SUPPORTING INFORMATION

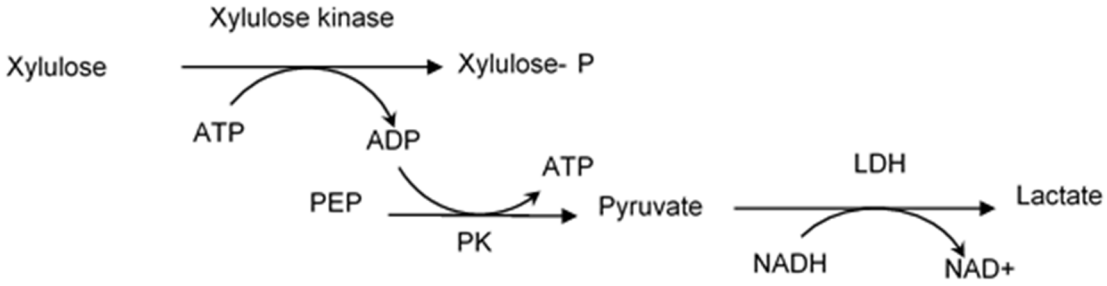
Design of Modular Protein Tags for Orthogonal Covalent Bond Formation at Specific DNA Sequences

Thang Minh Nguyen, Eiji Nakata, Masayuki Saimura, Huyen Dinh, and Takashi Morii*

Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan.

*To whom correspondence should be addressed:
Prof. Takashi Morii
Tel.: +81 774-38-3515
Fax: +81 774-38-3516
E-mail: t-morii@iae.kyoto-u.ac.jp

Table of contents



Scheme S1. Coupled enzyme assay for study the activity of XK derivatives.

Table S1. Equilibrium disassociation constants (K_D) for the complexes of modular adaptors with ODNs.

Modular Adaptors	K_D (nM)	
	ODN-AZ	ODN-ZF
ZF-SNAP	> 1000	63 ± 18*
AZ-SNAP	61 ± 30	> 1000
AZ-CLIP	65 ± 2	> 1000
AZ-Halo	118 ± 6	> 1000

*The value from ref 8.

Table S2. Kinetic parameters for the cross-linking reaction between 5'-^{32}P-end-labeled ODN derivatives and modular adaptors.

Rate constant k ($M^{-1}s^{-1}$)		**Modular adaptors**			
		ZF-SNAP	**AZ-SNAP**	**AZ-CLIP**	**AZ-Halo**
ODN derivatives	**ODN-ZF-BG**	**(3.8 ± 1.2) × 10^5***	(9.5 ± 0.6) × 10^4	62 ±13	n.d.
	ODN-AZ-BG	(2.6 ± 0.2) × 10^4	**(9.1 ± 2.4) × 10^5**	77 ± 18	n.d.
	ODN-AZ-BC	76 ± 0.8	304 ± 9	**(5.0 ± 0.7) × 10^5**	n.d.
	ODN-AZ-CH	55 ± 11	147 ± 25	17 ± 24	**(4.0 ± 1.0) × 10^5**

*The value from ref 8. n.d. : not detected.

Table S3. The staple strands including zif268 and AZP4 binding sites with substrates modified T (T^{BG}, T^{CH}, T^{BC} for T modified with BG, BC, and CH, respectively). The zif268 and AZP4 binding sites on the staple strands were colored in red and blue, respectively.

Oligo name	Sequence (from 5' to 3')
5j-AZ-BC	GCTGAGAGACAGACAA CTTATGCCACGTAGCGTT^{BC}TTCGCTACGTGGCATAAG TATTTTTAACGCTCATGGAAATA
8E-AZ-CH	CTACTAA CTTATGCCACGTAGCGTT^{CH}TTCGCTACGTGGCATAAG TGACCATTAGATACAACGAGTAGA
11D-ZF-BG	AACAGGTC CTTACGCCCACGCGCG TT^{BG}TT CGCGCGTGGGCGTAAG GAACCAGACCGGAAGATTCGAGC
24D-AZ-BC	GGACAGAT CTTATGCCACGTAGCGTT^{BC}TTCGCTACGTGGCATAAG AAATTGTGTCGAAATCTGTATCAT

Table S4. Total numbers and yields of the DNA scaffolds assembled with the modular adaptors or modular adaptor fused enzymes analyzed by AFM.

DNA scaffold	Modular adaptor derivatives	Number of well-formed DNA scaffold	Numbers and yields of the modified cavities			Coassembly yield	AFM image
			Cavity I	Cavity II	Cavity III		
I-1AH/II-1ZS/III-1AC	AZ-Halo	151	145 **[96%]**	1 **[1%]**	4 **[3%]**	n.a.	Figure 3c
	ZF-SNAP	179	1 **[1%]**	166 **[93%]**	5 **[3%]**	n.a.	Figure 3d
	AZ-CLIP	147	1 **[96%]**	5 **[3%]**	137 **[93%]**	n.a.	Figure 3e
I-1AH/II-1ZS/III-1AC	AZ-Halo	157	155 **[99%]**	5 **[3%]**	6 **[4%]**	n.a.	Figure 4 1st step
	AZ-Halo and ZF-SNAP	129	125 **[97%]**	120 **[93%]**	4 **[3%]**	118 **[91%]**	Figure 4 2nd step
	AZ-Halo, ZF-SNAP and AZ-CLIP	101	100 **[99%]**	92 **[91%]**	100 **[99%]**	92 **[91%]**	Figure 4 3rd step
I-1AH/II-1ZS/III-1AC	AZ-Halo, ZF-SNAP and AZ-CLIP	153	153 **[100%]**	146 **[95%]**	151 **[98%]**	142 **[93%]**	Figure 5b
	AZ-Halo, ZF-SNAP and AZ-CLIP	153	149 **[97%]**	139 **[91%]**	143 **[93%]**	133 **[87%]**	Figure 5c
I-1AH/II-1ZS/III-1AC	AH-XK	142	132 **[93%]**	1 **[1%]**	3 **[2%]**	n.a.	Figure S9b
	ZS-XR	174	2 **[1%]**	162 **[93%]**	1 **[1%]**	n.a.	Figure S9c
	AC-XK	163	1 **[1%]**	3 **[2%]**	147 **[90%]**	n.a.	Figure S9d
I-4XR/II-4XDH/III-1XK	ZS-XR, G-XDH and AC-XK	151	150 **[99%]**	126 **[83%]**	147 **[95%]**	125 **[83%]**	Figure 6d
I-4XR/I-4XDH/I-1XK	ZS-XR, G-XDH and AC-XK	167	167 **[100%]**	10 **[6%]**	6 **[4%]**	144 **[86%]**	Figure 6e

n.a. : not applicable

Table S5. Kinetic parameters for XK, AC-XK and AH-XK for the phosphorylation of xylulose

	XK	AC-XK	AH-XK
K_m for xylulose (μM)	203 ± 20	180 ± 21	223 ± 25
k_{cat} (s^{-1})	257 ± 23	210 ± 15	240 ± 21
k_{cat}/K_m ($mM^{-1}.s^{-1}$)	1271 ± 55	1146 ± 194	1030 ± 84

Table S6. Quantitation of the cofactors by HPLC

On the scaffold	Scaffold name	In bulk solution	ATP(μM)	ADP(μM)	NADH(μM)	NAD^+(μM)
ZS-XR	I-4XR	G-XDH AC-XK	587 ± 18	406 ± 21	73 ± 51	1910 ± 45
ZS-XR G-XDH	I-4XR/II-4XDH	AC-XK	557 ± 37	442 ± 37	132 ± 26	1867 ± 37
ZS-XR G-XDH AC-XK	I-4XR/II-4XDH/III-1XK	-	522 ± 53	477 ± 52	116 ± 15	1883 ± 11
ZS-XR G-XDH	I-4XR/I-4XDH	AC-XK	510 ± 24	497 ± 29	81 ± 5	1921 ± 12
ZS-XR G-XDH AC-XK	I-4XR/I-4XDH/I-1XK	-	459 ± 8	542 ± 10	102 ± 6	1897 ± 10

Table S7. Amino acid sequences and molecular weights of modular adaptor derivatives.

Derivatives	Amino acid sequence	calculated molecular weight (Da)
AZ-SNAP	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPFQCRICMRNFSRSDSLTRHIRTHTGEKPFAC DICGRKFAESDNRKTHTKIHTGEKEFGGSGGSMDKDCEMKRTTLDSPLGKLELSGCEQGLHEIIFLGKG TSAADAVEVPAPAAVLGGPEPLMQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLLKVV KFGEVISYSHLAALAGNPAATAAVKTALSGNPVPILIPCHRVVQGDLDVGGYEGGLAVKEWLLAHEGHR LGKPGLGGGSGGSHHHHHH	32,181
AZ-CLIP	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPFQCRICMRNFSRSDSLTRHIRTHTGEKPFAC DICGRKFAESDNRKTHTKIHTGEKEFGGSGGSMDKDCEMKRTTLDSPLGKLELSGCEQGLHRIIFLGKG TSAADAVEVPAPAAVLGGPEPLIQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLLKVV KFGEVISESHLAALVGNPAATAAVNTALDGNPVPILIPCHRVVQGDSDVGPYLGGLAVKEWLLAHEGHR LGKPGLGGGSGGSHHHHHH	32,196
AZ-Halo	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPFQCRICMRNFSRSDSLTRHIRTHTGEKPFAC DICGRKFAESDNRKTHTKIHTGEKEFGGSGGSMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLF LHGNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHD WGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPM GVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPG VLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGEPTTEDLYFQSDNAIA HHHHHH	47,768
AC-XK	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPFQCRICMRNFSRSDSLTRHIRTHTGEKPFAC DICGRKFAESDNRKTHTKIHTGEKEFGGSGGSMDKDCEMKRTTLDSPLGKLELSGCEQGLHRIIFLGKG TSAADAVEVPAPAAVLGGPEPLIQATAWLNAYFHQPEAIEEFPVPALHHPVFQQESFTRQVLWKLLKVV KFGEVISESHLAALVGNPAATAAVNTALDGNPVPILIPCHRVVQGDSDVGPYLGGLAVKEWLLAHEGHR LGKPGLGGGSGGSMLCSVIQRQTREVSNTMSLDSYYLGFDLSTQQLKCLAINQDLKIVHSETVEFEKDL PHYHTKKGVYIHGDTIECPVAMWLEALDLVLSKYREAKFPLNKVMAVSGSCQQHGSVYWSSQAESLLEQ LNKKPEKDLLHYVSSVAFARQTAPNWQDHSTAKQCQEFEECIGGPEKMAQLTGSRAHFRFTGPQILKIA QLEPEAYEKTKTISLVSNFLTSILVGHLVELEEADACGMNLYDIRERKFSDELLHLIDSSSKDKTIRQK LMRAPMKNLIAGTICKYFIEKYGFNTNCKVSPMTGDNLATICSLPLRKNDVLVSLGTSTTVLLVTDKYH PSPNYHLFIHPTLPNHYMGMICYCNGSLARERIRDELNKERENNYEKTNDWTLFNQAVLDDSESSENEL GVYFPLGEIVPSVKAINKRVIFNPKTGMIEREVAKFKDKRHDAKNIVESQALSCRVRISPLLSDSNASS QQRLNEDTIVKFDYDESPLRDYLNKRPERTFFVGGASKNDAIVKKFAQVIGATKGNFRLETPNSCALGG CYKAMWSLLYDSNKIAVPFDKFLNDNFPWHVMESISDVDNENWDRYNSKIVPLSELEKTLIGGSGGSHH HHHH	100,901
AH-XK	MKTGEKRPYACPVESCDRRFSQSNDLTRHIRIHTGQKPFQCRICMRNFSRSDSLTRHIRTHTGEKPFAC DICGRKFAESDNRKTHTKIHTGEKEFGGSGGSMAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPVLF LHGNPTSSYVWRNIIPHVAPTHRCIAPDLIGMGKSDKPDLGYFFDDHVRFMDAFIEALGLEEVVLVIHD WGSALGFHWAKRNPERVKGIAFMEFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPM GVVRPLTEVEMDHYREPFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPG VLIPPAEAARLAKSLPNCKAVDIGPGLNLLQEDNPDLIGSEIARWLSTLEISGEPTTEDLYFQSDNAIA MLCSVIQRQTREVSNTMSLDSYYLGFDLSTQQLKCLAINQDLKIVHSETVEFEKDLPHYHTKKGVYIHG DTIECPVAMWLEALDLVLSKYREAKFPLNKVMAVSGSCQQHGSVYWSSQAESLLEQLNKKPEKDLLHYV SSVAFARQTAPNWQDHSTAKQCQEFEECIGGPEKMAQLTGSRAHFRFTGPQILKIAQLEPEAYEKTKTI SLVSNFLTSILVGHLVELEEADACGMNLYDIRERKFSDELLHLIDSSSKDKTIRQKLMRAPMKNLIAGT ICKYFIEKYGFNTNCKVSPMTGDNLATICSLPLRKNDVLVSLGTSTTVLLVTDKYHPSPNYHLFIHPTL PNHYMGMICYCNGSLARERIRDELNKERENNYEKTNDWTLFNQAVLDDSESSENELGVYFPLGEIVPSV KAINKRVIFNPKTGMIEREVAKFKDKRHDAKNIVESQALSCRVRISPLLSDSNASSQQRLNEDTIVKFD YDESPLRDYLNKRPERTFFVGGASKNDAIVKKFAQVIGATKGNFRLETPNSCALGGCYKAMWSLLYDSN KIAVPFDKFLNDNFPWHVMESISDVDNENWDRYNSKIVPLSELEKTLIGGSGGSHHHHHH	116,473

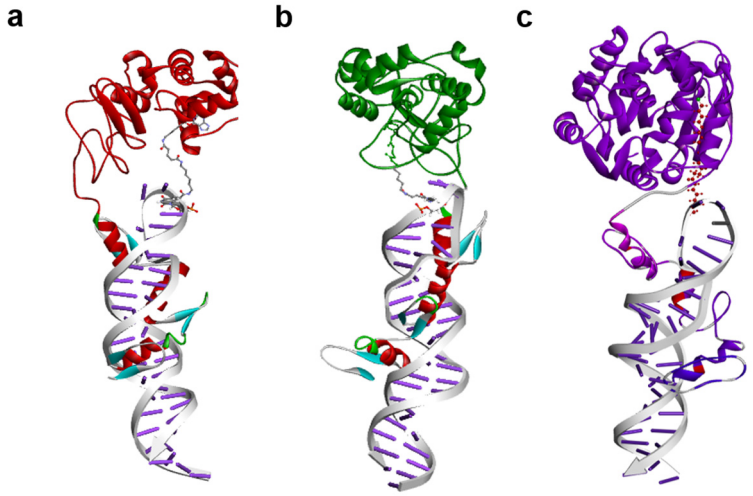


Figure S1. Possible models for the complex of (a) AZ-SNAP with ODN-AZ-BG, (b) AZ-CLIP with ODN-AZ-BC, (c) AZ-Halo with ODN-AZ-CH, respectively, based on the crystal structure of the complex between zif268 and ODN (PDB ID : 1ZAA) and the complex of SNAP-tag with BG (PDB ID:3KZY), and Halo-tag with CH (PDB ID : 1CQW). The models were constructed by using Discovery Studio (version 3.1, Accelrys Inc.).

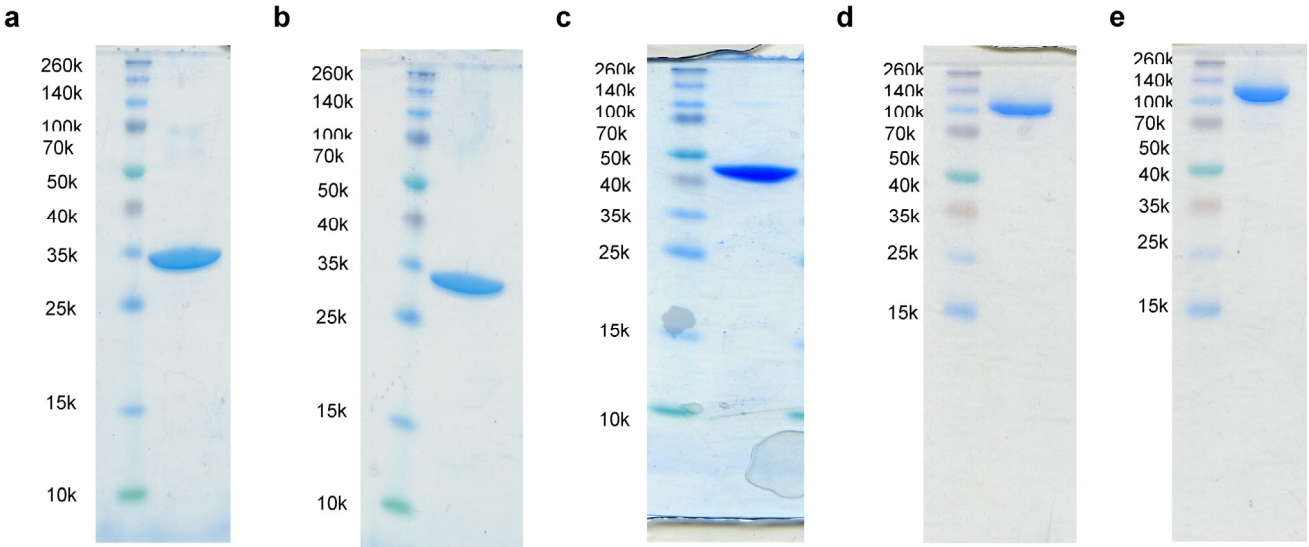


Figure S2. SDS-PAGE analyses of purified modular adaptors and modular adaptor-fused enzymes. (a) AZ-SNAP, (b) AZ-CLIP, (c) AZ-Halo, (d) AC-XK, (e) AH-XK. Amino acid sequences and molecular weights of these proteins are shown in Table S7.

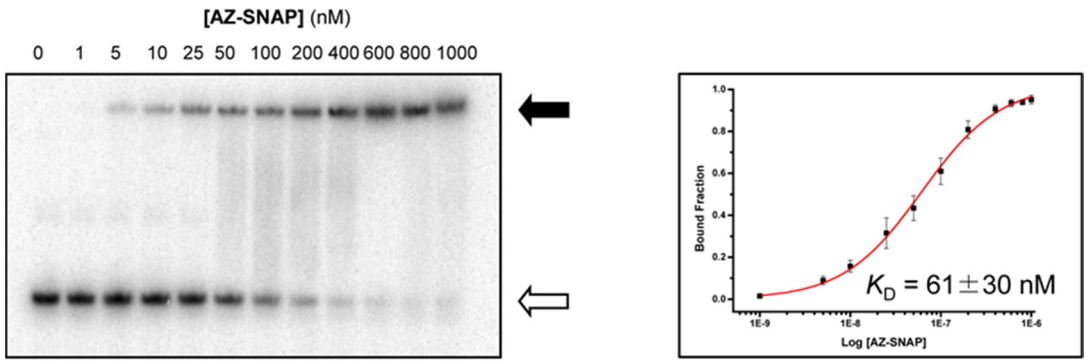


Figure S3. An autoradiogram shows the electrophoretic mobility shift titration of AZ-SNAP to ODN-AZ in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM $MgCl_2$, 1mM DTT, 1μM $ZnCl_2$ 0.02 % Tween 20, and 200 nM BSA at ambient temperature (left). Open arrow and filled arrow denote free ODN-AZ and AZ-SNAP bound ODN-AZ, respectively. A semilogarithmic plot shows the fractions of 5'-^{32}P-labeled ODN-AZ bound to AZ-SNAP (right). An equilibrium dissociation constant obtained is listed in Table S1.

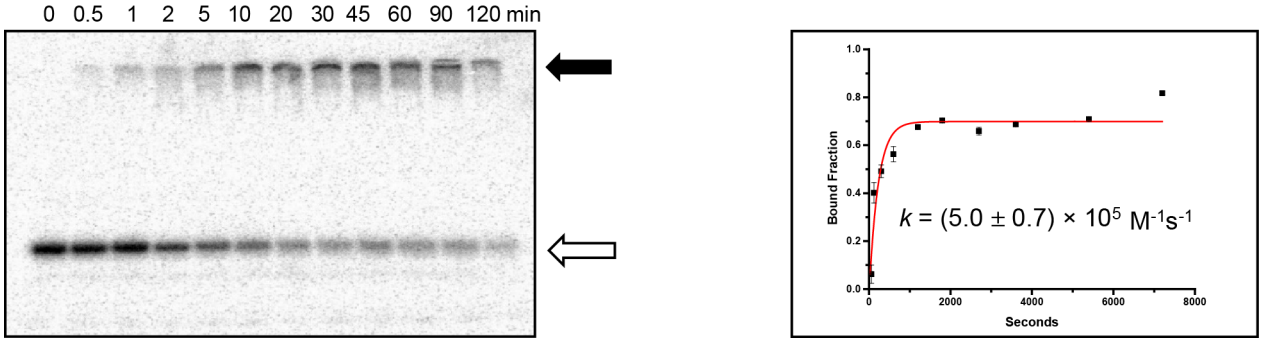


Figure S4. An autoradiogram shows denaturing gel electrophoretic analysis of the cross-linking reactions of 5'-^{32}P-end-labeled ODN-AZ-BC with AZ-CLIP (10 nM) (left). Open arrow and filled arrow denote ODN-AZ-BC and AZ-CLIP bound ODN-AZ-BC, respectively. A time-course plot for the crosslinking reaction of ODN-AZ-BC and AZ-CLIP to obtain the rate constant (***k***) (right). The determined rate constant is listed in Table S2.

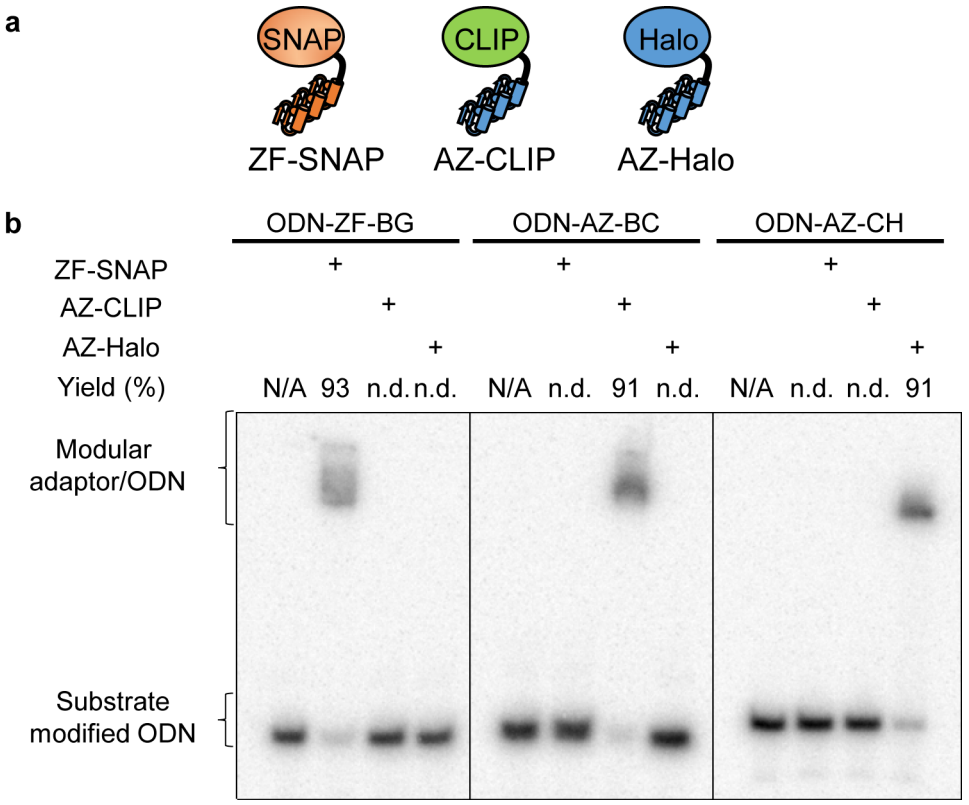


Figure S5. (a) Combination of modular adaptors to validate their orthogonal reactions to target sites. (b) Denaturing PAGE analyses of the crosslinking reaction by modular adaptors (ZF-SNAP, AZ-CLIP and AZ-Halo) and the substrate modified ODN (ODN-ZF-BG, ODN-AZ-BC, and ODN-AZ-CH), respectively. Each 5'-^{32}P-end-labled ODN (ODN-ZF-BG, ODN-AZ-BC or ODN-AZ-CH) was incubated with a modular adaptor (100 nM : ZF-SNAP, AZ-CLIP or AZ-Halo) for 30 min in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM $MgCl_2$, 1 mM DTT, 1 μM $ZnCl_2$, 0.02% Tween 20, 200 nM BSA and 100 nM calf thymus DNA at ambient temperature. N/A : not applicable, n.d: not detectable.

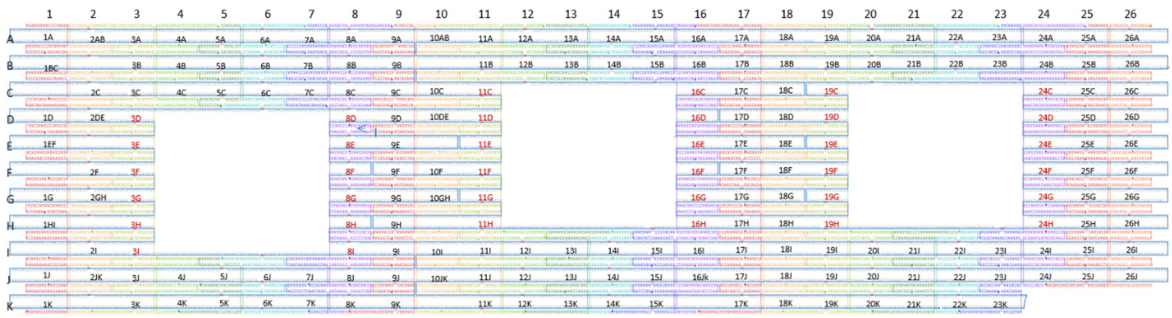


Figure S6. An illustration shows the shape and addresses of the DNA origami scaffold used in this study. Nucleotide sequences of all staple strands were shown in previous report[11] and Table S3.

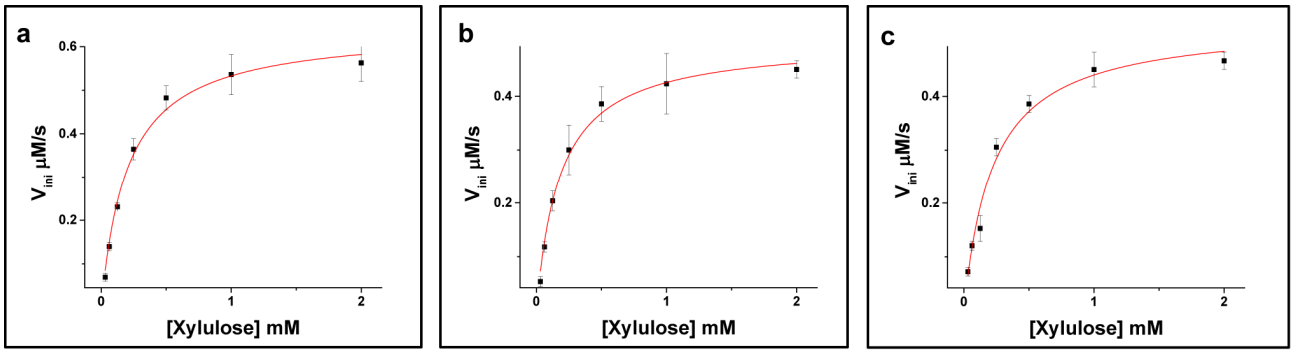


Figure S7. Michaelis-Menten plots for the phosphorylation of xylulose by (a) XK, (b) AC-XK, or (c) AH-XK. Enzymatic reactions with the same concentration of enzyme (2 nM) were performed at 25 °C in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM $MgCl_2$, 1 mM DTT, 1 µM $ZnCl_2$, 0.02% Tween20, 100 mM NaCl, 1.1 mM ATP, 0.2 mM NADH, 2.3 mM Phosphoenolpyruvate, 4.8 U/ml PK and 4.5 U/ml LDH, and the indicated concentration of xylulose (0.03 to 2mM). The reaction was started by an addition of xylulose.

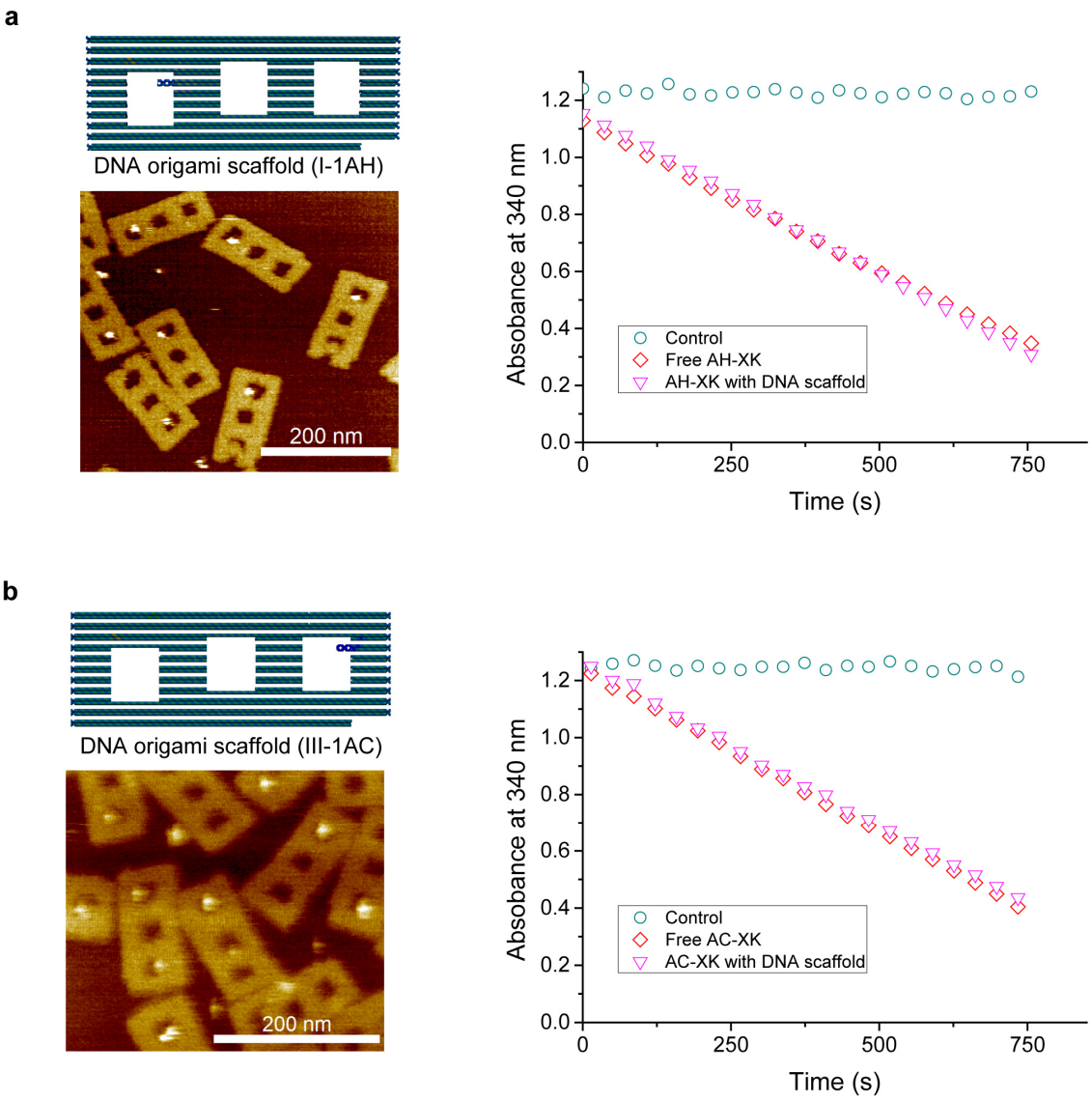


Figure S8. Effect of the DNA scaffold on the catalytic activity of XK. (a) An illustration of the DNA scaffold holding a binding site modified by CH (for AH-XK) and an AFM image of AH-XK bound on the DNA scaffold (left). Time-course profiles of the oxidation of NADH monitored by absorbance at 340 nm (right). (b) An illustration of the DNA scaffold holding a binding site modified by BC (for AC-XK) and an AFM image of AC-XK bound on the DNA scaffold (left). Time-course profiles of the oxidation of NADH monitored by absorbance at 340 nm (right).

Enzyme assay was carried out in a solution containing AH-XK or AC-XK in the absence (Free AH-XK or Free AC-XK, red diamonds) or presence of the DNA scaffold (AH-XK with the DNA scaffold or AC-XK with the DNA scaffold, pink triangles). Prior to the assay for XK activity, 2 nM AH-XK (or AC-XK) was incubated with or without 10 nM DNA scaffold for 30 minutes on ice in a buffer (pH 7.6) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM $MgCl_2$, 1 mM DTT, 1 μM $ZnCl_2$, and 0.02% Tween20. Enzyme reactions were carried out with the same enzyme concentration (1 nM) at 25 °C in a buffer (pH 7.6) containing 100 mM NaCl, 1.1 mM ATP, 0.2 mM NADH, 2.3 mM PEP, 4.8 U/ml PK and 4.5 U/ml LDH. The reactions were started by the addition of 2 mM xylulose.

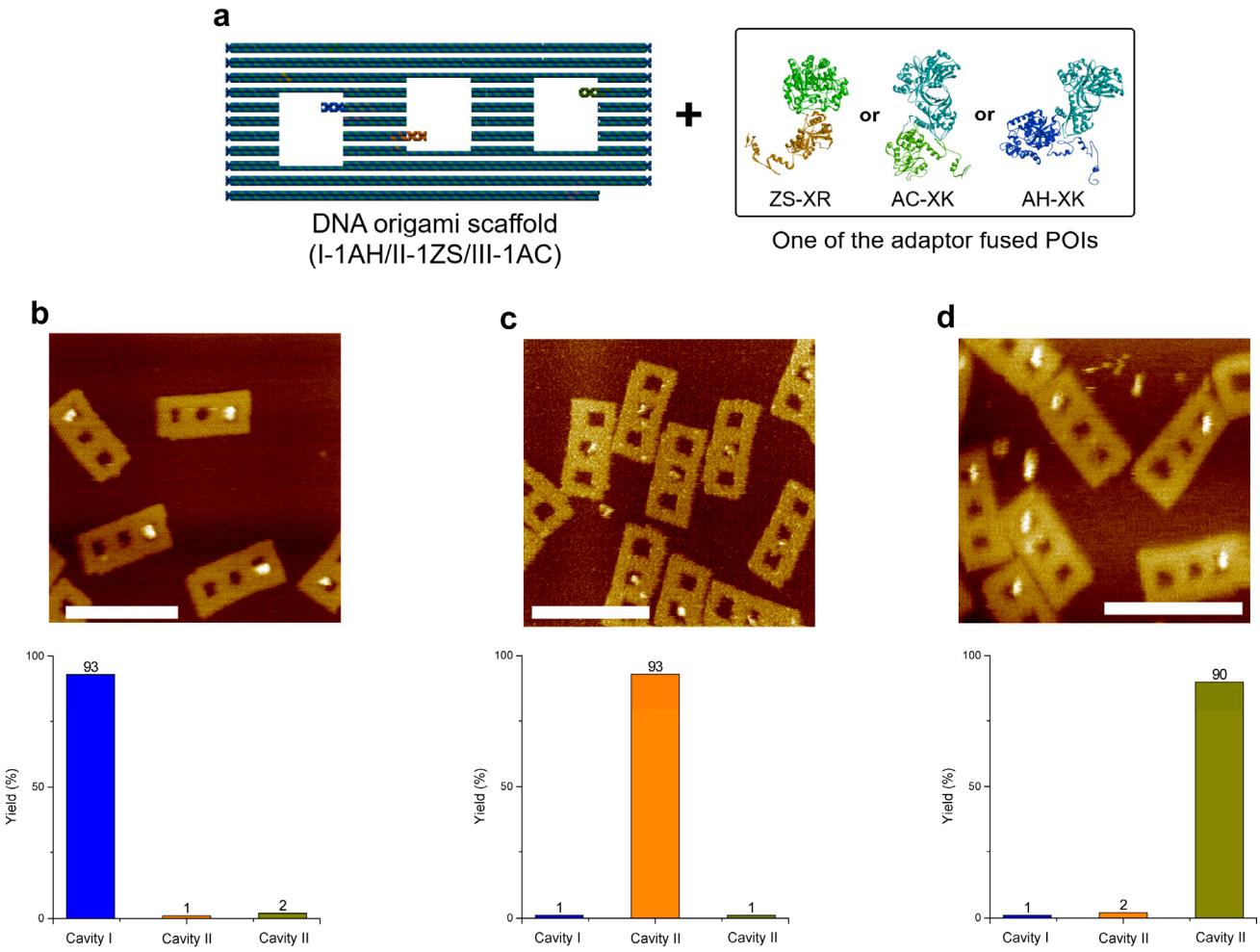


Figure S9. (a) An illustration of orthogonal assembly of three adaptor fused POIs (ZS-XR, AC-XK, AH-XK). (b-d) AFM images and binding yield of (b) AH-XK, (c) ZS-XR, and (d) AC-XK on the DNA scaffold (I-1AH/II-1ZS/III-1AC), respectively. 5 nM of DNA scaffold was incubated with 5 molar equivalent of AH-XK, ZS-XR or AC-XK for 30 minutes on ice in a buffer (pH 8.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM $MgCl_2$, 1 mM DTT, 1 μM $ZnCl_2$ and 0.02% Tween20. Crosslinking yields were estimated by counting the number of cavities occupied by the modular adaptors (Table S7). Scale bars represent 200 nm.

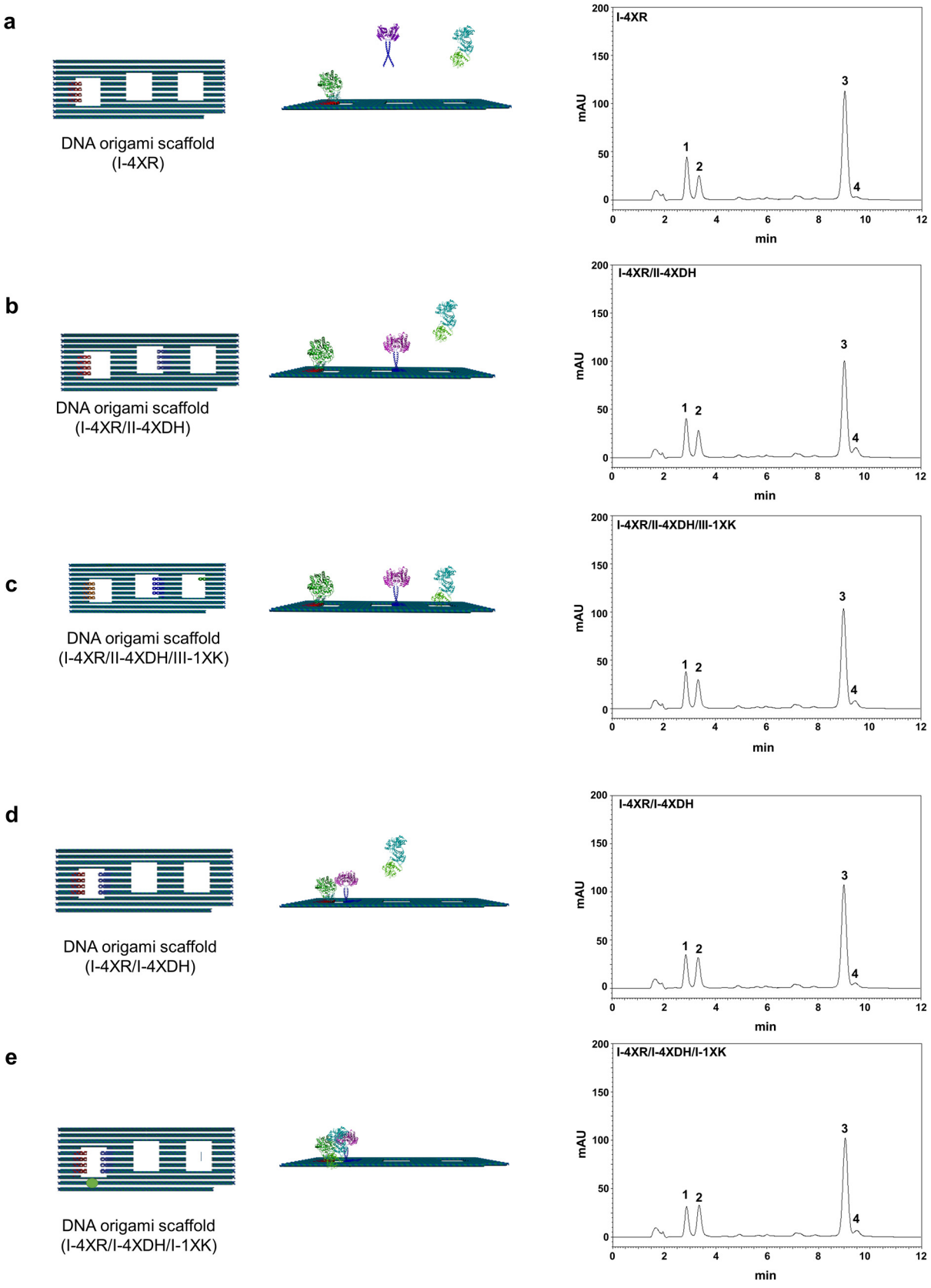


Figure S10. HPLC analyses (detected by UV at 260 nm) of cofactors in the three enzyme cascade reaction. Peaks 1, 2, 3 and 4 indicated ATP, ADP, NAD^+ and NADH, respectively. The analyses were carried out with the reaction mixture incubated for 24 hours (a) with ZS-XR located on the scaffold (I-4XR) and G-XDH and

AC-XK in bulk solution, (b) with ZS-XR and G-XDH located on the scaffold (I-4XR/I-4XDH) and AC-XK in bulk solution, (c) ZS-XR, G-XDH and AC-XK located on the scaffold (I-4XR/I-4XDH/I-1XK), (d) with ZS-XR and G-XDH located on the scaffold (I-4XR/II-4XDH) and AC-XK in bulk solution, and (e) ZS-XR, G-XDH and AC-XK located on the scaffold (I-4XR/II-4XDH/III-1XK), respectively. Each reaction mixture for (a)-(e) contained 26 nM ZS-XR, 26 nM G-XDH and 6.5 nM AC-XK in a buffer (pH 7.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM $MgCl_2$, 1 mM DTT, 1 μM $ZnCl_2$, 0.02% Tween20, 2 mM NADH and 1 mM ATP. Reaction was started by an addition of 200 mM xylose.

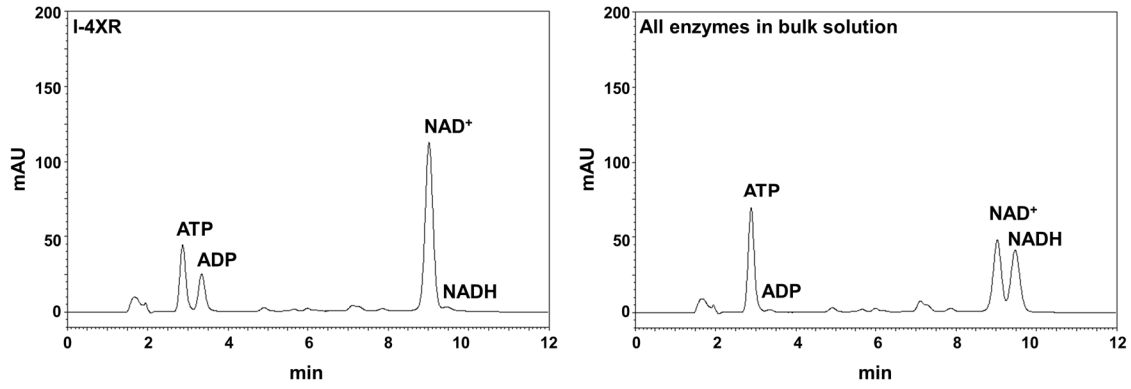


Figure S11. Comparison of the system with the first enzyme XR being loaded on the scaffold (I-4XR) and the one with all enzymes in bulk solution. HPLC analyses (detected by UV at 260 nm) of cofactors in the three-enzyme cascade reaction. The analyses were carried out with the reaction mixture incubated for 24 hours (a) with ZS-XR located on the scaffold (I-4XR) and G-XDH and AC-XK in the bulk solution and (b) ZS-XR, G-XDH and AC-XK in the bulk solution, respectively. Each reaction mixture contained 26 nM ZS-XR, 26 nM G-XDH and 6.5 nM AC-XK in a buffer (pH 7.0) containing 40 mM Tris-HCl, 20 mM acetic acid, 12.5 mM $MgCl_2$, 1 mM DTT, 1 μM $ZnCl_2$, 0.02% Tween20, 2 mM NADH and 1 mM ATP. Reaction was started by an addition of 200 mM xylose.

Discussion : In the cascade reaction by the three enzymes, the system with the first enzyme XR being loaded on the scaffold (I-4XR) was more efficient than the reaction system with all enzymes being free in the solution as shown in Figure S11, where "I-4XR" produced more ADP than "All enzymes in bulk solution". The result indicated that XR could be more stable on the DNA scaffold than in the bulk solution because the consumption of NADH was also lower in the case of “All enzymes in bulk solution”. The stability of XDH and XK was similar on the DNA scaffold and in the bulk solution (data not shown). Using the system with all enzymes being free in solution as a control experiment would overestimate the actual spatial effect of assembly for three enzymes. Therefore, we considered that the system with XR being loaded on the DNA scaffold would be an appropriate control experiment.