Supporting Information

The Protein-Templated Synthesis of Enzyme-Generated Aptamers
J. Vacacela, A.-L. Schaap-Johansen, P. Manikova, P. Marcatili, M. Prado, Y. Sun, J. Ashley*
Author Contributions

J.V. Data curation:Supporting; Investigation:Supporting; Writing – review & editing:Supporting
A.-L.S.-J. Data curation:Supporting; Formal analysis:Supporting
P.M. Formal analysis:Supporting; Methodology:Supporting
P.M. Supervision:Supporting; Writing – review & editing:Supporting
M.P. Supervision:Supporting; Writing – original draft:Supporting; Writing – review & editing:Supporting
Y.S. Methodology:Supporting; Supervision:Supporting; Writing – original draft:Equal; Writing – review & editing:Equal
J.A. Conceptualization:Lead; Data curation:Lead; Formal analysis:Lead; Funding acquisition:Lead; Investigation:Lead; Methodology:Lead; Supervision:Lead; Writing – original draft:Lead; Writing – review & editing:Lead
Experimental Section

MATERIAL AND METHODS

The DNA initiator sequences, 5’ ATC AGT TCG AGC AGA TGA GC’3, 5’ FAM - ATC AGT TCG AGC AGA TGA GC’3, 5’ Bio - ATC AGT TCG AGC AGA TGA GC’3 and the 10 - 100 nt oligonucleotide ladder, and the reverse primer CCA GAC TGC GAG CGT TTT TTT TTT B, were purchased from IDT. Individual polynucleotide sequences (Excluding the initiator sequence and poly A region with biotin tags and the control sequence (SC01) 5’ ATC AGT TCG ACT GGG GGA TAT GAG AAG TGA GAG ACG TCG GAC CGT GAG GAA TCG GGG GAA GGA GCG AGG ATG AGG GAG TGG CAG GGG TAC GGG AGT GGT GCC GGA GGG TTG GCG ATT GGG CGT GAG GGG TGT AAG ACA CCG TTG GAG GCG TAA CCT GAA GT’3 were synthesized using IDT’s Ultramer oligonucleotide synthesis service. Terminal deoxynucleotidyl transferase (TdT), dNTPs, and human lactoferrin were purchased from Sigma Aldrich. SYBR gold stain and neutravidin were purchased from Thermo Scientific. 5x FIREPol® Master Mix and sample loading buffer were purchased from Solis Biodyne and used for all PCR reactions. The Buffer SB1: 50 mM Tris-HCl, 250 mM KCl and 7.5 mM MgCl₂ pH 7.4 (5X) were used for all TdT imprinting and non imprinting reactions and library selection steps. The purification of oligonucleotides was performed using a PCR clean-up kit purchased from Macherey-Nagel GmbH. All electrophoretic mobility shift assays (EMSA); (5-6%). All SPR chips, were purchased from BioNavis (Finland).

Templated synthesis of protein binding enzyme generated aptamers

TdT catalysed pre-polymerisation mixtures were performed by preparing 400 µl solutions containing, 0.5 µM of the initiator sequence, 75 - 400 µM of dNTPs (dATP X µM, Y µM dCTP, Y µM dTTP and Z µM dGTP) in 1 x SB1 and template protein (0.5 µM). Corresponding non-templated EGA mixtures were formed in the
same manner as for templated mixtures but in the absence of the template. Reactions were initiated by the addition of 1-2 U/µl of TdT and the entire mixture was allowed to incubate at room temperature for 0.5 – 2 hours. Each reaction was terminated using 4 µl of 0.2 M EDTA or heat at 75 °C for 10 minutes to form each EGA mixture. Table S1 shows the reaction parameters and incubation times of each batch of pre-polymerisation mixture. Upon termination of the enzymatic reactions, resultant libraries were purified and concentrated using the oligonucleotide clean-up kit (Machary Nagel) and eluted into 20-60 µl of elution buffer (5 mM Tris-HCl). The size range of the EGAs were monitored using a 5 % denaturing acrylamide gel and visualised by staining with 1 x SYBR gold stain.

Rapid amplification of variable ends (RAVE) and next generation sequencing

EGA protein complexes were gel extracted, pooled together, purified and concentrated using the oligonucleotide clean-up kit (Machary Nagel) and eluted into 60 µl of elution buffer (5 mM Tris HCl). These sequences were converted to dsDNA using a RAVE assay. A poly (A) tail was introduced at the 3’end of the polynucleotides using TdT tailing reaction. TdT (1–2 U/µl) was incubated with solutions (20 µl) containing dATP, 10 mM Tris–HCl, 50 mM KCl and 1.5 mM MgCl₂ buffer, the extracted bound polynucleotide sequences (10 µl) for 0.5–2 h followed by termination of the reaction by heating the solution to 75°C for 10 min. The resultant extended sequences were used directly as the template for subsequent PCR reactions. 20 µl solutions containing 1× master mix (10 µl), 0.1–0.5 µM (1 µl) of the forward and reverse primers and 1–5 µl of the DNA-poly (A) templates were prepared. PCR reactions were performed over 30–40 cycles consisting of a denaturing step at 94°C for 30 s, an annealing step at 45–65°C for 30 s and an extension step of 72°C for 30 s. The resultant dsDNA libraries were purified again and were sequenced using a Illumina Hiseq2500 using a 250bp pair end read (Macrogen, South Korea) and is summarised in Table S5. All sample libraries were prepared using an Illumina TruSeq Nano DNA kit prior to sequencing. NGS data was analysed on a Galaxy suite based on previous workflows[1]. Sequences were filtered for quality and converted to FASTA file. The sequences were then trimmed to remove the initiator sequence and poly (A) region. The NGS data was then processed by copy number and ranked. Motifs were determined using MEME suite[2]. Regions demonstrating poly G repeats of >8 were filtered out. The top 200 sequences by copy number were analysed by secondary structure using mFold (7.5 mM Mg ion concentration and 150 mM Na ion concentration). Sequences were filtered out that showed a ΔG of > -10. Sequences were also analysed by QGRS Mapper[3]. Sequences showing no G score or a score below 10 were excluded. The corresponding non-templated aptamer sequences were also sequenced and a NT aptamer sequences were chosen based on the corresponding size to protein templated aptamers.

Gel analysis of human lactoferrin binding towards protein templated enzyme generated aptamer mixtures and individual sequences.

Resultant protein template EGA mixtures and the corresponding non-templated EGAs mixtures (7 µl) were placed in 1 x SB1 (4 µl) and refolded by heating the solution to 94 °C and cooling at a rate of 0.5 °C s⁻¹.
0.250 µM lactoferrin protein for 1 hour at room temperature. The resultant complex was then separated on a 5% native EMSA acrylamide gel and visualised by staining with 1 x SYBR gold stain. Selectivity studies were performed by incubating protein template enzyme generated aptamers and non-templated enzyme generated aptamers with human lactoferrin, trypsin and human serum albumin respectively and analysing on a 5% EMSA with SYBR gold stain. For individual protein-templated EGA sequences identified from next generation sequencing, each candidate sequence was resynthesized and incubated (250 nM) with each protein (1000 nM) respectively. Complexes were separated and analysed on a 7% EMSA acrylamide gel and visualised by staining with 1 x SYBR gold stain. A corresponding Scrambled sequence (SC01) was used as the control.

**SPR binding analysis of enzyme generated aptamer mixtures and individual sequences**

SPR binding analysis was performed on a MP-SPR Navi™ 200 OTSO SPR instrument. Firstly, a self-assembly monolayer was prepared by incubating bare gold chips into 5 mM of mercaptodecanoic acid 11-MUA in ethanol for 24 hours. The resultant chips were then washed using water and ethanol, dried using nitrogen and docked into the instrument. The instrument was primed with the run buffer DI water prior to immobilization of the aptamer and the flow rate was set at 20 µl / min. For the immobilization of the protein templated EGA mixtures, non-templated EGAs mixtures, each batch was immobilized onto each channel respectively by injecting an aqueous solution of NHS and EDC (80 µl). Neutravidin was then injected (80 µl, 50 µg / ml) in 50 mM sodium acetate buffer pH 5.0 onto the activated chip surface. The unreacted activated ester groups were then blocked by injecting 1M ethanolamine pH 8.3 (80 µl). Protein templated EGA mixtures were injected on to the analyte channel (channel 1) and non-templated EGA mixtures were injected onto the reference channel (channel 2). Kinetic analysis was performed by priming the SPR instrument with 1x SB1 run buffer and setting the flow rate to 20 µl / min. Injections of 0 – 350 nM of human lactoferrin (80 µl) or other proteins were performed on both channels and the relative response of the protein templated EGA mixtures after the removal of the response of the reference channel containing the immobilized NIA was recorded. For the analysis of individual protein-templated EGA sequences identified from next generation sequencing, 1µM of the biotin tagged individual sequence (80 µl) in SB1 were injected on the analyte channel (channel 1), while the scrambled sequence (SC01) was injected onto the reference channel (channel 2) respectively. Injections of 0 – 350 nM of human lactoferrin (80 µl) or other proteins were performed on both channels and the relative response of the individual sequences after the removal of the response of the reference channel containing the scrambled sequence was recorded. The kinetic binding parameters were fitted against the relative response using 1:1 model or bivalent model using the trace drawer data analysis software. The binding affinities (K_D) were determined from the association and dissociation rates where k_a is the association rate and k_d is the dissociation rate. All SPR experiments were performed in duplicate. The specificity of the protein templated EGAs, non-templated EGAs and individual protein templated EGA sequences identified from NGS were determined by measuring the absolute response of different concentrations of each protein and performing equilibrium analysis using trace viewer.
FIDA 1 binding analysis of enzyme generated aptamer mixtures and individual sequences

Binding affinity studies of enzyme-generated mixtures for both the FAM labelled protein templated EGA mixtures and non-templated EGA mixtures were performed on a FIDA 1 instrument (FidaBio, Denmark) adapted from previous studies[4–6]. 50 µl solutions containing different concentrations of human lactoferrin (0.6 - 500nM) were incubated with fixed concentrations of protein templated EGA mixtures (1000X dilution) and non-templated EGA mixtures in 1 x SB1 buffer respectively (indicator solution). Corresponding 50 µl samples containing the series dilution of human lactoferrin (0.6 – 500 nM) in 1X SB1 buffer only was also prepared (Analyte solution). 75 µm ID capillaries were conditioned by pressure injection with 1M NaOH for 45 seconds (3500 mbar) followed by the 1x SB1 assay buffer for 75 seconds (3500 mbar). The analysis was performed using the premixing mode by filling the capillary with the analyte solution for 20 seconds (3500 mbar), followed by injection of the indicator solution for 10 seconds (50 mbar). The fluorescence signal was measured by injecting the analyte solution by pressure driven flow for 180 seconds (400 mbar). The hydrodynamic radius was determined based on the peak shape of the eluted fluorescence peak. The change in hydrodynamic radii was plotted against the concentration of the analyte and a 1:1 model kinetic model was fitted against the data. The binding affinities of individual sequences were performed in the same manner but using 10nM of FAM labelled aptamer.

CD-spectroscopy

10 µM of EGA 4T was suspended in SB1 and placed in 1 mm path length quartz cells. A J-1500 Circular Dichroism Spectrophotometer was used for all CD measurements. Four scans were averaged from 310 to 210 nm at a scan rate of 100 nm min−1, a 1 s response time and 10 nm bandwidth. The baseline signal of SB1 was subtracted from the spectrum.
Table 1S Parameters for TdT catalysed molecular imprinting experiments.

|                      | 1a     |     | 1b     |     | 2a     |     | 2b     |     |
|----------------------|--------|-----|--------|-----|--------|-----|--------|-----|
|                      | T-EGA  | NT-EGA | T-EGA  | NT-EGA | T-EGA  | NT-EGA | T-EGA  | NT-EGA |
| Initiator Sequence (F Primer 10µM) | 20 µl  | (0.5 µM) | 20 µl  | (0.5 µM) | 20 µl  | (0.5 µM) | 20 µl  | (0.5 µM) |
| 5x SB1               | 80 µl  |     | 80 µl  |     | 80 µl  |     | 80 µl  |     |
| 1mM dTTP             | 20 µl  | (50 µM) | -     |     | -     |     | 20 µl  | (50 µM) |
| 1mM dCTP             | 20 µl  | (50 µM) | -     |     | -     |     | 20 µl  | (50 µM) |
| 1mM dATP             | -      |     | 20 µl  | (50 µM) | -     |     | 20 µl  | (50 µM) |
| 1mM dGTP             | -      |     | 20 µl  | (50 µM) | 20 µl  | (50 µM) | -     |     |
| Template protein Lactoferrin | 38.5 µl | (0.5 µM) | -     |     | -     |     | 38.5 µl | (0.5 µM) |
| TdT 400 U/µl         | 1 µl   |     | 1 µl   |     | 1 µl   |     | 1 µl   |     |
| Water                | 200.5 µl | 239 µl | 200.5 µl | 239 µl | 200.5 µl | 239 µl | 200.5 µl | 239 µl |
| Total volume         | 400 µl | 400 µl | 400 µl | 400 µl | 400 µl | 400 µl | 400 µl | 400 µl |
| Time of Reaction     | 1 hour | 1 hour | 1 hour | 1 hour | 1 hour | 1 hour | 1 hour | 1 hour |
Table 2S Parameters for TdT catalysed molecular imprinting experiments.

|                      | 3a          | 3b          | 4a          | 4b          |
|----------------------|-------------|-------------|-------------|-------------|
|                      | T-EGA       | NT-EGA      | T-EGA       | NT-EGA      |
| Initiator Sequence   |             |             |             |             |
| (F Primer 10µM)      | 20 µl (0.5 µM) | 20 µl (0.5 µM) | 20 µl (0.5 µM) | 20 µl (0.5 µM) |
| 5x SB1               | 80 µl       | 80 µl       | 80 µl       | 80 µl       |
| 1mM dTTP             | 10 µl (25 µM) | 10 µl (25 µM) | 20 µl (50 µM) | 20 µl (50 µM) |
| 1mM dCTP             | 20 µl (50 µM) | 20 µl (50 µM) | 10 µl (25 µM) | 10 µl (25 µM) |
| 1mM dATP             | 20 µl (50 µM) | 20 µl (50 µM) | 10 µl (25 µM) | 10 µl (25 µM) |
| 1mM dGTP             | 10 µl (25 µM) | 10 µl (25 µM) | 20 µl (50 µM) | 20 µl (50 µM) |
| Template protein     |             |             |             |             |
| Lactoferrin          | 38.5 µl (0.5 µM) | - | 38.5 µl (0.5 µM) | - |
| TdT 400 U/µl         | 1 µl        | 1 µl        | 1 µl        | 1 µl        |
| Water                | 200.5 µl    | 239 µl      | 200.5 µl    | 239 µl      |
| Total volume         | 400 µl      | 400 µl      | 400 µl      | 400 µl      |
| Time of Reaction     | 1 hour      | 1 hour      | 1 hour      | 1 hour      |
Table 3S Parameters for TdT catalysed molecular imprinting experiments.

|                      | 5a T-EGA | 5a NT-EGA | 5b T-EGA | 5b NT-EGA |
|----------------------|----------|-----------|----------|-----------|
| **Initiator Sequence (F Primer 10µM)** | 20µl (0.5 µM) | 20 µl (0.5 µM) | 20µl (0.5 µM) | 20 µl (0.5 µM) |
| **5x SB1**           | 80 µl    | 80 µl     | 80 µl    | 80 µl     |
| **1mM dTTP**         | 20 µl (50µM) | 20 µl (50µM) | 20 µl (50µM) | 20 µl (50µM) |
| **1mM dCTP**         | 20 µl (50 µM) | 20 µl (50 µM) | 20 µl (50 µM) | 20 µl (50 µM) |
| **1mM dATP**         | 5 µl (12.5 µM) | 5 µl (12.5 µM) | 5 µl (12.5 µM) | 5 µl (12.5 µM) |
| **1mM dGTP**         | 20 µl (50µM) | 20 µl (50µM) | 20 µl (50µM) | 20 µl (50µM) |
| **Template protein Lactoferrin** | 38.5 µl (0.5 µM) | - (1 µM) | 77 µl (1 µM) | - (1 µM) |
| **TdT 400 U/µl**     | 1 µl     | 1 µl      | 1 µl     | 1 µl      |
| **Water**            | 195.5 µl | 234 µl    | 157 µl   | 234 µl    |
| **Total volume**     | 400 µl   | 400 µl    | 400 µl   | 400 µl    |
| **Time of Reaction** | 1 hour   | 1 hour    | 1 hour   | 1 hour    |
Table S4: Summary of the lead protein T-EGA sequences found from next generation sequencing and corresponding NT-EGA controls.
### Table S5: Raw Data Stats from the Illumina 2500 NGS platform.

| Name | Sequence | Gibbs free energy (kcal mole\(^{-1}\)) using mFold | G-score using QGRS Mapper | SPR Binding Affinity \(K_D\) (nM) | \(k_{on}\) (1/M*s) | \(k_{off}\) (1/s) |
|------|----------|---------------------------------------------------|--------------------------|-------------------------------|----------------|----------------|
| EGA 1T | /5BiosG/GTCGGGCAGGGATCGCCTGCAGGGGATCGGCCGCGGCGGCGGGCGGCTGG | -24.3 | 40, 21 | 23 | 2.1e4 | 4.9e-4 |
| EGA 2T | /5BiosG/GTCCGGCGGTGCGCCTGCAGGGGATCGGCCGCGGCGGCTGG | -21.47 | 37, 42, 14, 20 | 34 | 1.6e4 | 3.6e-4 |
| EGA 3T | /5BiosG/GATTGGCGGCCGCGCCGCGGGGATCGGCCGCGGCGGCTGG | -20.3 | 58, 21, 39 | 14 | 4.0e4 | 5.6e-4 |
| EGA 4T | /5BiosG/GTTGCTCGGGGGTGGGCGGGGCAGGGGATCGGCCGCGGCGGCTGG | -13.0 | 62, 21, 17 | 5.4 | 3.6e4 | 1.9e-4 |
| EGA NT 4 | /5 BiosG/ATGACCGGGAGGGGGGGGCGGCGGTCCGGGGGATCGGCCGCGGCGGCTGG | -11.33 | 58, 21, 16 | 57 | 1.2e4 | 7.1e-4 |

| Name       | Total Reads | Total Read Bases (bp) | GC (%) | AT (%) |
|------------|-------------|-----------------------|--------|--------|
| T-EGA      | 63231288    | 15,871,053,288        | 56.46  | 43.54  |
| NT-EGA     | 85988146    | 21,583,024,646        | 49.17  | 50.83  |
Figure S1 (A) 5% denaturing gel of synthesized templated-EGAs and non-templated EGAs for 5a and 5b; (B) 5% EMSA of templated-EGAs and non-templated EGAs with composition 5a and (C) EMSA of templated-EGAs and non-templated EGAs of composition 5b.

Figure S2 SPR Sensorgrams showing the absolute response of HSA (A) and Trypsin (B) against templated-EGAs and non-templated EGAs and HSA (C) and Trypsin (D) against non-templated EGAs.
Figure S3 Affinity binding measurements of (A) protein templated EGA mixtures (T-EGAs) and (B) non-templated EGA mixtures (NT-EGAs) using the FIDA 1 instrument.
Figure S4  (A) Analysis of RAVE PCR products of protein templated EGA mixtures by Agilent Bioanalyzer, (B) size distribution of sequenced protein templated EGA mixtures sequences and Binding motifs identified from NGS data and (C) Motif logos for templated EGAs generated using MEME Suite.
Figure S5 (A) Analysis of RAVE PCR products of Non templated EGA mixtures by Agilent Bioanalyzer, (B) size distribution of sequenced non templated EGA mixtures sequences and Binding motifs identified from NGS data and (C) Motif logos for non templated EGAs generated using MEME Suite.
Figure S6 Binding affinity studies of EGA 4T (A) and EGA 4NT (B) using the FIDA 1 platform.
Figure S7 SPR Sensorgrams showing the relative response of HSA (A) and Trypsin (B) towards individual enzyme generated aptamer sequence EGA 4T; (C) Confirmative EMSA (7 %) of sequence EGA 4T and SC01 (250 nM) towards human lactoferrin, trypsin, bovine lactoferrin HSA, Lysozyme, and hemoglobin (1000 nM each).
Figure S8 CD Spectrum of EGA 4T.

References

[1] W. H. Thiel, *Mol. Ther. - Nucleic Acids* **2016**, 5, e345.
[2] T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, *Nucleic Acids Res.* **2015**, 43, W39–W49.
[3] O. Kikin, L. D’Antonio, P. S. Bagga, *Nucleic Acids Res.* **2006**, 34, 676–682.
[4] M. E. Pedersen, J. Østergaard, H. Jensen, *ACS Omega* **2020**, 5, 10519–10524.
[5] N. N. Poulsen, N. Z. Andersen, J. Østergaard, G. Zhuang, N. J. Petersen, H. Jensen, *Analyst* **2015**, 140, 4365–4369.
[6] M. E. Pedersen, R. M. S. Haegebaert, J. Østergaard, H. Jensen, *Sci. Rep.* **2021**, 11, 1–10.