Article:
Johari, Y.B., Brown, A.J., Alves, C.S. et al. (2019) CHO genome mining for synthetic promoter design. Journal of Biotechnology, 294. pp. 1-13. ISSN: 0168-1656

https://doi.org/10.1016/j.jbiotec.2019.01.015

Title: CHO Genome Mining for Synthetic Promoter Design

Authors: Yusuf B. Johari, Adam J. Brown, Christina S. Alves, Yizhou Zhou, Chapman M. Wright, Scott D. Estes, Rashmi Kshirsagar, David C. James

PII: S0168-1656(19)30022-7
DOI: https://doi.org/10.1016/j.jbiotec.2019.01.015
Reference: BIOTEC 8350

To appear in: *Journal of Biotechnology*

Received date: 15 November 2018
Revised date: 7 January 2019
Accepted date: 9 January 2019

Please cite this article as: Johari YB, Brown AJ, Alves CS, Zhou Y, Wright CM, Estes SD, Kshirsagar R, James DC, CHO Genome Mining for Synthetic Promoter Design, *Journal of Biotechnology* (2019), https://doi.org/10.1016/j.jbiotec.2019.01.015

CHO Genome Mining for Synthetic Promoter Design

Yusuf B. Johari[a], Adam J. Brown[a], Christina S. Alves[b], Yizhou Zhou[b], Chapman M. Wright[b,1], Scott D. Estes[b,2], Rashmi Kshirsagar[b], David C. James[a,*]

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

[b] Cell Culture Development, Biogen Inc., Cambridge, MA 02142, USA

[1] present address: Generation Bio, Cambridge, MA 02142, USA

[2] present address: Codiak Biosciences, Cambridge, MA 02139, USA

* to whom correspondence should be addressed.

Tel: +44 114 222 7505; E-mail: d.c.james@sheffield.ac.uk

Highlights

- Bioinformatic analysis of endogenous CHO promoter sequences.
- Linking transcriptomic and genomic datasets for bioprocess-directed promoter design.
- *In silico* DOE-based design method for synthetic promoter construction.
- Design promoters exhibited 2.5-fold increase in activity over CMV-IE promoter.
- Promoters designed to function in context of a biphasic fed-batch production process.

Abbreviations: CHO, Chinese hamster ovary; CMV, cytomegalovirus; DOE, design of experiment; FPKM, fragments per kilobase of transcript per million mapped reads; GS,

glutamine synthetase; IVCD, integral viable cell density; ORF, open reading frame; PIC, pre-initiation complex; qP, cell specific production rate; RNA-Seq, next-generation RNA sequencing; SEAP, secreted alkaline phosphatase; SV40, Simian Virus 40; TF, transcription factor; TFRE, transcription factor regulatory element; TSS, transcriptional start site.

Keywords: synthetic promoter; Chinese hamster ovary cells; recombinant protein expression; CHO genome; transcriptomics; synthetic biology.

ABSTRACT

Synthetic promoters are an attractive alternative for use in mammalian hosts such as CHO cells as they can be designed *de novo* with user-defined functionalities. In this study, we describe and validate a method for bioprocess-directed design of synthetic promoters utilizing CHO genomic sequence information. We designed promoters with two objective features, (i) constitutive high-level recombinant gene transcription, and (ii) upregulated transcription under mild hypothermia or late-stage culture. CHO genes varying in transcriptional activity were selected based on a comparative analysis of RNA-Seq transcript levels in normal and biphasic cultures in combination with estimates of mRNA half-life from published genome scale datasets. Discrete transcription factor regulatory elements (TFREs) upstream of these genes were informatically identified and functionally screened *in vitro* to identify a subset of TFREs with the potential to support high activity recombinant gene transcription during biphasic cell culture processes. Two libraries of heterotypic synthetic promoters with varying TFRE combinations were then designed *in silico* that exhibited a maximal 2.5-fold increase in transcriptional strength over the CMV-IE promoter after transient transfection into host CHO-K1 cells. A subset of synthetic promoters was then used to create stable transfectant

pools using CHO-K1 cells under glutamine synthetase selection. Whilst not achieving the maximal 2.5-fold increase in productivity over stable pools harboring the CMV promoter, all stably transfected cells utilizing synthetic promoters exhibited increased reporter production — up to 1.6-fold that of cells employing CMV, both in the presence or absence of intron A immediately downstream of the promoter. The increased productivity of stably transfected cells harboring synthetic promoters was maintained during fed-batch culture, with or without a transition to mild hypothermia at the onset of stationary phase. Our data exemplify that it is important to consider both host cell and intended bioprocess contexts as design criteria in the *de novo* construction of synthetic genetic parts for mammalian cell engineering.

1. INTRODUCTION

Chinese hamster ovary (CHO) cells remain the most widely used cell factory for production of recombinant therapeutic proteins. Improvements in expression vector design and clonal selection methods, combined with cell and process engineering approaches (e.g. use of small-molecules) have led to product titers in the multiple grams per liter range (Kunert and Reinhart 2016; Zhou et al. 2018). Of cellular synthetic processes underpinning cell specific production rate, recombinant transgene transcription has one of the largest impacts (O'Callaghan et al. 2010). Strong recombinant gene transcription in CHO cells is still dominated by viral sequences such as the cytomegalovirus immediate-early promoter (CMV-IE) despite potentially undesirable attributes such as cell-cycle dependency and epigenetic silencing (Brightwell et al. 1997; Kim et al. 2011). Endogenous promoters such as the eIF1α promoters are typically relatively large which can limit transfection efficiencies or use in multigene vectors. For example, the Chinese hamster EF1α promoter (Running-Deer and Allison 2004) supports a high level of expression but requires the use of two distinct multi-kb elements.

Various efforts have been carried out to improve promoter characteristics in CHO cells. Examples include insertion of a core CpG island element (Mariati et al. 2014) and a CHO-K1E77 regulatory element (Kang et al. 2016) into CMV which resulted in enhanced stable expression. Other attempts to engineer natural promoters for increased production are exemplified via the creation of "super core" by incorporating extra elements into the CMV core promoter (Juven-Gershon et al. 2006) or the use of "dual core" by insertion of an additional core sequence into CHO S100a6 promoter (Thaisuchat et al. 2011), as well as via truncation of a CHO endogenous promoter (Chen et al. 2013). However, these modifications do not eliminate the drawbacks of currently used viral promoters and engineered endogenous promoters, nor do they radically expand the set of existing mammalian cell engineering tools. In this context, synthetic promoters are an attractive alternative as they can be custom-designed to engineer recombinant gene transcription predictably, removing functionally ill-defined and uncontrollable elements in expression vectors. Indeed, we have previously reported synthetic promoters that offer stable, precise control of recombinant transcriptional activity in CHO cells (Brown et al. 2014, 2017). These genetic constructs were designed by (i) analysis of CHO cell transcription factor (TF) expression profiles, and (ii) determination of the transcription factor regulatory element (TFRE) compositions of commonly utilized viral promoters. However, no previous studies have analyzed CHO genomic sequences to identify active sequence motifs associated with endogenous promoters. We hypothesized that mining CHO genomic information (e.g. Xu et al. 2011) would identify new transcriptionally active TFREs to enable the design of synthetic promoters with novel functionalities.

Whilst it is known that thousands of endogenous promoters function together to define the transcriptional landscape of mammalian cells (Carninci et al. 2006), it is still a considerable challenge to identify specific genomic regulatory elements that may be utilized to control expression of heterologous recombinant genes in a specific context. For example,

CHO cells undergo physiological transitions during a cell culture process associated with discrete alterations in cellular transcriptional activity. Thus we aimed to create a workflow based on bioprocess-relevant CHO genomic information to identify homologous regulatory elements that could be used to design synthetic promoters specifically useful for CHO cell based production processes. Synthetic promoters for mammalian systems have been constructed from endogenous (human) sequence components, but the expression levels observed were still below or only comparable to that of CMV (Cheng and Alper 2016), and there is no information on how we can specifically design promoters to function in a biomanufacturing setting (e.g. a biphasic process). The data from Cheng and Alper (2016) implied that each host cell system requires a bespoke synthetic promoter for optimal expression. In this regard, recent advancements in next-generation sequencing technology and CHO reference genome annotations (Wright and Estes 2014) represent a novel tool for construction of synthetic promoters with functionally predictable characteristics.

In this study we describe a procedure to mine CHO genomic information to design synthetic promoters that can support high-level, biphasic recombinant protein production. Through systematic bioinformatic analysis of 'omic data streams and *in vitro* screening we identified novel CHO endogenous TFREs that are potentially active under different bioprocess conditions. Based on this, we utilized an *in silico* experimental design strategy to efficiently construct synthetic promoters that harness CHO cell transcriptional capacity. We were able to construct, for the first time, synthetic promoters exhibiting upregulated activity during a biphasic production process, generating an increased titer of recombinant protein over that obtained using a CMV promoter. Additionally, analysis of stable transfectants revealed that synthetic promoters can be designed with minimal "promoter–promoter interference" that are compatible with the use of introns. Whilst this study demonstrates effective *de novo* creation of optimized, bespoke promoters for CHO cell based

biomanufacturing, we anticipate that similar approaches could be used to create synthetic promoters for diverse other applications requiring expression of recombinant genes in mammalian cells.

2. MATERIALS AND METHODS

2.1 RNA-Seq and *in silico* Analysis of Transcription Factor Regulatory Elements

RNA-Seq FASTQ data of duplicate CHO-K1SV fed-batch bioreactor cultures at different time points (Days 4, 9 and 11) and under different conditions (constant 37°C and a temperature shift to 32°C at Day 7) were provided by Biogen (Cambridge, MA). Galaxy (usegalaxy.org) and R software were used to analyze the RNA-Seq data according to Trapnell et al. (2012) using CHO-K1 GFF3 and FASTA files (RefSeq 2014) from www.CHOgenome.org as well as Biogen's in-house CHO-K1 GTF and FASTA files. The transcriptional activity, T_a, of a gene was calculated as follows:

$$T_a = k_d m = (\ln 2 / t_{1/2}) m \qquad \text{(Eq. 1)}$$

where k_d is the mRNA decay rate constant (h^{-1}), m is the mRNA transcript abundance (FPKM) and $t_{1/2}$ is the mRNA half-life (h). k_d and $t_{1/2}$ values were derived from Schwanhäusser et al. (2011) and Sharova et al. (2009) (see below and Supplementary Figure S1). Transcriptional start sites (TSSs) were determined using annotated 5'UTRs and Genomatix Gene Regulation software suite (MatInspector Release 8.2 and MatBase Version 9.3; Genomatix, Munich, Germany) was used to analyze the region –1000 to +200 relative to the TSS against Genomatix-defined mouse promoter background to find putative transcription factor regulatory elements (TFREs).

2.2 TFRE-Reporter Vector Construction

A minimal hCMV-IE1 core promoter from the human cytomegalovirus (CMV) was synthesized (Eurofins Genomics, Ebersberg, Germany), PCR amplified (Phusion high-fidelity DNA polymerase; NEB, Hitchin, UK), and purified (QIAquick PCR Purification kit; Qiagen, Crawley, UK). The PCR products were then digested with HindIII and EcoRI enzymes (NEB), gel extracted (QIAquick Gel Extraction kit; Qiagen) and inserted directly upstream of the secreted alkaline phosphatase (SEAP) open reading frame (ORF) of a promoterless pSEAP2-Basic vector (Clontech, Oxford, UK). The CMV core promoter sequence used was as follows: 5'-AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGAC-3'.

To create TFRE reporter plasmids, synthetic oligonucleotides containing 6× repeat copies of the TFRE consensus sequences in Table 1 were synthesized, PCR amplified and inserted into KpnI and HindIII sites upstream of the CMV core promoter. To create core promoter TFRE reporter plasmids, the CMV core promoter sequence between +19 and +39 relative to the TSS was replaced by MTE and DPE consensus sequences (Table 1). Mutation of TATA-box was performed by replacing TATATAA with ACGTCCG. For example, the modified core promoter containing mutated TATA-box, MTE and DPE (changes underlined) was as follows: 5'-AGGTCACGTCCGGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCGAGCGGAGCAGACGTGCCTCCATAGAAGAC-3'. Core promoters were synthesized and inserted directly upstream of the SEAP ORF of a promoterless pSEAP2-Basic vector as described above. Clonally derived plasmids were purified using a QIAGEN Plasmid Midi kit (Qiagen). The sequence of all plasmid constructs were confirmed by DNA sequencing.

2.3 Synthetic Promoter Construction

To create synthetic promoters, synthetic genes containing combinations of specific TFREs were designed *in silico* (Design-Expert 10; Stat-Ease, Minneapolis, MN) using two-level factorial designs (reduced two-factor interaction models; Table 2). The positions of the TFRE blocks within the promoters were randomly arranged using R software in forward orientation of 5' DNA strand. Synthetic genes were synthesized (Eurofins Genomics) and inserted into KpnI and HindIII sites upstream of the CMV core promoter. A full length CMV promoter containing the same CMV core (−601 to +50 relative to the TSS) was also synthesized and inserted into KpnI and EcoRI sites upstream of the SEAP ORF. The pSEAP2-Control vector harboring a Simian Virus (SV40) early promoter/enhancer was obtained from Clontech. The sequence of all plasmid constructs was confirmed by DNA sequencing.

2.4 CHO Cell Culture

A $GS^{-/-}$ CHO-K1SV cell line was provided by Biogen. Cells were cultured in Biogen proprietary medium supplemented with 6 mM L-glutamine (Sigma, Poole, UK) in Erlenmeyer flasks (Corning, Acton, MA) maintained at 37°C, 140 rpm under 5% CO_2. Sub-cultures were seeded at 2×10^5 viable cells mL^{-1} every 3–4 days. Cell concentration and viability were measured using a Vi-CELL XR (Beckman Coulter, Brea, CA).

2.5 Transient Transfection

Transfections were conducted using an Amaxa Cell Line Nucleofector Kit V system (Lonza, Basel, Switzerland) as previously described (Johari et al. 2015). 5×10^6 cells per cuvette were electroporated with 4–6 μg DNA and transferred to a TubeSpin bioreactor tube (TPP, Trasadingen, Switzerland) containing 10 mL pre-warmed culture media. Transfected cells were cultured for 48 h at 37°C, 170 rpm under 5% CO_2. To determine transfection efficiency, cells were transfected with pmaxGFP vector (Lonza) using the same procedure and analyzed using Attune Acoustic Focusing Cytometer (Thermo Fisher, Waltham, MA).

2.6 Generation of Stable Pools

A pEE12.4 vector containing an SV40 promoter-driven glutamine synthetase (GS) gene was provided by Biogen. To create the stable expression vectors, the SEAP gene was cloned by PCR (Q5 high-fidelity 2× master mix; NEB) and inserted into BsiWI and EcoRI sites, the minimal CMV core promoter was cloned without or with an 898 bp CMV intron A downstream of the core and inserted into HindIII and BsiWI sites upstream of the SEAP ORF, and the synthetic genes of the synthetic promoters were cloned and inserted into BsrGI and HindIII sites upstream of the CMV core promoter. All vector constructs were confirmed by sequencing. Stable transfections were conducted using circular plasmids and Amaxa Cell Line Nucleofector Kit V system as described above. Transfected cells were transferred to a TubeSpin bioreactor tube containing 10 mL glutamine-free media and cultured at 37°C, 170 rpm under 5% CO_2 with media change every 3–4 days until the cell viability reached >95% and then cryopreserved.

2.7 Fed-Batch Stable Production

CHO cells were revived and cultured as described above for three passages prior to fed-batch production. Cells were seeded at 4×10^5 viable cells mL^{-1} into 125 mL Erlenmeyer flasks at a working volume of 28 mL using Biogen proprietary medium supplemented with anti-clumping agent (Life Technologies) at 1:250 (v/v) dilution. Fed-batch cultures were maintained at 37°C, 140 rpm under 5% CO_2 and 75% humidity for 10–11 days until cell viability dropped below 80%. Daily nutrient supplementation was performed using a total of 35–40% (v/v) Biogen proprietary feed throughout culture. For cultures maintained at reduced culture temperature, cells were shifted to 32°C on Day 6.

2.8 Determination of Recombinant SEAP Production

Prior to analysis, culture medium was filtered through a 0.45 µm Spin-X centrifuge tube filter (Sigma). SEAP concentration was measured using SensoLyte pNPP Alkaline Phosphatase Assay kit (AnaSpec, Fremont, CA) according to the manufacturer's instructions. The cell specific production rate (qP) was calculated as described in O'Callaghan et al. (2010).

2.9 Measurement of Recombinant SEAP mRNA Copy Number

1.5×10^6 viable cells were collected by centrifugation at 200×g for 5 min. Cell pellets were immediately resuspended in 300 µL of RNAprotect Cell Reagent (Qiagen) and stored at 4°C. Total RNA was extracted using RNeasy Plus Mini kit in combination with QIAshredder homogenizer (Qiagen) according to the manufacturer's instructions. gDNA-free RNA was converted to cDNA and quantified by one-step RT-PCR using an Applied Biosystems 7500 Fast Real-Time PCR system (Thermo Fisher) by the relative quantification method. Reverse transcription and PCR reactions were performed in a final volume of 25 µL using Power SYBR Green RNA-to-C_T 1-Step kit (Thermo Fisher), containing 0.5 µL of RNA and 100 nM of forward and reverse primer each. Primers were designed using Applied Biosystems Primer Express 3.0 software (Thermo Fisher) as follows: SEAP Fwd 5'-CCATATGTGGCTCTGTCCAA-3', Rev 5'-GTCTGGAAGTTGCCCTTGAC-3'; GS Fwd 5'-CGCAGAGATCCCAACAAGCT-3', GS Rev 5'-TGCAGGCTTCCGGTTGTACT-3'; and GAPDH Fwd 5'-TGCCACCCAGAAGACTGTG-3', Rev 5'-GTGGATGCAGGGATGATGTT-3'. The PCR thermal cycle profile was as follows: 48°C for 30 min; 95°C for 10 min; 40 cycles of 95°C for 15 s, and 60°C for 1 min; followed by melt curve analysis to confirm the specificity of amplification. Relative SEAP gene expression was determined using the $\Delta\Delta C_t$ method normalized to endogenous GAPDH mRNA levels.

3. RESULTS

3.1 Bioinformatic Identification of Transcription Factor Regulatory Elements in CHO-K1 Endogenous Promoters

The level of transcriptional activation mediated by a discrete promoter sequence is largely determined by its unique composition of TFREs, and the availability of cognate TFs. Accordingly, design rules for constructing synthetic elements with desired functionalities can be obtained by either (i) determination of host cell TF repertoires under appropriate experimental conditions, or (ii) analysis of TFRE compositions of endogenous promoters exhibiting suitable expression dynamics. Whilst synthetic promoters have been created for CHO cells by analysis of TF expression levels (Brown et al. 2017), CHO genomic (i.e. endogenous TFRE) information has never been utilized, potentially limiting opportunities for synthetic promoter design. We therefore devised a methodical, comprehensive approach for informed design of synthetic promoters based on CHO genomic sequence information (summarized in Figure 1). This work flow enables identification of endogenous gene promoters (and associated TFREs) with desirable functional features, thus allowing *de novo* design of synthetic promoters with related, bioprocess-relevant characteristics. To profile endogenous gene expression in CHO cells, we generated RNA-Seq data (Figure 1, Step 1) using CHO-K1SV fed-batch bioreactor cultures sampled at Days 4, 9 and 11 post-inoculation with and without a shift to 32°C at Day 7. Bioinformatic analysis (Figure 1, Step 2) was performed with two objectives; (i) to identify genes with high transcriptional activity (FPKM/h) at different culture phases/temperature, and (ii) to identify potentially active CHO genomic TFREs upstream of genes varying in transcriptional activity. Based on RNA-Seq differential gene expression analysis of ~27,600 annotated CHO-K1 genes, most changes in mRNA transcript abundance ($q < 0.05$) were observed between log (Day 4) and stationary

(Day 9) phases of culture, where 13.2% of the genes were upregulated and 13.8% were downregulated. In contrast, comparison of mRNA abundance in cells under hypothermic or normal culture conditions (Day 9, 32°C vs. Day 9, 37°C) identified only relatively minor changes in gene expressions (2.6% upregulated and 3.6% downregulated). There were no apparent changes in mRNA abundance (0.07–0.66% up/downregulated) between early stationary (Day 9) and late stationary (Day 11) phases for both 37°C and 32°C cultures.

With respect to the second objective (identification of active endogenous TRFEs), it was necessary to quantify the relative transcriptional activity of CHO genes. As the abundance of a given mRNA is a function of both gene transcription and mRNA decay rates (Cheadle et al. 2005; where the latter can vary between ~1 and 30 h), in order to compare the transcriptional activity of CHO genes we estimated CHO mRNA stabilities based on published data for other mammalian cells, as none exists for CHO cells at the genome scale. Averaged mRNA half-lives deriving from empirical analyses of murine NIH3T3 and embryonic stem cells from Schwanhäusser et al. (2011) and Sharova et al. (2009) respectively were used to estimate and rank the relative transcriptional activities (expressed as FPKM/h; Equation 1) of ~7,600 CHO genes. We note that there was a very significant (albeit weak) correlation between the two half-life datasets ($r = 0.492$, $p < 2.2\times10^{-16}$; Supplementary Figure S1). We acknowledge that it may be expected that (in general) mRNA stability will change for cells exposed to mild hypothermia. However, in the absence of any genome-scale datasets we assumed for the purpose of this analysis that the stability of CHO mRNAs at 32°C was generally proportionate to that observed at 37°C.

Further analysis of the top 1% of genes with high transcriptional activity revealed that most exhibited a decrease from log to stationary phase but exhibited no significant change when shifted from 37°C to 32°C (data not shown). In order to identify TFREs that were also active at stationary phase *and* under hypothermia, the top 200 genes from each

phase/condition (i.e. Day 4; Day 9, 37°C; and Day 9, 32°C) were selected. A subset of 50 genes that occurred in all three datasets was then selected to form a high transcriptional activity group. Corresponding medium and low transcriptional activity groups were created by selection of the mid 50 genes (exhibiting a median level of transcriptional activity) and bottom 50 genes with $T_a > 0.0005$ FPKM/h of the ranked genes, respectively. Supplementary Table S1 lists the selected genes in each group. We note that the largest fold change increases in transcriptional activity (during log to stationary, or 37°C to 32°C) occurred in genes with very low transcriptional activity (which still remained low), hence providing less useful information for design of active synthetic promoters (data not shown).

For each selected CHO gene, a transcriptional start site (TSS) was estimated based on annotated CHO 5'UTR sequences. We note that available *in silico* TSS prediction tools were largely inaccurate (Abeel et al. 2009) whilst experimentally verified annotation of CHO TSSs (e.g. using RNA-Seq method; Jakobi et al. 2014) was still in infancy. Relative to the TSS a −1000 to +200 bp segment was extracted from the CHO genomic sequences for putative TFRE analysis. TFRE identification (Figure 1, Step 3) was performed using the Genomatix Overrepresented TFBS (MatInspector/MatBase) tool against a murine promoter background with the promoter Z-score threshold set to >2.0 ($p < 0.05$). The result revealed 144, 93 and 72 discrete TFREs in the high, medium and low transcriptional activity groups respectively. To minimize false positives (i.e. inactive TFREs; Figure 1, Step 4) in the high activity group, we filtered out TFREs that also occurred in the medium/low activity group (Figure 2) as well as non-conserved TFREs that occurred only in ≤10% of genes. To further TFRE pool complexity, we selected only up to 2 TFREs with the highest Z-scores from each TF family. Table 1 lists the final set of 32 TFREs incorporated into the functional screen, where 29 TFREs were enhancer elements and 3 TFREs were core promoter elements. The TF

matrix/family, frequency and Z-score of the selected TFREs are detailed in Supplementary Table S2.

3.2 Determination of Transcription Factor Regulatory Element Activity in CHO Cells

The relative transcriptional activity of each enhancer TFRE (Table 1) in CHO-K1SV cells was determined as previously described (Brown et al. 2014) using a SEAP reporter construct that contained six repeat copies of a specific TFRE in series, upstream of a minimal hCMV-IE1 core promoter (–34 to +50 relative to the TSS, containing a TATA box and an Inr motif). Transient transfection of plasmid DNA by nucleofection yielded transfection efficiency of ~94% (measured using a vector encoding GFP, flow cytometry data not shown). Measurement of SEAP expression deriving from transfection of each TFRE-reporter plasmid after 48 h culture post-transfection is shown in Figure 3A. This analysis identified 11 TFREs that significantly increased expression (>5-fold) over basal expression from the minimal core promoter (NFκB, NFκB-p65, GABPβ, DMP1, AhR/ARNT, USF1, STAT3, Sp1, ZBED1, Pax3 and EGR-1). Amongst these, only 3 TFREs — NFκB, USF1 (E-box) and Sp1 (GC-box) have been identified for our previous viral- and TF-derived synthetic promoter constructs (where the former was the most active element in CHO cells; Brown et al. 2014, 2017). Therefore identification of CHO genomic TRFEs significantly expanded the CHO synthetic promoter design toolbox. As there were no correlations with respect to the specific TFRE frequency of occurrence in the high activity group (Supplementary Table S2), we infer that gene transcriptional activity is largely a function of the context-specific mechanism of discrete TF-mediated transcriptional activation and/or affinity of a given TF for its cognate TFRE. In summary, this screening exercise yielded a pool of transcriptionally active TFREs that were subsequently used to construct synthetic promoters.

Regulatory elements within core promoter regions (typically −50 to +50 relative to the TSS) bind cognate general TFs to form transcription pre-initiation complexes (PICs). It was hypothesized that optimized core promoter functionality would be achieved by maximizing core regulatory element numbers, where these elements function synergistically to increase transcription initiation rates (Juven-Gershon and Kadonaga 2010). To investigate the use of the core promoter to achieve high expression level, we created and characterized modified CMV core promoters that contain TATA, Inr, MTE and/or DPE motifs. The synthetic MTE and DPE elements were incorporated into core promoters with strictly maintained spacing from Inr, i.e. at +19 and +30 relative to the TSS (Kutach and Kadonaga 2000; Lim et al. 2004). As TATA boxes were not identified in our bioinformatic analysis, we also created TATA-less core promoters by mutating the element as described in Lim et al. (2004). The core promoters were each subcloned into a promoterless vector with a SEAP reporter gene. The core promoter activity (in absence of enhancer elements) after 48 h culture is shown in Figure 3B. This analysis shows that the MTE+DPE core did not significantly improve the basal expression of CMV core with <2-fold increment observed (compared to up to 63-fold increase achieved by enhancer elements), whilst mutating the TATA box led to a 55% reduction in core basal expression. When combined with a CMV enhancer, the best modified core promoter (MTE+DPE) yielded a similar SEAP productivity compared to the wild-type CMV core (data not shown). This is possibly due to TF–TFRE interactions within enhancers decreasing the influence of core-mediated processes such as PIC formation and transcription initiation (Juven-Gershon and Kadonaga 2010). We conclude that the efficacy of core promoter engineering is much more limited compared to enhancer region engineering, and that synthetic core promoters would need to exhibit considerably improved functionality to replace the CMV core in synthetic promoter libraries.

3.3 Construction of Strong Synthetic Promoters using Design-of-Experiments Reveals CHO Intrinsic Transcriptional Capacity

Given our previous finding that mammalian synthetic promoters can be constructed according to relatively simple design rules (Brown et al. 2017), we hypothesized that we could build promoters (i.e. TFRE combinations) *in silico,* obviating the requirement to screen hundreds of randomly assembled TFRE compositions. To test this, a two-level fractional factorial design was employed to systematically construct synthetic promoters containing different combinations of TFRE blocks. This design was chosen as it assumes that interaction effects between two or more variables are small as compared to main effects, and importantly allows analysis of many different factors (TFREs) using a limited number of experiments (promoter variants) which can then be projected into stronger designs using a subset of the significant factors (Carlson and Carlson 2005). In order to create strong synthetic promoters we utilized the TFREs identified as transcriptionally active in CHO-K1SV cells (Figure 3A) with bias towards highly active TFREs. To reduce the design space only one TFRE from each family was utilized (e.g. NFκB-p65 subunit was omitted), as well as the less active ZBED1, Pax3 and EGR-1 TFREs. With respect to the former, we reasoned that the use of TFREs from different TF families may have advantageous "modular" combinatorial effects (Cartharius et al. 2005), whereas NFκB and its p65 subunit might induce transcription in a similar manner. Based on our previous work we also employed the following design rules: (i) maximum copy number of any discrete TFRE = 6 (i.e. to minimize off-target effects on host cell performance (Brown et al. 2017), (ii) copy number of Sp1 per promoter = 1 (small quantities of Sp1 having been shown to support high transcriptional activities within heterogeneous promoter architectures (Brown et al. 2014)), and (iii) the relative order of constituent TFREs was random, separated by minimal spacers (Brown et al. 2017).

Nine synthetic promoter constructs with the varying TFRE compositions shown in Table 2 (Library 1) were chemically synthesized and inserted upstream of the minimal CMV core promoter in SEAP reporter plasmids. A control CMV promoter reporter plasmid was constructed using the CMV-IE1 promoter (–601 to +50 relative to the TSS, i.e., the complete hCMV-IE1 enhancer containing the distal, proximal and core promoter regions, henceforth referred to as CMV) upstream of the SEAP ORF. The commonly used Simian Virus 40 (SV40) early promoter/enhancer (henceforth referred to as SV40) was also tested as a comparison. To ensure that SEAP production kinetics were not affected by recombinant DNA overload (Johari et al. 2015), cells were transfected with a 33% lower DNA dosage compared to the TFRE screening process. The transient SEAP production and mRNA data (Figure 4A) shows that all Library 1 synthetic promoters yielded very strong reporter expression, where the least active promoter (1/01) achieved a cell specific SEAP production rate (qP) similar to that observed for the CMV control vector which is over twice the size of 1/01. The strongest Library 1 synthetic promoter (1/09) exhibited a 2.5-fold or 8.3-fold increase in transcriptional activity compared to the CMV and SV40 promoters respectively. Again, this promoter (473 bp) is 25% smaller than CMV, indicative of far more efficient transcriptional activation under these assay conditions. However, the data in Figure 4A also indicate that synthetic promoter mediated transcription becomes saturated as it approaches a level ~2.5-fold greater than that exhibited by the CMV promoter. For example, promoter 1/09, which contains 50% more TFRE blocks than promoter 1/05 (Table 2), displayed only a 16% increase in expression.

To investigate the relative activity, as well as synergism between TFREs we performed a design-of-experiment (DOE) analysis using ‘logit transformation’ to emulate saturation of reporter expression by setting the upper bound transcriptional output to a value equivalent to 2.5-fold the activity of the CMV promoter. Table 3 (Library 1) summarizes the

reduced two-factor interaction model employed with an inconsequential NFκB–STAT3 interaction parameter removed to improve model fit. Highly significant model was obtained ($p < 0.001$) for this first library with an insignificant "lack of fit" indicating comparable variance of modeled and empirical data. Additionally, the model achieved a good agreement between the predicted and adjusted R^2 values (<0.2 difference) validating the reliability of the model in predicting the output response to the combination of effectors. The model for this first library derived from the relative qP analysis in terms of TFRE block numbers is as follows:

$$qP = 0.93(NF\kappa B) + 0.90(GABP\beta) + 0.59(DMP1) + 0.23(AhR/ARNT) + 0.18(USF1) + 0.07(STAT3) + 1.20 \quad \text{(Eq. 2)}$$

The positive parameter coefficients indicate that all TFREs functioned as transactivators of reporter gene expression. Additionally, the statistical analysis confirmed that NFκB, GABPβ and to a lesser extent DMP1 significantly influenced expression ($p < 0.05$; Table 3, Library 1) although it is possible that these were contributed by positive interactions of two discrete TFREs (see the alias structure in Table 3). In contrast, AhR/ARNT, USF1 and STAT3 were insignificant factors ($p > 0.05$; partly due to the fewer block numbers) which also indirectly indicate that there were no significant interactions among the strong TFREs (NFκB, GABPβ, DMP1) or among the weak TFREs (AhR/ARNT, USF1, STAT3).

Derived from the analysis of Library 1 promoters, we created a second library of promoters in an effort to augment the transcriptional activity which could be achieved by further biasing the components towards highly significant (active) TFREs. Specifically, eight promoters were constructed using a two-level fractional factorial design with the block number of highly significant NFκB and GABPβ increased to seven, while the number of DMP1 and AhR/ARNT blocks made constant at mid-levels (4 and 2 block copies, respectively) and Sp1 omitted. Two additional promoter constructs (2/09 and 2/10) favoring

NFκB, GABPβ, DMP1 and AhR/ARNT were also constructed. Library 2 synthetic promoters were created as described previously and their TFRE block compositions are listed in Table 2. The relative transcriptional activity of Library 2 promoters is shown in Figure 4B. These promoters exhibited increased activity with the mean expression level (relative to CMV) shifted from 1.9-fold for first library to 2.1-fold for Library 2 promoters. However, as per Library 1, reporter expression saturated at ~2.5-fold CMV, implying that the synthetic promoters had reached the limit of CHO-intrinsic transactivation capacity under these conditions. The model for the second library promoters derived from the relative qP analysis in terms of TFRE block numbers is as follows:

$$qP = 1.46(NF\kappa B) + 1.21(GABP\beta) + 0.41(USF1) + 0.41(STAT3) + 0.57 \quad \text{(Eq. 3)}$$

The statistical analysis (Table 3, Library 2) confirmed that NFκB and GABPβ were the most active contributors to the gene expression (with no confounding factors). USF1 and STAT3 remained insignificant factors although these elements were still required for strong "basal" expression relative to CMV. Despite the higher model resolution (alias structure), the analysis also indicated that the TFRE interaction parameters were insignificant and thus were removed in order to improve model fitness. Combining all observations made above, we conclude that promoter strength is primarily a function of (i) transcriptional activation mechanism of a specific TFRE, and (ii) copy number of a specific TFRE — corroborating our previous TF-based promoter designs (Brown et al. 2017).

3.4 Analysis of Synthetic Promoter Function for Stable Recombinant Protein Production

In order to determine if synthetic promoters can be deployed in a stable expression format and operate effectively under different culture conditions/phases, we generated stably transfected pools of CHO cells and evaluated their relative functional activity through a

biphasic fed-batch production process. Stable transfectant pools (rather than clonally derived populations) were employed to compare the relative stable activity of synthetic promoters as they provide a population average readout of synthetic promoter transactivation with less interference from integration-specific effects and clone-to-clone phenotypic variation. Further, previous studies have shown that the use of 3' intron immediately downstream of transcriptional start site can augment CMV promoter-driven recombinant protein production (Mariati et al. 2010) where they function as translation-enhancing elements (Skoko et al. 2011). We therefore examined the effects of a wild-type CMV promoter intron A when coupled with synthetic promoters for stable SEAP expression. We hypothesized that intronic sequence downstream of a synthetic promoter influenced transcriptional activity and/or gene expression generally. A panel of four strong synthetic promoters with varying TFRE composition were selected (specifically promoters 1/05, 1/08, 1/09 and 2/08), and these were inserted into the pEE12.4 stable vector backbone encoding SEAP either with or without an 898 bp intron A immediately downstream of the promoter. Vector constructs encoded glutamine synthetase (GS) under the control of an SV40 promoter as the selectable marker in the $GS^{-/-}$ (knockout) host cell line. We note that promoter 2/10, despite its distinctive composition (Table 2), could not be tested for stable performance due to protracted gene synthesis turnaround (~3 months) that arose from sequence-specific DNA synthesis difficulties. This emphasizes a potentially unforeseen issue in synthetic biology, i.e. difficulty in chemical synthesis when contextual sequence is varied.

Four transfection pools for each promoter were created, with host cell viability 2 h post-transfection ranging from 79 to 83% (data not shown). During the selection process in glutamine-free medium, we observed that all cell populations that employed synthetic promoters had a better recovery rate compared to both CMV populations, either with or without intron A (Figure 5A). Whilst cell viability across all stable pools was similar at 24 h

post transfection (50±6%), CHO cell populations transfected with constructs harboring synthetic promoters achieved a ~26% higher viability at 7 days post-transfection than cells transfected with CMV promoter constructs and were fully recovered by Day 13 (viability >95%) compared to CMV pools which were recovered by Day 15.

Stable SEAP production during a 4-day batch culture was measured after the fourth passage post-cell revival. Whilst nearly all cell populations transfected with synthetic promoters exhibited stable SEAP production in excess of that observed from cells stably transfected with the CMV promoters, the extent to which SEAP production was increased relative to that deriving from CMV containing vectors was not as substantial as previously observed for transiently transfected cells (Figure 4). For example, SEAP production from cell transfected with synthetic promoter 1/09(−intron A) was only 1.6-fold that of cells transfected with CMV(−intron A) in stable mode rather than 2.5-fold that of CMV in transient mode (Figure 4). Nevertheless, three of the synthetic promoters tested (1/08, 1/09 and 2/08) still demonstrated a significant increase in reporter production over CMV. Comparison of SEAP production by cells transfected with promoters upstream of intron A (+) compared to their intronless (−) counterparts using a two-tailed *t*-test revealed no significant differences in all cases ($p \geq 0.21$; Figure 5B). This was not completely unexpected as previous studies showed that although CMV(+intron A) generally performed better, the effects were also protein-dependent (Chapman et al. 1991). Whilst this study suggests that posttranscriptional SEAP mRNA processing was not a critical factor affecting productivity under these conditions, it also indicates that introns may be employed downstream of synthetic promoters with no detrimental effects.

As stable productivity mediated by synthetic promoters (relative to that mediated by CMV) was below that predicted by transient transfection, coupled with the observation that cells transfected with synthetic promoters recovered growth more rapidly, we tested the

hypothesis that the more complex CMV promoter repressed SV40 mediated transcription of the GS selection marker via promoter-promoter interference (Curtin et al. 2008) to a greater extent than synthetic promoters. This would result in reduced GS mRNA in CMV-transfected cells associated with more stringent coupling of recombinant gene transcription to cell growth (Fan et al. 2013). Therefore, we further quantified SEAP and GS mRNA levels in a subset of stable transfectants (all lacking intron A) by qRT-PCR. This analysis (Figure 5C) revealed that GS mRNA copy number was generally about 96% less than SEAP mRNA copy number when these two promoters were utilized together to create stable transfectants. Even considering potential differences in SEAP and GS mRNA stability, this is significantly less GS mRNA than would be expected. For example, where CMV and SV40 promoters are tested separately, SV40 mediated expression is typically about one third of that obtained with CMV (e.g. Figure 4A). However, the abundance of GS mRNA in all synthetic promoter stable transfectant pools was significantly higher (between 34 to 63%) than in CMV-derived stable transfectants (Figure 5C). From this we infer, as we have previously reported (Davies et al. 2011), that the higher level of GS expression resulted in less stringency of stable transfectant selection, thus the better general recovery of cells transfected with synthetic promoters post-transfection but relatively less stable expression (relative to CMV) than would have been predicted by transient expression. Our data do not suggest that synthetic promoters do not interfere at all with other co-located promoters, but that the significant added complexity of the "natural" CMV promoter likely results in more pronounced interference.

Lastly, we hypothesized that progression of CHO cells through a production process, with the associated dynamic variation in cell physiology may alter the relative abundance of endogenous TFs affecting synthetic promoter activity thus affecting synthetic promoter performance. In order to determine if the synthetic promoters displayed improved

biomanufacturing performance, we evaluated CMV, 1/09 and 2/08 stable transfectant pools (−intron A) for their relative functional capability through an 11-day fed-batch production process with and without a shift to 32°C at Day 6 post-inoculation (Figure 6). Stable pools utilizing synthetic promoters appeared to have slightly lower integral viable cell density (IVCD) compared to CMV although this was largely insignificant (<9%; $p \geq 0.10$), suggesting that the use of synthetic promoters specific for a small number of TFs may compete with the host cell genome for transactivation, thus reducing host cell proliferation and density. Despite the reduction in IVCD, the synthetic promoters were still more advantageous, particularly promoter 1/09 where it achieved 72% and 38% increases in qP at log phase (Day 0–6) and stationary phase (Day 6–9) respectively under normal culture condition. This resulted in a 1.5-fold increase in product titer (Figure 6B). We surmise the late stage culture performance was supported by DMP1 and Sp1 as these TFREs were found exclusively in a subset of 7 genes with log-to-stationary upregulation (Supplementary Table S1, high transcriptional activity group) during bioinformatic analysis (data not shown). Further, qP of cells stably transfected with promoters 1/09 and 2/08 increased when the culture temperature was reduced to 32°C (Figure 6B), demonstrating for the first time that synthetic promoters can be designed to specifically function in a biphasic production process. Again, despite the reduction in IVCD, the promoters achieved up to 1.3-fold increase in titer compared to the CMV 32°C culture, validating our selection of TFRE combinations that were designed to maximize productivity. These data also substantiate our general hypothesis that promoter activity can be specifically regulated by designing endogenous sequences to bind TFs with dynamic expression profiles. We anticipate that synthetic promoters constructed with heterogeneous mixture of TFREs specifically screened for activity in stationary phase and hypothermic conditions (Figure 1, Step 5), coupled with an appropriate (weak)

synthetic/SV40 promoter driving a selection marker, would significantly increase recombinant protein production.

4. DISCUSSION

In this study we have mined CHO genomic information to construct synthetic promoters designed specifically to function in a biphasic production process. A key feature of our work is the incorporation of whole bioprocess performance into synthetic promoter design. This method enables the construction of optimized, bespoke promoters from a vast design space for different biomanufacturing and synthetic biology applications using mammalian cells. For example, RNA-Seq analysis has revealed that different chemical inducers (sodium butyrate, hexamethylene bisacetamide, caffeine) have distinct impact on endogenous gene expression of CHO cells (Fomina-Yadlin et al. 2015) and we anticipate that bioinfomatic analysis of these transcriptomic datasets would potentially yield TFRE pools that are particularly useful for designing inducible gene-expression systems. The major challenge with such *de novo* designs is the difficulty in identifying TFREs underpinning the more complex regulation governing endogenous gene expression systems (e.g. synergistic activation; Matsusaka et al. 1993) which would require massive screens of "di-TFRE" activities under specific bioprocess conditions, preventing rational selection or determination of their functionality. However, as the quality and volume of CHOmics data continues to increase, and, given the progressive development of TFRE database and analysis tools, finding such TF interactions *in silico* may indeed be tractable.

The synthetic promoters presented in this study exhibited saturation of activity at 2.5-fold over CMV, which we speculate (based on this and previous studies) is the maximal level of CHO specific recombinant gene transcription. This is likely as CMV itself is a relatively potent promoter where recombinant mRNA transcripts have been shown to account 16% of

total mRNA in the cell, and a combination of inducer treatment and hypothermic condition increased the expression level to up to 45% of total mRNA (Fomina-Yadlin et al. 2015). Whilst the latter is not surprising due to radical changes in the cells "transcriptional landscape", the former is also a very significant proportion considering that CHO cells express thousands of genes. Further, expression of many cell growth and housekeeping genes appears to negatively correlate to recombinant transcript levels (Fomina-Yadlin et al. 2015; Sha et al. 2018) suggesting that the use of recombinant gene promoter resulted in titration of TFs away from endogenous genes affecting their expression levels. Therefore, the strength of a promoter may be effectively constrained by the effects it imposes on cellular functions. This could partially explain the observed slight reduction in cell growth and peak VCD of cells employing the synthetic promoters (Figure 6A) as well as the general inverse relationship between cell specific growth rate and qP (Dinnis and James 2005; Johari et al. 2015).

In addition to promoter-proximal elements and enhancers, endogenous gene expression is regulated through the integrated action of various *cis*-regulatory modules such as silencers, insulators and tethering elements. However, many of these elements are not required for recombinant protein manufacturing *in vitro* and contribute to unpredictable/uncontrollable gene expression (e.g., Thaisuchat et al. 2011). Specifically designed, rather than evolved, our synthetic promoters do not have such functionally redundant elements, permitting precise regulation of recombinant transcriptional activity in CHO cells via specific TFRE stoichiometry and with relatively small promoter size. Moreover, *de novo* design of promoters also offers additional advantages beyond high level transcriptional activity. For example we indirectly showed that synthetic promoters can be designed with reduced promoter interference effects by constructing promoters containing distinct TFRE types, making them highly desirable in applications such as multigene vectors.

Although we did not deliberately design and test our promoters for expression stability in stable transfectants, our previous work (Brown et al. 2017) has also demonstrated that it is entirely feasible to minimize the number of CpG dinucleotides within each construct, thus preventing the formation of methylation-mediated epigenetic silencing associated with production instability (Kim et al. 2011). Further, as gene silencing can also be caused by deletion of DNA segments via homologous recombination (Jasin and Rothstein 2013), TFREs can be organized in particular configurations to specifically avoid the occurrence of repeat sequences.

Numerous families of TFs have been identified, and their activities are regulated by a variety of mechanisms including interactions with other TF proteins (TF–TF) or DNA binding elements (TF–TFRE), as well as the posttranslational modification state of the TFs (Lambert et al. 2018; Pan et al. 2010). Our comparative transient expression analyses revealed that the synthetic promoter activity was primarily a function of a promoter's relative composition of NFκB, GABPβ and DMP1 TFREs (Tables 2 and 3). However, considering the composition of promoters 1/09 and 2/08, our stable expression data did not support the conclusion that NFκB, GABPβ and DMP1 TFRE blocks alone could support high transcriptional activity in biphasic fed-batch production processes — clearly a combination of a range of TFREs at the appropriate stoichiometry is necessary. For example the relatively weaker promoter 2/08 was highly biased towards NFκB and GABPβ and lacked Sp1 sites. NFκB, a transcription factor associated with uncontrolled proliferation and suppressed apoptosis in cancer cells (Dolcet et al. 2005) is likely to be downregulated in stationary phase and was reported not to be induced by mild hypothermia in HeLa cells (Eskla et al. 2018). Unsurprisingly, bioinformatic analysis in this study indicated that endogenous promoters with high transcriptional activity throughout the bioproduction process contained more discrete TFRE types (Figure 2), whereas CMV contains at least 12 identifiable TFRE types (Brown et

al. 2015) making it (sub-optimally) active in stationary phase and hypothermic condition. Therefore, whilst as few as 4 TFRE types are required to achieve high transcriptional power under a specific condition (e.g. promoter 2/10; Brown et al. 2014), increased promoter complexity may confer additional advantages in dynamic bioprocess conditions.

Lastly, the synthetic promoter design methodology described here is applicable to other industrially relevant cell lines or the development of novel gene therapies with higher specific activity. With regard to the latter, given the fact that TFREs are derived from endogenous components, it may possible to design promoters entirely *in silico* from 'omic datasets, obviating the need to test and characterize each TFRE *in vitro* using model cells (which themselves may have altered transcriptional landscape compared to cells *in vivo*). With regard to the former, the lack of high-quality, annotated reference genome sequences of such cell lines can be circumvented by using reference genome homologs. For instance, transcriptomic analyses of HEK293 (Lin et al. 2014) and BHK cell lines (Johnson et al. 2014) against human and Chinese hamster genomes, respectively, revealed a high degree of homology. Further, production cell lines share some major phenotypic traits, e.g. their acquired ability to proliferate indefinitely and adapt to suspension culture. Therefore, the regulation mechanisms of housekeeping/major genes utilized in this study (e.g. Actb, Gapdh, Hspa5) are likely to be comparable across different cell line hosts and consequently we expect there will be significant TFRE overlaps between them. Through *in vitro* use of parallel high-throughput screening techniques, tens of TFRE parts can be characterized simultaneously, enabling confident prediction of their functionality. Accordingly, we should now be able to equip any mammalian cell with new machinery and processing capability optimally suited for a specific intended purpose. Underpinning this advance is our ability to expand the promoter design space using the emergent tools of 'omic and synthetic biology,

whilst combinatorial, context-dependent empirical modeling will further assist the construction of the optimal promoter for a specific application.

Declarations of interest: None.

The authors declare that there is no conflict of interest.

Acknowledgements: This work was supported by Biogen. The authors thank Joseph Longworth (University of Sheffield), Bryan St. Germain (Biogen) and Joost Groot (Biogen) for their technical support.

REFERENCES

Abeel, T., Van de Peer, Y., Saeys, Y., 2009. Toward a gold standard for promoter prediction evaluation. Bioinformatics 25, i313–i320. https://doi.org/10.1093/bioinformatics/btp191

Brightwell, G., Poirier, V., Cole, E., Ivins, S., Brown, K.W., 1997. Serum-dependent and cell cycle-dependent expression from a cytomegalovirus-based mammalian expression vector. Gene 194, 115–123. https://doi.org/10.1016/S0378-1119(97)00178-9

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 Res. 45, 10906-10919. https://doi.org/10.1093/nar/gkx768

Brown, A.J., Sweeney, B., Mainwaring, D.O., James, D.C., 2014. Synthetic promoters for CHO cell engineering. Biotechnol. Bioeng. 111, 1638–1647. https://doi.org/10.1002/bit.25227

Brown, A.J., Sweeney, B., Mainwaring, D.O., James, D.C., 2015. NF-κB, CRE and YY1 elements are key functional regulators of CMV promoter-driven transient gene expression in CHO cells. Biotechnol. J. 10, 1019–1028. https://doi.org/10.1002/biot.201400744

Carlson, R., Carlson, J.E., 2005. Two-level fractional factorial designs, in: Rutan, S., Walczak, B. (Eds.), Design and optimization in organic synthesis. Elsevier B.V., Amsterdam, pp. 119–168. https://doi.org/10.1016/S0922-3487(05)80010-3

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engström, P.G., Frith, M.C., Forrest, A.R., Alkema, W.B., Tan, S.L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S.M., Wells, C.A., Orlando, V., Wahlestedt, C., Liu, E.T., Harbers, M., Kawai, J., Bajic, V.B., Hume, D.A., Hayashizaki, Y., 2006. Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet. 38, 626–635. https://doi.org/10.1038/ng1789

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., Werner, T., 2005. MatInspector and beyond: promoter analysis based on

transcription factor binding sites. Bioinformatics 21, 2933–2942. https://doi.org/10.1093/bioinformatics/bti473

Chapman, B.S., Thayer, R.M., Vincent, K.A., Haigwood, N.L., 1991. Effect of intron A from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian cells. Nucleic Acids Res. 19, 3979–3986. https://doi.org/10.1093/nar/19.14.3979

Cheadle, C., Fan, J., Cho-Chung, Y.S., Werner, T., Ray, J., Do, L., Gorospe, M., Becker, K.G., 2005. Stability regulation of mRNA and the control of gene expression. Ann. N. Y. Acad. Sci. 1058, 196–204. https://doi.org/10.1196/annals.1359.026

Chen, J., Haverty, J., Deng, L., Li, G., Qiu, P., Liu, Z., Shi, S., 2013. Identification of a novel endogenous regulatory element in Chinese hamster ovary cells by promoter trap. J. Biotechnol. 167, 255–261. https://doi.org/10.1016/j.jbiotec.2013.07.001

Cheng, J.K., Alper, H.S., 2016. Transcriptomics-guided design of synthetic promoters for a mammalian system. ACS Synth. Biol. 5, 1455–1465. https://doi.org/10.1021/acssynbio.6b00075

Curtin, J.A., Dane, A.P., Swanson, A., Alexander, I.E., Ginn, S.L., 2008. Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct. Gene Ther. 15, 384–390. https://doi.org/10.1038/sj.gt.3303105

Davies, S.L., O'Callaghan, P.M., McLeod, J., Pybus, L.P., Sung, Y.H., Rance, J., Wilkinson, S.J., Racher, A.J., Young, R.J., James, D.C., 2011. Impact of gene vector design on the control of recombinant monoclonal antibody production by CHO cells. Biotechnol. Prog. 27, 1689–1699. https://doi.org/10.1002/btpr.692

Dinnis, D.M., James, D.C., 2005. Engineering mammalian cell factories for improved recombinant monoclonal antibody production: lessons from nature? Biotechnol. Bioeng. 91, 180–189. https://doi.org/10.1002/bit.20499

Dolcet, X., Llobet, D., Pallares, J., Matias-Guiu, X., 2005. NF-κB in development and progression of human cancer. Virchows Arch. 446, 475–482. https://doi.org/10.1007/s00428-005-1264-9

Eskla, K.L., Porosk, R., Reimets, R., Visnapuu, T., Vasar, E., Hundahl, C.A., Luuk, H., 2018. Hypothermia augments stress response in mammalian cells. Free Radic. Biol. Med. 121, 157–168. https://doi.org/10.1016/j.freeradbiomed.2018.04.571

Fan, L., Kadura, I., Krebs, L.E., Larson, J.L., Bowden, D.M., Frye, C.C., 2013. Development of a highly-efficient CHO cell line generation system with engineered SV40E promoter. J. Biotechnol. 168, 652–658. https://doi.org/10.1016/j.jbiotec.2013.08.021

Fomina-Yadlin, D., Mujacic, M., Maggiora, K., Quesnell, G., Saleem, R., McGrew, J.T., 2015. Transcriptome analysis of a CHO cell line expressing a recombinant therapeutic protein treated with inducers of protein expression. J. Biotechnol. 212, 106–115. https://doi.org/10.1016/j.jbiotec.2015.08.025

Jakobi, T., Brinkrolf, K., Tauch, A., Noll, T., Stoye, J., Pühler, A., Goesmann, A., 2014. Discovery of transcription start sites in the Chinese hamster genome by next-generation RNA sequencing. J. Biotechnol. 190, 64–75. https://doi.org/10.1016/j.jbiotec.2014.07.437

Jasin, M., Rothstein, R., 2013. Repair of strand breaks by homologous recombination. Cold Spring Harb. Perspect. Biol. 5, a012740. https://doi.org/10.1101/cshperspect.a012740

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. Biotechnol. Bioeng. 112, 2527–2542. https://doi.org/10.1002/bit.25687

Johnson, K.C., Yongky, A., Vishwanathan, N., Jacob, N.M., Jayapal, K.P., Goudar, C.T., Karypis, G., Hu, W.S., 2014. Exploring the transcriptome space of a rcombinant BHK cell line through next generation sequencing. Biotechnol. Bioeng. 111, 770–781. https://doi.org/10.1002/bit.25135

Juven-Gershon, T., Cheng, S., Kadonaga, J.T., 2006. Rational design of a super core promoter that enhances gene expression. Nat. Methods 3, 917–922. https://doi.org/10.1038/nmeth937

Juven-Gershon, T., Kadonaga, J.T., 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev. Biol. 339, 225–229. https://doi.org/10.1016/j.ydbio.2009.08.009

Kang, S.Y., Kim, Y.G., Kang, S., Lee, H.W., Lee, E.G., 2016. A novel regulatory element (E77) isolated from CHO-K1 genomic DNA enhances stable gene expression in Chinese hamster ovary cells. Biotechnol. J. 11, 633–641. https://doi.org/10.1002/biot.201500464

Kim, M., O'Callaghan, P.M., Droms, K.A., James, D.C., 2011. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol. Bioeng. 108, 2434–2446. https://doi.org/10.1002/bit.23189

Kunert, R., Reinhart, D., 2016. Advances in recombinant antibody manufacturing. Appl. Microbiol. Biotechnol. 100, 3451–3461. https://doi.org/10.1007/s00253-016-7388-9

Kutach, A.K., Kadonaga, J.T., 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. Mol. Cell. Biol. 20, 4754–4764. https://doi.org/10.1128/MCB.20.13.4754-4764.2000

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., Weirauch, M.T., 2018. The human transcription factors. Cell 172, 650–665. https://doi.org/10.1016/j.cell.2018.01.029

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., Kadonaga, J.T., 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes Dev. 18, 1606–1617. https://doi.org/10.1101/gad.1193404

Lin, Y.C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van de Peer, Y., Tavernier, J., Callewaert, N., 2014. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. Nat. Commun. 5, 4767. https://doi.org/10.1038/ncomms5767

Mariati, Ng, Y.K., Chao, S.H., Yap, M.G., Yang, Y., 2010. Evaluating regulatory elements of human cytomegalovirus major immediate early gene for enhancing transgene expression levels in CHO K1 and HEK293 cells. J. Biotechnol. 147, 160–163. https://doi.org/10.1016/j.jbiotec.2010.02.022

Mariati, Yeo, J.H., Koh, E.Y., Ho, S.C., Yang, Y., 2014. Insertion of core CpG island element into human CMV promoter for enhancing recombinant protein expression stability in CHO cells. Biotechnol. Prog. 30, 523–534. https://doi.org/10.1002/btpr.1919

Matsusaka, T., Fujikawa, K., Nishio, Y., Mukaida, N., Matsushima, K., Kishimoto, T., Akira S., 1993. Transcription factors NF-IL6 and NF-κB synergistically activate transcription of the inflammatory cytokines, interleukin 6 and interleukin 8. Proc. Natl. Acad. Sci. U.S.A. 90, 10193–10197. https://doi.org/10.1073/pnas.90.21.10193

O'Callaghan, P.M., McLeod, J., Pybus, L.P., Lovelady, C.S., Wilkinson, S.J., Racher, A.J., Porter, A., James, D.C., 2010. Cell line specific control of recombinant monoclonal antibody production by CHO cells. Biotechnol. Bioeng. 106, 938–951. https://doi.org/10.1002/bit.22769

Pan, Y., Tsai, C.J., Ma, B., Nussinov, R., 2010. Mechanisms of transcription factor selectivity. Trends Genet. 26, 75–83. https://doi.org/10.1016/j.tig.2009.12.003

Running-Deer, J., Allison, D.S., 2004. High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1α gene. Biotechnol. Prog. 20, 880–889. https://doi.org/10.1021/bp034383r

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., Selbach, M., 2011. Global quantification of mammalian gene expression control. Nature 473, 337–342. https://doi.org/10.1038/nature10098

Sha, S., Bhatia, H., Yoon, S., 2018. An RNA-seq based transcriptomic investigation into the productivity and growth variants with Chinese hamster ovary cells. J. Biotechnol. 271, 37–46. https://doi.org/10.1016/j.jbiotec.2018.02.008

Sharova, L.V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., Ko, M.S., 2009. Database for mRNA half-life of 19977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res. 16, 45–58. https://doi.org/10.1093/dnares/dsn030

Skoko, N., Baralle, M., Tisminetzky, S., Buratti, E., 2011. InTRONs in biotech. Mol. Biotechnol. 48, 290–297. https://doi.org/10.1007/s12033-011-9390-x

Thaisuchat, H., Baumann, M., Pontiller, J., Hesse, F., Ernst, W., 2011. Identification of a novel temperature sensitive promoter in CHO cells. BMC Biotechnol. 11, 51. https://doi.org/10.1186/1472-6750-11-51

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578. https://doi.org/10.1038/nprot.2012.016

Wright, C., Estes, S., 2014. Next-generation bioprocess: an industry perspective of how the 'omics era will affect future biotherapeutic development. Pharm. Bioprocess. 2, 371–375. https://doi.org/10.4155/pbp.14.41

Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M.R., Neff, N., Passarelli, B., Koh, W., Fan, H.C., Wang, J., Gui, Y., Lee, K.H., Betenbaugh, M.J., Quake, S.R., Famili, I., Palsson, B.O., Wang, J., 2011. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat. Biotechnol. 29, 735–741. https://doi.org/10.1038/nbt.1932

Zhou, Y., Raju, R., Alves, C., Gilbert, A., 2018. Debottlenecking protein secretion and reducing protein aggregation in the cellular host. Curr. Opin. Biotechnol. 53, 151–157. https://doi.org/10.1016/j.copbio.2018.01.007

FIGURE LEGENDS

Figure 1. Summary of bioinformatic analysis (Steps 1–4) followed by *in vitro* screening (Step 5) to obtain active TFREs for the design of CHO synthetic promoters for biphasic production processes.

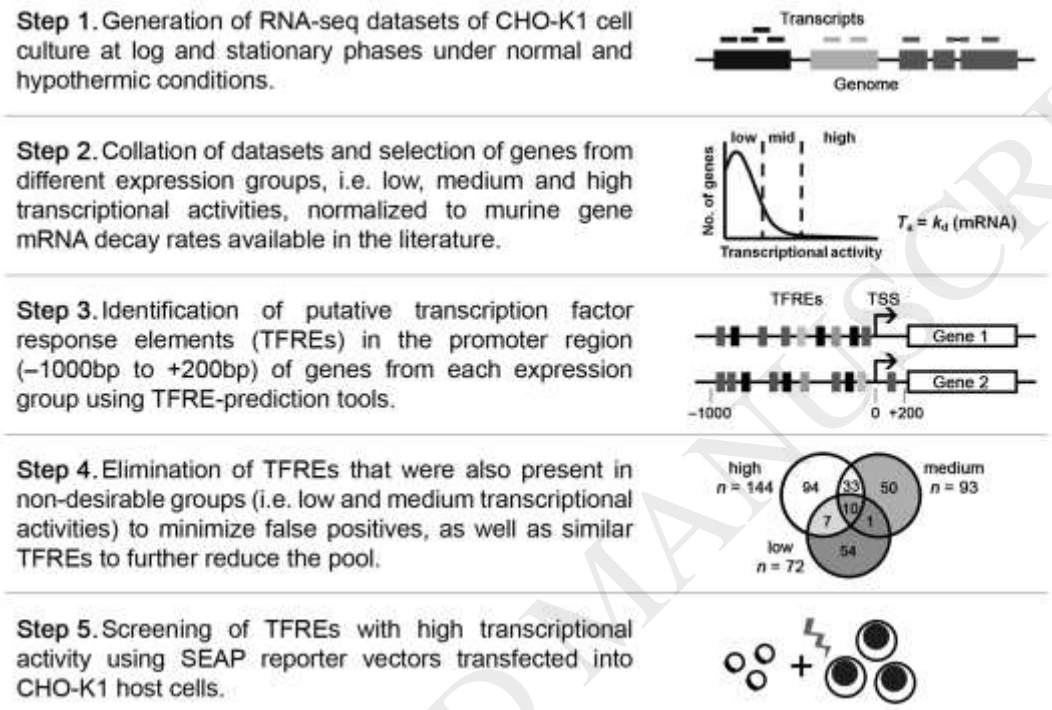


Figure 2. Distribution of discrete transcription factor regulatory elements (TFREs) across high, medium and low transcriptional activity groups. CHO-K1 endogenous promoters with high, medium and low transcriptional activities (50 promoters in each group; Supplementary Table S1) were surveyed for the presence of discrete TFREs using Genomatix Gene Regulation software. The region −1000 to +200 relative to predicted TSSs was analyzed against a murine promoter background to find overrepresented TFREs in each group. In order to identify potentially active TFREs in the high activity group, TFREs that also occurred in the medium and/or low activity groups were excluded and the remaining TFREs were narrowed down to 32 as described in text. DNA sequences of the selected TFREs are listed in Table 1.

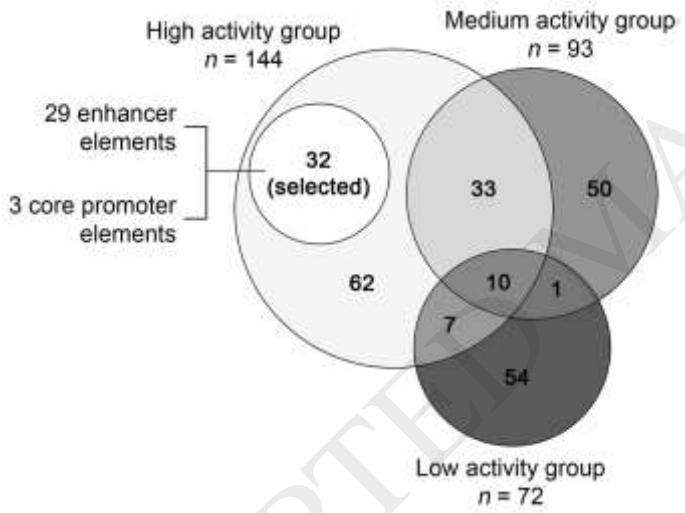


Figure 3. Screening discrete transcription factor regulatory element activity in CHO-K1 cells. **A:** Six copies of each enhancer elements (as described in Table 1) were cloned in series upstream of a minimal CMV core promoter in reporter vectors encoding SEAP. **B:** The minimal CMV core promoter (containing a TATA-box and Inr) was modified to include core promoter elements MTE and/or DPE at specific positions from Inr while TATA-less core promoters were created by mutating the TATA box sequence. The core promoters were inserted upstream of the SEAP ORF of promoterless, enhancerless vectors. 6 µg of plasmid was transfected into 5×10^6 CHO-K1SV cells by electroporation followed by culture in TubeSpin bioreactors at 37°C. SEAP level in cell culture supernatant was measured 48 h post-transfection. Data are expressed as a fold-change with respect to the cell specific production rate (qP) of a vector containing only a minimal, wild-type CMV core promoter (Core). Data shown are the mean value ± standard deviation of three biological and three technical replicates.

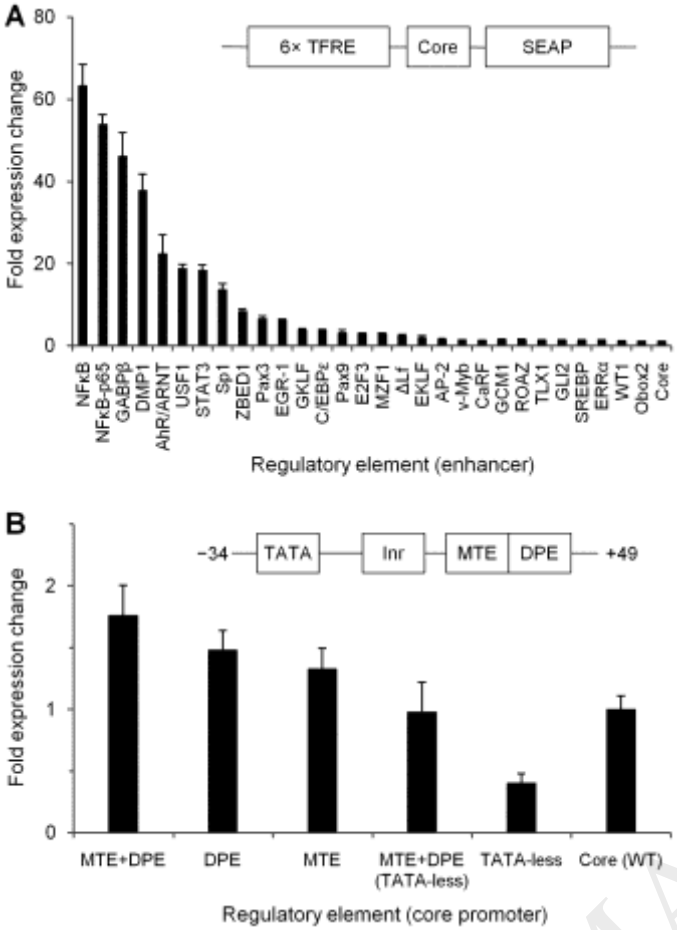


Figure 4. Measurement of synthetic promoter activity in CHO-K1 cells. Nine synthetic promoters comprising Library 1 (**A**) and ten synthetic promoters comprising Library 2 (**B**; TFRE compositions described in Table 2) were transfected into CHO-K1 cells (4 µg plasmid per 5×10^6 cells). SEAP level in cell culture supernatant and SEAP mRNA level in cells were analyzed 48 h post-transfection. Data are expressed as a fold-change with respect to the qP (■) and mRNA abundance (□) exhibited by the control CMV promoter. In **A**, SEAP expression driven by the SV40 promoter was also tested and in **B**, the SEAP expression from the most active promoter from the first library (1/09) is shown as a striped bar. Data shown are the mean value ± standard deviation of two independent experiments each performed in duplicate.

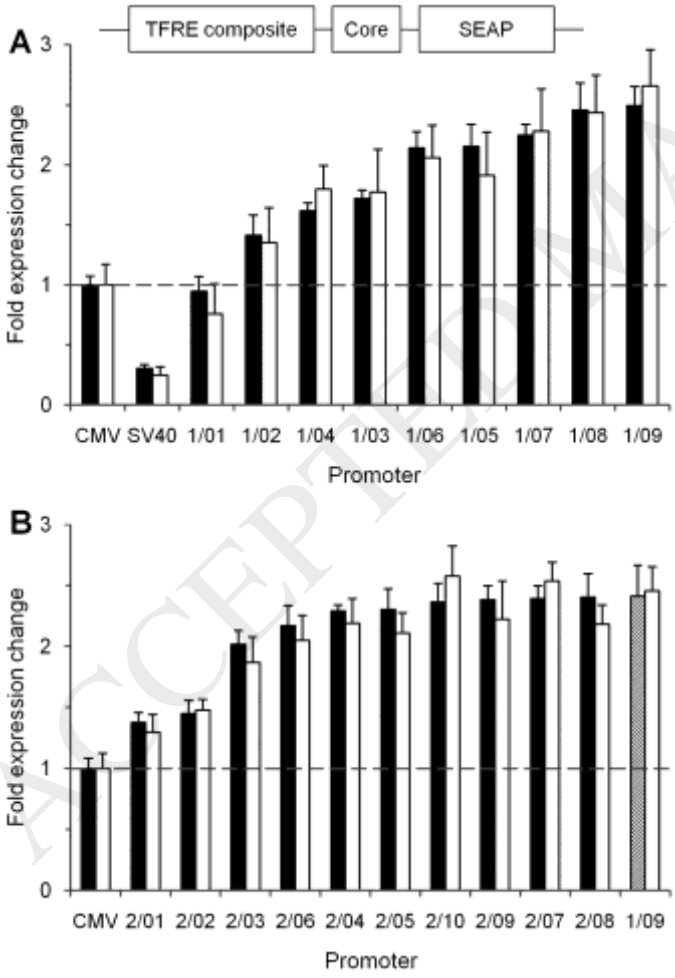
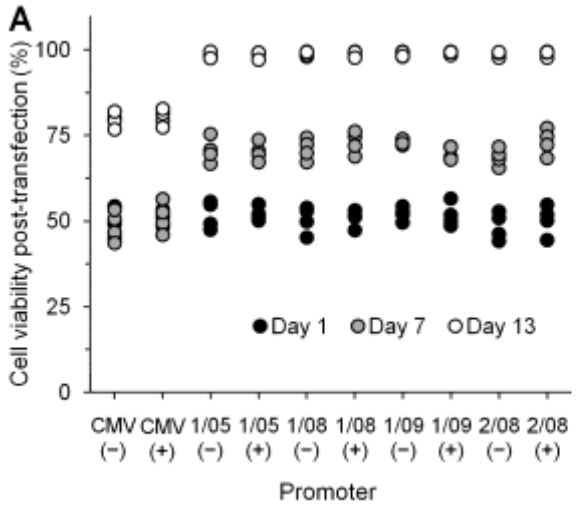
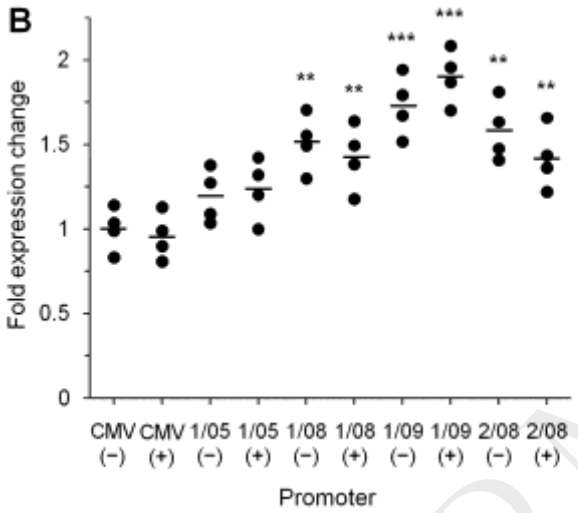


Figure 5. Generation and analysis of CHO stable transfectant pools expressing recombinant SEAP under the control of natural and synthetic promoters**.** For each promoter, 5×10^6 $GS^{-/-}$ CHO-K1 cells were electroporated in quadruplicate with 4 µg plasmid containing a GS gene driven by an SV40 promoter and a SEAP gene driven by either a CMV or synthetic promoter without (−) or with (+) intron A, followed by selection in glutamine-free media. Transfected cells were cultured in TubeSpin bioreactors at 37°C with cell viability measured 2 h post-transfection 81±4% (data not shown). Culture media was changed every 3–4 days until the cell viability reached >95% and the stable pools were cryopreserved. **A:** Measurement of cell viability at Day 1 (black circles), Day 7 (gray circles) and Day 13 (open circles) post-transfection. Each circle represents a transfection pool. **B:** Measurement of SEAP expression (96 h culture) on the fourth passage post-cell revival. Each circle represents a stable pool and the horizontal bars represent the mean of the SEAP expression for each promoter. Data is normalized with respect to expression of CMV(−). **C:** 1.5×10^6 cells from the four stable pools of a specific promoter in **B** were combined and the SEAP mRNA and GS mRNA (expressed by SV40) levels were quantified. Data shown is the mean value ± standard deviation of three technical replicates normalized with respect to expression of CMV(−). Mean values significantly different (two-tailed Students *t*-test) from the expression of CMV(−) values are indicated by asterisks (**P*<0.05, ***P*<0.01, ****P*<0.001).





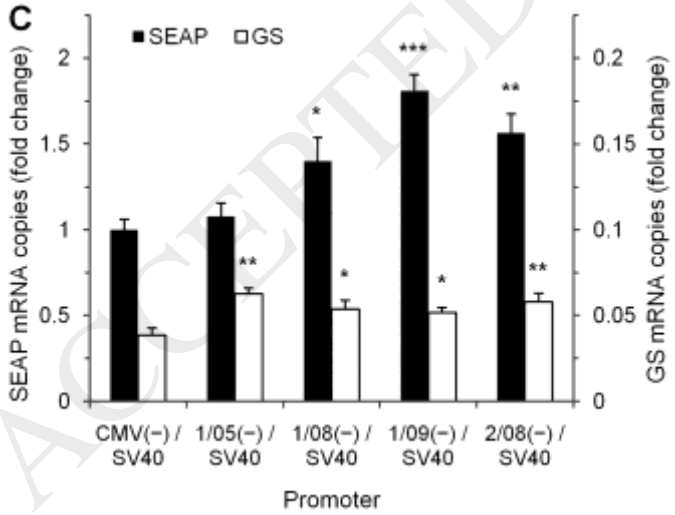
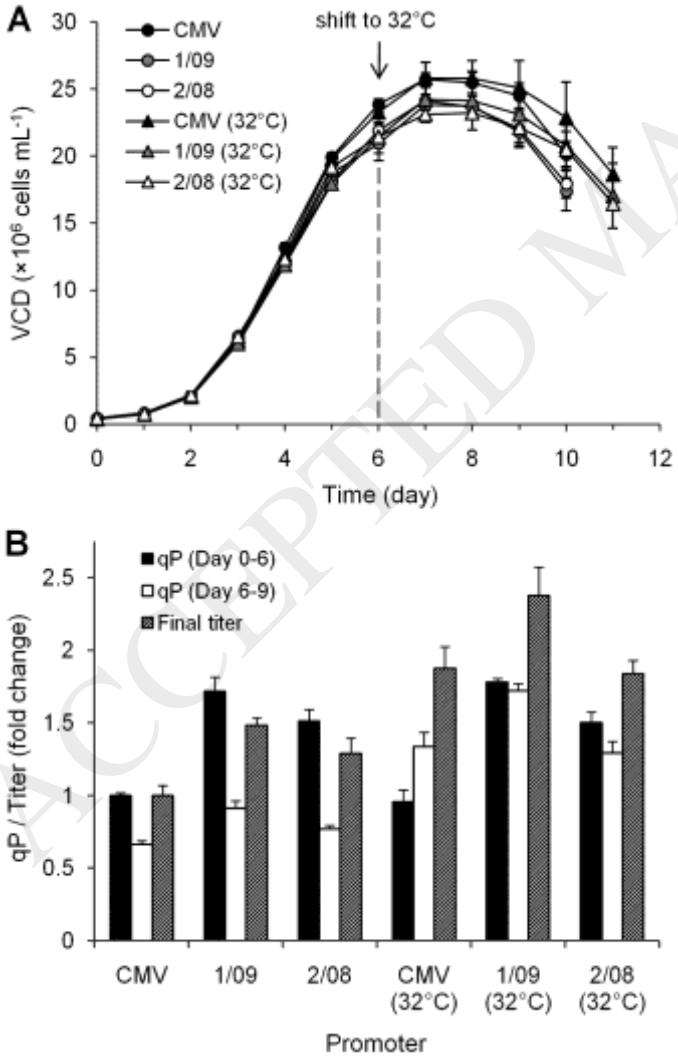


Figure 6. Synthetic promoters exhibit improved biomanufacturing performance during fed-batch culture. Stable transfectant pools expressing recombinant SEAP under the control of the CMV promoter or synthetic promoters 1/09 and 2/08 were subjected to a fed-batch production process in Erlenmeyer flasks. **A:** Viable cell concentration was measured over the course of 10 days for cultures under normal 37°C condition or 11 days for cultures with a shift to 32°C at Day 6 until cell viability dropped below 80%. **B:** qP at log phase (Day 0–6) and stationary phase (Day 6–9) were quantified by measuring the integral viable cell density (IVCD) and SEAP titer at Days 6 and 9. Data are normalized with respect to CMV culture under normal condition. Data shown are the mean value ± standard deviation of two biological and two technical replicates.



LIST OF TABLES

Table 1. DNA sequence of potentially active transcription factor regulatory elements identified by bioinformatic survey of endogenous CHO-K1 promoters. Measurement of the relative ability of TFREs to activate transcription of recombinant SEAP genes in CHO-K1 cells is shown in Figure 3.

Transcription factor regulatory element	Binding site
Enhancer elements	
Activator protein 2 (AP-2)	TTGCCTGGGGCTAT
Aryl hydrocarbon receptor and nuclear translocator heterodimer (AhR/ARNT)	GTTGCGTGCGAA
Calcium-response factor (CaRF)	AGAACGAGGCA
CCAAT-enhancer binding protein epsilon (C/EBPε)	ATTGCGCAAT
Cyclin D-interacting Myb-like protein (DMP1)	GACCCGGATGTAG
Delta-lactoferrin (ΔLf)	GGCACTTGC
E2F transcription factor 3 (E2F3)	GCTCGGCGCCAAAC
Early growth response protein 1 (EGR-1)	ATGCGTGGGCGT
Erythroid Krueppel-like factor (EKLF)	CAGGGAGGGTG
Estrogen-related receptor alpha (ERRα)	TCCAAGGTCACA
GA-binding protein beta (GABPβ)	CCCCGGAAGTGAC
GLI family zinc finger protein 2 (GLI2)	GACCACCCAAG
Glial cells missing homolog 1 (GCM1)	AAACCCGCATAT
Gut-enriched Krueppel-like factor (GKLF)	ATCACAGGATT
Myeloid zinc finger 1 (MZF1)	TGGTGGGGAGGGG
Nuclear factor kappaB (NFκB)	TGGGACTTTCCA
Nuclear factor kappaB p65 (NFκB-p65)	TTGGGGATTTCCCA
Oocyte specific homeobox 2 (Obox2)	TAGTTAATCCCCCT
Paired box protein Pax-3 (Pax3)	TCGTCACGCTTCA

Paired box protein Pax-9 (Pax9)	GTCACGCATGACTGC
Rat Olf-1/EBF-associated Zn finger protein (ROAZ)	GGCACCCAAGGGTGA
Signal transducer/activator of transcription 3 (STAT3)	ATTTCCCGGAAATG
Sterol regulatory element binding protein 1 & 2 (SREBP)	AATCACCCCACTGC
Stimulating protein 1 (Sp1)	AGGGGCGGGGT
T-cell leukemia homeobox 1 (TLX1)	CGGTAATTGG
Upstream stimulating factor 1 (USF1)	GGGTCACGTGG
Viral Myb protein (v-Myb)	TTTAACGGCAA
Wilms tumor suppressor (WT1)	TGCGTGGGAGTAG
Zinc finger BED-type containing 1 (ZBED1)	TGTCGCGACA
Core promoter elements	
Downstream promoter element (DPE)	AGACGTGCCT
Initiator (Inr)	TCAGAT
Motif ten element (MTE)	CCGAGCGGAGC

Table 2. Composition of specific TFRE copies in synthetic promoters. A two-level factorial design with a center point (marked with an asterisk [*]) using six factors (NFκB, GABPβ, DMP1, AhR/ARNT, USF1, STAT3) was used to design the first library of nine synthetic promoters with different TFRE compositions biased towards more active TFREs. A reduced two-factor interaction model was employed with the low and high levels for NFκB, GABPβ and DMP1 set to two and six copies, respectively, and the low and high levels for AhR/ARNT, USF1 and STAT3 set to one and three copies, respectively. Additionally, one copy of Sp1 is added to each promoter. Derived from the expression analysis of the first library promoters (Figure 4A), a two-level factorial design using four factors (NFκB, GABPβ, USF1, STAT3) was used to design a second library of eight promoters with different TFRE compositions further biased towards highly active TFREs. DMP1 and AhR/ARNT were added at mid-levels to support basal expression. Two potentially strong promoter designs not covered in the DOE (marked with double asterisks [**]) were also constructed. The arrangement of TFRE copies was randomized using R software, chemically synthesized and inserted upstream of the minimal CMV core promoter in SEAP reporter plasmids.

Promoter	**NFκB**	**GABPβ**	**DMP1**	**AhR/ ARNT**	**USF1**	**STAT3**	**Sp1**	**Total**
Library 1								
1/01	2	2	2	3	3	3	1	16
1/02	2	2	6	3	1	1	1	16
1/03	2	6	2	1	3	1	1	16
1/04	6	2	2	1	1	3	1	16
1/05*	4	4	4	2	2	2	1	19
1/06	2	6	6	1	1	3	1	20
1/07	6	2	6	1	3	1	1	20
1/08	6	6	2	3	1	1	1	20

1/09	6	6	6	3	3	3	1	28
Library 2								
2/01	3	3	4	2	1	1	–	14
2/02	3	3	4	2	3	3	–	18
2/03	3	7	4	2	1	3	–	20
2/04	3	7	4	2	3	1	–	20
2/05	7	3	4	2	1	3	–	20
2/06	7	3	4	2	3	1	–	20
2/07	7	7	4	2	1	1	–	22
2/08	7	7	4	2	3	3	–	26
2/09**	7	6	5	4	3	2	–	27
2/10**	7	7	7	7	–	–	–	28

Table 3. Summary of DOE analysis of effectors and effector interactions controlling SEAP gene expression in synthetic promoters. Using cell specific production rate (qP) as output variable, reduced two-factor interaction (2FI) models were used to compare the impact of specific TFREs and the presence of interactions among them, with $p < 0.05$ considered to be significant. The alias structures/confounding factors are as follows: First synthetic promoter library; A = A + BD + CE, B = B + AD + CF, C = C + AE + BF, D = D + AB + EF, E = E + AC + DF, F = F + BC + DE, AF = AF + BE + CD; Second synthetic promoter library; A = A, B = B, C = C, D = D, AB = AB + CD, AC = AC + BD, AD = AD + BC.

Factor	**Sum of squares**	***p*-value**	**Signif.**	**Model predictability (Pred/AdjR^2)**
Library 1				
Model	50.82	<0.001	Yes	0.42/0.61
A: NFκB	20.65	<0.001	Yes	
B: GABPβ	19.55	<0.001	Yes	
C: DMP1	8.48	0.012	Yes	
D: AhR/ARNT	1.28	0.296	No	
E: USF1	0.76	0.417	No	
F: STAT3	0.10	0.766	No	
AF: (NFκB)(STAT3)	–	–	–	
Library 2				
Model	94.05	0.001	Yes	0.36/0.51
A: NFκB	50.94	0.001	Yes	
B: GABPβ	35.00	0.004	Yes	
C: USF1	4.12	0.281	No	
D: STAT3	3.99	0.289	No	
AB: (NFκB)(GABPβ)	–	–	–	
AC: (NFκB)(USF1)	–	–	–	
AD: (NFκB)(STAT3)	–	–	–	