

# Novel insights into construct toxicity, strain optimization, and primary sequence design for producing recombinant silk fibroin and elastin-like peptide in *E. coli*

Alexander Connor<sup>a,b</sup>, Caleb Wigham<sup>a,b</sup>, Yang Bai<sup>b</sup>, Manish Rai<sup>a,b</sup>, Sebastian Nassif<sup>a</sup>,  
Mattheos Koffas<sup>a,b,\*\*</sup>, R. Helen Zha<sup>a,b,\*</sup>

<sup>a</sup> Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA

<sup>b</sup> Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA

## ARTICLE INFO

### Keywords:

Recombinant silk  
Strain engineering  
Toxicity  
*E. coli*  
Spider dragline silk

## ABSTRACT

Spider silk proteins (spidroins) are a remarkable class of biomaterials that exhibit a unique combination of high-value attributes and can be processed into numerous morphologies for targeted applications in diverse fields. Recombinant production of spidroins represents the most promising route towards establishing the industrial production of the material, however, recombinant spider silk production suffers from fundamental difficulties that includes low titers, plasmid instability, and translational inefficiencies. In this work, we sought to gain a deeper understanding of upstream bottlenecks that exist in the field through the production of a panel of systematically varied spidroin sequences in multiple *E. coli* strains. A restriction on basal expression and specific genetic mutations related to stress responses were identified as primary factors that facilitated higher titers of the recombinant silk constructs. Using these findings, a novel strain of *E. coli* was created that produces recombinant silk constructs at levels 4–33 times higher than standard BL21(DE3). However, these findings did not extend to a similar recombinant protein, an elastin-like peptide. It was found that the recombinant silk proteins, but not the elastin-like peptide, exert toxicity on the *E. coli* host system, possibly through their high degree of intrinsic disorder. Along with strain engineering, a bioprocess design that utilizes longer culturing times and attenuated induction was found to raise recombinant silk titers by seven-fold and mitigate toxicity. Targeted alteration to the primary sequence of the recombinant silk constructs was also found to mitigate toxicity. These findings identify multiple points of focus for future work seeking to further optimize the recombinant production of silk proteins and is the first work to identify the intrinsic disorder and subsequent toxicity of certain spidroin constructs as a primary factor related to the difficulties of production.

## 1. Introduction

Spider silk proteins (spidroins) are of intense interest to engineers and researchers due to their high value material properties and utility in diverse applications. Orb-weaving spiders produce up to seven different types of silk, with dragline (major ampullate) silk serving as a safety line and framework of the web. Dragline silk fibers are five times stronger by weight than steel and three times tougher than the top-quality man-made fiber Kevlar (Hardy et al., 2008; Fink and Zha, 2018). Additionally, silk is biodegradable and biocompatible, and silk proteins can be processed into numerous morphologies including coatings, hydrogels,

and tissue scaffolds (Hardy et al., 2008; Fink and Zha, 2018; Gosline et al., 1999). As such, the applications of silk proteins range from next-generation body armor to optofluidic devices and even coatings for food preservation (Tsioris et al., 2010; Gould, 2002; Marelli et al., 2016). While silk protein from the *Bombyx mori* silkworm is farmed at scale for the textiles industry, dragline silk cannot be readily obtained through farming, as spiders are territorial and cannibalistic (Tokareva et al., 2013). Thus, researchers have used recombinant production to obtain proteins that mimic or directly copy the sequences of natural dragline spidroins. Recombinant production currently represents the most promising method for producing dragline spidroins at scale while also

\* Corresponding author. Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA.

\*\* Corresponding author. Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA.

E-mail addresses: [koffam@rpi.edu](mailto:koffam@rpi.edu) (M. Koffas), [zhar@rpi.edu](mailto:zhar@rpi.edu) (R.H. Zha).

<https://doi.org/10.1016/j.mec.2023.e00219>

Received 13 October 2022; Received in revised form 6 December 2022; Accepted 24 January 2023

Available online 3 February 2023

2214-0301/© 2023 The Authors. Published by Elsevier B.V. on behalf of International Metabolic Engineering Society. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

presenting the ability to rationally design protein sequences with targeted properties<sup>8-10</sup>.

The unique properties of dragline spidroins arise from specific peptide motifs, chemical interactions, and hierarchical organization that are highly conserved among orb-weaving spiders. Natural dragline fibers are composed of two proteins, Major Ampullate Spidroin 1 (MaSp1) and Major Ampullate Spidroin 2 (MaSp2) in a ratio of approximately three MaSp1 for every two MaSp2 (Heim et al., 2009; Sarkar et al., 2019). Major ampullate spidroins are generally quite large at 250–350 kDa and take the form of a segmented copolymer with small non-repetitive N and C-terminal domains that flank a large repetitive core domain. The repetitive domain of dragline spidroins represents approximately 90% of the total protein, with repeating units that are typically 33–45 amino acids long (Heim et al., 2009; Sarkar et al., 2019; Gatesy et al., 2001). The repeat unit of MaSp1 is characterized by a tandem alanine repeat ( $A_n$ , where  $n \sim 6-9$ ) adjacent to a glycine-rich region that contains GGX motifs, where X often represents tyrosine (Y), glutamine (Q), or leucine (L). The repeat unit of MaSp2 also contains tandem alanine repeats, but its glycine-rich region is high in proline (P) and contains GPGXX and GGX motifs, where X often represents Q, Y, L, G, or serine (S) (Teulé et al., 2009; Guerette et al., 1996; Malay et al., 2017). In both MaSp1 and MaSp2, the tandem alanine segments assemble into beta-sheet nanocrystals, and the glycine-rich regions form an amorphous matrix during fiber spinning. The interplay between these crystalline and amorphous domains endows spider silk with many of its unique properties, including a combination of high tensile strength and toughness (Sarkar et al., 2019; Tokareva et al., 2014).

Recombinant silk has been produced in a diverse set of host organisms, including bacteria, yeast, mammalian cells, insect cells, transgenic plants, and transgenic animals (Wohlrab et al., 2014; Chung et al., 2012). Common practice in the field is to create synthetic spider silk genes that combine spidroin amino acid motifs (GGX ( $A_n$ ), etc.), in ways that mimic the repetitive core of a natural dragline sequence. This is due to the difficulty in obtaining exact copies of full-length dragline spidroin genes by PCR, as they are long, repetitive, and have a high GC content (Wohlrab et al., 2014; Chung et al., 2012). Recombinantly produced dragline spidroins generally have anywhere from 2 to 196 repeats of a relatively short “monomer” segment (typically around 35 amino acids) and may or may not include non-repetitive terminal domains. Most efforts to produce recombinant spidroins have suffered from low titers, preventing the production and utilization of artificial spider silk at a commercial scale (Whittall et al., 2021; Edlund et al., 2018). Additionally, expressing recombinant spidroins in bacteria is often plagued by a high degree of plasmid instability, inclusion body formation, low solubility of the spidroin constructs, and transcriptional and translational errors (Yang et al., 2016; Chung et al., 2012). These issues, particularly the low titers, correlate with recombinant spidroin size, which further limits the production of useful materials, as increasing spidroin size has been shown to increase the mechanical properties of resultant fibers (Whittall et al., 2021; Wohlrab et al., 2014; Chung et al., 2012; Edlund et al., 2018; Bowen et al., 2018). However, two recent works have shown that high titers of recombinant spidroins are possible. Using an *E. coli* host system, Yang et al. achieved a titer of 3.6 g/L for 200 kDa dragline spidroin in a bioreactor kept at 16 °C. The researchers also employed a secondary plasmid to upregulate glycyl-tRNA supply (Yang et al., 2016). Furthermore, Schmuck et al. documented a titer of 14.5 g/L for a small recombinant spidroin using an *E. coli* host system in a bioreactor. This 33 kDa recombinant spidroin only contained two monomer repeats in its primary sequence, but it could be spun into fibers that exhibited mechanical properties similar to much larger recombinant spidroins (Schmuck et al., 2021).

Despite these promising efforts to increase titer, the relatively large body of work surrounding recombinant silk production has not been able to fully understand the challenges in expressing dragline spidroin. Yang et al. demonstrated that glycyl-tRNA upregulation can facilitate higher titers of large, glycine-rich dragline spidroins. However, their

work also showed that decreasing the culture temperature from 30 °C to 16 °C, independent of glycyl-tRNA upregulation, was sufficient to increase titers nearly an order of magnitude from 0.36 g/L to 2.9 g/L through an unknown mechanism (Yang et al., 2016). Moreover, Bhattacharyya et al. recently showed that while increasing tRNAGly and tRNAPro was helpful for producing certain spidroin sequences, it could be non-beneficial or even toxic to the cells when other spidroin sequences were expressed (Bhattacharyya et al., 2021). These findings indicate that mechanisms aside from translational difficulties need to be addressed. Work related to strain optimization of viable industrial hosts and hypotheses related to upstream production bottlenecks are sparse within the recombinant silk field. Likewise, little work outside of increasing spidroin length has been done to understand how alterations to the primary sequence affects upstream expression outcomes such as cell growth, plasmid maintenance, and titer. A thorough analysis of the *E. coli* platform using multiple strains and spidroin primary sequences can move the field forward, since many studies in this space have only used a “default” *E. coli* BL21 strain (Whittall et al., 2021; Schmuck et al., 2021; Yang et al., 2016; Teulé et al., 2009; Bowen et al., 2018; Bhattacharyya et al., 2021; Cai et al., 2020; Wei et al., 2020; Andersson et al., 2017; Xia et al., 2010; Fredriksson et al., 2009; Humenik et al., 2014). As a long-standing host choice for synthetic biology, the *E. coli* platform has become highly diversified with numerous distinct strains. Several *E. coli* strains present characteristics that can elucidate additional cellular mechanisms to target for increasing recombinant spidroin titers.

In this study, we sought to uncover a deeper understanding of how to construct microbial strains for optimal spidroin production. We also aimed to understand how the design of a recombinant spidroin construct affects production efficiency. To that end, we studied the expression of a panel of dragline spidroins that varied systematically in size and primary sequence in ten *E. coli* strains to characterize the relationships between upstream outcomes, host characteristics, and primary spidroin sequence. Our results suggest that the expression of MaSp2-mimetic recombinant spidroins exerts toxicity on *E. coli*, as shown through a negative effect on cell growth and plasmid maintenance, but that these effects could be attenuated with targeted primary sequence alteration. For further mechanistic insight, we extended our work towards expressing elastin-like peptide (ELP), a repetitive protein of interest for tissue engineering and drug delivery applications. Recombinant ELPs share several similarities with dragline spidroins, including a polymeric structure, a high glycine and proline content, and an ability to self-assemble when triggered by external stimuli (Schmuck et al., 2021; Sarkar et al., 2019; Varanko et al., 2020). The ELP, however, exerted no observable toxicity on host cells during expression. We hypothesized that this is due to a higher intrinsic disorder for our spidroin constructs compared to the ELP. Eliminating basal expression, or refraining from gene induction, was found to play a key role in obtaining high titers of toxic recombinant spidroin constructs but had no beneficial effect for ELP production. Additionally, we identified multiple genes related to stress response as new targets for future strain optimization that aims to increase titers of toxic recombinant proteins. Using these findings, we developed a novel strain of *E. coli* that combines both a restriction on basal expression and unique genetic mutations to facilitate high spidroin titer and plasmid maintenance without sacrificing the ability to use a strong promoter or exert control over gene induction. This novel strain offered no benefit for ELP production, indicating that future work should address production bottlenecks that may be unique to recombinant spidroin constructs.

## 2. Materials and methods

### 2.1. Bacterial strains and plasmids

The *E. coli* strains used for protein expressions are listed in [Supplementary Table 1](#). Chemically competent cells of NEB Stable Competent *E. coli* (New England Biolabs, Ipswich, MA) were used for gene cloning.

Restriction enzymes and T4 DNA ligase were purchased from Thermo Fisher Scientific (Waltham, MA) and DNA manipulations were carried out using the manufacturer's protocols. Extraction of plasmid DNA from *E. coli* and from agarose gel was performed using the E.Z.N.A. Plasmid DNA Mini Kit I and the E.Z.N.A. Gel Extraction kit, respectively (Omega Bio-tek, Norcross, GA). Plasmids were introduced into the expression strains via electroporation using the Gene Pulser Xcell Electroporation System at (Bio-Rad, Hercules, CA). Transformants were isolated on LB agar plates supplemented with either 100 µg/ml ampicillin or 100 µg/ml ampicillin and 25 µg/ml chloramphenicol for strains that carry an additional chloramphenicol selectable vector, namely pLysS, pGro7, RosettaGami, and SoluBL21-pLysS.

Genes for the monomer units of the A5 (GPGQAAAAAGPGQQGPGQQGPGEQGPGSG) and A10 (GPGQAAAAAAAAGPGQQGPGQQGPGEQGPGSG) silk constructs were codon optimized for *E. coli* and purchased as synthetic genes inserted into the XhoI and BamHI sites of pBluescript II SK(+) (Genscript, Piscataway, NJ). The compatible, but non regenerable, restriction sites XmaI and Kpn2I were included at the beginning and end of the genes, respectively. This allowed for the iterative duplication of the A5 or A10 genes using a previously described restriction cloning procedure (Teulé et al., 2009). The resultant genes were polymerized forms of their original monomer sequence (ex. A5 16mer) and coded for proteins that had tandem repeats of either the A5 or A10 sequence. For protein expressions, genes were cloned into the pET-19b expression vector (Addgene, Watertown, MA) using the XhoI and BamHI restriction sites. The A4Y1 ELP was codon optimized and purchased as a synthetic gene inserted into the NdeI and BamHI sites of pET-19b (Genscript, Piscataway, NJ). For synthesis of the A10 4mer BWT construct, cDNA of the N and C terminal regions of *L. hesperus* MaSp1 (major ampullate silk protein 1) were purchased as synthetic genes codon optimized for *E. coli* and inserted into the XhoI and BamHI restriction sites of pBluescript II SK(+) (Genscript, Piscataway, NJ). The *L. hesperus* N-terminus was cloned into the beginning of the A10 4mer gene using the NdeI (on the pET-19b backbone) and XmaI restrictions sites, followed by subsequent cloning of the *L. hesperus* C-terminus to the end of the A10 4mer gene using the Kpn2I and BamHI restriction sites. All expression plasmids were sequenced through Genewiz (South Plainfield, NJ).

## 2.2. Protein expressions

All protein expressions took place at 37 °C in LB media (Lennox broth, Sigma Aldrich, St. Louis, MO) supplemented with either 100 µg/ml ampicillin or 100 µg/ml ampicillin and 25 µg/ml chloramphenicol for strains that carry an additional chloramphenicol selectable vector, namely pLysS, pGro7, RosettaGami, and SoluBL21-pLysS. Strains were cultivated overnight at 37 °C in 5 ml of LB media placed in 15 ml conical tubes at an angle of ~55° with 225 rpm shaking. One milliliter of overnight cultures was used to inoculate 50 ml expression cultures placed in 250 ml Erlenmeyer flasks with 225 rpm shaking. Recombinant protein production was induced with 1 mM isopropyl-β-D-thiogalactoside (IPTG; Sigma Aldrich, St. Louis, MO) when the cultures reached an OD600 of 0.6–0.8, typically 2–3 h after inoculation. For experiments where titer and plasmid maintenance were quantified, expression cultures were harvested 4 h after induction. For experiments that generated 10.25-h growth curves, cells were put through the same procedure as above, but cultures were not harvested until 10.25 h after inoculation.

## 2.3. Lysis, purification, and titer quantification

After protein expressions, cell cultures were transferred to 50 ml conical tubes and centrifuged at 7000 rcf for 15 min at 4 °C. The supernatant was decanted, and cells were resuspended in 5 ml of lysis buffer (50 mM Tris-HCl (pH 8), 10 mM MgCl<sub>2</sub>, 10 mM NaCl, 1 mM PMSF) and frozen at –20 °C overnight. The frozen cell pellets were thawed at 37 °C and mixed with 2.5 mg of lysozyme (Sigma Aldrich, St.

Louis, MO). The cells were then placed on ice and sonicated with the Fisherbrand Model 120 Sonic Dismembrator (Fischer Scientific, Pittsburgh, Pennsylvania) using a 1/8-inch probe at 70% amplitude with a 1 s on and 1 s off cycle for 5 min.

For purification of the A5 and A10 constructs, the lysates were heated to 80 °C for 10 min in a water bath and then centrifuged for 10 min at 20,000 rcf. This procedure removed a majority of the native *E. coli* proteins from the lysate, resulting in a sample that was primarily comprised of the heat resistant A5 and A10 constructs. For purification of the ELP, the lysates were heated to 60 °C for 15 min in a water bath to remove native *E. coli* proteins. From here, the lysates were placed on ice for 30 min to take advantage of the temperature sensitive solubility of the A4Y1, which aggregates above temperatures of 37 °C but solubilizes at colder temperatures. After cooling on ice, the ELP lysates were centrifuged for 10 min at 20,000 rcf to remove denatured *E. coli* proteins. The heat treatment purification for the silk and ELP proteins was sufficient for isolating discrete bands of the recombinant proteins that were free of nearby contaminating bands on SDS-PAGE for subsequent titer analysis (Fig. S1). Purification of the A10 4mer BWT required the use of its histidine tag. Lysates of the A10 4mer BWT cultures were mixed with one part binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 8) and passed through 2 ml of HisPur Ni-NTA Resin (Thermo Fisher Scientific, Waltham, MA) that was placed in a 10 ml gravity-flow polypropylene chromatography column. The column was washed first with 6 ml of wash buffer 1 (20 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 8) and then 6 ml of wash buffer 2 (40 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 8). The A10 4mer BWT was then collected by passing 4 ml of elution buffer (300 mM imidazole, 20 mM Tris-HCl, pH 8) through the column.

All purified protein samples were detected via SDS PAGE using NuPAGE 10% Bis-Tris Midi Gels (Thermo Fisher Scientific, Waltham, MA) run in 1x NuPAGE MES SDS Running Buffer (Thermo Fisher Scientific, Waltham, MA) for 60 min at 120 V. Ten microliters of a sample was mixed with 10 µl NuPAGE LDS Sample Buffer (Thermo Fisher Scientific, Waltham, MA) prior to running the SDS PAGE. Gels were stained with SimplyBlue SafeStain (Thermo Fisher Scientific, Waltham, MA), and then destained in distilled water. To quantify the titer of silk and ELP protein, the relative intensity of the SDS PAGE bands was correlated to known protein concentrations. Standards of bovine serum albumin (BSA) ranging in concentration from 50 µg/ml to 2000 µg/ml were included in the SDS PAGE protocol, to serve as known standards for quantification of SDS PAGE band intensity with the ChemiDoc XRS + Image Lab software (Bio-Rad, Hercules, California). The imaging software and BSA standards run with each SDS PAGE allowed for the creation of a standards curve, which was used to quantify the concentration of silk or ELP protein in the samples based on the intensity of their SDS PAGE bands.

## 2.4. Plasmid maintenance assay

To assess the level of plasmid maintenance that strains had during recombinant silk or ELP expression, 100 µl of an expression culture was taken at the time of induction and then again at 2 and 4 h after induction. The 100 µl culture sample was serially diluted in fresh LB media, and 100 µl of this serial dilution was plated onto non-selective LB plates which were incubated overnight at 37 °C to obtain several hundred isolated colonies. For strains BL21, pLysS, pGro7, Origami, SoluBL21, and SoluBL21-pLysS a 10<sup>4</sup>-fold serial dilution was necessary to obtain the proper number of isolated colonies. For strains RosettaGami, BLR, HMS174, and MG1655 a 10<sup>2</sup> dilution was necessary to obtain enough colonies for the plasmid maintenances assay. After colonies from each strain had grown overnight on the non-selective LB plates, 50 colonies were patch plated onto LB plates with 100 µg/ml ampicillin using sterile pipette tips. The plates were incubated overnight at 37 °C and the percent plasmid maintenance was calculated by counting the number of colonies that had grown on the selective plate and dividing by 50.

## 2.5. FTIR analysis

Protein secondary structure confirmations were analyzed using a Bruker Vertex 70 FT-IR with the attenuated total reflectance accessory Bruker BioATR II (Bruker, Billerica, MA). The FT-IR spectra was collected with OPUS 8.1 software; sample absorbance was collected between  $800\text{ cm}^{-1}$  and  $4000\text{ cm}^{-1}$ , with 128 scans, and a resolution of  $4\text{ cm}^{-1}$ . Background subtractions were taken prior to each sample measurement and at least four trials were performed for each protein ( $n = 4$ ). Deconvolution of the amide I region ( $1600\text{--}1700\text{ cm}^{-1}$ ) was performed using the second derivative analysis and peak fitting in Igor Pro 8.04. Peaks were assigned to protein secondary structure as follows: beta-sheet ( $1610\text{--}1635$ ,  $1695\text{--}1700\text{ cm}^{-1}$ ), alpha-helix ( $1647\text{--}1664\text{ cm}^{-1}$ ), random coil ( $1635\text{--}1645\text{ cm}^{-1}$ ), and beta-turn ( $1666\text{--}1695\text{ cm}^{-1}$ ) (Yang et al., 2015; Lammel et al., 2010).

## 2.6. Genomic sequencing of SoluBL21

Genomic DNA of *E. coli* SoluBL21 (Genlantis, San Diego, CA) was extracted from a culture grown in LB media using the PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific, Waltham, MA). Uniquely indexed Illumina sequencing libraries with 200–300 bp insert length were prepared from 2  $\mu\text{g}$  gDNA using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA) following the manufacturer's instructions. The libraries were quantified on a Qubit 4.0 fluorometer using Qubit 1x ds DNA HS Assay kit (Thermo Fisher Scientific, Waltham, MA). Library size was analyzed by an Agilent 2100 Bioanalyzer with DNA 1000 Kit (Agilent, Santa Clara, CA). Pooled libraries were sequenced on an Illumina iSeq platform with iSeq 100 i1 Reagent v2 (Illumina, San Diego, CA) generating 2x150 bp paired-end reads at Genomics Research Core of Center for Biotechnology and Interdisciplinary Studies at Rensselaer Polytechnic Institute (Troy, NY). Illumina sequence data were quality assessed using FastQC v0.11.9 and filtered by removing low quality and adapter read regions using Trimmomatic v0.38 (Bolger et al., 2014). The *de novo* genome assembly was carried out using SPAdes v3.14.1 (Bankevich et al., 2012). Assembled contig sequences were aligned against the reference genome (GCF\_013166975.1\_ASM1316697v1\_genomic.fna) using BWA-MEM to verify genetic differences between strains SoluBL21 and BL21 (Li, 2013).

## 3. Results and discussion

### 3.1. Design of recombinant spidroin constructs and screening of ten *E. coli* strains

To screen recombinant spidroin expression in different *E. coli* strains, nine commercially available strains were used in this work (Supplementary Table 1). Some strains were chosen based on factors previously shown or hypothesized to affect recombinant silk production, such as codon usage and inclusion body formation. These strains include RosettaGami B, which has upregulation of seven tRNAs for rare codons including those for glycine and proline, as well as BLR, which has a  $\text{recA}^-$  mutation that may facilitate increased stability of plasmids containing repetitive sequences (Yedahalli et al., 2016; Huemmerich et al., 2004; Karla et al., 2005). The strain SoluBL21 has been developed through directed evolution to produce soluble protein when its ancestral strain, BL21(DE3), does not yield detectable soluble product. Likewise, strain pGro7 expresses a chaperone protein that prevents inclusion body formation and promotes soluble production (SoluBL21 Competent *E. coli*, 2022; Nishihara et al., 1998). The remaining strains were chosen to potentially identify unknown targets for the optimization of recombinant spidroin expression. This includes strains HMS174 and MG1655, which unlike other strains tested, are from the K-12 *E. coli* lineage instead of the B lineage (Hausjell et al., 2018; Leal-Egaña et al., 2012; Bachmann, 1972). Strain pLysS restricts basal expression, while strain Origami B contains mutations that change the cytoplasmic environment

and cellular stress responses through alterations to the thiol-redox equilibrium, glutathione metabolism, and oxidative stress response (Studier and Moffatt, 1986; Kong and Guo, 2014; Lobstein et al., 2016; Kunert et al., 1990). All strains used in this study were DE3 lysogens and proteins were expressed from the pET19b vector under control of the T7 promoter. We additionally studied a tenth hybrid strain, SoluBL21-pLysS, which we developed by combining features from pLysS and SoluBL21.

Four different *de novo* designed spidroin constructs were expressed in these *E. coli* strains, with titer, plasmid maintenance, and OD600 measured as expression outcomes. The primary sequences and polymeric structure of the spidroin constructs are depicted in Table 1. To assess the effect of protein size, recombinant spidroins were designed to have either four or sixteen identical monomer units in tandem (referred to as 4mers and 16mers, respectively). To assess the effect of modulating primary sequence, two different monomer units of 35 amino acids were designed, with one containing a segment of five tandem alanine residues (A5) and the other a segment of ten tandem alanine residues (A10). The remaining amino acids in the monomer sequences consists of multiple GPGQQ motifs (four for the A5 monomer and three for the A10 monomer) and single GPGEQ and GPGSG motifs. Both monomer units were designed based on naturally occurring primary sequences found in the MaSp2 dragline spidroin of orb-weaving spiders (Teulé et al., 2009; Guerette et al., 1996; Malay et al., 2017). Modulating the tandem alanine length and total construct length were chosen as focal points to study the effect of construct design on expression outcomes, as these are primary factors implicated in determining the material properties of recombinant silks (Bowen et al., 2018; Zhao et al., 2021; Bratzel and Buehler, 2012; Heidebrecht et al., 2015). All constructs expressed have an identical starting sequence that is present on the pET-19b expression vector, which contains a 10x histidine tag for purification followed by an enterokinase cleavage sequence (cleavage of the 10x histidine tag using the enterokinase site was not employed in this work).

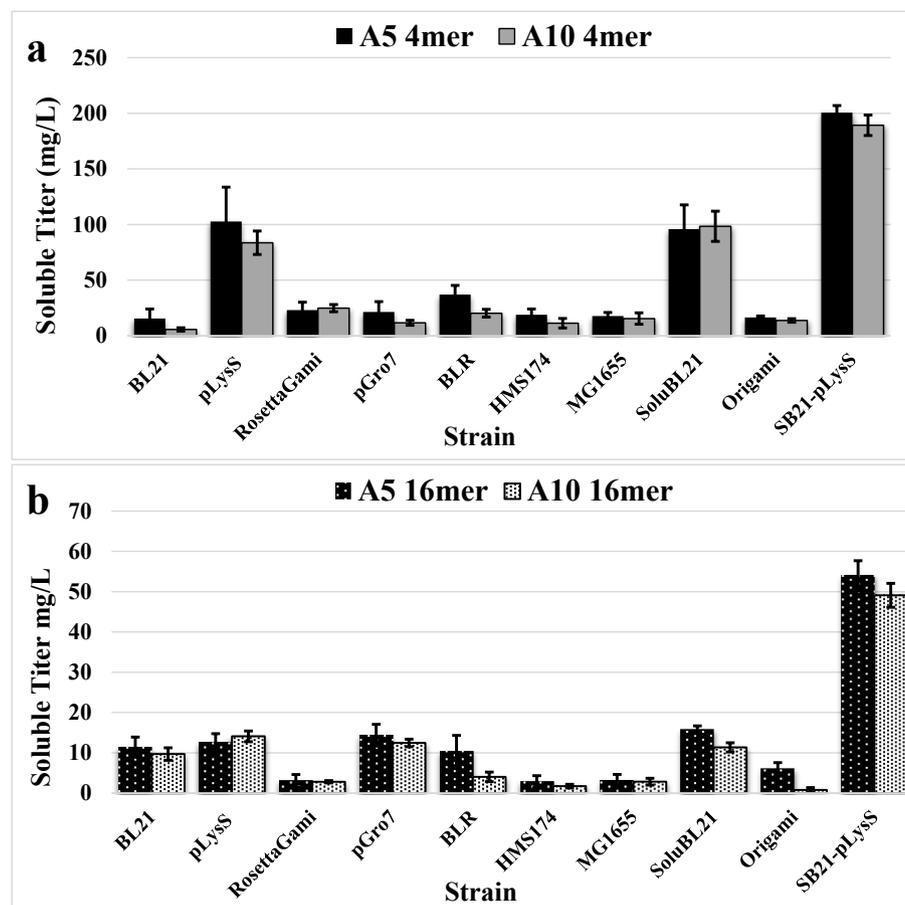
### 3.2. Recombinant silk titers in ten *E. coli* strains

Expression of the A5 and A10 constructs took place in LB media at  $37\text{ }^\circ\text{C}$  for 4 h. Initial findings showed that pLysS and SoluBL21 had higher production levels for the smaller recombinant spidroins (A5 4mer and A10 4mer). These strains yielded approximately 80–100 mg/L of A5 and A10 4mer protein, producing at levels several times above the other strains (Fig. 1a). Considering these results, the pLysS plasmid from the pLysS strain was extracted and transformed into SoluBL21 to form a novel hybrid strain, SoluBL21-pLysS, to investigate whether the advantages of the parent strains could be synergistic. The hybrid strain was able to produce the small spidroins at 201 ( $\pm 6$ ) and 189 ( $\pm 10$ ) mg/L for the A5 4mer and A10 4mer, respectively. These titers are approximately 70 mg/L higher than the reported shake flask titer for a recombinant spidroin that was recently produced at the highest titer ever reported for a recombinant spidroin in a bioreactor (Schmuck et al., 2021; Andersson et al., 2017). Interestingly, these titers are approximately twice that of either parent strain. Moreover, when compared to BL21, which is a typical strain used for spidroin expression, these titers represent a 13X increase for the A5 4mer and a 33X increase for the A10 4mer (Fig. 1a).

The titers for both the A5 and A10 16mers were lower than that of the 4mers for nearly all strains, which is consistent with previous findings showing an inverse relationship between yield and spidroin length (Xia et al., 2010). Despite displaying some of the highest titers for the 16mers, at 11–15 mg/L, strains pLysS and SoluBL21 showed no appreciable advantage over BL21, pGro7, or BLR, which all yielded similar titers (Fig. 1b). Interestingly, the SoluBL21-pLysS hybrid strain outperformed both of its parent strains for producing the 16mers, with titers of 53 ( $\pm 4$ ) and 49 ( $\pm 3$ ) mg/L for the A5 16mer and A10 16mer, respectively. This is approximately a four-fold increase in 16mer titer versus the parent strains. Several strains, including RosettaGami, HMS174, MG1655, and Origami B were barely capable of producing

**Table 1**  
Primary sequence of recombinant spidroins.

Spidroin	Primary Sequence	# of Repeats	Molecular Weight (kDa)
A5 4mer	MGHHHHHHHHHSSGHIDDDDKHMLEHMPG (GPGQQAAAAAGPGQQGPGQQGPGEQGPGSG) <sub>n</sub> TSGS	n = 4	16.1
A5 16mer	MGHHHHHHHHHSSGHIDDDDKHMLEHMPG (GPGQQAAAAAGPGQQGPGQQGPGEQGPGSG) <sub>n</sub> TSGS	n = 16	52.7
A10 4mer	MGHHHHHHHHHSSGHIDDDDKHMLEHMPG (GPGQQAAAAAGPGQQGPGQQGPGEQGPGSG) <sub>n</sub> TSGS	n = 4	15.6
A10 16mer	MGHHHHHHHHHSSGHIDDDDKHMLEHMPG (GPGQQAAAAAGPGQQGPGQQGPGEQGPGSG) <sub>n</sub> TSGS	n = 16	50.9



**Fig. 1.** (a) Soluble titer for the small recombinant spidroins, A5 4mer (black) and A10 4mer (gray), for ten strains of *E. coli* in LB media. Error bars represent standard deviations from the mean values of three replicates. (b) Soluble titer for the large recombinant spidroins, A5 16mer (black with white dots) and A10 16mer (white with black dots), in ten strains of *E. coli* in LB media. Error bars represent standard deviations from the mean values of three replicates.

detectable levels of the 16mers, as shown by titers of 5 mg/L or less. The poor performance of RosettaGami B for both the 4mers and 16mers is of particular interest, as a leading hypothesis in the field relates the low titers of silk proteins to translational difficulties caused by the overabundance of a select few amino acids in spidroins, mainly glycine (Yang et al., 2016; Bowen et al., 2018; Xia et al., 2010). Recent work has shown that upregulating glycyl-tRNA has favorable effects on the production of spidroins, however, the upregulation of glycine and proline tRNAs inherent in RosettaGami appeared to have no positive effect on titer (Yang et al., 2016; Bowen et al., 2018; Xia et al., 2010). Moreover, we observed no translation truncation of the A5 and A10 spidroins in SDS PAGE analysis (Fig. S1). Translational truncation, or early termination of the nascent recombinant spidroin from the ribosomal complex, has

been observed by multiple researchers and is indicative of translational bottlenecks in silk production (Bhattacharyya et al., 2021; Xia et al., 2010; Fahnestock and Irwin, 1997). These findings support the hypothesis that translational difficulties alone cannot fully explain the difficulties in obtaining high titers of recombinant silk.

### 3.3. Recombinant ELP production in BL21, pLysS, SoluBL21, and SoluBL21-pLysS

To investigate if the increased spidroin titers achieved with the hybrid SoluBL21-pLysS strain could be extended to other repetitive, structural proteins, an elastin-like peptide (ELP) was produced in strains BL21, pLysS, SoluBL21, and SoluBL21-pLysS. The recombinant ELP,

A4Y1, was chosen for production in these four strains based on a balance between similarity and difference when compared to the A5 4mer primary sequence (Table S2). The A5 4mer and the A4Y1 ELP both have a 4mer polymeric structure, along with a near identical molecular weight and glycine, proline, and alanine contents. Furthermore, both recombinant spidroins and ELPs are known to self-assemble into supramolecular materials when triggered by external stimuli (Schmuck et al., 2021; Sarkar et al., 2019; Varanko et al., 2020). However, key differences are that the A5 4mer has tandem alanines ( $A_n$ ) while the ELP has alanine residues distributed throughout the construct. Additionally, the A5 4mer has a high amount of glutamine (21%), which the ELP lacks entirely, and the ELP has a high percentage of valine (16%) and some tyrosine (3%), both of which are missing from the A5 4mer.

Fig. 2a shows that strains pLysS, SoluBL21, and SoluBL21-pLysS offered no advantage over BL21 for the titer of the ELP construct. The soluble titers for these three strains were similar, at  $\sim 240$  mg/L. SoluBL21 performed the worst out of the four strains with a titer of  $196 (\pm 8)$  mg/L. In all cases, the A5, A10, and ELP constructs were expressed primarily in the soluble fraction of the lysate, with only negligible amounts found in the pellet ( $< 2$  mg/L for all strains). Notably, the titer for the ELP in BL21 is over 15 times higher than for the A5 4mer under identical expression conditions. This finding, as well as the observation that the hybrid strain offered no advantage, were both unexpected when considering the high degree of similarity between the A5 4mer and the A4Y1 ELP. As peptides of identical molecular weight and similar polymeric structure, a possible explanation for these outcomes may lie in structural differences that result from primary sequence variation, the mainly the disparity between valine and glutamine content.

### 3.4. Plasmid maintenance during expression of silk and ELP constructs

During a recombinant protein expression, cells can potentially lose the plasmid vector that was transformed into them. Plasmid loss is indicative of excessive metabolic burden, which may stem from repetitive recombinant DNA sequences or toxic recombinant protein products and is exacerbated by depletion of antibiotic selection factors (Yang et al., 2016; Dumon-Seignovert et al., 2004; Corchero and Villaverde, 1998). Plasmid-free cells can continue to divide, leading to a substantial decrease in the overall number of cells in a culture that are producing recombinant protein. A high level of plasmid maintenance is a critical factor for achieving high titers, particularly for high-density cell cultivation in bioreactors (Yang et al., 2016). Fig. 3 shows the plasmid maintenance of the A5 and A10 constructs in the ten *E. coli* strains, and Fig. 2b shows the plasmid maintenance of the ELP construct in BL21,

pLysS, SoluBL21, and SoluBL21-pLysS. Both figures represent the plasmid maintenance at the end of a 4-h expression in LB media. Fig. 4 shows that plasmid maintenance decreased in 75% of cases when a strain transitioned from a 4mer to a 16mer expression (ex. pLysS A5 4mer compared to pLysS A5 16mer), while increasing in only 10% of cases. Instability of recombinant silk plasmids has been previously reported, but to our knowledge this is the first work to observe a correlation between plasmid instability and number of genetic repeats across multiple spidroin designs and *E. coli* strains (Yang et al., 2016; Chung et al., 2012; Edlund et al., 2018). Although plasmid maintenance of the 16mers decreased substantially for pLysS compared to the 4mers, the hybrid strain maintained the ability of SoluBL21 to maintain these 16mer vectors at 85% or above.

This data suggests that a high titer of recombinant spidroin required high plasmid maintenance, though high plasmid maintenance does not necessarily lead to high titer in every case. The strains that yielded the highest titers of the 4mer proteins, pLysS, SoluBL21, and SoluBL21-pLysS, all exhibit a plasmid maintenance of 90% or higher. Likewise, the strain that yielded the highest titers for the 16mers, SoluBL21-pLysS, exhibited one of the highest overall plasmid maintenance levels for the 16mers. However, several strains, namely MG1655, BLR, and HMS174, exhibited a high level of plasmid maintenance in some cases (upwards of 90%) but relatively low titers for all constructs. This indicates that factors outside of the maintenance of spidroin vectors are at least partially responsible for low titers. For the BL21, RosettaGami, pGro7, and Origami strains, there was nearly a complete loss of the plasmid during spidroin expression. In contrast, Fig. 2b shows that strain BL21 exhibited a plasmid maintenance of 43% during the ELP expression, which is over 14 times higher than maintenance during the expression of the A5 4mer or any other spidroin. This is in accordance with the 15 times higher level of production that BL21 was able to achieve for the ELP versus the A5 4mer. The strains pLysS, SoluBL21, and SoluBL21-pLysS exhibited levels of plasmid maintenance for the ELP at or near 100%, similar to their behavior during spidroin expressions.

### 3.5. Cell growth during expression of spidroin and ELP constructs

We investigated cell growth over a 4-h expression of the four different spidroin constructs (Fig. 4). While all cultures were induced for protein expression at an OD600 of 0.6–0.8, the final OD600 at the end of the 4-h expression was highly variable among the strains, ranging from 1.58 to 3.86. The high-producing strains, pLysS, SoluBL21, and SoluBL21-pLysS, showed final OD600s that were at or near the median of the dataset obtained for the ten strains (median of 2.07). The OD600

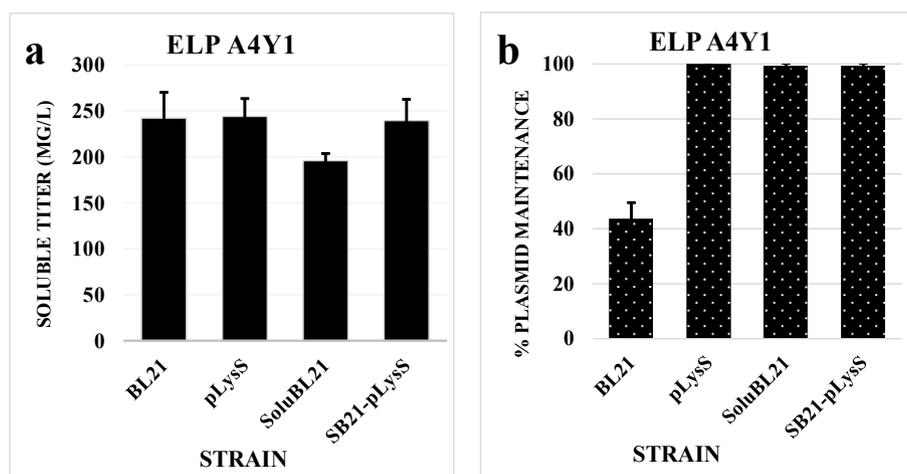
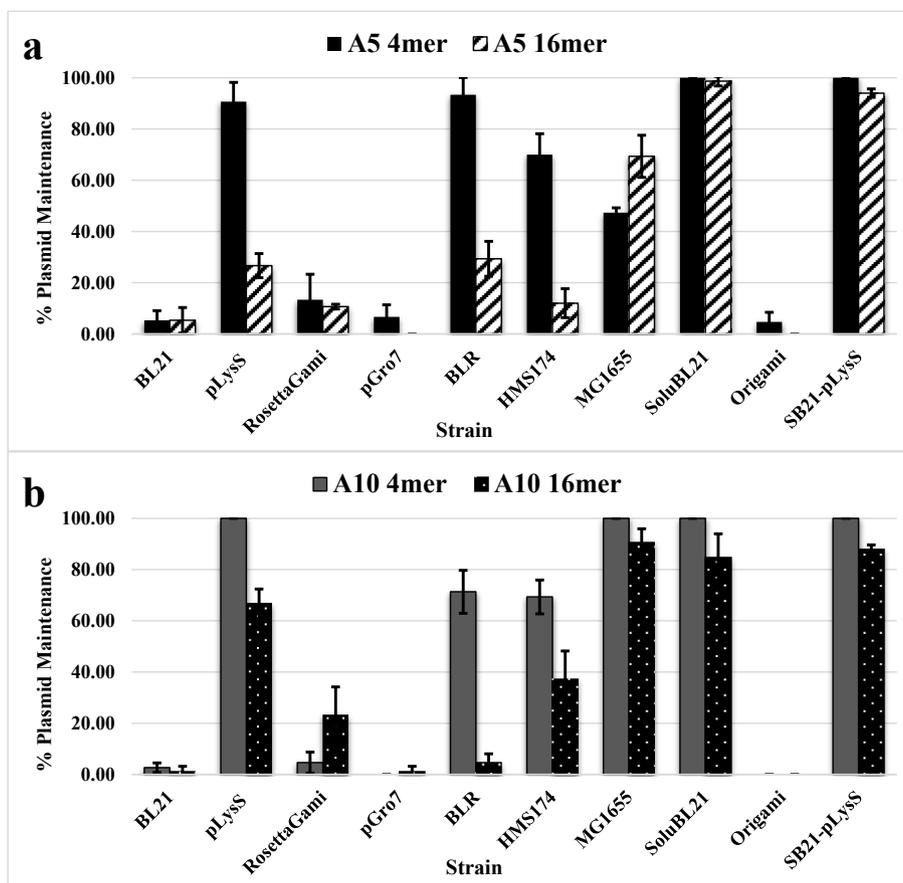
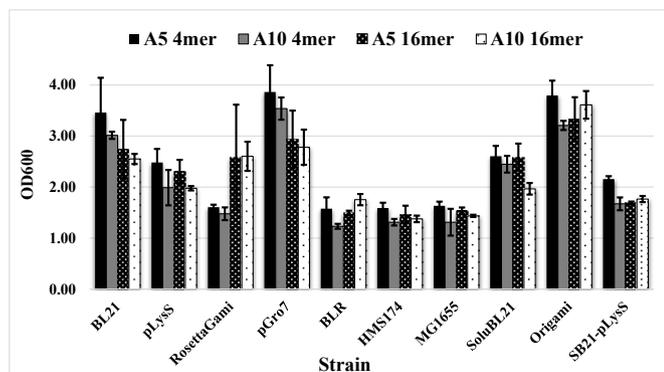


Fig. 2. (a) Titer of the A4Y1 ELP in BL21, pLysS, SoluBL21, and SoluBL21-pLysS (SB21-pLysS) strains. Error bars represent standard deviations from the mean values of three replicates. (b) Plasmid maintenance of the ELP construct in the strains BL21, pLysS, SoluBL21, and SoluBL21-pLysS (SB21-pLysS) at the end of a 4-h expression in LB media. Error bars represent standard deviations from the mean values of three replicates.



**Fig. 3.** Plasmid maintenance of the A5 4mer (blue), A5 16mer (gray), A10 4mer (orange), and A10 16mer (yellow) in the ten *E. coli* strains at the end of a 4-h expression in LB media. Error bars represent standard deviations from the mean values of three replicates.



**Fig. 4.** OD600 of cultures directly before and after a 4-h expression of the four different silk constructs. All cultures were induced for protein expression at an OD600 of 0.6–0.8, the final OD600 at the end of the 4-h expression ranged from 1.58 to 3.86. Error bars represent standard deviations from the mean values of three replicates.

at the end of a 4-h expression did not show an obvious relationship to spidroin titer, as strains with poor titers showed both higher and lower final OD600s than pLysS, SoluBL21, and SoluBL21-pLysS. Notably, strains BL21, pGro7, and Origami B, which grew the most during spidroin expression by routinely reaching final OD600s of above 3, were also the strains that showed the lowest levels of plasmid maintenance in addition to low titers. This is likely due to the degradation of ampicillin in the media, which allows non-plasmid bearing cells that are potentially fitter to proliferate. This phenomenon is particularly applicable in cases where the recombinant construct is harmful or toxic to the cells

(Dumon-Seignovert et al., 2004; Corchero and Villaverde, 1998). This problem is further exacerbated by using ampicillin over other antibiotics as the selection pressure, since the product of the beta-lactamase gene conferring resistance to ampicillin is secreted, with studies showing that rapid plasmid loss and growth of non-plasmid bearing cells can be difficult to prevent, even in cases where additional ampicillin is added to the culture (Dumon-Seignovert et al., 2004; Corchero and Villaverde, 1998; Sieben et al., 2016; Rosano and Ceccarelli, 2014). Thus, the increased growth rates of strains that have lost the vector and produce little silk protein may suggest that expression of the A5 and A10 spidroins exerts toxicity on the cells.

The possibility that expressing A5 and A10 spidroins causes host cell toxicity was further supported by observations made during the plasmid maintenance assay, in which 0.1 ml of a 10,000x culture dilution (generated through serial dilutions) was plated for colony counting. For most strains observed, this procedure resulted in several hundred single colonies on LB agar plates. However, the strains that showed moderate to high levels of plasmid maintenance but low titers and inhibited growth after induction (RosettaGami, BLR, HMS174, and MG1655) displayed a lack of colony forming units on LB plates using this protocol. Compared to other strains at the same OD600, cultures of RosettaGami, BLR, HMS174, and MG1655 required a 100x (instead of a 10,000x) dilution of a culture sample to obtain enough isolated colonies for the plasmid maintenance assay (minimum of 50 colonies required). This lack of colony-forming-units after recombinant protein induction is a documented effect of toxic protein expression in cases where the vector is still maintained (Dumon-Seignovert et al., 2004; Rosano and Ceccarelli, 2014; Onodera et al., 1996; Saïda et al., 2006; Kwon et al., 2015). Support for a toxicity effect from the spidroins, but not necessarily the ELP, can be seen when the final OD600 for cells expressing ELP vs A5

4mer. (Fig. S2). The final OD600 at the end of a 4-h expression for the ELP was higher for pLysS, SoluBL21, and SoluBL21-pLysS (+0.22, 0.72, 0.575, respectively), even with a higher titer of ELP than A5 4mer. There was a decrease in the final OD600 of  $-0.64$  for BL21 during ELP expression when compared to A5 4mer. However, this can likely be attributed to the large difference in titers between the two proteins, with 15 times more ELP than A5 4mer made by BL21. Protein overexpression at a high level, independent of the toxicity of the construct, is associated with decreases in cell growth (James et al., 2021a).

### 3.6. Toxicity of recombinant spidroin constructs

To further study the potential toxicity of the A5 and A10 spidroins when compared to the ELP, growth under a variety of conditions was measured for strain SoluBL21-pLysS. If a recombinant product is toxic to the host, then growth during production will be decreased or completely inhibited, leading to an earlier stationary phase at a lower OD600 when compared to uninduced or empty vector controls (Dumon-Seignovert et al., 2004; Rosano and Ceccarelli, 2014; Onodera et al., 1996; Saïda et al., 2006; Kwon et al., 2015). Fig. 5 shows the OD600 of SoluBL21-pLysS in LB media for 10.25 h under a variety of conditions and starting from the point of inoculation (2% v/v inoculum from an overnight culture). The uninduced cultures and empty vector controls reach a late exponential or early stationary phase after 10.25 h and at an OD600 of approximately 3.72–4.18. The cultures induced for ELP production follow these curves closely in both rate of growth and overall

growth, as there is no significant decrease in the slope of the curve after induction, and the final OD600 of 3.42 is only an 8% decrease from that of the induced empty vector control (3.72). These findings support the conclusion that ELP expression does not significantly affect cell health and fitness, even at titers above 200 mg/L. However, when expressing an A5 or A10 spidroin, there is a significant decrease in the slope of the growth curve directly after induction at 1.75 h. Then, a stationary phase is reached up to 5 h sooner and the final OD600 decreases on average by  $-61\%$  and  $-57\%$  when compared to the uninduced or induced empty vector controls, respectively. When expressing the A10 4mer, there is potentially a death phase (consistent decrease in OD600) that begins approximately 4 h after induction.

### 3.7. Disorder of recombinant spidroins as a primary factor for toxicity and low titers

Our data on growth kinetics, plasmid maintenance, and expression levels among the strains, suggests a toxicity of the A5 and A10 spidroin constructs that can be addressed with proper strain engineering to obtain higher titers. As this is in sharp contrast to the ELP, we performed additional work to understand why the expression outcomes between A5 4mer and ELP, which have similar polymeric structure and amino acid composition, were profoundly different. An important observation throughout this work was the aberrant mobility of purified spidroins through an SDS-PAGE gel (Fig. S1). The A5 4mer and ELP both have molecular weights of 16 kDa, but the A5 4mer shows an apparent weight

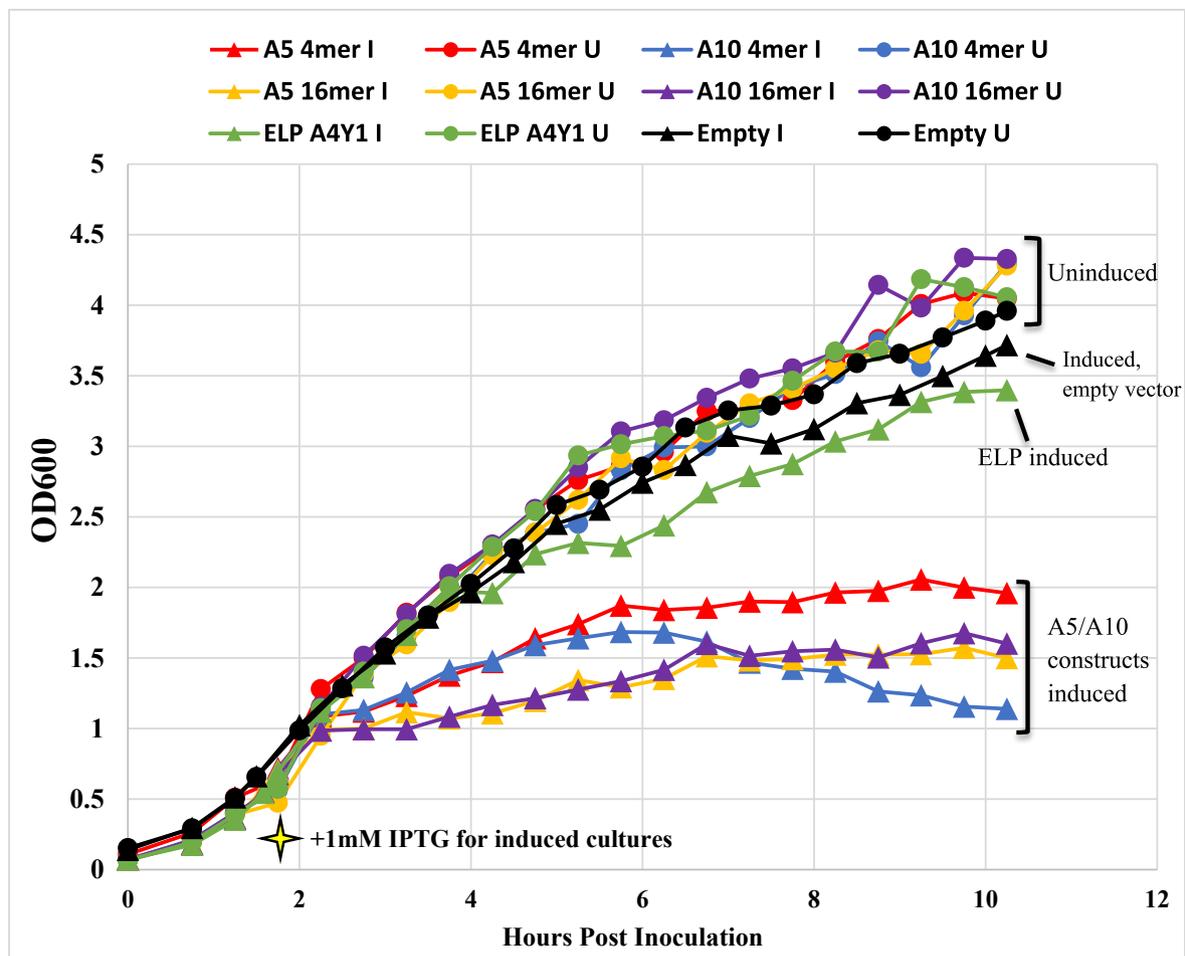


Fig. 5. OD600 of SoluBL21-pLysS in LB media for 10.25 h under a variety of conditions. T = 0 indicates time of inoculation (2% v/v inoculum from an overnight culture). For induced cultures, induction occurred with 1 mM IPTG at 1.75 h post inoculation. Growth curves for all silk constructs and the ELP under induced (triangles, I) and uninduced (circles, U) conditions were generated. Growth curves for the strain harboring the empty expression vector (Empty) that lacks a recombinant gene were also generated with and without induction. Data points on the curve represent the mean OD600 for that time point from three replicates.

of 38 kDa on an SDS PAGE, whereas the ELP appears at 14 kDa, which is close to its true molecular weight. A high degree of aberrant mobility has been shown to be a unique characteristic of disordered proteins and is observed for the three other spidroins as well (Fig. S1) (Pedersen et al., 2020; Tan et al., 2021). Analysis of amino acid content indeed suggests that the spidroins may be more disordered than the ELP. Amino acids can be ranked in terms of their propensity to promote structural order or disorder (from most order-promoting to least order-promoting): W, F, Y, I, M, L, V, N, C, T, A, G, R, D, H, Q, K, S, E, P (Campen et al., 2008). While both the spidroin and ELP constructs have similar glycine and proline content (~33% and ~15%, respectively), the spidroins also contain serine, glutamate, and a high proportion of glutamine (more than 20% for the A5 constructs), all of which are lacking in the ELP sequence (Table S2). Furthermore, the ELP sequence contains a higher proportion of amino acids that promote structural order, including valine and tyrosine (both are absent from the A5 4mer).

Computational tools predict a much greater likelihood of disorder in our spidroins versus the ELP (Figs. S3 and S4). The IUPRED2A disorder predictor shows near 100% probability that all residues within the A5 4mer sequence are in regions of structural disorder. In contrast, the ELP sequence fluctuates around approximately 50% disorder probability throughout, with the probability of disorder never rising above 75% (Fig. S3) (Mészáros et al., 2018). The DisMeta computational disorder calculator suggests the entire length of the A5 and A10 spidroin constructs as having a 90–100% chance of being disordered (Huang et al., 2014). In contrast, the ELP contains large regions that have much lower probability of disorder, including sections with disorder probability of 10% or less (Fig. S4). Additionally, FTIR and secondary structure analysis show that the A5 4mer contains a 37% ( $\pm 11$ ) random coil content, compared to 0% ( $\pm 0$ ) for the ELP (Fig. 6. Raw, reproduced, and deconvoluted spectra can be seen in Fig. S5), where random coils can be associated with structural disorder and conformational flexibility in proteins (James et al., 2021b; Choi et al., 2011; Marsh and Forman-Kay, 2010).

We hypothesize that the disordered nature of the spidroin constructs is a main factor underlying their toxicity, low titers, and low plasmid maintenance when compared to the ELP. Overexpression of disordered proteins has been shown to exert toxicity in both *D. melanogaster* and *C. elegans* in a dose-dependent manner (Pedersen et al., 2020; Vavouri

et al., 2009). Production of disordered recombinant proteins has also been shown to yield negligible titers and cause toxicity in *E. coli* and yeast species such as *S. cerevisiae* (Treusch and Lindquist, 2012; Liang et al., 2010; Hwang et al., 2012; Guo et al., 2021). It is hypothesized that this toxicity may result from promiscuous and harmful binding interactions by disordered proteins within the intracellular milieu (Pedersen et al., 2020; Vavouri et al., 2009; Treusch and Lindquist, 2012; Liang et al., 2010; Hwang et al., 2012; Guo et al., 2021). To our knowledge, this is the first work to show toxicity resulting from protein disorder as a primary factor causing unfavorable outcomes during recombinant spidroin expression. This is of particular interest to the field, as many recombinant MaSp2-mimetic spidroins produced by other groups are similar to our A5 and A10 constructs in that they are high in disorder-promoting amino acids (proline, glutamine, glycine, serine) and low in order-promoting amino acids with low titers consistently reported (Table S3).

### 3.8. Addressing spidroin toxicity with strain engineering and experimental design

Our hybrid SoluBL21-pLysS strain showed substantially improved recombinant spidroin expression compared to other *E. coli* strains (Fig. 1). To understand the underlying cellular mechanisms, a series of experiments examining basal expression were performed. Basal expression refers to the expression of a recombinant gene without induction, which can cause plasmid loss and subsequent low titers if the gene product is toxic. Strain pLysS (and SoluBL21-pLysS by default) is designed to exert tight control over basal expression through the action of the pLysS vector. The product of the pLysS vector is T7 lysozyme, which inhibits action of T7 polymerase and prevents basal expression of recombinant genes placed on pET vectors. Upon addition of IPTG T7 polymerase concentration increases and overcomes the inhibition of pLysS (Rosano and Ceccarelli, 2014). Fig. 7a presents experimental verification that basal expression was strongly restricted in strains pLysS and SoluBL21-pLysS, showing either an absence or low levels (<7 mg/L) of A5 4mer protein expression from an overnight culture grown without IPTG. Basal expression was also attenuated in strain SoluBL21, although slightly less so than for the aforementioned strains (11 ( $\pm 3$ ) mg/L from an overnight culture without IPTG). Strain BL21 exhibited copious basal expression, with overnight cultures (18 h) lacking any IPTG induction yielding an average A5 4mer titer of 102 ( $\pm 11$ ) mg/L. Cultures of BL21 inoculated from an overnight culture and grown to OD600 of 0.6–0.8 also showed basal expression at 8 ( $\pm 1$ ) mg/L. Moreover, when the strain was put through the same 4-h expression protocol as previously used, save for the addition of IPTG, the titer was 83 ( $\pm 9$ ) mg/L. These basal expression titers for the A5 4mer in BL21 are significant in that they are approximately equal to that of the pLysS and SoluBL21 strains, however, they are only half of what the high-performance SoluBL21-pLysS strain yielded.

Growth curves for strain BL21 give further insight into the effect of basal spidroin expression. Fig. 7b shows OD600 over 10.25 h for the strain under a variety of conditions, including growth with and without inducer and empty vector controls. This data shows that basal expression of the A5 4mer construct had almost no negative effect on the growth of the BL21 strain when compared to controls, even though 83 ( $\pm 9$ ) mg/L of A5 4mer protein had accumulated intracellularly by 6.75 h into the experiment. Interestingly, the induced strain showed no significant deviation in growth from the uninduced curves, however the titer at 6.75 h total culture time (4 h post induction with 1 mM IPTG) was only 15 ( $\pm 8$ ) mg/L. Thus, counterintuitively, the BL21 strain produces approximately 5.5 times more silk protein when it is not induced with IPTG. Fig. S6 shows that this outcome is likely a result of changes in plasmid maintenance in response to induction. The plasmid maintenance at 2.75 h (directly before induction) for both uninduced and induced are roughly the same (50%), albeit low due to the basal expression of a toxic gene (Dumon-Seignovert et al., 2004; Sieben et al.,

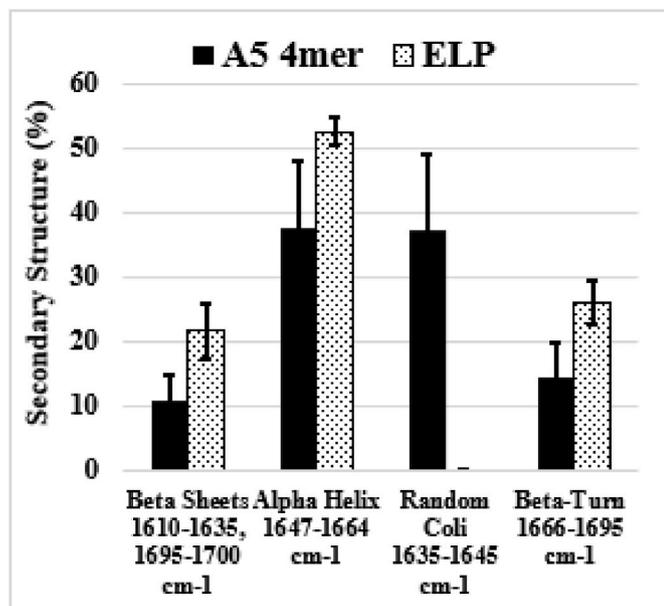
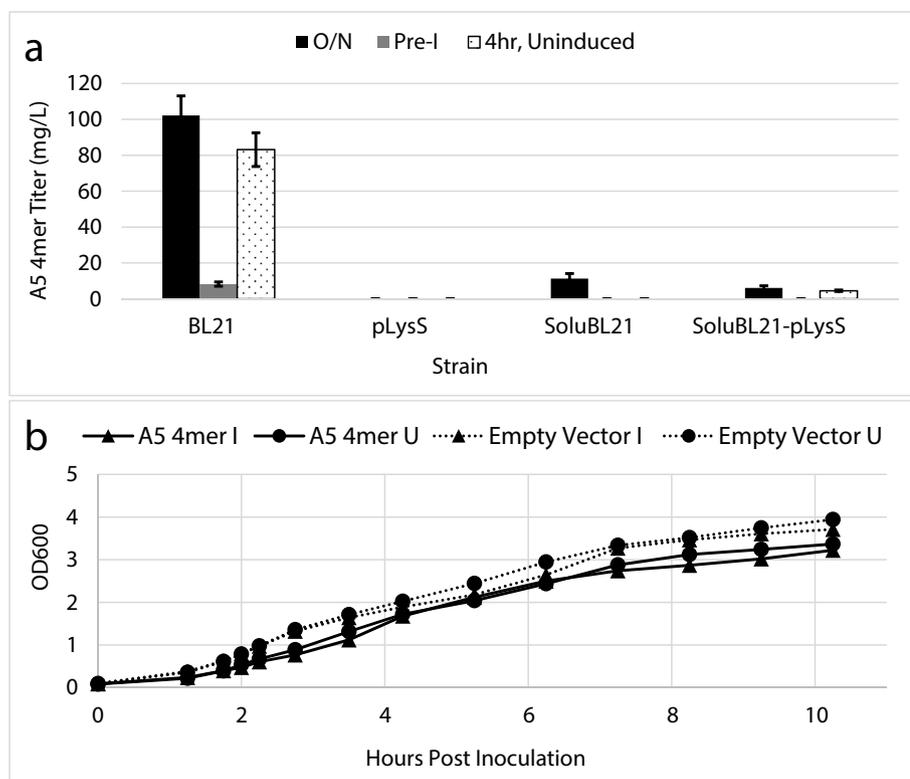


Fig. 6. Secondary structure analysis derived from FTIR spectra for the A5 4mer and ELP proteins. Error bars represent standard deviations from the mean values of three replicates.



**Fig. 7. (a)** Basal Expression of A5 4mer in strains BL21, pLysS, SoluBL21, and SoluBL21. Basal expression from a culture grown overnight (18 h) after inoculation from a single colony (blue bar). Basal expression from a culture inoculated from an overnight liquid culture (2% v/v inoculum), grown to OD600 of 0.6–0.8 and harvested (orange bar). The orange bar thus represents basal expression directly before what would have been the induction point for a typical expression culture. Basal expression during a 4-h expression without addition of IPTG (gray bar). The gray bar shows basal expression from a culture inoculated from an overnight liquid culture (2% v/v inoculum), grown to OD600 of 0.6–0.8, and allowed to grow for four more hours without addition of IPTG. Error bars represent standard deviations from the mean values of three replicates. **(b)** A5 4mer Basal Expression Growth Curve for strain BL21. OD600 over 10.25 h for strain BL21 with the A5 4mer or empty vector under induced (I) or uninduced (U) conditions. Cultures were inoculated with 2% v/v inoculum from an overnight culture. Induction (if applicable) occurred with 1 mM IPTG at 2.75 h total culture time when the OD600 was 0.6–0.8. Empty vector corresponds to strain BL21 transformed with the pET19b vector that has had the silk gene excised. Data points on the curve represent the mean OD600 for that time point from three replicates.

2016; Kwon et al., 2015). However, after induction with 1 mM IPTG, the plasmid maintenance of the induced culture decreases to just 3% after 4 h while the uninduced maintains the vector at 26%. This finding suggests that the leakiness, and likely strength, of a promoter plays an integral role in the successful production of toxic recombinant spidroins. In uninduced conditions, the BL21 strain is already expressing the recombinant silk gene, meaning that subsequent additions of IPTG may overwhelm cellular machinery beyond its ability to handle toxic protein expression, causing severe plasmid loss and low titers. Likewise, since the pLysS, SoluBL21, and SoluBL21-pLysS strains do not exhibit much basal expression they exhibit high plasmid maintenance and are likely more tolerant of the increase in promoter activity caused by IPTG (Yang et al., 2016; Dumon-Seignovert et al., 2004; Corchero and Villaverde, 1998; Kwon et al., 2015).

In this context, weaker promoters, or a strong promoter in combination with a restriction on basal expression, may be easily implemented tools for optimizing spidroin expression. The BL21 strain and strong promoters, mainly the bacteriophage T7 promoter, are commonplace in the recombinant silk field (Table S3), yet they may be a reason for the fundamental problems related to low titers, biomass accumulation, and plasmid stability (Huemmerich et al., 2004; Scheibel, 2004; Lewis et al., 1996; Prince et al., 1995; Mulinti et al., 2022; Malay et al., 2020; Teulé et al., 2007). Much in the same way that temperature downregulation mitigates toxicity of a recombinant construct and increases spidroin titer in *E. coli* (Schmuck et al., 2021; Yang et al., 2016), tightly regulating basal expression through strain or plasmid engineering can further enhance bacterial production of recombinant spidroins (Schmuck et al., 2021; Yang et al., 2016; Sørensen and Mortensen, 2005; Kawai et al., 2019; Mujacic et al., 1999). In cases where this is not feasible, forgoing an inducer and using longer culturing times may yield more favorable outcomes for strains or vector systems exhibiting basal expression of a recombinant spidroin or toxic protein. Furthermore, these methods may yield beneficial economic outcomes in scenarios where the relatively high cost of IPTG outweighs the cost of longer culturing times (Edlund et al., 2018).

Notwithstanding, restricting basal expression alone does not fully explain why the hybrid strain SoluBL21-pLysS performed better than either of its parent strains. The SoluBL21 strain was developed partly for toxic protein expression. Along with having a restriction of basal expression, sequencing of the strain reveals key mutations that differentiate it from BL21. We identified mutations on 47 genes in SoluBL21. Many of these mutations occur on uncharacterized proteins or on genes related to mannitol or glycerol metabolism, which are unlikely to affect silk production in LB media to our knowledge. However, there are 14 genes with mutations that are directly involved in stress responses in *E. coli*. These genes include *yggW*, *yedY*, *yedW*, *yedY*, *speC*, *speB*, *uspC*, *hchA*, *loiP*, *mltC*, *envZ*, *ompR*, *yhgF*, *hupB*. These genes play roles in stress responses pertaining to heat, reactive oxygen species, cell surface damage, salt changes, acid exposure, carbonyls, osmotic changes, putrescine production, nutrient starvation, ethanol exposure, radiation, and the SOS response (Table S4) (Nonaka et al., 2006; Urano et al., 2015, 2017; Gennaris et al., 2015; Tabor and Tabor, 1985; Kurihara et al., 2005; Schneider et al., 2013; Hafner et al., 1979; Gustavsson et al., 2002; Mujacic et al., 2004; Subedi et al., 2011; Mujacic and Baneyx, 2006, 2007; Weber et al., 2006; Huang et al., 2008; Hagiwara et al., 2003; Pomposiello et al., 2003; Tokishita et al., 1992; Chakraborty et al., 2017; Barbieri et al., 2013; Byrne et al., 2014; Goshima et al., 1990; Stojkova et al., 2019; Maslowska et al., 2019).

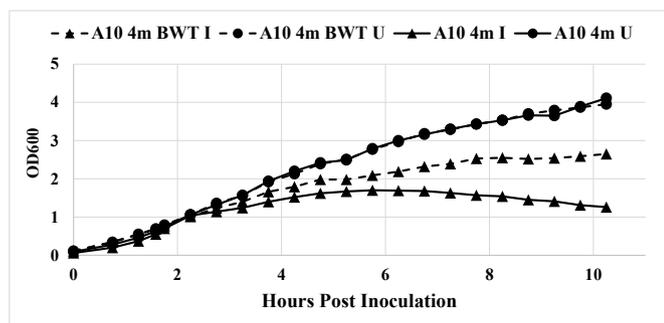
Several of the mutations in SoluBL21 occur on genes responsible for extensive and broad stress pathway signaling within *E. coli*. This includes the *envZ/ompR* two-component system, in which a mutation on the *envZ* gene causes constant phosphorylation of the *ompR* transcriptional regulator. This results in a decreased repression of several stress response pathways, including those related to osmotic and acid stress (Tokishita et al., 1992; Chakraborty et al., 2017; Barbieri et al., 2013). Furthermore, mutations in the *speC* and *speB* genes may alter putrescine production pathways, with putrescine production representing a fundamental way that several organisms, including *E. coli*, respond to a myriad of harmful conditions (Tabor and Tabor, 1985; Kurihara et al., 2005; Schneider et al., 2013; Hafner et al., 1979). Likewise, the *uspC*

(Universal stress protein C) gene is mutated, potentially promoting a more favorable cell phenotype in response to silk protein production, as *uspC* is induced by a diversity of stress factors that includes nutrient starvation (of multiple types), oxidative stress, DNA damage, radiation, heat shock, and ethanol exposure (Gustavsson et al., 2002). There are also mutations in the DNA-binding region of *hupB*, a transcriptional factor that controls expression of 8% of the entire genome in regions where the genes are associated with adaptations to harsh environments, including the SOS response system, and oxidative and radiative stress systems (Goshima et al., 1990; Stojkova et al., 2019). It stands to reason that these mutations promote a cell phenotype that is more tolerant to the expression of toxic and disordered proteins, and results in the favorable characteristics observed for strains SoluBL21 and SoluBL21-pLysS. Additionally, identification of these mutations provides previously unknown targets for continued strain optimization and supports the conclusion that strains which harbor targeted adjustments to their stress response genotype are necessary for improving recombinant spidroin production.

### 3.9. Addressing spidroin toxicity with targeted protein design

Recent work in the recombinant silk field has initiated a deeper exploration into the non-repetitive terminal domains that are present in natural spidroins. These terminal domains, typically 100–200 amino acids in length, can promote solubility and controlled self-assembly of recombinant spidroins (Andersson et al., 2017). Moreover, the highest reported titer for a recombinant silk protein in a bacterial system used a construct that included terminal domains flanking a repetitive core domain (Schmuck et al., 2021). Previous work supports the idea that the terminal domains can dimerize and promote micelle formation among silk proteins, with the repetitive and potentially disordered core domain enveloped by more ordered terminal domains (Römer and Scheibel, 2008; Schwarze et al., 2013). We hypothesized that this phenomenon may shield a recombinant host strain from the toxic effects of a disordered protein sequence by sequestering it into a compact micelle.

Thus, we explored the use of terminal domains on increasing titers and decreasing toxicity in our system. We inserted cDNA copies of the terminal regions of *L. hesperus* (western black widow) MaSp1 dragline silk at either end of the A10 4mer gene to form the A10 4mer BWT construct (Fig. S7). The 43 kDa A10 4mer BWT protein was readily produced and purified from the SoluBL21-pLysS strain at  $96 (\pm 12)$  mg/L, a titer that is approximately twice that of the similarly sized (50 kDa) A10 16mer which also contains an identical repeat sequence. Fig. 8 shows the growth of strain SoluBL21-pLysS over 10.25 h during induced expression of the A10 4mer and A10 4mer BWT constructs, with uninduced cultures serving as controls. Including the black widow termini



**Fig. 8.** Cell growth during expression of the A10 4mer (solid line – with induction (I), triangles – without induction (U), circles) and A10 4mer BWT (dashed line – with induction (I), triangles – without induction (U), circles) constructs with and without induction. For the induced curves, induction occurred with the addition of 1 mM IPTG at 1.75 h at an OD<sub>600</sub> of 0.6–0.8. Data points on the curve represent the mean OD<sub>600</sub> for that time point from three replicates.

increases the final OD<sub>600</sub> versus the original A10 4mer by 110%. Unlike the original A10 4mer expressions, during expression of A10 4mer BWT, the strain exhibits a later stationary, and no death phase is observed over the 10.25 h. FTIR data on the constructs shows that there is a substantial shift in protein structure when the terminal domains are included, with A10 4mer BWT demonstrating an increase in alpha helices and a decrease in beta turns (Fig. S8). Decreased disorder of A10 4mer BWT versus A10 4mer is supported by observation of more normal mobility in SDS PAGE, where the 43 kDa A10 4mer BWT shows an apparent molecular weight of 48 kDa while the 15 kDa A10 4mer shows an apparent weight of 27 kDa (Figs. S1 and S9). Furthermore, computational disorder predictions for the A10 4mer BWT show large regions where individual residues have less than 10% probability of disorder, while the original A10 4mer construct is shown to have 90–100% probability of disorder at all residues (Fig. S4). These findings support the hypothesis that including the terminal regions either decreases structural disorder and/or shields the disordered core domain from the host organism, facilitating increased biomass accumulation while maintaining a high titer. Thus, modulating primary sequence with spidroin terminal domains is another strategy that can be used to decrease toxicity and potentially increase titers of recombinant spidroins, while likely maintaining the construct's ability to exhibit silk-mimetic material properties (Schmuck et al., 2021; Andersson et al., 2017). More broadly, these findings may provide valuable insight for producing toxic or disordered proteins in general. Particularly in situations where a toxic recombinant protein sequence cannot be altered due to requirements of a specific application, the fusion of flanking spidroin terminal domain sequences (which could be subsequently removed through the inclusion of cleavage sequences) provides a tool to mitigate toxicity.

## 4. Conclusion

We have produced a panel of recombinant spidroin constructs in ten strains of *E. coli*, and results show that both a restriction on basal expression and genetic mutations on multiple genes related to stress responses facilitated increased titers and plasmid maintenance. These attributes were combined to create a novel strain of *E. coli* that produces recombinant spidroins at levels 4–33 times higher levels than standard BL21. Interestingly, this novel strain offered no benefit for the production of an ELP with structural and compositional similarities to the recombinant spidroins. It was found that expression of the spidroins, but not the ELP, exerted toxicity on the *E. coli* host system. To our knowledge, our results are the first to suggest that this toxicity, which may result from the intrinsic disorder of the spidroin product, is a primary factor causing low titers of recombinant spidroins. Counterintuitively, a bioprocess design that excludes the addition of inducer and uses longer culturing times was found to mitigate toxicity and increase spidroin titers in strain BL21 by 5–7 times. Targeted alteration of the spidroin primary sequence using cDNA of natural spidroin terminal domains was also found to mitigate toxicity and increase final OD<sub>600</sub> by 110%. Mutations on multiple genes related to stress response systems in *E. coli* such as *yggW*, *yedY*, *yedW*, *yedY*, *speC*, *speB*, *uspC*, *hchA*, *loiP*, *mltC*, *envZ*, *ompR*, *yhgF*, *hupB* were identified as unique attributes in the strain that produced the highest levels of spidroins and represent genes previously unknown to affect spidroin production. In conclusion, we have put forth a new hypothesis related to the persistent difficulty in producing high titers of certain recombinant spidroins, and we have outlined multiple points of focus for the future optimization of spidroin production that includes basal expression, promoter strength, induction strength, construct design, and specific genomic alterations.

## Author contributions

Alexander Connor: Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Visualization, Writing – Original draft preparation, revision and editing. Caleb Wigham: Formal Analysis,

Investigation for FTIR work. Yang Bai: Formal Analysis, Investigation, for genomic sequencing work. Manish Rai: Formal Analysis, for genomic sequencing work. Sebastian Nassif: Investigation, for growth curve data. Mattheos Koffas: Conceptualization, Resources, Funding Acquisition, Supervision, Writing – revision and editing. R. Helen Zha: Conceptualization, Resources, Funding Acquisition, Supervision, Writing – revision and editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

Data will be made available on request.

### Acknowledgements

This work received funding from NSF MCB#2036768. The authors would like to acknowledge Derek Nelson for supplying the A4Y1 ELP vector sequence and as well as Jasmine Yang for assistance with protein disorder predictions.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.mec.2023.e00219>.

### References

- 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., 2017. Biomimetic spinning of artificial spider silk from a chimeric minispidroin. *Nat. Chem. Biol.* 13 (3), 262–264. <https://doi.org/10.1038/nchembio.2269>.
- Bachmann, B.J., 1972. Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol. Rev.* 36 (4), 525–557. <https://doi.org/10.1128/br.36.4.525-557.1972>.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol. : a journal of computational molecular cell biology* 19 (5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Barbieri, C.M., Wu, T., Stock, A.M., 2013. Comprehensive analysis of OmpR phosphorylation, dimerization, and DNA binding supports a canonical model for activation. *J. Mol. Biol.* 425 (10), 1612–1626. <https://doi.org/10.1016/j.jmb.2013.02.003>.
- Bhattacharyya, G., Oliveira, P., Krishnaji, S.T., Chen, D., Hinman, M., Bell, B., Harris, T. I., Ghazitatabaei, A., Lewis, R.V., Jones, J.A., 2021. Large Scale Production of Synthetic Spider Silk Proteins in *Escherichia coli*, vol. 183. *Protein Expression and Purification*. <https://doi.org/10.1016/j.pep.2021.105839>.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bowen, C.H., Dai, B., Sargent, C.J., Bai, W., Ladiwala, P., Feng, H., Huang, W., Kaplan, D. L., Galazka, J.M., Zhang, F., 2018. Recombinant spidroins fully replicate primary mechanical properties of natural spider silk. *Biomacromolecules* 19 (9), 3853–3860. <https://doi.org/10.1021/acs.biomac.8b00980>.
- Bratzel, G., Buehler, M.J., 2012. Sequence-structure correlations in silk: poly-Ala repeat of *N. clavipes* MaSp1 is naturally optimized at a critical length scale. *J. Mech. Behav. Biomed. Mater.* 7, 30–40. <https://doi.org/10.1016/j.jmbm.2011.07.012>.
- Byrne, R.T., Chen, S.H., Wood, E.A., Cabot, E.L., Cox, M.M., 2014. *Escherichia coli* genes and pathways involved in surviving extreme exposure to ionizing radiation. *J. Bacteriol.* 196 (20) <https://doi.org/10.1128/JB.01589-14>.
- Cai, H., Chen, G., Yu, H., Tang, Y., Xiong, S., Qi, X., 2020. One-step heating strategy for efficient solubilization of recombinant spider silk protein from inclusion bodies. *BMC Biotechnol.* 20 (1) <https://doi.org/10.1186/s12896-020-00630-1>.
- Campen, A., Williams, R.M., Brown, C.J., Meng, J., Uversky, V.N., Dunker, A.K., 2008. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.* 15 (9), 956–963. <https://doi.org/10.2174/092986608785849164>.
- Chakraborty, S., Winardhi, R.S., Morgan, L.K., Yan, J., Kenney, L.J., 2017. Non-canonical activation of OmpR drives acid and osmotic stress responses in single bacterial cells. *Nat. Commun.* 8 (1) <https://doi.org/10.1038/s41467-017-02030-0>.
- Choi, U.B., McCann, J.J., Weninger, K.R., Bowen, M.E., 2011. Beyond the random coil: stochastic conformational switching in intrinsically disordered proteins. *Structure* (London, England : 1993) 19 (4), 566–576. <https://doi.org/10.1016/j.str.2011.01.011>.
- Chung, H., Kim, T.Y., Lee, S.Y., 2012. Recent advances in production of recombinant spider silk proteins. *Curr. Opin. Biotechnol.* 23 (6), 957–964. <https://doi.org/10.1016/j.copbio.2012.03.013>.
- Corchero, J.L., Villaverde, A., 1998. Plasmid maintenance in *Escherichia coli* recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol. Bioeng.* 58 (6), 625–632.
- Dumon-Seignover, L., Cariot, G., Vuillard, L., 2004. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* 37 (1), 203–206. <https://doi.org/10.1016/j.pep.2004.04.025>.
- Edlund, A.M., Jones, J., Lewis, R., Quinn, J.C., 2018. Economic feasibility and environmental impact of synthetic spider silk production from *Escherichia coli*. *New biotechnology* 42, 12–18. <https://doi.org/10.1016/j.nbt.2017.12.006>.
- Fahnestock, S.R., Irwin, S.L., 1997. Synthetic spider dragline silk proteins and their production in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 47 (1), 23–32. <https://doi.org/10.1007/s002530050883>.
- Fink, T.D., Zha, R.H., 2018. Silk and silk-like supramolecular materials. *Macromol. Rapid Commun.* 39, 1700834 <https://doi.org/10.1002/marc.201700834>.
- Fredriksson, C., Hedhammar, M., Feinstein, R., Nordling, K., Kratz, G., Johansson, J., Huss, F., Rising, A., 2009. Tissue response to subcutaneously implanted recombinant spider silk: an in vivo study. *Materials* 2 (4), 1908–1922. <https://doi.org/10.3390/ma2041908>.
- Gatesy, J., Hayashi, C., Motriuk, D., Woods, J., Lewis, R., 2001. Extreme diversity, conservation, and convergence of spider silk fibroin sequences. *Science* 291 (5513), 2603–2605. <https://doi.org/10.1126/science.1057561>.
- Gennaris, A., Ezratty, B., Henry, C., Agrebi, R., Vergnes, A., Oheix, E., Bos, J., Leverrier, P., Espinosa, L., Szewczyk, J., Vertommen, D., Iranzo, O., Collet, J.F., Barras, F., 2015. Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons. *Nature* 528 (7582), 409–412. <https://doi.org/10.1038/nature15764>.
- Goshima, N., Kohno, K., Imamoto, F., Kano, Y., 1990. HU-1 mutants of *Escherichia coli* deficient in DNA binding. *Gene* 96 (1), 141–145. [https://doi.org/10.1016/0378-1119\(90\)90355-u](https://doi.org/10.1016/0378-1119(90)90355-u).
- Gosline, J.M., Guerette, P.A., Ortlepp, C.S., Savage, K.N., 1999. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J. Exp. Biol.* 202 (23) <https://doi.org/10.1242/jeb.202.23.3295>.
- Gould, P., 2002. Exploiting spiders silk. *Mater. Today* 5 (12). [https://doi.org/10.1016/S1369-7021\(02\)01238-5](https://doi.org/10.1016/S1369-7021(02)01238-5).
- Guerrero, P.A., Ginzinger, D.G., Weber, B.H.F., Gosline, J.M., 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science* 272 (5258). <https://doi.org/10.1126/science.272.5258.112>.
- Guo, H., Xu, N., Prell, M., Königs, H., Hermanns-Sachweh, B., Lüscher, B., Kappes, F., 2021. Bacterial Growth Inhibition Screen (BGIS): harnessing recombinant protein toxicity for rapid and unbiased interrogation of protein function. *FEBS Lett.* 595, 1422–1437. <https://doi.org/10.1002/1873-3468.14072>.
- Gustavsson, N., Diez, A., Nyström, T., 2002. The universal stress protein paralogs of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage. *Mol. Microbiol.* 43 (1), 107–117. <https://doi.org/10.1046/j.1365-2958.2002.02720.x>.
- Hafner, E.W., Tabor, C.W., Tabor, H., 1979. Mutants of *Escherichia coli* that do not contain 1,4-diaminobutane (putrescine) or spermidine. *J. Biol. Chem.* 254 (24), 12419–12426.
- Hagiwara, D., Sugiura, M., Oshima, T., Mori, H., Aiba, H., Yamashino, T., Mizuno, T., 2003. Genome-wide analyses revealing a signaling network of the RcsC-YojN-RcsB phosphorelay system in *Escherichia coli*. *J. Bacteriol.* 185 (19), 5735–5746. <https://doi.org/10.1128/JB.185.19.5735-5746.2003>.
- Hardy, J.G., Römer, L.M., Scheibel, T.R., 2008. Polymeric materials based on silk proteins. *Polymer* 49 (20). <https://doi.org/10.1016/j.polymer.2008.08.006>.
- Hausjell, J., Weissensteiner, J., Molitor, C., Halbwirther, H., Spadiut, O., 2018. *E. coli* HMS174(DE3) is a sustainable alternative to BL21(DE3). *Microb. Cell Factories* 17 (1), 169. <https://doi.org/10.1186/s12934-018-1016-6>.
- Heidebrecht, A., Eisoldt, L., Diehl, J., Schmidt, A., Geffers, M., Lang, G., Scheibel, T., 2015. Biomimetic fibers made of recombinant spidroins with the same toughness as natural spider silk. *Adv. Mater.* 27, 2189–2194. <https://doi.org/10.1002/adma.201404234>.
- Heim, M., Keerl, D., Scheibel, T., 2009. Spider silk: from soluble protein to extraordinary fiber. *Angew. Chem., Int. Ed.* 48 (Issue 20) <https://doi.org/10.1002/anie.200803341>.
- Huang, Y., Dong, K., Zhang, X., Zhang, B., Hou, L., Chen, N., Chen, S., 2008. Expression and regulation of the yggG gene of *Escherichia coli*. *Curr. Microbiol.* 56 (1), 14–20. <https://doi.org/10.1007/s00284-007-9030-7>.
- Huang, Y.J., Acton, T.B., Montelione, G.T., 2014. DisMeta: a meta server for construct design and optimization. *Methods Mol. Biol.* 1091, 3–16. [https://doi.org/10.1007/978-1-62703-691-7\\_1](https://doi.org/10.1007/978-1-62703-691-7_1).
- Huemmerich, D., Helsen, C.W., Quedzuweit, S., Oschmann, J., Rudolph, R., Scheibel, T., 2004. Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry* 43 (42), 13604–13612. <https://doi.org/10.1021/bi048983q>.
- Humenik, M., Magdeburg, M., Scheibel, T., 2014. Influence of repeat numbers on self-assembly rates of repetitive recombinant spider silk proteins. *J. Struct. Biol.* 186 (3), 431–437. <https://doi.org/10.1016/j.jsb.2014.03.010>.

- Hwang, P.M., Pan, J.S., Sykes, B.D., 2012. A PagP fusion protein system for the expression of intrinsically disordered proteins in *Escherichia coli*. *Protein Expr. Purif.* 85 (1), 148–151. <https://doi.org/10.1016/j.pep.2012.07.007>.
- James, J., Yarnall, B., Koranteng, A., Gibson, J., Rahman, T., Doyle, D.A., 2021a. Protein over-expression in *Escherichia coli* triggers adaptation analogous to antimicrobial resistance. *Microb. Cell Factories* 20 (1), 13. <https://doi.org/10.1186/s12934-020-01462-6>.
- James, J., Yarnall, B., Koranteng, A., Gibson, J., Rahman, T., Doyle, D.A., 2021b. Protein over-expression in *Escherichia coli* triggers adaptation analogous to antimicrobial resistance. *Microb. Cell Factories* 20 (1), 13. <https://doi.org/10.1186/s12934-020-01462-6>.
- Karla, A., Lively, M.O., Paetzel, M., Dalbey, R., 2005. The identification of residues that control signal peptidase cleavage fidelity and substrate specificity. *J. Biol. Chem.* 280 (8), 6731–6741. <https://doi.org/10.1074/jbc.M413019200>.
- Kawai, S., Kawamoto, J., Ogawa, T., Kurihara, T., 2019. Development of a regulatable low-temperature protein expression system using the psychrotrophic bacterium, *Shewanella livingstonensis* Ac10, as the host. *Biosci. Biotechnol. Biochem.* 83 (11), 2153–2162. <https://doi.org/10.1080/09168451.2019.1638754>.
- Kong, B., Guo, G.L., 2014. Soluble expression of disulfide bond containing proteins FGF15 and FGF19 in the cytoplasm of *Escherichia coli*. *PLoS One* 9 (1). <https://doi.org/10.1371/journal.pone.0085890>.
- Kunert, K.J., Cresswell, C.F., Schmidt, A., Mullineaux, P.M., Foyer, C.H., 1990. Variations in the activity of glutathione reductase and the cellular glutathione content in relation to sensitivity to methylviologen in *Escherichia coli*. *Arch. Biochem. Biophys.* 282 (2), 233–238. [https://doi.org/10.1016/0003-9861\(90\)90110-k](https://doi.org/10.1016/0003-9861(90)90110-k).
- Kurihara, S., Oda, S., Kato, K., Kim, H.G., Koyanagi, T., Kumagai, H., Suzuki, H., 2005. A novel putrescine utilization pathway involves gamma-glutamylated intermediates of *Escherichia coli* K-12. *J. Biol. Chem.* 280 (6), 4602–4608. <https://doi.org/10.1074/jbc.M411114200>.
- Kwon, S.K., Kim, S.K., Lee, D.H., Kim, J.F., 2015. Comparative genomics and experimental evolution of *Escherichia coli* BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. *Sci. Rep.* 5, 16076. <https://doi.org/10.1038/srep16076>.
- Lammel, A.S., Hu, X., Park, S.H., Kaplan, D.L., Scheibel, T.R., 2010. Controlling silk fibroin particle features for drug delivery. *Biomaterials* 31 (16). <https://doi.org/10.1016/j.biomaterials.2010.02.024>.
- Leal-Egaña, A., Lang, G., Mauere, C., Wickinghoff, J., Weber, M., Geimer, S., Scheibel, T., 2012. Interactions of fibroblasts with different morphologies made of an engineered spider silk. *Protein. Adv. Eng. Mater.* 14, B67–B75. <https://doi.org/10.1002/adem.201180072>.
- Lewis, R.V., Hinman, M., Kothakota, S., Fournier, M.J., 1996. Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins. *Protein Expr. Purif.* 7 (4), 400–406. <https://doi.org/10.1006/prep.1996.0060>.
- Li, H., 2013. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. <https://doi.org/10.48550/arXiv.1303.3997> arXiv preprint arXiv:1303.3997.
- Liang, M., Pang, C.N.I., Li, S.S., Wilkins, M.R., 2010. Proteins deleterious on overexpression are associated with high intrinsic disorder, specific interaction domains, and low abundance. *J. Proteome Res.* 9 (3), 1218–1225. <https://doi.org/10.1021/pr900693e>.
- Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P., Berkmen, M., 2016. Erratum to: sHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Factories* 15 (1), 124. <https://doi.org/10.1186/s12934-016-0512-9>.
- Malay, A.D., Arakawa, K., Numata, K., 2017. Analysis of repetitive amino acid motifs reveals the essential features of spider dragline silk proteins. *PLoS One* 12 (8), e0183397. <https://doi.org/10.1371/journal.pone.0183397>.
- Malay, A.D., Suzuki, T., Katashima, T., Kono, N., Arakawa, K., Numata, K., 2020. Spider silk self-assembly via modular liquid-liquid phase separation and nanofibrillation. *Sci. Adv.* 6 (45), eabb6030. <https://doi.org/10.1126/sciadv.abb6030>.
- Marelli, B., Brenckle, M.A., Kaplan, D.L., Omenetto, F.G., 2016. Silk fibroin as edible coating for perishable food preservation. *Sci. Rep.* 6. <https://doi.org/10.1038/srep25263>.
- Marsh, J.A., Forman-Kay, J.D., 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* 98 (10), 2383–2390. <https://doi.org/10.1016/j.bpj.2010.02.006>.
- Maslowska, K.H., Makiela-Dzbenka, K., Fijałkowska, I.J., 2019. The SOS system: a complex and tightly regulated response to DNA damage. *Environ. Mol. Mutagen.* 60 (4), 368–384. <https://doi.org/10.1002/em.22267>.
- Mészáros, B., Erdos, G., Dosztányi, Z., 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* 46 (W1), W329–W337. <https://doi.org/10.1093/nar/gky384>.
- Mujacic, M., Baneyx, F., 2006. Regulation of *Escherichia coli* hchA, a stress-inducible gene encoding molecular chaperone Hsp31. *Mol. Microbiol.* 60 (6), 1576–1589. <https://doi.org/10.1111/j.1365-2958.2006.05207.x>.
- Mujacic, M., Baneyx, F., 2007. Chaperone Hsp31 contributes to acid resistance in stationary-phase *Escherichia coli*. *Appl. Environ. Microbiol.* 73 (3), 1014–1018. <https://doi.org/10.1128/AEM.02429-06>.
- Mujacic, M., Cooper, K.W., Baneyx, F., 1999. Cold-inducible cloning vectors for low-temperature protein expression in *Escherichia coli*: application to the production of a toxic and proteolytically sensitive fusion protein. *Gene* 238 (2), 325–332. [https://doi.org/10.1016/S0378-1119\(99\)00328-5](https://doi.org/10.1016/S0378-1119(99)00328-5).
- Mujacic, M., Bader, M.W., Baneyx, F., 2004. *Escherichia coli* Hsp31 functions as a holding chaperone that cooperates with the DnaK-DnaJ-GrpE system in the management of protein misfolding under severe stress conditions. *Mol. Microbiol.* 51 (3), 849–859. <https://doi.org/10.1046/j.1365-2958.2003.03871.x>.
- Mulinti, P., Diekjürgen, D., Kurtzeborn, K., Balasubramanian, N., Stafslin, S.J., Grainger, D.W., Brooks, A.E., 2022. Anti-coagulant and antimicrobial recombinant heparin-binding major ampullate spidroin 2 (MaSp2) silk protein. *Bioengineering* (Basel, Switzerland) 9 (2), 46. <https://doi.org/10.3390/bioengineering9020046>.
- Nishihara, K., Kanemori, M., Kitagawa, M., Yanagi, H., Yura, T., 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cry2, in *Escherichia coli*. *Appl. Environ. Microbiol.* 64 (5), 1694–1699. <https://doi.org/10.1128/AEM.64.5.1694-1699.1998>.
- Nonaka, G., Blankschien, M., Herman, C., Gross, C.A., Rhodius, V.A., 2006. Regulation and promoter analysis of the *E. coli* heat-shock factor, sigma32, reveals a multifaceted cellular response to heat stress. *Genes Dev.* 20 (13), 1776–1789. <https://doi.org/10.1101/gad.1428206>.
- Ondera, O., Roses, A.D., Tsuji, S., Vance, J.M., Strittmatter, W.J., Burke, J.R., 1996. Toxicity of expanded polyglutamine-domain proteins in *Escherichia coli*. *FEBS Lett.* 399 (1–2), 135–139. [https://doi.org/10.1016/S0014-5793\(96\)01301-4](https://doi.org/10.1016/S0014-5793(96)01301-4).
- Pedersen, C.P., Seiffert, P., Brakti, I., Bugge, K., 2020. Production of intrinsically disordered proteins for biophysical studies: tips and tricks. *Methods Mol. Biol.* 2141, 195–209. [https://doi.org/10.1007/978-1-0716-0524-0\\_9](https://doi.org/10.1007/978-1-0716-0524-0_9).
- Pomposiello, P.J., Koutsolioutsou, A., Carrasco, D., Demple, B., 2003. SoxRS-regulated expression and genetic analysis of the yggX gene of *Escherichia coli*. *J. Bacteriol.* 185 (22), 6624–6632. <https://doi.org/10.1128/JB.185.22.6624-6632.2003>.
- Prince, J.T., McGrath, K.P., DiGirolamo, C.M., Kaplan, D.L., 1995. Construction, cloning, and expression of synthetic genes encoding spider dragline silk. *Biochemistry* 34 (34), 10879–10885. <https://doi.org/10.1021/bi00034a022>.
- Römer, L., Scheibel, T., 2008. The elaborate structure of spider silk: structure and function of a natural high performance fiber. *Prion* 2 (4), 154–161. <https://doi.org/10.4161/pri.2.4.7490>.
- Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* 5, 172. <https://doi.org/10.3389/fmicb.2014.00172>.
- Saïda, F., Uzan, M., Odaert, B., Bontems, F., 2006. Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. *Curr. Protein Pept. Sci.* 7 (1), 47–56. <https://doi.org/10.2174/138920306775474095>.
- Sarkar, A., Connor, A.J., Koffas, M., Zha, R.H., 2019. Chemical synthesis of silk-mimetic polymers. *Materials* 12 (24), 4086. <https://doi.org/10.3390/ma12244086>.
- Scheibel, T., 2004. Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. *Microb. Cell Factories* 3 (1), 14. <https://doi.org/10.1186/1475-2859-3-14>.
- Schmuck, B., Greco, G., Barth, A., Pugno, N.M., Johansson, J., Rising, A., 2021. High-yield production of a super-soluble miniature spidroin for biomimetic high-performance materials. *Mater. Today* 50, 16–23. <https://doi.org/10.1016/J.MATTOD.2021.07.020>.
- Schneider, B.L., Hernandez, V.J., Reitzer, L., 2013. Putrescine catabolism is a metabolic response to several stresses in *Escherichia coli*. *Mol. Microbiol.* 88 (3), 537–550. <https://doi.org/10.1111/mmi.12207>.
- Schwarze, S., Zwettler, F., Johnson, C., et al., 2013. The N-terminal domains of spider silk proteins assemble ultrafast and protected from charge screening. *Nat. Commun.* 4, 2815. <https://doi.org/10.1038/ncomms3815>.
- Sieben, M., Steinhorn, G., Müller, C., Fuchs, S., Ann Chin, L., Regestein, L., Büchs, J., 2016. Testing plasmid stability of *Escherichia coli* using the continuously operated shaken BIOreactor system. *Biotechnol. Proc.* 32, 1418–1425. <https://doi.org/10.1002/btpr.2341>.
- SoluBL21 Competent *E. coli*. Genlantis, product manual. <https://lib.store.yahoo.net/lib/yhst-131428861332406/SoluBL21-Toxic-Clones.pdf>. (Accessed 22 September 2022). Accessed.
- Sørensen, H.P., Mortensen, K.K., 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Factories* 4, 1. <https://doi.org/10.1186/1475-2859-4-1>.
- Stojkova, P., Spidlova, P., Stulik, J., 2019. Nucleoid-associated protein HU: a lilliputian in gene regulation of bacterial virulence. *Front. Cell. Infect. Microbiol.* 9, 159. <https://doi.org/10.3389/fcimb.2019.00159>.
- Studier, F.W., Moffatt, B.A., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189 (1), 113–130. [https://doi.org/10.1016/0022-2836\(86\)90385-2](https://doi.org/10.1016/0022-2836(86)90385-2).
- Subedi, K.P., Choi, D., Kim, I., Min, B., Park, C., 2011. Hsp31 of *Escherichia coli* K-12 is glyoxalase III. *Mol. Microbiol.* 81 (4), 926–936. <https://doi.org/10.1111/j.1365-2958.2011.07736.x>.
- Tabor, C.W., Tabor, H., 1985. Polyamines in microorganisms. *Microbiol. Rev.* 49 (1), 81–99. <https://doi.org/10.1128/mr.49.1.81-99.1985>.
- Tan, F., Sun, N., Zhang, L., Wu, J., Xiao, S., Tan, Q., Uversky, V.N., Liu, Y., 2021. Functional characterization of an unknown soybean intrinsically disordered protein in vitro and in *Escherichia coli*. *Int. J. Biol. Macromol.* 166, 538–549. <https://doi.org/10.1016/j.ijbiomac.2020.10.211>.
- Teulé, F., Furin, W.A., Cooper, A.R., Duncan, J.R., Lewis, R., 2007. Modifications of spider silk sequences in an attempt to control the mechanical properties of the synthetic fibers. *J. Mater. Sci.* 42 (21). <https://doi.org/10.1007/s10853-007-1642-6>.
- Teulé, F., Cooper, A.R., Furin, W.A., Bittencourt, D., Rech, E.L., Brooks, A., Lewis, R.V., 2009. A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning. *Nat. Protoc.* 4 (3), 341–355. <https://doi.org/10.1038/nprot.2008.250>.
- Tokareva, O., Michalczychen-Lacerda, V.A., Rech, E.L., Kaplan, D.L., 2013. Recombinant DNA production of spider silk proteins. *Microb. Biotechnol.* 6 (6), 651–663. <https://doi.org/10.1111/1751-7915.12081>.

- Tokareva, O., Jacobsen, M., Buehler, M., Wong, J., Kaplan, D.L., 2014. Structure-function-property-design interplay in biopolymers: spider silk. *Acta Biomater.* 10 (4), 1612–1626. <https://doi.org/10.1016/j.actbio.2013.08.020>.
- Tokishita, S., Kojima, A., Mizuno, T., 1992. Transmembrane signal transduction and osmoregulation in *Escherichia coli*: functional importance of the transmembrane regions of membrane-located protein kinase, EnvZ. *J. Biochem.* 111 (6), 707–713. <https://doi.org/10.1093/oxfordjournals.jbchem.a123823>.
- Treusch, S., Lindquist, S., 2012. An intrinsically disordered yeast prion arrests the cell cycle by sequestering a spindle pole body component. *J. Cell Biol.* 197 (3), 369–379. <https://doi.org/10.1083/jcb.201108146>.
- Tsioris, K., Tilburey, G.E., Murphy, A.R., Domachuk, P., Kaplan, D.L., Omenetto, F.G., 2010. Functionalized-silk-based active optofluidic devices. *Adv. Funct. Mater.* 20, 1083–1089. <https://doi.org/10.1002/adfm.200902050>.
- Urano, H., Umezawa, Y., Yamamoto, K., Ishihama, A., Ogasawara, H., 2015. Cooperative regulation of the common target genes between H<sub>2</sub>O<sub>2</sub>-sensing YedVW and Cu<sup>2+</sup>-sensing CusSR in *Escherichia coli*. *Microbiology (Reading, England)* 161 (Pt 4), 729–738. <https://doi.org/10.1099/mic.0.000026>.
- Urano, H., Yoshida, M., Ogawa, A., Yamamoto, K., Ishihama, A., Ogasawara, H., 2017. Cross-regulation between two common ancestral response regulators, HprR and CusR, in *Escherichia coli*. *Microbiology (Reading, England)* 163 (2), 243–252. <https://doi.org/10.1099/mic.0.000410>.
- Varanko, A.K., Su, J.C., Chilkoti, A., 2020. Elastin-like polypeptides for biomedical applications. *Annu. Rev. Biomed. Eng.* 22, 343–369. <https://doi.org/10.1146/annurev-bioeng-092419-061127>.
- Vavouri, T., Semple, J.I., Garcia-Verdugo, R., Lehner, B., 2009. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell* 138 (1), 198–208. <https://doi.org/10.1016/j.cell.2009.04.029>.
- Weber, A., Kögl, S.A., Jung, K., 2006. Time-dependent proteome alterations under osmotic stress during aerobic and anaerobic growth in *Escherichia coli*. *J. Bacteriol.* 188 (20), 7165–7175. <https://doi.org/10.1128/JB.00508-06>.
- Wei, S.P., Qian, Z.G., Hu, C.F., Pan, F., Chen, M.T., Lee, S.Y., Xia, X.X., 2020. Formation and functionalization of membraneless compartments in *Escherichia coli*. *Nat. Chem. Biol.* 16 (10), 1143–1148. <https://doi.org/10.1038/s41589-020-0579-9>.
- Whittall, D.R., Baker, K.v., Breitling, R., Takano, E., 2021. Host systems for the production of recombinant spider silk. In: *Trends in Biotechnology*, vol. 39. Elsevier Ltd, pp. 560–573. <https://doi.org/10.1016/j.tibtech.2020.09.007>, 6.
- Wohlrab, S., Thamm, C., Scheibel, T., 2014. The power of recombinant spider silk proteins. *Biotechnology of silk*. In: *Biologically-Inspired Systems*, vol. 5. Springer, Dordrecht, pp. 179–201. [https://doi.org/10.1007/978-94-007-7119-2\\_10](https://doi.org/10.1007/978-94-007-7119-2_10).
- Xia, X.-X., Qian, Z.-G., Ki, C.S., Park, Y.H., Kaplan, D.L., Lee, S.Y., 2010. In: *Native-sized Recombinant Spider Silk Protein Produced in Metabolically Engineered Escherichia coli Results in a Strong Fiber*, vol. 107, pp. 14059–14063. <https://doi.org/10.1073/pnas.1003366107>, 32.
- Yang, H., Yang, S., Kong, J., Dong, A., Yu, S., 2015. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat. Protoc.* 10 (3), 382–396. <https://doi.org/10.1038/nprot.2015.024>.
- Yang, Y.X., Qian, Z.G., Zhong, J.J., Xia, X.X., 2016. Hyper-production of large proteins of spider dragline silk MaSp2 by *Escherichia coli* via synthetic biology approach. *Process Biochem.* 51 (4), 484–490. <https://doi.org/10.1016/j.procbio.2016.01.006>.
- Yedahalli, S.S., Rehmann, L., Bassi, A., 2016. Expression of exo-inulinase gene from *Aspergillus Niger* 12 in *E. coli* strain Rosetta-gami B (DE3) and its characterization. *Biotechnol. Prog.* 32, 629–637. <https://doi.org/10.1002/btpr.2238>.
- Zhao, S., Ye, X., Wu, M., Ruan, J., Wang, X., Tang, X., Zhong, B., 2021. Recombinant silk proteins with additional polyalanine have excellent mechanical properties. *Int. J. Mol. Sci.* 22 (4), 1513. <https://doi.org/10.3390/ijms22041513>.