ACS central science

http://pubs.acs.org/journal/acscii

Research Article

# Optimized Loopable Translation as a Platform for the Synthesis of Repetitive Proteins

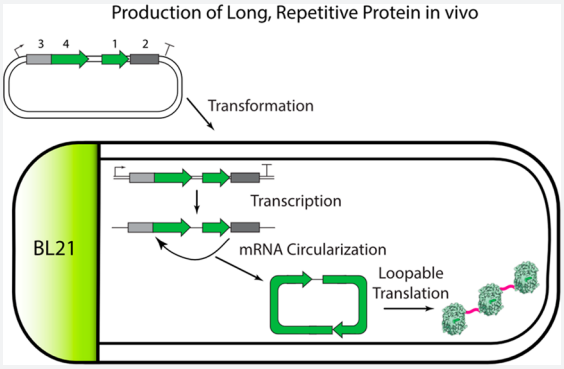
Sea On Lee, Qi Xie, and Stephen D. Fried*

Cite This: *ACS Cent. Sci.* 2021, 7, 1736–1750

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** The expression of long proteins with repetitive amino acid sequences often presents a challenge in recombinant systems. To overcome this obstacle, we report a genetic construct that circularizes mRNA *in vivo* by rearranging the topology of a group I self-splicing intron from T4 bacteriophage, thereby enabling "loopable" translation. Using a fluorescence-based assay to probe the translational efficiency of circularized mRNAs, we identify several conditions that optimize protein expression from this system. Our data suggested that translation of circularized mRNAs could be limited primarily by the rate of ribosomal initiation; therefore, using a modified error-prone PCR method, we generated a library that concentrated mutations into the initiation region of circularized mRNA and discovered mutants that generated markedly higher expression levels. Combining our rational improvements with those discovered through directed evolution, we report a loopable translator that achieves protein expression levels within 1.5-fold of the levels of standard vectorial translation. In summary, our work demonstrates loopable translation as a promising platform for the creation of large peptide chains, with potential utility in the development of novel protein materials.




## INTRODUCTION

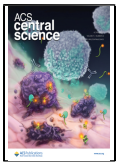
Proteins serve as the building blocks for functional, tunable materials across the tree of life. Spider silks (composed of spidroins[1−6]), connective matrices (predominantly composed of collagens[7−9]), biofilms,[10−13] and squid ring teeth (composed of SRT proteins[14−16]) are four examples of materials with distinct mechanical properties and biological functions whose properties are genetically encoded through their respective protein sequences. Nevertheless, there remain many opportunities and challenges toward the capacity to synthetically recapitulate (or improve upon) the biological processes responsible for the creation of these materials.

Recently, there have been several successes in the design of novel globular cagelike protein materials, generally based on motifs that self-assemble through symmetry.[17−20] On the other hand, natural fibrous proteins typically consist of highly repetitive low-complexity regions within polypeptides with long chain lengths and display self-assembly on many length scales, spanning from the nanometer (protein−protein interactions) to the micrometer (phase separation) and the millimeter (filamentization). Proteins of this type are generally less amenable to rational design and also are typically challenging to express.

Despite the sequence and functional diversity of extant fibrous proteins attested in nature,[1−16] some unifying features of these proteins are the presence of long, repetitive, low-complexity regions. Pioneering work by Kaplan and co-workers[1,6] demonstrated an approach to reproduce such a protein architecture in a recombinant expression vector. In their strategy, a short repetitive unit is autoligated into tandem repeats by availing of self-complementary sticky ends from two restriction enzymes.[1,6] The judicious choice of two enzymes that create identical sticky ends but recognize distinct recognition sites enables directionality and prevents the formation of inverse repeats (Figure 1A).

More recently, work by Demirel and co-workers appropriated rolling circle amplification (RCA)[21] as an alternative strategy to generate tandem repeat proteins, an approach that has been used to express large SRT polypeptides (42 repeats; 1260 residues) in *Escherichia coli* (Figure 1B).[14] Nevertheless, existing technologies to generate large repetitive proteins for recombinant expression in microbial hosts suffer from a number of major limitations: (i) Highly repetitive DNA sequences suffer from genetic instability due to spontaneous recombination in microbial hosts. (ii) Large plasmids containing tandem repeat proteins can be inconvenient for

**Received:** May 11, 2021
**Published:** September 24, 2021



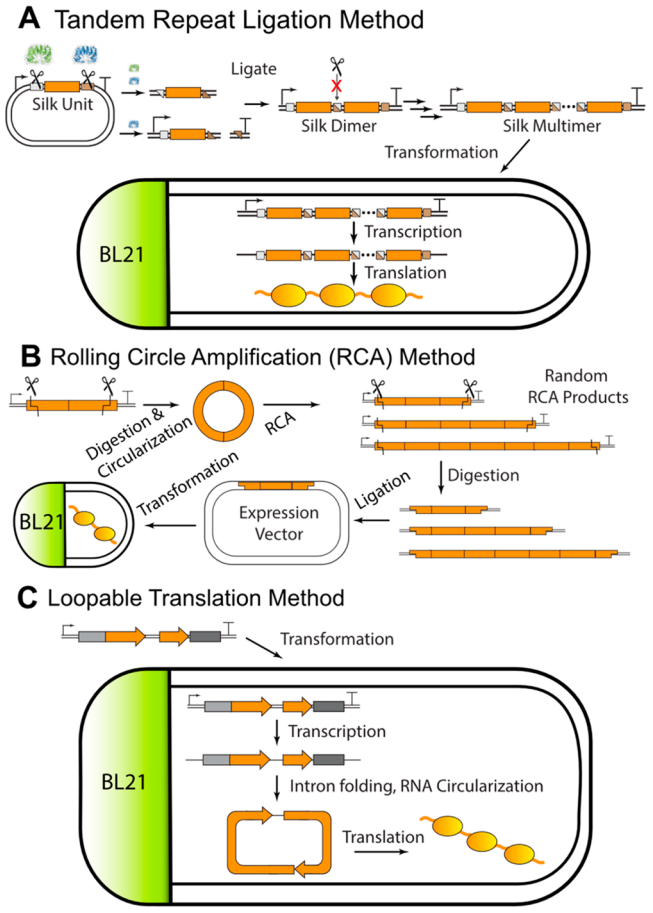



**Figure 1.** Approaches to generate tandem repeat proteins. (A) A plasmid encoding a monomeric unit is cut with two restriction enzymes to generate an insert. The original plasmid is cut with one of the two restriction enzymes to create self-complementary sticky ends, which allows ligation of the insert back into its original vector, creating a plasmid that encodes a tandem repeat. Upon ligation in a tail-to-head manner, the restriction site is destroyed, allowing directional cloning (inverse repeats can be removed by redigesting). This process can be repeated to generate larger $2^n$ multimers. (B) A short segment of DNA encoding a monomeric unit is digested and circularized *in vitro*. The circular DNA serves as the template for rolling circle amplification and generates a mixture of concatemers with different lengths. The desired size of the repetitive sequences can be selected by gel extraction, digested, and cloned back into an expression vector. (C) In loopable translation, a plasmid is created in which a segment of DNA encoding a monomeric unit is incorporated within a permuted self-splicing Group I intron (gray blocks). The plasmid is transformed into an expression strain, and its gene is transcribed into RNA, which is circularized. The circular mRNA translates into a repetitive protein product through ribosome looping.

routine molecular biology manipulations (e.g., PCR, sequencing, and transformation). (iii) The metabolic cost to replicate and transcribe large DNA and mRNA molecules undermines the goal of redirecting resources toward the creation of proteins. To address these limitations, we report an alternative approach based on loopable translation (Figure 1C).

In contrast to previous approaches to generate tandem repeat proteins (Figure 1A,B), the loopable translator does not require large repetitive DNA constructs; rather, it uses a permuted group I self-splicing intron to circularize an mRNA transcript, thereby allowing ribosomes to translate a region of interest many times in succession as a single polypeptide chain.

Expanding on the pioneering work of Ares and co-workers[22−24] on permuted introns, we demonstrate a reporter system that can quantify the efficiency of circular translation, allowing us to optimize its performance through rational considerations and by directed evolution. Our results suggest that such a system is a promising platform for the creation of fibrous proteins with a repetitive sequence and hierarchical structure.

## RESULTS

**Creation of a Loopable Translator Coupled to a Fluorescence Reporter.** Ares and co-workers demonstrated that the self-splicing group I intron within the thymidylate synthase (td) gene of T4 phage retains activity following circular permutation at the P6a stem (Figure 2A).[22−26] In this reorganized topology, the same two phosphoryl transfer reactions result in the circularization of an internal region, in contrast with the wild-type intron which catalyzes the splicing of two flanking exons. Such a circular mRNA can then be translated into a large repetitive protein. However, this remarkable activity is challenging to quantify, and indeed, tandem repeats of green fluorescent protein (GFP) are not fluorescent (possibly because of efficient nonradiative energy transfer within the polymer). In our design, superfolder GFP (sfGFP) was interrupted after its internal helix (at residue 52) with a T4 td intron (including 15 nucleotides of td exonic context flanking the intron). The exonic sequences were included to ensure proper interactions with the internal guide sequences (IGSs) of the T4 td intron (circularly permuted at P6a for circularization; Figure 2A,B). We then added a ribosome binding site (RBS) before sfGFP's N-terminus and a TEV protease cleavage site following sfGFP's C-terminus (Figure 2A,B) and cloned the sequence into an expression vector in which transcription was controlled by an arabinose promoter and an rrnB-T1 terminator. This plasmid (pBAD-$td_{3'}$-$sfGFP_{53\text{-}end}$-TEV-RBS-DB-$sfGFP_{1-52}$-$td_{5'}$; note that the downstream box sequence (DB) refers to a sequence reported to enhance translational efficiency[27,28]) was then cotransformed into *E. coli* BL21(DE3) with pRK793 (expressing a soluble form of TEV protease (TEVP)).[29,30] To abbreviate, we refer to our designed loopable translator construct as pBAD-tdTEVDB. In this system, circularized mRNA translates a GFP concatemer; however, the activity of TEVP liberates GFP monomers from the chain allowing for fluorescence to be monitored *in vivo* (Figure 2B).

pBAD-tdTEVDB was designed in a way that a nonfluorescent GFP fragment (residue 1−52) would be formed upon translation of a noncircularized RNA (translation was designed to terminate with a UAA stop codon that follows immediately after the end of the 5′ exonic sequence (5′ Ex); Figure 2A,B). Thus, a failure of the mRNA molecule to circularize will result in truncated GFP, which is nonfluorescent and serves as a negative control (Figure 2C). On the other hand, circularization (along with TEVP activity) will result in full-length sfGFP with a 10 amino acid scar between residues 52 and 53, arising from the natural exonic sequences of td (shown in orange in Figure 2A or as orange blocks in Figure 2B,C). This GFP is still fluorescent, and we use it as a positive control (Figure 2C). Preliminary fluorescence assays demonstrated that pBAD-tdTEVDB produces significant levels of fluorescence above baseline levels (Figure 2D, Figure S1; $P < 0.0001$ by Student's $t$ test), confirming that mRNA circularization occurs, in accordance with previous investi-


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

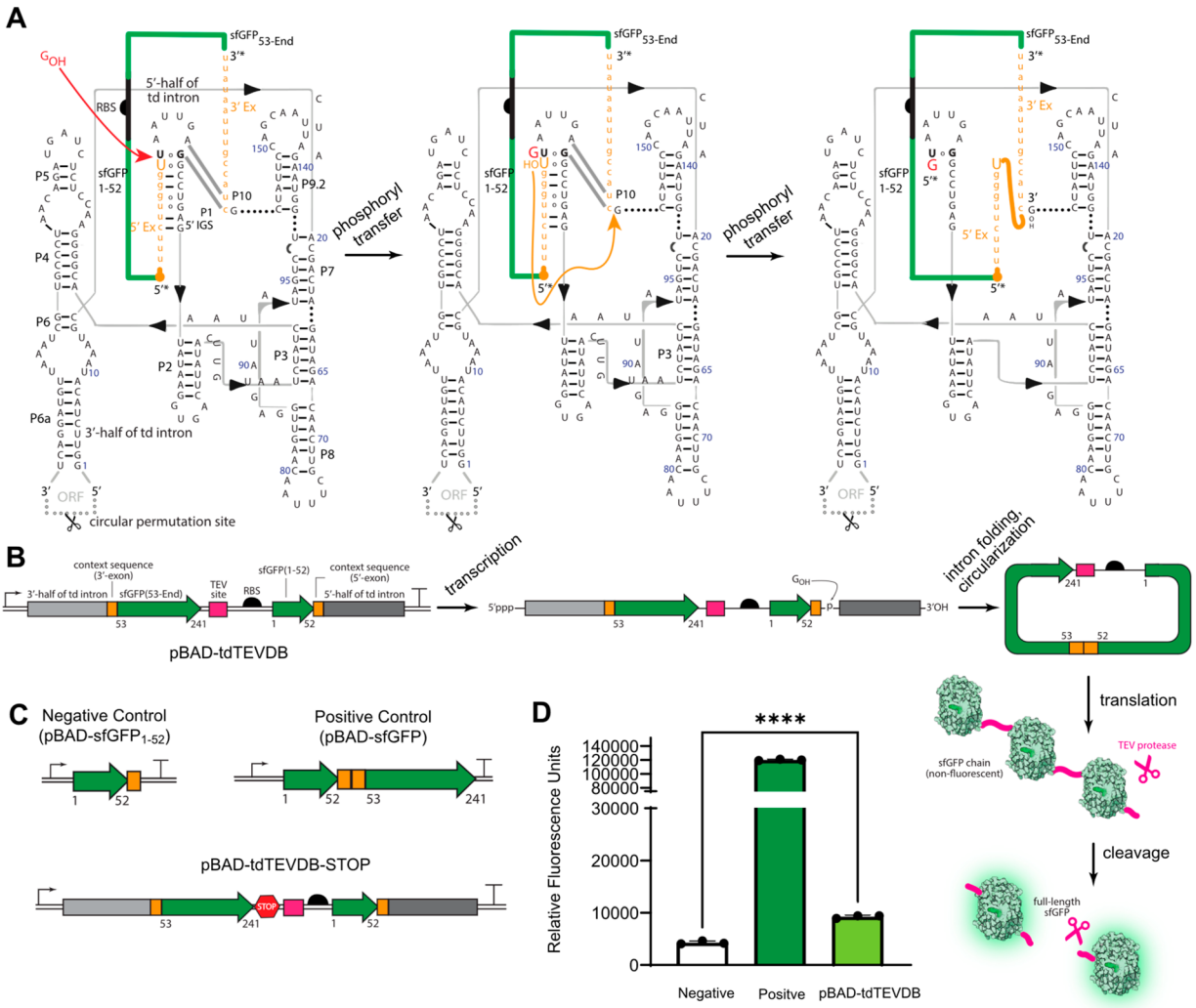



**Figure 2.** Design of a loopable translator with a coupled fluorescence GFP reporter. (A) Sequence, secondary structure, and mechanism of the self-splicing Group I intron from the thymidylate synthase (td) gene of T4 bacteriophage. This intron catalyzes two consecutive, site-specific phosphoryl transfer reactions, which, in the natural form, results in the splicing of the two flanking exons (orange). The 5′ splice site (marked with a red arrow) is selected by base-pairing between the 5′ exonic sequence (5′ Ex) and an internal guide sequence (IGS) at the beginning of the intron (P1), while the 3′ splice site (marked with an orange arrow) is selected by a short 2 bp stem formed between the 3′ exonic sequence (3′ Ex) and the edge of the P1 loop (termed P10, shown by two gray lines). An exogenous guanosine (red) is required to initiate the reaction and becomes prepended to the 5′-terminus of the intron sequence (black). 5′* and 3′* denote the termini of the RNA molecule in its natural form. In our construct, the circular permutation in the ORF region of P6a results in two new termini (labeled 5′ and 3′), and a permuted GFP sequence (green) is incorporated between 3′* and 5′*. In this reorganized topology, intron activity will result in circularization of the internal "exonic" region. (B) Design of a GFP fluorescence reporter system for RNA circularization. A plasmid is created (pBAD-tdTEVDB) in which a permuted superfolder GFP (sfGFP) gene is incorporated within the permuted intron. The GFP is split such that the N-terminal portion (residues 1−52) is placed downstream of the C-terminal portion (residues 53−241). In between these two coding regions are inserted a TEV protease site and a ribosome binding site (RBS) along with an enhancing downstream box (DB) in frame with the GFP coding sequence. After intron folding and RNA circularization, an mRNA is formed which is competent to recruit ribosomes and generate full-length GFP. Because polymeric GFP is found to have low fluorescence, this system is coexpressed with TEV protease, which can liberate fluorescent GFP monomers from the primary chain product. (C) A negative control plasmid (pBAD-sfGFP$_{1-52}$) encodes the protein product that would form in the absence of circularization (with the same promoter, origin, and selectable marker as pBAD-tdTEVDB). A positive control plasmid (pBAD-sfGFP) encodes the protein product that would form upon circularization. Importantly, it differs from a wild-type sfGFP in that it has a 10-residue long "scar sequence" in between residues 52 and 53, which result from exonic context sequences (represented as orange boxes) that were retained to ensure proper base-pairing with the IGSs. pBAD-tdTEVDB-STOP is identical to pBAD-tdTEVDB except that a stop codon is placed at the end of GFP. (D) pBAD-tdTEVDB generates a fluorescence signal that is significantly higher than the background level ($P < 0.0001$ by Student's $t$ test) but represents 4.2% of the signal of the corresponding positive control ($n = 3$). Fluorescence measurements were conducted in biological triplicate.

gators' work on this system.[22−24] Nevertheless, our fluorometric readout makes it clear that significantly less GFP (4.2%; Figure S1) is produced compared to a positive control in which proteins are expressed via standard (vectorial) translation, suggesting that either (i) RNA circularization, (ii) ribosomal initiation, or (iii) TEV activity is limiting. To test for possible toxicities associated with loopable translation, the three constructs corresponding to a negative control, positive control, and pBAD-tdTEVDB were expressed in BL21+pRK793 cells at 30 °C, and growth curves were


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

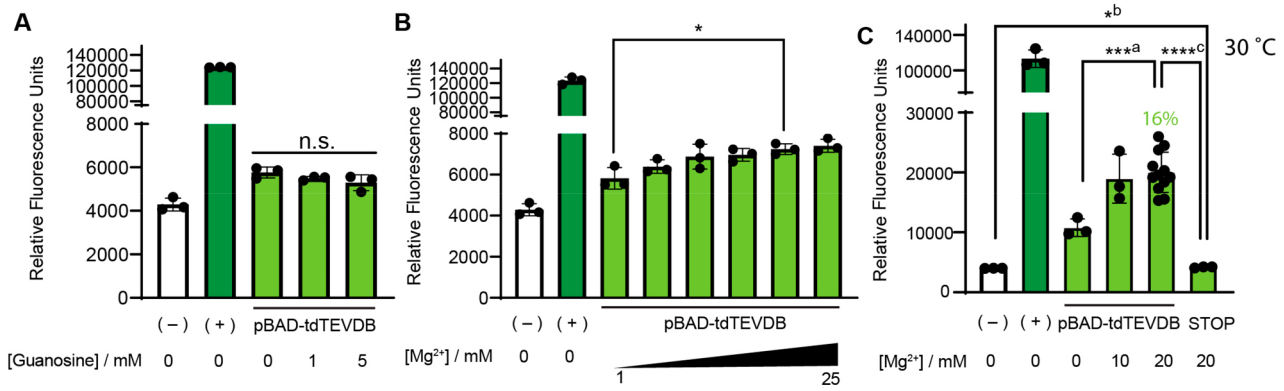



**Figure 3.** High Mg and reduced temperature enhance loopable translation. (A) Bar chart showing the levels of fluorescence of several constructs (negative control (−), positive control (+), and pBAD-tdTEVDB) expressed at 37 °C with varying concentrations of guanosine supplemented to the growth media during expression assays. Additional guanosine had no significant effect on the fluorescence signal generated by pBAD-tdTEVDB. (B) Bar chart showing the levels of fluorescence of the same set of constructs expressed at 37 °C with varying concentrations of $MgCl_2$ supplemented to the growth media during expressions assays. Higher $MgCl_2$ concentrations had a beneficial effect; relative to 1 mM $MgCl_2$, fluorescence levels in 20 mM $MgCl_2$ were 1.2-fold higher ($P$-value = 0.01). (C) Bar chart showing the levels of fluorescence of the same set of constructs with varying $MgCl_2$ supplemented to the growth media, but with expressions carried out at 30 °C instead of 37 °C. Lower temperature had a significantly beneficial effect on fluorescence signal (2.8-fold at 20 mM $MgCl_2$, $P$-value <0.0001). Moreover, at the lower temperature, the fluorescence signal significantly benefitted from high concentrations of $MgCl_2$ (a: 1.9-fold, $P$-value = 0.0006). Under these conditions, the loopable translator generated 16% of the fluorescence of the positive control. Fluorescence levels from pBAD-tdTEVDB-STOP (see Figure 2C, labeled "STOP") were slightly higher than those from the negative control (b: 1.05-fold, $P$-value = 0.04), but much less than the loopable translator (c: 4.8-fold, $P$-value <0.0001). Fluorescence measurements were conducted in biological triplicate; statistical tests were conducted with Student's $t$ test.

measured (Figure S2). We found no significant difference in growth rates across these three strains.

**Rational Improvement of Activity from Loopable Translator.** We next sought to improve the activity of pBAD-tdTEVDB by considering a number of factors that could influence the three considerations above. Group I self-splicing introns require two cofactors: magnesium cations and a guanosine nucleoside.[31,32] Guanosine is required because the intron uses a two-step mechanism that begins with the free nucleoside cleaving the phosphodiester linkage between the 5′-exon and the intron by prepending itself to the 5′-most nucleotide of the intron, and divalent cations are required to stabilize the tertiary fold of the catalytic core.[25,31−34] We found that GFP expression from pBAD-tdTEVDB was not strongly dependent on guanosine (Figure 3A); however, fluorescence could be markedly improved by 147% (from 5.09% to 7.46% of the positive control) by increasing the concentration of $MgCl_2$ in the growth media up to 20 mM (Figure 3B). These findings are broadly consistent with the facts that intracellular total guanosine ($(p)(p)(p)G_{OH}$) concentration is quite high in *E. coli* (5 mM, higher than the intron's $K_M$[35]); however, *free* $Mg^{2+}$ in the cytosol is relatively scarce (0.01−1 mM), and many *in vitro* studies of ribozymes show a strong reliance on (and benefit from) high divalent cation concentrations.[35]

Strikingly, we observed an enhancement in GFP expression when cells were incubated at a lower growth temperature (30 °C), with the fluorescence levels approaching 16% of the positive control (Figure 3C). We speculate two potential reasons why this could occur: (i) TEVP is an aggregation-prone protein, and reduced temperatures could improve its solubility in the *E. coli* cytoplasm; and (ii) lower temperatures could shift equilibria to favor certain catalytic elements on the intron, some of which consist of only two base pairs (e.g., P10; cf. Figure 2A).[36]

As a control, we also performed a fluorescence assay with a modified construct in which a stop codon is inserted after the GFP coding sequence (pBAD-tdTEVDB-STOP, Figure 2C), such that each ribosomal initiation event on circularized mRNA would only result in a single GFP protein being synthesized (noncircularized mRNA would still result in truncated GFP). Fluorescence levels decreased markedly, down to levels only slightly higher than the baseline level associated with the negative control (Figure 3C). This result suggests that the fluorescence signal achieved by pBAD-tdTEVDB arises from ribosomes transiting circular mRNAs many times without stopping. Using the slightly higher (but still statistically significant) fluorescence signal of pBAD-tdTEVDB-STOP relative to the negative control (1.04-fold, $P$-value = 0.04 by Student's $t$ test), we can estimate the loop count (number of transits per initiation) to be ∼90 ± 55 using

$$\text{loop count} = \frac{F_{\text{loop}} - F_{-}}{F_{\text{stop}} - F_{-}}$$

where $F$ corresponds to average fluorescence, and "loop", "−", and "stop" correspond to the tdTEVDB, negative control, and tdTEVDB-STOP constructs, respectively.

A number of other variations we considered did not improve the efficiency of GFP fluorescence, including an attempt to stabilize the permutation site at P6a (Figure S3, Supporting Information Text 1), flanking the TEV cleavage site with flexible linkers (GGSGGGSGG) to facilitate the access of TEV protease (Figure S4, Supporting Information Text 2), the replacement of a classic TEV protease with several mutants that have been reported to be more active (Figure S5, Supporting Information Text 3),[37−39] and the replacement of the RBS with an internal ribosome entry site (IRES) that has been reported to efficiently recruit bacterial ribosomes (Figure S6, Supporting Information Text 4).[40] We were able to detect a very low activity using an alternative circularization strategy that is based on ribozymes and RtcB ligation (Figure S7, Supporting Information Text 5),[41] though we note that this strategy was developed in mammalian cells, and not in *E. coli*.


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

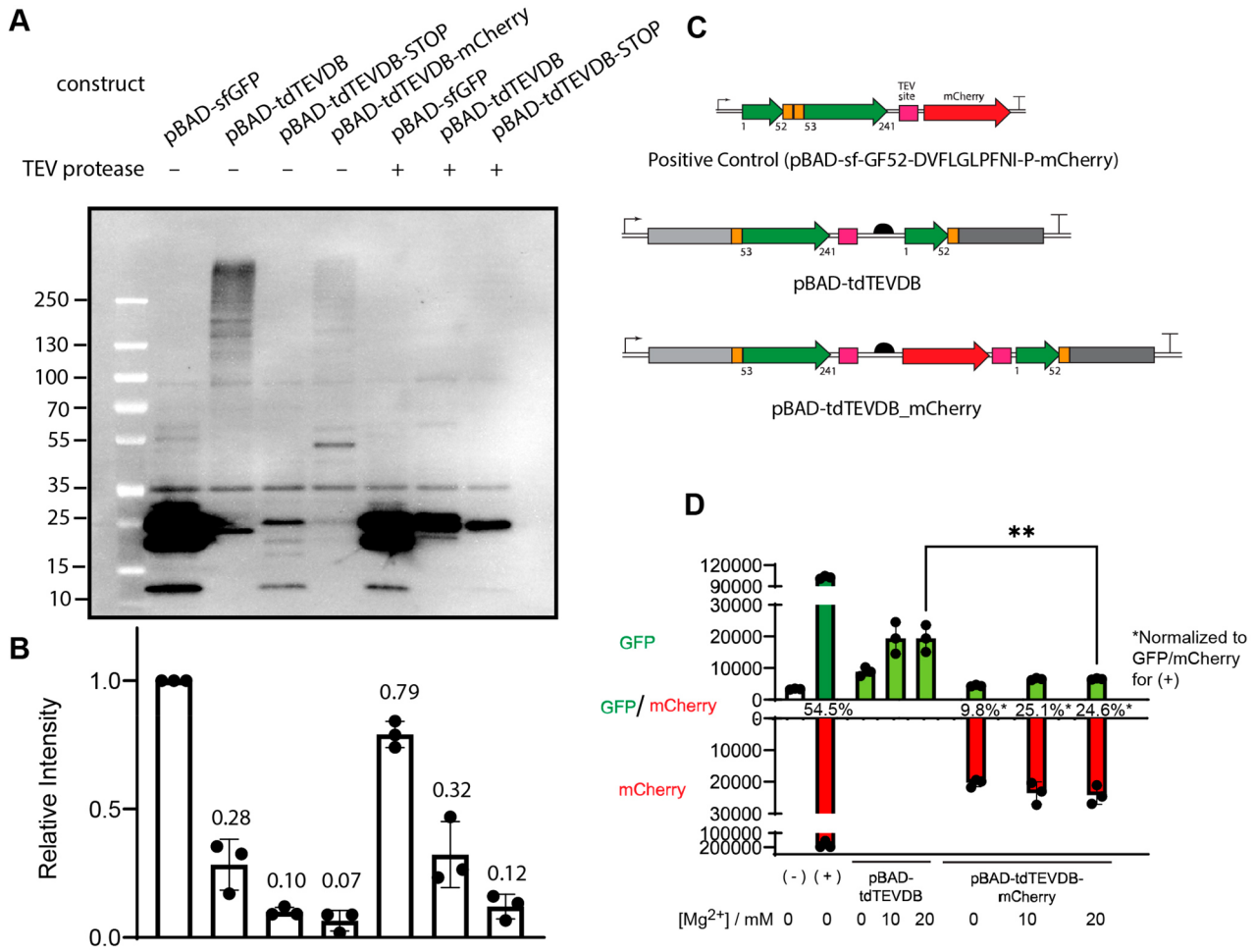



**Figure 4.** Verification of RNA circularization and polyGFP synthesis. (A) Anti-His Western blot image showing the protein products of pBAD-sfGFP, pBAD-tdTEVDB, pBAD-tdTEVDB-STOP, and pBAD-tdTEVDB-mCherry that were expressed with and without the pRK793 plasmid (expressing TEV protease). pBAD-tdTEVDB in the absence of TEV protease generated proteins of high molecular weight. Expression in the presence of TEV protease generates a species with the molecular weight of monomeric GFP (25 kDa). (B) Densitometry analysis showing average intensities relative to a positive control, for three biological replicates of the Western blot. (C) Construct maps of the positive control for the mCherry experiment (pBAD-sfGFP(52-DVFLGLPFNI)-mCherry), pBAD-tdTEVDB, and pBAD-tdTEVDB-mCherry. pBAD-tdTEVDB-mCherry was designed to form a larger mRNA loop compared to that of pBAD-tdTEVDB, in which GFP would be expressed only upon circularization, whereas mCherry would be expressed independently of circularization. (D) Bar chart showing the levels of fluorescence from the constructs shown in part C at 30 °C with varying concentrations of $MgCl_2$. The positive control shows the native difference in fluorescence between GFP and mCherry and serves as a normalization factor. With pBAD-tdTEVDB-mCherry, 25% of the target mRNA achieved circularization at 20 mM $Mg^{2+}$. The larger mRNA loop was detrimental for circularization, leading to lower levels of GFP fluorescence compared to those of pBAD-tdTEVDB ($P$ = 0.0066 by Student's $t$ test).

**Visualization of Concatemer Formation and Verifying mRNA Circularization.** Because polymeric GFP (polyGFP) is not fluorescent,[23] our fluorescence assays require the activity of TEVP to generate a readable signal. In contrast, polyGFP can be visualized directly as a protein of high molecular weight via Western blot. Indeed, we found that, in the absence of the plasmid that expresses TEVP, pBAD-tdTEVDB generates a set of protein products with high molecular weights (>250 kDa) that are specific for an anti-His antibody (the GFP coding sequence contains a His-tag, Figure 4A, Figure S11). In contrast, in the presence of TEVP, pBAD-tdTEVDB generates a single band at the expected molecular weight for monomeric GFP (25 kDa), identical to a positive control in which GFP is expressed by linear translation (Figure 4A). Moreover, the introduction of a stop codon within the mRNA loop completely abrogates the high-molecular-weight features, as expected, and also significantly reduces protein expression level—showing that polyGFP synthesis is dependent on a ribosome iterating numerous times on circular mRNA. Densitometry analysis conducted on blots of biological triplicates (Figure 4B) showed that, relative to linear translation, overall protein expression from the loopable translator was down ∼3.6-fold, and protein expression from pBAD-tdTEVDB-STOP was down ∼10-fold, consistent with what we found in fluorescence assays (Figure 3C).

To provide independent verification that mRNA is circularized by the split-intron architecture, we designed a dual-fluorescence ratiometric construct (Figure 4C). The construct, called pBAD-tdTEVDB-mCherry, places a full-length mCherry open reading frame between the RBS and the N-terminal fragment of sfGFP. In this construct, mCherry is expressed independently of circularization while GFP, as before, requires mRNA circularization to be fully synthesized. From this construct, we found that the GFP signal was 4-fold


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

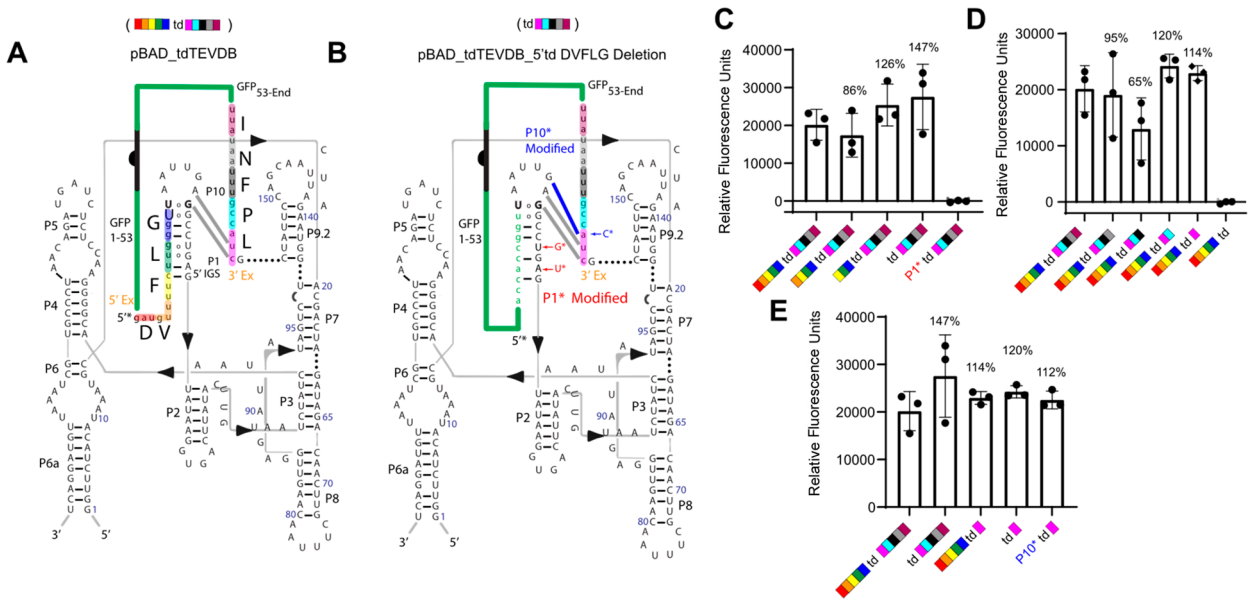



**Figure 5.** Minimal context requirements for loopable translator. (A) Secondary structure of the loopable translator highlighting the 10 amino acid "scar sequence" that is incorporated into the loop because of the inclusion of 30 nt of exonic context retained from the natural td gene (15 nt from the original 5′ exon, and 15 nt from the original 3′ exon). Each codon triplet is color coded, and deletions of this context sequence are represented through their corresponding color blocks in bar charts C−E. (B) Secondary structure that is formed when all 15 nt (coding for DVFLG) of the 5′ exonic context sequence are deleted. The nucleotides that correspond to residues 50−52 of GFP (green) replace the original 5′ exon to pair with the IGS. Nucleotides marked with * (in red) represent compensatory mutations in the P1 IGS (termed the modified P1*). (C) Bar chart showing the level of fluorescence from a truncation series in which the 5′ context sequence was deleted one codon at a time. (D) Bar chart showing the level of fluorescence from a truncation series in which the 3′ context sequence was deleted one codon at a time. (E) Finding the minimal context requirements. Overall, the entire 5′ context sequence can be deleted, and all but the last 3 nucleotides of the 3′ context sequence (which form P10) can be deleted. Strengthening P10 (P10*) did not have a beneficial effect. All fluorescence measurements were conducted in biological triplicate.

less than that of mCherry (after correcting for background and intrinsic difference in sfGFP's and mCherry's fluorescence intensity, Figure 4D), suggesting that 25% of the tdTEVDB-mCherry mRNA is circularized in the steady state. We also found that the overall GFP fluorescence signal in this larger 1563-nt loop was significantly reduced (3-fold) relative to the smaller 816-nt loop created by pBAD-tdTEVDB, which is consistent with the Western blot showing 4-fold reduced polyGFP from pBAD-tdTEVDB-mCherry relative to pBAD-tdTEVDB (Figure 4A,B). Tentatively, our explanation for this observation is that forming the larger loop has a higher entropy cost to bring the two portions of the split-intron together and assemble the active ribozyme. If this difference in GFP signal can be ascribed solely to the fraction of mRNA molecules that are circularized, these data suggest that as much as ∼75% of the tdTEVDB mRNA is circularized in the steady state. Although there could be other factors involved in GFP's translational efficiency from these two circular RNAs, Northern blot analysis of the RNA from cells harbouring pBAD-tdTEVDB found that 50% of the GFP-containing RNA was circularized (Figure S8). To summarize, evidence from Western blots, Northern blots, and our dual-fluorescence assay supports the view that the split-intron can achieve a steady-state fractional circularization level of 50−75% of 800-nt regions. We found this level suitable for our purposes, though it is noted that fractional circularization could be potentially optimized with further directed evolution of the intron.

**Identifying the Minimal Context Requirements for Circularization.** The design of pBAD-tdTEVDB would incorporate a 10 amino acid scar at the junction site where looping occurs. Hence, we asked what the minimal "exon context" requirements for intron activity are by "walking back" the 15 nucleotides of native exon context on each flank. Classic experiments on self-splicing introns demonstrated the essentiality of the final U·G wobble pair at the 5′ splice-site and the formation of the P1 stem between the 5′-exon and the beginning of the intron sequence (5′ IGS).[25] We successively deleted td's 5′-exon context (GAU/GUU/UUC/UUG/GGU, encoding DVFLG) one codon at a time (Figure 5A,C and Figure S9) and discovered that, remarkably, all 5 codons of td's 5′-exon context could be removed, resulting in a slight improvement in fluorescence activity. Inspection of the RNA molecule created following this deletion revealed that coincidentally sfGFP's residues 50−52 are Thr-Thr-Gly, which are encoded by ACC/ACC/GGU, a sequence that conserves the three critical nucleotides (GGU) immediately before the intron (forming two strong base pairs, then a wobble base pair, Figure 5B). Moreover, like the wild-type intron, this sequence creates four total Watson−Crick base pairs along P1. We wondered if the circularization efficiency could be further improved by stabilizing P1, which we tested by mutating IGS nucleotides (U12 and A14) to a base pair with the native sfGFP nucleotides (P1* in red in Figure 5B). However, surprisingly, the mutation reduced fluorescence down to baseline levels (Figure 5C), suggesting that it abrogates circularization activity. These observations imply that the presence of several (possibly two) non-Watson−Crick base pairs in P1 is necessary for activity, rather than incidental. Collectively, these observations suggest that the minimal requirements for proper selection of the 5′ exon fragment by P1 are (i) that it must end in GGU and (ii) that the preceding four nucleotides must make two Watson−Crick (but not more) base pairs with the intron's 5′-IGS.


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

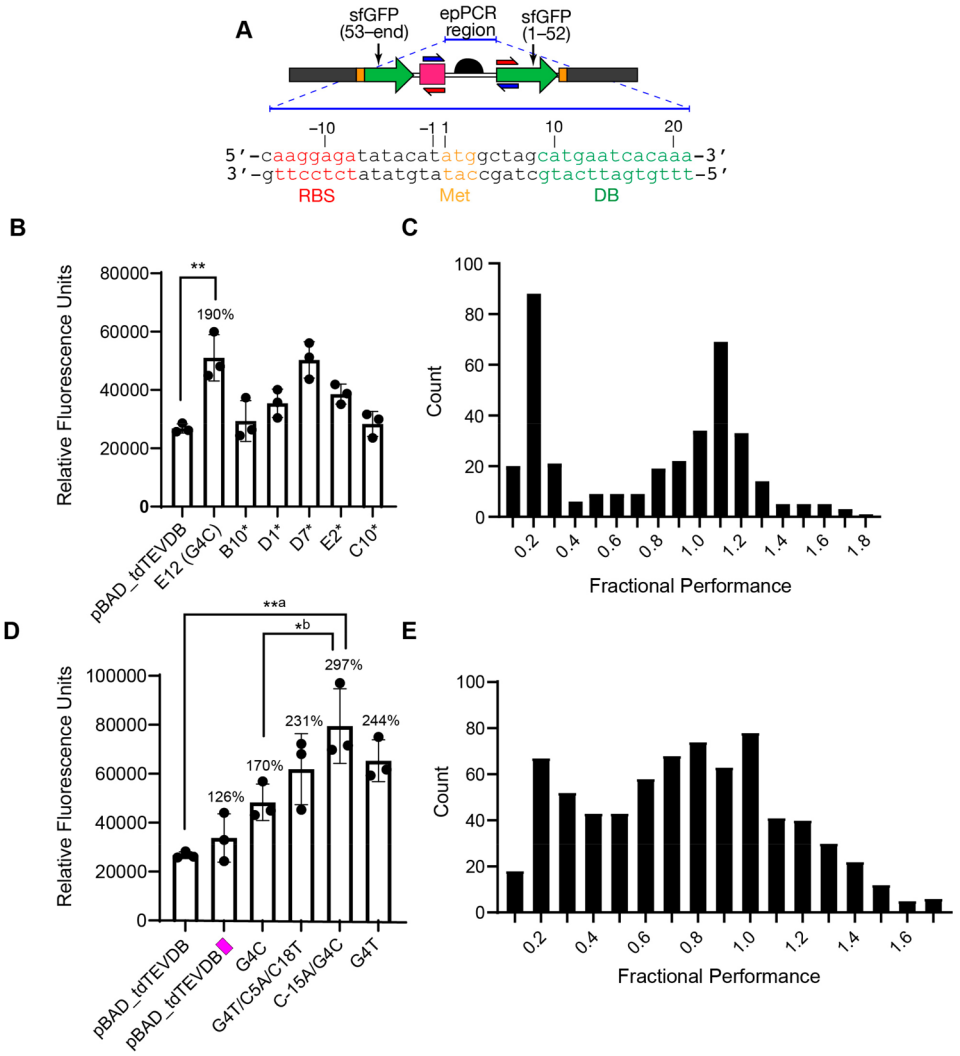



**Figure 6.** Directed evolution on the initiation sequence of loopable translator with minimum context. (A) Error-prone PCR was performed on a short 36 bp amplicon encoding the initiation region [ribosome binding site (RBS), spacer, initiator methionine, and a downstream box (DB)] of the GFP reporter gene. (B) Bar chart showing the fluorescence signals from the top-performing constructs selected from the first round of directed evolution ($n$ = 3 biological replicates following primary screen). A construct with a point mutation (G4C) generated a signal that is significantly higher than the wild-type (pBAD-tdTEVDB) (1.9-fold; $P$ = 0.0065). (C) Histogram showing the fluorescence signal of 372 constructs screened in the first round of directed evolution relative to the wild-type. (D) Bar chart showing the fluorescence signal from the top-performing constructs selected from the second round of directed evolution ($n$ = 3 biological replicates following primary screen). A construct with an additional mutation (C-15A/G4C) generated a signal that was ca. 3-fold higher than the wild-type (a: $P$ = 0.0039) and ca. 2-fold higher than the single mutant (b: $P$ = 0.033). (E) Histogram showing the fluorescence signal of 720 constructs screened in the second round of directed evolution relative to G4C. All statistical tests were conducted using Student's $t$ test

We also successively abrogated td's 3′-exon context region one codon at a time (Figure 5A,D; Figure S9) and confirmed that only the initial three nucleotides (CUA; the ones closest to the exon and marked in magenta) are required for activity. This is consistent with previous work[36] which showed that the selection of the 3′-splice site is controlled by a combination of a tertiary structure and a short (2 bp) stem (called P10) formed between the 3′-exon (3′ Ex) and P1's loop. A deletion of the first 3 nucleotides of the 3′-exon (CUA) abrogated all circularization activity (Figure 5D), confirming the importance of these nucleotides for splice site selection and therefore circularization.

Putting these observations together, we next sought to identify the minimal exon context requirements for intron activity, so that future use of the loopable translator system would impose minimal scar sequences at the junction site (Figure 5E). Combining the insights from these two deletion series, we found that complete removal of the 5′ td-exon context and removal of the 3′ td-exon context up to the first three nucleotides (CUA) were tolerated by the permuted td intron and, in fact, provided a small (20%) activity increase with respect to the wild-type (Figure 5E). Next, to determine if intron activity could be enhanced by strengthening P10, we modified the sequence of the initial three nucleotides from CUA to CUC, which would form a third Watson−Crick base pair with the P1 loop (P10* in blue in Figure 5B,E). The resulting construct generated a signal that was slightly lower than the construct with only CUA, suggesting that two base pairs of complementarity are indeed the optimal strength for this critical interaction.

Combining the requirement for CUA immediately after the intron and the requirement of GGU immediately before the


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

intron, our data imply that mRNA circularization can be accomplished with a scar as small as two amino acids at the junction region.

**Improving Initiation on Circular mRNA by Directed Evolution.** Because 50−75% of mRNA generated by pBAD-tdTEVDB is circularized (cf. Figure 4 and Figure S8), we reasoned that initiation of ribosomes on circular mRNA could be limiting the efficiency of protein expression. In *E. coli*, translational initiation is mediated by several factors. The interaction between the Shine−Dalgarno sequence (i.e., RBS) and the anti-Shine−Dalgarno on 16S rRNA is a well-known determinant, and many expression vectors use "strong" RBSs to increase expression.[42−44] Additionally, ribosomal protein S1 (bS1) plays a critical—but less well understood—role in mediating initiation,[45,46] which we hypothesized could be less adept at initiating translation on circular mRNA, because S1 binds upstream of the RBS, which, for most mRNA molecules, would be proximal to the 5′ terminus. On the other hand, on circular mRNAs, RBSs would not be proximal to any terminus. To test this hypothesis, we examined whether directed evolution on the initiation region (Figure 6A) would enable the creation of a modified initiation sequence better suited for circular mRNA.

Specifically, we sought to introduce several mutations into a short 36 bp region that covers the canonical RBS, its spacer, the initiator codon, and the downstream box. We found that conventional error-prone PCR methods were ill-suited for achieving the desired mutation density in this small amplicon (1−2 mutations per 36 bp) while also maintaining low levels of bias;[47−50] hence, we developed a modified error-prone PCR method based on iterating between dilution and reamplification with error-prone touchdown PCR to suppress accumulation of incorrect products. The details of this modified method are reported elsewhere.[51]

This dilution/reamplification error-prone PCR method was used to introduce mutations into a 36 base pair region, which was then cloned back into the pBAD-tdTEVDB vector linearized with the reverse complement primers (red primers, in Figure 6A), via in-fusion cloning (see the Materials and Methods section). The library DNA was isolated and retransformed into BL21(DE3) cells harboring the pRK793 plasmid, and hundreds of colonies were selected from agar plates at random for inoculation into 96-well plates to screen for beneficial mutations by fluorescence. In the first screen, we assayed fluorescence from 372 constructs and found a bimodal distribution (Figure 6C) with many clones having a neutral phenotype (fractional performance close to 1) or with very low activity (fractional performance close to 0.2). A small number of variants had improved levels of translational efficiency, up to 1.9-fold, including a variant E12 (which contained the mutation G4C, immediately following initiator methionine; Figure 6B). Because other high performers by this screen contained mixtures of sequences upon performing sequence analysis, we focused on G4C as a starting point for a second round of diversification and screening. We used the pBAD-tdTEVDB$^{G4C}$ plasmid as a template for error-prone PCR to incorporate further mutations in the initiation region and cloned back into the pBAD-tdTEVDB backbone to perform a second screen. This time, we assayed 720 constructs and discovered mutants that improved translational efficiency 1.8-fold again (Figure 6E). The best performing mutant, C-15A/G4C contains an additional mutation that apparently "extends" its RBS sequence to now include a run of 8 purines; other successful clones contained other mutants at position 4. Overall, the double mutant C-15A/G4C in the initiation region is ca. 3-fold more active at producing GFP relative to our starting construct (Figure 6D). These experiments appear to support our hypothesis that ribosomal initiation rates hampered translation on circular mRNA, though specific alterations to the initiation region can mitigate this limitation. Finally, if we combine these improvements in translational initiation with lowered temperature and increased $Mg^{2+}$ concentrations, the result is a ∼14-fold improvement in translational efficiency relative to the original loopable translator (cf. Figure 6D and Figure 2D, ca. 75 000 and 5260 relative fluorescence units, respectively, following background subtraction). Overall, protein expression from these optimized circular mRNAs is within a factor of 1.5 relative to the expression levels from standard linear mRNA.

## DISCUSSION AND FUTURE DIRECTION

Loopable translation has several traits that could make it an attractive option for the preparation of highly repetitive proteins. Because it does not require the creation of repetitive DNA sequences, the resulting genetic constructs would be expected to be more stable and less susceptible to homologous recombination. Moreover, the absence of repetitive sequences renders these plasmids amenable to many more molecular biology and synthetic biology manipulations—such as PCR, Gibson assembly, and recombineering—which are challenging (or impossible) to conduct with highly repetitive DNA sequences. These features should allow for greater interoperability; that is, it is facile for any researcher to "swap out" the GFP sequence in pBAD-tdTEVDB with an arbitrary sequence via a single Gibson assembly. All of these features additionally make this system much more amenable to creating combinatorial libraries and performing directed evolution, as we have demonstrated. Due to the universal nature of the autocatalytic intron splicing reaction (requiring only magnesium and guanosine as cofactors), we expect that circularization should be functional in a range of microbial hosts. Finally, the repeat number (loop count) from such a system could be very high—as the data with the pBAD-tdTEVDB-STOP construct (Figure 3C) would suggest—possibly only limited by the presumably rare occurrence of spontaneous frameshifting[52] or possibly nonspecific peptide release from the ribosome.

Initially, our study focused on coupling mRNA circularization to a fluorescence-based reporter (sfGFP), so expression from circular mRNA could be quantified under a range of conditions and for a variety of constructs. This reporter proved to be valuable in identifying the importance of a high concentration of magnesium (20 mM) and lower growth temperature (30 °C) for a high level of expression, likely because both factors help folding and stability of the intron, which consists of several "delicate" components, notably P10 which specifies the 3′ splice site in the catalytic machinery with only 2 base pairs. The benefit of these factors is supported by the higher RNA circularization levels evidenced in our Northern blot compared to previous reports (cf. Figure S8 and refs 22−24). Although expression levels from this system were initially low, we hypothesized that improved translational efficiency could be achieved via modification of the initiation region. The finding that we were able to enhance translational efficiencies nearly 3-fold upon screening a relatively small number of clones could suggest that greater improvements


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

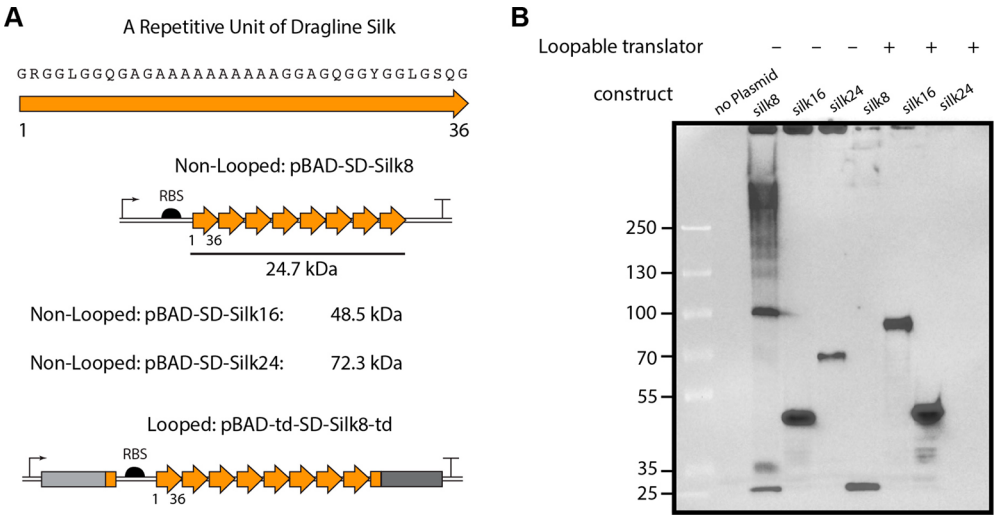



**Figure 7.** Applying pBAD-tdTEVDB to producing spider silk spidroin. (A) Illustration of the nonlooped and looped constructs containing repetitive units of dragline silk. A repetitive unit with 36 amino acids was chosen as a monomeric unit. Six constructs were generated, comprising 8, 16, or 24 tandem repeats of a repetitive protein unit derived from major ampullate spidroin protein 1 (MaSp1), cloned into either a standard expression vector or the loopable translator. (B) Anti-His Western blot image demonstrating the protein products of the six constructs. Apparent low molecular weights from the loopable translator could be due to high insolubility and/or instability of the resulting protein products.

could be achieved with further rounds of directed evolution, an effort that we are currently undertaking.

The next major direction for this line of research is to deploy the loopable translator to create long repetitive fibrous proteins that could potentially have enhanced material properties (see the Materials and Methods section). In preliminary experiments, we were able to clone sequences corresponding to the major ampullate spidroin protein 1 (MaSp1) into the loopable translator motif and attempted to express high-molecular-weight spidroins (Figure 7A). Surprisingly, we detected that nonlooped MaSp1 proteins could generate species of high molecular weight (>250 kDa), while looped MaSp1 proteins resulted in very low expression levels overall (Figure 7B). One possible explanation for these findings is that the spidroin sequence that we used in this study is very hydrophobic and aggregation-prone,[53−56] meaning that even proteins of low nominal molecular weights might form aggregates relatively easily, while higher-molecular-weight spidroins (such as those that would be expected to form from the loopable translator) might form insoluble entities that cannot be stably expressed in the *E. coli* cytosol and instead engage various forms of cellular stress−response which terminate their translation and initiate their degradation.

Future studies will therefore have to balance the potential benefits of long fibrous proteins with the requirement of maintaining solubility and stability within the cytosol prior to downstream material processing. This might be possible by incorporating solubilizing domains, which naturally occur at the N- and C-terminus of spidroin proteins and play important roles in maintaining these proteins in a soluble micelle-like state before they are concentrated and pulled into fibers in a spider's spinerets.[57−59] Cloned into our loopable translator motif, the solubilizing end domains will be expressed many times along with the fibrous monomeric unit and would potentially allow the expression of long repetitive fibrous protein materials with desired properties. It may also be that the loopable translator would be better suited at translating very long proteins that consist of domains which are initially soluble and then fold-switch into an aggregating form, seeing as our results have demonstrated the viability of expressing soluble domains as long concatemers with relative ease (cf. Figure 4). Curli proteins such as *E. coli* CsgA are potentially attractive in this regard as they have been extensively engineered[10,11,60,61] and are translated in a primary form that maintains solubility in the *E. coli* periplasm prior to amyloid growth nucleated following secretion.[62,63] Looped extracellular matrix (ECM) proteins such as fibronectin[64] and collagen[65] are other potential targets for applications as tissue engineering scaffolds,[66] where the loopable translator system would provide the benefits of facile genetic tuning and incorporation of signaling protein modules. Hence, we are optimistic for the future outlook of protein materials research efforts building off the tools that we have herein described.

## MATERIALS AND METHODS

**Cloning pBAD-tdTEVDB Plasmid Construct and Controls.** To create the negative control plasmid (pBAD-sfGFP1-52), a plasmid encoding the full-length superfolder GFP (pBAD-sfGFP) was used as a template for PCR and amplified with primers Delete53-f and Delete53-r (Table S1 and Figure S10) using the Q5 DNA polymerase (NEB) according to the manufacturer's protocol. The construct was designed in a way that the truncated GFP is under the control of the same arabinose-inducible promoter and terminator. The PCR product was assessed by 0.8% agarose (0.5× TBE) gel electrophoresis for the correct molecular weight, DpnI (NEB) digested (10 U were added to the PCR reaction and incubated at 37 °C for 30 min and then 80 °C for 20 min to inactivate), column purified using the DNA clean-and-concentrate kit (Zymo) according to the manufacturer's protocol, and quantified using a Nanodrop OneC (Thermo). The DNA was ligated using the QuickChange strategy by directly transforming it into chemically competent 10-beta cells (NEB) without *in vitro* ligation. PCR products that are described as being ligated by QuickChange mean that the primers were designed to encode ∼15 nucleotides of homology at the 5′ and 3′ termini, enabling ligation *in vivo* by 10-beta cells' endogenous recombinases (see Table S1 for details).


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

Next, 1 μL of purified DNA was combined with 25 μL of competent cells, incubated on ice for 25 min, subject to a heat pulse at 42 °C in a water bath for 40 s, and then returned to ice for 2 min. Next, cells were recovered by inoculation into 1 mL of SOC media and incubated at 37 °C for 1 h with agitation (700 rpm in a thermomixer), and then, 100 μL of the transformant was spread out on selective plates consisting of LB agar supplemented with 15 μg/mL tetracycline using coli rollers. After ∼16 h of incubation at 37 °C, colonies were selected and inoculated into 5 mL of LB supplemented with 15 μg/mL tetracycline and incubated overnight (∼16 h) at 37 °C with agitation (220 rpm). Cells from the saturated overnight cultures were collected by centrifugation (3200*g* for 15 min at 4 °C), and plasmid DNA was isolated using the ZR plasmid miniprep kit (Zymo) according to the manufacturer's protocol. In this (and in all further molecular cloning manipulations), we would typically isolate plasmid DNA from 4−6 colonies and subject it to sequencing analysis by Sanger sequencing (GeneWiz, following GeneWiz's specifications). A single clone with the correct DNA sequence would then be used for further steps.

The positive control was created through a QuickChange approach to introduce the two td-exon context sequences (encoding the decapeptide DVFLGLPFNI) at the junction between residues 52 and 53. PCR was conducted on pBAD-sfGFP as a template and using the primers TD_Context_Insert-f and TD_Context_Insert-r (Table S1 and Figure S10). The PCR product was purified and transformed as described above. This construct creates a replica of the GFP molecule that would be created by the loopable translator.

A geneblock for the prototype loopable translator (td-FP-G-td) was ordered from IDT and cloned after the araBAD promoter in pBAD through Gibson assembly. A linearized pBAD vector was generated by PCR using pBAD-sfGFP as a template and oligos clone-pBADGFP-f and clone-pBADGFP-r as primers. The vector fragment was assessed by 0.8% agarose (0.5× TBE) gel electrophoresis for the correct molecular weight, DpnI (NEB) digested (10 U was added to the PCR reaction and incubated at 37 °C for 30 min and then 80 °C for 20 min to inactivate), and column purified (Zymo). The prototype geneblock was designed to have the permuted sfGFP (residues 53−241 and then 1−52) flanked by the permuted td intron of T4 bacteriophage (P6a−P9.2, P1−P6a). The td-FP-G-td geneblock has a His tag at the end of GFP residue 241, a stop codon at the end of the His tag, and a ribosome binding site (RBS) in front of GFP residue 1. The geneblock and the linearized vector were ligated using HiFi 2x× Gibson Master Mix (NEB) according to the manufacturer's protocol and transformed into chemically competent 10 beta cells as described above.

Because polyGFP is not fluorescent,[23,24] we modified this system in a few more ways. A TEV cleavage site was installed after the His-tag, and the stop codon was removed using PCR with primers TEV-f and TEV-r via QuikChange. We also installed the downstream box sequence[27,28] after the start codon using PCR with primers InsertDB-F and InsertDB-R via QuikChange (Table S1 and Figure S10).[23,24] Following this sequence of alterations, the resulting construct corresponds to the pBAD-tdTEVDB drawn in Figure 2C and interrogated in Figure 3. The full nucleotide sequence of this construct is given in Figure S11.

To create pBAD-tdTEVDB-STOP, we reintroduced the TAA stop codon between GFP and the TEV site by PCR with primers TEV_DB_STOP_F and TEV_DB_STOP_R via blunt-end ligation (Table S1 and Figure S10).

**Fluorescence Assays.** First, chemically competent BL21-(DE3) (NEB) cells were transformed with pRK793 (encoding TEV protease),[30] a gift from the laboratory of Doug Barrick (JHU Biophysics department). 1 μL of pRK793 plasmid DNA and 25 μL of competent BL21(DE3) cells were mixed together in a microfuge tube by aspiration, incubated on ice for 25 min, subject to heat pulse at 42 °C in a water bath for 40 s, and then returned to ice for 2 min. Next, cells were recovered by inoculation into 1 mL of SOC media and incubated at 37 °C for 1 h with agitation (700 rpm in a thermomixer), and then, 100 μL of the transformant was spread out on selective plates consisting of LB agar supplemented with 100 μg/mL ampicillin using coli rollers. After ∼16 h of incubation at 37 °C, one colony was inoculated into 5 mL of LB supplemented with 100 μg/mL ampicillin and incubated overnight (∼16 h) at 37 °C with agitation (220 rpm). The overnight culture was inoculated into 1 L of LB with 100 μg/mL ampicillin in a 2 L baffled sterile flask to a starting $OD_{600}$ of 0.02−0.04 to make homebrew BL21(DE3)+pRK793 competent cells. The BL21-(DE3)+pRK793 day culture was incubated at 37 °C with agitation (220 rpm) until the OD reached 0.4−0.6. When the desired OD was achieved, the culture was transferred into 500 mL centrifuge bottles and incubated on ice for 20 min (after this step, the culture was handled either on ice or in a cold room). During the incubation, three buffers (200 mL of 100 mM $MgCl_2$, 200 mL of 100 mM $CaCl_2$, and 60 mL of 85 mM $CaCl_2$, 15% glycerol) were prepared and prechilled on ice. The ice-incubated culture was centrifuged at 3000*g* for 15 min at 4 °C. The supernatant was decanted off, and cell pellets in each 500 mL bottle were resuspended in 100 mL of cold 100 mM $MgCl_2$. The resuspension was centrifuged at 2000*g* for 15 min at 4 °C. The supernatant was decanted off, and cell pellets in each 500 mL bottle were resuspended in 100 mL of cold 100 mM $CaCl_2$. The resuspension was incubated on ice for 20−40 min and centrifuged at 2000*g* for 15 min at 4 °C. The supernatant was decanted off, and cell pellets in each 500 mL bottle were resuspended in 25 mL of cold 85 mM $CaCl_2$ and 15% glycerol, transferred to a 50 mL falcon tube, and spun at 1500*g* for 15 min at 4 °C. The supernatant was decanted off, and cell pellets in the 50 mL tube were resuspended in 4 mL of cold 85 mM $CaCl_2$ and 15% glycerol. The 4 mL of competent BL21(DE3)+pRK793 cells was aliquoted into 40 microfuge tubes, containing 105 μL apiece, which were flash frozen in liquid nitrogen and stored at −80 °C until further use.

To begin a fluorescence assay, an aliquot of BL21-(DE3)+pRK793 was thawed on ice. 1 μL of pBAD plasmid DNA and 25 μL of competent BL21(DE3)+pRK793 cells were mixed together in a microfuge tube by aspiration, incubated on ice for 25 min, subject to a heat pulse at 42 °C in a water bath for 40 s, and then returned to ice for 2 min. Next, cells were recovered by inoculation into 1 mL of SOC media and incubated at 37 °C for 1 h with agitation (700 rpm in a thermomixer), and then, 40 μL of the transformant was spread out on selective plates consisting of LB agar supplemented with 100 μg/mL ampicillin and 15 μg/mL tetracycline using coli rollers. After ∼16 h of incubation at 37 °C, (typically) three separate colonies were inoculated into 3× 200 μL LB supplemented with 100 μg/mL ampicillin and 15 μg/mL tetracycline in a clear-bottom 96-well plate (Costar). After inoculation, the plate was sealed with parafilm and incubated overnight (∼16 h) at 37 °C with agitation (220 rpm), and


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

these created the biological triplicates used for fluorescence measurements. For repeat measurements, separate colonies would be selected from the same agar plate, which were stored at 4 °C and would be reused for at most 1 week.

The final $OD_{600}$ values were measured for the overnight cultures using a plate reader instrument (SpectraMax iD3 from Molecular Devices) and were used to subculture down to a starting $OD_{600}$ of 0.05 in 200 μL of induction media (LB supplemented with 15 μg/mL tetracycline, 100 μg/mL ampicillin, 0.1 mM IPTG, and 0.2% arabinose) in a clear-bottom 96-well plate. 96-well plates containing biological triplicates for each of the conditions considered would be loaded into an iD3 microplate reader preincubated at 37 °C. The plate reader recorded both growth ($OD_{600}$ by absorbance) as well as GFP fluorescence (excitation at 488 nm, emission at 535 nm, PMT gain at 500 V with an integration time of 20 ms from 5 mm from the plate) every 10 min over 16 h. In between measurements, the plate reader agitated the plate with orbital mixing (high; 577 rpm).

In later experiments, plate reader measurements would be conducted at 30 °C (instead of 37 °C). Additionally, induction media was additionally supplemented with various concentrations of guanosine (Sigma) or $MgCl_2$.

To analyze the data, the absorbance graphs of the cultures were first examined to see if any data points needed to be removed due to absence of growth. Based on early experiments, it was found that the fluorescence time point after 50 000 s of growth was representative of maximal expression levels before cell death, and hence, this time point was used throughout this study. Fluorescence values at 50 000 s were compiled in Graphpad Prism 9, which was used to generate the bar charts shown in Figures 2−6. Statistical analysis was conducted using Student's *t* test (assuming normally distributed populations with equal variances), as implemented in Prism 9.

**Protein Expression and Western Blot.** 1 μL of each pBAD plasmid DNA was mixed together with either 25 μL of competent BL21(DE3) cells or 25 μL of competent BL21-(DE3)+pRK793 cells, respectively, in a microfuge tube by aspiration, incubated on ice for 25 min, subject to a heat pulse at 42 °C in a water bath for 40 s, and then returned to ice for 2 min. Next, the cells were recovered by inoculation into 1 mL of SOC media and incubated at 37 °C for 1 h with agitation (700 rpm in a thermomixer), and 40 μL of the transformant was spread out on a corresponding selective plate using coli rollers; BL21(DE3) transformants were spread on a selective plate consisting of LB agar supplemented with 15 μg/mL tetracycline only, and BL21(DE3)+pRK793 transformants were spread on a selective plate consisting of LB agar supplemented with 100 μg/mL ampicillin and 15 μg/mL tetracycline. One colony from each plate was inoculated into 5 mL of LB supplemented with the corresponding antibiotics in a 14 mL sterile round-bottom tube (ThermoFisher) and grown overnight (∼16 h). These overnight cultures were subcultured down to a starting $OD_{600}$ of 0.05 in 50 mL of LB media supplemented with corresponding antibiotics, inducers (0.2% arabinose and 0.1 mM IPTG), and 20 mM $Mg^{2+}$ in sterile 250 mL Erlenmeyer flasks. The day cultures were grown for ∼4 h at 30 °C with agitation (220 rpm) to final ODs of 1.0 to 1.2, aliquoted into 1.5 mL microfuge tubes by 1 mL, spun down at 3000*g* for 15 min at 4 °C, and stored at −20 °C until future use.

For lysis, the cell pellets were thawed on ice for 15 min and resuspended in 1 mL of 1× PBS. 30 μL of each resuspension was transferred to a microfuge tube, mixed with 7.5 μL of 5× Tris-glycine-SDS loading buffer via vortexing, heated in a 90 °C water bath for 5 min, and incubated on ice for 2 min. 30 μL of each lysate was loaded onto precast Novex WedgeWell 8−16% Tris-glycine Mini protein gels (ThermoFisher Scientific) with 3 μL of prestained PAGE ruler as the ladder (ThermoFisher Scientific; 26619) and was separated by electrophoresis at 140 V for approximately 1 h, using 1× Tris-glycine-SDS electrophoresis running buffer (BioRad). The resulting gel was incubated for approximately 2−3 min in 0.8× Tris-glycine buffer (BioRad) and 20% (v/v) methanol. The gel was trimmed to remove the wells and foot, and electroblotting was performed using an iBlot 2 gel transfer device (ThermoFisher) and its matching transfer packet (Invitrogen; iBlot 2 Transfer Stacks, PVDF, regular size), according to the manufacturer's protocol (7 min; 20 V). After electroblotting, the PVDF membrane was incubated in 15 mL of 5% (w/v) nonfat milk (Nestle—Carnation, instant nonfat dry milk)−TBST solution for approximately 1 h with rocking to block the membrane. 5% nonfat milk−TBST solution was made by combining 1× TBS (Quality Biological; pH 7.4) with evaporated milk and Tween 20 to a final concentration of 0.1% (v/v). The blocked membrane was incubated ∼16 h in 8 mL of diluted primary anti-His antibody (mouse; Invitrogen) solution at 4 °C with rocking; the diluted primary antibody was made by diluting the antibody by 1:1000 in 5% (w/v) nonfat milk−TBST. The membrane was rinsed in 1× TBST 3 times for 10 min each and incubated in 8 mL of diluted secondary antimouse-HRP antibody (goat; Invitrogen) solution at room temperature for 40 min with rocking; the diluted secondary antibody was made by diluting the antibody 1:10 000 in 5% (w/v) nonfat milk−TBST. The incubated membrane was rinsed in 1× TBST 3 times for 10 min each, incubated for 1 min in 800 μL of chemiluminescence reagents (Signal West Femto Maximum Sensitive Substrate; ThermoFisher Scientific) that were mixed in a 1:1 ratio, and then, images were acquired using a ChemiDoc Touch imaging system (BioRad).

**Making Constructs for a Dual-Fluorescence Ratiometry Assay.** A pRK5 plasmid containing a gene that encodes mCherry was obtained as a gift from the laboratory of TaekJip Ha (JHU Biophysics department). Using this plasmid as a template, two new constructs were made by PCR via Gibson assembly: a new positive control with mCherry (pBAD-sfGFP52-DVFLGLPFNI-P-mCherry) and pBAD-tdTEVDB with mCherry (pBAD-td-FP-TEVDB-mCherry-TEV-G-td; denoted as pBAD-tdTEVDB-mCherry) (Figure 4C). For the new positive control, a linearized vector was made by PCR using the original positive control (pBAD-sf-GF_DVFLGLPFNI_P) as the template and primers, PosCTRL_BB_F and tRBSdX_BB_R (Figure S10 and Table S1). The vector was then assessed by 0.8% agarose (0.5× TBE) gel electrophoresis and purified as described above. The mCherry insert was made by PCR using primers, mCherry-tRBSdX-F and mCherry-posCTRL_R, assessed by 0.8% agarose (0.5× TBE) gel electrophoresis, and purified as described above. The linearized vector and the inset were ligated using HiFi 2× Gibson Master Mix (NEB) according to the manufacturer's protocol and transformed into chemically competent 10 beta cells as described above.

For mCherry-containing pBAD-tdTEVDB, pBAD-tdTEVDB was linearized by PCR using primers, TEV-DB-


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

BB-F and TEV-DB-BB-R (Figure S10 and Table S1), assessed by 0.8% agarose (0.5× TBE) gel electrophoresis, and purified as described above. The mCherry insert was made by PCR using primers, mCherry-TEV-DB-F and mCherry-R (Figure S10 and Table S1), assessed by 0.8% agarose (0.5× TBE) gel electrophoresis, and purified as described above. The linearized vector and the inset were ligated using HiFi 2× Gibson Master Mix (NEB) according to the manufacturer's protocol and transformed into chemically competent 10 beta cells as described above. The two new constructs were sequence verified, before being transformed into BL21(DE3)+pRK793 cells and assessed using the fluorescence assay method described below in the Dual-Fluorescence Ratiometry Assay section.

**Dual-Fluorescence Ratiometry Assay.** 1 μL of a set of pBAD plasmids (negative control, pBAD-sfGF52-DVFLGLPFNI-P-mCherry, pBAD-tdTEVDB, and pBAD-tdTEVDB-mCherry) was transformed into BL21+pRK793 as described above. 40 μL of the transformant was spread out on selective plates consisting of LB agar supplemented with 100 μg/mL ampicillin and 15 μg/mL tetracycline using coli rollers. After ∼16 h of incubation at 37 °C, three separate colonies were inoculated into 3× 200 μL of LB supplemented with 100 μg/mL ampicillin and 15 μg/mL tetracycline in a clear-bottom 96-well plate (Costar). After inoculation, the plate was sealed with parafilm and incubated overnight (∼16 h) at 37 °C with agitation (220 rpm), and these created the biological triplicates used for fluorescence measurements. The final $OD_{600}$ values were measured for the overnight cultures using a plate reader instrument (SpectraMax iD3 from Molecular Devices) and were used to subculture down to a starting $OD_{600}$ of 0.05 in 200 μL of induction media (LB supplemented with 15 μg/mL tetracycline, 100 μg/mL ampicillin, 0.1 mM IPTG, 0.2% arabinose, and 20 mM $Mg^{2+}$) in a clear-bottom 96-well plate. 96-well plates containing biological triplicates were loaded into an iD3 microplate reader preincubated at 30 °C and measured for both growth ($OD_{600}$ by absorbance) and GFP and mCherry fluorescence (GFP, excitation at 488 nm, emission at 535 nm; mCherry, excitation at 588 nm, emission at 650 nm) every 10 min over 16 h. The rest of the setting, including PMT and the speed of orbital mixing, stayed the same.

To obtain the ratio of the native fluorescence of GFP and mCherry, the corresponding backgrounds (GFP and mCherry signals from the negative control) were subtracted from the fluorescence signals (GFP and mCherry) generated from the positive control (pBAD-sfGF52-DVFLGLPFNI-P-mCherry). For the positive control, GFP generated signals that were 54% of the level of mCherry (Figure 4D). This served as a normalization factor. For pBAD-tdTEVDB-mCherry, the corresponding backgrounds were subtracted from each of the fluorescence signals. The background-subtracted GFP signals were divided by the background-subtracted mCherry signals and then divided by the normalization factor (0.54) to generate the normalized GFP/mCherry ratio, indicating circularization efficiency.

**Library Generation.** To improve ribosomal initiation on the circular mRNA, error-prone PCR (epPCR) was performed on the initiation region of pBAD-tdTEVDB plasmid to install 1 or 2 point mutations on the target region. The starting point for these mutations was the pBAD-tdTEVDB variant with the trimmed-down context regions (DVFLG deleted from the 5′-exon context and PFNI deleted from the 3′-exon context, see Figure 5). The commercial GeneMorph II Random Mutagenesis kit (Agilent) was used to perform epPCR, but using a modified protocol that involved performing several iterations of dilution and reamplification with a touchdown PCR protocol (see ref 51 for details). The forward and reverse primers SD_lib_insert_F/R (Table S1 and Figure S10) were designed to span the initiation sequence. In the first round of epPCR, the manufacturer's protocol was adopted except that template plasmid was first diluted $10^9$-fold, and only 1 attagram (ag; 1 ag = $10^{-18}$) of plasmid DNA was used as a template in the PCR. The first epPCR product (primary amplicon) was size-verified by 1.2% agarose gel and diluted 1000-fold using Millipore water, and 0.5 μL of the diluent (ca. 2 pg) was used to seed a reamplification epPCR in a 25 μL scale that otherwise had the same primer, polymerase, and dNTP concentrations. Unlike the first epPCR, the reamplification PCR used a touchdown protocol in which the annealing temperature started at 65 °C in the first cycle and decreased by 0.5 °C in each cycle. Under these optimal reaction conditions found, the reamplification was repeated a total of nine times, in each case diluting the product by 1000-fold and using 0.5 μL of the diluent to seed the next PCR. After the final reamplification, the PCR products were subjected to a digest with DpnI and column-purified. Separately, backbone fragments were amplified using the primers SD_lib_BB_F/R (Table S1 and Figure S10) with a high-fidelity DNA polymerase (Q5 polymerase, NEB) that were designed to have homology arms with the flanks of the mutagenized insert. The insert was then cloned into the backbone via in-fusion cloning (TaKaRa Bio Inc.), and the ligated products were transformed into chemically competent 10 beta cells as described previously. The transformant was directly inoculated into liquid culture, from which plasmid DNA was purified via midi-prep to generate a plasmid library.

**Directed Evolution.** 1 μL of the plasmid library was transformed into 25 μL of BL21(DE3)+pRK793 competent cells as previously described, and 40 μL of the transformant was spread over selective plates of LB agar supplemented with 15 μg/mL tetracycline and 100 μg/mL ampicillin using coli rollers: 4 separate plates were prepared for the first round of directed evolution, and 8 plates were prepared for the second round of directed evolution. As a reference sample, pBAD-tdTEVDB was also transformed into 25 μL of BL21-(DE3)+pRK793 competent cells as previously described. After ∼16 h of incubation at 37 °C, three colonies of pBAD-tdTEVDB were manually picked from the plate using sterile pipet tips and were inoculated into the first three wells (A1−A3) of each of the four Nunc 96-Well Polypropylene DeepWell plates (2 mL well capacity; Fisher Scientific); each well contained 1.2 mL of LB media supplemented with 15 μg/mL tetracycline and 100 μg/mL ampicillin. The colonies on the plasmid library plates were automatically picked by sterilized needles of the RapidPick Colony Picking system (Wagner Life Science) through its corresponding software (Picker 5.0.1); 372 colonies were picked for the first round of directed evolution, and 720 colonies were picked for the second round. The picked colonies were inoculated into the identical selective media in each of the remaining wells of the 96 DeepWell plates. After inoculation, the 96 DeepWell plates were covered with breathable sheets [Breathe-EASIER 6" × 3.25" (cat. no. BERM-2000); Diversified BioTech] and incubated at 37 °C overnight with agitation (220 rpm). After ∼16 h of incubation, 5 μL of each overnight culture was inoculated into 200 μL of LB supplemented 15 μg/mL


https://doi.org/10.1021/acscentsci.1c00574
ACS Cent. Sci. 2021, 7, 1736−1750

tetracycline, 100 μg/mL ampicillin, 0.1 mM IPTG, 0.2% arabinose, and 20 mM $MgCl_2$ on black flat-bottom 96-well plates [4 plates; Caplugs Evergreen Labware Products (cat. no. 290-8195-Z1F)] using a 20 μL multichannel pipet (Eppendorf). The inoculated cultures in the black 96-well plates were again covered with breathable sheets and incubated at 30 °C with agitation (300 rpm) for 13 h (the 13 h time point was selected for measuring fluorescent signals based on the signal curve patterns of constructs observed up to that point). At the 13 h time point, all plates were taken out from the shaker and directly screened for fluorescent signals (excitation at 488 nm; emission at 535 nm; without lid) by a Spark multimode plate reader (Tecan). Top-performing constructs were chosen and directly inoculated into LB supplemented with 15 μg/mL tetracycline and 100 μg/mL ampicillin, from which constructs were purified via mini-prep and sent off for Sanger sequencing. Once sequence-verified, 1 μL of each of the purified constructs was rephenotyped by transformation into 25 μL BL21-(DE3)+pRK793 competent cells and used to set up a fluorescence assay in biological triplicates.

**Safety Statement.** No unexpected or unusually high safety hazards were encountered.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.1c00574.

> Additional methods, data, and figures including fluorescence levels, performances, $OD_{600}$ values, max growth rate, structures, illustrations, Northern blot image, densitometry results, flow chart, and sequences (PDF)

### Accession Codes

Plasmid Availability: pBAD-tdTEVDB_C-15A/G4C (the reporter plasmid following two rounds of directed evolution) is available on AddGene.

## ■ AUTHOR INFORMATION


### Corresponding Author

**Stephen D. Fried** − *Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, United States;* orcid.org/0000-0003-2494-2193; Email: sdfried@jhu.edu

### Authors

**Sea On Lee** − *Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, United States;* orcid.org/0000-0001-6462-8751

**Qi Xie** − *Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, United States;* orcid.org/0000-0002-3591-8749

Complete contact information is available at: https://pubs.acs.org/10.1021/acscentsci.1c00574


### Author Contributions

Conceptualization was performed by S.D.F. Methodology was provided by S.O.L. and S.D.F. Investigation was performed by S.O.L., Q.X., and S.D.F. Visualization was performed by S.O.L., Q.X., and S.D.F. S.D.F. supervised the work. S.O.L. and S.D.F wrote the manuscript.

### Notes

The authors declare the following competing financial interest(s): The loopable translator is currently under review by JHTV for a pending patent.

## ■ ACKNOWLEDGMENTS

We acknowledge HFSP (RGY0074/2019) for funding, as well as start-up funding from Johns Hopkins.

## ■ REFERENCES


(1) Xia, X. X.; Qian, Z. G.; Ki, C. S.; Park, Y. H.; Kaplan, D. L.; Lee, S. Y. Native-Sized Recombinant Spider Silk Protein Produced in Metabolically Engineered Escherichia Coli Results in a Strong Fiber. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (32), 14059−14063.

(2) Bowen, C. H.; Dai, B.; Sargent, C. J.; Bai, W.; Ladiwala, P.; Feng, H.; Huang, W.; Kaplan, D. L.; Galazka, J. M.; Zhang, F. Recombinant Spidroins Fully Replicate Primary Mechanical Properties of Natural Spider Silk. *Biomacromolecules* **2018**, *19* (9), 3853−3860.

(3) Rising, A.; Johansson, J. Toward Spinning Artificial Spider Silk. *Nat. Chem. Biol.* **2015**, *11* (MAY), 309−315.

(4) Omenetto, F. G.; Kaplan, D. L. New Opportunities for an Ancient Material. *Science* **2010**, *329* (5991), 528−531.

(5) Andersson, M.; Jia, Q.; Abella, A.; Lee, X. Y.; Landreh, M.; Purhonen, P.; Hebert, H.; Tenje, M.; Robinson, C. V.; Meng, Q.; Plaza, G. R.; Johansson, J.; Rising, A. Biomimetic Spinning of Artificial Spider Silk from a Chimeric Minispidroin. *Nat. Chem. Biol.* **2017**, *13* (3), 262−264.

(6) Prince, J. T.; McGrath, K. P.; DiGirolamo, C. M.; Kaplan, D. L. Construction, Cloning, and Expression of Synthetic Genes Encoding Spider Dragline Silk. *Biochemistry* **1995**, *34* (34), 10879−10885.

(7) Vacanti, J. P.; Langer, R. Tissue Engineering: The Design and Fabrication of Living Replacement Devices for Surgical Reconstruction and Transplantation. *Lancet* **1999**, *354*, S32−S34.

(8) Paramonov, S. E.; Gauba, V.; Hartgerink, J. D. Synthesis of Collagen-like Peptide Polymers by Native Chemical Ligation. *Macromolecules* **2005**, *38* (18), 7555−7561.

(9) Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78*, 929−958.

(10) Barnhart, M. M.; Chapman, M. R. Curli Biogenesis and Function. *Annu. Rev. Microbiol.* **2006**, *60*, 131−147.

(11) Abdali, Z.; Aminzare, M.; Zhu, X.; Debenedictis, E.; Xie, O.; Keten, S.; Dorval Courchesne, N. M. Curli-Mediated Self-Assembly of a Fibrous Protein Scaffold for Hydroxyapatite Mineralization. *ACS Synth. Biol.* **2020**, *9* (12), 3334−3343.

(12) Diehl, A.; Roske, Y.; Ball, L.; Chowdhury, A.; Hiller, M.; Molière, N.; Kramer, R.; Stöppler, D.; Worth, C. L.; Schlegel, B.; Leidert, M.; Cremer, N.; Erdmann, N.; Lopez, D.; Stephanowitz, H.; Krause, E.; van Rossum, B. J.; Schmieder, P.; Heinemann, U.; Turgay, K.; Akbey, Ü.; Oschkinat, H. Structural Changes of TasA in Biofilm Formation of Bacillus Subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (13), 3237−3242.

(13) Knowles, T. P. J.; Buehler, M. J. Nanomechanics of Functional and Pathological Amyloid Materials. *Nat. Nanotechnol.* **2011**, *6* (8), 469−479.

(14) Vural, M.; Lei, Y.; Pena-Francesch, A.; Jung, H.; Allen, B.; Terrones, M.; Demirel, M. C. Programmable Molecular Composites of Tandem Proteins with Graphene Oxide for Efficient Bimorph Actuators. *Carbon* **2017**, *118*, 404−412.

(15) Pena-Francesch, A.; Akgun, B.; Miserez, A.; Zhu, W.; Gao, H.; Demirel, M. C. Pressure Sensitive Adhesion of an Elastomeric Protein Complex Extracted from Squid Ring Teeth. *Adv. Funct. Mater.* **2014**, *24* (39), 6227−6233.

(16) Pena-Francesch, A.; Jung, H.; Segad, M.; Colby, R. H.; Allen, B. D.; Demirel, M. C. Mechanical Properties of Tandem-Repeat Proteins Are Governed by Network Defects. *ACS Biomater. Sci. Eng.* **2018**, *4* (3), 884−891.

(17) Bale, J. B.; Gonen, S.; Liu, Y.; Shedffler, W.; King, N. P.; Baker, D. Accurate Design of Megadalton-Scale Two-Component Icosahe-

dral Protein Complexes. *Science (Washington, DC, U. S.)* **2016**, *353* (6297), 389−394.
(18) King, N. P.; Jacobitz, A. W.; Sawaya, M. R.; Goldschmidt, L.; Yeates, T. O. Structure and Folding of a Designed Knotted Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (48), 20732−20737.
(19) Hershewe, J. M.; Wiseman, W. D.; Kath, J. E.; Buck, C. C.; Gupta, M. K.; Dennis, P. B.; Naik, R. R.; Jewett, M. C. Characterizing and Controlling Nanoscale Self-Assembly of Suckerin-12. *ACS Synth. Biol.* **2020**, *9* (12), 3388−3399.
(20) Votteler, J.; Ogohara, C.; Yi, S.; Hsia, Y.; Nattermann, U.; Belnap, D. M.; King, N. P.; Sundquist, W. I. Designed Proteins Induce the Formation of Nanocage-Containing Extracellular Vesicles. *Nature* **2016**, *540* (7632), 292−295.
(21) Jung, H.; Pena-Francesch, A.; Saadat, A.; Sebastian, A.; Kim, D. H.; Hamilton, R. F.; Albert, I.; Allen, B. D.; Demirel, M. C. Molecular Tandem Repeat Strategy for Elucidating Mechanical Properties of High-Strength Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (23), 6478−6483.
(22) Ford, E.; Ares, M., Jr Synthesis of Circular RNA in Bacteria and Yeast Using RNA Cyclase Ribozymes Derived from a Group I Intron of Phage T4. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (8), 3117−3121.
(23) Perriman, R.; Ares, M. Circular MRNA Can Direct Translation of Extremely Long Repeating- Sequence Proteins in Vivo. *RNA* **1998**, *4* (9), 1047−1054.
(24) Perriman, R. Circular MRNA Encoding for Monomeric and Polymeric Green Fluorescent Protein. *Methods Mol. Biol.* **2002**, *183*, 69−85.
(25) Shub, D. A.; Gott, J. M.; Xu, M. Q.; Lang, B. F.; Michel, F.; Tomaschewski, J.; Pedersen-Lane, J.; Belfort, M. Structural Conservation among Three Homologous Introns of Bacteriophage T4 and the Group I Introns of Eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (4), 1151−1155.
(26) Umekage, S.; Uehara, T.; Fujita, Y.; Suzuki, H.; Kikuchi, Y. In Vivo Circular RNA Expression by the Permuted Intron-Exon Method. *Innov. Biotechnol.* **2012**, *108* (4), 354−356.
(27) Sprengart, M. L.; Fuchs, E.; Porter, A. G. The Downstream Box: An Efficient and Independent Translation Initiation Signal in Escherichia Coli. *EMBO J.* **1996**, *15* (3), 665−674.
(28) Etchegaray, J.-P.; Inouye, M. Translational Enhancement by an Element Downstream of the Initiation Codon in Escherichia Coli*. *J. Biol. Chem.* **1999**, *274*, 10079.
(29) Van Den Berg, S.; Löfdahl, P. Å.; Härd, T.; Berglund, H. Improved Solubility of TEV Protease by Directed Evolution. *J. Biotechnol.* **2006**, *121* (3), 291−298.
(30) Kapust, R. B.; Tözsér, J.; Fox, J. D.; Anderson, D. E.; Cherry, S.; Copeland, T. D.; Waugh, D. S. Tobacco Etch Virus Protease: Mechanism of Autolysis and Rational Design of Stable Mutants with Wild-Type Catalytic Proficiency. *Protein Eng., Des. Sel.* **2001**, *14* (12), 993−1000.
(31) Cech, T. R. Self-Splicing of Group I Introns. *Annu. Rev. Biochem.* **1990**, *59*, 543−568.
(32) Guo, F.; Gooding, A. R.; Cech, T. R. Structure of the Tetrahymena Ribozyme. *Mol. Cell* **2004**, *16* (3), 351−362.
(33) Michel, F.; Hanna, M.; Green, R.; Bartel, D. P.; Szostak, J. W. The Guanosine Binding Site of the Tetrahymena Ribozyme. *Nature* **1989**, *342* (6248), 391−395.
(34) Legault, P.; Herschlag, D.; Celander, D. W.; Cech, T. R. Mutations at the Guanosine-Binding Site of the Tetrahymena Ribozyme Also Affect Site-Specific Hydrolysis. *Nucleic Acids Res.* **1992**, *20* (24), 6613−6619.
(35) Varik, V.; Oliveira, S. R. A.; Hauryliuk, V.; Tenson, T. HPLC-Based Quantification of Bacterial Housekeeping Nucleotides and Alarmone Messengers PpGpp and PppGpp. *Sci. Rep.* **2017**, *7* (1), 1−12.
(36) Chen, X.; Mohr, G.; Lambowitz, A. M. The Neurosporea Crassa CYT-18 Protein C-Terminal RNA-Binding Domain Helps Stabilize Interdomain Tertiary Interactions in Group I Introns. *RNA* **2004**, *10* (4), 634−644.
(37) Wei, L.; Cai, X.; Qi, Z.; Rong, L.; Cheng, B.; Fan, J. In Vivo and in Vitro Characterization of TEV Protease Mutants. *Protein Expression Purif.* **2012**, *83* (2), 157−163.
(38) Fang, J.; Chen, L.; Cheng, B.; Fan, J. Engineering Soluble Tobacco Etch Virus Protease Accompanies the Loss of Stability. *Protein Expression Purif.* **2013**, *92* (1), 29−35.
(39) Sanchez, M. I.; Ting, A. Y. Directed Evolution Improves the Catalytic Efficiency of TEV Protease. *Nat. Methods* **2020**, *17* (2), 167−174.
(40) Colussi, T. M.; Costantino, D. A.; Zhu, J.; Donohue, J. P.; Korostelev, A. A.; Jaafar, Z. A.; Plank, T. D. M.; Noller, H. F.; Kieft, J. S. Initiation of Translation in Bacteria by a Structured Eukaryotic IRES RNA. *Nature* **2015**, *519* (7541), 110−113.
(41) Litke, J. L.; Jaffrey, S. R. Highly Efficient Expression of Circular RNA Aptamers in Cells Using Autocatalytic Transcripts. *Nat. Biotechnol.* **2019**, *37* (6), 667−675.
(42) Levin-Karp, A.; Barenholz, U.; Bareia, T.; Dayagi, M.; Zelcbuch, L.; Antonovsky, N.; Noor, E.; Milo, R. Quantifying Translational Coupling in E. Coli Synthetic Operons Using RBS Modulation and Fluorescent Reporters. *ACS Synth. Biol.* **2013**, *2* (6), 327−336.
(43) Zhang, Y.; Chen, H.; Zhang, Y.; Yin, H.; Zhou, C.; Wang, Y. Direct RBS Engineering of the Biosynthetic Gene Cluster for Efficient Productivity of Violaceins in E. Coli. *Microb. Cell Fact.* **2021**, *20* (1), 1−13.
(44) Zhang, X.; Lin, Y.; Wu, Q.; Wang, Y.; Chen, G. Q. Synthetic Biology and Genome-Editing Tools for Improving PHA Metabolic Engineering. *Trends Biotechnol.* **2020**, *38* (7), 689−700.
(45) Sørensen, M. A.; Fricke, J.; Pedersen, S. Ribosomal Protein S1 Is Required for Translation of Most, It Not All, Natural MRNAs in Escherichia Coli in Vivo. *J. Mol. Biol.* **1998**, *280* (4), 561−569.
(46) Boni, I. V.; Lsaeva, D. M.; Musychenko, M. L.; Tzareva, N. V. Ribosome-Messenger Recognition: MRNA Target Sites for Ribosomal Protein S1. *Nucleic Acids Res.* **1991**, *19* (1), 155−162.
(47) Rasila, T. S.; Pajunen, M. I.; Savilahti, H. Critical Evaluation of Random Mutagenesis by Error-Prone Polymerase Chain Reaction Protocols, Escherichia Coli Mutator Strain, and Hydroxylamine Treatment. *Anal. Biochem.* **2009**, *388* (1), 71−80.
(48) McInerney, P.; Adams, P.; Hadi, M. Z. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Mol. Biol. Int.* **2014**, *2014*, 1−8.
(49) Patrick, W. M.; Firth, A. E.; Blackburn, J. M. User-Friendly Algorithms for Estimating Completeness and Diversity in Randomized Protein-Encoding Libraries. *Protein Eng., Des. Sel.* **2003**, *16* (6), 451−457.
(50) Vanhercke, T.; Ampe, C.; Tirry, L.; Denolf, P. Reducing Mutational Bias in Random Protein Libraries. *Anal. Biochem.* **2005**, *339* (1), 9−14.
(51) Lee, S. O.; Fried, S. D. An Error-Prone PCR Method for Small Amplicons. *Anal. Biochem.* **2021**, *628*, 114266.
(52) Farabaugh, P. J.; Björk, G. R. How Translational Accuracy Influences Reading Frame Maintenance. *EMBO J.* **1999**, *18* (6), 1427−1434.
(53) Sarr, M.; Kronqvist, N.; Chen, G.; Aleksis, R.; Purhonen, P.; Hebert, H.; Jaudzems, K.; Rising, A.; Johansson, J. A Spidroin-Derived Solubility Tag Enables Controlled Aggregation of a Designed Amyloid Protein. *FEBS J.* **2018**, *285* (10), 1873−1885.
(54) Vincenz-Donnelly, L.; Holthusen, H.; Körner, R.; Hansen, E. C.; Presto, J.; Johansson, J.; Sawarkar, R.; Hartl, F. U.; Hipp, M. S. High Capacity of the Endoplasmic Reticulum to Prevent Secretion and Aggregation of Amyloidogenic Proteins. *EMBO J.* **2018**, *37* (3), 337−350.
(55) Abelein, A.; Chen, G.; Kitoka, K.; Aleksis, R.; Oleskovs, F.; Sarr, M.; Landreh, M.; Pahnke, J.; Nordling, K.; Kronqvist, N.; Jaudzems, K.; Rising, A.; Johansson, J.; Biverstål, H. High-Yield Production of Amyloid-$\beta$ Peptide Enabled by a Customized Spider Silk Domain. *Sci. Rep.* **2020**, *10* (1), 1−10.
(56) Landreh, M.; Andersson, M.; Marklund, E. G.; Jia, Q.; Meng, Q.; Johansson, J.; Robinson, C. V.; Rising, A. Mass Spectrometry

Captures Structural Intermediates in Protein Fiber Self-Assembly. *Chem. Commun.* **2017**, *53* (23), 3319−3322.
(57) Jin, H. J.; Kaplan, D. L. Mechanism of Silk Processing in Insects and Spiders. *Nature* **2003**, *424* (6952), 1057−1061.
(58) Schwarze, S.; Zwettler, F. U.; Johnson, C. M.; Neuweiler, H. The N-Terminal Domains of Spider Silk Proteins Assemble Ultrafast and Protected from Charge Screening. *Nat. Commun.* **2013**, *4*, 4.
(59) Ittah, S.; Cohen, S.; Garty, S.; Cohn, D.; Gat, U. An Essential Role for the C-Terminal Domain of a Dragline Spider Silk Protein in Directing Fiber Formation. *Biomacromolecules* **2006**, *7* (6), 1790−1795.
(60) Nguyen, P. Q.; Courchesne, N.-M. D.; Duraj-Thatte, A.; Praveschotinunt, P.; Joshi, N. S. Engineering Living Materials: Prospects and Challenges for Using Biological Systems to Direct the Assembly of Smart Materials. *Adv. Mater.* **2018**, *30* (12), 139−148.
(61) Chen, A. Y.; Deng, Z.; Billings, A. N.; Seker, U. O. S.; Lu, M. Y.; Citorik, R. J.; Zakeri, B.; Lu, T. K. Synthesis and Patterning of Tunable Multiscale Materials with Engineered Cells. *Nat. Mater.* **2014**, *13* (5), 515−523.
(62) Nguyen, P. Q.; Botyanszki, Z.; Tay, P. K. R.; Joshi, N. S. Programmable Biofilm-Based Materials from Engineered Curli Nanofibres. *Nat. Commun.* **2014**, *5*, 1−10.
(63) Duraj-Thatte, A. M.; Manjula-Basavanna, A.; Courchesne, N. M. D.; Cannici, G. I.; Sánchez-Ferrer, A.; Frank, B. P.; van't Hag, L.; Cotts, S. K.; Fairbrother, D. H.; Mezzenga, R.; Joshi, N. S. Water-Processable, Biodegradable and Coatable Aquaplastic from Engineered Biofilms. *Nat. Chem. Biol.* **2021**, *17*, 732.
(64) Dubey, G.; Mequanint, K. Conjugation of Fibronectin onto Three-Dimensional Porous Scaffolds for Vascular Tissue Engineering Applications. *Acta Biomater.* **2011**, *7* (3), 1114−1125.
(65) Glowacki, J.; Mizuno, S. Collagen Scaffolds for Tissue Engineering. *Biopolymers* **2008**, *89* (5), 338−344.
(66) Boccaccini, A. R.; Blaker, J. J. Bioactive Composite Materials for Tissue Engineering Scaffolds. *Expert Rev. Med. Devices* **2005**, *2* (3), 303−317.


https://doi.org/10.1021/acscentsci.1c00574
*ACS Cent. Sci.* 2021, 7, 1736−1750