Supporting Information for

Design of DNA-Conjugated Polypeptide-based Capture Probes for the Anchoring of Proteins to DNA Matrices

Ryan M. Schweller, Pamela E. Constantinou, Nicholas W. Frankel, Priyanka Narayan and Michael R. Diehl*

Department of Chemistry, and Department of Bioengineering, Rice University, 6100 Main Street, MS 142 Houston, TX 77005.

Supporting Information

List of Contents

Detailed Materials and Methods
Table S1. List of Nucleotide Sequences Used for Conjugation and Self-Assembly
Table S2. List of Amino Acid Sequences for Artificial Proteins and Leucine Zipper Fusions
Figure S1. MALDI-TOF Spectrum of Artificial Proteins
Figure S2. Chromatograph and SDS-PAGE Gel of Artificial Protein-DNA Conjugates
Figure S3. Full Non-Denaturing PAGE Gel for Multi-Protein Assembly on a DNA Scaffold
Figure S4. 8 Hour Time Point Array and Corresponding Profile Plot

MATERIALS AND METHODS

General. All reagents were purchased from Sigma-Aldrich unless otherwise noted. All oligonucleotides were purchased from IDT DNA. The Alexa647-anti-green fluorescent protein (GFP) was purchased from Molecular Probes. Anti-GST antibodies were purchased from GE Healthcare and labeled in house using a Cy3 mono-reactive dye pack (GE Healthcare) and the resulting dye to antibody ratio (6 to 1) was determined according to the manufacturer's protocols.

All cloning procedures were carried out in the *E. coli* strain XL1-Blue (Stratagene). Expression was performed using BL21 (DE3) strains (Novagen). The sequences of the plasmids produced from each cloning step were verified using restriction mapping and DNA sequence analysis.

Primers sequences used to construct synthetic genes were designed using a web-based software tool (DNAWorks) (*1*). DNA sequences used for conjugation, surface-functionalization, and to form molecular scaffolds based on single DNA duplicies were designed using a program called SEQUIN (*2*). This software was used to minimize sequence repeats in the different strands used for conjugation and spotting, as well as those incorporated into the DNA scaffolds. All oligonucleotides were individually purified using denaturing PAGE protocols.

Plasmid Construction. Genes encoding for each domain of the artificial proteins including the basic portion of the zipper complex (Z_R) and the elastic 'mid-block domain' or ELS fragment (VPGVG VPG**S**G VPGVG VPG**S**G VPGVG) were constructed from synthetic oligonucleotides through PCR. During this process a C-terminal cysteine and a stop codon was introduced into the gene. The resulting fragments were directionally ligated into a pQE60 vector (Qiagen). This gene was then polymerized stepwise into larger Z_R-$(ELS)_2$-C, Z_R-$(ELS)_4$-C and Z_R-$(ELS)_6$-C constructs using the recursive directional ligation (RDL) procedure described in reference (*3*) with slight modifications. At each intermediate cloning step, two gene fragments were isolated from each vector by digestion with *Nco*I and *Bgl*I or *PflM*I and *Hind*III restriction endonucleases. These two fragments were then simultaneously inserted directionally into the *Nco*I and *Hind*III sites of a new pQE60 vector.

The recombinant target proteins containing Z_E fusions GST-Z_E, GFP-Z_E, calmodulin-Z_E (CAM-Z_E), and PSE-4-Z_E were all prepared using standard cloning procedures. Each of these genes contained a C-terminal Z_E and 6xHistidine (6xHis) for purification.

Protein Expression and Purification. To avoid proteolytic degradation of the Z_R fragment of the artificial proteins, and to facilitate Ni^{2+}-NTA purification, Z_R-$(ELS)_6$-C was co-expressed with a Z_E-6xHis gene that was created in pQE60 and then inserted into the *Nhe*I sites of a pREP4 vector (Qiagen). Both vectors contained the inducible, bacteriophage T5 promoter of the pQE plasmid. For expression, cells were grown in terrific broth (TB; 24 g of casein hydrolysate, 12 g of yeast extract, 4 mL glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4), and antibiotics (50 mg/L ampicillin and 35 mg/L kanamycin). The cells were cultured at 37 ˚C until an OD_{600} of 1 was reached. Afterwards, expression was induced for 5 hours by adding 1 mM IPTG. Cells were then pelleted (5,000 RPM and 4 ˚C) an resuspended in lysis buffer (8 M urea, 10 mM Tris-HCl, 100 mM NaH_2PO_4, pH 8.0) and allowed to lyse overnight at 4 ˚C with vigorous stirring. The cell lysate was clarified by centrifugation and the supernatant was applied to a column containing Ni^{2+}-NTA resin. After washing with lysis buffer, the artificial protein was eluted using a buffer containing 6 M guanidine-HCl, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 8.0. This buffer disassociates the Z_E/Z_R complex, allowing the Z_R-$(ELS)_6$-C to be eluted while the majority of the Z_E-6xHis proteins remain on the column. To remove excess Z_E-6xHis, the eluent was run through a second Ni^{2+}-NTA column. The flow through of this column was dialyzed with frequent water exchanges for 3 days and lyophilized. The protein purity was verified using SDS-PAGE and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectroscopy. The MALDI-TOF of the polymers gave peaks at 18652 m/z and 9323 m/z corresponding to the M^+ and M^{+2} polymer species, respectively. The Z_E-6xHis polypeptide was not present at detectable levels using either SDS-PAGE or MALDI-TOF analyses.

Cells expressing the target proteins were cultured similarly to those used for artificial protein production. However, expression was induced overnight at 23 ˚C. Purification of the

target proteins were carried out using standard native lysis and Ni^{2+}-NTA purification protocols with the exception of the PSE-4-Z_E. Cells containing these proteins lysed using an osmotic shock procedure that allows the outer membrane of cells to be broken using an osmotic shock buffer (30% sucrose, 1 mM EDTA, 30 mM Tris, pH 8.0). Protein expression and purification was verified using SDS-PAGE analysis while protein concentrations were determined using the Bradford method. For some experiments, clarified lysates were saved, aliquoted, and stored at -80 ˚C for experiments with unpurified proteins.

Artificial Protein-DNA conjugation. Artificial proteins were labeled with DNA as described in reference (*4*) with slight modifications. Briefly, concentrated stock solutions of Z_R-$(ELS)_6$-C (10 mg/mL) were prepared by dissolving the lyophilized polymers in 8 M Urea, 100 mM NaH_2PO_4, 10 mM Tris, pH 7.2. Disulfide bonds between polymers were then reduced by adding 400 mM tris(2-carboxyethyl)phosphine (TCEP) and incubated at 37 ˚C for 1.5 hours. Simultaneously, amine-labeled oligonucleotides, 20 bp in length, were reacted for one hour at 37 ˚C with the heterobifunctional crosslinker, sulfosuccinimidyl 4-[*N*-maleimidomethyl]cyclohexane-1-carboxylate (sulfo-SMCC). Afterwards, the reduced artificial protein and the DNA solutions were passed through two separate NAP-5 desalting columns, combined, and incubated for 2 hours at room temperatures and then overnight at 4 ˚C. The following day, the Z_R-$(ELS)_6$-C-ssDNA conjugates were isolated using a HiTrapQ ion-exchange column and fast protein liquid chromatography (FPLC). The peak fractions containing the conjugate were pooled and dialyzed with frequent water exchanges for 3 days and lyophilized. The polymers were then redissolved in Tris-EDTA (TE) buffer (pH 7.4) at a concentration of approximately 35 μM. The actual conjugate stock concentrations were determined from the OD_{260} using an extinction coefficient

from the appended ssDNA's extinction coefficient (ε_{260} = 185,000 M^{-1} cm^{-1}). Conjugate stock solutions were aliquoted and stored at -20 ˚C until needed.

DNA-Templated Protein Microarray Fabrication. DNA microarrays were fabricated by printing 5'-amine modified oligonucleotides on silyated (amine-aldehyde) functionalized slides (CEL Associates Inc). Oligonucleotide stocks were diluted appropriately with spotting buffer (40% glycerol in PBS, pH 6.6). For the single protein arrays, the arrays were printed in 90, 50, and 10 μM concentrations and a control strand was printed at 90 μM. The multi-protein arrays were all printed at 50 μM. A custom built microarrayer outfitted with SMP3 pins (Telechem Int.) was used for all printing. After printing, the slides were placed in a humid chamber for 45 minutes, and then allowed to dry outside the chamber for 10 minutes. The remaining free aldehyde groups were quenched in a solution of 25% ethanol and 0.25% $NaBH_4$ in PBS for 5 minutes. The slides were then thoroughly rinsed with milli-Q water (Millipore) and blocked using pre-hybridization buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA, 1% Tween 20, 0.1 mg/mL denatured herring sperm DNA, 0.1 mg/mL BSA, 3% dry milk) with gentle agitation for 2 hours. The arrays were then washed 3 times for 10 minutes each in wash buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA, 0.01% Tween 20), thoroughly rinsed with milli-Q water and dried under a nitrogen stream.

Artificial protein-DNA conjugates were used to direct the assembly of target proteins onto the DNA-spotted microarrays by first pre-incubating the target Z_E-fusion proteins individually with the artificial protein conjugates. A three-fold molar excess of target protein was typically maintained during these incubations. A reaction chamber was prepared by affixing a Gene Frame (AB Gene) to the microarrays. Mixtures of artificial proteins and target proteins

were brought to a final volume of 100 μL with hybridization buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM EDTA, 0.01% Tween 20, 0.1 mg/mL denatured herring sperm DNA, 0.5% dry milk) and perfused into the reaction chamber. For multi-protein array fabrication, two solutions of polymer-target protein complexes were prepared separately, combined into one solution, and introduced to the reaction chamber. The slides were left at room temperature to incubate overnight. Afterwards, the Gene Frames were removed and the slides were washed three times with wash buffer, rinsed with milli-Q water, and dried under nitrogen stream. New Gene Frames were applied to each array and a solution containing Alexa647-anti-GFP antibodies or both Alexa647-anti-GFP and Cy3-anti-GST antibodies were added at concentrations of 1 μg/mL for two hours. The slides were then washed 3 times for 10 minutes in wash buffer, rinsed briefly with milli-Q water and dried. The slides were immediately imaged on a GenePix 4000B microarray scanner (Molecular Devices) using a gain of 450.

The possibility of zipper exchange was also examined through a time-series study. Here, we prepared two leucine-zipper complexes: (1) a GFP-Z_E protein captured using a DNA-polypeptide conjugate and (2) an unconjugated Z_R-$(ELS)_6$-C polypeptide complexed to a GST-Z_E protein. The two captured proteins were mixed and incubated for periods of 1, 2, 4, or 8 hours before being applied to a printed microarray. Arrays were washed probed and scanned as in the two-channel experiments described above.

Surface-Immobilized β-Lactamase Assays. All reagents used to prepare substrates for these assays were dissolved in TE buffer (pH 7.4). 96-well plates were incubated with biotin-BSA (1 mg/mL) for 10 minutes. The plates were then blocked with casein (1 mg/mL) followed by BSA (1 mg/mL) for 1 minute each, and washed twice with TE buffer. Next, streptavidin (0.5 mg/mL)

was added to each well and incubated for 10 minutes. Afterwards, the streptavidin solutions were aspirated off and the plates were rinsed briefly with the casein solution and then twice with TE buffer. Biotin functionalized DNA was added to the appropriate wells to a final volume of 50 μL and incubated for 30 minutes. Certain wells were designated for controls that either omitted the DNA deposition or contained a non-complementary DNA strand. Excess DNA that did not bind the surface was removed by washing with casein and BSA as before. To immobilize enzymes, PSE-4-Z_E was pre-incubated in 3 fold excess of the Z_R-$(ELS)_6$-C-ssDNA conjugate for 30 minutes. The PSE-4-Z_E and conjugate solution were applied to the appropriate wells. The plate was left to incubate for 2 hours and then protein solutions were aspirated from the wells. After washing 4 times with TE buffer, each well was brought to a final volume of 100 μL by adding TE buffer. Subsequently, nitrocefin was added to each well, bringing the wells to a final volume of 200 μL and a final substrate concentration to 150 μM. The plates were scanned using a Tecan Infinite 200 microplate reader. Absorbance readings at 485 nm were monitored every 50 seconds for 3 hours at 25 ˚C. The steady-state reaction rates of the immobilized β-Lactamase enzymes were determined via linear regression analysis.

Synthesis and Characterization of Multi-Protein Assemblies. The DNA scaffold consisted of a single duplex of DNA (34 bases; 10 nm in length) flanked by unique single-strand 'overhangs' comprised of 20 bases on either end. To create scaffolds, strands were mixed stoichiometrically as estimated by A_{260}, and annealed from 95 ˚C to RT, over 1.5 hrs in TAE/Mg^{2+} buffer (40 mM Tris, 40 mM Acetic Acid, 2 mM EDTA, 12.5 mM Mg^{2+} Acetate, pH 8) for a final concentration of 1.8 μM. Target protein (either CaM-Z_E or GFP-Z_E) and its respective Z_R-$(ELS)_6$-C-ssDNA conjugate were mixed in a 1:1 ratio and incubated at 4 ˚C for at least 1 hr. DNA scaffold was

added in a 1:2 or 1:1 ratio with the protein mixture (depending on whether it was the single or two protein system) and incubated at RT for 1.5 hrs, then 4 ˚C for 30 min. A 7% non-denaturing PAGE gel was run at 200 V at 4 ˚C (SE 600 Ruby, GE Healthcare) in TAE/Mg^{2+} buffer. The DNA/protein complexes were visualized by staining with Stains-All (Sigma).

(1) Hoover, D. M., and Lubkowski, J. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res. 30*, e43.

(2) Seeman, N. C. (1990) De novo design of sequences for nucleic acid structural engineering. *J. Biomol. Struct. Dyn. 8*, 573-581.

(3) Meyer, D. E., and Chilkoti, A. (2002) Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: examples from the elastin-like polypeptide system. *Biomacromolecules 3*, 357-367.

(4) Kukolka, F., and Niemeyer, C. M. (2004) Synthesis of fluorescent oligonucleotide--EYFP conjugate: towards supramolecular construction of semisynthetic biomolecular antennae. *Org. Biomol. Chem. 2*, 2203-2206.

Table S1. List of nucleotide sequences used for conjugation and self-assembly. All sequences are listed in the 5' – 3' direction. AmMC6 indicates a 5' amino modifier on carbon 6 and 5Biosg represents a 5' biotin.

Sequence Name	**Nucleic Acid Sequence**
Conjugate1	/5AmMC6/CCAATGCGGTCTATCCAGCC
Conjugate2	/5AmMC6/CTACGGCAACTGTGGTCATC
scConjugate1	/5AmMC6/TTTTTTTTTTTTTTTTTTTTGGCTGGATAGACCGCATTCG
scConjugate2	/5AmMC6/TTTTTTTTTTTTTTTTTTTTGATGACCACAGTTGCCGTAG
sControl	/5AmMC6/TTTTTTTTTTTTTTTTTTTTCCGAGCAGGCGTGAGTCGTC
kcConjugate1	/5Biosg/TTTTTTTTTTTTTTTTTTTTGGCTGGATAGACCGCATTCG
kControl	/5Biosg/TTTTTTTTTTTTTTTTTTTTGATGACCACAGTTGCCGTAG
Scaffold Overhang1	CGTAGCAGGCACATCGTTGGCTGGATAGACCGCATTCG
Scaffold Overhang2	CGTAGCAGGCACATCGTTGATGACCACAGTTGCCGTAG
Scaffold1	GTCACGGACTGAGCGT
Scaffold2	CGATGTGCCT
Scaffold3	GCTACGACGCTC
Scaffold4	AGTCCGTGACTTGGCTGGATAGACCGCATTCG

Table S2. Amino acid sequences of the artificial polypeptide components.

Sequence Name	**Amino Acid Sequence**
Z_E	LEIEAAALEQENTALETEVAELEQEVQRLENIVSQYRTRYGPL
Z_R	LEIRAAALRRRNTALRTRVAELRQRVQRLRNEVSQYETRYGPL
ELS	VPGVGVPGSGVPGVGVPGSGVPGVGV
Z_R-$(ELS)_6$-C	MVGS-Z_R-GSNHGVG-$(ELS)_6$-VPGWLRSC

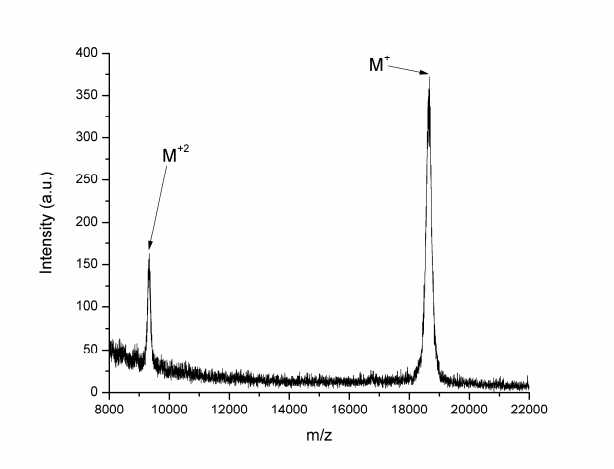

Figure S1. MALDI-TOF spectrum for the Z_R-$(ELS)_6$-C protein. The spectrum shows peaks at 9323 m/z and 18652 m/z, corresponding to the M^{+2} and M^{+} peaks, respectively. The calculated theoretical mass of the full-length polymers is 18793.5 Da, and 18662.3 Da if the N-terminal methionine is cleaved.

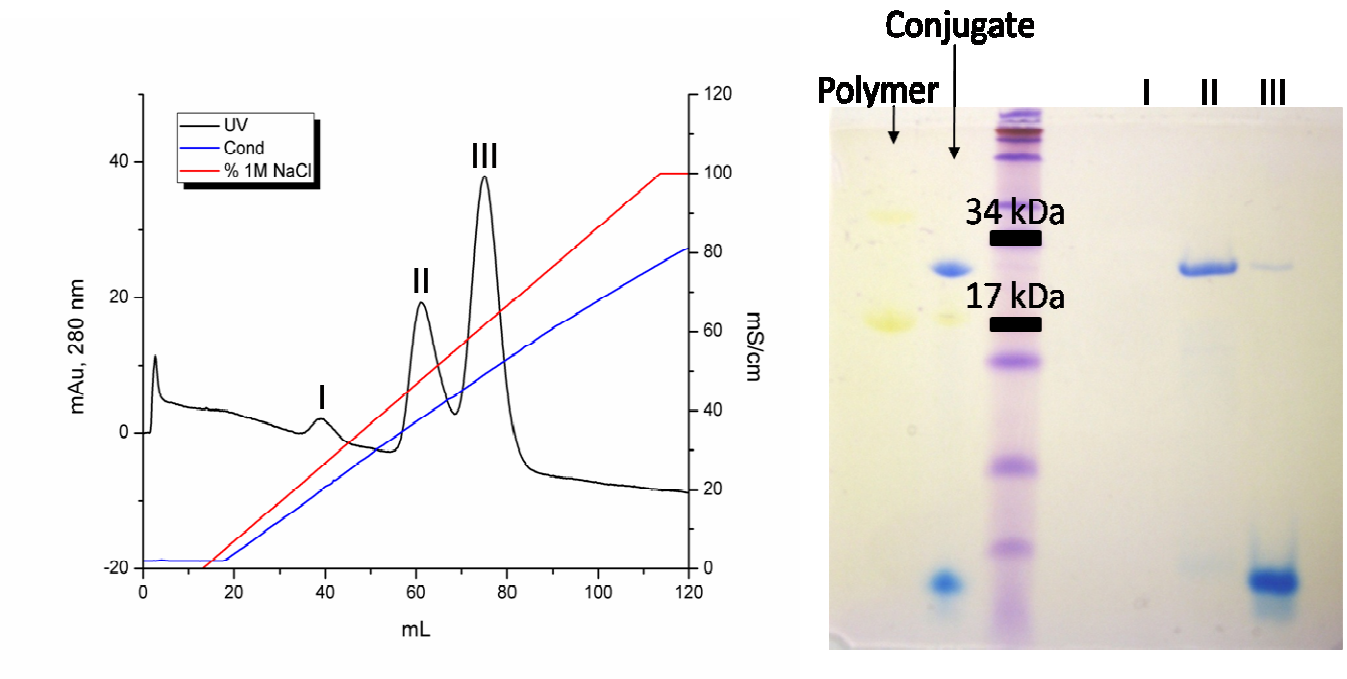

Figure S2. FPLC chromatogram and SDS-PAGE gel for conjugate synthesis and purification, indicating the absorption measured at 280 nm, gradient implemented, and measured conductivity. The chromatogram shows the characteristic 3 peaks from the crosslinking reaction: I. Unconjugated polymer; II. Conjugate; III. Unconjugated DNA. The corresponding gel shows purified polymer, the crosslinking reaction products (Conjugate), and a sample of each of the 3 peaks. The Stains-All stains DNA blue and protein yellow. The shift in the DNA band shows the presence of the conjugate in peak II.

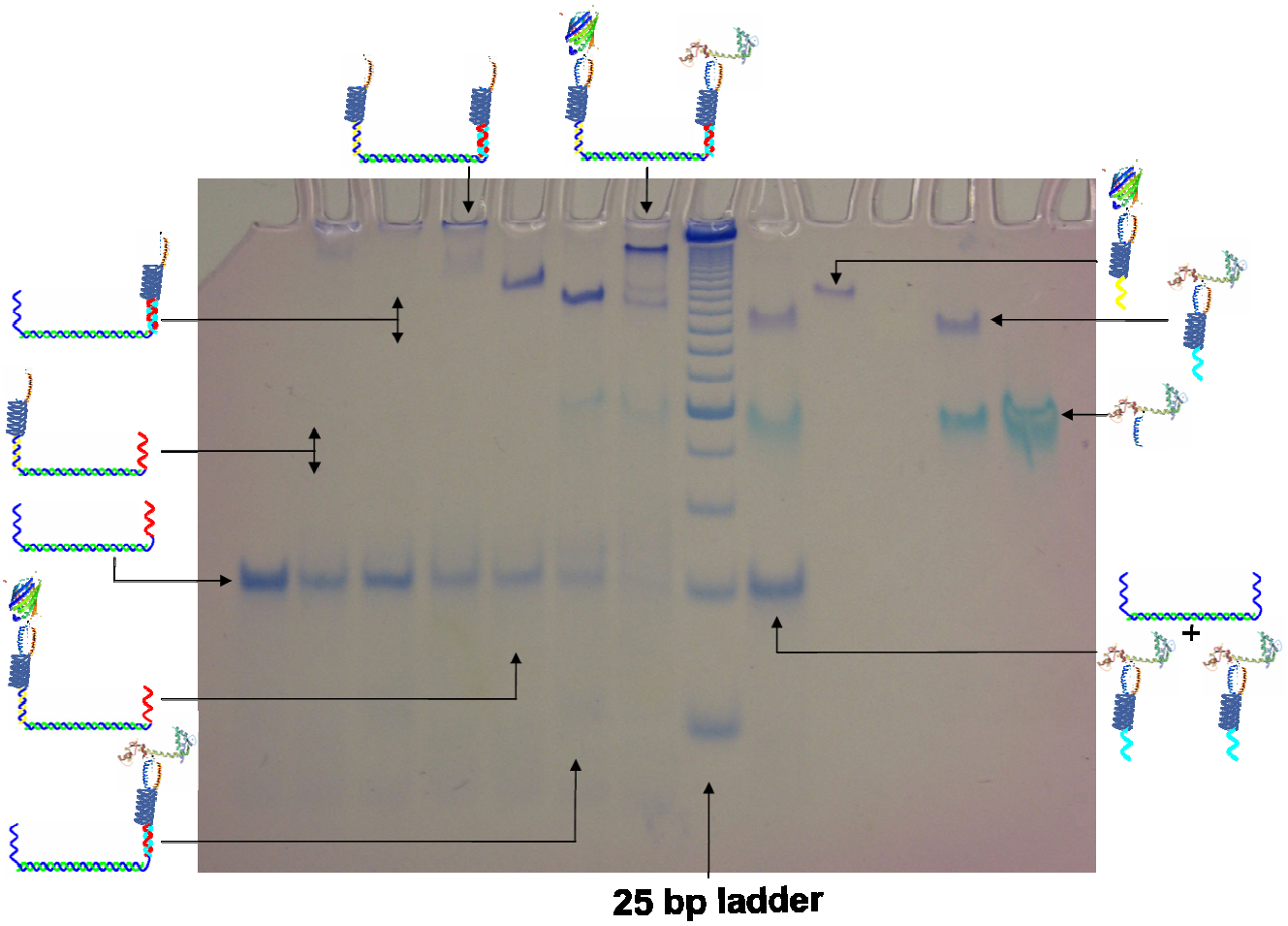

Figure S3. A non-denaturing PAGE gel displaying each of the components of the assembly individually as well as intermediate structures. The corresponding components of each lane are depicted at the periphery of the gel. From this analysis, we are able to adjust the volumes of individual protein stocks to compensate for errors in protein concentration measurements, in order to drive assembly formation more quantitatively.

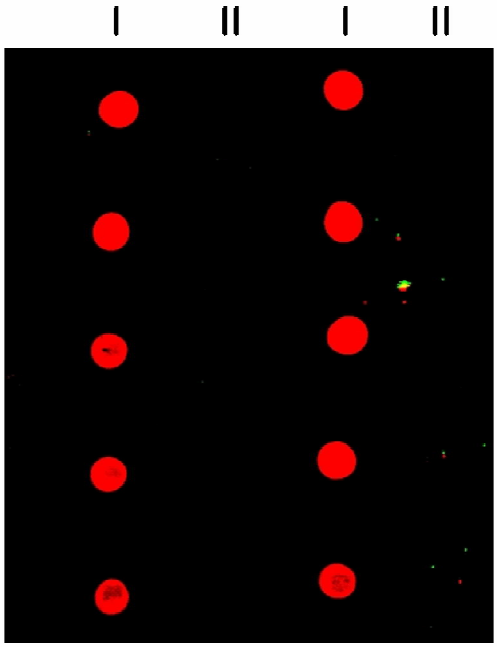

Figure S4. Microarray experiment showing the leucine zippers do not exchange between target proteins. Two protein complexes: (1) a GFP-Z_E protein captured using a DNA-polypeptide conjugate, and (2) a GST-Z_E complexed with an unconjugated Z_R-$(ELS)_6$-C polypeptide were mixed together for periods of 1-8 hours to test exchange. Afterwards, this solution was incubated for 16 hours over DNA-spotted array. Without exchange, the GST-Z_E protein should not immobilize on our arrays, and any exchange should result in the appearance of a GST-positive signal. All of the arrays show selective capture of the GFP-Z_E. We did not observe evidence of exchange in any experiments. The 8 hour time point is shown in the figure. The positive control strand was printed in lane I (the GFP positive lane), while a non-complementary strand is was printed in lane II : The GST signal (green) was lower than background in all spots in the array.