
Article:
Brown, A.J. orcid.org/0000-0002-3290-4560, Gibson, S.J. orcid.org/0000-0002-3228-9983, Hatton, D. et al. (2 more authors) (2018) Whole synthetic pathway engineering of recombinant protein production. Biotechnology and Bioengineering. ISSN 0006-3592

https://doi.org/10.1002/bit.26855

Whole synthetic pathway engineering of recombinant protein production

Adam J. Brown[1*], Suzanne J. Gibson[2], Diane Hatton[2], Claire L. Arnall[1], David C. James[1*]

[1]Department of Chemical and Biological Engineering, University of Sheffield, University of Sheffield, Mappin St., Sheffield, S1 3JD, England

[2]Biopharmaceutical Development, MedImmune, Cambridge, CB21 6GH, England

*Corresponding authors:

David C. James, Department of Chemical and Biological Engineering, University of Sheffield, Mappin St., Sheffield, S1 3JD

Telephone: +44-(0)114-222-7505, Email: d.c.james@sheffield.ac.uk

Adam J. Brown, Department of Chemical and Biological Engineering, University of Sheffield, Mappin St., Sheffield, S1 3JD

Telephone: +44-(0)114-222-7594, Email: adam.brown@sheffield.ac.uk

Keywords: protein production, synthetic biology, pathway engineering, DNA parts, cell chassis

Abstract

The output from protein biomanufacturing systems is a function of total host cell biomass synthetic capacity and recombinant protein production per unit cell biomass. In this study we describe how these two properties can be simultaneously optimized via design of a product-specific combination of synthetic DNA parts to maximize flux through the protein synthetic pathway and the use of a host cell chassis with an increased capability to synthesise both cell and product biomass. Using secreted alkaline phosphatase (SEAP) production in Chinese hamster ovary cells as our example, we demonstrate how an optimal composition of input components can be assembled from a minimal toolbox containing rationally designed promoters, untranslated regions, signal peptides, product coding sequences, cell chassis and genetic effectors. Product titer was increased 10-fold, compared to a standard reference system by i) identifying genetic components that acted in concert to maximize the rates of SEAP transcription, translation and translocation, ii) selection of a cell chassis with increased biomass synthetic capacity and iii) engineering the host cell factory's capacity for protein folding and secretion. This whole synthetic pathway engineering process to design optimal expression cassette-chassis combinations should be applicable to diverse recombinant protein and host cell type contexts.

Introduction

Synthetic biology promises to revolutionize biotechnology by enabling the rational design of genetic constructs (parts) and cells (chassis) with predictable user-defined functions. It is particularly applicable to recombinant protein manufacturing, where the engineered system comprises just two key biological components, a recombinant gene expression cassette and a host-cell factory, and two critical system outputs, product yield and quality. Although diverse eukaryotic and prokaryotic cell-types are used to produce recombinant proteins, the core biomanufacturing process remains constant. A cell factory must create and maintain cellular biosynthetic capacity, and utilize it to synthesize a complex protein product. As shown in Figure 1, product yields are a function of the rates of five key biosynthetic steps (four if product membrane translocation is not required), each of which can be controlled by re-design of either the cell chassis or a discrete genetic component. Strategies to enhance product yield have traditionally been limited to increasing the output from a single one of these steps, typically resulting in relatively modest improvements (Davy et al., 2017; Wells & Robinson, 2017; Xiao et al., 2014). Given recent advances in DNA-part and chassis engineering, we should be able to move beyond modulation of individual cellular processes, and create a new paradigm for biomanufacturing where the entire product biosynthetic pathway is specifically designed to maximize system output.

Designing an optimal chassis-expression cassette combination is complicated by the fact that the quantitative output of synthetic genetic constructs (e.g. the amount of mRNA transcribed) can be inversely correlated with host cell factory performance characteristics (Ceroni et al., 2015; Gorochowski et al., 2016; Scott et al., 2010). Heterologous protein expression imposes a metabolic load on the host cell, requiring consumption of limited cellular resources, such as polymerases, ribosomes and protein folding chaperones (Borkowski et al., 2016; Raveh et al., 2016; Weiße et al., 2015). Any attempt to increase the flux through a step in the product biosynthetic pathway via DNA-part design, for example by utilizing a stronger promoter, exerts a concomitant increase in chassis burden. This can be manifested as a reduction in cell growth, an increase in cell stress/death,

and the induction of bottlenecks in the recombinant protein synthetic pathway, particularly in protein folding and secretion (Delic et al., 2014; Dong et al., 1995; Hoffmann & Rinas, 2004). Accordingly, it is challenging to create ideal biomanufacturing systems where total cell biomass, and protein production per unit cell biomass, are both maximized.

Efforts to optimize the chassis-expression cassette interactome (all direct and indirect interactions between the host cell and rDNA or rDNA-derived synthetic intermediates) have focused on reducing the burden that synthetic constructs impose on the host cell. For example, using inducible expression systems, product biosynthesis can be selectively 'turned-off' during the early stages of culture to facilitate maximized accumulation of cellular capacity (Møller et al., 2017; Shrestha & Hildebrand, 2017). The overall biosynthetic cost of recombinant protein production may also be reduced by the use of orthogonal transcriptional/ translational machineries, where the associated metabolic load is preferable to the over-consumption of particularly limited cellular resources (Bervoets et al., 2018; De Jong et al., 2017; Segall-Shapiro et al., 2014). Further, negative feedback loops containing stress-responsive promoters can be designed to automatically tune product expression levels down when burden causes deleterious effects to chassis performance (Ceroni et al., 2018; Dragosits et al., 2012). While these approaches enable improvements in overall system output, they typically increase cell biomass creation and maintenance at the expense of flux through the product biosynthetic pathway. Accordingly, expression cassette design, and therefore product yields, are constrained by the limitations of the chassis. Although cell factory capacities can be increased via genetic engineering, the performance of the designed chassis tends to be highly context-specific, dependent on the product, cell strain, expression cassette, and production process (Delic et al., 2014; Fischer et al., 2015; Hansen et al., 2017; Waegeman & Soetaert, 2011). We therefore hypothesized that in order to create an ideal biomanufacturing system the expression cassette should first be designed to maximize flux through product gene transcription, mRNA translation and protein translocation, before the chassis is subsequently specifically optimized to handle the associated metabolic load. While context-dependency and incomplete understanding of

cell factory and DNA-part functionalities prevents forward engineering of such systems (Cardinale & Arkin, 2012; Pasotti & Zucca, 2014), we reasoned that for a given product, an optimal combination of synthetic elements, each controlling a discrete step in the synthetic pathway (e.g. promoters, untranslated regions, signal peptides, product coding sequences, chassis and protein folding enhancers) could be identified by rational screening of a minimal toolbox of components. Here, using Secreted alkaline phosphatase production in Chinese Hamster Ovary cells as our example, we show that i) application of this approach enables simultaneous maximization of both cellular biomass capacity and cell specific productivity, and ii) synthetic manipulation of the entire product biosynthetic pathway results in substantially greater recombinant protein yields than engineering the cell or DNA parts in isolation. In doing so, we illustrate a process for designing optimal chassis-expression cassette combinations that should be applicable to diverse product-host cell type bioproduction contexts.

Materials and methods

Synthetic part design

Promoters: To generate 100RPU_PH, the proximal promoter region of 100RPU.2 (GenBank accession number (GAN) LC270637, nucleotides 1 - 201) was added to the 5' terminus of 100RPU.1 (GAN LC270636). To create 100RPU.1_SC2, 100RPU.1_SC3, and 100RPU.1_SC4, the proximal promoter region of 100RPU.1 (nucleotides 1 – 209) was fused to super core promoter 2 (Juven-Gershon et al., 2006), 3 (Even et al., 2016) and 4 (a hCMV-IE1 core promoter modified to contain DCE (D.-H. Lee et al., 2005) and DTIE (Marbach-Bar et al., 2015) elements) respectively. The transcription factor binding site, and core promoter element, composition of each synthetic promoter is shown in Supplementary Figure 1. Sequences have been deposited with the DNA Data Bank of Japan (accession numbers LC380022 – LC380025).

Protein coding sequences: To generate SEAP_MO, the 'native' secreted alkaline phosphatase (SEAP) coding sequence (signal peptide + mature protein coding sequence; GAN U89938.1, nucleotides 272 – 1831) was re-designed for expression in Chinese hamster ovary (CHO) cells using the GeneOptimizer ™ algorithm (Raab et al., 2010). SEAP_CAI and SEAP_HT were created using the Codon Optimization On-Line (COOL) software tool (Chin et al., 2014), according to the following, respective, design settings: i) maximization of the sequence's codon adaption index value, and ii) maximization of 'folding instability' and codon context score between nucleotides 1 – 50 and 51 – 1560 respectively. Finally, SEAP_MSS was designed using EuGene (Gaspar et al., 2012), employing minimization of secondary structure free energy as the sole design parameter. Synthetic gene sequences were analyzed to check for the presence of splice sites (NNsplice (Reese et al., 1997); detection threshold = 0.6), transcription termination motifs (AATAAA; 100% match), transcription initiation motifs (MatInspector (Cartharius et al., 2005) matrix library = general core promoter elements, core/ matrix similarity threshold = 1.00/ 0.95), and mRNA instability elements (ATTTA, GTTTG, TTTTT, TTTGTTT, TTTTAAA; 100% match). Undesirable sequence features were removed by substituting an appropriate codon, where all possible codon-replacement solutions were tested to identify those that had minimal impact on desired design criteria. Sequences have been deposited with the DNA Data Bank of Japan (accession numbers LC380026 – LC380029).

Signal peptides: 1313 experimentally-verified mammalian (human and murine) signal peptide sequences were extracted from the Signal Peptide Website (www.signalpeptide.de). Examination of these sequences revealed that the majority of constructs (717) were between 18 and 24 amino acids (aa) in length. Signal peptides in this size range were analyzed to determine the modal length and amino acid usage bias of constituent N-terminal, Hydrophobic, and C-terminal regions (see Fig. S2). Using this information, thousands of synthetic subcomponents were randomly generated according to the following design rules: N-terminal regions - length = 4 aa, composite aa = A, G, K, L, M (only in position one), P, R, S; Hydrophobic regions – length = 13

aa, composite aa = A, L, V; C-terminal regions – length = 5 aa, composite aa = A, G, L, P, S. These sub-parts, and those from the previously described 'super' synthetic signal peptide Secrecon (Barash et al., 2002) were then randomly combined to create 10,000, 21-aa-long novel components (Fig. S2). To determine predicted cleavage sites and relative "signal peptide-ness" (D-scores), each synthetic signal peptide was screened in silico using SignalP 4.1 (www.cbs.dtu.dk/services/SignalP (Petersen et al., 2011)). The four components with the highest D-score values were selected for in vitro testing, alongside secrecon. Güler-Gane et al. recently showed that the secretion efficiency of recombinaint proteins can be generally increased by inserting two alanine residues at the signal peptide-mature protein junction (Güler-Gane et al., 2016), and accordingly this design feature was added to all five synthetic components (see Table S1 for sequences).

Expression cassette construction

A standard reference expression cassette was designed to contain the following components: hCMV-IE1 promoter (GAN M60321.1, nucleotides 517 – 1193); hCMV-IE1 5' UTR (GAN M60321.1, nucleotides 1194 - 2103); 'perfect' kozak consensus sequence (GCCACC); native SEAP signal peptide (GAN U89938.1, nucleotides 272 - 323); native mature SEAP coding sequence (GAN U89938.1, nucleotides 324 - 1831); and an SV40 3' UTR (GAN LT727517.1, nucleotides 1449 – 1676). This construct was synthesized and cloned into the pMA vector (GeneArt, Regensburg, Germany) to create pMA-SRS. Designed synthetic parts, and combinations thereof, were synthesized and inserted into pMA-SRS, replacing corresponding standard reference components. Note that i) when synthetic signal peptides were tested in isolation, GeneOptimizer ™ was utilized to generate signal peptide coding sequences that were inserted directly upstream of the native mature SEAP coding sequence, and ii) when gene optimization methods were applied, the entire protein coding sequence (i.e. signal peptide + mature SEAP coding sequence) was re-designed.

Construction of effector gene expression plasmids

Cypb (UniProt accession number (UAN P23284), Ero1a (UAN Q96HE7), Pdia1 (UAN P07237), Pdia4 (UAN P13667), Bip (UAN P11021), Calr (UAN P27797), Canx (UAN P27824), Hspa1a (UAN P0DMV8), Tor1a (UAN O14656), Atf6ac (UAN P18850, amino acids 1 - 377), Xbp1s (UAN P17861, isoform 2) and Cert (UAN Q9Y5P4) protein coding sequences were optimized for expression in CHO cells using GeneOptimizer ™, synthesized, and inserted into pMA-SRS, replacing the native SEAP signal peptide and mature protein coding sequence. Constitutively active versions of transactivators were utilized, corresponding to 50 kDa and 54 kDa forms of ATF6A (ATF6Ac) and XBP1 (XBP1s) respectively.

CHO cell culture

CHO-S_C1.80 (evolved chassis (ECH) 1), CHO-S_C2.200 (ECH2), CHO-S_C4.200 (ECH3), CHO-S_C11.80 (ECH4) and CHO-S_C17.200 (ECH5) were obtained from A. Fernandez-Martell (Fernandez-Martell et al., 2018). These cell lines, and standard CHO-S cells (Thermo Fisher Scientific, Paisley, UK), were routinely cultured in CD-CHO medium (Thermo Fisher Scientific) at 37°C in 5% (v/v) CO_2 in vented Erlenmeyer flasks (Corning, UK), shaking at 140 rpm, and subcultured every 3-4 days at a seeding density of 2 x 10^5 cells/ml. Cell concentration and viability were determined by an automated Trypan Blue exclusion assay using a Vi-Cell cell viability analyser (Beckman-Coulter, High Wycombe, UK).

Characterizing the performance of discrete expression cassette-chassis combinations

A high-throughput component-palette screening platform was optimized, with respect to DNA load, Lipofectamine concentration and cell seeding density, to facilitate maximal SEAP production titers. Specifically, 8 x 10^5 cells from a mid-exponential phase culture were seeded into individual wells of a 24 well plate (Nunc, Stafford, UK), and subsequently transfected with DNA-lipid complexes comprising 500 ng of DNA and 3 µl of Lipofectamine (Thermo Fisher Scientific). Transfected cells were incubated for 72 h prior to quantification of SEAP protein expression using the Sensolyte pNPP SEAP colorimetric reporter gene assay kit (Cambridge Biosciences, Cambridge, UK). Total cellular capacity (integral of viable cell concentration) and cell specific

productivity (qP) were calculated as described by Khoo and Al-Rubeai (cell growth was measured at 24 h intervals; Khoo & Al-Rubeai, 2009).

Fed-batch transient transfection

Two hours prior to transfection 7 x 10^6 cells from mid-exponential phase CHO cell cultures were seeded into 50 mL CultiFlask bioreactors (Sartorius, Surrey, UK) at a working volume of 10 ml. Cells were transfected with DNA-lipid complexes, comprising 40 µg of DNA and 160 µl of Lipofectamine. To prevent cell clumping, 36 µl of Anti-Clumping Agent (Thermo Fisher Scientific) was added 24 hr post transfection. Fed-batch cultures were maintained for six days by nutrient supplementation with 10% v/v CHO CD Efficient Feed B (Thermo Fisher Scientific) on day 0, 1, 4 and 5. Cell viabilities were measured on day 4 and 5 to confirm that they had not dropped below 70%.

Results and discussion

Creating a minimal synthetic component design-palette to maximize recombinant protein yield

Forward engineering of protein biomanufacturing systems is restricted by the context-dependent function of genetic constructs and cell factories (Cardinale & Arkin, 2012; Nikel et al., 2014; Pasotti & Zucca, 2014). Accordingly, to assemble an expression cassette-cell combination which maximizes flux through each step in product biosynthesis, component libraries must first be tested to identify parts and chassis with desired functionalities. However, given that i) well-characterized parts and chassis are available for all commonly-utilized bioproduction hosts, and ii) it is relatively simple to create novel components with user-defined functions, it should be possible to rapidly perform such screens using minimal design-palettes. While some part types and design methods may vary depending on the specific protein-cell type combination, the design-test-build process that we present here will be generally applicable in all biomanufacturing contexts. To demonstrate the power of this approach we created a component-palette to maximize SEAP (a human

glycoprotein) production in Chinese hamster ovary (CHO) cells, the predominant biopharmaceutical expression host (Fig. 2).

Gene transcription. Context-specific maximization of recombinant gene transcription requires identification of an appropriate promoter part. As highly active promoters have been created for use in all bioproduction hosts, a component with the required functionality can be simply selected by screening a small panel (e.g. < 10) of previously characterized synthetic constructs (Brown & James, 2015; Gilman & Love, 2016; Redden et al., 2014). Alternatively, a 'gold standard' part can be used directly if it has been specifically designed/ shown to maximize transcriptional output in diverse expression conditions. A component with this functionality has reportedly been constructed for use in CHO cells. Synthetic promoter 100RPU.1 was specifically designed in silico, using transcription factor binding sites (TFBSs) with requisite functionalities, to exhibit maximal transcriptional activity and long-term expression stability, without imposing off-target effects on CHO cell growth or viability (Brown et al., 2017). However, this construct was not explicitly shown to exhibit the highest promoter activity possible in CHO cells. Accordingly, we included this part in our component-palette, but also rationally designed four 100RPU.1-variants to test if transcriptional output could be further increased (Fig. S1). We hypothesized that the output from 100RPU.1 could be enhanced by either increasing the number of TFBSs in the proximal region (limited to 14 in the original design), or replacing the non-engineered minimal core region (taken from the hCMV-IE1 promoter). We therefore i) constructed a hybrid promoter (100RPU_PH) containing the proximal regions from both 100RPU.1 and a second highly active synthetic part, 100RPU.2 (Brown et al., 2017), and ii) created elements (100RPU.1_SC2, 100RPU.1_SC3, 100RPU.1_SC4) where the hCMV-IE1 core promoter was substituted by varying 'super cores' that are specifically designed to maximize transcription initiation rates (Even et al., 2016; Juven-Gershon et al., 2006) (see Fig. S1).

mRNA translation. Translational output, a function of mRNA stability and translation initiation/ elongation rates, is determined collectively by the protein coding sequence and 5'/3'

untranslated regions (UTR). The function of key 5' UTR elements controlling translation initiation (e.g. Shine-Dalgarno sequences in prokaryotes, Kozak sequences in eukaryotes) is well understood and can be predictably controlled (Kozak, 1987; Reeve et al., 2014). However, relatively little is known about how other UTR sequence features affect mRNA half-life and translation rate. It is therefore difficult to rationally select or design UTR-parts that are fully optimized for a specific context or purpose. Accordingly, we rationalized that for these components we would instead have to utilize existing constructs that are known to exhibit desirable functionality. For this example, we employed parts that are commonly used in high-producing CHO cell biomanufacturing systems, the SV40 3'UTR and a hCMV-IE1 5'UTR that has been modified to contain a 'perfect' Kozak consensus sequence (hCMV-IE1+PKS).

Design principles governing protein coding sequence performance are similarly poorly understood. While many critical mRNA sequence features have been identified, including codon context, local/ global secondary structure, instability motifs and GC content, multivariate optimization of these interdependent parameters to maximize translational output is currently intractable (Gould et al., 2014; Parret et al., 2016). However, unlike UTRs, a tried and tested component cannot be used 'off the shelf' as new constructs must be generated for each novel product. While it is not possible to rationally build optimal protein coding sequences, we reasoned that a component exhibiting near-maximal translational output could be identified by testing multiple parts optimized according to divergent design rules. Here, we created synthetic constructs with varying characteristics using four distinct design methods that have previously been shown to increase protein expression levels when compared to non-optimized native sequences. Specifically, we employed the following design strategies: i) minimization of mRNA secondary structures (i.e. maximizing the construct's minimum free energy; SEAP_MSS) (Kim et al., 2010), ii) codon optimization according to usage bias in the host cell-type (i.e. maximizing the gene's codon adaptation index value; SEAP_CAI) (Gustafsson et al., 2004), iii) multiparameter optimization of codon usage, secondary structure and GC-content using a sliding window approach (SEAP_MO)

(Fath et al., 2011), and iv) independent engineering of the mRNA head and tail sections to minimize secondary structure and optimize codon context (i.e. match the codon pair usage pattern of host cell genes) respectively (SEAP_HT) (B. K.-S. Chung et al., 2013; Kudla et al., 2009). In each case the formation of undesirable sequence features known to negatively affect component performance, such as splice sites and mRNA instability elements, were specifically precluded (Gould et al., 2014; Parret et al., 2016).

Protein translocation. Signal peptides control the rate of protein translocation across endoplasmic reticulum and cytoplasmic membranes in eukaryotic and prokaryotic cells respectively. The performance of signal peptide parts is typically unpredictable as the design rules governing their function are poorly defined (Brockmeier et al., 2006; Haryadi et al., 2015; Obst et al., 2017). In particular, features underpinning construct affinity for secretory pathway components, such as the signal recognition particle, are not well understood. Moreover, it is not known why signal peptide function is highly protein-specific, a phenomenon that is likely determined by amino acid sequences at the signal peptide-product N terminal junction. Accordingly, it is not possible to rationally select or design signal peptides in order to maximize the translocation rate of specific proteins. However, trial and error testing has identified parts that exhibit robust performance with diverse protein partners, and panels of these constructs can be screened to find optimal signal peptide-product combinations (Güler-Gane et al., 2016; Haryadi et al., 2015; Kober et al., 2013). For this example, such screening was unnecessary as previous studies have already identified an appropriate signal peptide for the specific context of SEAP production in mammalian cells (Barash et al., 2002; Güler-Gane et al., 2016). We therefore included this part (Secrecon) in our component palette. Additionally, to test if SEAP translocation rates could be further increased, we applied basic information regarding signal peptide sequence features to create four novel synthetic constructs. Fundamental signal peptide design principles were determined by analyzing the size, structure (i.e. average length of the N-terminal, Hydrophobic, and C-terminal regions) and amino acid usage patterns of experimentally verified mammalian sequences. Using these design rules, we

created thousands of synthetic parts and tested them in silico using SignalP 4.1, a neural network-based algorithm that quantifies the "signal peptide-ness" (D-score) of amino acid sequences (Petersen et al., 2011). While some studies have shown no relationship between signal peptide prediction scores in silico and functional outputs in vitro (i.e. levels of secreted recombinant protein) (Brockmeier et al., 2006), others have found that these properties are positively correlated in certain contexts, including SEAP production in HEK293 cells (Barash et al., 2002; Mori et al., 2015). Accordingly, in an attempt to maximize SEAP secretion efficiency in our system, we graded synthetic signal peptides according to their D-score values and selected the four top-ranking parts (SSP 1 – 4; see Fig. S2).

Cellular biomass capacity. The total cellular capacity available for protein biosynthesis is a function of the cell factory's ability to both accumulate and maintain biomass. A wide range of cell-variants with enhanced biomass capacities have been created for most protein production hosts, for example by deriving cell lines/strains with increased proliferation rates (Schlegel et al., 2017; Wurm, 2013). However, the biomass performance of a discrete chassis is highly context-specific, dependent on the cell's macromolecular composition (e.g. transcriptional, translational, and glycolytic capacities) and the unique metabolic load imposed by each product-expression cassette combination (Mahalik et al., 2014; O'Callaghan et al., 2010). These two properties are typically poorly defined, and moreover, there is limited understanding of how discrete product biosynthetic processes affect biomass output from given host cell backgrounds. Accordingly, it is not currently possible to rationally select or design a cell factory in order to maximize biomass capacity in a specific biomanufacturing context. We therefore reasoned that multiple candidate-chassis combinations will need to be tested to identify a cell host that functions optimally when paired with a particular designed expression cassette. In this case, we utilized five functionally-distinct CHO cell chassis that were specifically created to exhibit high cellular biomass capacities. Fernandez-Martell et al. recently utilized a directed evolution approach to generate 22 clonally-derived CHO-S subclones with enhanced specific growth rates (Fernandez-Martell et al., 2018). For the majority

of these cell lines, an increase in biomass accumulation during exponential phase growth was associated with an undesirable decrease in cellular biomass maintenance during stationary phase. However, five subclones (ECH1 – ECH5) were shown to exhibit the optimal combination of rapid proliferation and extended maintenance of high cell biomass concentration in multiple bioproduction contexts, and, accordingly, we selected these chassis for our component palette.

Protein folding and secretion. Regardless of the chassis used, if an expression cassette is specifically designed to maximize flux through the product biosynthetic pathway, protein folding (and secretion where applicable) machinery may become exhausted. Indeed, the capacity for this cellular process is commonly found to be limiting in protein biomanufacturing systems (Baeshen et al., 2015; Delic et al., 2014; Zhou et al., 2018). Inefficient conversion of unfolded polypeptide to final protein product restricts system output, and can adversely affect cell function by inducing cellular stress responses (Gasser et al., 2008; Hetz et al., 2015; Hetz & Papa, 2017). These problems can be overcome by genetically engineering the cellular protein folding and secretion machinery, however, the relative benefits of over-expressing discrete effector genes are context-specific, dependent on product folding pathways, chassis macromolecular compositions, and recombinant protein expression levels (Delic et al., 2014; Hansen et al., 2017). Accordingly, to identify engineering solutions that maximize product folding and secretion rates, a wide range of effectors need to be screened in each specific biomanufacturing system. Due to the widespread use of synthetic engineering approaches, catalogues of potential candidate genes are available for all commonly-used bioproduction host cell-types. Here, we selected twelve components of the protein folding and secretion machinery whose overexpression has previously been shown to increase product yields in CHO cells, including foldases (Cypb (Johari et al., 2015), Ero1a (Cain et al., 2013), Pdia1 and Pdia4 (Borth et al., 2005)), chaperones (Bip (Johari et al., 2015), Hspa1a (Y. Y. Lee et al., 2009), Tor1a(Jossé et al., 2010), Calr and Canx (J. Y. Chung et al., 2004)), transactivators (Atf6ac and Xbp1s (Pybus et al., 2014)) and a lipid transporter (Cert (Florin et al., 2009)).

Identification of context-specific parts that maximize flux through product transcription, translation, and translocation.

To create an expression cassette driving maximal rates of SEAP transcription, translation and translocation, we tested the function of our in silico-designed DNA-parts in vitro. We reasoned that the component palette screening system needed to be high-throughput, while still maintaining the environmental context of industrial scale protein production processes. Accordingly, we utilized a relatively rapid (72 h), small scale (24 well plate) screening process where conditions were i) optimized to maximize product yields (i.e. DNA load, transfection conditions, cell seeding density) and ii) designed to match standard CHO cell biomanufacturing systems as much as possible (i.e. growth media, temperature, CO_2 atmosphere, cell growth dynamics). To evaluate the performance of discrete synthetic expression cassette-chassis combinations, we created a standard reference system (SRS; Fig. 2B) comprising components that are either native to the product (SEAP-native signal peptide and mature protein coding sequence) or commonly-used in industry (CHO-S cell chassis, CMV-IE1 promoter + 5' UTR, optimal kozak sequence, SV40 3'UTR).

Designed promoters, signal peptides and protein coding sequences were chemically synthesized and individually inserted into the SRS, replacing corresponding incumbent parts. As shown in Figure 3, synthetic modulation of SEAP transcription, translation and translocation enabled significant increases in product yield, where the best-performing constructs increased production by 2.05x, 1.38x and 1.48x respectively compared to the SRS. As parts enhanced SEAP titer without significantly affecting cell growth (final cell concentrations and culture viabilities varied by < 10% compared to the SRS, data not shown) we presumed that productivity gains resulted from increased flux through the SEAP biosynthetic pathway. However, we did not specifically measure SEAP transcription, translation and translocation levels as we reasoned that i) component performance was ultimately characterized by total system output, and, accordingly, ii) further system characterization would unnecessarily increase screening time, cost and complexity.

Similarly, we did not perform N-terminal sequence analysis to confirm predicted signal peptide cleavage sites, although we note that this would be necessary in some contexts, including biopharmaceutical manufacturing (Gibson et al., 2017).

Overall, 11/14 synthetic parts facilitated an increase in SEAP yield, thereby validating our component design and selection strategies. The poor performance of the other three constructs may have been due to context-specific factors, such as product amino acid sequence. For example, while the expression level of certain proteins can be increased by minimizing mRNA secondary structures (Kim et al., 2010), application of this design method to the SEAP coding sequence (SEAP_MSS) resulted in near-complete abrogation of protein production (Fig. 3C). Although this unpredictability in part performance necessitates screening multiple constructs to identify those that exhibit desired functionality in discrete biomanufacturing contexts, we have shown here that this can be achieved using a minimal component palette ($\leq$ 4 constructs per part type). Moreover, part design and selection methods will further improve with multiple iterations of the design-build-test-learn process.

To increase the rate of multiple product biosynthetic processes simultaneously, we created expression cassettes containing combinations of the best performing synthetic components. As shown in Figure 3D, concurrent enhancement of SEAP transcription, translation and translocation enabled significantly higher titers than systems where only one or two of these rates were optimized. The ideal combination of promoter, protein coding sequence and signal peptide increased SEAP yield by 302% compared to the SRS, thereby highlighting the potential benefits of designing these three part-types in unison. However, while simultaneously increasing the flux through product transcription, translation and translocation should, in theory, synergistically increase system output, incorporating best performing parts into an optimal expression cassette only had an additive effect on SEAP titer. This may have been due to functional redundancy between SSP2 and SEAP_MO as signal peptide coding sequences can affect mRNA translation rates, particularly at the level of initiation. However, utilizing these two parts together without

100RPU_PH (i.e. used in conjunction with the standard hCMV-IE1 promoter) had a synergistic effect on SEAP production, indicating that SSP2 predominantly functioned by increasing protein translocation. 100RPU_PH: SSP2 and 100RPU_PH: SEAP_MO combinations also synergystically increased system output, suggesting that discrete component performances were not significantly affected by physical context (e.g. sequences at part junctions). These data show that when any two rates were increased, product yields were within -12% of that expected if effects were multiplicative, compared to a -28% difference when all three were enhanced. We therefore reasoned that i) the metabolic load imposed on the host cell by concurrent maximization of product transcription, translation and translocation rates created a bottleneck in the recombinant protein synthetic pathway, most likely in protein folding and secretion, and ii) as a result, we had reached the maximum production gains that were possible via DNA-part design. To confirm this, we created an expression cassette containing the second-best performing construct for each component-type (100RPU.1, Secrecon, SEAP_CAI). As shown in Figure 3D, using this promoter-signal peptide-protein coding sequence combination resulted in the same level of system output as 100RPU_PH: SSP2: SEAP_MO. Accordingly, we determined that further increases in SEAP yield would require improvements in CHO cell chassis performance.

Selection and engineering of a chassis to maximize both cellular biomass capacity and cell-specific productivity.

Having identified DNA parts that maximize flux through the SEAP biosynthetic pathway, we next sought to derive a chassis that was specifically optimized to function in conjunction with them. To achieve this, we first tested the optimized expression cassette in combination with our panel of five CHO cell lines that have previously been shown to display desirable biomass phenotypes. Utilization of ECH1, ECH2, or ECH4 significantly enhanced SEAP yield, where the best performing cell line (ECH2) increased production 1.9x compared to the standard CHO-S cell factory (Fig. 4A). Production gains were approximately proportional to increases in total cellular

capacity (measured as the integral of viable cell concentration), indicating that these chassis were able to exhibit improved rates of biomass accumulation and maintenance without suffering an associated decrease in cell-specific productivity. Indeed, the level of SEAP production per cell was slightly increased in ECH1 and ECH2.

To determine the effect of using cell capacity enhancement as a stand-alone approach for improving SEAP titer, we tested each chassis in concert with the standard reference expression cassette. This analysis showed that i) ECH2 facilitated a 2.2x increase in product yield relative to the SRS, and ii) the rank order of chassis performance was the same when using either combination of DNA parts (see Fig. S3). Given the latter, we concluded that relative differences in SEAP production between distinct cell chassis were not a result of varying output from discrete genetic components. Indeed, while altering the chassis carries a risk of significantly affecting part functionalities, we assume that this is relatively minor given that it would require substantial changes in the activity of specific transcriptional, translational or translocational machinery components (e.g. distinct transcription factors or tRNAs). We postulated that the high cellular capacity/low SEAP production phenotype displayed by ECH3 and ECH5 may have been a function of unequal resource allocation, whereby these chassis produced cellular biomass at the expense of SEAP biosynthesis. The poor performance of these cell factories, which could also have been due to product or process specific factors, highlights the requirement to test a panel of chassis in each unique biomanufacturing context. However, as demonstrated here, if cell lines are rationally selected based on well-characterized phenotypes, a chassis with desired functionality can be identified from a relatively small number of candidates.

Finally, we attempted to enhance SEAP production-per-cell by increasing the chassis' capacity for protein folding and secretion. The twelve candidate genetic effectors were each transfected into ECH2 alongside the optimized 100RPU_PH: SSP2: SEAP_MO expression cassette. Given that genetic engineering outcomes are dependent on effector expression levels, where gene dosages need to be carefully tailored to achieve desired phenotypes without negatively

affecting other important cellular functions (Johari et al., 2015; Pybus et al., 2014; Tastanova et al., 2016), we tested each component at three different concentrations. To ensure a fair comparison between the control (i.e. SEAP reporter plasmid transfected in isolation) and genetically engineered systems, we i) kept DNA load constant across all transfections in order to maintain optimized experimental conditions, and ii) did not use empty vector to normalize SEAP reporter plasmid copy numbers (i.e. to prevent artificially hampering control system performance). No significant improvements in SEAP yield were observed when effector genes were used at a 1:40 or 1:20 molar ratio relative to the product gene (data not shown). However, product titer increased 1.39-fold compared to the control system when Atf6ac and Seap were co-transfected at a 1:10 ratio (Fig. 4B). Subsequent testing of this effector over an extended concentration range identified that SEAP yield was maximized (161% increase over control) by using an Atf6ac: Seap ratio of 1: 6.66, where increased production was a result of enhanced cell specific productivity (i.e. cellular biomass capacity was unaffected; Fig. 4C).

ATF6A is a TF that upregulates expression of protein folding machinery components, such as foldases and chaperones, under conditions of ER stress (Brewer, 2014). Expression of a constitutively active version (ATF6Ac) of this effector forces cells to increase their protein folding capacity. Accordingly, we reasoned that ATF6Ac enhanced SEAP titer by alleviating a bottleneck in product folding. Unsurprisingly, ATF6Ac (nor any other effector) had no effect on SEAP production when used in conjunction with the standard reference expression cassette (data not shown), as the level of product expression driven by this construct (i.e > 3-fold lower than the optimized expression cassette) was already efficiently handled by the chassis' existing ER capacity (Hansen et al., 2017). We therefore concluded that i) maximizing SEAP transcription, translation, and translocation resulted in an excess of unfolded polypeptide, but this load was not sufficient to induce substantial ER stress, and ii) triggering an artificial unfolded protein response (UPR) facilitated conversion of this material to finished product, thereby increasing SEAP yield. While the other eleven components had no effect on SEAP production, overexpression of discrete

chaperones/foldases, and activation of alternative UPR arms, have been shown to be beneficial in other settings (Baeshen et al., 2015; Delic et al., 2014; Fischer et al., 2015). Although unpredictable context-dependent factors (e.g. product-type, process conditions) prevent rational or generic engineering of protein folding machineries to improve product titers, as shown here, an optimal, system-specific solution can be determined by screening a wide variety of diverse genetic effectors.

The final optimized chassis-expression cassette combination increased SEAP yield by 924% compared to the SRS. As summarized in Figure 5, synthetic modification of the entire product biosynthetic pathway facilitated substantially greater SEAP titers than engineering the cell, vector, or discrete process steps in isolation. These results therefore validate our design strategy to use minimal palettes of synthetic components (just 31 in this example) to derive optimal protein biomanufacturing systems by simultaneously maximizing both cellular capacity and product biosynthetic flux. To confirm that the system assembled via small-scale, high-throughput screening maintained its functionality in more industrial-like production contexts, we evaluated its performance in a 6-day fed-batch process in mini-bioreactors. This analysis showed that the optimized system achieved a similarly high increase in SEAP production (12.37-fold) relative to the SRS in this new environmental context. While we assume that synthetic chassis-expression cassette combinations will also maintain their functionalities in stable expression platforms, we did not definitively prove this by creating and testing clonal cell lines. Indeed, we reasoned that this time-consuming process, typically employed to enhance product expression in mammalian systems, may be unnecessary if transient production yields can be increased by an order of magnitude (although we note that for biopharmaceutical manufacturing this would require regulatory approval of transient expression approaches).

Concluding remarks

The design process that we have presented here for creating optimal expression cassette-chassis combinations should be applicable to diverse bioproduction contexts. Given the universality of the

core product biosynthetic pathway, the design-test-build methodology described could be used in conjunction with a wide range of commonly-utilized biomanufacturing host cell-types. However, larger component palettes may be required for more complex protein structures, including common therapeutic product-types such as multi-subunit proteins (e.g. monoclonal antibodies) and "difficult to express" non-natural, engineered protein formats (e.g. fusion proteins). For these products, it will likely be beneficial to use genetic components with more diverse quantitative outputs (e.g. a wide range of transcriptional activities) in order to enable identification of optimal, protein-specific ratios of biosynthetic process rates and/or polypeptide chain expression levels.

System outputs may be further increased by including additional part-types in component-design palettes. For example, non-coding RNAs and chemical effectors can be rationally selected to enhance cellular biomass capacity and/or recombinant protein production per cell (Allen et al., 2008; Fischer et al., 2014). Further, for some products it will be beneficial to employ genetic effectors of additional cellular processes. Indeed, while we have focussed on maximizing recombinant protein yield, product quality can be significantly enhanced via glycosylation engineering (Costa et al., 2014; Wang et al., 2018). Accordingly, where glycosylation profiles are known determinants of key product attributes, such as bioactivity or serum half-life, appropriate genetic effectors could be utilized (i.e. in conjunction with an expression cassette-chassis combination that maximizes product titer) to achieve desirable, user-defined glyco-modifications. Moreover, while we determined suitable effector gene dosages from a limited range of discrete concentration points, system performance may be optimized by using auto-regulatory feedback loops to precisely fine-tune effector expression levels. For example, UPR stress-responsive promoters could be employed to dynamically coordinate production of protein folding machinery components with ER capacity limitations (Ruijter et al., 2016). Finally, where HT robotic screening systems are available, optimal synthetic part-assemblies may be identified more rapidly by exploring the entire component-combination design space in a single experiment. By

simultaneously testing 1000s of promoter-signal peptide-protein coding sequence-chassis-genetic effector compositions, ideal biomanufacturing systems could be derived in a matter of weeks.

Acknowledgements

This work was supported by MedImmune and the Biotechnology and Biological Sciences Research Council (grant no. BB/K020986/1).

References

Allen, M. J., Boyce, J. P., Trentalange, M. T., Treiber, D. L., Rasmussen, B., Tillotson, B., . . . Reddy, P. (2008). Identification of novel small molecule enhancers of protein production by cultured mammalian cells. Biotechnology and Bioengineering, 100(6), 1193-1204.

Baeshen, M. N., Al-Hejin, A., Bora, R. S., Ahmed, M., Ramadan, H., Saini, K. S., . . . Redwan, E. M. (2015). Production of biopharmaceuticals in E. coli: current scenario and future perspectives. Journal of Microbiology and Biotechnology, 25, 953-962.

Barash, S., Wang, W., & Shi, Y. (2002). Human secretory signal peptide description by hidden Markov model and generation of a strong artificial signal peptide for secreted protein expression. Biochemical and Biophysical Research Communications, 294(4), 835-842.

Bervoets, I., Van Brempt, M., Van Nerom, K., Van Hove, B., Maertens, J., De Mey, M., & Charlier, D. (2018). A sigma factor toolbox for orthogonal gene expression in Escherichia coli. Nucleic Acids Research, 46, 2133-2144.

Borkowski, O., Ceroni, F., Stan, G.-B., & Ellis, T. (2016). Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. Current Opinion in Microbiology, 33, 123-130.

Borth, N., Mattanovich, D., Kunert, R., & Katinger, H. (2005). Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line. Biotechnology Progress, 21(1), 106-111.

Brewer, J. W. (2014). Regulatory crosstalk within the mammalian unfolded protein response. Cellular and Molecular Life Sciences, 71(6), 1067-1079.

Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T., & Eggert, T. (2006). Systematic screening of all signal peptides from Bacillus subtilis: a powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria. Journal of Molecular Biology, 362(3), 393-402.

Brown, A. J., Gibson, S. J., Hatton, D., & James, D. C. (2017). In silico design of context-responsive mammalian promoters with user-defined functionality. Nucleic Acids Research, 45(18), 10906-10919.

Brown, A. J., & James, D. C. (2015). Precision control of recombinant gene transcription for CHO cell synthetic biology. Biotechnology Advances, 34, 492-503.

Cain, K., Peters, S., Hailu, H., Sweeney, B., Stephens, P., Heads, J., . . . Dickson, A. (2013). A CHO cell line engineered to express XBP1 and ERO1-Lα has increased levels of transient protein expression. Biotechnology Progress, 29(3), 697-706.

Cardinale, S., & Arkin, A. P. (2012). Contextualizing context for synthetic biology–identifying causes of failure of synthetic biological systems. Biotechnology journal, 7(7), 856-866.

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., . . . Werner, T. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics, 21(13), 2933-2942.

Ceroni, F., Algar, R., Stan, G.-B., & Ellis, T. (2015). Quantifying cellular capacity identifies gene expression designs with reduced burden. Nature Methods, 12(5), 415.

Ceroni, F., Boo, A., Furini, S., Gorochowski, T. E., Borkowski, O., Ladak, Y. N., . . . Ellis, T. (2018). Burden-driven feedback control of gene expression. Nature Methods, 15, 387-393.

Chin, J. X., Chung, B. K.-S., & Lee, D.-Y. (2014). Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design. Bioinformatics, 30(15), 2210-2212.

Chung, B. K.-S., Yusufi, F. N., Yang, Y., & Lee, D.-Y. (2013). Enhanced expression of codon optimized interferon gamma in CHO cells. Journal of Biotechnology, 167(3), 326-333.

Chung, J. Y., Lim, S. W., Hong, Y. J., Hwang, S. O., & Lee, G. M. (2004). Effect of doxycycline-regulated calnexin and calreticulin expression on specific thrombopoietin productivity of recombinant chinese hamster ovary cells. Biotechnology and Bioengineering, 85(5), 539-546.

Costa, A. R., Rodrigues, M. E., Henriques, M., Oliveira, R., & Azeredo, J. (2014). Glycosylation: impact, control and improvement during therapeutic protein production. Critical Reviews in Biotechnology, 34(4), 281-299.

Davy, A. M., Kildegaard, H. F., & Andersen, M. R. (2017). Cell factory engineering. Cell systems, 4(3), 262-275.

De Jong, H., Geiselmann, J., & Ropers, D. (2017). Resource reallocation in bacteria by reengineering the gene expression machinery. Trends in Microbiology, 25(6), 480-493.

Delic, M., Göngrich, R., Mattanovich, D., & Gasser, B. (2014). Engineering of protein folding and secretion—strategies to overcome bottlenecks for efficient production of recombinant proteins. Antioxidants & redox signaling, 21(3), 414-437.

Dong, H., Nilsson, L., & Kurland, C. G. (1995). Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. Journal of Bacteriology, 177(6), 1497-1504.

Dragosits, M., Nicklas, D., & Tagkopoulos, I. (2012). A synthetic biology approach to self-regulatory recombinant protein production in Escherichia coli. Journal of Biological Engineering, 6(1), 2.

Even, D. Y., Kedmi, A., Basch-Barzilay, S., Ideses, D., Tikotzki, R., Shir-Shapira, H., . . . Juven-Gershon, T. (2016). Engineered promoters for potent transient overexpression. PloS One, 11(2), e0148918.

Fath, S., Bauer, A. P., Liss, M., Spriestersbach, A., Maertens, B., Hahn, P., . . . Wagner, R. (2011). Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. PloS One, 6(3), e17596.

Fernandez-Martell, A., Johari, Y. B., & James, D. C. (2018). Metabolic phenotyping of CHO cells varying in cellular biomass accumulation and maintenance during fed-batch culture. Biotechnology and Bioengineering, 115(3), 645-660.

Fischer, S., Buck, T., Wagner, A., Ehrhart, C., Giancaterino, J., Mang, S., . . . Handrick, R. (2014). A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells. Biotechnology journal, 9(10), 1279-1292.

Fischer, S., Handrick, R., & Otte, K. (2015). The art of CHO cell engineering: A comprehensive retrospect and future perspectives. Biotechnology Advances, 33(8), 1878-1896.

Florin, L., Pegel, A., Becker, E., Hausser, A., Olayioye, M. A., & Kaufmann, H. (2009). Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells. Journal of Biotechnology, 141(1-2), 84-90.

Gaspar, P., Oliveira, J. L., Frommlet, J., Santos, M. A., & Moura, G. (2012). EuGene: maximizing synthetic gene design for heterologous expression. Bioinformatics, 28(20), 2683-2684.

Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodríguez-Carmona, E., Baumann, K., . . . Lang, C. (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microbial cell factories, 7(1), 11.

Gibson, S., Bond, N., Milne, S., Lewis, A., Sheriff, A., Pettman, G., . . . Hatton, D. (2017). N-terminal or signal peptide sequence engineering prevents truncation of human monoclonal antibody light chains. Biotechnology and Bioengineering, 114(9), 1970-1977.

Gilman, J., & Love, J. (2016). Synthetic promoter design for new microbial chassis. Biochemical Society Transactions, 44(3), 731-737.

Gorochowski, T. E., Avcilar-Kucukgoze, I., Bovenberg, R. A., Roubos, J. A., & Ignatova, Z. (2016). A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes. ACS synthetic biology, 5(7), 710-720.

Gould, N., Hendy, O., & Papamichail, D. (2014). Computational tools and algorithms for designing customized synthetic genes. Frontiers in bioengineering and biotechnology, 2, 41.

Güler-Gane, G., Kidd, S., Sridharan, S., Vaughan, T. J., Wilkinson, T. C., & Tigue, N. J. (2016). Overcoming the refractory expression of secreted recombinant proteins in mammalian cells through modification of the signal peptide and adjacent amino acids. PloS One, 11(5), e0155340.

Gustafsson, C., Govindarajan, S., & Minshull, J. (2004). Codon bias and heterologous protein expression. Trends in Biotechnology, 22(7), 346-353.

Hansen, H. G., Pristovšek, N., Kildegaard, H. F., & Lee, G. M. (2017). Improving the secretory capacity of Chinese hamster ovary cells by ectopic expression of effector genes: lessons learned and future directions. Biotechnology Advances, 35(1), 64-76.

Haryadi, R., Ho, S., Kok, Y. J., Pu, H. X., Zheng, L., Pereira, N. A., . . . Yang, Y. (2015). Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. PloS One, 10(2), e0116878.

Hetz, C., Chevet, E., & Oakes, S. A. (2015). Proteostasis control by the unfolded protein response. Nature Cell Biology, 17(7), 829.

Hetz, C., & Papa, F. R. (2017). The unfolded protein response and cell fate control. Molecular Cell, 69, 169-181.

Hoffmann, F., & Rinas, U. (2004). Stress induced by recombinant protein production in Escherichia coli. Advances in Biochemical Engineering and Biotechnology, 89, 73-92.

Johari, Y. B., Estes, S. D., Alves, C. S., Sinacore, M. S., & James, D. C. (2015). Integrated cell and process engineering for improved transient production of a “difficult-to-express” fusion protein by CHO cells. Biotechnology and Bioengineering, 112, 2527-2542.

Jossé, L., Smales, C. M., & Tuite, M. F. (2010). Transient expression of human TorsinA enhances secretion of two functionally distinct proteins in cultured Chinese hamster ovary (CHO) cells. Biotechnology and Bioengineering, 105(3), 556-566.

Juven-Gershon, T., Cheng, S., & Kadonaga, J. T. (2006). Rational design of a super core promoter that enhances gene expression. Nature Methods, 3(11), 917.

Khoo, S. H. G., & Al-Rubeai, M. (2009). Detailed understanding of enhanced specific antibody productivity in NS0 myeloma cells. Biotechnology and Bioengineering, 102(1), 188-199.

Kim, H. J., Lee, S. J., & Kim, H.-J. (2010). Optimizing the secondary structure of human papillomavirus type 16 L1 mRNA enhances L1 protein expression in Saccharomyces cerevisiae. Journal of Biotechnology, 150(1), 31-36.

Kober, L., Zehe, C., & Bode, J. (2013). Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnology and Bioengineering, 110(4), 1164-1173.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Research, 15(20), 8125-8148.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science, 324(5924), 255-258.

Lee, D.-H., Gershenzon, N., Gupta, M., Ioshikhes, I. P., Reinberg, D., & Lewis, B. A. (2005). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. Molecular and Cellular Biology, 25(21), 9674-9686.

Lee, Y. Y., Wong, K. T., Tan, J., Toh, P. C., Mao, Y., Brusic, V., & Yap, M. G. (2009). Overexpression of heat shock proteins (HSPs) in CHO cells for extended culture viability and improved recombinant protein production. Journal of Biotechnology, 143(1), 34-43.

Mahalik, S., Sharma, A. K., & Mukherjee, K. J. (2014). Genome engineering for improved recombinant protein expression in Escherichia coli. Microbial cell factories, 13(1), 177.

Marbach-Bar, N., Bahat, A., Ashkenazi, S., Golan-Mashiach, M., Haimov, O., Wu, S.-Y., . . . Levine, A. J. (2015). DTIE, a novel core promoter element that directs start site selection in TATA-less genes. Nucleic Acids Research, 44(3), 1080-1094.

Møller, T. S., Hay, J., Saxton, M. J., Bunting, K., Petersen, E. I., Kjærulff, S., & Finnis, C. J. (2017). Human β-defensin-2 production from S. cerevisiae using the repressible MET17 promoter. Microbial cell factories, 16(1), 11.

Mori, A., Hara, S., Sugahara, T., Kojima, T., Iwasaki, Y., Kawarasaki, Y., . . . Nakano, H. (2015). Signal peptide optimization tool for the secretion of recombinant protein from Saccharomyces cerevisiae. Journal of Bioscience and Bioengineering, 120(5), 518-525.

Nikel, P. I., Martínez-García, E., & De Lorenzo, V. (2014). Biotechnological domestication of pseudomonads using synthetic biology. Nature Reviews Microbiology, 12(5), 368.

O'Callaghan, P. M., McLeod, J., Pybus, L. P., Lovelady, C. S., Wilkinson, S. J., Racher, A. J., . . . James, D. C. (2010). Cell line-specific control of recombinant monoclonal antibody production by CHO cells. Biotechnology and Bioengineering, 106(6), 938-951.

Obst, U., Lu, T. K., & Sieber, V. (2017). A modular toolkit for generating Pichia pastoris secretion libraries. ACS synthetic biology, 6(6), 1016-1025.

Parret, A. H., Besir, H., & Meijers, R. (2016). Critical reflections on synthetic gene design for recombinant protein expression. Current Opinion in Structural Biology, 38, 155-162.

Pasotti, L., & Zucca, S. (2014). Advances and computational tools towards predictable design in biological engineering. Computational and Mathematical Methods in Medicine, 2014, 369681.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods, 8(10), 785.

Pybus, L. P., Dean, G., West, N. R., Smith, A., Daramola, O., Field, R., . . . James, D. C. (2014). Model-directed engineering of “difficult-to-express” monoclonal antibody production by Chinese hamster ovary cells. Biotechnology and Bioengineering, 111(2), 372-385.

Raab, D., Graf, M., Notka, F., Schödl, T., & Wagner, R. (2010). The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. Systems and Synthetic Biology, 4(3), 215-225.

Raveh, A., Margaliot, M., Sontag, E. D., & Tuller, T. (2016). A model for competition for ribosomes in the cell. Journal of The Royal Society Interface, 13(116), 20151062.

Redden, H., Morse, N., & Alper, H. S. (2014). The synthetic biology toolbox for tuning gene expression in yeast. FEMS Yeast Research, 15, 1.

Reese, M. G., Eeckman, F. H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in Genie. Journal of Computational Biology, 4(3), 311-323.

Reeve, B., Hargest, T., Gilbert, C., & Ellis, T. (2014). Predicting translation initiation rates for designing synthetic biology. Frontiers in bioengineering and biotechnology, 2, 1.

Ruijter, J. C., Koskela, E. V., & Frey, A. D. (2016). Enhancing antibody folding and secretion by tailoring the Saccharomyces cerevisiae endoplasmic reticulum. Microbial cell factories, 15(1), 87.

Schlegel, S., Genevaux, P., & de Gier, J.-W. (2017). Isolating Escherichia coli strains for recombinant protein production. Cellular and Molecular Life Sciences, 74(5), 891-908.

Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., & Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. Science, 330(6007), 1099-1102.

Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D., & Voigt, C. A. (2014). A ‘resource allocator’for transcription based on a highly fragmented T7 RNA polymerase. Molecular Systems Biology, 10(7), 742.

Shrestha, R. P., & Hildebrand, M. (2017). Development of a silicon limitation inducible expression system for recombinant protein production in the centric diatoms Thalassiosira pseudonana and Cyclotella cryptica. Microbial cell factories, 16(1), 145.

Tastanova, A., Schulz, A., Folcher, M., Tolstrup, A., Puklowski, A., Kaufmann, H., & Fussenegger, M. (2016). Overexpression of YY1 increases the protein production in mammalian cells. Journal of Biotechnology, 219, 72-85.

Waegeman, H., & Soetaert, W. (2011). Increasing recombinant protein production in Escherichia coli through metabolic and genetic engineering. Journal of Industrial Microbiology & Biotechnology, 38(12), 1891-1910.

Wang, Q., Chung, C. Y., Chough, S., & Betenbaugh, M. J. (2018). Antibody glycoengineering strategies in mammalian cells. Biotechnology and Bioengineering, 115, 1378-1393.

Weiße, A. Y., Oyarzún, D. A., Danos, V., & Swain, P. S. (2015). Mechanistic links between cellular trade-offs, gene expression, and growth. Proceedings of the National Academy of Sciences, 112(9), E1038-E1047.

Wells, E., & Robinson, A. S. (2017). Cellular engineering for therapeutic protein production: product quality, host modification, and process improvement. Biotechnology journal, 12(1), 1600105.

Wurm, F. M. (2013). CHO quasispecies—implications for manufacturing processes. Processes, 1(3), 296-311.

Xiao, S., Shiloach, J., & Betenbaugh, M. J. (2014). Engineering cells to improve protein expression. Current Opinion in Structural Biology, 26, 32-38.

Zhou, Y., Raju, R., Alves, C., & Gilbert, A. (2018). Debottlenecking protein secretion and reducing protein aggregation in the cellular host. Current Opinion in Biotechnology, 53, 151-157.

Figure legends

Figure 1: Recombinant protein production system output is determined by synthetically engineerable input components. Product yields are a function of five key biosynthetic process rates, each of which can be modified by designing either the chassis or discrete genetic parts. UTR = untranslated region; RBS = ribosome binding site; CDS = coding sequence.

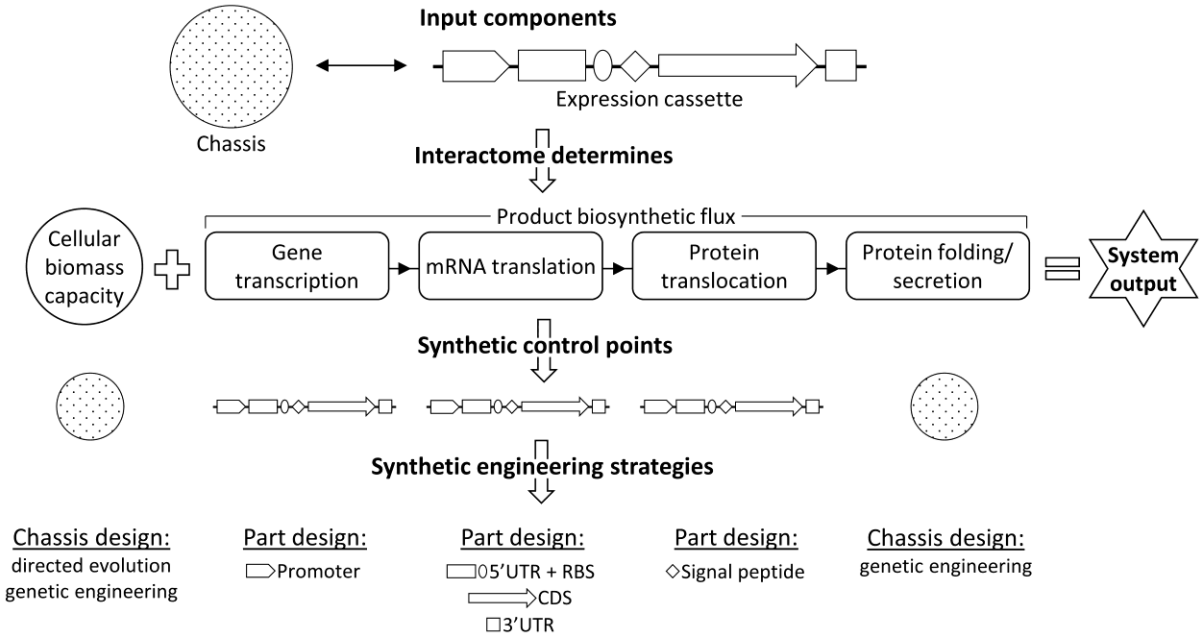
Figure 2: A minimal synthetic component-palette for maximizing SEAP production in CHO cells. Synthetic components were rationally designed and selected to enable derivation of a biomanufacturing system in which both cellular capacity and product biosynthetic flux are maximized **(A)**. A system comprising commonly-utilized, and native, product-specific components was used as a reference standard for evaluating the performance of discrete synthetic expression cassette-chassis combinations **(B)**. UTR = untranslated region; RBS = ribosome binding site; CDS = coding sequence; PKS = perfect Kozak sequence.

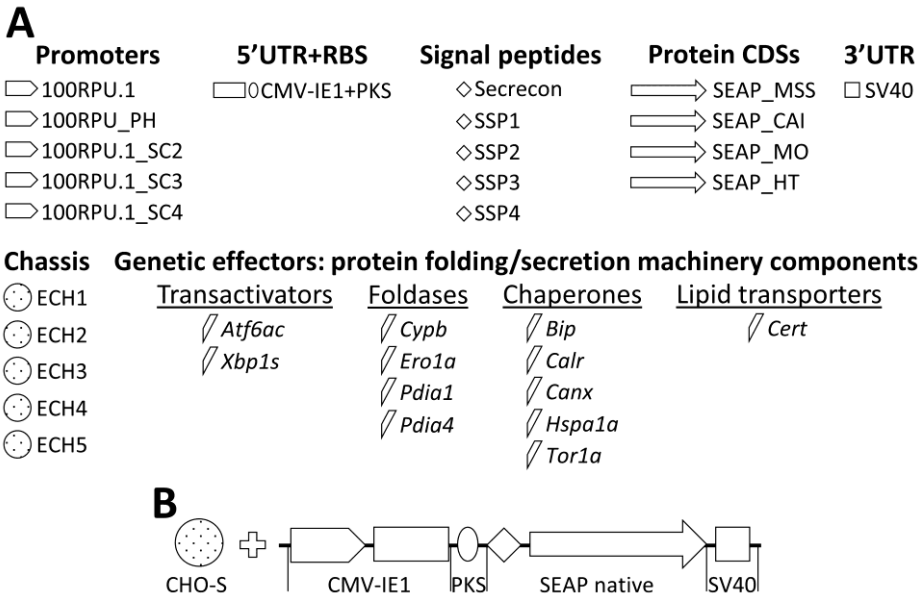
Figure 3: Assembling an expression cassette that maximizes flux through SEAP transcription, translation and translocation. Designed (D) parts were synthesized and individually inserted into a system comprising standard reference (SR) components (see Figure 2). The function of synthetic promoters (panel **A**), signal peptides (panel **B**) and SEAP coding sequences (panel **C**) was evaluated in a 72 hr transient production process, and best-performing parts were subsequently used in combination (panel **D**). Data are expressed as a percentage of the production exhibited by the standard reference system; statistically significant differences are indicated by asterisks (one-way ANOVA followed by Dunnett's test: *P < 0.05, **P < 0.01, ***P < 0.001). Values represent the mean ± SD of three independent experiments (n = 3, each performed in triplicate).

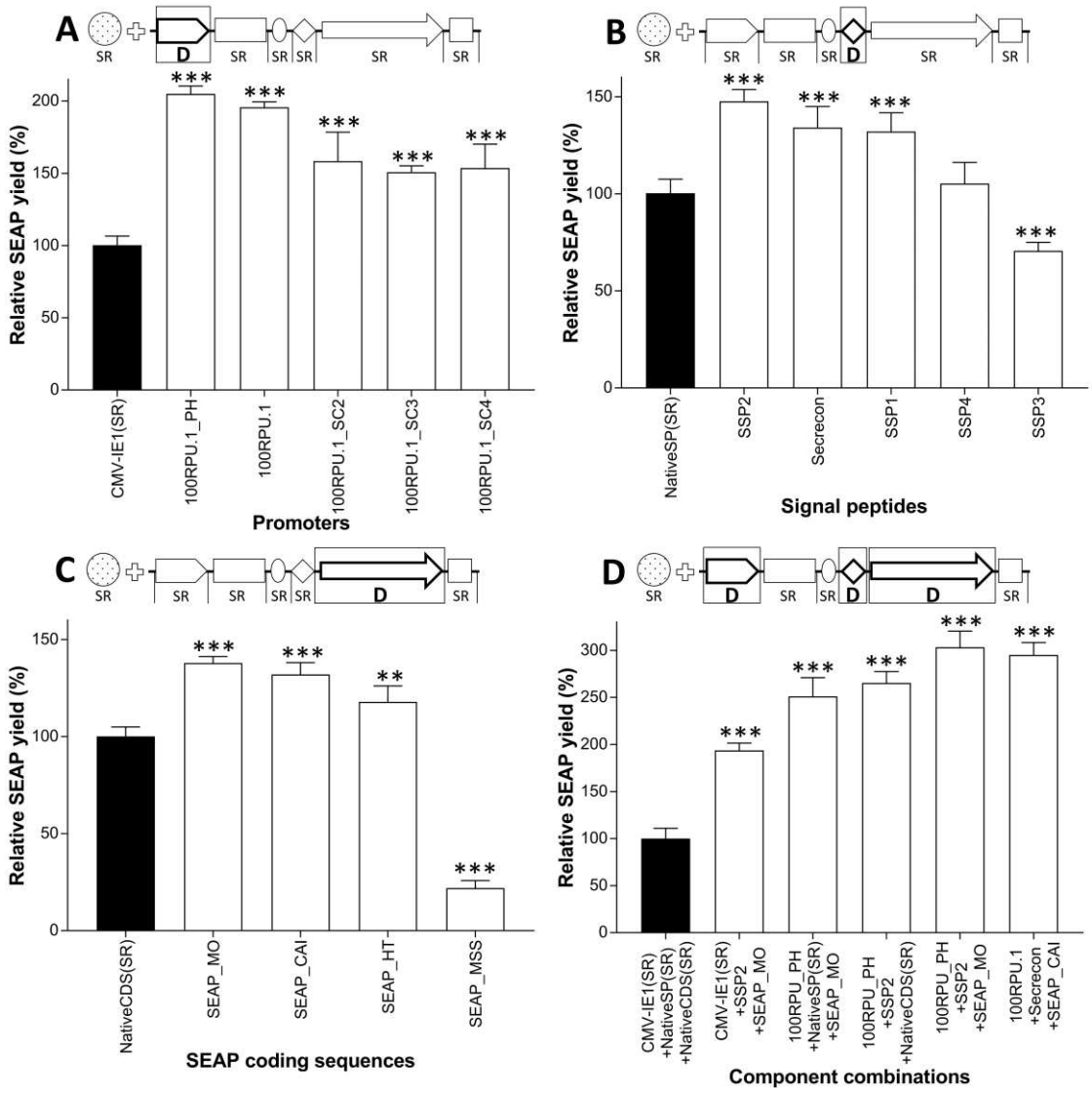
Figure 4: Selecting and engineering a chassis to maximize cellular biomass capacity and cell-specific productivity. A SEAP reporter vector containing an optimal combination of synthetic genetic components (shown at the top of the panel; see Figure 3D) was transiently transfected into varying CHO cell chassis **(A)**. Cell growth was measured at 24 h intervals to calculate integral viable cell concentration (IVCC) values. SEAP production was quantified 72 h post-transfection, and cell-specific productivities (qP) were determined. Data are expressed as a percentage of the production achieved using a standard CHO-S cell line; statistically significant differences are indicated by asterisks (one-way ANOVA followed by Dunnett's test: *P < 0.05, **P < 0.01, ***P < 0.001). The optimized SEAP-

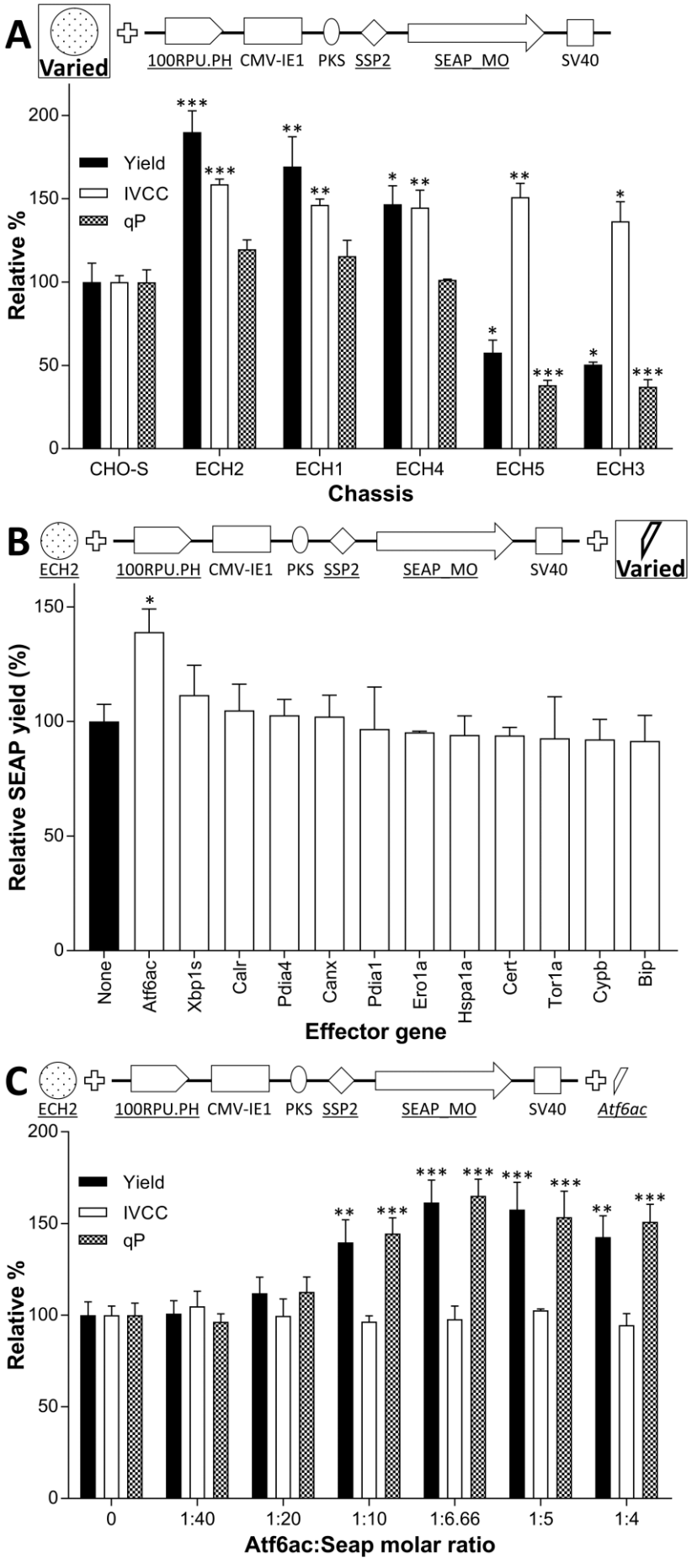
reporter vector was then co-transfected into the best-performing chassis (ECH2) alongside either **B)** varying effector gene expression plasmids at a molar ratio of 10: 1, or **C)** an ATF6Ac expression plasmid at varying molar ratios. Data are expressed as a percentage of the production achieved when the SEAP reporter plasmid was used in isolation; statistically significant differences are indicated by asterisks (one-way ANOVA followed by Dunnett's test: *P < 0.05, **P < 0.01, ***P < 0.001). The total amount of transfected DNA was kept constant in all cases (empty vector was not used to normalize SEAP-reporter plasmid copy numbers). In A, B, and C, values represent the mean ± SD of two, two, and three independent experiments respectively (n = 2, 2, 3, each performed in triplicate).

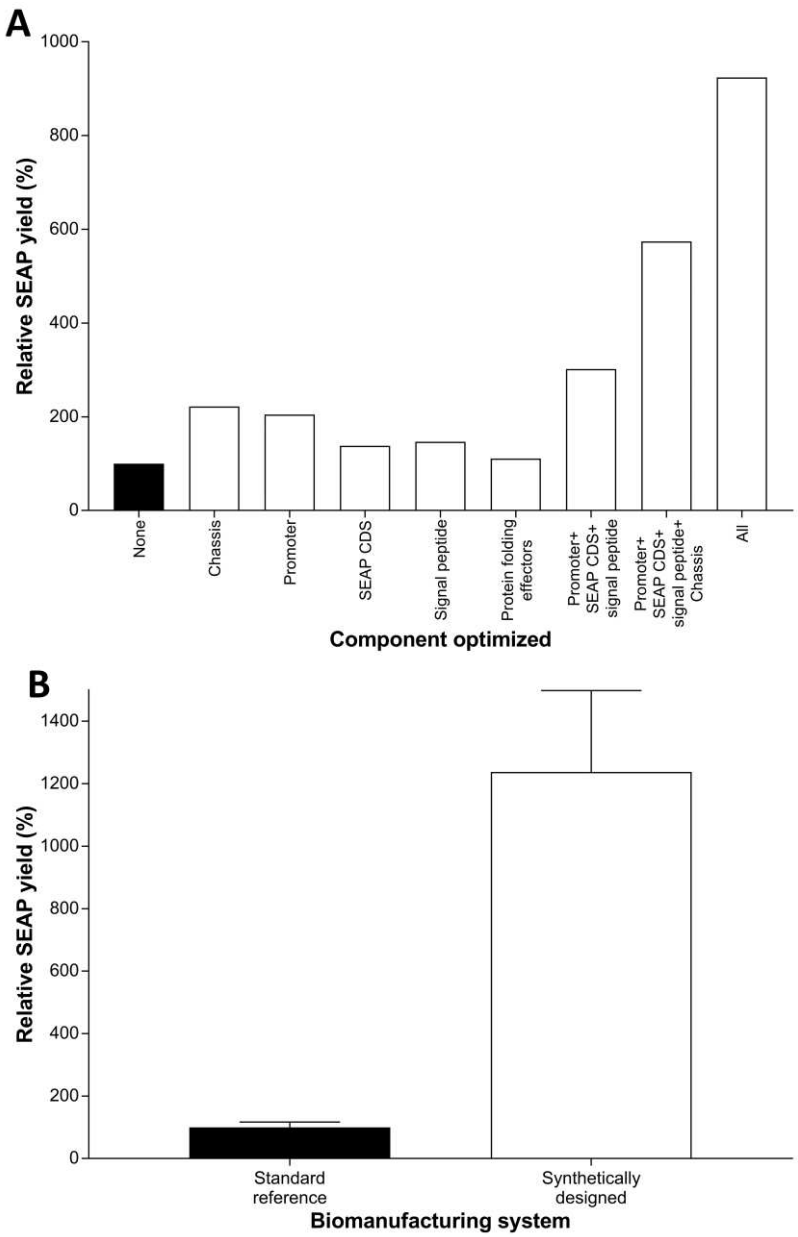
Figure 5: The performance of an optimally designed synthetic expression cassette-chassis combination is maintained in a scaled-up, fed-batch production process. A) Summary of the effect discrete system engineering strategies had on SEAP yield in the context of a 72 h production process in 24-well plates. **B)** A SEAP reporter vector containing an optimal combination of synthetic genetic components, and an ATF6Ac expression plasmid, were co-transfected into ECH2 CHO cells at a molar ratio of 6.66: 1 (system composition = 'all components optimized'). SEAP production was measured over the course of a 6-day fed-batch process in tube-spin bioreactors. Data are expressed as a percentage of the production exhibited by the standard reference system (see Figure 2). Values represent the mean ± SD of two independent experiments (n = 2, each performed in duplicate).


Input components
Chassis
Expression cassette
Interactome determines
Product biosynthetic flux
Cellular biomass capacity
Gene transcription
mRNA translation
Protein translocation
Protein folding/ secretion
System output
Synthetic control points
Synthetic engineering strategies
Chassis design:
directed evolution
genetic engineering
Part design:
Promoter
Part design:
5'UTR + RBS
CDS
3'UTR
Part design:
Signal peptide
Chassis design:
genetic engineering


A
Promoters
5'UTR+RBS
Signal peptides
Protein CDSs
3'UTR
100RPU.1
100RPU_PH
100RPU.1_SC2
100RPU.1_SC3
100RPU.1_SC4
CMV-IE1+PKS
Secrecon
SSP1
SSP2
SSP3
SSP4
SEAP_MSS
SEAP_CAI
SEAP_MO
SEAP_HT
SV40
Chassis
Genetic effectors: protein folding/secretion machinery components
ECH1
ECH2
ECH3
ECH4
ECH5
Transactivators
Atf6ac
Xbp1s
Foldases
Cypb
Ero1a
Pdia1
Pdia4
Chaperones
Bip
Calr
Canx
Hspa1a
Tor1a
Lipid transporters
Cert
B
CHO-S
CMV-IE1
PKS
SEAP native
SV40


A
SR
D
SR
SR
SR
SR
SR
Relative SEAP yield (%)
0
50
100
150
200
***
CMV-IE1(SR)
100RPU.1_PH
100RPU.1
100RPU.1_SC2
100RPU.1_SC3
100RPU.1_SC4
Promoters
B
SR
SR
SR
SR
D
SR
SR
Relative SEAP yield (%)
0
50
100
150
***
NativeSP(SR)
SSP2
Secrecon
SSP1
SSP4
SSP3
Signal peptides
C
SR
SR
SR
SR
SR
D
SR
Relative SEAP yield (%)
0
50
100
150
***
**
NativeCDS(SR)
SEAP_MO
SEAP_CAI
SEAP_HT
SEAP_MSS
SEAP coding sequences
D
SR
D
SR
SR
D
D
SR
Relative SEAP yield (%)
0
50
100
150
200
250
300
***
CMV-IE1(SR) +NativeSP(SR) +NativeCDS(SR)
CMV-IE1(SR) +SSP2 +SEAP_MO
100RPU_PH +NativeSP(SR) +SEAP_MO
100RPU_PH +SSP2 +NativeCDS(SR)
100RPU_PH +SSP2 +SEAP_MO
100RPU.1 +Secrecon +SEAP_CAI
Component combinations


A
Varied
100RPU.PH
CMV-IE1
PKS
SSP2
SEAP_MO
SV40
Yield
IVCC
qP
Relative %
CHO-S
ECH2
ECH1
ECH4
ECH5
ECH3
Chassis
B
ECH2
Varied
Relative SEAP yield (%)
None
Atf6ac
Xbp1s
Calr
Pdia4
Canx
Pdia1
Ero1a
Hspa1a
Cert
Tor1a
Cypb
Bip
Effector gene
C
Atf6ac
Atf6ac:Seap molar ratio
0
1:40
1:20
1:10
1:6.66
1:5
1:4


A
Relative SEAP yield (%)
1000
800
600
400
200
0
None
Chassis
Promoter
SEAP CDS
Signal peptide
Protein folding effectors
Promoter+ SEAP CDS+ signal peptide
Promoter+ SEAP CDS+ signal peptide+ Chassis
All
Component optimized
B
Relative SEAP yield (%)
1400
1200
1000
800
600
400
200
0
Standard reference
Synthetically designed
Biomanufacturing system