Standard Intein Gene Expression Ramps (SIGER) for protein-independent expression control

Authors

Maxime Fages-Lartaud[1], Yasmin Mueller[1], Florence Elie[1], Gaston Coutarde[1], Martin Frank Hohmann-Marriott[1,2]

Affiliation

[1]Department of Biotechnology and Food Science, Norwegian University of Science and Technology, Trondheim, N-7491, Norway.

[2]United Scientists CORE (Limited), Dunedin 9016, Aotearoa - New Zealand.

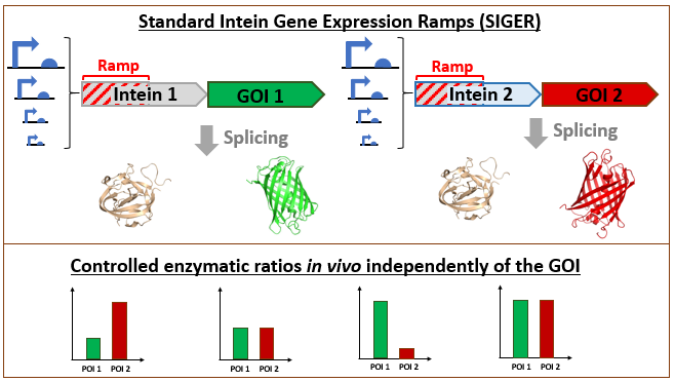
Contact (corresponding author)

Email: martin.hohmann-marriott@ntnu.no

Abstract

Coordination of multi-gene expression is one of the key challenges of metabolic engineering for the development of cell factories. Constraints on translation initiation and early ribosome kinetics of mRNA are imposed by features at the start of the gene, referred to as the "gene ramp", such as rare codons and mRNA secondary structures forming in relation with the 5'UTR. These features strongly influence translation yield and protein quality by regulating ribosome distribution on mRNA strands. The utilization of genetic expression sequences, such as promoters and 5'UTRs in combination with different target genes leads to a wide variety of gene ramp compositions with irregular translation rates leading to unpredictable levels of protein yield and quality. Here, we present the Standard Intein Gene Expression Ramps (SIGER) system for controlling protein expression. The SIGER system uses inteins to uncouple a characterized gene ramp from a target protein. We generated sequence-specific gene expression sequences for two inteins (DnaB and DnaX) that display defined levels of protein expression. Additionally, we used inteins that possess the ability to release the C-terminal fusion protein *in vivo* to avoid impairment of protein functionality by the fused intein. Overall, our results show that SIGER systems are unique tools to mitigate the undesirable effects of gene ramp variation and to control the relative ratios of enzymes involved in molecular pathways.

Graphical abstract



Introduction

Cell factories are central components of biotechnology for the production of recombinant proteins and biochemicals that find numerous applications in pharmaceutical, agricultural, food, cosmetic and chemical industries[1–3]. The choice of the cellular host depends on the application and is crucial to obtain high product yield[4] and desired protein properties such as solubility, secretion, glycosylation and other post-translational modifications[5]. The most common cell factories include various bacteria[6,7], yeast[8,9], filamentous fungi, plants[10], as well as insect and mammalian cells[11,12]. The ability of cell factories to produce quality protein in high yields is determined by the type of genetic expression system, the characteristics of the strain and its adaptability to large-scale cultivation processes[13]. For a given host, the choice of genetic expression system is fundamental because it determines the maximum yield, affects protein quality, and protein expression characteristics. For the production of biomolecules, the expression of enzymes involved in a metabolic pathway must be tuned to balance the relative enzymatic activities, thus avoiding burdensome overexpression of proteins and toxicity of intermediate metabolites[14–18]. Unsuitable genetic expression systems can be the source of substantial cell toxicity and protein misfolding leading to the failure of biotechnological endeavors.

Current gene expression systems comprise constitutive and inducible promoters[6,19] coupled with native or computationally designed 5'UTRs[20–22]. Single genes are expressed with a monocistronic cassette and multiple genes are expressed with series of monocistronic or polycistronic constructs. For example, the monocistronic pET system (DE3/T7), which is one of the predominant protein expression systems in *E. coli*[23], consists of an IPTG-inducible T7 polymerase gene integrated into the *E. coli* genome and a pET vector containing a T7 promoter expressing the gene of interest (GOI). However, genetic expression systems are not always successful in expressing certain GOIs or ensuring the functionality of the protein of interest (POI)[6]. The main causes of failure are related to defects in translation, protein folding, protein translocation, mRNA stability, plasmid sustainability and cell viability[6]. The aforementioned transcriptional, translational and protein maturation problems often originate from the genetic expression system and the lack of complementarity between genetic elements or adaptation to the host organism.

Mechanisms governing translation initiation involve mRNA stability, mRNA unfolding and ribosome entry-sites accessibility[24–27] (*e.g.* Shine-Dalgarno[28] (SD) or Kozak[29] sequence). Translation efficiency is highly dependent on the mRNA secondary structures formed within the 5'UTR and between the 5'UTR and the start of the coding sequence (CDS). The region of the CDS that influences translation initiation is known as the gene "ramp" and comprises the first 100 to 150 nucleotides[30,31]. The gene ramp influences ribosomal entry onto the mRNA and modulates early translation rates (*Figure 1a*). The limitation in early translation rates at 5' genes termini was hypothesized to allow ribosomal spacing, therefore avoiding traffic jams, ribosome fall-offs and aborted translation[31–33]. The nucleotide composition of the ramp modulates the formation of mRNA secondary structures that influence translation initiation kinetics[34]. Additionally, the gene ramp is enriched in rare codon clusters that decrease early translation rates[30,35–37] and are involved in determining local mRNA secondary structures[30,38]. Positively charged amino acids, such as lysine, which are also overrepresented in ramp codons, interact with negatively charged residues in the ribosomal exit tunnel decreasing translation rates[39,40]. These characteristics of 5' gene termini properties are linked to conserved evolutionary mechanisms that are shared across many microorganisms[31,37].

The complementarity between the 5'UTR and the gene ramp influences translation yield and protein folding[41–43]. For an established genetic expression system, replacing the GOI changes the resulting secondary structure occurring between the 5'UTR and the ramp (*Figure 1b*). These structural variations affect the outcome of protein production. There are computational tools to help predict 5'UTRs sequences suitable for a given coding sequence[20–22]. However, these tools only provide predictions for individual cases, which have to be experimentally verified, whereas standardized *a priori* experimentally-validated systems would be preferable. In addition, codon optimization of GOIs can affect the composition of rare codons within the ramp[37], leading to perturbations in translation rates that produce insoluble or misfolded proteins[41–43].

The principal solution deployed to circumvent incompatibility between genetic elements and rescue protein solubility is the use of N-terminal fusion protein tags. These tags increase the POI's solubility, provide a compatible buffer sequence with genetic expression parts, and some tags can be used for

downstream purification process[44,45]. Most solubility tags require the use of specific proteases to liberate the protein of interest (POI) from the fusion construct[46]. One exception is the family of 2A self-cleaving linker peptides[47] that display the ability to excise themselves from a fusion protein, releasing the POI from the fusion tag. Inteins are another type of self-cleaving proteins that are used in protein purification[48,49] (*e.g.* the IMPACT system from New England Biolabs). Inteins are naturally occurring autocatalytic proteins that possess the ability to excise themselves from a larger protein, ligating the two flanking proteins (exteins) together in the splicing process[51]. The protein splicing process is spontaneous, occurs post-translationally and does not require the intervention of exogenous factors or proteases[51]. In the case of the IMPACT system, the *Sce* VMA intein is combined with a chitin-binding domain to bind the appropriate resin, and exhibit N-terminal cleavage in the presence of DTT or β-Mercaptoethanol[50]. Split inteins are another type of inteins, composed of an N-intein and C-intein, each fused to an extein, individually translated. After, translation, the N- and C-intein fragments assemble non-covalently to provide the canonical intein structure and carry out *trans* protein splicing[51,52]. In addition, inteins have been engineered to selectively release the N- or C-terminal peptide by mutating key catalytic residues. These unique properties enable a wide variety of applications such as enzyme activation, protein ligation, production of cyclic peptides, protein purification systems, biosensors and reporter systems[51,53–55]. One example of such an application is the simultaneous production of two equimolar POIs by using a dual-intein system composed of the respective N- and C-terminal cleavage properties of *Ssp DnaE* and *Ssp DnaB*[56].

Here, we used C-terminal cleaving inteins to design Standard Intein Gene Expression Ramps (SIGERs). Mini-inteins are short genes that can provide the necessary properties of a gene ramp while releasing the POI *in vivo*. We used the geneEE method[57] to create artificial promoter/5'UTRs that interact with the intein sequences in order to obtain a wide variety of gene ramps displaying a broad range of protein expression levels. This way, our SIGER systems fulfill the genetic complementarity requirements of a gene ramp, thus ameliorating protein folding and solubility, as well as offering the possibility to adapt expression levels to avoid cellular burden. Since translation initiation imposes major constraints on protein yield and protein quality, SIGER systems offer a buffer genetic region that allows exchanging

GOIs without affecting the level of protein expression. Furthermore, we show that coupling different SIGER systems in the expression host permits facile control of enzymatic ratios to balance metabolic pathways. The standard intein systems tightly controls the production of discrete enzymes *in vivo* in the desired quantity, without concerns for complementarity between genetic expression elements. Overall, we developed two SIGER systems and explored their ability to fulfill ramp properties and control multi-enzyme expression levels.

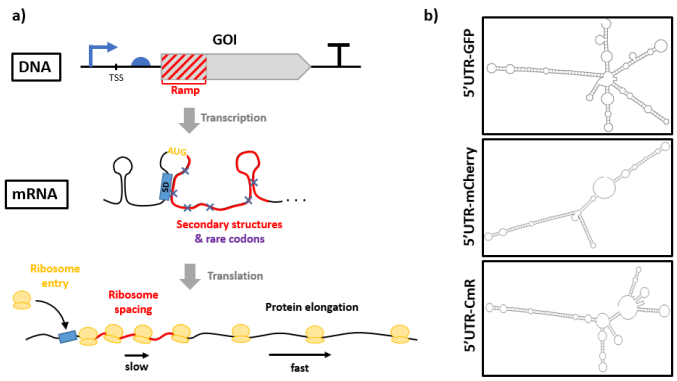


Figure 1. Characteristics of a gene ramp and implications for protein expression. a) Representation of a genetic cassette. The gene ramp corresponds to the first 100 to 150 nucleotides of the gene downstream a 5'UTR. The gene ramp forms secondary structures with the 5'UTR and within itself that imposes constraints on translation initiation. Additionally, rare codons with slow ribosome decoding are over-represented in the gene ramp. These characteristics affect translation initiation and slow early translation rates. The precise control of ribosome kinetics during translation initiation influences translation yield and protein quality. b) Example of the variation in secondary structures formed between an identical 5'UTR and three different gene ramps (*sfGFP*, *mCherry* and *cat*) predicted using the web-based RNAfold tool (http://rna.tbi.univie.ac.at/). The exchange of GOI downstream the 5'UTRs dramatically affects the resulting mRNA secondary structure. The mRNA secondary structures influence the efficiency of translation initiation and protein yield.

Results and Discussion

Design of the standard intein gene expression ramp (SIGER)

The principle of a standard gene expression ramp is to buffer the complementarity effect of the gene expression sequences (GES), composed of a promoter + 5'UTR, with the gene of interest. We selected two C-terminal cleaving mini-inteins, *DnaB*[58] and *DnaX*[59], of 159 and 140 amino acids, respectively, to serve as gene expression ramps (for sequences see *Supplementary Table S2*). The length of the intein genes is around 450 bp, which is long enough to absorb the properties of a ramp that affects usually the first 150 bp after the start codon (*Figure 2a*). The POI is released from the intein by autocatalytic cyclization of the C-terminal Asn residue (*Figure 2b*), N159 and N140 for DnaB and DnaX respectively. The efficiency of the reaction is dependent on the residue adjacent to the C-terminal asparagine. Ser-Arg residues at the C-terminal end of inteins have been experimentally demonstrated to result in high C-terminal self-cleavage activity[60]. Therefore, a C-extein amino acid linker, SRGP[56], was added to the C-terminal end of DnaB and DnaX. In this way, the intein gene fragment should function as the *in vivo* gene ramp and the resulting intein translation product can release the C-terminal fusion protein of interest (POI). Using SIGER, a discrete POI is produced, and thus avoiding potential defects in functionality due to the steric hindrance by the N-terminal fusion protein.

We first aimed to create artificial GES tailored to the *DnaB* and *DnaX* sequences, which possess a wide range of expression. For this, we used the gene expression engineering method (GeneEE) that consists in placing a DNA segment of 200 random nucleotide (200N) directly upstream a protein-coding sequence[57]. GeneEE provides a functional GES for 30 to 40% of sequences, a wide range of expression and gene-tailored 5'UTRs in *E. coli*[57]. The application of GeneEE created a multitude of GES tailored to the *DnaB* and *DnaX* sequences that displayed different levels of expression (*Figure 2a*).

The combined properties of inteins and the GeneEE method can provide interesting gene expression tools that define expression levels independently of the GOI sequence. The standard intein gene expression ramps (SIGER) circumvent issues related to the complementarity of gene ramps and GOI, as well as release discrete proteins. The GeneEE method can provide intein-tailored GES candidates

with fine-tuned protein expression characteristics that limit cell toxicity due to burdensome protein expression. Finally, the combination of characterized SIGER systems can permit to define the level of expression of different POIs to balance metabolic pathways.

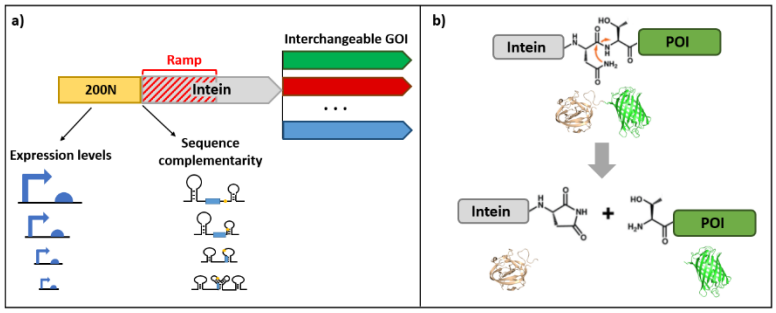


Figure 2. Characteristics of the standard intein gene expression ramp (SIGER) systems. a) Genetic organization of SIGER systems. The inteins fulfill the role of standard gene ramps and the 200N random DNA fragment provides intein-tailored gene expression sequences (GES). The nucleotide sequence of GES 5'UTRs produces different mRNA secondary structures in relation to the sequences of the inteins and each GES displays various expression levels. b) Autocatalytic C-terminal cleavage of inteins. The cyclization of the N-terminal Asn residue releases the POI *in vivo*.

Fabrication of intein-tailored promoter library

We used the GeneEE method to generate a library of GES adapted to the sequence of *DnaB* and *DnaX*. To do so, we inserted the 200N DNA fragment in front of *DnaB-GFP* and *DnaX-GFP* by Golden Gate assembly and transformed it into *E. coli* (see *Methods* and *Supplementary Figure S1*). Positive clones (n=186) displaying seemingly green fluorescence under UV light were selected and grown overnight in 96 well plates. The fluorescence measurement of the 186 positive clones for each *DnaB-GFP* and *DnaX-GFP* are presented in *Figure 3*. The artificial GES resulted in a wide range of expression levels spanning over one order of magnitude. For each intein, 10 strains were selected to represent characteristic levels of expression. After eliminating strains showing inconsistencies, such as strains with large variability in GFP fluorescence levels, we obtained seven and eight gene-tailored GES for *DnaB* and *DnaX*, respectively (GES were named B_P1 to B_P7 for *DnaB* and X_P1 to X_P8 for *DnaX*) (*Figure 4*). All the GES identified with *DnaB-GFP* and *DnaX-GFP* were sequenced (*Supplementary*

Table S3). Using the complimentary GES and intein ramps, we investigated the effect of exchanging the GOI on the gene expression levels.

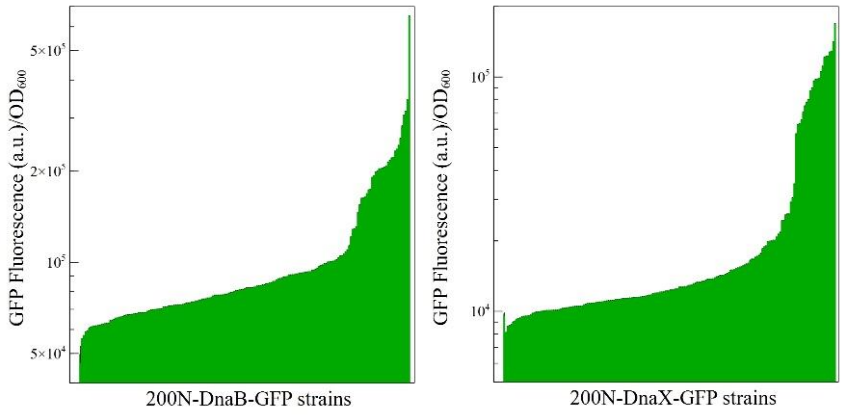


Figure 3. Fluorescence intensity of the GES library tailored for *DnaB* and *DnaX*. The histograms represent single fluorescence measurements for the GES library expressing *DnaB-GFP* and *DnaX-GFP*. For each intein, 186 strain expressing various levels of GFP were analyzed. The first histogram bar (position 1) represents the average auto-fluorescence of a negative control performed in triplicates and standard deviation is shown in black (calculated for this sample only).

Gene expression intensity is mostly independent of the GOI

Ten strains for each inteins were selected to represent distinct thresholds values of GFP protein expression (*Figure 3*). From these 10 strains, the levels of gene expression for seven (*DnaB*) and eight (*DnaX*) strains were defined as conserved because they showed consistent fluorescence levels when the measurements were performed in triplicates (*Figure 4*). In order to evaluate the genetic buffering effect of SIGER systems, we expressed *mCherry* and the chloramphenicol resistance gene *cat*. The GOIs *mCherry* and *cat* were inserted downstream both inteins (see *Methods* and *Supplementary Figure S1*), to create *DnaB-mCherry, DnaB-cat, DnaX-mCherry* and *DnaX-cat*. Then, we extracted the seven and eight GES identified with *DnaB-GFP* and *DnaX-GFP* respectively to and placed them in front of the new gene cassettes (see *Methods* and *Supplementary Figure S1*). Each construct was transformed into *E. coli* and positive clones were grown overnight in triplicates in 96 well plates. The fluorescence of clones containing *DnaB-mCherry* and *DnaX-mCherry* was quantified (*Figure 4*). The strains containing *DnaB-cat* and *DnaX-cat* were diluted 100 times and replica-plated on LB-agar plates supplemented with increasing chloramphenicol concentrations to evaluate their resistance profile (*Figure 4* and

Supplementary Figure S2). The fluorescence of strains carrying weak GES was confirmed by fluorescence microscopy (*Supplementary Figure S3*).

For *DnaB-mCherry*, a first observation is that fluorescence levels are generally higher than expected from the *DnaB-GFP* range although the relative strength ranking of GES remained the same (*Figure 4a*). This slight *mCherry* overexpression could be due to the newly identified alternative translation start site of *mCherry* that produces a short functional protein isoform[61]. The short mCherry isoform produces significant background fluorescence when the reporter is used as a C-terminal fusion partner. Although the background fluorescence is supposedly the same across strains, the production of the short mCherry isoform blurs the relative differences in the expression range.

The chloramphenicol gradient also confirms the relative strength of each *DnaB*-specific GES (*Figure 4a*), with the exception of B_P2 that presents a high chloramphenicol resistance (400 µg/mL) when the expectations would place it closer to B_P1 (250 µg/mL) in respect to fluorescence measurements of GFP and mCherry. Sanger sequencing of B_P2 from *DnaB-cat* revealed that the GES was a different sequence than the GES in B_P2 from *DnaB-GFP* (*Supplementary Table S3*). This discrepancy can be explained in the following way, when using the GeneEE method, after cloning of the 200N DNA fragment, several plasmids can be transformed into one *E. coli* cell. The different GES can be sustained in one *E. coli* cell due to the relatively small nucleotide variation of the plasmid (only in the 200N region). After prolonged growth, different cell populations can emerge bearing only one of the plasmid isoforms. Here, the PCR amplification of the GES regions of *DnaB-GFP* may have extracted different GES copies. The subsequent cloning onto new gene cassettes can yield strains with an unexpected sequence. This effect yielded two different *DnaB-P2-cat* strains with different representation in the population. The selection on chloramphenicol favored the underrepresented strain that possessed a stronger promoter.

For *DnaX-mCherry*, the expression range is conserved with respect to *DnaX-GFP* except for X_P4 and X_P6 (*Figure 4b*). As for *DnaB-P2-cat*, Sanger sequencing of *DnaX-P4-mCherry* revealed a different DNA sequence than *DnaX-P4-GFP*, thus explaining the lower fluorescence than expected. However, for *DnaX-P6-mCherry*, DNA sequencing provided the same sequence as *DnaX-P4-GFP*. The

fluorescence defect of *DnaX-P6-mCherry* may originate from another random mutation on the plasmid. The chloramphenicol resistance profile of *DnaX-cat* strains confirms the expression range obtained with *DnaX-GFP*, although it was not possible to differentiate between expression levels for X_P6, X_P7 and X_P8 because of insufficient method sensitivity (*Figure 4b*).

Overall, the expression range of the SIGER systems established with *GFP* were conserved when the GOI was exchanged with *mCherry* or *cat*. Apart from the sustained multiple GES issue, data shows that the different GES coupled to each intein provide standard gene expression cassettes with defined, consistent and predictable translation levels. We then investigated the combination of SIGER systems to tune multi-enzyme expression.

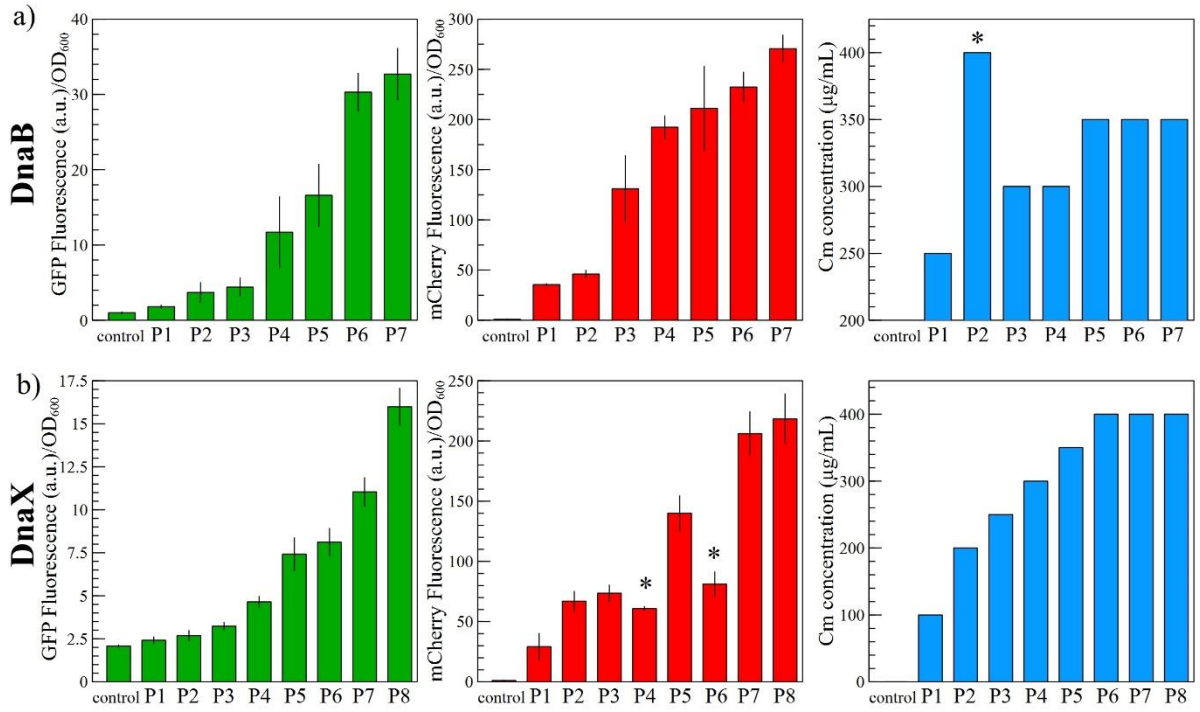


Figure 4. Expression of different GOIs with the DnaB and DnaX SIGER systems. The GES specific for *DnaB* and *DnaX* were transferred to two other gene cassettes to express *mCherry* and *cat*. The fluorescence measurements of GFP (green) and mCherry (red), and the chloramphenicol resistance profile of each strain (blue) are presented for the a) *DnaB* and b) *DnaX* SIGER systems. The expression range selected with GFP is generally conserved when a different GOI is placed after the inteins. The reasons for the observed exceptions (indicated with an asterisk) are discussed in the main text.

Coupling SIGER systems for balancing multi-gene expression in vivo

The developed SIGER systems are unique tools that enable fine-tuning of the expression of different discrete enzymes in the desired quantity *in vivo*. In order to test the accuracy of SIGER systems in controlling multi-enzyme expression, we coupled the *DnaB-mCherry* and *DnaX-GFP* cassettes on the same plasmid. The two intein cassettes were coupled on a level 2 vector (Lv2), containing a chloramphenicol resistance marker, by using the level assembly method developed by Fages-Lartaud *et al.* for pathway assembly[62] (see *Methods* and *Supplementary Figure S1*). In brief, the *DnaB* and *DnaX* gene cassettes were constructed on level 1 plasmid (Lv1), bearing an ampicillin marker, and containing BbsI restriction sites on each side of the gene cassettes. Then, the complementarity of the BbsI scars was used to assemble each cassette on the Lv2 vector. As a demonstration, we selected three GES for each intein that represent distinct levels of expression. For *DnaB-mCherry*, we used B_P2 as a weak GES, B_P3 as an intermediate GES and B_P7 as a strong GES. For *DnaX-GFP*, we used X_P2 as a weak GES, X_P5 as an intermediate GES and X_P7 as a strong GES. Each *DnaB* cassette was coupled to the three *DnaX* cassettes, resulting in nine genetic combinations. Each coupled construct was transformed into *E. coli* and cells were selected on LB-agar plates supplemented with 30 µg/mL chloramphenicol. The *E. coli* strains carrying each of the nine constructs were grown overnight in triplicates in 96 well plates and the fluorescence of mCherry and GFP was measured for each strain. The results are presented in *Figure 5*.

The first observation is that the expression intensities of *DnaB-mCherry* and *DnaX-GFP* were conserved when compared to single cassette measurements (see *Figure 4*), with the exception of B_P3 in the coupled system possessing B_P3 and X_P5. DNA sequencing revealed that B_P3 of that specific construct contained an unexpected sequence, thus explaining the discrepancy in fluorescence intensity. Beside this issue, the *DnaB* SIGER displayed three distinct levels of expression from low expression to 8.2 and 13.8-fold increase (*Figure 5)*. The *DnaX* SIGER also showed three expression levels with a 1.6 and 3.7-fold increase relatively to the weak GES (*Figure 5)*. The fluorescence of weak GES and the cellular co-localization of *mCherry* and *GFP* was confirmed by fluorescence microscopy (*Supplementary Figure S3*).

Secondly, results show that coupling SIGER systems enables control of the level of expression of two discrete proteins. The nine GES combinations of the *DnaB* and *DnaX* SIGER systems present various levels of relative enzymatic expression. Based on these results, we show that conceivably SIGER systems may be applied to control the expression of enzymes in a metabolic pathway to limit cellular burden or toxicity without the need to engineer specific GES for each GOI.

The last verification regarding the functionality of SIGER systems was to confirm the ability of the C-terminal cleaving inteins to release the POI. Indeed, one of the objectives of the SIGER systems is to produce *in vivo* discrete POIs without interference from an N-terminal fusion partner. Therefore, we investigated the efficacy of C-terminal cleavage by *DnaB* and *DnaX in vivo*.

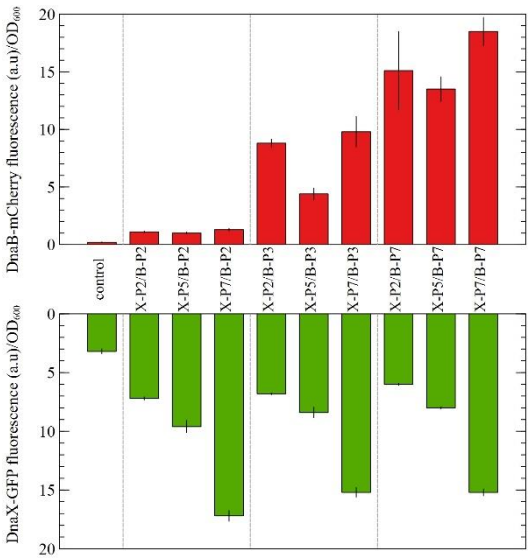


Figure 5. Controlled enzymatic expression by coupling two SIGER systems. The *DnaB-mCherry* and *DnaX-GFP* gene cassettes, each expressing at three different levels, were coupled on the same plasmid to assess the ability of SIGER systems to control multi-enzymatic expression. The histograms represent the fluorescence levels of the weak, intermediate and strong GES (respectively B_P2, B_P3 and B_P7 for DnaB; and X_P2, X_P5 and X_P7 for DnaX). The SIGER systems present very good control over the protein expression of any POI *in vivo*.

Assessment of in vivo cleavage

In order to assess the efficiency of intein-mediated *in vivo* C-terminal cleavage, we created two mutants of *DnaB* and *DnaX* possessing an N-terminal truncation (see *Methods*). The truncation renders the intein cleavage non-functional by preventing the N-terminal asparagine cyclization and the release of the POI.

DnaB-GFP and *DnaX-GFP*, expressed by a strong GES, and their mutated versions were transferred into *E. coli* BL21 for protein production at 30°C in canonical Erlenmeyer flasks. Cell cultures were harvested by centrifugation, cell content was released by sonication and cell debris were eliminated by centrifugation. Proteins present in the soluble fractions were identified by SDS-PAGE followed by Western Blot to detect the 6xHis tag of GFP. Furthermore, soluble fractions were chromatographically purified using a HisTrap HP Ni-sepharose column and evaluated on SDS-PAGE prior to LC-MS analysis (*Supplementary Figure S4)*.

The Western Blot results are presented in *Figure 6*. For *DnaB-GFP*, a clear GFP band is observable at 28 kDa and a faint band is noticeable around 45 kDa. This result shows that DnaB in vivo cleavage is efficient and that the *DnaB*-SIGER system releases the POI. The 45-kDa protein corresponding to the uncleaved DnaB-GFP fusion protein was analyzed by LC-MS after protein purification and SDS-PAGE. The analysis confirms that the 45-kDa protein is the uncleaved DnaB-GFP fusion protein (see *Supplementary Figure S5*).

The mutated version of *DnaB-GFP* displays a clear 43-kDa band corresponding to the uncleaved fusion protein, but also an unexpected 28-kDa protein around the size of GFP. This last-mentioned protein might be the result of an alternative translation start site occurring just upstream the *GFP* sequence. The mutated DnaB-GFP protein does not contain internal methionine residues that could result in the production of a GFP isoform. However, there are a several valine residues, such as V136 and V144, that could serve as alternative start codons, especially given the presence of an SD sequence (GGAG), six nucleotides upstream of V136. LC-MS analysis of the purified unexpected 28-kDa protein confirmed that it is GFP with a small N-terminal peptide elongation (at least DLTVPGPR) (see *Supplementary Figure S5*). This result support the hypothesis of an alternative start codon close to the gene start producing a GFP isoform. Nevertheless, this artifact does not interfere with the conclusion regarding the *in vivo* cleavage performance of DnaB-GFP.

The *DnaX-GFP* immunoblotting presents several bands, a strong 28-kDa band of the discrete GFP protein, a lighter 43-kDa band related to the uncleaved DnaX-GFP protein, and a couple intermediate bands (*Figure 6*). The relatively strong intermediate band may originate from an internal alternative

translation start site created by the methionine residue M77 of DnaX that results in a 35-kDa protein. The intermediate isoforms were not present on the SDS-PAGE performed after purification (*Supplementary Figure S4*). Therefore, a differential protein migration could also be the cause of the intermediate bands appearing on the Western Blot. The Western Blot and the SDS-PAGE performed with purified fractions both suggest an incomplete cleavage of the DnaX-GFP fusion protein. The reason for the low cleavage efficiency may be either low cleavage kinetics or suboptimal protein production conditions for full cleavage of DnaX-GFP. The 43-kDa protein was confirmed to correspond to the uncleaved DnaX-GFP fusion protein by LC-MS analysis (see *Supplementary Figure S5*).

The truncation of *DnaX* introduces a stop codon at the start of the *GFP* gene by creating a frameshift. Unfortunately, an in-frame GFP isoform can still be produced from the mutated *DnaX-GFP* mRNA due to the presence of an unexpected methionine located nine amino acids before the start of the *GFP* gene. In addition, an SD-like sequence, GGAA, is present seven nucleotide before the mentioned methionine. The combination of these elements resulted in the production of a GFP isoform possessing a nine amino acid N-terminal extension (*Figure 6*).

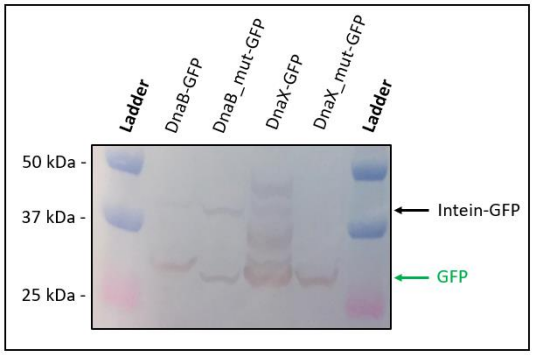


Figure 6. Western blot of the soluble protein fractions of the *DnaB* and *DnaX* SIGER systems and their truncated versions. The molecular weight of the bands in the ladder are indicated on the left side and sample names are indicated on top. The uncleaved Intein-GFP protein and the released free GFP are indicated with black and green arrows respectively. The DnaB-GFP protein displays efficient dissociation. The truncated version, DnaB_mut-GFP, shows the expected size of an uncleaved fusion protein. The DnaX-SIGER system presents several bands including uncleaved fusion protein, free GFP and intermediate bands. The origin of these intermediate bands is discussed in the text, but they were absent from the SDS-PAGE (*Supplementary Figure S4*). Regardless, the DnaX-GFP protein shows only

partial *in vivo* cleavage. The GFP bands present for the mutated DnaB and DnaX versions originates from alternative translation start sites that were confirmed by LC-MS analysis.

The analysis of intein C-terminal cleavage revealed a nearly complete *in vivo* cleavage of DnaB and GFP and an incomplete excision of DnaX from the fusion construct. Intein autocatalytic cleavage is highly dependent on conditions such as pH and temperature. Therefore, it is possible to optimize the release of the POI *in vivo* by controlling the conditions during protein expression. This may help in releasing the POI from the DnaX intein. Nevertheless, the two SIGER systems displayed the intended *in vivo* C-terminal cleavage despite the variable efficiency.

Conclusion

In this study, we used the self-cleavage properties of inteins to create gene expression systems that allow standardization of expression for different proteins. With the GeneEE method, we generated libraries of GES tailored to *DnaB* and *DnaX* sequences. A subset of GES was selected to represent distinct expression levels. These GES coupled to *DnaB* and *DnaX* constitute SIGER systems that control protein expression levels independently of the GOI. We demonstrate that SIGER systems can be used to balance the expression of multiple enzymes in vivo. The main advantage of using inteins in the SIGER system is that the POI is released from the fusion construct *in vivo* , precluding interference by fused protein tags on enzyme activity. Moreover, the intein gene ramp ensures optimal complementarity of genetic expression sequences with GOIs. The *DnaB* SIGER system showed efficient *in vivo* cleavage, while the *DnaX* SIGER system presented incomplete cleavage. Inteins cleavage occurs spontaneously and is not host-specific. Therefore, once established, SIGER systems may be applied across a wide range of organisms. The methodology presented here can be applied to other C-terminal cleaving inteins to create new SIGER systems that may be superior to the *DnaB* and *DnaX* SIGER used in this study. SIGER systems represent a new synthetic biology tool that will open new doors for applications in the expression of proteins and the establishment of metabolic pathways.

Acknowledgments.

This study was supported by a PhD fellowship awarded to M.F-L by the Faculty of Natural Sciences of the Norwegian University of Science and Technology. This work was financed by the Novo Nordisk Foundation (Grant NNF18OC0032242). We thank our partners at Vectron Biosolutions AS for technical assistance.

Conflict of interest.

The authors declare no conflict of interests.

Material and Methods

Materials

Escherichia coli DH5-α (New England Biolabs) was used as the cloning and testing strain in this work. *Escherichia coli* BL21 was used as the protein expression strain for downstream protein purification. Cells were grown in LB-Lennox (10 g/L casein peptone (Oxoid, ThermoFisher Scientific), 5 g/L yeast extract (Oxoid, ThermoFisher Scientific), 5 g/L NaCl (VWR) and supplemented with 15 g/L agar (Oxoid, ThermoFisher Scientific) for agar plates) supplemented with the appropriate antibiotics (Sigma-Aldrich). All enzymes were purchased from New England Biolabs. Primers were ordered from Eurofins Genomics or Sigma-Aldrich (primer list: *Supplementary Table S1*). PCR reactions were performed with Q5 polymerase (NEB) unless specified otherwise. Colony PCR reactions were performed with *Taq* polymerase (NEB). QiaQuick PCR purification kit (Qiagen) and QIAprep plasmid Miniprep kit (Qiagen) were used for the purification of PCR products and plasmid DNA respectively. DNA sequences were confirmed by Sanger sequencing performed by Eurofins Genomics.

Construction of standard cassettes

Genes encoding the mini-inteins ssp-DnaB[58] and ssp-DnaX[59,63] were synthesized by Twist Bioscience. The DNA fragments were flanked with biobrick prefix and suffix for PCR and Gibson assembly purposes. The fragment also contained two SapI restriction sites for insertion of the gene of interest after the intein, and two BsaI restriction sites to transfer the fusion gene cassette to a new backbone with a selected promoter (see *Supplementary Figure S1*). The synthesized DNA fragments were amplified by PCR with primers MFL 25 & 26 and subsequently purified (see primer list in *Supplemental Table S1*). A pUC8 backbone was amplified by PCR with primers MFL 334 & 335, containing complementary overhangs to the synthesized DNA fragments. Each mini-intein DNA fragment was assembled onto the pUC8 backbone by Gibson assembly[64] (see *Supplementary Figure S1a*). Then, 10 µL of each assembly mixture was chemically transformed into competent *E. coli* (heat shock 42°C, 45s) and cells were plated on LB-agar plates containing 100 µg/mL ampicillin. Correct insertion was assessed by colony PCR with primers MFL 25 & 26. Positive clones were grown overnight at 37°C in

5 mL LB containing 100 µg/mL ampicillin, the respective plasmids were purified by Miniprep and the sequence was confirmed by Sanger DNA sequencing.

Three genes of interest were used to test the intein standard expression system, *sfGFP*, *mCherry* and the chloramphenicol acetyltransferase gene *cat*. *sfGFP* was PCR amplified in two parts from the biobrick (BBa_I746916) with primers MFL 355 & 356 and MFL 357 & 132 respectively, in order to eliminate an internal BbsI restriction site and insert a C-terminal 6-His tag onto *sfGFP*. The *E. coli* codon optimized *mCherry* gene was synthesized by Twist Bioscience and all internal BbsI restriction sites were removed by using synonymous codons. The *mCherry* gene was amplified with primers MFL 351 & 354. The *cat* gene was amplified from the pXMJ19 plasmid with primers MFL 359 & 318. The PCR amplifications conferred each gene with upstream and downstream SapI restriction sites, with CCA and TAA scars respectively, complementary to the SapI scars of the intein DNA fragments. Each gene was assembled onto the pUC8-DnaB and pUC8-DnaX plasmids by cycles of SapI restriction (2 min, 37°C) and ligation with T4 ligase (2 min, 16°C) as presented in the Start-Stop assembly method[65] (see *Supplementary Figure S1b*). 10 µL of each Start-Stop assembly mixture was chemically transformed into *E. coli* and cells were plated on LB-agar plates containing 100 µg/mL ampicillin. Colony PCR with primers MFL 25 & 26 was performed to check the correct insertion of the genes of interest. Positive clones were grown overnight at 37°C in 5 mL LB containing 100 µg/mL ampicillin and the respective plasmids were purified and sequenced. The resulting plasmids pUC8-DnaB-GOI and pUC8-DnaX-GOI were used as template for the amplification of the standard cassettes.

Creation of promoter library and promoter selection

To create a promoter library adapted to *DnaB* and *DnaX* gene sequence, we use the GeneEE method[57] that utilizes 200-nucleotide long DNA fragment of random composition to generate gene-tailored promoters and 5'UTRs. In brief, a single stranded DNA fragment of 200 random nucleotides (200N), synthesized by Integrated DNA Technology (Louvain, Belgium), was amplified by PCR with primer MFL 25 & 26. The resulting double stranded DNA fragment contained BsaI restriction sites with an upstream 5'-TGCC-3' scar and a downstream 5'-NATG-3' scar. The pUC8-DnaB-GFP and pUC8-DnaX-GFP were used as templates for PCR amplification of the *DnaB-GFP* and *DnaX-GFP* gene

fragments (primers MFL 25 & 26) followed by treatment with DpnI overnight at 37°C. The *DnaB-GFP* and *DnaX-GFP* gene fragments both contained an upstream 5'-AATG-3' BsaI scar, but a different downstream BsaI scar, 5'-AGTT-3' for *DnaB-GFP* and 5'-TCAA-3' for *DnaX-GFP*. The random 200N DNA fragment and the intein-GFP fragment were assembled on pUC19 backbones with corresponding BsaI scars by a 3-piece Golden Gate assembly[66] (see *Supplementary Figure S1c*). Two *E. coli* transformations were performed for each Golden Gate mixture and cells were plated on LB-agar plates containing 100 µg/mL ampicillin. For both *DnaB-GFP* and *DnaX-GFP*, colonies seemingly displaying green fluorescence under UV light were picked and grown overnight at 37°C at 800 rpm into two 96-well plates containing LB supplemented with 100 µg/mL ampicillin.

Cells were transferred into black 96-well plates with transparent bottom (Thermo Scientific) and fluorescence was measured with an Infinite M200 Pro TECAN fluorimeter (Noax Lab AS). GFP was excited with a wavelength of 488 nm and fluorescence emission was detected at 526 nm. The fluorescence values were normalized by the OD_{600} of the corresponding well. For both *DnaB-GFP* and *DnaX-GFP*, 10 strains representing different expression levels were selected and grown in triplicates overnight at 37°C with 800 rpm agitation in 96-well plates with LB supplemented with 100 µg/mL ampicillin. Strains displaying inconsistent fluorescence values when replicated were abandoned. Seven strains for *DnaB-GFP* and eight for *DnaX-GFP* were conserved to represent the expression range and each GES was sequenced (see *Supplementary Table S3*).

Promoter transfer to other gene cassettes

In order to assess whether the strength of GES remains equivalent when other genes are fused to the inteins, we tested the *mCherry* gene and the chloramphenicol resistance (CmR) gene *cat*. Each GES selected for *DnaB-GFP* and *DnaX-GFP* was extracted by PCR with primers MFL 330 & 331 and MFL 332 & 333, respectively. These primers contain BsaI recognition sites that confer an upstream 5'-TGCC-3' scar and a downstream 5'-AATG-3' scar on each side of the GES sequences. The *DnaB/X-mCherry* and *DnaB/X-cat* cassettes were amplified by PCR with primer MFL 25 & 26 from the pUC8-Intein-GOI plasmids described above, and digested with DpnI overnight at 37°C. As for *GFP*, the extracted GES sequences and the Intein-GOI cassettes were assembled on pUC19 backbones with

corresponding BsaI scars by a 3-piece Golden Gate assembly[66] (see *Supplementary Figure S1c*). Each Golden Gate mixture was transformed into *E. coli* and plated on LB-agar plates containing 100 µg/mL ampicillin for *mCherry* constructs and 10 µg/mL chloramphenicol for *cat* constructs. Colonies selected on chloramphenicol and visibly red colonies under UV light were grown in triplicates overnight at 37°C with 800 rpm agitation in 96-well plates in LB media supplemented with the appropriate antibiotic.

Cells carrying *DnaB/X-mCherry* plasmids were transferred into black 96-well plates with transparent bottom to measure their respective fluorescence with an Infinite M200 Pro TECAN fluorimeter. The fluorescence of mCherry was monitored with a wavelength couple of 576/610nm and fluorescence values were normalized by the OD_{600} of the corresponding well. Cells displaying fluorescence levels deviating from expectations based on GFP fluorescence levels were grown overnight in 5 mL LB supplemented with 100 µg/mL ampicillin for plasmid purification and sequencing.

Cells carrying *DnaB/X-cat* plasmids grown in a 96-well plate were diluted 100 times into a new 96-well plate containing LB supplemented with 10 µg/mL chloramphenicol. The dilutions were stamped on 15 cm diameter LB-agar plates containing increasing chloramphenicol concentrations (10, 30, 50, 75, 100, 150, 200, 250, 300, 350, 400 and 500 µg/mL) using a 96-pin replicator. Cells were incubated overnight at 37°C and photographed with a Canon camera EOS M.

Coupling of intein cassettes

The gene cassettes described above with *DnaB-GOI* and *DnaX-GOI* were assembled on pUC19 backbones from the previously described pathway assembly method (Fages-Lartaud *et al*[62]). The principles of the method that are utilized in this study are described hereafter. In the first level, promoter libraries or selected promoters are assembled with a single gene on a respective pUC19 backbone (Lv1 plasmids) (see *Supplementary Figure S1c*). Each Lv1 plasmid contains an upstream and a downstream BbsI site. Here, the Lv1 carrying *DnaB-GOIs* contains and upstream 5'-TGCC-3' BbsI scar and a downstream 5'-AGTT-3' BbsI scar; and the Lv1 carrying *DnaX-GOIs* contains and upstream 5'-AGTT-3' BbsI scar and a downstream 5'-TCAA-3' BbsI scar. The *DnaB-GOI* and *DnaX-GOI* cassettes can be

coupled on a Lv2 plasmid containing matching 5'-TGCC-3' and 5'-TCAA-3' BbsI scars that is selectable on chloramphenicol (see *Supplementary Figure S1d*).

Different combinations of *DnaB-mCherry* and *DnaX-GFP* with various GES intensities were tested. The respective Lv1 plasmids were mixed together with a Lv2 backbone and subject to 50 Golden Gate-like assembly cycles for BbsI restriction (2 min, 37°C) and T4 ligase ligation (2 min, 16°C)[67 62]. The resulting assembly mixtures were transformed into *E. coli* cells that were plated on LB-agar plates containing 30 µg/mL chloramphenicol and incubated overnight at 37°C. Positive clones were grown in triplicates, overnight at 37°C with 800 rpm agitation, in 96-well plates in LB supplemented with 30 µg/mL chloramphenicol. Cells were transferred into black 96-well plates with transparent bottom to measure the fluorescence of GFP and mCherry as described previously. GES sequences were sequenced in constructs that resulted in discrepant fluorescence intensities compared to the expectations based on fluorescence measurements of the single-gene constructs.

Fluorescence microscopy

The strains carrying the weakest single-gene GES (P1) and the weakest paired-gene GES (P2) were analyzed by fluorescence microscopy to confirm protein expression and cellular co-localization. One microliter of overnight culture was analyzed with an inverted microscope (Zeiss Axio Observer.Z1, 14 2.3.64.0) possessing a 20x air objective (NA 0.8). The GFP and mCherry filters were applied to measure the fluorescence of both proteins. Image processing was performed with Zeiss image analysis software (2.3.64.0).

Intein inactivation for the assessment of protein cleavage

In order to verify the correct C-terminal cleavage of intein, we compared intein cleavage with inactive, C-terminal truncated variants of *DnaB* and *DnaX*. The plasmids carrying a strong GES expressing *DnaB-GFP* and *DnaX-GFP* were used for the analysis of C-terminal cleavage. The Lv1-P7-*DnaB-GFP* and Lv1-P8-*DnaX-GFP* were amplified by PCR with primers MFL 633 & 132 and MFL 634 & 132 respectively. The PCR creates a C-terminal truncation of the *DnaB* and *DnaX* genes and primers contain SapI sites necessary for Start/Stop assembly. PCR products were digested with DpnI overnight at 37°C

and subsequently purified. The single fragments were subject to 50 cycles of SapI digestion (2 min, 37°C) and ligation with T4 ligase (2 min, 16°C) as described before. A volume of 10 µL of each Start/Stop assembly mixture was transformed into *E. coli* cells. Transformants were selected on LB-agar plates containing 100 µg/mL ampicillin and incubated overnight at 37°C. Several clones were grown overnight at 37°C in 5 mL of LB medium containing 100 µg/mL ampicillin and their respective plasmids were purified and sequenced. The *sfGFP* gene contains a C-terminal 6-His tag suitable for protein purification.

Protein production and purification

The Lv1-P7-*DnaB-GFP* and Lv1-P8-*DnaX-GFP* plasmids and their mutated versions were transformed into *E. coli* BL21 for protein production. Positive clones were grown overnight at 37°C in 5 mL LB medium containing 100 µg/mL ampicillin. A volume of 2 mL of overnight culture was used to inoculate canonical Erlenmeyer flasks containing 200 mL of the same medium. The cultures were incubated in a shaking incubator at 30ºC and 225 rpm for 24 hours. Cells were harvested by centrifugation (4500 × g, 10 min) followed by cell lysis using pulsed sonication in 10 mL lysis buffer (50 mM Tris-HCl, 50 mM NaCl, 0.05 % Triton X-100, pH 8.0) supplemented with 1 tablet EDTA-free cOmplete ULTRA protease inhibitor (Roche). Cell debris were removed by centrifugation (30,000 × g, 30 min). The supernatant was sterilized by filtration through a 0.2 µm Nalgene sterile vacuum filter unit (ThermoFischer). Lysates were used for Western blot analysis after SDS-PAGE migration (see below). Samples for LC-MS analysis were purified on Nickel-sepharose column and analyzed on SDS-PAGE (see below).

Protein purification

Lysates were purified using a 1 mL HisTrap HP Ni-sepharose column (GE Healthcare Life Sciences) connected to an ÄKTA pure protein purification system (Cytiva). The column was equilibrated with 5 CV Buffer A (50 mM Tris-HCl, 300 mM NaCl, pH 8.0) before loading the supernatant onto the column. Impurities were removed by washing with Buffer A for 10 CV. His-tagged proteins were eluted using a gradient of 0-100% Buffer B (50 mM Tris-HCl, 300 mM NaCl, 400 mM imidazole, pH 8.0) over 40 CV. All steps were performed using a flow-rate of 1 mL/min. GFP-containing fractions containing were identified using absorbance at 488 nm.

Protein analysis

SDS-PAGE gels were run under denaturing conditions using SurePAGE™ Bis-Tris 12% gels (GenScript) and Tris-MES-SDS Running buffer (GenScript). The gel used for Western blotting was not stained, whereas the gel used for proteomics analysis was stained using the eStain™ L1 protein staining system (GenScript). Precision Plus Protein™ standards (Bio-Rad Laboratories) were used for the identification of target proteins in both gels.

After SDS-PAGE, proteins were transferred from the unstained gel to polyvinylidene fluoride membranes using the Trans-Blot turbo transfer system (Bio-Rad Laboratories). An iBind™ Flex system (Invitrogen/ThermoFischer) was used for blocking and antibody incubation according to the manufacturer's protocol. The primary antibody (6xHis tag monoclonal antibody; Invitrogen/ThermoFischer), secondary antibody (Polyclonal rabbit anti-mouse immunoglobulins/horseradish peroxidase; Dako/Agilent) and iBind™ Flex solutions were added to their corresponding reservoirs in the iBind™ Flex system. Following incubation for 1 h at room temperature, membranes were rinsed in water prior to immunodetection with 3,3',5,5'-tetramethylbenzidine (Sigma-Aldrich/Merck).

Sample Preparation for LC-MS

Please note that (unless otherwise specified): volumes are to cover the gel bands, also liquids are removed after each incubation. The excised gel bands of interest were cut in smaller pieces (3-5 mm3) and were de-stained by incubation for 15 minutes in 50 mM ammonium bicarbonate (ABC), 50% methanol. Then, they were shrunk with acetonitrile for 15 minutes. The samples were reduced by 10 mM DTT in 25 mM ABC at 56°C for 45 minutes, alkylated by 55 mM iodoacetamide in 25 mM ABC at room temperature in the dark for 45 minutes, washed with 50 mM ABC, 50% methanol for 10 minutes and shrunk with acetonitrile. Then, 12.5 ng/µL trypsin in 50 mM ABC was added to the gel pieces and incubated for 30 minutes on ice. After removal of the liquid, 50 mM ABC was added, and samples were digested by trypsin at 37°C overnight. Peptides were collected and dried in a vacuum concentrator at

room temperature. Dried peptides were reconstituted in 50 µL 0.1% formic acid in water and shaken at 6 degrees Celsius at 900 rpm for 1.5 hour. Samples were centrifuged at 16,000 × g for 10 minutes and 40 µL of the supernatants were transferred to MS-vials for LC-MS analysis.

LC-MS analysis

LC-MS analysis was performed on an EASY-nLC 1200 UPLC system (Thermo Scientific) interfaced with an Q Exactive mass spectrometer (Thermo Scientific) via a Nanospray Flex ion source (Thermo Scientific). Peptides were injected onto an Acclaim PepMap100 C18 trap column (75 µm i.d., 2 cm long, 3 µm, 100 Å, Thermo Scientific) and further separated on an Acclaim PepMap100 C18 analytical column (75 µm i.d., 50 cm long, 2 µm, 100 Å, Thermo Scientific) using a 120-minute multi-step gradient (90 min 5%-40% B, 15 min 40%-100% B, 15 min at 100% B; where B is 0.1 % formic acid and 80% CH_3CN and A is 0.1 % formic acid) at 250 nl/min flow. Peptides were analyzed in positive ion mode under data-dependent acquisition using the following parameters: Electrospray voltage 1.9 kV, HCD fragmentation with normalized collision energy 25. Each MS scan (200 to 2000 m/z, 2 m/z isolation width, profile) was acquired at a resolution of 70,000 FWHM in the Orbitrap analyzer, followed by MS/MS scans at resolution 17,500 (2 m/z isolation width, centroid) triggered for the 8 most intense ions, with a 60 s dynamic exclusion and analyzed in the Orbitrap analyzer. Charge exclusion was set to unassigned, 1, >5.

Processing of LC-MS Data

Proteins were identified by processing the LC-MS data using Thermo Scientific Proteome Discoverer (Thermo Scientific) version 2.5. The following search parameters were used: enzyme specified as trypsin with maximum two missed cleavages allowed; acetylation of protein N-terminus with methionine loss, oxidation of methionine, and deamidation of asparagine/glutamine were considered as dynamic and carbamidomethylation of cysteine as static post-translational modifications; precursor mass tolerance of 10 parts per million with a fragment mass tolerance of 0.02 Da. Sequest HT node queried the raw files against sequences for expected proteins (DnaB GFP, DnaX GFP, BM GFP, GFP C-terminal) and a common LC-MS contaminants database. For downstream analysis of these peptide-

spectrum matches (PSMs), for protein and peptide identifications the PSM FDR was set to 1% and as high and 5% as medium confidence, thus only unique peptides with these confidence thresholds were used for final protein group identification and to label the level of confidence, respectively.

References

(1) Sanchez, S.; Demain, A. Special Issue on the Production of Recombinant Proteins. *Biotechnol. Adv.* **2012**, *30* (5), 1100–1101. https://doi.org/10.1016/j.biotechadv.2011.12.004.

(2) Headon, D. R.; Walsh, G. The Industrial Production of Enzymes. *Biotechnol. Adv.* **1994**. https://doi.org/10.1016/0734-9750(94)90004-3.

(3) Puetz, J.; Wurm, F. M. Recombinant Proteins for Industrial versus Pharmaceutical Purposes: A Review of Process and Pricing. *Processes* **2019**, *7* (8), 476. https://doi.org/10.3390/pr7080476.

(4) Liu, L.; Yang, H.; Shin, H. D.; Chen, R. R.; Li, J.; Du, G.; Chen, J. How to Achieve High-Level Expression of Microbial Enzymes: Strategies and Perspectives. *Bioengineered*. July 2013, pp 212–223. https://doi.org/10.4161/bioe.24761.

(5) Makrides, S. C. Strategies for Achieving High-Level Expression of Genes in Escherichia Coli. *Microbiol. Rev.* **1996**, *60* (3), 512–538.

(6) Overton, T. W. Recombinant Protein Production in Bacterial Hosts. *Drug Discovery Today*. Elsevier Ltd 2014, pp 590–601. https://doi.org/10.1016/j.drudis.2013.11.008.

(7) Ferrer-Miralles, N.; Villaverde, A. *Bacterial Cell Factories for Recombinant Protein Production; Expanding the Catalogue*; 2013; Vol. 12.

(8) Gupta, V.; Sengupta, M.; Prakash, J.; Tripathy, B. C. Production of Recombinant Pharmaceutical Proteins. In *Basic and Applied Aspects of Biotechnology*; Springer Singapore, 2017; pp 77–101. https://doi.org/10.1007/978-981-10-0875-7_4.

(9) Tripathi, N. K.; Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Frontiers in Bioengineering and Biotechnology*. Frontiers Media S.A. December 20, 2019. https://doi.org/10.3389/fbioe.2019.00420.

(10) Burnett, M. J. B.; Burnett, A. C. Therapeutic Recombinant Protein Production in Plants: Challenges and Opportunities. *Plants People Planet*. Blackwell Publishing Ltd March 1, 2020, pp 121–132. https://doi.org/10.1002/ppp3.10073.

(11) Wurm, F. M. Production of Recombinant Protein Therapeutics in Cultivated Mammalian Cells. *Nature Biotechnology*. November 2004, pp 1393–1398. https://doi.org/10.1038/nbt1026.

(12) O'Flaherty, R.; Bergin, A.; Flampouri, E.; Mota, L. M.; Obaidi, I.; Quigley, A.; Xie, Y.; Butler, M. Mammalian Cell Culture for Production of Recombinant Proteins: A Review of the Critical Steps in Their Biomanufacturing. *Biotechnology Advances*. Elsevier Inc. November 1, 2020. https://doi.org/10.1016/j.biotechadv.2020.107552.

(13) Wehrs, M.; Tanjore, D.; Eng, T.; Lievense, J.; Pray, T. R.; Mukhopadhyay, A. Engineering Robust Production Microbes for Large-Scale Cultivation. *Trends Microbiol.* **2019**, *27* (6), 524–537. https://doi.org/10.1016/j.tim.2019.01.006.

(14) Pitera, D. J.; Paddon, C. J.; Newman, J. D.; Keasling, J. D. Balancing a Heterologous Mevalonate Pathway for Improved Isoprenoid Production in Escherichia Coli. *Metab. Eng.* **2007**. https://doi.org/10.1016/j.ymben.2006.11.002.

(15) Paddon, C. J.; Westfall, P. J.; Pitera, D. J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M. D.; Tai, A.; Main, A.; Eng, D.; Polichuk, D. R.; Teoh, K. H.; Reed, D. W.; Treynor, T.; Lenihan, J.; Jiang, H.; Fleck, M.; Bajad, S.; Dang, G.; Dengrove, D.; Diola, D.; Dorin, G.; Ellens, K. W.; Fickes, S.; Galazzo, J.; Gaucher, S. P.; Geistlinger, T.; Henry, R.; Hepp, M.; Horning, T.; Iqbal, T.; Kizer, L.; Lieu, B.; Melis, D.; Moss, N.; Regentin, R.; Secrest, S.; Tsuruta, H.; Vazquez, R.; Westblade, L. F.; Xu, L.; Yu, M.; Zhang, Y.; Zhao, L.; Lievense, J.; Covello, P. S.; Keasling, J. D.; Reiling, K. K.; Renninger, N. S.; Newman, J. D. High-Level

Semi-Synthetic Production of the Potent Antimalarial Artemisinin. *Nature* **2013**, *496* (7446), 528–532. https://doi.org/10.1038/nature12051.

(16) Kumaran Ajikumar, P.; Xiao, W.-H.; Tyo, K. E. J.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G. Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia Coli. *Science (80-.).* **2010**, *330*, 70–74.

(17) Tan, S. Z.; Prather, K. L. Dynamic Pathway Regulation: Recent Advances and Methods of Construction. *Current Opinion in Chemical Biology*. Elsevier Ltd December 1, 2017, pp 28–35. https://doi.org/10.1016/j.cbpa.2017.10.004.

(18) Galanie, S.; Thodey, K.; Trenchard, I. J.; Filsinger Interrante, M.; Smolke, C. D. Complete Biosynthesis of Opioids in Yeast. *Science (80-.).* **2015**, *349* (6252), 1095–1100. https://doi.org/10.1126/science.aac6383.

(19) Engstrom, M. D.; Pfleger, B. F. Transcription Control Engineering and Applications in Synthetic Biology. *Synthetic and Systems Biotechnology*. KeAi Communications Co. September 1, 2017, pp 176–191. https://doi.org/10.1016/j.synbio.2017.09.003.

(20) Salis, H. M.; Mirsky, E. A.; Voigt, C. A. Automated Design of Synthetic Ribosome Binding Sites to Control Protein Expression. *Nat. Biotechnol.* **2009**, *27* (10), 946–950. https://doi.org/10.1038/nbt.1568.

(21) Salis, H. M. The Ribosome Binding Site Calculator. In *Methods in Enzymology*; Academic Press Inc., 2011; Vol. 498, pp 19–42. https://doi.org/10.1016/B978-0-12-385120-8.00002-4.

(22) Wang, Y.; Wang, H.; Wei, L.; Li, S.; Liu, L.; Wang, X. Synthetic Promoter Design in Escherichia Coli Based on a Deep Generative Network. *Nucleic Acids Res.* **2020**, *48* (12). https://doi.org/10.1093/nar/gkaa325.

(23) Studier, F. W. Use of Bacteriophage T7 Lysozyme to Improve an Inducible T7 Expression System. *J. Mol. Biol.* **1991**, *219* (1). https://doi.org/10.1016/0022-2836(91)90855-Z.

(24) Espah Borujeni, A.; Channarasappa, A. S.; Salis, H. M. Translation Rate Is Controlled by Coupled Trade-Offs between Site Accessibility, Selective RNA Unfolding and Sliding at Upstream Standby Sites. *Nucleic Acids Res.* **2014**, *42* (4), 2646–2659. https://doi.org/10.1093/nar/gkt1139.

(25) Espah Borujeni, A.; Salis, H. M. Translation Initiation Is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. *J. Am. Chem. Soc.* **2016**, *138* (22), 7016–7023. https://doi.org/10.1021/jacs.6b01453.

(26) Espah Borujeni, A.; Cetnar, D.; Farasat, I.; Smith, A.; Lundgren, N.; Salis, H. M. Precise Quantification of Translation Inhibition by MRNA Structures That Overlap with the Ribosomal Footprint in N-Terminal Coding Sequences. *Nucleic Acids Res.* **2017**, *45* (9), 5437–5448. https://doi.org/10.1093/nar/gkx061.

(27) Mignone, F.; Gissi, C.; Liuni, S.; Pesole, G. Untranslated Regions of MRNAs. *Genome Biol.* **2002**, *3* (3), 1–10.

(28) Shine, J.; Dalgarno, L. The 3'-Terminal Sequence of Escherichia Coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites (Terminal Labeling/Stepwise Degradation/Protein Synthesis/Suppression). *Proc. Natl. Acad. Sci.* **1974**, *71* (4), 1342–1346.

(29) Kozak, M. Point Mutations Define a Sequence Flanking the AUG Initiator Codon That Modulates Translation by Eukaryotic Ribosomes. *Cell* **1986**, *44* (2). https://doi.org/10.1016/0092-8674(86)90762-2.

(30) Tuller, T.; Waldman, Y. Y.; Kupiec, M.; Ruppin, E. Translation Efficiency Is Determined by

Both Codon Bias and Folding Energy. *Proc. Natl. Acad. Sci.* **2010**, *107* (8), 3645–3650. https://doi.org/10.1073/pnas.0909910107.

(31) Tuller, T.; Carmi, A.; Vestsigian, K.; Navon, S.; Dorfan, Y.; Zaborske, J.; Pan, T.; Dahan, O.; Furman, I.; Pilpel, Y. An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell* **2010**, *141*, 344–354. https://doi.org/10.1016/j.cell.2010.03.031.

(32) Zhang, S.; Goldman, E.; Zubay, G. Clustering of Low Usage Codons and Ribosome Movement. *J. Theor. Biol.* **1994**, *170* (4). https://doi.org/10.1006/jtbi.1994.1196.

(33) Fredrick, K.; Ibba, M. How the Sequence of a Gene Can Tune Its Translation. *Cell* **2010**. https://doi.org/10.1016/j.cell.2010.03.033.

(34) Shah, P.; Ding, Y.; Niemczyk, M.; Kudla, G.; Plotkin, J. B. Rate-Limiting Steps in Yeast Protein Translation. *Cell* **2013**, *153*, 1589–1601. https://doi.org/10.1016/j.cell.2013.05.049.

(35) Goodman, D. B.; Church, G. M.; Kosuri, S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science (80-.).* **2013**, *342* (6157). https://doi.org/10.1126/science.1241934.

(36) Tuller, T.; Zur, H. Multiple Roles of the Coding Sequence 5' End in Gene Expression Regulation. *Nucleic Acids Res.* **2015**, *43* (1), 13–28. https://doi.org/10.1093/nar/gku1313.

(37) Villada, J. C.; Brustolini, O. J. B.; Silveira, W. B. da. Integrated Analysis of Individual Codon Contribution to Protein Biosynthesis Reveals a New Approach to Improving the Basis of Rational Gene Design. *DNA Res.* **2017**, *24* (4), 419–434. https://doi.org/10.1093/dnares/dsx014.

(38) Bentele, K.; Saffert, P.; Rauscher, R.; Ignatova, Z.; Blüthgen, N. Efficient Translation Initiation Dictates Codon Usage at Gene Start. *Mol. Syst. Biol.* **2013**, *9* (675). https://doi.org/10.1038/msb.2013.32.

(39) Charneski, C. A.; Hurst, L. D. Positively Charged Residues Are the Major Determinants of Ribosomal Velocity. *PLoS Biol.* **2013**, *11* (3). https://doi.org/10.1371/journal.pbio.1001508.

(40) Lu, J.; Deutsch, C. Electrostatics in the Ribosomal Tunnel Modulate Chain Elongation Rates. *J. Mol. Biol.* **2008**, *384* (1). https://doi.org/10.1016/j.jmb.2008.08.089.

(41) Crombie, T.; Swaffield, J. C.; Brown, A. J. P. Protein Folding within the Cell Is Influenced by Controlled Rates of Polypeptide Elongation. *J. Mol. Biol.* **1992**, *228* (1). https://doi.org/10.1016/0022-2836(92)90486-4.

(42) Rosano, G. L.; Ceccarelli, E. A. Rare Codon Content Affects the Solubility of Recombinant Proteins in a Codon Bias-Adjusted Escherichia Coli Strain. *Microb. Cell Fact.* **2009**, *8* (41). https://doi.org/10.1186/1475-2859-8-41.

(43) Jacobson, G. N.; Clark, P. L. Quality over Quantity: Optimizing Co-Translational Protein Folding with Non-'optimal' Synonymous Codons. *Current Opinion in Structural Biology*. Elsevier Ltd June 1, 2016, pp 102–110. https://doi.org/10.1016/j.sbi.2016.06.002.

(44) Walls, D.; Loughran, S. T. Tagging Recombinant Proteins to Enhance Solubility and Aid Purification; 2011. https://doi.org/10.1007/978-1-60761-913-0_9.

(45) Young, C. L.; Britton, Z. T.; Robinson, A. S. Recombinant Protein Expression and Purification: A Comprehensive Review of Affinity Tags and Microbial Applications. *Biotechnol. J.* **2012**, *7* (5). https://doi.org/10.1002/biot.201100155.

(46) Mishra, V. Affinity Tags for Protein Purification. *Curr. Protein Pept. Sci.* **2020**, *21* (8). https://doi.org/10.2174/1389203721666200606220109.

(47) Liu, Z.; Chen, O.; Wall, J. B. J.; Zheng, M.; Zhou, Y.; Wang, L.; Ruth Vaseghi, H.; Qian, L.; Liu, J. Systematic Comparison of 2A Peptides for Cloning Multi-Genes in a Polycistronic

Vector. *Sci. Rep.* **2017**. https://doi.org/10.1038/s41598-017-02460-2.

(48) Sharma, S. S.; Chong, S.; Harcum, S. W. Intein-Mediated Protein Purification of Fusion Proteins Expressed under High-Cell Density Conditions in E. Coli. *J. Biotechnol.* **2006**, *125* (1). https://doi.org/10.1016/j.jbiotec.2006.01.018.

(49) Lahiry, A.; Fan, Y.; Stimple, S. D.; Raith, M.; Wood, D. W. Inteins as Tools for Tagless and Traceless Protein Purification. *J. Chem. Technol. Biotechnol.* **2018**. https://doi.org/10.1002/jctb.5415.

(50) Banki, M.; Wood, D. W. Inteins and Affinity Resin Substitutes for Protein Purification and Scale Up. *Microb. Cell Fact.* **2005**, *4* (1). https://doi.org/10.1186/1475-2859-4-32.

(51) Shah, N. H.; Muir, T. W. Inteins: Nature's Gift to Protein Chemists. *Chemical Science*. 2014. https://doi.org/10.1039/c3sc52951g.

(52) Lin, Y.; Li, M.; Song, H.; Xu, L.; Meng, Q.; Liu, X. Q. Protein Trans-Splicing of Multiple Atypical Split Inteins Engineered from Natural Inteins. *PLoS One* **2013**. https://doi.org/10.1371/journal.pone.0059516.

(53) Topilina, N. I.; Mills, K. V. Recent Advances in in Vivo Applications of Intein-Mediated Protein Splicing. *Mobile DNA*. 2014. https://doi.org/10.1186/1759-8753-5-5.

(54) Wood, D. W.; Camarero, J. A. Intein Applications: From Protein Purification and Labeling to Metabolic Control Methods. *Journal of Biological Chemistry*. 2014. https://doi.org/10.1074/jbc.R114.552653.

(55) Pavankumar, T. Inteins: Localized Distribution, Gene Regulation, and Protein Engineering for Biological Applications. *Microorganisms* **2018**. https://doi.org/10.3390/microorganisms6010019.

(56) Zhang, B.; Rapolu, M.; Liang, Z.; Han, Z.; Williams, P. G.; Su, W. W. A Dual-Intein Autoprocessing Domain That Directs Synchronized Protein Co-Expression in Both Prokaryotes and Eukaryotes. *Sci. Rep.* **2015**. https://doi.org/10.1038/srep08541.

(57) Lale, R.; Tietze, L.; Nesje, J.; Onsager, I.; Engelhardt, K.; Fai Alex, C.; Akan, M.; Hummel, N.; Kalinowski, J.; Rückert, C.; Frank, M. A Universal Method for Gene Expression Engineering. https://doi.org/10.1101/644989.

(58) Wu, H.; Xu, M.-Q.; Liu, X.-Q. Protein Trans-Splicing and Functional Mini-Inteins of a Cyanobacterial DnaB Intein. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1998**, *1387* (1–2). https://doi.org/10.1016/S0167-4838(98)00157-5.

(59) Qi, X.; Meng, Q.; Liu, X. Q. Spontaneous C-Cleavage of a Mini-Intein without Its Conserved N-Terminal Motif A. *FEBS Lett.* **2011**, *585* (15), 2513–2518. https://doi.org/10.1016/j.febslet.2011.06.035.

(60) Mathys, S.; Evans, T. C.; Chute, I. C.; Wu, H.; Chong, S.; Benner, J.; Liu, X.-Q.; Xu, M.-Q. Characterization of a Self-Splicing Mini-Intein and Its Conversion into Autocatalytic N- and C-Terminal Cleavage Elements: Facile Production of Protein Building Blocks for Protein Ligation. *Gene* **1999**, *231* (1–2). https://doi.org/10.1016/S0378-1119(99)00103-1.

(61) Fages-Lartaud, M.; Lisa, T.; Elie, F.; Lale, R.; Hohmann-Marriott, M. F. MCherry Contains a Fluorescent Protein Isoform That Interferes with Its Reporter Function.

(62) Fages-Lartaud, M.; Lale, R.; Hohmann-Marriott, M. F. Universal Method for Activating and Balancing Gene Expression in Metabolic Pathways .

(63) Liu, X.-Q.; Hu, Z. Identification and Characterization of a Cyanobacterial DnaX Intein. *FEBS Lett.* **1997**, *408* (3). https://doi.org/10.1016/S0014-5793(97)00393-1.

(64) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6* (5), 343–345. https://doi.org/10.1038/nmeth.1318.

(65) Taylor, G. M.; Mordaka, P. M.; Heap, J. T. Start-Stop Assembly: A Functionally Scarless DNA Assembly System Optimized for Metabolic Engineering. *Nucleic Acids Res.* **2018**. https://doi.org/10.1093/nar/gky1182.

(66) Engler, C.; Kandzia, R.; Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS One* **2008**, *3* (11). https://doi.org/10.1371/journal.pone.0003647.

(67) Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS One* **2009**, *4* (5). https://doi.org/10.1371/journal.pone.0005553.