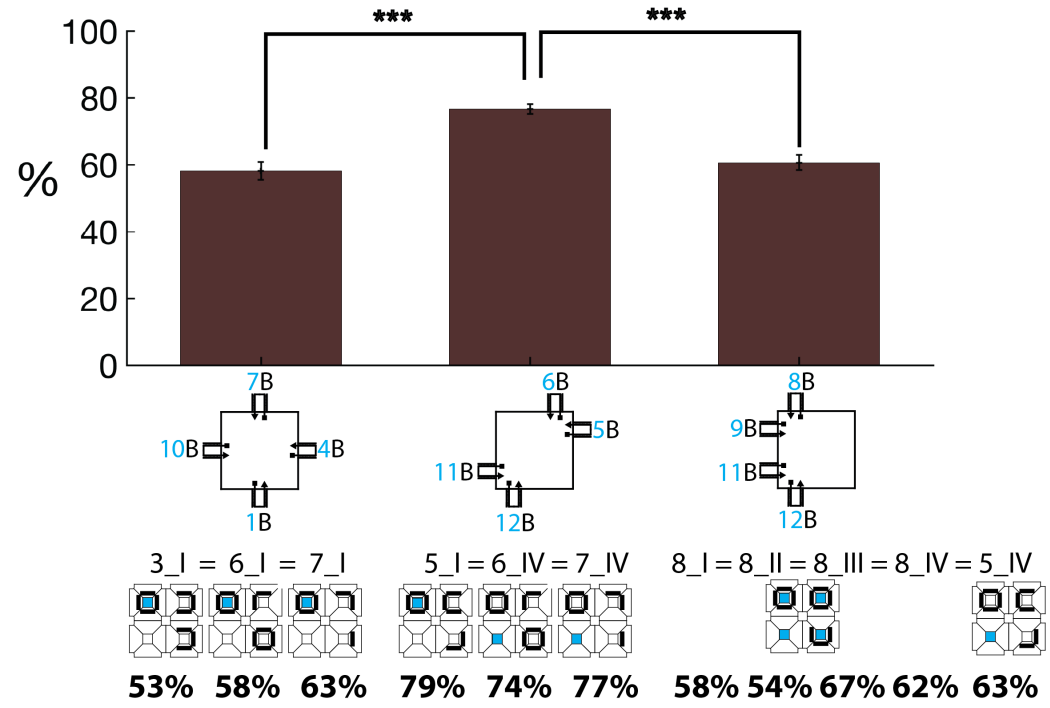
Supplementary Figure 16. Comparison between the binding yields of streptavidin to cavities with two different aptamer configurations both in crystals and 2 x 2 arrays. Top: Schematic representation of the SAA1 aptamer in the cavity of a 2D crystal and 2 x 2 array (highlighted in blue) and the corresponding binding yields. Bottom: AFM images of 2D crystalline arrays formed from DNA origami structures with the two different aptamer configurations. Example images of the 2 x 2 arrays are shown in Supplementary Figures 13 - 15.



Supplementary Figure 17. Comparison between the binding yield of streptavidin incubated either with 2D crystal/cavity structures in solution or deposited on mica. The bar graph compares the binding yield for different incubation procedures for the SAA1 aptamer configuration (2, 3, 5, 6) shown on the top: The bar on the left corresponds to 2D crystals, which were incubated with streptavidin for 30 min at 37ºC and then deposited in mica for imaging (cf. Section S2). The bar in the middle shows the binding yield of 2D crystals deposited on mica in the absence of protein, followed by addition of 5µl of 40 nM STV solution in 1XFB and 60 µl of 1X folding buffer. The sample was then incubated at 37°C for 30 min and imaged afterwards. The right bar represents the binding yield of single structures incubated with streptavidin in solution for 30 min before imaging. The bar graphs state the mean ± SEM (n=3). For the solution-incubated samples, three independent AFM sessions at different times were performed. For the crystal incubation with protein on the surface, three measurements were performed at three different positions on the mica surface in the same experiment. Analysis via a t-student test indicates statistically significant differences in the binding yields (*$P < 0.05$, **$P < 0.01$). Both 2D crystals give a higher binding yield than the isolated structures, while solution incubation appears to form better than surface incubation.

Note that the experiment was performed with a different batch of streptavidin proteins, resulting in overall reduced binding yields compared to those stated in Supplementary Figure 16. This gives an indication that the largest source of variability in our experiments probably is the protein quality and concentration. Importantly, comparison of barcoded structures under identical conditions is not affected by this variation and allows us to make robust statements about binding yield differences.

a



b

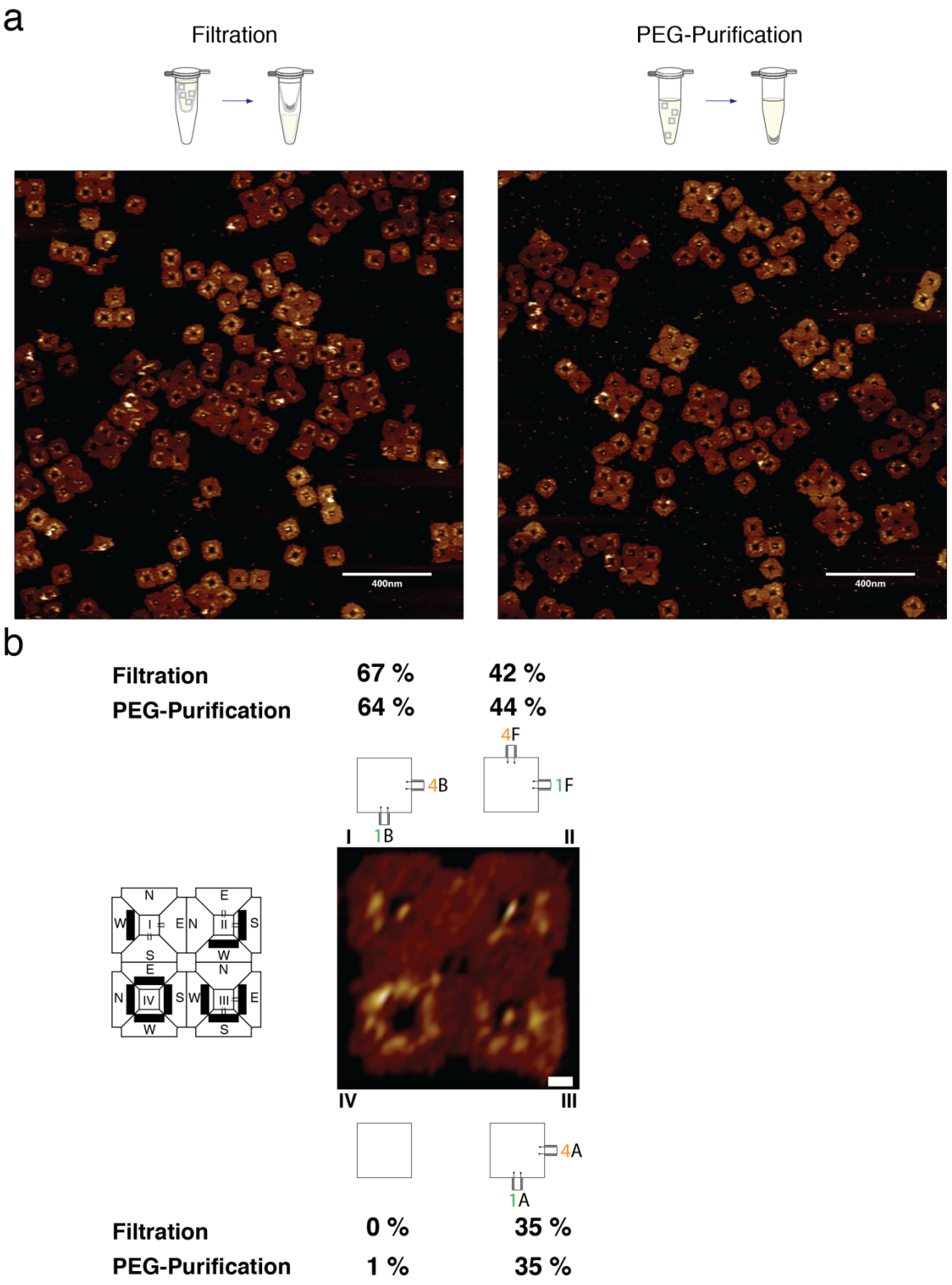
SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
(1,4,7,10)	3	174,505238	58,1684128	21,5633098
(5,6,11,12)	3	230,072029	76,6906762	6,66688588
(8,9,11,12)	5	303,12862	60,625724	25,1565094

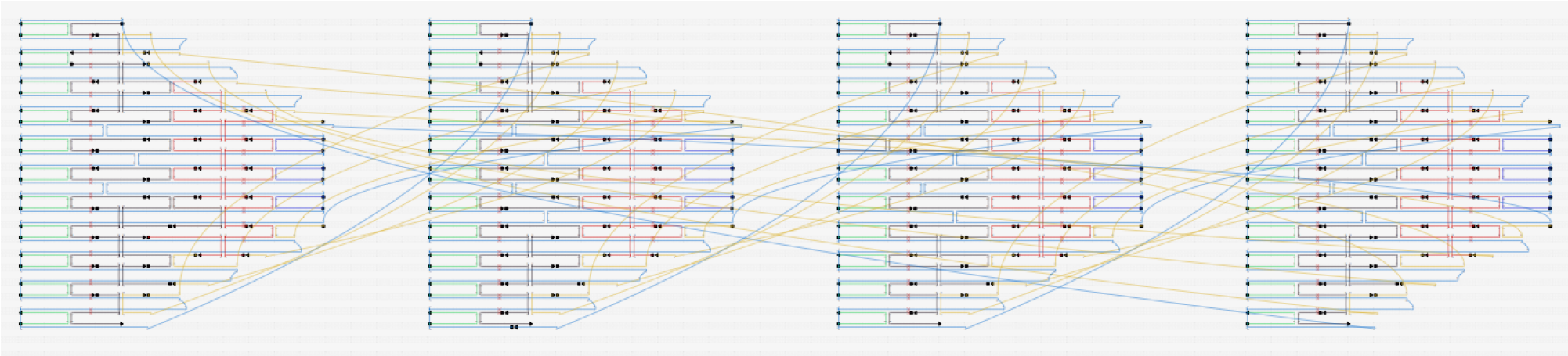
ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	640,862205	2	320,431102	16,318716	0,00150195	4,45897011
Within Groups	157,086429	8	19,6358036			
Total	797,948634	10				

Supplementary Figure 18. Comparison between three streptavidin aptamer binding configurations to estimate binding yield uncertainty within the same experiment. Analysis of variance (ANOVA) and a t-student test reveal statistically significant differences in binding yield. **a,** Schematic representations of three SAA1 aptamer configurations with corresponding binding yields. As indicated, the same configuration was present in different cavities of barcoded 2x2 arrays. The percentages given below the structures show the binding yield of the cavities highlighted in blue. The bar graphs represent the mean ± SEM of the three configurations (n=3, 3 and 5). P-values were calculated with student t-test (***P < 0.001). **b,** Summary of the data analyzed using a one-way ANOVA test to assess statistical significance.



Supplementary Figure 19. Comparison of the binding yield with nanostructures purified with either column-filtration (left) or PEG-purification (right). **a,** Scheme and AFM image of the 2 x 2 arrays incubated with protein for 30 min. In the case of PEG-purification, we observed slightly more unbound protein on the mica surface. **b,** Scheme and binding yield of the proteins in the four configurations.



Supplementary Figure 20. Cadnano design for the square-shaped nanostructure. Staples used for barcoding are colored in red. From left to the right: trapezoids West, South, East and North.

3. Supplementary Tables

Supplementary Table 1. Quantification of binding yield of α-Thrombin to 2 x 2 arrays 1-6. The length and type of spacer is indicated with characters A-E. HD22 is located in coordinate 1 and the position of the TBA1 aptamer changed in the 24 configurations. See figure 2. for more details.

spacer		AA (26bp spacer)							AA (26bp spacer)			
coordinates	negative	1,2	1,3	1,4 (90º)	1,5	1,6	1,7 (180º)	1,8	1,9	1,10	1,11	1,12
2 x 2		**1**			**2**				**3**			
position	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
total structures	205	202	205	208	214	213	203	209	205	209	205	211
protein bound	15	44	50	81	67	67	30	37	56	73	49	31
%	**7**	**22**	**24**	**39**	**31**	**31**	**15**	**18**	**27**	**35**	**24**	**15**

spacer	CC (30 bp spacer-stiff)		DD (20bp flex)		EE (20bp spacer)		BB (20 bp spacer)		CD (hd22 C tba D)		dc hd22 flex tba stiff	
coordinates	1,4 (90º)	1,7 (180º)	1,7 (180º)	1,4 (90º)	1,4 (90º)	1,7 (180º)	1,7 (180º)	1,4 (90º)	1,4 (90º)	1,7 (180º)	1,7 (180º)	1,4 (90º)
2 x 2	**4**				**5**				**6**			
position	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
total structures	207	209	220	219	203	210	211	213	215	215	216	221
protein bound	48	32	40	103	49	31	56	135	92	45	30	62
%	**23**	**15**	**18**	**47**	**24**	**15**	**27**	**63**	**43**	**21**	**14**	**28**

Supplementary Table 2. Quantification of binding yield of alpha-Thrombin to 2 x 2s with 4T or 6T linker between the stem and the aptamer. The length of the spacer is denoted with the A, B and F cases. The configuration tested was 90º, in which HD22 is located in coordinate 1 and TBA1 aptamer in 4. See figure 2. and supplementary figure 7 and 14 for more details.

2 x 2	4T vs 6T			
Column filtration	**I**	**II**	**III**	**IV**
total structures	210	215	222	226
bound protein	141	90	78	0
%	**67**	**42**	**35**	**0**

PEG purification	**I**	**II**	**III**	**IV**
total structures	211	195	207	217
bound	136	86	73	3
%	**64**	**44**	**35**	**1**

Supplementary Table 3. Quantification of binding yield of alpha-Thrombin to 2 x 2s with only HD22 or HD22 and TBA1 aptamer. The length of the spacer stem is 20 bp. The single-stranded linker between the stem and the aptamer contains four thymidines. See supplementary figure 8 for more details.

2 x 2	1 aptamer vs 2 aptamer			
Nanostructure	**I**	**II**	**III**	**IV**
total structures	241	249	187	213
bound protein	24	165	4	7
%	**10**	**66**	**2**	**3**

Supplementary Table 4. Quantification of binding yield of streptavidin to 2 x 2s 1-8 with the following configurations. The length of the spacer stem is denoted with A (26 bp) or B (20 bp). The coordinates show the 4 positions where the SAA1-3 protrude. See figure 3 for more details.

spacer	A (26bp spacer)				A (26bp spacer)				B (20 bp spacer)				B (20 bp spacer)			
coordinates	1,4,7,10	2,5,8,11	2,3	2,3,5,12	2,38,9	5,7,10,12	1,2,8,9	2,3,5,6	1,4,7,10	2,5,8,11	2,3	2,3,5,12	2,38,9	5,7,10,12	1,2,8,9	2,3,5,6
2 x 2	**1**				**2**				**3**				**4**			
position	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
total structures	201	202	213	200	205	204	204	210	193	200	212	206	183	211	211	202
bound protein	83	108	73	100	102	96	82	114	103	160	97	135	110	128	138	166
%	**41**	**53**	**34**	**50**	**50**	**47**	**40**	**54**	**53**	**80**	**46**	**66**	**60**	**61**	**65**	**82**

spacer	B (20 bp spacer)				B (20 bp spacer)				B (20 bp spacer)				B (20 bp spacer)			
coordinates	5,6,11,12	1,4,6,11	4,5,11,12	8,9,11,12	1,4,7,10	1,4,7,10	1,4,7,10	5,6,11,12	1,4,7,10			5,6,11,12	8,9,11,12	8,9,11,12	8,9,11,12	8,9,11,12
2 x 2	**5**				**6**				**7**				**8**			
position	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
total structures	198	203	207	207	200	209	206	201	182	214	211	197	201	206	200	200
bound protein	157	145	159	130	117	15	29	149	114	14	17	151	118	126	138	133
%	**79**	**71**	**77**	**63**	**59**	**7**	**14**	**74**	**63**	**7**	**8**	**77**	**59**	**61**	**69**	**67**

5. Supplementary Discussion

We here give a brief overview of concepts and theoretical considerations related to multivalent binding which are relevant in the context of our origami-based binding cavity.

5.1 Multivalent binding

We assume that a ligand protein P can bind to two aptamers A_1 and A_2, which are connected by a scaffold structure S. The individual equilibrium constants for binding to the aptamers are given by K_1 and K_2, i.e.,

$$\mathrm{P} + \mathrm{A_i} \rightleftharpoons \mathrm{P{-}A_i} \tag{1}$$

with

$$K_i = \frac{[\mathrm{P} - \mathrm{A_i}]}{[\mathrm{P}][\mathrm{A_i}]} \tag{2}$$

In the context of the aptamer scaffold, we have:

$$\mathrm{P} + \mathrm{A_1{-}S{-}A_2} \rightleftharpoons \mathrm{P{-}A_1{-}S{-}A_2} \tag{3}$$

$$\mathrm{A_1{-}S{-}A_2} + \mathrm{P} \rightleftharpoons \mathrm{A_1{-}S{-}A_2{-}P} \tag{4}$$

The corresponding equilibrium constants will, in general, be affected by the scaffold structure - typically a reduction in binding strength is observed (see below).
Both complexes can intramolecularly transition into a complex, in which the ligand is bound to both aptamers, i.e.

$$\mathrm{P{-}A_1{-}A_2} \quad \rightleftharpoons \quad \mathrm{A_1{-}P{-}A_2} \tag{5}$$

$$\mathrm{A_1{-}A_2{-}P} \quad \rightleftharpoons \quad \mathrm{A_1{-}P{-}A_2}, \tag{6}$$

where we omitted the scaffold S in order to simplify the notation (cf. Scheme 1) - the resulting bivalent complexes are identical. We denote the equilibrium constants for these intramolecular reactions by K_2' and K_1', i.e.,

$$K_1' = \frac{[\mathrm{A_1} - \mathrm{P} - \mathrm{A_2}]}{[\mathrm{A_1} - \mathrm{A_2} - \mathrm{P}]} \tag{7}$$

$$K_2' = \frac{[\mathrm{A_1} - \mathrm{P} - \mathrm{A_2}]}{[\mathrm{P} - \mathrm{A_1} - \mathrm{A_2}]} \tag{8}$$

The effective equilibrium constant K_{12} is then given by

$$K_{12} = K_1 K_2' = K_2 K_1' \tag{9}$$

The observable association constant is given by the ratio of the concentration of bound proteins and that of unbound proteins and scaffolds:

$$
\begin{aligned}
K_{obs} &= \frac{[\mathsf{P}]_{\text{bound}}}{[\mathsf{P}][\mathsf{A_1}-\mathsf{A_2}]} && (10)\\
&= \frac{[\mathsf{P}-\mathsf{A_1}-\mathsf{A_2}]+[\mathsf{A_1}-\mathsf{A_2}-\mathsf{P}]+[\mathsf{A_1}-\mathsf{P}-\mathsf{A_2}]}{[\mathsf{P}][\mathsf{A_1}-\mathsf{A_2}]} && (11)\\
&= K_1+K_2+\frac{[\mathsf{A_1}-\mathsf{P}-\mathsf{A_2}]}{[\mathsf{P}][\mathsf{A_1}-\mathsf{A_2}]} = K_1+K_2+K_{12} && (12)\\
&= K_1(1+K_2')+K_2(1+K_1') && (13)
\end{aligned}
$$

In case of effective scaffolding, $K_1', K_2' \gg 1$, and thus

$$K_{obs} = K_1K_2' + K_2K_1' \tag{14}$$

Whereas K_1, K_2 relate to second order association reactions and depend on the concentration of the ligand P, K_1', K_2' are (dimensionless) first order equilibrium constants, which depend on the 'local concentration' of the ligand within the complex. This local concentration is determined by the orientation and distance of the aptamers and their mechanical properties (their flexibility). The closer P is held by one aptamer in the vicinity of the other in the correct orientation, the higher the probability of formation of the second aptamer-ligand bond, and also the lower the entropy loss upon bond formation.

5.2 Influence of flexibility - a 1D toy model

In order to qualitatively understand the role of flexibility of the aptamers and their linkers, we here consider a one-dimensional toy model, in which two binders are attached to rigid walls via springs with spring constants k_1, k_2. The springs are elastic elements that represent, e.g., stretching/bending of a double helical element or entropic elasticity of single-stranded DNA etc. The effective binding energy as a function of the spring extensions is then given by:

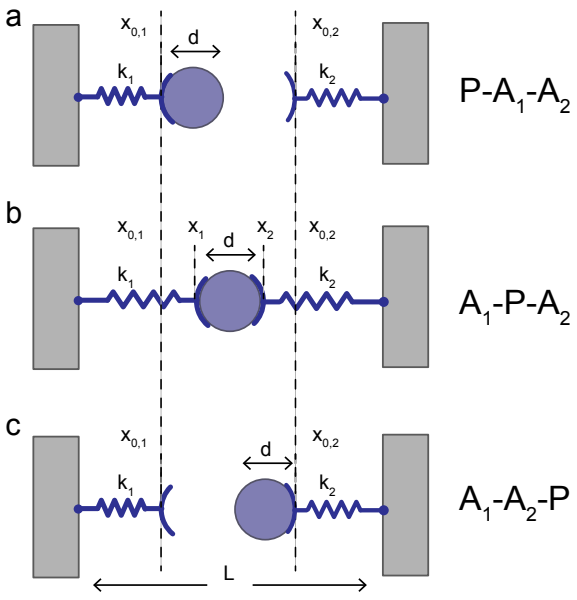
$$E_{eff}(x_1,x_2) = \epsilon_1 + \frac{1}{2}k_1(x_1-x_{0,1})^2 + \epsilon_2 + \frac{1}{2}k_2(x_2-x_{0,2})^2, \tag{15}$$

where $\epsilon_1, \epsilon_2 < 0$ are the binding energies for binding to only one of the ligands. If both ligands are bound, the diameter of the molecule has to fulfill $d = x_2 - x_1$, and therefore:

$$E_{eff}(x_1) = \epsilon_1 + \frac{1}{2}k_1(x_1-x_{0,1})^2 + \epsilon_2 + \frac{1}{2}k_2(x_1+d-x_{0,2})^2 \tag{16}$$

We find the equilibrium position by taking the derivative with respect to x_1:

$$\frac{dE_{eff}(x_1)}{dx_1} = k_1(x_1-x_{0,1}) + k_2(x_1+d-x_{0,2}) \stackrel{!}{=} 0, \tag{17}$$



Scheme 1. 1D Model of two aptamers connected to a scaffold cavity of length L via springs with spring constants k_1, k_2 and equilibrium positions $x_{0,1}, x_{0,2}$. a: Protein P (with diameter d) bound only to the left aptamer (Aptamer 1), b: protein bound to both aptamers and c: P bound to the right aptamer.

and hence:

$$\tilde{x}_1 = \frac{k_1 x_{0,1} + k_2(x_{0,2} - d)}{k_1 + k_2} \tag{18}$$

The equilibrium energy therefore is given by:

$$E_{eff}(\tilde{x}_1) = \epsilon_1 + \epsilon_2 + \frac{k_1 k_2 (x_{0,1} - x_{0,2} + d)^2}{2(k_1 + k_2)} \tag{19}$$

Binding to both ligands will be energetically favorable when E_{eff} is lower than either of the individual energies ϵ_1, ϵ_2, i.e.

$$E_{eff} \overset{!}{<} \min(\epsilon_1, \epsilon_2) \tag{20}$$

5.3 Flexibility - good or bad?

In order to optimize binding of the protein to the two aptamers, we wish to minimize the effective energy of binding. This can be achieved by tuning the length of the linkers so that the difference of their equilibrium extensions $x_{0,2} - x_{0,1}$ is as close to the optimum distance d as possible.
In practice it may be difficult to achieve this. In case of a large distance mismatch, the elastic energy contribution can be reduced by choosing a rather flexible linker with low k.
But wouldn't then an infinitely flexible linker ($k = 0$) be the best? That this is not the case is related to the thermal fluctuations of such a flexible linker. We know that a spring will fluctuate

around its equilibrium position with a variance $var(x) = \sqrt{\frac{k_B T}{k}}$. Thus the softer the spring the larger these fluctuations, which tend to destroy the binding complex. In other words, the energy may be minimized for small k, but the minimum will become locally very shallow, and thus easy to escape from.

This argument can also be put differently in terms of 'local concentrations' (here we follow - in a highly simplified manner - a similar argument that was previously made by Crothers & Metzger for multivalent antigen-antibody interactions[8]. In contrast to their treatment, in 1D we do not have to worry about the orientation of the binding partners). We assume that the equilibrium for bond formation is given by

$$K = S(x)dxK_0, \tag{21}$$

where $S(x)$ is the probability for one aptamer being at position x where binding is possible, and dx is the (1D) interaction volume. K_0 is the equilbrium for bond formation, when the binding partners are within the required distance and interaction volume.

For an infinitely flexible attachment, the probability is simple given by $S(x) = 1/L$ as the aptamer will have equal probability to be found anywhere in the box. For a more rigid spring,

$$S(x) = \sqrt{\frac{k}{2\pi k_B T}} e^{-k(x-x_0)^2/2k_B T}, \tag{22}$$

where we assume that the aptamer is sufficiently localized so that we can use a Gaussian approximation. The standard deviation is given by $\sigma = \sqrt{k_B T/k}$. A rough approximation would be that the probability to find the particle within the interval $[x_0 - \sigma; x_0 + \sigma]$ is $S(x) \approx 1/2\sigma = \sqrt{k/4k_B T}$, and $S(x) = 0$ outside.

Thus the ratio of the equilibrium constants will be

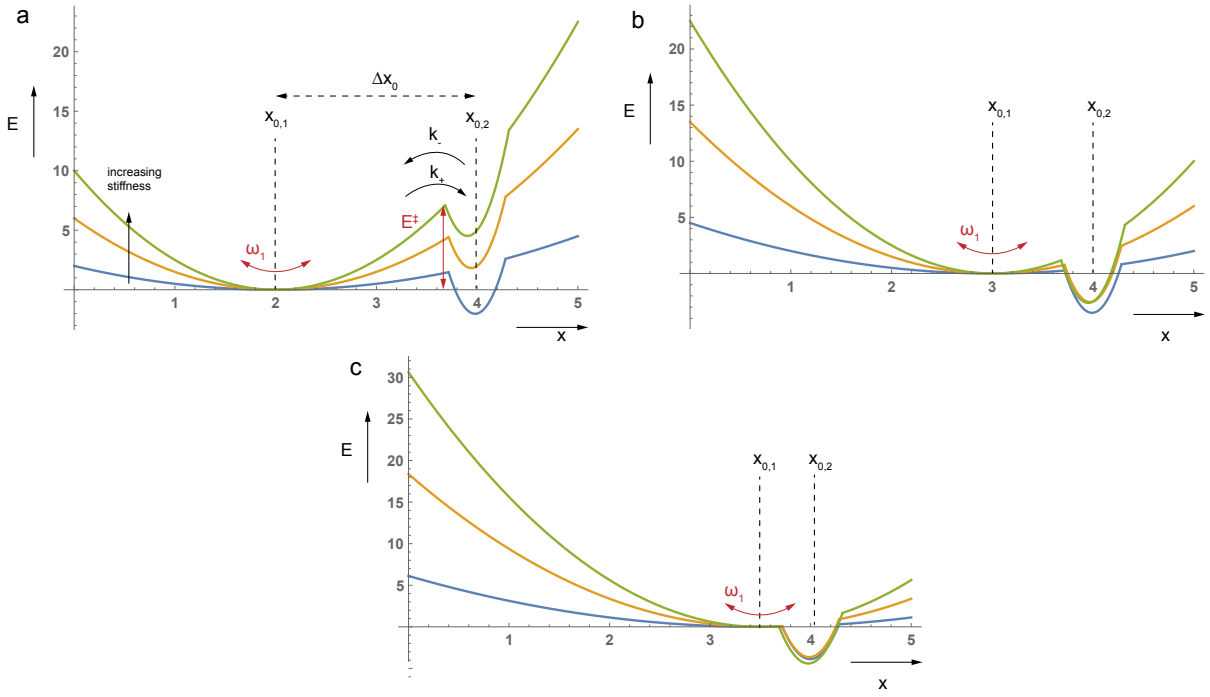
$$\frac{K_{rigid}}{K_{sloppy}} \approx \sqrt{\frac{kL^2}{2k_B T}}, \tag{23}$$

close to the x_0 and zero otherwise. This indicates that when the distances of the aptamers and the lengths of the linkers are 'right', more rigid linkers result in better binding. As a corollary, we see that it is also better when two aptamers with flexible linkers are co-localized simply in a smaller volume L.

5.4 Kinetic viewpoint

We will briefly discuss the topic of binding in the cavity also from a kinetic viewpoint. For instance, the equilibrium

$$\mathsf{A_1{-}A_2{-}P} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathsf{A_1{-}P{-}A_2} \tag{24}$$



Scheme 2. Schematic energy landscape for linker 1/aptamer 1, when aptamer 2 is attached via a rigid linker (i.e., cannot move) and the protein is bound to either of the aptamers. The spring and the binding potential are modeled as parabolic potential wells. ω_1 is the frequency of the spring's vibration, and $E^{\ddagger}$ is the effective energy barrier that has to be overcome to allow simultaneous binding of the protein to both aptamers. a: For a large length mismatch Δx_0 the energy barrier becomes larger for larger spring constant k. b, c: For smaller distances the barrier becomes small and less dependent on k.

can be described in terms of the rate constants via:

$$K_2' = \frac{k_+}{k_-} \tag{25}$$

The reaction involves stretching of the aptamer linkers 1 and 2 in order to enable binding of the protein by both aptamers. Both rates will be affected by the flexibility of the linkers. For simplicity, we will assume that linker 2 is completely rigid, but linker 1 has some flexibility.

In Scheme 2, we depict a schematic potential landscape for aptamer/linker 1 for this situation. We assume that linker 1 has an equilibrium position $x_{0,1}$ and can stretch to reach aptamer 2, where simultaneous binding to both aptamers will provide some binding energy $-\epsilon < 0$. For illustrative purposes, we here model the effective aptamer binding potential as a parabolic potential with a much shorter range than the spring potential of linker 1 (i.e., with $\xi \gg k_1$):

$$E(x) = \frac{k_1}{2}(x - x_{0,1})^2 + min(0, -\epsilon + \frac{\xi}{2}(x - x_{0,2})^2) \tag{26}$$

This potential has a cusp-like energy barrier at $x^{\ddagger} \approx x_{0,2} - \sqrt{2\epsilon/\xi}$ with height $E^{\ddagger} = \frac{k_1}{2}(\Delta x_0 - \sqrt{2\epsilon/\xi})^2$, where $\Delta x_0 := x_{0,2} - x_{0,1}$.

We further assume that the binding process can be described by a Kramers process, in which aptamer 1/linker 1 overcome the energy barrier in a diffusive process. For cusp potentials as depicted in Scheme 2, the Kramers rate takes the simple form[9]:

$$k_+ = \frac{\omega_1}{2\pi} \times e^{-E^{\ddagger}/k_B T}, \tag{27}$$

where $\omega_1 = (k_1/m)^{1/2}$ and m is the corresponding mass of the molecule. For large Δx_0 - i.e., corresponding to mismatched linker lengths - the energy barrier becomes larger with the spring constant k_1, diminishing the k_+ rate for rigid linkers. On the other hand, for well matched linker lengths $\Delta x_0 \approx 0$ the energy barrier vanishes ($E^{\ddagger} \approx 0$) and the k_+ rate is determined by the pre-exponential factor, which becomes *larger* for more rigid linkers!

For the reverse rate k_- the opposite holds true: For large Δx_0, the energy barrier for the reverse process becomes smaller for larger k_1, and thus a more rigid linker - a larger spring force on the complex - increases the off-rate. This is also in agreement with the concept of force-induced unbinding processes[10].

5.5 Quantitation of binding equilibria

5.5.1 General relationships

As mentioned above, binding of a target protein to an aptamer is described by the simple chemical equilibrium:

$$\mathrm{P} + \mathrm{A} \rightleftharpoons \mathrm{C}, \tag{28}$$

where $\mathrm{C} = \mathrm{P}{-}\mathrm{A}$ is the aptamer-protein complex. The total concentration of protein and aptamer are constant: $[P]_0 = [P] + [C]$ and $[A]_0 = [A] + [C]$ and thus the dissociation constant is given by:

$$K_d = \frac{[A][P]}{[C]} = \frac{([A]_0 - [C])([P]_0 - [C])}{[C]} \tag{29}$$

The equilibrium complex concentration is therefore given by

$$[C] = \frac{1}{2} \cdot \Big([A]_0 + [P]_0 + K_d) - \sqrt{([A]_0 + [P]_0 + K_d)^2 - 4[A]_0[P]_0}\Big) \tag{30}$$

$$= \frac{1}{2} \cdot ([A]_0 + [P]_0 + K_d) \cdot \left(1 - \sqrt{1 - \frac{4[A]_0[P]_0}{([A]_0 + [P]_0 + K_d)^2}}\right) \tag{31}$$

and the binding fraction is:

$$\theta = \frac{[C]}{[A]_0} = \frac{[A]_0 + [P]_0 + K_d}{2[A]_0} \cdot \left(1 - \sqrt{1 - \frac{4[A]_0[P]_0}{([A]_0 + [P]_0 + K_d)^2}}\right) \tag{32}$$

On the other hand

$$\theta = \frac{[C]}{[A] + [C]} = \frac{1}{1 + [A]/[[C]} = \frac{1}{1 + K_d/[P]} = \frac{[P]}{[P] + K_d} \tag{33}$$

Note that Eq. 33 gives the binding yield as a function of unbound protein, whereas Eq. 32 gives θ as a function of the experimentally controllable total aptamer and protein concentrations.
We can solve Eq. 32 for K_d, which results in:

$$K_d = \frac{(1-\theta)([P]_0 - \theta[A]_0)}{\theta} \tag{34}$$

5.5.2 Values for our experimental parameters

In order to relate our experimental results to an 'effective' binding strength, we can naively use the relations derived for aptamer binding given in the previous section. In our experiments, origami concentrations are set to $[A]_0 = 3$ nM, whereas thrombin is added at a concentration of $[P]_0 = 40$ nM. With a $K_d = 100$ nM for TBA1, we would expect a binding yield of $\theta \approx 28\%$, and for HD22 - with a reported $K_d = 0.25$ nM, we would even expect a yield of $\theta \approx 99.3\%$! In stark contrast, in experiments with only the HD22 aptamer, we found a yield of merely $\theta_{exp} \approx 10\%$, pointing to a much higher effective K_d.
Reversing the argument, if we consider $\theta_{exp} \approx 10\%$ as the measured binding yield for the HD22 aptamer and $\theta_{exp} \approx 70\%$ for the best configuration of both aptamers in the cavity, this would correspond to a decrease of "effective" K_d from ≈ 360 nM to ≈ 16 nM.

5. Supplementary references

1 Pettersen, E. F. *et al*. UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).

2 Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E. & Tulinsky, A. The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* **268**, 17651-17654 (1993).

3 Pica, A. *et al*. Through-bond effects in the ternary complexes of thrombin sandwiched by two DNA aptamers. *Nucl. Acids Res.* **45**, 461-469 (2017).

4 Munteanu, M. Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends Biochem. Sci.* **23**, 341-347 (1998).

5 Douglas, S. M. *et al*. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucl. Acids Res.* **37**, 5001-5006 (2009).

6 Tikhomirov, G., Petersen, P. & Qian, L. Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns. *Nature* **552**, 67-71 (2017).

7 Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302 (2006).

8 Crothers, D. M. & Metzger, H. The influence of polyvalency on the binding properties of antibodies. Immunochem. 9, 341-357 (1972).

9 Hänggi, P., Talkner, P. & Borkovec, M. Reaction-rate theory: fifty years after Kramers. Rev. Mod. Phys. 62, 251-341 (1990).

10 Evans, E. & Ritchie, K. Dynamic strength of molecular adhesion bonds. Biophys. J. 72, 1541-1555 (1997).