

PURE mRNA display and cDNA display provide rapid detection of consensus binding motif via high-throughput sequencing

Sabrina Galinanes Reyes¹, Yutetsu Kuruma², Mai Fujimi³, Masako Yamazaki⁴, Sumie Eto⁴, Satoshi Tamaki⁴, Asaki Kobayashi⁵, Ryo Mizuuchi⁶, Lynn Rothschild⁷, Mark Ditzler⁷, and Kosuke Fujishima³

¹University of Glasgow

²Japan Agency for Marine-Earth Science and Technology

³Tokyo Institute of Technology

⁴MOLCURE, Inc

⁵INSERM

⁶JST

⁷NASA Ames Research Center

June 4, 2020

Abstract

The recombinant *in vitro* translation system known as the PURE system has been used in a variety of cell-free experiments such as expression of native and *de novo* proteins as well as various display methods to select for functional polypeptides. We developed a refined PURE-based display method for the preparation of stable mRNA and cDNA-peptide conjugates and validated its utility for *in vitro* selection. Our conjugate formation efficiency exceeded 40%, followed by gel purification to allow minimum carry-over of components from the translation system to the downstream assay enabling clean and efficient random peptide sequence screening. As a proof-of-concept, we chose the commercially available anti-FLAG M2 antibody as a target molecule for validation. Starting from approximately 1.7×10^{12} random sequences, a round-by-round high-throughput sequencing showed clear enrichment of the FLAG epitope DYKDDD as well as revealing consensus FLAG binding motif DYK(D/L/N)(L/Y/D/N/F)D. Enrichment of core FLAG motifs lacking one of the four key residues (DYKxxD) indicates that Tyr (Y) and Lys (K) appear as the two key residues essential for binding. Furthermore, comparison between mRNA display and cDNA display method resulted in overall similar performance with slightly higher enrichment for mRNA display. The consistency of two different display methods achieved by commercially available PURE system will be useful for future studies to explore sequence and functional space of diverse polypeptides.

Keywords

mRNA display, cDNA display, PURE system, peptide sequence screening, FLAG consensus motif

Introduction

Recombinant *in vitro* transcription-translation (TX-TL) procedures, also known as cell-free systems, have become increasingly common in recent years for a variety of biochemical and molecular procedures, mainly due to their versatility, and particularly in synthetic biology (Hammerling, Krüger, & Jewett, 2019; Keasling, 2012; Perez, Stark, & Jewett, 2016; Villarreal & Tan, 2017). These systems allow flexibility in the reaction condition of parameters for protein production, such as the transcription and translation machinery, as well as minimizing the possible limitations that carry working with living cells (Carlson, Gan, Hodgman,

& Jewett, 2012; Hodgman & Jewett, 2012; Silverman, Karim, & Jewett, 2019). Two main types of cell-free systems exist: those that derive from cell extracts and those that use purified recombinant proteins. The cell lysate-based systems were the first to be developed, and employ extracts from a variety of living organisms such as, *Escherichia coli*, yeast, fall armyworm, wheat germ, tobacco, rabbit reticulocytes and HeLa cell line (Anderson, Straus, & Dudock, 1983; Buntru, Vogel, Spiegel, & Schillberg, 2014; Ezure, Suzuki, & Ando, 2014; Jackson & Hunt, 1983; Sun et al., 2013; Wang, Zhao, & Zhao, 2014; Yadavalli & Sam-Yellowe, 2015). The application of using cell lysates quickly gained popularity as it tackled constraints encountered when performing synthetic biology research using living organisms, such as expression of protein that exhibits cellular toxicity or protein production under growth toxic compound (Bowie et al., 2020; Kay & Jewett, 2020; Tinafar, Jaenes, & Pardee, 2019). These systems often produce high yield of protein products; however, some issues still remain due to components that actively degrade the mRNA and proteins (nucleases, proteases), and the presence of additional unknown factors included in the cell lysate. While ongoing research continues to mitigate these issues (Didovyk, Tonooka, Tsimring, & Hasty, 2017; Fujiwara & Doi, 2016), cell-free systems using recombinant protein elements and purified ribosomes such as PURE system (Kuruma & Ueda, 2015; Y Shimizu et al., 2001; Yoshihiro Shimizu, Kuruma, Kanamori, & Ueda, 2014), offer a contaminant-free alternative with a final significant protein yield (Kazuta, Matsuura, Ichihashi, & Yomo, 2014). Lately, more research has focused on simplification, robustness and low-cost for reconstruction (Lavickova & Maerkl, 2019) as well as altering the energy source (Wang et al., 2019) in the PURE system.

These advantages of recombinant systems are particularly beneficial when one needs a control over conditions for high-throughput screening and directed evolution of peptide/proteins (Contreras-Llano & Tan, 2018; Dodevski, Markou, & Sarkar, 2015; Fujii et al., 2014). Different display methods (phage display (Ledsgaard, Kilstrup, Karatt-Vellatt, McCafferty, & Laustsen, 2018), yeast display (Boder & Wittrup, 1997; Cherf & Cochran, 2015), ribosome display (Zahnd, Amstutz, & Plückthun, 2007), liposome display (Fujii, Matsuura, Sunami, Kazuta, & Yomo, 2013), DNA display (Bertschinger & Neri, 2004; Doi & Yanagawa, 1999; Yonezawa, Doi, Kawahashi, Higashinakagawa, & Yanagawa, 2003), cDNA display (Naimuddin & Kubo, 2016; Yamaguchi et al., 2009), mRNA display (Nemoto, Miyamoto-Sato, Husimi, & Yanagawa, 1997; Roberts & Szostak, 1997; Seelig, 2011)) use various strategies to couple genotype to phenotype and as such have become indispensable tools for directed evolution. Among the display methods, *in vitro* approaches such as mRNA, cDNA, and ribosome display can screen the highest number of molecules (up to 10^{13}) to be tested because they are not limited by the efficiency of transformation or transfection. In the case of mRNA display, screening of large libraries is achieved by creating a covalent phenotype-genotype linkage between an mRNA and the polypeptide it encodes using puromycin (Takahashi, Austin, & Roberts, 2003). Moreover, *in vitro* reactions can be easily modified to suit a specific environment for functional screening (Josephson, Ricardo, & Szostak, 2014).

The utility of mRNA display is limited by the relative instability of mRNA-protein conjugates, especially in cell lysate-based translation systems, due to the presence of proteases and especially ribonucleases (Hino et al., 2008; Opyrchal, Anderson, Sokoloski, Wilusz, & Wilusz, 2005; Shin & Noireaux, 2010). The use of RNase inhibitors and nuclease-free chemicals can help minimize the degradation of RNA components (Newton, Cabezas-Perusse, Tong, & Seelig, 2020; Seelig, 2011). The recent advent of reconstituted contaminant-free PURE translation system has made *in vitro* display methods more popular for screening antibodies (Kanamori, Fujino, & Ueda, 2014; Nagumo, Fujiwara, Horisawa, Yanagawa, & Doi, 2016) and functional peptidomimetics (Bashiruddin & Suga, 2015). Since the PURE system operates primarily with reconstituted components, it offers increased stability of mRNA-protein conjugates. In addition, the cDNA display method, which converts translated mRNA-peptide conjugates into mRNA/cDNA-peptide conjugates, is advantageous under conditions where RNA instability is an issue during the selection step, such as targeting cell surface antigens under the presence of cellular ribonucleases (Ueno & Nemoto, 2012; Yamaguchi et al., 2009).

Here, we have established robust PURE system-based mRNA display and cDNA display methods, and compare their performance to screen for FLAG epitope sequences against anti-FLAG M2 antibody (Roosild, Castronovo, & Choe, 2006). Next-generation sequencing has been recently used along with display technology to provide an overview of sequence distribution (Fujimori et al., 2012). Hence, we performed round-by-round

deep sequencing to validate our method by monitoring stepwise enrichment patterns of the core FLAG epitope motifs and contribution of key residues that would otherwise be difficult to trace by traditional sequencing methods.

Materials and Methods

Linear DNA library construction

Linear DNA libraries were constructed by annealing two single stranded DNA (ssDNA) oligos and extension by Klenow polymerase. Oligos were ordered through FASMAC Inc., (Supplementary Table 1). A final concentration of 3 μM of Display-F and Display-FLAG-random-R oligos were used for the synthesis of FLAG-random library, as well as Display-F and Display-10aa-random-R oligos for the synthesis of 10aa-random library and Display-F and FLAG-control-R oligo for the synthesis of fixed FLAG (DYKDDDDK) sequence for western blot analysis. Oligo pairs were mixed with final 200 μM dNTPs in Klenow buffer (Takara Bio) and annealed by heating at 92 for 30 seconds followed by cooling to room temperature. Then 10U of Klenow Fragment (Large Fragment *E. coli* DNA Polymerase I; Takara Bio) was added to the reaction mix and an extension step was performed at 37 for 1 hour, followed by inactivation of the enzyme at 50 for 15 min. The Linear DNA libraries (FLAG-random and 10aa-random) were further column purified using the Wizard SV Gel and PCR Clean-Up System (Promega) and quantified by Nanodrop2000c (Thermo Fisher Scientific).

In vitro transcription and DNA-tag ligation

In vitro transcription was performed using HiScribe T7 Quick High Yield RNA Synthesis Kit (NEB) with approximately 500 ng (11 pmol, 6.6×10^{12} molecules) of linear DNA library at 37 for 6 h according to the manufacturer's protocol. The transcribed RNA was cleaned up using the NucleoSpin[®] RNA Clean-up kit (Macherey-Nagel). To anneal the mRNA library with a puromycin-FITC DNA tag, a 100 μL reaction was set up consisting of final concentrations of 4 μM transcribed mRNA, 6 μM puromycin-FITC DNA tag (Supplementary Table 1) and 1 mM ATP in T4 ligation buffer (NEB). The reaction was incubated at 90 for 30 sec, and then cooled to room temperature at 1 per second to correctly anneal the mRNA library with the puromycin-FITC DNA tag. Next, 3U of T4 PNK Kinase (NEB) and 20U of T4 RNA ligase (NEB) were added to the reaction mixture and further incubated at 25 for 30 min.

Gel electrophoresis, electroelution and ethanol precipitation

Gel electrophoresis of the linear DNA library samples were performed on a 2% agarose Tris-Acetate-EDTA (TAE) gel, run at constant 100V. Transcribed mRNA libraries and their FITC-puromycin-FITC DNA tag ligated products (mRNA-tag) were resolved using mini size (8 ml) 8 M urea 6% polyacrylamide Tris-borate-EDTA (TBE) gel at a constant 50 mA. SYBR Gold staining (Thermo Fisher Scientific) was used to detect DNA and mRNA products. For gel excision, ligation product was run on a large-scale urea-TBE gel (40 ml). A clean scalpel was used to cut off a gel strip containing mRNA-tag under a blue light LED transilluminator (FUJIFILM Wako Pure Chemical). Gel purification was performed using standard electroelution method using a Model 422 Electro-Eluter (Bio-Rad). First, the gels were crushed into small pieces and placed inside the Electro-Eluter tubes (Bio-Rad) with membrane cap (12 kDa cut-off) attached. Tubes were filled with 1x TBE buffer and run at 10 mA/tube for 30 min. Then, the polarity was reversed for 1 min to release the mRNA-tag from the membrane. Eluted sample solution was recovered and subjected to ethanol precipitation. For ethanol precipitation, 10% sample volume of 3 M sodium acetate was added to the eluted sample solution, then 3 volumes of 99.5% ethanol (FUJIFILM Wako Pure Chemical) was added. After mixing, the was incubated overnight at -20 to help increase the yield, and the mixture then centrifuged at 4, 20,000 x g for 60 min. The supernatant was carefully discarded, and 1 ml of 70% ethanol was added to the white pellet. A second centrifugation was performed at 4, 20,000 x g for 15 min. The supernatant was again carefully discarded, and the pellet was air-dried at room temperature for 30 min. Finally, the pellet was resuspended in RNase free water and quantified by a Nanodrop 2000c.

In vitro translation using cell-free system for mRNA display

Two μg of the mRNA-tag sample was used for *in vitro* translation using PURExpress 1.0 (Gene Frontier)

in the 100 μ l reaction mix. The optimized condition for efficient mRNA-peptide conjugate formation are as follows: initial translation at 37 for 30 min, followed by addition of a salt mix to a final concentration of 32.5 mM $MgCl_2$ and 375 mM KCl, and then a second incubation at 37 for 60 min. After this step, an equal volume of 2x Laemmli sample buffer (Bio-Rad) was added, mixed and centrifuged at 10,000 x g, 1 min at room temperature to remove the salt precipitant. The supernatant containing the translated product was recovered and further resolved on a 3.5% stacking - 10% resolving polyacrylamide SDS-PAGE gel at constant 50 mA. 8 M urea was added to the resolving gel only. Afterwards, the gel band corresponding to mRNA-peptide conjugate was excised, purified using the Electro-Eluter. and precipitated with ethanol, as described in the previous section.

Optimization of mRNA-peptide conjugate formation

The formation rate of mRNA-peptide conjugate using PUREflex[®]1.0 was optimized through iterations of three different reaction conditions: translation time, salt mix concentration, and incubation time after salt mix addition. All reactions were performed at 10 μ l scale reaction. For the translation time, we tested 0, 5, 15, 30 and 60 minutes at 37 , prior to the addition of salt mix (final conc, 32.5 mM $MgCl_2$ and 375 mM KCl) followed by 60 minutes of incubation. For the optimization of salt mix concentration, we first conducted 30 minutes translation at 37 , then either added no salt or 1x, 1/2x, 1/4x, 1/8x concentration of 65 mM $MgCl_2$ and 750 mM KCl concentration originally reported to promote covalent bond between puromycin and polypeptide chain (Naimuddin et al., 2011; Naimuddin & Kubo, 2016). For the incubation time after salt addition, we initially performed 30 minutes translation, added salt to final 32.5 mM $MgCl_2$, 375 mM KCl concentration, then tested 0, 10, 30, 60 and 90 minutes of incubation time.

Trypsin digestion

Trypsin digestion was carried out by using 75 ng mRNA-tag sample (with and without translation under optimal condition), 1 mM $CaCl_2$, 50 mM ammonium bicarbonate, and 0.5 μ g of Trypsin Gold (Promega). Each reaction was incubated at 37 for 30 min and resolved in a 3.5% stacking - 10% resolving polyacrylamide SDS-PAGE gel, with 8 M urea added to the resolving gel.

Reverse transcription for cDNA synthesis

Reverse transcription (RT) was performed against purified mRNA-peptide conjugate to obtain mRNA/cDNA duplexes for cDNA display using PrimeScript One Step RT-PCR Kit Ver.2 (Takara Bio). The initial peptide library (0th round) was prepared by mixing FLAG-random and 10aa-random mRNA-peptide conjugate samples to a 1:10,000 molar ratio. Approximately 200 ng ($\sim 2 \times 10^{12}$ molecules) of the mRNA-peptide initial library was reverse transcribed in 50 μ l reaction volume with final 0.4 μ M of RTPCR-R primer in 1 step buffer (Takara Bio) supplied by the kit. Tubes were incubated at 80 for 30 seconds, then cooled down to 4 for primer annealing. After, the PrimeScript 1 step Enzyme Mix (Takara Bio) was added and incubated at 48 for 20 min. We also prepared RT(-) tube as a negative control without adding the enzyme mix. After the incubation, both samples were purified using the NucleoSpin RNA Clean-up XS kit (Macherey-Nagel) and quantified by Nanodrop.

Western blot analysis

Western blot analysis was performed to confirm the physical linkage of peptides and their cognate mRNA sequences in the form of mRNA-peptide or mRNA/cDNA-peptide conjugates. A control mRNA sequence (Supplementary Table 1) encoding a FLAG epitope (DYKDDDDK) was ligated with puromycin-FITC DNA tag and translated under optimized condition (see previous section). Samples were run on a 10% polyacrylamide SDS-PAGE gel at constant 30 mA until the samples stack and reach the separating gel, and at constant 50 mA for additional 40 minutes. The samples in the gel were semi-dry transferred to a PDVF membrane (Bio-Rad) at constant 110 mA for 60 minutes. Blocking was carried out using the PVDF Blocking Reagent for Can Get Signal (Toyobo) and, the membrane was incubated using Anti DYKDDDDK tag monoclonal antibody peroxidase-conjugated (FUJIFILM Wako Pure Chemical) and Can Get Signal, solution 2 (Toyobo) with a 1: 12,000 ratio, respectively. The membrane was then revealed with chemiluminescence

using ECL Select Western Blotting Detection Reagent (GE Healthcare).

Selection using anti-FLAG M2 antibody

First, 40 μ l of Anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) were washed two times with TBS-T buffer (50 mM Tris HCl, pH 7.4, with 150 mM, NaCl, 1 mM EDTA, 0.2% Tween 20), using the DynaMag-2 magnetic stand (Thermo Fisher Scientific). A total of 100 ng of mRNA-peptide initial library or mRNA/cDNA-peptide conjugate initial library samples were diluted with 500 μ l of TBS-T buffer. Diluted samples were mixed with washed beads and incubated at room temperature with gentle mixing in a circular rotator for 1 hour to allow binding. After the incubation, the supernatant was removed, and the beads were further washed three times with TBS-T and additional three times with TBS buffer (without Tween 20). Through 1 to 3 round, beads were resuspended in 50 μ l TBS buffer and solutions were directly applied to the RT-PCR reaction (see next section for detail) to generate double strand DNA library for the next selection round and sequencing. For the 4th round samples, we performed elution in a stepwise manner using FLAG Peptide (Sigma-Aldrich) with increasing concentrations from 4 μ g/ml to 20 μ g/ml to 100 μ g/ml. Each of the eluted fractions, along with the remaining beads, were analyzed with RT-PCR.

RT-PCR for amplifying selected sequences

Total 10 μ l of resuspended magnetic beads with bound mRNA-, mRNA/cDNA-peptide conjugates were used as a substrate for 50 μ l RT-PCR reaction volume using PrimeScript One Step RT-PCR Kit Ver.2 with 0.4 μ M RTPCR-F and RTPCR-R primers (Supplementary Table 1). The reaction mix was incubated at 80 for 30 seconds, then cooled down to 4 for primer annealing, followed by the addition of PrimeScript enzyme and incubation at 48 for 20 min to complete the reverse transcription reaction. Reaction mix was then split into total 8 tubes to recover samples at different PCR cycles (0, 5, 10, 15, 20, 25, 30 and 35) to check the amplification efficiency. The PCR condition was set as follows: initial denaturation at 94 for 60 second, followed by 30 cycles of 94 for 15 seconds, 62 for 30 seconds and 72 for 25 seconds, then final extension at 68 for 30 seconds. In order to avoid over amplification of DNA, the reaction samples were run on a 2% TAE agarose gel stained by SYBR Gold and the optimal amplification cycle was determined based on the band intensity and the absence of non-specific amplified products. Finally, 50 μ l RT-PCR reaction was once again conducted with optimal PCR cycles (20 cycles for round 1 and 2, 15 cycles for round 3) and the product was purified using Wizard SV Gel and PCR Clean-Up System for the next round of selection and sequencing.

Next-generation sequencing

Next-generation sequencing (NGS) was conducted using Illumina MiSeq for every round of DNA library (0 to 4 rounds). Prior to sequence library preparation, quantification was carried out using Qubit 2.0 Fluorometer and Qubit 1X dsDNA HS Assay Kit (Thermo Fisher Scientific). A total 50 ng of the DNA library sample was used as an input for the NEBNext Ultra II DNA Library Prep with Sample Purification Beads (NEB). We also used NEBNext Multiplex Oligos for Illumina (Index Primers Set 1) (NEB) to multiplex our samples. The length of prepped DNA Library was determined by using Agilent 2100 Bioanalyzer (Agilent Technologies), quantified using Qubit 2.0 Fluorometer and run on MiSeq illumina platform using the Miseq Reagent Kit v3 600 cycles.

Sequence analysis

Sequence files were retrieved from the MiSeq illumina platform as a raw FASTQ file. Sequences corresponding to the coding regions were extracted using a pattern search program written by Perl script. Only the sequences with an 8 nucleotide perfect match upstream of the start codon and downstream of the random region are considered as reliable for downstream analysis. The reliable sequences (in FASTQ format) were further processed using FASTAptamer-count (Alam, Chang, & Burke, 2015) to rank sort sequences based on their read counts. Consensus sequence logos were created using WebLogo 3 (Crooks, Hon, Chandonia, & Brenner, 2004) based on the top 50 reads with specific conditions (fixed letter and positions).

Results

Overview of PURE system-based display method

We have refined and optimized the mRNA and cDNA display method utilizing the commercially available PUREfrex1.0 system (*in vitro* TX-TL system with recombinant elements (Y Shimizu et al., 2001; Yoshihiro Shimizu, Kanamori, & Ueda, 2005)) and screened approximately 10^{12} peptide sequences in a single experiment (Figure 1). One of the important features of our refined method is the stability of the mRNA-tag product during translation via RNase-free PUREfrex and two separate gel purification steps that assure the high purity of mRNA-tag for translation as well as peptide-conjugated products. Elimination of PURE components (RNA, enzyme, polyamine, cofactors, and ions) from the mRNA and cDNA-peptide conjugates avoids non-specific interaction among components, peptide conjugates and target during the downstream binding assay. Anti-FLAG M2 antibody was chosen as a suitable target to test the performance of our refined display method given the nature of its short octapeptide FLAG epitope (DYKDDDDK), known crystal structure (Roosild et al., 2006), and the evidence that previously three different display methods (phage display, DNA display and ribosome display) were able to enrich the core FLAG epitope motif (Miceli, DeGraaf, & Fischer, 1994; Osada, Shimizu, Akbar, Kanamori, & Ueda, 2009; Srila & Yamabhai, 2013; Yonezawa et al., 2003). The overall procedure takes total approximately 21 hours (mRNA display) or 22 hours (cDNA display), requiring standard four working days for one round of selection (Figure 1). High-throughput sequencing was performed after collecting multiple rounds of DNA library samples.

Random DNA library design

We designed two different random DNA libraries for the purpose of monitoring the trajectory of various sequences per round. The FLAG-random library contains fixed codons encoding first five letters of FLAG epitope (DYKDDDDK) followed by three consecutive degenerate “RRN” codons corresponding to seven amino acids (Asn, Lys, Asp, Glu, Ser, Arg and Gly). Whereas the 10aa-random library consists of ten degenerate codons harboring degenerate “VNN” and “NNY” codons corresponding to 16 and 17 different amino acids, respectively (Supplementary Figure 1). Expected sequence variation of the FLAG-random and 10aa-random library are 343 and 1.7×10^{12} , respectively. By using the FLAG-random library as a control, we expected to confirm the enrichment of the full FLAG epitope sequence in the early rounds. On the other hand, the 10aa-random library enables us to validate the performance of our display method by exploring large sequence space for finding a suite of sequences that are able to bind to anti-FLAG M2 antibody. Both DNA libraries contain an upstream T7 promoter for *in vitro* transcription, and a ribosome binding site (RBS) for *in vitro* translation. The leader sequence at the 3’ region is complementary to the puromycin-FITC DNA tag for efficient ligation. The only difference between the two libraries are the randomized sequence regions (Supplementary Figure 1).

In vitro transcription and ligation

After the DNA library construction, *in vitro* transcription was performed, followed by column purification and ligation to the puromycin-FITC DNA tag. Ligation was carried out via the efficient Y-ligation method (Nishigaki, Taguchi, Kinoshita, Aita, & Husimi, 1998) to connect the 3’ end of single strand mRNA, and the phosphorylated 5’ end of the DNA tag (Supplementary Figure 2A). Ligated products were run on 6 % polyacrylamide TBE gel with 8 M urea to confirm the band shift after the ligation. Incorporation of the DNA tag was confirmed by SYBR Gold staining of the mRNA molecule and the FITC fluorescence detection of the puromycin-FITC DNA tag (Supplementary Figure 2B). The ligation resulted in approximately 50% of the mRNA molecules ligated to the DNA tag. Ligated mRNA-tag product was further gel purified and electroeluted to eliminate unligated products. The elimination of free DNA tag is crucial for the downstream translation reaction since puromycin can interfere with ribosome. Further mRNA-tag was recovered by ethanol precipitation in a quantity of several μg .

Optimization of conjugate formation during *in vitro* translation

Recovered mRNA-tag was subjected to *in vitro* translation using a commercial cell-free PUREfrex 1.0 kit to form mRNA-peptide conjugates. PUREfrex is known for its high purity for eliminating nucleases and proteases and thus ideal for the application for mRNA display. Indeed, throughout our experiment we never

encountered mRNA-tag degradation during the translation reaction. The formation rate of the mRNA-peptide conjugate also depends on the incubation conditions during and after the translation to promote efficient incorporation of puromycin to the stalled ribosome at the mRNA-DNA boundary. Previous studies have applied rare codons at the 3' end of mRNA coding region (Nagumo et al., 2016) or used long linker in the DNA tag (Naimuddin & Kubo, 2016) to improve the conjugate formation. In this study, we have adopted both of these features by placing rare GGA codon at the 3' end and included polyA (18 nt) in the puromycin-FITC DNA tag. By using the gel purified 10aa-random mRNA-tag as a substrate, we carried out the optimization of conjugate formation via iteration of three different parameters: translation time, salt mix (KCl and MgCl₂) concentration, and incubation time after salt addition. Translation time reflects the ribosome loading, translation of the coding region which ends up stalling at the 3' end due to lack of stop codon. We tested translation at 37 from 0 to 60 minutes and found that 5 minutes of translation is sufficient to reach over 40% conjugate formation and the extent of conjugation does not significantly increase with longer incubation times (Figure 2A). Next, various salt mix concentrations were tested during post translation incubation. Previously, a high concentration of KCl has been shown to improve the accessibility of ribosome-bound peptidyl-tRNA to puromycin (Van Der Mast & Bloemers, 1973) and has been adopted in the recent mRNA and cDNA display methods (Naimuddin & Kubo, 2016; Seelig, 2011; Yamaguchi et al., 2009). We found that without salt addition, conjugate formation is less than 10%, however the addition of 32.5 mM MgCl₂ and 375 mM KCl greatly increased the formation rate to above 40% (Figure 2B). Furthermore, we found the incubation time after salt mix addition to be important for puromycin to react with peptidyl-tRNA to form covalent link between mRNA and peptide. Here, 60 minutes of incubation at 37 gave the highest yield with a formation rate of 48.5% (Figure 2C). With the above optimized condition, mRNA-peptide conjugates were synthesized, and gel purified from the SDS-urea polyacrylamide gel. Throughout all conditions we tested, we did not observe any mRNA degradation nor mRNA-peptide conjugate degradation during the translation and incubation steps (Figure 2A and 2C).

Confirmation of mRNA- and mRNA/cDNA-peptide conjugates

We confirmed the presence of synthesized peptide on the mRNA- and mRNA/cDNA-peptide conjugates with two different methods. First, trypsin digestion (targeting Lys and Arg) was applied to the 10aa-random mRNA-tag product with and without translation and revealed on a urea-SDS PAGE. All peptides contain multiple Lys residues at the fixed C-terminus region (Supplementary Figure 1) adjacent to the DNA tag, hence the digested product size should become almost equivalent to the mRNA-tag. As a result, the upper band in the post-translation product disappeared after 30 minutes of incubation, indicating that the band was indeed corresponding to the mRNA-peptide conjugate product (Figure 2D). Alternatively, by using the mRNA-tag product derived from fixed FLAG-control sequence (Supplementary Figure 1), the presence of conjugated peptides before and after the cDNA synthesis was confirmed by western blot analysis (Figure 2E). FITC fluorescence signal from the membrane transferred samples and chemiluminescence signal via anti-FLAG-HRP antibody clearly overlap at the same position, indicating the formation of both peptide conjugates. Note that under the SDS-PAGE condition, the double-stranded mRNA/cDNA-peptide conjugate migrated towards the cathode faster than the single-stranded mRNA-peptide conjugate due to the doubled charge from the phosphate backbone.

In vitro selection and RT-PCR

To evaluate the performance of our mRNA and cDNA display method, we performed a binding assay against anti-FLAG M2 antibody magnetic beads under neutral pH condition as described in the Materials and Methods section. For both methods, FLAG-random and 10aa-random conjugate samples were mixed to a 1:10,000 molar ratio. After the binding step and repetitive wash step, beads were recovered and directly subjected to RT-PCR. To avoid the PCR bias via over amplification, we checked the amplified product every 5 cycles (0 to 35 cycles) to determine the appropriate cycle for double-strand DNA library synthesis, (Supplementary Figure 3). Amplified DNA products from both display methods became visual at the 15th cycle, and were saturated after 20 cycles. As the selection round progressed, band intensity of the 15th cycle became stronger, implying the quantity of bead-bound conjugate products had increased over the course of

selection. We also observed a non-specific band appearing around 200 bp, however this band remained weak throughout multiple rounds, thus we proceeded with the RT-PCR reaction. PCR was carried out for 20 cycles for the 1st and 2nd round libraries and 16 cycles for the 3rd round for both display methods.

High-throughput sequencing analysis

High-throughput sequence data was collected from the Miseq system and parsed through a Perl script to extract the coding region. Extracted sequences were further analyzed by the FastAptamer software (Alam et al., 2015) to count and rank the unique sequences based on their number of reads. We first focused on the top 50 most abundant sequences from each round of mRNA display to confirm the enrichment of FLAG epitope during the selection. Sequences from the FLAG-random library (DYKDDxxx) resulted in a fixed Asp (D) residue at the 6th position immediately after the first round of selection, followed by positively charged 7th position with Arg (R) and Lys (K) and a variable 8th position (Figure 3A). After three rounds of selection, sequences derived from the 10aa-random library converge on a conserved core FLAG motif (DYKxxD) in two different positions, in one case starting at position 2 and in the other case starting at position 5 (Figure 3B). This binding motif has been reported in the previous *in vitro* selection studies to have strong affinity for anti-FLAG M2 antibody (Miceli et al., 1994; Osada et al., 2009; Srila & Yamabhai, 2013; Yonezawa et al., 2003). By focusing on only those sequences with the core FLAG motif, we further found features in other variable residues such as weakly conserved Glu (E) immediately before the FLAG motif, (D/L/N) and (D/F/L/N/Y) located between the conserved Lys (K) and Asp (D) with over 80% probability followed by downstream Pro (P) residue (Figure 3C and 3D).

We next compared the enrichment rate of various FLAG motifs detected through mRNA display with that of cDNA display. The expected frequency for the full FLAG epitope sequence (DYKDDDDK) in the initial library (0th round) is 0.29 reads per million, and indeed we were not able to detect the epitope sequence from the 0th round in our sample of 1,768,526 sequences. Although as the rounds progressed, we observed continuous enrichment up to 1479 and 198 reads per million for the 3rd round mRNA and cDNA display, respectively (Figure 4A). When we focused on the consensus FLAG motif (DYKDDDDxx) with fixed Asp (D) residue at the 6th position, we observed marked enrichment of 261,982 and 31,689 reads per million, suggesting that 6th position significantly enhances binding to the anti-FLAG M2 antibody (Figure 4B). For the final 4th round, we performed competitive elution with the FLAG epitope octapeptide (DYKDDDDK) by increasing the concentration in a stepwise manner. The majority of the selected sequences were replaced by sequences derived from 10aa-random library harboring consensus FLAG motif: DYK(D/L/N)(D/F/L/N/Y)D located at two different positions (Figure 4C and 4D).

The highest enrichment of consensus FLAG motifs was achieved from sequences recovered from the eluant with high concentrations of competitor FLAG peptide (20 or 100 $\mu\text{g/ml}$) reaching total 325,917 and 135,913 RPM for mRNA and cDNA display, respectively (Table 1). Under our selection condition, mRNA display shows higher enrichment rate of the consensus FLAG motif (average 28.7 fold per round) over cDNA display (average 19.1 fold per round). Additionally, with mRNA display, the abundance of the FLAG motif among the sequences that remain bound to the beads after all of the elution was consistently greater than with cDNA display (Figure 4 and Table 1, 100 $\mu\text{g/ml}$ vs beads). These results suggest that the mRNA/cDNA conjugate is more readily eluted from the beads.

Finally, the influence of the individual FLAG amino acid residues on antibody binding was evaluated through the comparison of sequences lacking one of the four residues of the core FLAG motif (DYKxxD). Enrichment of sequences lacking Tyr (Y) or Lys (K) residues showed negative enrichment in the first round indicating the importance of these residues for primary recognition of the anti-FLAG M2 antibody. The observed result is consistent with the previous selection (Miceli et al., 1994; Srila & Yamabhai, 2013; Yonezawa et al., 2003) and obtained M2 antibody structural data (Roosild et al., 2006).

Discussion

We have established and refined mRNA and cDNA display methods utilizing the cell-free translation system (PURE system) for efficient and stable mRNA and mRNA/cDNA-peptide conjugate synthesis and down-

stream selection (Figure 1). One of the highlights of our method is the capability of preventing carry over components through two separate electroelution steps to ensure the purity of mRNA and mRNA/cDNA-peptide conjugates. This could be useful if one intends to target the molecules present in the PURE system such as nucleotides and translation related RNA and proteins. Also in comparison to affinity beads purification, polypeptides do not require peptide affinity tags. Performance of the two display methods was validated by conducting an epitope search from approximately 10^{12} peptide sequences against the well-known anti-FLAG M2 antibody (Roosild et al., 2006). A rapid enrichment of the canonical FLAG epitope: DYKD-DDxx after three rounds of selection and further, was identified with high-throughput sequencing. This led to new insights into the variable residues, resulting in the discovery of the consensus FLAG motif sequences; DYK(DLN)(DFLNY)D. This motif appeared in both of the sites that were able to accommodate the motif within the restrictions imposed by the library design. Only three sites in the original library had degenerate codon sequences that can code for tyrosine, and of those three sites, one is too close to the C-terminal end of the random region to be part of the FLAG motif.

Two display methods resulted in overall similar performance, showing consistency of our method but with slightly higher enrichment for mRNA display over cDNA display. The mRNA display FLAG motif sequences were slower to elute from the beads than that of cDNA display, possibly because the RNA-conjugated peptides smaller size allows it access to more binding sites on the beads than the larger mRNA/cDNA-peptide conjugates. In contrast, the cDNA display library appears to have more non-specific binding on the remaining beads.

High-throughput sequencing has become a powerful approach for various *in vitro* selection and evolution experiments. We show that round-by-round sequencing and stepwise competitive elution with the FLAG peptide revealed consensus FLAG motif over other sequence variants by the 4th and final round. We also observed additional enrichment of several amino acids adjacent to the FLAG motif in a library dependent manner (FLAG-random vs 10aa-random) suggesting high-order epistasis between residues may exist to account for the optimal binding conformation. The resulting large sequence database also contains important information regarding the fitness of individual amino acid residues. By comparing the enrichment pattern of sequences with a single residue difference, we were able to highlight the most influential residues important for anti-FLAG M2 antibody recognition. Such patterns would be useful to further explore the high-affinity epitope candidates, along with the recent advancement of array based high-throughput and quantitative protein assay able to discover high-avidity epitope variants (Layton, McMahon, & Greenleaf, 2019). Hence, we expect this method can be applied to broad range of molecular targets to allow rapid detection of consensus binding motifs.

Acknowledgements

We thank Shota Nishikawa and Hidenori Watanabe of Tokyo Institute of Technology, and Stephen Fried and Bryan Lanning from Johns Hopkins University for their valuable comments on the manuscript.

Conflicts of Interests

The authors declare no conflict of interest.

References

- Alam, K. K., Chang, J. L., & Burke, D. H. (2015). FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Molecular Therapy - Nucleic Acids* . <https://doi.org/10.1038/mtna.2015.4>
- Anderson, C. W., Straus, J. W., & Dudock, B. S. (1983). Preparation of a cell-free protein-synthesizing system from wheat germ. *Methods in Enzymology* . [https://doi.org/10.1016/0076-6879\(83\)01044-7](https://doi.org/10.1016/0076-6879(83)01044-7)
- Bashiruddin, N. K., & Suga, H. (2015). Construction and screening of vast libraries of natural product-like macrocyclic peptides using in vitro display technologies. *Current Opinion in Chemical Biology* . <https://doi.org/10.1016/j.cbpa.2014.11.011>

- Bertschinger, J., & Neri, D. (2004). Covalent DNA display as a novel tool for directed evolution of proteins in vitro. *Protein Engineering, Design and Selection* . <https://doi.org/10.1093/protein/gzh082>
- Boder, E. T., & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology* . <https://doi.org/10.1038/nbt0697-553>
- Bowie, J. U., Sherkhanov, S., Korman, T. P., Valliere, M. A., Oppenorth, P. H., & Liu, H. (2020). Synthetic Biochemistry: The Bio-inspired Cell-Free Approach to Commodity Chemical Production. *Trends in Biotechnology* . <https://doi.org/10.1016/j.tibtech.2019.12.024>
- Buntru, M., Vogel, S., Spiegel, H., & Schillberg, S. (2014). Tobacco BY-2 cell-free lysate: An alternative and highly-productive plant-based in vitro translation system. *BMC Biotechnology* . <https://doi.org/10.1186/1472-6750-14-37>
- Carlson, E. D., Gan, R., Hodgman, C. E., & Jewett, M. C. (2012). Cell-free protein synthesis: Applications come of age. *Biotechnology Advances* , 30 (5), 1185–1194. <https://doi.org/10.1016/j.biotechadv.2011.09.016>
- Cherf, G. M., & Cochran, J. R. (2015). Applications of yeast surface display for protein engineering. *Methods in Molecular Biology* . https://doi.org/10.1007/978-1-4939-2748-7_8
- Contreras-Llano, L. E., & Tan, C. (2018). High-throughput screening of biomolecules using cell-free gene expression systems. *Synthetic Biology* . <https://doi.org/10.1093/synbio/ysy012>
- Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research* . <https://doi.org/10.1101/gr.849004>
- Didovyk, A., Tonooka, T., Tsimring, L., & Hasty, J. (2017). Rapid and Scalable Preparation of Bacterial Lysates for Cell-Free Gene Expression. *ACS Synthetic Biology* . <https://doi.org/10.1021/acssynbio.7b00253>
- Dodevski, I., Markou, G. C., & Sarkar, C. A. (2015). Conceptual and methodological advances in cell-free directed evolution. *Current Opinion in Structural Biology* . <https://doi.org/10.1016/j.sbi.2015.04.008>
- Doi, N., & Yanagawa, H. (1999). STABLE: Protein-DNA fusion system for screening of combinatorial protein libraries in vitro. *FEBS Letters* . [https://doi.org/10.1016/S0014-5793\(99\)01041-8](https://doi.org/10.1016/S0014-5793(99)01041-8)
- Ezure, T., Suzuki, T., & Ando, E. (2014). A cell-free protein synthesis system from insect cells. *Methods in Molecular Biology* . https://doi.org/10.1007/978-1-62703-782-2_20
- Fujii, S., Matsuura, T., Sunami, T., Kazuta, Y., & Yomo, T. (2013). In vitro evolution of α -hemolysin using a liposome display. *Proceedings of the National Academy of Sciences of the United States of America* . <https://doi.org/10.1073/pnas.1314585110>
- Fujii, S., Matsuura, T., Sunami, T., Nishikawa, T., Kazuta, Y., & Yomo, T. (2014). Liposome display for in vitro selection and evolution of membrane proteins. *Nature Protocols* , 9 (7), 1578–1591. <https://doi.org/10.1038/nprot.2014.107>
- Fujimori, S., Hirai, N., Ohashi, H., Masuoka, K., Nishikimi, A., Fukui, Y., ... Miyamoto-Sato, E. (2012). Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data. *Scientific Reports* . <https://doi.org/10.1038/srep00691>
- Fujiwara, K., & Doi, N. (2016). Biochemical preparation of cell extract for cell-free protein synthesis without physical disruption. *PLoS ONE* . <https://doi.org/10.1371/journal.pone.0154614>
- Hammerling, M. J., Krüger, A., & Jewett, M. C. (2019). Strategies for in vitro engineering of the translation machinery. *Nucleic Acids Research* . <https://doi.org/10.1093/nar/gkz1011>
- Hino, M., Kataoka, M., Kajimoto, K., Yamamoto, T., Kido, J. I., Shinohara, Y., & Baba, Y. (2008). Efficiency of cell-free protein synthesis based on a crude cell extract from Escherichia coli, wheat germ, and rabbit reticulocytes. *Journal of Biotechnology* . <https://doi.org/10.1016/j.jbiotec.2007.08.008>

- Hodgman, C. E., & Jewett, M. C. (2012). Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering* . <https://doi.org/10.1016/j.ymben.2011.09.002>
- Jackson, R. J., & Hunt, T. (1983). Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. *Methods in Enzymology* . [https://doi.org/10.1016/S0076-6879\(83\)96008-1](https://doi.org/10.1016/S0076-6879(83)96008-1)
- Josephson, K., Ricardo, A., & Szostak, J. W. (2014). mRNA display: From basic principles to macrocycle drug discovery. *Drug Discovery Today* . <https://doi.org/10.1016/j.drudis.2013.10.011>
- Kanamori, T., Fujino, Y., & Ueda, T. (2014). PURE ribosome display and its application in antibody technology. *Biochimica et Biophysica Acta - Proteins and Proteomics* . <https://doi.org/10.1016/j.bbapap.2014.04.007>
- Kay, J. E., & Jewett, M. C. (2020). A cell-free system for production of 2,3-butanediol is robust to growth-toxic compounds. *Metabolic Engineering Communications* . <https://doi.org/10.1016/j.mec.2019.e00114>
- Kazuta, Y., Matsuura, T., Ichihashi, N., & Yomo, T. (2014). Synthesis of milligram quantities of proteins using a reconstituted in vitro protein synthesis system. *Journal of Bioscience and Bioengineering* . <https://doi.org/10.1016/j.jbiosc.2014.04.019>
- Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metab Eng* , 14 (1096-7184 (Electronic)), 189–195. <https://doi.org/10.1016/j.ymben.2012.01.004>
- Kuruma, Y., & Ueda, T. (2015). The PURE system for the cell-free synthesis of membrane proteins. *Nature Protocols* . <https://doi.org/10.1038/nprot.2015.082>
- Lavickova, B., & Maerkl, S. J. (2019). A Simple, Robust, and Low-Cost Method to Produce the PURE Cell-Free System. *ACS Synthetic Biology* . <https://doi.org/10.1021/acssynbio.8b00427>
- Layton, C. J., McMahon, P. L., & Greenleaf, W. J. (2019). Large-Scale, Quantitative Protein Assays on a High-Throughput DNA Sequencing Chip. *Molecular Cell* . <https://doi.org/10.1016/j.molcel.2019.02.019>
- Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., McCafferty, J., & Laustsen, A. H. (2018). Basics of antibody phage display technology. *Toxins* . <https://doi.org/10.3390/toxins10060236>
- Miceli, R. M., DeGraaf, M. E., & Fischer, H. D. (1994). Two-stage selection of sequences from a random phage display library delineates both core residues and permitted structural range within an epitope. *Journal of Immunological Methods* . [https://doi.org/10.1016/0022-1759\(94\)90097-3](https://doi.org/10.1016/0022-1759(94)90097-3)
- Nagumo, Y., Fujiwara, K., Horisawa, K., Yanagawa, H., & Doi, N. (2016). PURE mRNA display for in vitro selection of single-chain antibodies. *Journal of Biochemistry* . <https://doi.org/10.1093/jb/mvv131>
- Naimuddin, M., Kobayashi, S., Tsutsui, C., MacHida, M., Nemoto, N., Sakai, T., & Kubo, T. (2011). Directed evolution of a three-finger neurotoxin by using cDNA display yields antagonists as well as agonists of interleukin-6 receptor signaling. *Molecular Brain* . <https://doi.org/10.1186/1756-6606-4-2>
- Naimuddin, M., & Kubo, T. (2016). A High Performance Platform Based on cDNA Display for Efficient Synthesis of Protein Fusions and Accelerated Directed Evolution. *ACS Combinatorial Science* . <https://doi.org/10.1021/acscmbosci.5b00139>
- Nemoto, N., Miyamoto-Sato, E., Husimi, Y., & Yanagawa, H. (1997). In vitro virus: Bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Letters* . [https://doi.org/10.1016/S0014-5793\(97\)01026-0](https://doi.org/10.1016/S0014-5793(97)01026-0)
- Newton, M. S., Cabezas-Perusse, Y., Tong, C. L., & Seelig, B. (2020). In Vitro Selection of Peptides and Proteins - Advantages of mRNA Display. *ACS Synthetic Biology* . <https://doi.org/10.1021/acssynbio.9b00419>

- Nishigaki, K., Taguchi, K., Kinoshita, Y., Aita, T., & Husimi, Y. (1998). Y-ligation: An efficient method for ligating single-stranded DNAs and RNAs with T4 RNA ligase. *Molecular Diversity* . <https://doi.org/10.1023/A:1009644028931>
- Opyrchal, M., Anderson, J. R., Sokoloski, K. J., Wilusz, C. J., & Wilusz, J. (2005). A cell-free mRNA stability assay reveals conservation of the enzymes and mechanisms of mRNA decay between mosquito and mammalian cell lines. *Insect Biochemistry and Molecular Biology* . <https://doi.org/10.1016/j.ibmb.2005.08.004>
- Osada, E., Shimizu, Y., Akbar, B. K., Kanamori, T., & Ueda, T. (2009). Epitope mapping using ribosome display in a reconstituted cell-free protein synthesis system. *Journal of Biochemistry* . <https://doi.org/10.1093/jb/mvp027>
- Perez, J. G., Stark, J. C., & Jewett, M. C. (2016). Cell-free synthetic biology: Engineering beyond the cell. *Cold Spring Harbor Perspectives in Biology* . <https://doi.org/10.1101/cshperspect.a023853>
- Roberts, R. W., & Szostak, J. W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proceedings of the National Academy of Sciences of the United States of America* . <https://doi.org/10.1073/pnas.94.23.12297>
- Roosild, T. P., Castronovo, S., & Choe, S. (2006). Structure of anti-FLAG M2 Fab domain and its use in the stabilization of engineered membrane proteins. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* . <https://doi.org/10.1107/S1744309106029125>
- Seelig, B. (2011). mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. *Nature Protocols* . <https://doi.org/10.1038/nprot.2011.312>
- Shimizu, Y, Inoue, a, Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., & Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nature Biotechnology* , 19 (8), 751–755. <https://doi.org/10.1038/90802>
- Shimizu, Yoshihiro, Kanamori, T., & Ueda, T. (2005). Protein synthesis by pure translation systems. *Methods* . <https://doi.org/10.1016/j.ymeth.2005.04.006>
- Shimizu, Yoshihiro, Kuruma, Y., Kanamori, T., & Ueda, T. (2014). The PURE system for protein production. *Methods in Molecular Biology* , 1118 , 275–284. <https://doi.org/10.1007/978-1-62703-782-2-19>
- Shin, J., & Noireaux, V. (2010). Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system. *Journal of Biological Engineering* . <https://doi.org/10.1186/1754-1611-4-9>
- Silverman, A. D., Karim, A. S., & Jewett, M. C. (2019). Cell-free gene expression: an expanded repertoire of applications. *Nature Reviews Genetics* . <https://doi.org/10.1038/s41576-019-0186-3>
- Srila, W., & Yamabhai, M. (2013). Identification of amino acid residues responsible for the binding to anti-FLAGTM M2 antibody using a phage display combinatorial peptide library. *Applied Biochemistry and Biotechnology* . <https://doi.org/10.1007/s12010-013-0326-8>
- Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., & Noireaux, V. (2013). Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *Journal of Visualized Experiments : JoVE* , (79), e50762. <https://doi.org/10.3791/50762>
- Takahashi, T. T., Austin, R. J., & Roberts, R. W. (2003). mRNA display: Ligand discovery, interaction analysis and beyond. *Trends in Biochemical Sciences* . [https://doi.org/10.1016/S0968-0004\(03\)00036-7](https://doi.org/10.1016/S0968-0004(03)00036-7)
- Tinafar, A., Jaenes, K., & Pardee, K. (2019). Synthetic Biology Goes Cell-Free. *BMC Biology* . <https://doi.org/10.1186/s12915-019-0685-x>
- Ueno, S., & Nemoto, N. (2012). CDNA display: Rapid stabilization of mrna display. *Methods in Molecular Biology* . https://doi.org/10.1007/978-1-61779-379-0_8

Van Der Mast, C. A., & Bloemers, H. P. J. (1973). The puromycin reaction mediated by yeast ribosomes in high salt. *Molecular Biology Reports* . <https://doi.org/10.1007/BF00357646>

Villarreal, F., & Tan, C. (2017). Cell-free systems in the new age of synthetic biology. *Frontiers of Chemical Science and Engineering* . <https://doi.org/10.1007/s11705-017-1610-x>

Wang, P. H., Fujishima, K., Berhanu, S., Kuruma, Y., Jia, T. Z., Khusnutdinova, A. N., ... McGlynn, S. E. (2019). A Bifunctional Polyphosphate Kinase Driving the Regeneration of Nucleoside Triphosphate and Reconstituted Cell-Free Protein Synthesis. *ACS Synthetic Biology* . <https://doi.org/10.1021/acssynbio.9b00456>

Wang, X., Zhao, L., & Zhao, K. N. (2014). An optimized yeast cell-free lysate system for in vitro translation of human virus mRNA. *Methods in Molecular Biology* . https://doi.org/10.1007/978-1-62703-782-2_14

Yadavalli, R., & Sam-Yellowe, T. (2015). HeLa based cell free expression systems for expression of Plasmodium rhoptry proteins. *Journal of Visualized Experiments* . <https://doi.org/10.3791/52772>

Yamaguchi, J., Naimuddin, M., Biyani, M., Sasaki, T., Machida, M., Kubo, T., ... Nemoto, N. (2009). cDNA display: A novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions. *Nucleic Acids Research* . <https://doi.org/10.1093/nar/gkp514>

Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T., & Yanagawa, H. (2003). DNA display for in vitro selection of diverse peptide libraries. *Nucleic Acids Research* . <https://doi.org/10.1093/nar/gng119>

Zahnd, C., Amstutz, P., & Pluckthun, A. (2007). Ribosome display: Selecting and evolving proteins in vitro that specifically bind to a target. *Nature Methods* . <https://doi.org/10.1038/nmeth1003>

Figures

Hosted file

image1.emf available at <https://authorea.com/users/329657/articles/456613-pure-mrna-display-and-cdna-display-provide-rapid-detection-of-consensus-binding-motif-via-high-throughput-sequencing>

Figure 1. Schematic overview of mRNA and cDNA display using PUREfrex system. The system starts with a designed DNA library (randomized sequences) harboring T7 promoter and a leader sequence that is complementary to the DNA tag. DNA library is *in vitro* transcribed to mRNA library and undergo ligation with the puromycin-FITC DNA tag. Tagged product is then gel purified and translated using the PUREfrex 1.0 kit. The resulting mRNA-peptide conjugate is then gel purified. The purified product is then either used directly in the binding step (mRNA display) or first reverse transcribed to become an mRNA/cDNA-peptide conjugate (cDNA display). Both products are used for binding against anti-FLAG M2 magnetic beads and the selected conjugates are further reverse transcribed and sequenced using the Illumina MiSeq system. Preparation time (hours) for each step is shown in red with exception of Miseq sequencing (5 days) carried out only after all samples were obtained.

Hosted file

image2.emf available at <https://authorea.com/users/329657/articles/456613-pure-mrna-display-and-cdna-display-provide-rapid-detection-of-consensus-binding-motif-via-high-throughput-sequencing>

Figure 2. Optimization of translation efficiency and confirmation of mRNA/cDNA-peptide conjugates. Gel images of the mRNA-peptide conjugate formation under various translation conditions. (A) Difference in length of translation before salt addition. Number indicates different incubation time (in minutes) carried out at 37 degrees C. (B) Difference in salt concentrations with original concentration (1x) as 750 mM KCl, 65 mM MgCl₂ used in the previous studies (Naimuddin et al., 2011; Naimuddin & Kubo, 2016) (C) Difference in lengths of incubation time (in minutes) after salt addition. (D) Trypsin digestion of mRNA-tag and mRNA-peptide conjugate products. (E) Detection of FITC fluorescence signal from the

mRNA-tag, mRNA-peptide conjugate, and mRNA/cDNA-peptide conjugate products (left) and detection of FLAG epitope sequence using chemiluminescent Western blotting (right) from the same membrane.

Hosted file

image3.emf available at <https://authorea.com/users/329657/articles/456613-pure-mrna-display-and-cdna-display-provide-rapid-detection-of-consensus-binding-motif-via-high-throughput-sequencing>

Figure 3. Consensus logo of top 50 abundant sequences from each selection round of mRNA display. A round-by-round consensus logo was created using Weblogo3 (Crooks et al., 2004) using the top 50 most abundant sequences based on their RPM. Sequence logos derived from FLAG-random library (A), 10aa-random library (B), core FLAG motif bearing sequences within the 10aa-random library (C and D) are color coded based on their charge (negatively charged: red, positively charged: blue and others: black). For the 4th round, consensus logo was only created for the sequences eluted with highest FLAG epitope tag concentration (100 $\mu\text{g}/\text{ml}$).

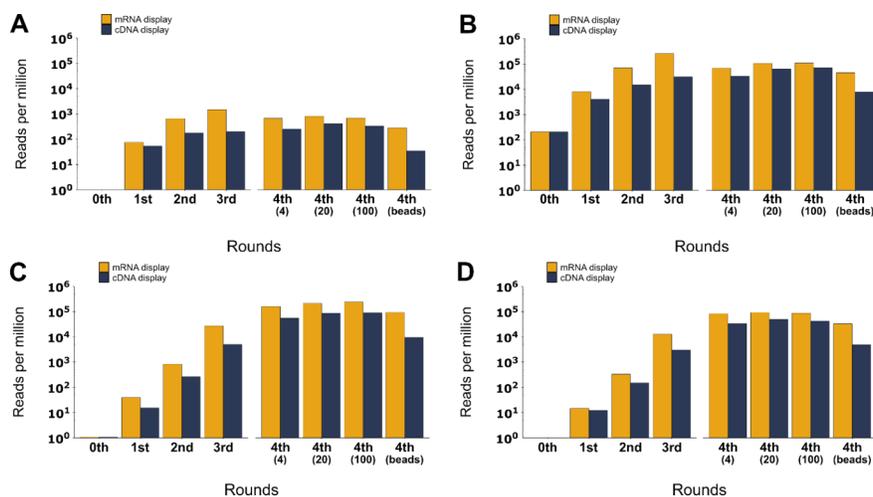


Figure 4. Round-by-round read counts for the enriched FLAG motifs. Sequence reads from initial library (0th) to 3rd round selection, 4th round with competitive elution with FLAG epitope tag at 4, 20, 100 $\mu\text{g}/\text{ml}$ concentration and remaining beads are counted for (A) DYKDDDDK, (B) DYKDDDDxx, (C) xDYK(D/L/N)(D/F/L/N/Y)Dxxx and (D) xxxxDYK(D/L/N)(D/F/L/N/Y)D sequences. For each round, read counts for mRNA display method (orange) and cDNA display (dark blue) are presented side-by-side.

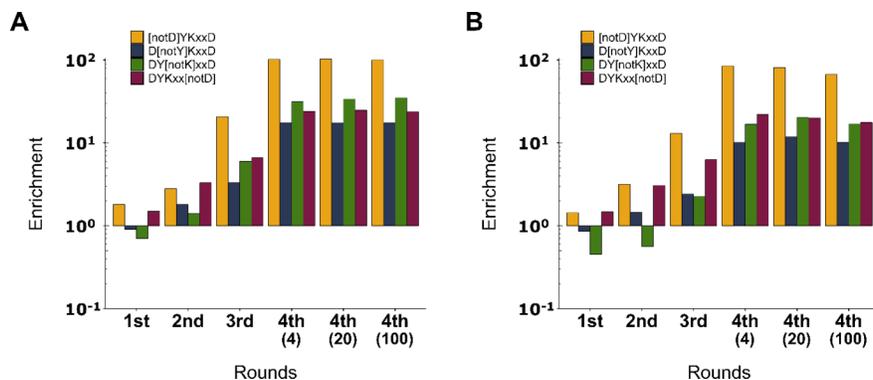
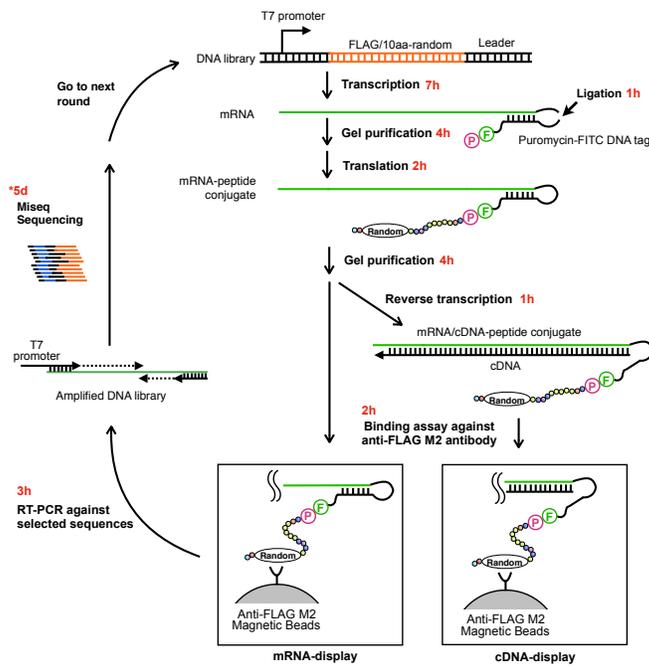
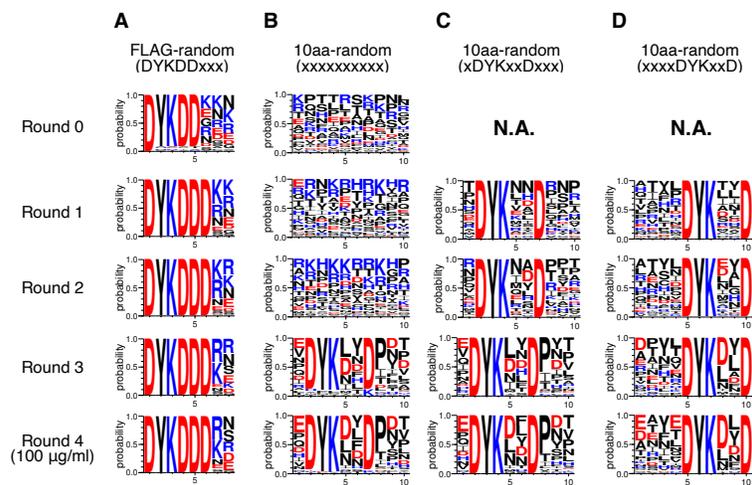
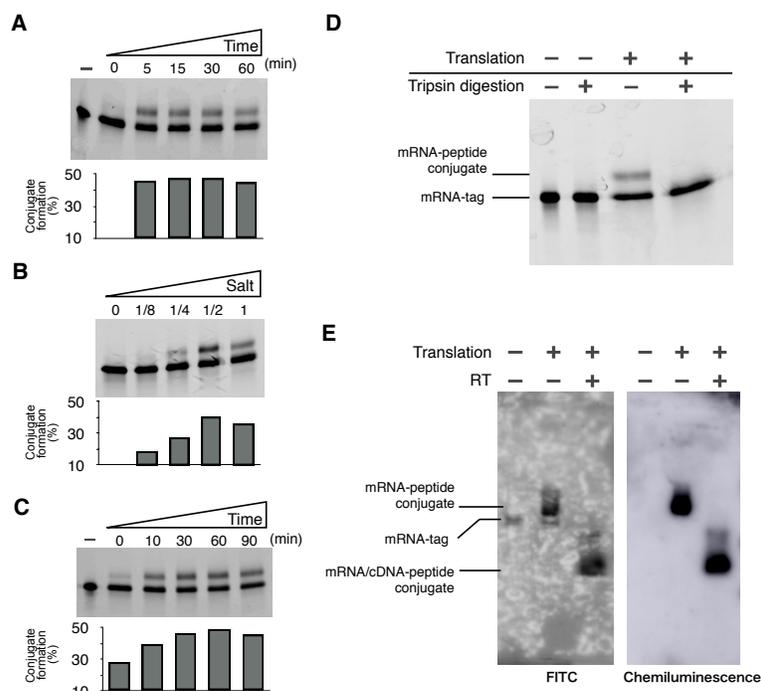
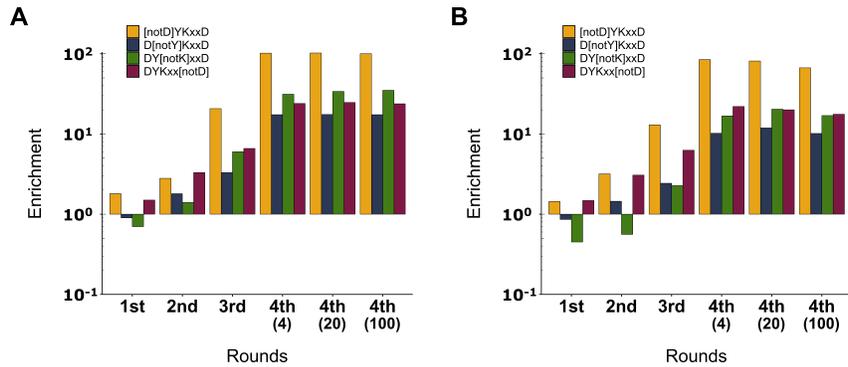
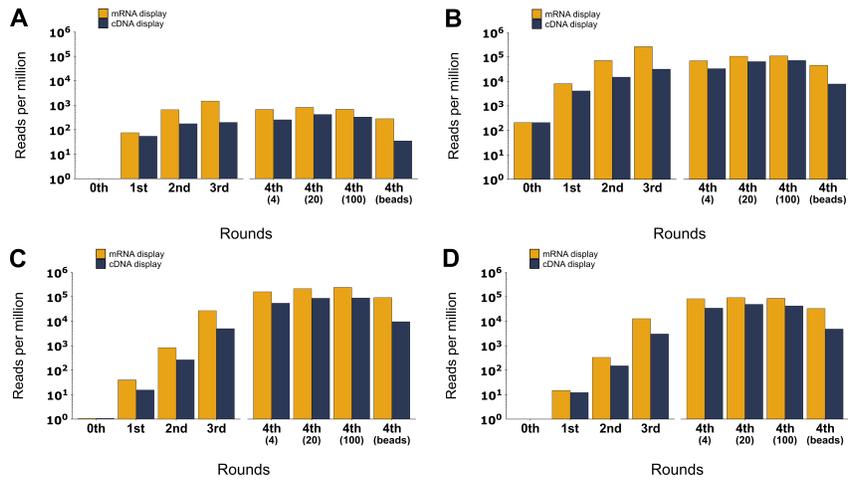


Figure 5. Round-by-round enrichment of the motifs lacking a single key residue from the core FLAG motif. Sequence reads for partial FLAG motif lacking each of the four key residues first D (yellow), Y (blue), K (green) and second D (dark red) are shown for each round during the (A) mRNA display and (B) cDNA display selection.







Hosted file

Table1-final.docx available at <https://authorea.com/users/329657/articles/456613-pure-mrna-display-and-cdna-display-provide-rapid-detection-of-consensus-binding-motif-via-high-throughput-sequencing>