Biotechnology Letters

Identifying and engineering promoters for high level and sustainable therapeutic recombinant protein production in cultured mammalian cells

--Manuscript Draft--

Manuscript Number:	BILE-D-14-00118R2
Full Title:	Identifying and engineering promoters for high level and sustainable therapeutic recombinant protein production in cultured mammalian cells
Article Type:	Review (invited by Niall Barron)
Section/Category:	Reviews
Keywords:	Mammalian cells; Promoter engineering; Promoter identification; Recombinant protein production
Corresponding Author:	Yuansheng Yang, Ph.D Bioprocessing Technology Institute SINGAPORE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Bioprocessing Technology Institute
Corresponding Author's Secondary Institution:	
First Author:	Steven C.L. Ho, Bachelor
First Author Secondary Information:	
Order of Authors:	Steven C.L. Ho, Bachelor
	Yuansheng Yang, Ph.D
Order of Authors Secondary Information:	
Manuscript Region of Origin:	SINGAPORE
Abstract:	Promoters are essential on plasmid vectors to initiate transcription of the transgenes when generating therapeutic recombinant protein expressing mammalian cell lines. High and sustained gene expression levels are desired during therapeutic protein production while fine control of gene expression is useful for cell engineering. As many promoters commonly exhibit cell and product specificity, there is a need for new promoters to be identified, optimized and carefully evaluated before use. Suitable promoters can be identified using techniques ranging from simple molecular biology methods to modern high-throughput omics screenings. Promoter engineering is often required after identification to either obtain high and sustained expression or to provide a wider range of gene expression. This review discusses some of the available methods to identify and engineer promoters for recombinant therapeutic protein expression in mammalian cells.

Section: Review

Identifying and engineering promoters for high level and sustainable therapeutic recombinant protein production in cultured mammalian cells

Steven C.L. Ho[1,2], Yuansheng Yang[1,3*]

[1] Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), 20 Biopolis Way, #06-01 Centros, Singapore 138668

[2] Division of Bioengineering, Department of Chemical and Biomolecular Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119620

[3]School of Chemical and Biomedical Engineering, Nanyang Technological University, N1.2-B2-33, 62 Nanyang Avenue, Singapore 637459

[*] Correspondence to:
Yuansheng Yang
Phone: (65)-64070825, Fax: (65)-64789561
Email: yang_yuansheng@bti.a-star.edu.sg

Abstract

Promoters are essential on plasmid vectors to initiate transcription of the transgenes when generating therapeutic recombinant proteins expressing mammalian cell lines. High and sustained levels of gene expression are desired during therapeutic protein production while fine control of gene expression is useful for cell engineering. As many promoters commonly exhibit cell and product specificity, there is a need for new promoters to be identified, optimized and carefully evaluated before use. Suitable promoters can be identified using techniques ranging from simple molecular biology methods to modern high-throughput omics screenings. Promoter engineering is often required after identification to either obtain high and sustained expression or to provide a wider range of gene expression. This review discusses some of the available methods to identify and engineer promoters for recombinant therapeutic protein expression in mammalian cells.

Keywords (up to 7):

Mammalian cells, Promoter engineering, Promoter identification, Recombinant protein production

Introduction

There is rapidly growing demand for existing products and a constant inflow of novel products in the recombinant therapeutic biologics market. Monoclonal antibodies (mAb), the best-selling class of biologics, accounted for sales of over $20.3 billion in the USA in 2011 (Aggarwal 2012). Biosimilar versions for some of the top therapeutic proteins are also being developed and getting approved for the market (Kling 2012). With the expected sustained growth of the biologics market in the near future, cell culture technologies for therapeutic recombinant protein production are required to improve as well.

Mammalian cells, such as Chinese hamster ovary (CHO) cells, human embryonic kidney cells (HEK293) and murine lymphoid cell lines, are commonly used for production of safe and effective biologics (Ho et al. 2013; Wurm 2004). Generating a therapeutic recombinant protein-producing cell line starts with transfecting the mammalian host cells with a plasmid vector carrying the gene for the respective recombinant protein (Birch and Racher 2006; Ho et al. 2013; Smale and Kadonaga 2003). Small amounts of transiently expressed product can be collected a few days after transfection for small scale assays during early stages of drug development (Pham et al. 2006). Stably-transfected cell lines that have undergone selection for plasmid vectors integrated into the host cell's genome are used for commercial scale therapeutic protein production at high yields and consistent quality (Barnes et al. 2003). It is desired that the recombinant protein expression be high in both transiently and stably transfected cell lines. There is an added requirement of sustained expression over at least two to three

months in stably-transfected cell lines during the scale-up process to ensure consistent product levels and quality (Barnes et al. 2003).

Obtaining sustained, high level recombinant protein production depends on many factors such as cell host, culture media, culture process, and expression vectors (Birch and Racher 2006; Costa et al. 2010; Ho et al. 2013). The promoter is an essential component of the expression vector which initiates transcription and can affect both expression level and stability. Promoters can be classified broadly into inducible and constitutive promoters. Constitutive promoters are commonly used for recombinant protein production, while inducible promoters are occasionally used either for cell engineering or expressing products toxic to the cell. This review will introduce the commonly used promoters and how novel promoters can be identified and engineered either for high and sustained gene expression or for cell engineering in mammalian cells, with emphasis placed on therapeutic protein production.

Promoters for therapeutic recombinant protein production

A typical promoter used for therapeutic recombinant protein production contains a core promoter element, an immediate upstream enhancer region and if required, other *cis*-acting regulatory elements. The core promoter is a minimal DNA sequence of around 80 nucleotides required for transcription initiation which extends approximately 35 base pairs upstream and downstream of the initiation site (Smale and Kadonaga 2003). Sequence motifs found in well-studied TATA-box based core promoters includes the TATA box (TATA box binding protein recognition sequence, TATAWAAR), BRE (Transcription factor IIB recognition sequence, SSRGCC), Inr (initiator sequence,

YYANWYY), MTE (motif ten element, CSARCSSAAC), DPE (downstream promoter element, RGWYV) (Juven-Gershon and Kadonaga 2010; Pedersen et al. 1999; Smale and Kadonaga 2003). These motifs allow anchoring of transcription factors to form a transcription machinery that drives transcription by RNA polymerase II (Juven-Gershon and Kadonaga 2010). Apart from TATA-box based promoters, CpG rich mammalian promoters which are more plastic and evolvable have been identified with modern genomic tools. These promoters have multiple transcription start sites as compared to focused start sites of TATA-box based promoters (Juven-Gershon and Kadonaga 2010). After transcription is initiated by the core promoter, the gene expression is augmented by DNA elements like enhancers and introns. Enhancers are able to recruit a variety of transcription factors to activate transcription and can function independent of location and orientation (Ong and Corces 2011). The enhancer bound transcription factors regulate promoter activity either by direct interaction with the basic transcription machinery (Malik and Roeder 2010) or by forming complexes and recruiting enzymes to remodel the chromatin structure to increase accessibility (Clapier and Cairns 2009).

There are currently countless promoters identified and catalogued for both eukaryotic and viral sources in online databases (Table 1). A promoter derived from the human cytomegalovirus major immediate early gene (hCMV) is the most commonly used promoter for high therapeutic recombinant protein expression in industry (Costa et al. 2010). Other promoters that can be used include promoters from the murine cytomegalovirus (mCMV), simian virus 40 (SV40), CHO and human elongation factor-1 α-1 gene (CHOEF1α, HEF1α) and a hybrid promoter of hCMV enhancer and chicken ß-actin promoter (CAG). These promoters are used on commercial plasmids for

Table 1

constructing vectors to express therapeutic recombinant proteins (Table 2). The SV40 promoter is considered as a weaker promoter among the above list and is mostly used to drive expression of the selection marker gene for more stringent selection of stably transfected high producing cells (Birch and Racher 2006).

Table 2

Due to the plethora of available promoters, the choice of promoters for recombinant protein expression and engineering of mammalian cells can be complicated as the promoters and *cis*-regulatory elements, such as introns, exhibit product specificity, cell line specificity and even differences in transient and stable conditions for recombinant protein expression in mammalian cells (Kim et al. 2002; Mariati et al. 2010; Xia et al. 2006). For instance, mCMV promoter is stronger than hCMV in a few reports (Addison et al. 1997; Kim et al. 2002). However, when these two promoters were compared for mAb production in CHO cells in Lonza's GS system, one of the commonly used vectors for mammalian gene expression in industry, mCMV, had lower expression than hCMV (Kalwy 2005). Another possible reason for conflicting reports in various studies is that different versions of truncated or full-length promoter were used. For instance, the hCMV promoter in some vectors only contains the enhancer and core promoter, while in others it could include the first intron and other cis-regulatory elements (Mariati et al. 2010). Evaluation should be performed when choosing promoters for different purposes instead of depending on the results published in literatures.

While expression sustainability is less frequently reported compared to expression level, it is another critical issue faced during recombinant protein production using stably transfected cell lines and cell engineering. There are multiple reports linking the loss of gene expression in CHO cells expressing therapeutic recombinant protein upon long term

culture to the use of the hCMV promoter (Chusainow et al. 2009; Kim et al. 2011; Osterlehner et al. 2011; Yang et al. 2010). Although the use of endogenous mammalian promoters is thought to be advantageous compared to viral promoters for transgene expression stability (Chen et al. 2013), there is a shortage of data comparing promoter expression stability in industrial applications. We have observed in our laboratory that the SV40 promoter was most resistant to gene silencing but gave the lowest expression when compared to hCMV, mCMV, and CHOEF1α in stably-transfected CHO cell clones (results not published). While therapeutic protein production requires the highest possible expression levels, cell engineering experiments to improve the therapeutic protein expression or quality can require protein expression, either at lower levels or under specific conditions, thereby further complicating promoter selection process (Blazeck and Alper 2013; Chen et al. 2013). Due to the absence of a promoter that performs well for all applications, work is always on-going to identify and engineer novel promoters to complement new products and cell lines.

Identifying novel promoters

Identifying a promoter that can achieve high and sustained gene expression is challenging, especially in scenarios where novel cell hosts are to be utilized. Promoters can either be isolated from viral sources that infect mammalian cells, for example, the hCMV (Thomsen et al. 1984) and mCMV promoters (Dorsch-Hasler et al. 1985) or from upstream sequences of the host's endogenous genes, like the HEF1α and CHOEF1α promoters (Deer and Allison 2004; Kim et al. 1990). Promoters can be identified using

methods ranging from using simple molecular biology techniques to utilizing high-throughput microarrays and genomics tools.

"Trapping" of promoters and regulatory elements can be performed to identify and study novel sequences (Fig. 1). This method of promoter identification is tedious due to its random nature but can be performed using basic molecular biology tools. A promoter trap utilizes a promoter-less vector carrying a green fluorescent protein (GFP) reporter gene (Chen et al. 2013). The vector was transfected into CHO cells and GFP expressing cells would signify that the vector was integrated adjacent to a transcriptionally active promoter. The upstream region was sequenced to identify a promoter which was 66% as active as the SV40 promoter. A variation of the promoter trap involved shotgun cloning of CHO genomic fragments into a promoter-less vector carrying an antibiotic selection marker gene. Genomic fragments which exhibit promoter activity can be isolated from clones which display antibiotic resistance (Pontiller et al. 2008). Enhancer trapping can be performed using vectors carrying a weak SV40 core promoter (Weber et al. 1984). Fragments from hCMV and mCMV were inserted in the enhancer trap, upstream of the SV40 core promoter, to substitute the missing enhancer. Sequences with enhancer properties that increase gene expression were identified (Boshart et al. 1985; Dorsch-Hasler et al. 1985).

Fig. 1

Another way to use the enhancer trap is to transfect cells with the enhancer-less vector. The transgenes would only be expressed when the vector is integrated downstream of an enhancer (Bulger and Groudine 2010). Although promoter and enhancer trapping is an arbitrary exercise, there is potential to identify novel promoters or regulatory segments

that cannot be easily identified by screening for conserved motifs or in-silico promoter prediction.

An alternative approach to identify strong promoters is based on the expression level of endogenous genes in mammalian cells. For instance, elongation factor-1 α-1 gene (EF1α) is a house-keeping gene and is highly expressed in different mammalian cells. HEF1α and CHOEF1α promoters had been isolated using cDNA probes to screen CHO-K1 lambda phage genomic libraries and confirmed to be strong promoters for high level transgene expression (Deer and Allison 2004; Uetsuki et al. 1989). CHOEF1α promoter when combined with its flanking elements gave reporter gene expression that was up to 35-fold higher than the hCMV promoter in CHO cells (Deer and Allison 2004). In another effort to identify endogenous promoters, a CHO DNA library was first built by blunt-end cloning chromosomal DNA fragments into promoter-less vector backbones. Primers binding to the first exon of highly expressed CHO genes were designed and used together with primers binding to the backbone of the vector used for generating the library to clone the 5'- flanking regions (Pontiller et al. 2010). The cloned 5'- ends were inserted into reporter luciferase vectors to check for promoter activity. While no promoters stronger than SV40 were identified in the above experiment, the method was effective in isolating flanking promoter regions of known genes.

While highly expressed endogenous proteins are obvious targets, genes with interesting properties like elevated expression under specific conditions can also contain useful promoter elements. One example is the mild-cold responsive enhancer (MCRE) found upstream of a cold-inducible RNA-binding protein (cirp). Cirp expression is induced during mild hypothermia. Using MCRE together with the hCMV promoter

increased expression of erythropoietin (EPO), a therapeutic protein, by 6-fold in cultures shifted to 32°C as compared to cultures at 37°C (Sumitomo et al. 2012). Other genes with expression driven by promoters of potential interest can be identified by culturing the cells in the desired conditions and using either oligo-dT probes to detect highly expressing mRNAs (Prentice et al. 2007) or high-throughput micro-arrays (Le et al. 2013; Thaisuchat et al. 2011). The ferritin heavy chain gene (FerH) was identified to be preferentially expressed at high levels during the stationary stages of culture using biotin tagged, oligo-dT generated probes from day 5 cell culture samples. The generated probes were allowed to hybridize to a cDNA library generated from the same cell sample. A more abundant mRNA species would give a more intense signal. The FerH promoter and its flanking regions gave expression three-fold higher than the hCMV promoter during late stages of culture to maximize protein expression during peak cell density (Prentice et al. 2007).

Therapeutic protein-producing mammalian cell cultures can be cultured under hypothermic conditions to slow down cell metabolism and extend cell viability for increased product yield. The S100a6 gene was identified in CHO microarray data to be highly expressed during hypothermia treatment and a 1.5 kb region upstream of the gene was isolated by screening CHO genomic lambda phage libraries (Thaisuchat et al. 2011). The identified promoter was two-fold better than the SV40 promoter during transient recombinant protein expression in hypothermic cultures. With the release of the CHO genome (Xu et al. 2011), tedious genomic library screens can be avoided as the upstream sequence of any endogenous genes can be located easily and cloned using PCR.

Microarray data was first used to identify genes highly expressed at late stages of cell

culture. Upstream sequences were then isolated based on the known genomic sequencing data. This method allowed identification of an interesting pTXnip dynamic promoter which was used to express a mGLUT5 fructose transporter (Le et al. 2013). pTXnip promoter allowed dynamic expression of the recombinant mGLUT5 gene, changing the expression levels in synchrony with the cell growth-staying low during early cell culture and increasing during the later stages of culture. Fructose consumption increased at later stages of culture when mGLUT5 expression was sufficiently high to avoid glucose depletion. While microarray data allows high-throughput screening of genes, it is limited to known genes and can be hard to implement for novel cell hosts which are not well studied.

The activity level of a promoter when expressing a gene depends on the interactions between protein *trans*-activating factors and the promoter sequence, allowing us to identify promoters by their DNA-protein activator interactions. Chromatin immunoprecipitation (ChIP) can be combined with massive parallel sequencing (NGS) to screen for promoters which bind to specific factors. Tissue specific transcription factors, like p300 for heart related enhancers, can be used to identify cell specific promoter elements (Blow et al. 2010). Performing either formaldehyde-assisted isolation of regulatory elements (FAIRE) or DNaseI enzymatic digestion to enrich samples combined with high-throughput assays, such as NGS and micro-arrays to identify the samples, are novel methods to identify regions of active chromatin, to enrich for active promoters and to isolate DNA regions that bind transcription factors (Furey 2012; Song et al. 2011). In FAIRE-seq, formaldehyde is added to cultured cells to cross-link DNA to proteins. The sample is sheared by sonication and DNA is extracted using phenol/chloroform for

identification (Giresi et al. 2007). DNase-seq enriches and identifies open chromosomal areas prone to DNaseI enzymatic digestion. These methods have mainly been applied to study and annotate the human genome. With human cell lines like PER.C6 (Kuczewski et al. 2011) being explored for therapeutic protein production, the promoters identified from these studies could be beneficial when human cell lines gain wider usage for biologics production.

With enhanced understanding of promoter sequence compositions, bioinformatics algorithms have been developed for sorting, analyzing and extracting succinct information from the large amounts of data obtained from genomics studies for prediction of novel promoters. The availability of the CHO genome allows application of these techniques to identifying promoters from the most commonly used mammalian host for biologics production. Basic screening of motifs like the TATA box is problematic in huge datasets. Instead, DNA structural properties like GC content, stabilizing energy, denaturation values and DNA-bending stiffness can be analyzed for promoter prediction (Abeel et al. 2008; Gan et al. 2012). Analysis can also be sequence based if there is a sufficiently large training set available. A support vector machine (SVM) utilized a previously generated set of data to produce~~t~~ k-mers that would identify enhancers without being restricted purely to transcription factor binding sequences (Lee et al. 2011). Several algorithms can also be used in sequence as a hybrid approach to identifying promoters (Lin and Li 2011).

Simply identifying a promoter region can be insufficient to achieve high therapeutic recombinant protein expression or to obtain the desired range of expression of

genes for cell line development and engineering. The promoter would require further engineering to optimize the activity and fine-tune the derived expression levels.

Engineering promoters to improve gene expression

Identified promoter regions can span several thousand base pairs and be cumbersome to use during vector construction. The promoters are subsequently engineered to only retain critical regulatory regions. Serial deletion analysis is one of the simpler ways to engineer an optimal promoter. Early deletion analysis was performed based on the sequence data and used restriction enzymes to systematically remove sequence repeats to study the effect of each segment to the strength of the promoter (Boshart et al. 1985; Dorsch-Hasler et al. 1985; Nelson et al. 1987). As knowledge of promoter structure improved, deletion analysis became more systematic and was based on deletion of known segments. The full length hCMV promoter was dissected to determine which regions are critical to transgene expression in CHO-K1 and HEK293 cells. Using the enhancer and core promoter as a control, it was observed that the full length promoter gave the highest stable EPO expression in CHO-K1 cells while addition of only the 1^{st} exon gave the highest expression in HEK293 cells (Mariati et al. 2010). Several other reports of promoter identification also performed deletion analysis to optimize the promoter (Prentice et al. 2007; Thaisuchat et al. 2011). The same process can also be applied to promoters for weakening its activity instead of strengthening. A series of SV40 promoters with different regions deleted were tested to weaken the expression of a glutamine synthetase for improved selection stringency of high producing cells (Fan et al. 2013).

The promoter can also be engineered by either removing sequences/motifs detrimental to expression or introducing factors beneficial to gene expression. For example, the PDX1 binding region in the hCMV promoter was shown to be a repressor and removal of the site by mutation increased expression by up to four-fold in transient luciferase experiments (Chao et al. 2004). Removal of CpG dinucleotide sites which are prone to gene silencing due to DNA methylation can improve transgene expression stability during long term culture (Swindle et al. 2004). An interesting way to introduce protein factors beneficial to gene expression is to link the factors to other DNA binding proteins. Recombinant protein expression was positively influenced in CHO cells, with increases in expression levels and stability, when the p300 histone acetyltransferase (HAT) domain was fused to a LexA-binding protein and a LexA-binding site was placed upstream of a hCMV promoter (Kwaks et al. 2005). In another example, an engineered zinc-finger protein (ZFP) transcription factor of the VP16 activation domain was fused to the ZFP DNA binding moiety to generate the novel factor ZFP-2392v. It was used to target a site upstream of the SV40 promoter, improving protein expression by up to two-fold (Reik et al. 2007). Positive effects were also observed after inserting binding sites for ZFP-2392v upstream of the hCMV promoter.

The enhancer and core promoter regions should be able to recruit transcription factors which work in synergy to improve transcription activity. As the enhancer and core promoter can be regarded as modular components, it is possible to fuse different enhancers and core promoters to enhance transcription activity or regulate expression (Blazeck and Alper 2013). For example, the combination of the hCMV enhancer with the chicken β-actin promoter provided transgene expression higher than the hCMV promoter

(Niwa et al. 1991). Use of a hCMV enhancer and HEF1α core promoter hybrid was beneficial to extending transgene expression for gene therapy in liver cells (Magnusson et al. 2011).

Another possible way to improve enhancer and promoter activity is by artificially directing and accelerating evolution of the sequences through the introduction of mutations. Enhancer libraries can be generated by polymerase cycling assembly using long overlapping oligonucleotides that have mutations introduced through a programmed level of degeneracy (Patwardhan et al. 2012). Promoter or enhancer libraries can also be generated utilizing DNA microarray printing. By controlling the depurination side reaction through a novel detritylation process, high quality 150-mer oligonucleotide libraries can be synthesized (LeProust et al. 2010). Enhancers with either single or multi-site mutagenesis were printed using the depurination-controlled process for systematic dissection studies and enhancer optimization (Melnikov et al. 2012). There is a need for high-throughput methods to efficiently screen the large libraries generated. One screening method incorporates fully degenerate short 20 base pair tags at the 3'-UTR end of enhancers in a library to create a catalog of enhancer-tag combinations for identification. Restriction sites can be added in between the enhancer and the tag for insertion of transgenes, leaving the tag between the transgene and the polyadenylation signal. The abundance of each tag after transfection and transcription can be used to determine the effect each modification has on enhancer activity (Melnikov et al. 2012; Patwardhan et al. 2012).

Synthetic enhancers and core promoters can be constructed either randomly or based on known promoter characteristics like transcription factor binding sites or CpG

content. Short synthetic oligonucleotides of binding sites for nuclear factor κB, activating protein-1, CArG binding factor A and nuclear factor Y were randomly ligated to generate a promoter with expression 3-fold higher than the hCMV promoter (Ogawa et al. 2007). In another study, promoter regions were given an alpha score based on how unlike a gene the nucleotide composition was. By replacing regions of a X-linked gene cancer/testis antigen 1A promoter with sequence that had double the alpha score resulted in increased expression (Grabherr et al. 2011). A novel synthetic promoter was generated by aligning the sequence of the hCMV and HEF1α promoters to obtain a consensus sequence which was modified to include a maximal number of transcription factor binding sites. This novel synthetic promoter with high transcription factor binding was combined with the hCMV enhancer to obtain transgene expression similar to the hCMV promoter but with improved long-term expression (Magnusson et al. 2011). In a study of the enhancer activity for random combinations of short DNA sequences, 100 base pair synthetic enhancers were generated by creating tandem repeats of every possible 10-mer DNA sequence by microarray printing. One enhancer worked twice as well as the human CMV enhancer in HeLa cells (Schlabach et al. 2010).

Introns can affect transcription by containing regions with enhancer- or repressor-like elements, containing splicing signals that enhance transcription initiation and RNA polymerase II activity or allowing formation of ordered nucleosome arrays around the promoter (Le Hir et al. 2003). Inclusion of introns into a basic enhancer-core promoter combination improves recombinant protein expression levels in mammalian cells (Kang et al. 2005; Kim et al. 2002; Mariati et al. 2010; Nott et al. 2003). Expression levels in CHO cells using the human and murine CMV promoters were improved by two-fold and

eight-fold, respectively, with addition of the first intron from the human EF-1α gene (Kim et al. 2002). A truncated version of the hCMV intron A, the Δ600 variant that was less than half the length of an original intron A with splicing acceptor sites preserved, was shown to perform as well the full length version for antibody expression in CHO-K1 cells (Quilici et al. 2013).

Promoter engineering maximizes the potential of both existing and newly identified promoters. A combination of different promoter engineering techniques might be required to yield the best results. For example, first shortening the promoter by deletion analysis and followed by designing hybrid combinations, introducing various mutations or testing the inclusion of introns to improve recombinant protein expression. With greater understanding of promoter function, designing synthetic promoters could be the most flexible option.

Conclusions

Promoter choice is critical to getting high and stable therapeutic recombinant protein expression and controlled expression levels for cell engineering in mammalian cells. While a long list of promoters have been identified and catalogued in online databases, any identified or engineered promoter still needs to be carefully characterized before use during protein production. A database of characterized promoters used either with, or derived from, mammalian cells, such as CHO cells, for therapeutic protein production is currently unavailable. Such a list would be a valuable resource for anyone trying to establish cell lines for biologics production. A gap also still exists before we are able to fully understand the complex regulation of promoter activity level and sustainability upon

integration into the chromosome. Bridging of this gap coupled with the promising work done on engineering synthetic promoters, would allow the design of custom promoters for any applications to be a feasible method to optimize protein expression level.

Acknowledgements

This work was supported by the Biomedical Research Council/Science and Engineering Research Council of A*STAR (Agency for Science, Technology and Research), Singapore.

References

Abeel T, Saeys Y, Bonnet E, Rouzé P, Van De Peer Y (2008) Generic eukaryotic core promoter prediction using structural features of DNA. Genome Res 18: 310-323.

Addison CL, Hitt M, Kunsken D, Graham FL (1997) Comparison of the human versus murine cytomegalovirus immediate early gene promoters for transgene expression by adenoviral vectors. J Gen Virol 78: 1653-1660.

Aggarwal S (2012) What's fueling the biotech engine - 2011 to 2012. Nat Biotechnol 30: 1191-1197.

Barnes LM, Bentley CM, Dickson AJ (2003) Stability of protein production from recombinant mammalian cells. Biotechnol Bioeng 81: 631-639.

Birch JR, Racher AJ (2006) Antibody production. Adv Drug Deliver Rev 58: 671-685.

Blazeck J, Alper HS (2013) Promoter engineering: Recent advances in controlling transcription at the most fundamental level. Biotechnol J 8: 46-58.

Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, et al (2010) ChIP-Seq identification of weakly conserved heart enhancers. Nat Genet 42: 806-810.

Boshart M, Weber F, Jahn G (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell 41: 521-530.

Bulger M, Groudine M (2010) Enhancers: The abundance and function of regulatory sequences beyond promoters. Dev Biol 339: 250-257.

Chao SH, Harada JN, Hyndman F, Gao X, Nelson CG, et al (2004) PDX1, a cellular homeoprotein, binds to and regulates the activity of human cytomegalovirus immediate early promoter. J Biol Chem 279: 16111-16120.

Chen J, Haverty J, Deng L, Li GH, Qiu P, et al (2013) Identification of a novel endogenous regulatory element in Chinese hamster ovary cells by promoter trap. J Biotechnol 167: 255-261.

Chusainow J, Yang YS, Yeo JH, Toh PC, Asvadi P, et al (2009) A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? Biotechnol Bioeng 102: 1182-1196.

Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. Annu Rev Biochem 78: 273-304.

Costa AR, Rodrigues ME, Henriques M, Azeredo J, Oliveira R (2010) Guidelines to cell engineering for monoclonal antibody production. Eur J Pharm Biopharm 74: 127-138.

Deer JR, Allison DS (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.

Dorsch-Hasler K, Keil GM, Weber F (1985) A long and complex enhancer activates transcription of the gene coding for the highly abundant immediate early mRNA in murine cytomegalovirus. Proc Natl Acad Sci USA 82: 8325-8329.

Fan L, Kadura I, Krebs LE, Larson J, Bowden D, et al (2013) Development of a highly-efficient CHO cell line generation system with engineered SV40E promoter. J Biotechnol 168: 652-658.

Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet 13: 840-852.

Gan Y, Guan J, Zhou S (2012) A comparison study on feature selection of DNA structural properties for promoter prediction. BMC Bioinform 13: 4.

Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 17: 877-885.

Grabherr MG, Pontiller J, Mauceli E, Ernst W, Baumann M, et al (2011) Exploiting nucleotide composition to engineer promoters. PLoS ONE 6: e20136.

Ho SCL, Tong YW, Yang Y (2013) Generation of monoclonal antibody-producing mammalian cell lines. Pharm Bioproc 1: 71-87.

Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol 339: 225-229.

Kalwy S (2005) Towards stronger gene expression - a promoter's tale. 19th European Society for Animal Cell Technology (ESACT) Meeting, Harrogate, England.

Kang M, Kim SY, Lee S, Lee YK, Lee J, et al (2005) Human β-globin second intron highly enhances expression of foreign genes from murine cytomegalovirus immediate-early promoter. J Microbiol Biotech 15: 544-550.

Kim DW, Uetsuki T, Kaziro Y, Yamaguchi N, Sugano S (1990) Use of the Human Elongation Factor-1-Alpha Promoter as a Versatile and Efficient Expression System. Gene 91: 217-223.

Kim M, O'Callaghan PM, Droms KA, James DC (2011) A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol Bioeng 108: 2434-2446.

Kim SY, Lee JH, Shin HS, Kang HJ, Kim YS (2002) The human elongation factor 1 alpha (EF-1α) first intron highly enhances expression of foreign genes from the murine cytomegalovirus promoter. J Biotechnol 93: 183-187.

Kling J (2012) Fresh from the biotech pipeline 2011. Nat Biotechnol 30: 128-131.

Kuczewski, M., Schirmer, E., Lain, B. and Zarbis-Papastoitsis, G. (2011) A single-use purification process for the production of a monoclonal antibody produced in a PER.C6 human cell line. Biotechnol J 6: 56-65.

Kwaks TH, Sewalt RG, van Blokland R, Siersma TJ, Kasiem M, et al (2005) Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. J Biotechnol 115: 35-46.

Le H, Vishwanathan N, Kantardjieff A, Doo I, Srienc M, et al (2013) Dynamic gene expression for metabolic engineering of mammalian cells in culture. Metab Eng 20: 212-220.

Le Hir H, Nott A, Moore MJ (2003) How introns influence and enhance eukaryotic gene expression. Trends Biochem Sci 28: 215-220.

Lee D, Karchin R, Beer MA (2011) Discriminative prediction of mammalian enhancers from DNA sequence. Genome Res 21: 2167-2180.

LeProust EM, Peck BJ, Spirin K, McCuen HB, Moore B, et al (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. Nucleic Acids Res 38: 2522-2540.

Lin H, Li QZ (2011) Eukaryotic and prokaryotic promoter prediction using hybrid approach. Theor Biosci 130: 91-100.

Magnusson T, Haase R, Schleef M, Wagner E, Ogris M (2011) Sustained, high transgene expression in liver with plasmid vectors using optimized promoter-enhancer combinations. J Gen Med 13: 382-391.

Malik S, Roeder RG (2010) The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet 11: 761-772.

Mariati, Ng YK, Chao SH, Yap MG, 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.

Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, et al (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol 30: 271-277.

Nelson JA, Reynolds-Kohler C, Smith BA (1987) Negative and positive regulation by a short segment in the 5'-flanking region of the human cytomegalovirus major immediate-early gene. Mol Cell Biol 7: 4125-4129.

Niwa H, Yamamura K, Miyazaki J (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene 108: 193-200.

Nott A, Meislin SH, Moore MJ (2003) A quantitative analysis of intron effects on mammalian gene expression. RNA 9: 607-617.

Ogawa R, Kagiya G, Kodaki T, Fukuda S, Yamamoto K (2007) Construction of strong mammalian promoters by random cis-acting element elongation. Biotechniques 42: 628-633.

Ong CT, Corces VG (2011) Enhancer function: New insights into the regulation of tissue-specific gene expression. Nat Rev Genet 12: 283-293.

Osterlehner A, Simmeth S, Göpfert U (2011) Promoter methylation and transgene copy numbers predict unstable protein production in recombinant chinese hamster ovary cell lines. Biotechnol Bioeng 108: 2670-2681.

Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, et al (2012) Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol 30: 265-270.

Pedersen AG, Baldi P, Chauvin Y, Brunak S (1999) The biology of eukaryotic promoter prediction - A review. Comput Chem 23: 191-207.

Pham P, Kamen A, Durocher Y (2006) Large-Scale transfection of mammalian cells for the fast production of recombinant protein. Mol Biotechnol 34: 225-237.

Pontiller J, Gross S, Thaisuchat H, Hesse F, Ernst W (2008) Identification of CHO endogenous promoter elements based on a genomic library approach. Mol Biotechnol 39: 135-139.

Pontiller J, Maccani A, Baumann M, Klancnik I, Ernst W (2010) Identification of CHO endogenous gene regulatory elements. Mol Biotechnol 45: 235-240.

Prentice HL, Tonkin CJD, Caamano L, Sisk WP (2007) High level expression of proteins using sequences from the ferritin heavy chain gene locus. J Biotechnol 128: 50-60.

Quilici LS, Silva-Pereira I, Andrade AC, Albuquerque FC, Brigido MM, et al (2013) A minimal cytomegalovirus intron A variant can improve transgene expression in different mammalian cell lines. Biotechnol Lett 35: 21-27.

Reik A, Zhou Y, Collingwood TN, Warfe L, Bartsevich V, et al (2007) Enhanced protein production by engineered zinc finger proteins. Biotechnol Bioeng 97: 1180-1189.

Schlabach MR, Hu JK, Li M, Elledge SJ (2010) Synthetic design of strong promoters. Proc Natl Acad Sci USA 107: 2538-2543.

Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. Annu Rev Biochem 72: 449-479.

Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, et al (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res 21: 1757-1767.

Sumitomo Y, Higashitsuji H, Liu Y, Fujita T, Sakurai T, et al (2012) Identification of a novel enhancer that binds Sp1 and contributes to induction of cold-inducible RNA-binding protein (cirp) expression in mammalian cells. BMC Biotechnol 12.

Swindle CS, Kim HG, Klug CA (2004) Mutation of CpGs in the murine stem cell virus retroviral vector long terminal repeat represses silencing in embryonic stem cells. J Biol Chem 279: 34-41.

Thaisuchat H, Baumann M, Pontiller J, Hesse F, Ernst W (2011) Identification of a novel temperature sensitive promoter in CHO cells. BMC Biotechnol 11.

Thomsen DR, Stenberg RM, Goins WF, Stinski MF (1984) Promoter-regulatory region of the major immediate early gene of human cytomegalovirus. Proc Natl Acad Sci USA 81: 659-663.

Uetsuki T, Naito A, Nagata S, Kaziro Y (1989) Isolation and characterization of the human chromosomal gene for polypeptide chain elongation factor-1 alpha. J Biol Chem 264: 5791-5798.

Weber F, de Villiers J, Schaffner W (1984) An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell 36: 983-992.

Wurm FM (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22: 1393-1398.

Xia W, Bringmann P, McClary J, Jones PP, Manzana W, et al (2006) High levels of protein expression using different mammalian CMV promoters in several cell lines. Protein Expr Purif 45: 115-124.

Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, et al (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 29: 735-741.

Yang Y, Mariati, Chusainow J, Yap MG (2010) DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. J Biotechnol 147: 180-185.

Table 1. Online promoter databases

EPD, Eukaryotic Promoter Database

http://epd.vital-it.ch/

Managed by the Swiss Institute of Bioinformatics, EPD is an annotated and experimentally validated collection of eukaryotic promoters from human, mouse, *D. melanogaster*, and zebrafish genomes. Over 200,000 promoters can be accessed in the database.

MPromDb, Mammalian Promoter Database

http://mpromdb.wistar.upenn.edu/

MPromDb is a curated database of annotated gene promoters identified from ChIP-seq experiment results. Datasets are obtained from six different human cell samples (CD4+TCells, HeLa S3, K562, NB4, Lymphyoblastoid, Jurkat) and mouse samples from five different mouse tissues and five different cell types.

PEDB, Mammalian Enhancer/Promoter Database

http://promoter.cdb.riken.jp/

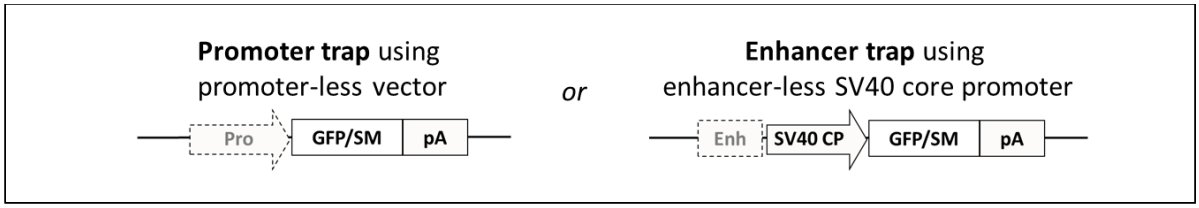
Mammalian promoter/enhancer database constructed by integrating information of conserved non-coding regions, transcriptional start sites and transcription factor biding sites to obtain a system level understanding of dynamic transcriptional regulation in mammals.

Table 2. Promoters used in commercial vectors for recombinant protein expression in mammalian cells.

Promoter	Vector	Supplier	Comments
hCMV (human cytomegalovirus promoter)	pcDNA3.1, pcDNA3.3/3.4-TOPO, pOptiVEC-TOPO, pBudCE4.1	Life Technologies (www.lifetechnologies.com)	Each vector uses different variants of the hCMV promoter. Refer to supplier database for more information.
	gWiz vector series	Genlantis (www.genlantis.com)	
	phCMV vector series		
	pCI/neo	Promega (www.promega.com)	
	pFLAG-CMV and pBICEP-CMV series of vectors	Sigma-Aldrich (www.sigmaaldrich.com)	
mCMV (murine cytomegalovirus promoter)	GS vector	Lonza (www.lonza.com)	Intron is included downstream of the promoter. Used in the popular GS expression system.
SV40 (simian virus 40 promoter)	pZeoSV2(+)	Life Technologies	SV40 is often used for expression of selection marker genes for stringent selection.
	pSI/neo	Promega	
	pcDNA3.1, pcDNA3.3-TOPO, pOptiVEC-TOPO, GS vector	Various	

	pBudCE4.1	Life Technologies	
HEF1α (human elongation factor 1α gene promoter)	pFUSE2-CLIg-hk pFUSE-CHIg-hG1	InvivoGen (www.invivogen.com)	Vectors specifically designed for Ig expression with light chain and heavy chain constant regions on the vector.
CHOEF1α (CHO elongation factor 1α gene promoter	pDEF38	CMC biologics (www.cmcbio.com)	5’ and 3’ flanking regions and the first intron of CHO EF-1α are included.
FerH, FerL (human ferritin heavy and light chain gene promoters)	pVITRO2-MCS	Invivogen	5’UTR is modified to reduce iron regulation of promoters.
CAG (CMV enhancer/ß-actin promoter)	pCAGG	Gene Bridges (www.genebridges.com)	

Fig. 1 "Trapping" of novel promoters and enhancers. Plasmid vectors that are either lacking a promoter or carrying a weak SV40 core promoter can be used as a promoter trap or enhancer trap respectively. The trap vectors can be used in two ways: (1) Directly transfecting the trap vector into the host cell. In this method, the green fluorescence protein (GFP) reporter or selection marker (SM) would only be expressed when the vector is integrated downstream of an endogenous promoter or enhancer, (2) Shotgun cloning DNA fragments from the target host into the trap vector before transfection. The fragments integrated would need to be a promoter or enhancer for the GFP or SM to be expressed. GFP screening or drug selection is subsequently performed to isolate positive cells. Sequence of the segments upstream of the trap vector or the ligated fragment can be identified by methods like nested PCR or genomic library screens. The novel promoter or enhancer would be further evaluated and engineered before use. Pro: Promoter; Enh: Enhancer; RDS: Random DNA segments.



Approach (1)

Trap vector transfected into cells.

Vector is integrated into chromosome.

RDS GFP/SM pA

RDS SV40 CP GFP/SM pA

Approach (2)

DNA from target is fragmented by enzymatic or physical means like sonication, nebulization.

Fragments ligated onto trap vector

RDS GFP/SM pA

RDS SV40 CP GFP/SM pA

Transfected into cells.

Antibiotics selection or screen of GFP reporter protein.

Surviving or selected cells have trap vectors either,
(1) integrated downstream of a regulatory element , or
(2) ligated to a regulatory element .

Isolate/identify sequences upstream of gene/SV40 CP

- Nested PCR from vector backbone
- Genomic library screen

Evaluation and engineering of promoter or enhancer

Fig. 1

For submission to Biotechnology Letters

Review article

Section in which the paper is to be considered: Animal cell technology

Identifying and engineering promoters for high level and sustainable therapeutic recombinant protein production in cultured mammalian cells

Steven C.L. Ho[1,2], Yuansheng Yang[1,3*]

[1] Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), 20 Biopolis Way, #06-01 Centros, Singapore 138668

[2] Division of Bioengineering, Department of Chemical and Biomolecular Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119620

[3]School of Chemical and Biomedical Engineering, Nanyang Technological University, N1.2-B2-33, 62 Nanyang Avenue, Singapore 637459

[*] Correspondence to:
Yuansheng Yang
Phone: (65)-64070825, Fax: (65)-64789561
Email: yang_yuansheng@bti.a-star.edu.sg

Abstract

Promoters are essential on plasmid vectors to initiate transcription of the transgenes when generating therapeutic recombinant protein expressing mammalian cell lines. High and sustained gene expression levels are desired during therapeutic protein production while fine control of gene expression is useful for cell engineering. As many promoters commonly exhibit cell and product specificity, there is a need for new promoters to be identified, optimized and carefully evaluated before use. Suitable promoters can be identified using techniques ranging from simple molecular biology methods to modern high-throughput omics screenings. Promoter engineering is often required after identification to either obtain high and sustained expression or to provide a wider range of gene expression. This review discusses some of the available methods ~~for~~to identify~~ing~~ and engineer~~ing~~ promoters for ~~use in~~ recombinant therapeutic protein expression in mammalian cells.

Keywords (up to 7):

Mammalian cells, Promoter engineering, Promoter identification, Recombinant protein production

Introduction

There is rapidly growing demand for existing products and a constant inflow of novel products in the recombinant therapeutic biologics market. Monoclonal antibodies (mAb), the best-selling class of biologics, accounted for sales of over $20.3 billion in the USA in 2011 (Aggarwal 2012). Biosimilar versions for some of the top therapeutic proteins are also being developed and getting approved for the market (Kling 2012). With the expected sustained growth of the biologics market in the near future, cell culture technologies for therapeutic recombinant protein production are required to improve as well.

Mammalian cells like the Chinese hamster ovary (CHO) cells, human embryonic kidney cells (HEK293) and murine lymphoid cell lines are commonly used for production of safe and effective biologics (Ho et al. 2013; Wurm 2004). Generating a therapeutic recombinant protein producing cell line starts with transfecting the mammalian host cells with a plasmid vector carrying the gene for the respective recombinant protein (Birch and Racher 2006; Ho et al. 2013; Smale and Kadonaga 2003). Small amounts of transiently expressed product can be collected a few days after transfection for small scale assays during early stages of drug development (Pham et al. 2006). Stably transfected cell lines which have undergone selection for plasmid vectors integrated into the host cell's genome are used for commercial scale therapeutic protein production at high yields and consistent quality (Barnes et al. 2003). It is desired that the recombinant protein expression be high in both transiently and stably transfected cell lines. There is an added requirement of sustained expression over a period of 2-3 months

in stably transfected cell lines during the scale up process to ensure consistent product levels and quality (Barnes et al. 2003).

Obtaining sustained, high level recombinant protein production depends on many factors such as cell host, culture media, culture process, and expression vectors (Birch and Racher 2006; Costa et al. 2010; Ho et al. 2013). The promoter is an essential component of the expression vector which initiates transcription and can affect both expression level and stability. Promoters can be broadly classified into inducible and constitutive promoters. Constitutive promoters are commonly used for recombinant protein production, while inducible promoters are occasionally used either for cell engineering or expressing products toxic to the cell. This review will introduce the commonly used promoters and how novel promoters can be identified and engineered either for high and sustained gene expression or for cell engineering in mammalian cells, with emphasis placed on therapeutic protein production.

Promoters for therapeutic recombinant protein production

A typical promoter used for therapeutic recombinant protein production contains a core promoter element, an immediate upstream enhancer region and if required, other *cis*-acting regulatory elements. The core promoter is a minimal DNA sequence of around 80 nucleotides required for transcription initiation which extends approximately 35 base pairs upstream and downstream of the initiation site (Smale and Kadonaga 2003). Sequence motifs found in well-studied TATA-box based core promoters includes the TATA box (TATA box binding protein recognition sequence, TATAWAAR), BRE (Transcription factor IIB recognition sequence, SSRGCC), Inr (initiator sequence,

YYANWYY), MTE (motif ten element, CSARCSSAAC), DPE (downstream promoter element, RGWYV) (Juven-Gershon and Kadonaga 2010; Pedersen et al. 1999; Smale and Kadonaga 2003). These motifs allow anchoring of transcription factors to form a transcription machinery that drives transcription by RNA polymerase II (Juven-Gershon and Kadonaga 2010). Apart from TATA-box based promoters, CpG rich mammalian promoters which are more plastic and evolvable have been identified with modern genomic tools. These promoters have multiple transcription start sites as compared to focused start sites of TATA-box based promoters (Juven-Gershon and Kadonaga 2010). After transcription is initiated by the core promoter, the gene expression is augmented by DNA elements like enhancers and introns. Enhancers are able to recruit a variety of transcription factors to activate transcription and can function independent of location and orientation (Ong and Corces 2011). The enhancer bound transcription factors regulate promoter activity either by direct interaction with the basic transcription machinery (Malik and Roeder 2010) or by forming complexes and recruiting enzymes to remodel the chromatin structure to increase accessibility (Clapier and Cairns 2009).

There are currently countless promoters identified and catalogued for both eukaryotic and viral sources in online databases (Table 1). A promoter derived from the human cytomegalovirus major immediate early gene (hCMV) is the most commonly used promoter for high therapeutic recombinant protein expression in industry (Costa et al. 2010). Other promoters that can be used include promoters from the murine cytomegalovirus (mCMV), simian virus 40 (SV40), CHO and human elongation factor-1 α-1 gene (CHOEF1α, HEF1α) and a hybrid promoter of hCMV enhancer and chicken ß-actin promoter (CAG). These promoters are used on commercial plasmids for

constructing vectors to express therapeutic recombinant proteins (Table 2). The SV40 promoter is considered as a weaker promoter among the above list and is mostly used to drive expression of the selection marker gene for more stringent selection of stably transfected high producing cells (Birch and Racher 2006).

Due to the plethora of available promoters, the choice of promoters for recombinant protein expression and engineering of mammalian cells can be complicated as the promoters and *cis*-regulatory elements, like introns, have shown to exhibit product specificity, cell line specificity and even differences in transient and stable conditions for recombinant protein expression in mammalian cells (Kim et al. 2002; Mariati et al. 2010; Xia et al. 2006). For instance, mCMV promoter was shown to be stronger than hCMV in a few papers (Addison et al. 1997; Kim et al. 2002). However, when these two promoters were compared for mAb production in CHO cells in Lonza's GS system™, one of the commonly used vectors for mammalian gene expression in industry, mCMV had lower expression than hCMV (Kalwy 2005). Another possible reason for conflicting reports in various studies is that different versions of truncated or full-length promoter were used. For instance, the hCMV promoter in some vector only contains the enhancer and core promoter, while others could include the first intron and other cis-regulatory elements (Mariati et al. 2010). Evaluation should ~~ideally~~ be performed when choosing promoters for different purposes instead of ~~purely~~ depending ~~based~~ on the results published in literatures.

While expression sustainability is less frequently reported compared to expression level, it is another critical issue faced during recombinant protein production using stably transfected cell lines and cell engineering. There are multiple reports linking the loss of

gene expression in CHO cells expressing therapeutic recombinant protein upon long term culture to the use of the hCMV promoter (Chusainow et al. 2009; Kim et al. 2011; Osterlehner et al. 2011; Yang et al. 2010). Although the use of endogenous mammalian promoters is thought to be advantageous compared to viral promoters for transgene expression stability (Chen et al. 2013), there is a shortage of data comparing promoter expression stability in industrial applications. We have observed in our lab that the SV40 promoter was most resistant to gene silencing but gave lowest expression when compared to hCMV, mCMV, and CHOEF1α in stably transfected CHO cell clones (results not published). While therapeutic protein production demands for the highest possible expression levels, cell engineering experiments to improve the therapeutic protein expression or quality can require protein expression either at lower levels or under specific conditions, further complicating promoter selection process (Blazeck and Alper 2013; Chen et al. 2013). Due to the absence of a promoter that performs well for all applications, work is always on-going to identify and engineer novel promoters to complement new products and cell lines.

Identifying novel promoters

Identifying a promoter that can achieve high and sustained gene expression is challenging, especially in scenarios where novel cell hosts are to be utilized. Promoters can either be isolated from viral sources that infect mammalian cells, for example, the hCMV (Thomsen et al. 1984) and mCMV promoters (Dorsch-Hasler et al. 1985) or from upstream sequences of the host's endogenous genes, like the HEF1α and CHOEF1α promoters (Deer and Allison 2004; Kim et al. 1990). Promoters can be identified using

methods ranging from using simple molecular biology techniques to utilizing high-throughput microarrays and genomics tools.

"Trapping" of promoters and regulatory elements can be performed to identify and study novel sequences (Fig. 1). This method of promoter identification is tedious due to its random nature but can be performed using basic molecular biology tools. A promoter trap utilizes a promoter-less vector carrying a green fluorescent protein (GFP) reporter gene (Chen et al. 2013). The vector was transfected into CHO cells and GFP expressing cells would signify that the vector was integrated adjacent to a transcriptionally active promoter. The upstream region was sequenced to identify a promoter which was 66% as active as the SV40 promoter. A variation of the promoter trap involved shotgun cloning of CHO genomic fragments into a promoter-less vector carrying an antibiotic selection marker gene. Genomic fragments which exhibit promoter activity can be isolated from clones which display antibiotic resistance (Pontiller et al. 2008). Enhancer trapping can be performed using vectors carrying a weak SV40 core promoter (Weber et al. 1984). Fragments from hCMV and mCMV were inserted in the enhancer trap, upstream of the SV40 core promoter, to substitute the missing enhancer. Sequences with enhancer properties that increase gene expression were identified (Boshart et al. 1985; Dorsch-Hasler et al. 1985). Another way to use the enhancer trap is to transfect cells with the enhancer-less vector. The transgenes would only be expressed when the vector is integrated downstream of an enhancer (Bulger and Groudine 2010). Although promoter and enhancer trapping is an arbitrary exercise, there is potential to identify novel promoters or regulatory segments that cannot be easily identified by screening for conserved motifs or in-silico promoter prediction.

An alternative approach to identify strong promoters is based on the expression level of endogenous genes in mammalian cells. For instance, elongation factor-1 α-1 gene (EF1α) is a house-keeping gene and known to be highly expressed in different mammalian cells. HEF1α and CHOEF1α promoters had been isolated using cDNA probes to screen CHO-K1 Lambda phage genomic libraries and confirmed to be strong promoters for high level transgene expression (Deer and Allison 2004; Uetsuki et al. 1989). CHOEF1α promoter when combined with its flanking elements gave reporter gene expression that was up to 35-folds higher than the hCMV promoter in CHO cells (Deer and Allison 2004). In another effort to identify endogenous promoters, a CHO DNA library was first built by blunt end cloning chromosomal DNA fragments into promoter-less vector backbones. Primers binding to the first exon of highly expressed CHO genes were designed and used together with primers binding to the backbone of the vector used for generating the library to clone the 5' flanking regions (Pontiller et al. 2010). The cloned 5' ends were inserted into reporter luciferase vectors to check for promoter activity. While no promoters stronger than SV40 were identified in the above experiment, the method is effective to isolate flanking promoter regions of known genes.

While highly expressed endogenous proteins are obvious targets, genes with interesting properties like elevated expression under specific conditions can also contain useful promoter elements. One example is the mild-cold responsive enhancer (MCRE) found upstream of a cold-inducible RNA-binding protein (cirp). Cirp expression is induced during mild hypothermia. Using MCRE together with the hCMV promoter increased expression of erythropoietin (EPO), a therapeutic protein, by 6-fold in cultures temperature shifted to 32°C as compared to cultures at 37°C (Sumitomo et al. 2012).

Other genes with expression driven by promoters of potential interest can be identified by culturing the cells in the desired conditions and using either oligo-dT probes to detect highly expressing mRNAs (Prentice et al. 2007) or high-throughput micro-arrays (Le et al. 2013; Thaisuchat et al. 2011). The ferritin heavy chain gene (FerH) was identified to be preferentially expressed at high levels during the stationary stages of culture using biotin tagged, oligo-dT generated probes from day 5 cell culture samples. The generated probes were allowed to hybridize to a cDNA library generated from the same cell sample. A more abundant mRNA species would give a more intense signal. The FerH promoter and its flanking regions gave expression three fold higher than the hCMV promoter during late stages of culture to maximize protein expression during peak cell density (Prentice et al. 2007). Therapeutic protein producing mammalian cell cultures can be cultured under hypothermic conditions to slow down cell metabolism and extend cell viability for increased product yield. The S100a6 gene was identified in CHO microarray data to be highly expressed during hypothermia treatment and a 1.5 kb region upstream of the gene was isolated by screening CHO genomic lambda phage libraries (Thaisuchat et al. 2011). The identified promoter was shown to be two fold better than the SV40 promoter during transient recombinant protein expression in hypothermic cultures. With the release of the CHO genome (Xu et al. 2011), tedious genomic library screens can be avoided as the upstream sequence of any endogenous genes can be located easily and cloned using PCR. Microarray data was first used to identify genes highly expressed at late stages of cell culture. Upstream sequences were then isolated based on the known genomic sequencing data. This method allowed identification of an interesting pTXnip dynamic promoter which was used to express a mGLUT5 fructose transporter (Le et al.

2013). pTXnip promoter allowed dynamic expression of the recombinant mGLUT5 gene, changing the expression levels in sync with the cell growth-staying low during early cell culture and increasing during the later stages of culture. Fructose consumption increased at later stages of culture when mGLUT5 expression was sufficiently high to avoid glucose depletion. While microarray data allows high-throughput screening of genes, it is limited to known genes and can be hard to implement for novel cell hosts which are not well studied.

The activity level of a promoter when expressing a gene depends on the interactions between protein *trans*-activating factors and the promoter sequence, allowing us to identify promoters by their DNA-protein activator interactions. Chromatin immunoprecipitation (ChIP) can be combined with massive parallel sequencing (NGS) to screen for promoters which bind to specific factors. Tissue specific transcription factors, like p300 for heart related enhancers, can be used to identify cell specific promoter elements (Blow et al. 2010). Performing either formaldehyde assisted isolation of regulatory elements (FAIRE) or DNaseI enzymatic digestion to enrich samples combined with high-throughput assays like NGS and micro-arrays to identify the samples are novel methods to identify regions of active chromatin, enrich for active promoters and isolate DNA regions that bind transcription factors (Furey 2012; Song et al. 2011). In FAIRE-seq, formaldehyde is added to cultured cells to crosslink DNA to proteins. The sample is sheared by sonication and DNA is extracted using phenol-chloroform for identification (Giresi et al. 2007). DNase-seq enriches and identifies open chromosomal areas prone to DNaseI enzymatic digestion. These methods have mainly been applied to study and annotate the human genome. With human cell lines like PER.C6 (Kuczewski et al. 2011)

being explored for therapeutic protein production, the promoters identified from these studies could be beneficial when human cell lines gain wider usage for biologics production.

With enhanced understanding of promoter sequence compositions, bioinformatics algorithms have been developed for sorting, analyzing and extracting succinct information from the large amounts of data obtained from genomics studies for prediction of novel promoters. The availability of the CHO genome allows application of these techniques to identifying promoters from the most commonly used mammalian host for biologics production. Basic screening of motifs like the TATA box is problematic in huge datasets. Instead, DNA structural properties like GC content, stabilizing energy, denaturation values and DNA-bending stiffness can be analyzed for promoter prediction (Abeel et al. 2008; Gan et al. 2012). Analysis can also be sequence based if there is a sufficiently large training set available. A support vector machine (SVM) utilized a previously generated data set to generate identifying k-mers to identify enhancers without being restricted purely to transcription factor binding sequences (Lee et al. 2011). Several algorithms can also be used in sequence as a hybrid approach to identifying promoters (Lin and Li 2011).

Simply identifying a promoter region can be insufficient to achieve high therapeutic recombinant protein expression or to obtain the desired range of expression of genes for cell line development and engineering. The promoter would require further engineering to optimize the activity and fine-tune the derived expression levels.

Engineering promoters to improve gene expression

Identified promoter regions can span several thousand base pairs and be cumbersome to use during vector construction. The promoters are subsequently engineered to only retain critical regulatory regions. Serial deletion analysis is one of the simpler ways to engineer an optimal promoter. Early deletion analysis was performed based on the sequence data and used restriction enzymes to systematically remove sequence repeats to study the effect of each segment to the strength of the promoter (Boshart et al. 1985; Dorsch-Hasler et al. 1985; Nelson et al. 1987). As knowledge of promoter structure improved, deletion analysis became more systematic and was based on deletion of known segments. The full length hCMV promoter was dissected to determine which regions are critical to transgene expression in CHO-K1 and HEK293 cells. Using the enhancer and core promoter as a control, it was observed that the full length promoter gave the highest stable EPO expression in CHO-K1 cells while addition of only the 1st exon gave the highest expression in HEK293 cells (Mariati et al. 2010). Several other reports of promoter identification also performed deletion analysis to optimize the promoter (Prentice et al. 2007; Thaisuchat et al. 2011). The same process can also be applied to promoters for weakening its activity instead of strengthening. A series of SV40 promoters with different regions deleted were tested to weaken the expression of a glutamine synthetase for improved selection stringency of high producing cells (Fan et al. 2013).

The promoter can also be engineered by either removing sequences/motifs detrimental to expression or introducing factors beneficial to gene expression. For example, the PDX1 binding region in the hCMV promoter was shown to be a repressor and removal of the site by mutation increased expression by up to four-fold in transient

luciferase experiments (Chao et al. 2004). Removal of CpG dinucleotide sites which are prone to gene silencing due to DNA methylation can improve transgene expression stability during long term culture (Swindle et al. 2004). An interesting way to introduce protein factors beneficial to gene expression is to link the factors to other DNA binding proteins. Recombinant protein expression was positively influenced in CHO cells, with increases in expression levels and stability, when the p300 histone acetyltransferase (HAT) domain was fused to a LexA binding protein and a LexA binding site was placed upstream of a hCMV promoter (Kwaks et al. 2005). In another example, an engineered zinc-finger protein (ZFP) transcription factor of the VP16 activation domain was fused to the ZFP DNA binding moiety to generate the novel factor ZFP-2392v. It was used to target a site upstream of the SV40 promoter, improving protein expression by up to two-fold (Reik et al. 2007). Positive effects were also observed after inserting binding sites for ZFP-2392v upstream of the hCMV promoter.

The enhancer and core promoter regions should be able to recruit transcription factors which work in synergy to improve transcription activity. As the enhancer and core promoter can be regarded as modular components, it is possible to fuse different enhancers and core promoters to enhance transcription activity or regulate expression (Blazeck and Alper 2013). For example, the combination of the hCMV enhancer with the chicken β-actin promoter provided transgene expression higher than the hCMV promoter (Niwa et al. 1991). Use of a hCMV enhancer and HEF1α core promoter hybrid was beneficial to extending transgene expression for gene therapy in liver cells (Magnusson et al. 2011).

Another possible way to improve enhancer and promoter activity is by artificially directing and accelerating evolution of the sequences through the introduction of mutations. Enhancer libraries can be generated by polymerase cycling assembly using long overlapping oligonucleotides that have mutations introduced through a programmed level of degeneracy (Patwardhan et al. 2012). Promoter or enhancer libraries can also be generated utilizing DNA microarray printing. By controlling the depurination side reaction through a novel detritylation process, high quality 150-mer oligonucleotide libraries can be synthesized (LeProust et al. 2010). Enhancers with either single or multi-site mutagenesis were printed using the depurination controlled process for systematic dissection studies and enhancer optimization (Melnikov et al. 2012). There is a need for high-throughput methods to efficiently screen the large libraries generated. One screening method incorporates fully degenerate short 20 base pair tags at the 3'UTR end of enhancers in a library to create a catalog of enhancer-tag combinations for identification. Restriction sites can be added in between the enhancer and the tag for insertion of transgenes, leaving the tag between the transgene and the polyadenylation signal. The abundance of each tag after transfection and transcription can be used to determine the effect each modification has on enhancer activity (Melnikov et al. 2012; Patwardhan et al. 2012).

Synthetic enhancers and core promoters can be constructed either randomly or based on known promoter characteristics like transcription factor binding sites or CpG content. Short synthetic oligonucleotides of binding sites for nuclear factor κB, activating protein-1, CArG binding factor A and nuclear factor Y were randomly ligated to generate a promoter with expression 3-fold higher than the hCMV promoter (Ogawa et al. 2007).

In another study, promoter regions were given an alpha score based on how unlike a gene the nucleotide composition was. By replacing regions of a X-linked gene cancer/testis antigen 1A promoter with sequence that had double the alpha score resulted in increased expression (Grabherr et al. 2011). A novel synthetic promoter was generated by aligning the sequence of the hCMV and HEF1α promoters to obtain a consensus sequence which was modified to include a maximal number of transcription factor binding sites. This novel synthetic promoter with high transcription factor binding was combined with the hCMV enhancer to obtain transgene expression similar to the hCMV promoter but with improved long-term expression (Magnusson et al. 2011). In a study of the enhancer activity for random combinations of short DNA sequences, 100 base pair synthetic enhancers were generated by creating tandem repeats of every possible 10-mer DNA sequence by microarray printing. One enhancer worked twice as well as the human CMV enhancer in HeLa cells (Schlabach et al. 2010).

Introns can affect transcription by containing regions with enhancer- or repressor-like elements, containing splicing signals that enhance transcription initiation and RNA polymerase II activity or allowing formation of ordered nucleosome arrays around the promoter (Le Hir et al. 2003). Inclusion of introns into a basic enhancer-core promoter combination has been shown to improve recombinant protein expression levels in mammalian cells (Kang et al. 2005; Kim et al. 2002; Mariati et al. 2010; Nott et al. 2003). Expression levels in CHO cells using the human and murine CMV promoters were improved by two-fold and eight-fold respectively with addition of the first intron from the human EF-1α gene (Kim et al. 2002). A truncated version of the hCMV intron A, the Δ600 variant that was less than half the length of an original intron A with splicing

acceptor sites preserved, was shown to perform as well the full length version for antibody expression in CHO-K1 cells (Quilici et al. 2013).

Promoter engineering maximizes the potential of both existing and newly identified promoters. A combination of different promoter engineering techniques might be required to yield the best results. For example, first shortening the promoter by deletion analysis and followed by designing hybrid combinations, introducing various mutations or testing the inclusion of introns to improve recombinant protein expression. With greater understanding of promoter function, designing synthetic promoters could be the most flexible option.

Conclusion

Promoter choice is critical to getting high and stable therapeutic recombinant protein expression and controlled expression levels for cell engineering in mammalian cells. While a long list of promoters have been identified and catalogued in online databases, any identified or engineered promoter still needs to be carefully characterized before use during protein production. A database of characterized promoters either used with or derived from mammalian cells for therapeutic protein production, like CHO cells, is currently unavailable. Such a list would be a valuable resource for anyone trying to establish cell lines for biologics production. A gap also still exists before we are able to fully understand the complex regulation of promoter activity level and sustainability upon integration into the chromosome. Bridging of this gap coupled with the promising work done on engineering synthetic promoters, would allow the design of custom promoters for any applications to be a feasible method to optimize protein expression level.

Acknowledgements

This work was supported by the Biomedical Research Council/Science and Engineering Research Council of A*STAR (Agency for Science, Technology and Research), Singapore.

References

Abeel, T., Saeys, Y., Bonnet, E., Rouzé, P. and Van De Peer, Y. (2008) Generic eukaryotic core promoter prediction using structural features of DNA. Genome Res 18: 310-323.

Addison, C. L., Hitt, M., Kunsken, D. and Graham, F. L. (1997) Comparison of the human versus murine cytomegalovirus immediate early gene promoters for transgene expression by adenoviral vectors. J Gen Virol 78: 1653-1660.

Aggarwal, S. (2012) What's fueling the biotech engine - 2011 to 2012. Nat Biotechnol 30: 1191-1197.

Barnes, L. M., Bentley, C. M. and Dickson, A. J. (2003) Stability of protein production from recombinant mammalian cells. Biotechnol Bioeng 81: 631-639.

Birch, J. R. and Racher, A. J. (2006) Antibody production. Adv Drug Deliver Rev 58: 671-685.

Blazeck, J. and Alper, H. S. (2013) Promoter engineering: Recent advances in controlling transcription at the most fundamental level. Biotechnol J 8: 46-58.

Blow, M. J., McCulley, D. J., Li, Z., Zhang, T., Akiyama, J. A., et al. (2010) ChIP-Seq identification of weakly conserved heart enhancers. Nat Genet 42: 806-810.

Boshart, M., Weber, F. and Jahn, G. (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell 41: 521-530.

Bulger, M. and Groudine, M. (2010) Enhancers: The abundance and function of regulatory sequences beyond promoters. Dev Biol 339: 250-257.

Chao, S. H., Harada, J. N., Hyndman, F., Gao, X., Nelson, C. G., et al. (2004) PDX1, a cellular homeoprotein, binds to and regulates the activity of human cytomegalovirus immediate early promoter. J Biol Chem 279: 16111-16120.

Chen, J., Haverty, J., Deng, L., Li, G. H., Qiu, P., et al. (2013) Identification of a novel endogenous regulatory element in Chinese hamster ovary cells by promoter trap. J Biotechnol 167: 255-261.

Chusainow, J., Yang, Y. S., Yeo, J. H., Toh, P. C., Asvadi, P., et al. (2009) A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? Biotechnol Bioeng 102: 1182-1196.

Clapier, C. R. and Cairns, B. R. (2009) The biology of chromatin remodeling complexes. Annu Rev Biochem 78: 273-304.

Costa, A. R., Rodrigues, M. E., Henriques, M., Azeredo, J. and Oliveira, R. (2010) Guidelines to cell engineering for monoclonal antibody production. Eur J Pharm Biopharm 74: 127-138.

Deer, J. R. and Allison, D. S. (2004) High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1α gene. Biotechnol Progr 20: 880-889.

Dorsch-Hasler, K., Keil, G. M. and Weber, F. (1985) A long and complex enhancer activates transcription of the gene coding for the highly abundant immediate early mRNA in murine cytomegalovirus. P Natl Acad Sci USA 82: 8325-8329.

Fan, L., Kadura, I., Krebs, L. E., Larson, J., Bowden, D., et al. (2013) Development of a highly-efficient CHO cell line generation system with engineered SV40E promoter. J Biotechnol 168: 652-658.

Furey, T. S. (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet 13: 840-852.

Gan, Y., Guan, J. and Zhou, S. (2012) A comparison study on feature selection of DNA structural properties for promoter prediction. BMC Bioinformatics 13: 4.

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. and Lieb, J. D. (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 17: 877-885.

Grabherr, M. G., Pontiller, J., Mauceli, E., Ernst, W., Baumann, M., et al. (2011) Exploiting nucleotide composition to engineer promoters. PLoS ONE 6: e20136.

Ho, S. C. L., Tong, Y. W. and Yang, Y. (2013) Generation of monoclonal antibody-producing mammalian cell lines. Pharm Bioprocess 1: 71-87.

Juven-Gershon, T. and Kadonaga, J. T. (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol 339: 225-229.

Kalwy, S. (2005) Towards stronger gene expression - a promoter's tale. 19th European Society for Animal Cell Technology (ESACT) Meeting, Harrogate, England.

Kang, M., Kim, S. Y., Lee, S., Lee, Y. K., Lee, J., et al. (2005) Human β-globin second intron highly enhances expression of foreign genes from murine cytomegalovirus immediate-early promoter. J Microbiol Biotechn 15: 544-550.

Kim, D. W., Uetsuki, T., Kaziro, Y., Yamaguchi, N. and Sugano, S. (1990) Use of the Human Elongation Factor-1-Alpha Promoter as a Versatile and Efficient Expression System. Gene 91: 217-223.

Kim, M., O'Callaghan, P. M., Droms, K. A. and James, D. C. (2011) A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol Bioeng 108: 2434-2446.

Kim, S. Y., Lee, J. H., Shin, H. S., Kang, H. J. and Kim, Y. S. (2002) The human elongation factor 1 alpha (EF-1α) first intron highly enhances expression of foreign genes from the murine cytomegalovirus promoter. J Biotechnol 93: 183-187.

Kling, J. (2012) Fresh from the biotech pipeline 2011. Nat Biotechnol 30: 128-131.

Kuczewski, M., Schirmer, E., Lain, B. and Zarbis-Papastoitsis, G. (2011) A single-use purification process for the production of a monoclonal antibody produced in a PER.C6 human cell line. Biotechnol J 6: 56-65.

Kwaks, T. H., Sewalt, R. G., van Blokland, R., Siersma, T. J., Kasiem, M., et al. (2005) Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. J Biotechnol 115: 35-46.

Le, H., Vishwanathan, N., Kantardjieff, A., Doo, I., Srienc, M., et al. (2013) Dynamic gene expression for metabolic engineering of mammalian cells in culture. Metab Eng 20: 212-220.

Le Hir, H., Nott, A. and Moore, M. J. (2003) How introns influence and enhance eukaryotic gene expression. Trends Biochem Sci 28: 215-220.

Lee, D., Karchin, R. and Beer, M. A. (2011) Discriminative prediction of mammalian enhancers from DNA sequence. Genome Res 21: 2167-2180.

LeProust, E. M., Peck, B. J., Spirin, K., McCuen, H. B., Moore, B., et al. (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. Nucleic Acids Res 38: 2522-2540.

Lin, H. and Li, Q. Z. (2011) Eukaryotic and prokaryotic promoter prediction using hybrid approach. Theor Biosci 130: 91-100.

Magnusson, T., Haase, R., Schleef, M., Wagner, E. and Ogris, M. (2011) Sustained, high transgene expression in liver with plasmid vectors using optimized promoter-enhancer combinations. J Gene Med 13: 382-391.

Malik, S. and Roeder, R. G. (2010) The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet 11: 761-772.

Mariati, Ng, Y. K., Chao, S. H., Yap, M. G. and 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.

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., et al. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol 30: 271-277.

Nelson, J. A., Reynolds-Kohler, C. and Smith, B. A. (1987) Negative and positive regulation by a short segment in the 5'-flanking region of the human cytomegalovirus major immediate-early gene. Mol Cell Biol 7: 4125-4129.

Niwa, H., Yamamura, K. and Miyazaki, J. (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene 108: 193-200.

Nott, A., Meislin, S. H. and Moore, M. J. (2003) A quantitative analysis of intron effects on mammalian gene expression. RNA 9: 607-617.

Ogawa, R., Kagiya, G., Kodaki, T., Fukuda, S. and Yamamoto, K. (2007) Construction of strong mammalian promoters by random cis-acting element elongation. Biotechniques 42: 628-633.

Ong, C. T. and Corces, V. G. (2011) Enhancer function: New insights into the regulation of tissue-specific gene expression. Nat Rev Genet 12: 283-293.

Osterlehner, A., Simmeth, S. and Göpfert, U. (2011) Promoter methylation and transgene copy numbers predict unstable protein production in recombinant chinese hamster ovary cell lines. Biotechnol Bioeng 108: 2670-2681.

Patwardhan, R. P., Hiatt, J. B., Witten, D. M., Kim, M. J., Smith, R. P., et al. (2012) Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol 30: 265-270.

Pedersen, A. G., Baldi, P., Chauvin, Y. and Brunak, S. (1999) The biology of eukaryotic promoter prediction - A review. Comput Chem 23: 191-207.

Pham, P., Kamen, A. and Durocher, Y. (2006) Large-Scale transfection of mammalian cells for the fast production of recombinant protein. Mol Biotechnol 34: 225-237.

Pontiller, J., Gross, S., Thaisuchat, H., Hesse, F. and Ernst, W. (2008) Identification of CHO endogenous promoter elements based on a genomic library approach. Mol Biotechnol 39: 135-139.

Pontiller, J., Maccani, A., Baumann, M., Klancnik, I. and Ernst, W. (2010) Identification of CHO Endogenous Gene Regulatory Elements. Mol Biotechnol 45: 235-240.

Prentice, H. L., Tonkin, C. J. D., Caamano, L. and Sisk, W. P. (2007) High level expression of proteins using sequences from the ferritin heavy chain gene locus. J Biotechnol 128: 50-60.

Quilici, L. S., Silva-Pereira, I., Andrade, A. C., Albuquerque, F. C., Brigido, M. M., et al. (2013) A minimal cytomegalovirus intron A variant can improve transgene expression in different mammalian cell lines. Biotechnol Lett 35: 21-27.

Reik, A., Zhou, Y., Collingwood, T. N., Warfe, L., Bartsevich, V., et al. (2007) Enhanced protein production by engineered zinc finger proteins. Biotechnol Bioeng 97: 1180-1189.

Schlabach, M. R., Hu, J. K., Li, M. and Elledge, S. J. (2010) Synthetic design of strong promoters. P Natl Acad Sci USA 107: 2538-2543.

Smale, S. T. and Kadonaga, J. T. (2003) The RNA polymerase II core promoter. Annu Rev Biochem 72: 449-479.

Song, L., Zhang, Z., Grasfeder, L. L., Boyle, A. P., Giresi, P. G., et al. (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res 21: 1757-1767.

Sumitomo, Y., Higashitsuji, H., Liu, Y., Fujita, T., Sakurai, T., et al. (2012) Identification of a novel enhancer that binds Sp1 and contributes to induction of cold-inducible RNA-binding protein (cirp) expression in mammalian cells. BMC Biotechnol 12.

Swindle, C. S., Kim, H. G. and Klug, C. A. (2004) Mutation of CpGs in the murine stem cell virus retroviral vector long terminal repeat represses silencing in embryonic stem cells. J Biol Chem 279: 34-41.

Thaisuchat, H., Baumann, M., Pontiller, J., Hesse, F. and Ernst, W. (2011) Identification of a novel temperature sensitive promoter in cho cells. BMC Biotechnol 11.

Thomsen, D. R., Stenberg, R. M., Goins, W. F. and Stinski, M. F. (1984) Promoter-regulatory region of the major immediate early gene of human cytomegalovirus. P Natl Acad Sci USA 81: 659-663.

Uetsuki, T., Naito, A., Nagata, S. and Kaziro, Y. (1989) Isolation and characterization of the human chromosomal gene for polypeptide chain elongation factor-1 alpha. J Biol Chem 264: 5791-5798.

Weber, F., de Villiers, J. and Schaffner, W. (1984) An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell 36: 983-992.

Wurm, F. M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol 22: 1393-1398.

Xia, W., Bringmann, P., McClary, J., Jones, P. P., Manzana, W., et al. (2006) High levels of protein expression using different mammalian CMV promoters in several cell lines. Protein Expres Purif 45: 115-124.

Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., et al. (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol 29: 735-741.

Yang, Y., Mariati, Chusainow, J. and Yap, M. G. (2010) DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. J Biotechnol 147: 180-185.

Table 1. Online promoter databases

EPD, Eukaryotic Promoter Database http://epd.vital-it.ch/ Managed by the Swiss Institute of Bioinformatics, EPD is an annotated and experimentally validated collection of eukaryotic promoters from human, mouse, *D. melanogaster*, and zebrafish genomes. Over 200,000 promoters can be accessed in the database.
MPromDb, Mammalian Promoter Database http://mpromdb.wistar.upenn.edu/ MPromDb is a curated database of annotated gene promoters identified from ChIP-seq experiment results. Datasets are obtained from 6 different human cell samples (CD4+TCells, HeLa S3, K562, NB4, Lymphyoblastoid, Jurkat) and mouse samples from 5 different mouse tissues and 5 different cell types.
PEDB, Mammalian Enhancer/Promoter Database http://promoter.cdb.riken.jp/ Mammalian promoter/enhancer database constructed by integrating information of conserved non-coding regions, transcriptional start sites and transcription factor biding sites to obtain a system level understanding of dynamic transcriptional regulation in mammals.

Formatted Table

Table 2. Promoters used in commercial vectors for recombinant protein expression in mammalian cells.

Promoter	Vector	Supplier	Comments
hCMV (human cytomegalovirus promoter)	pcDNA3.1, pcDNA3.3/3.4-TOPO, pOptiVEC-TOPO, pBudCE4.1	Life Technologies (www.lifetechnologies.com)	Each vector uses different variants of the hCMV promoter. Refer to supplier database for more information.
	gWiz vector series	Genlantis (www.genlantis.com)	
	phCMV vector series	Promega (www.promega.com)	
	pFLAG-CMV and pBICEP-CMV series of vectors	Sigma-Aldrich (www.sigmaaldrich.com)	
mCMV (murine cytomegalovirus promoter)	GS vector	Lonza (www.lonza.com)	Intron is included downstream of the promoter. Used in the popular GS expression system.
SV40 (simian virus 40 promoter)	pZeoSV2(+)	Life Technologies	SV40 is often used for expression of selection marker genes for stringent selection.
	pSI/neo	Promega	
	pcDNA3.1, pcDNA3.3-TOPO, pOptiVEC-TOPO, GS vector	Various	

Formatted Table

HEF1α (human elongation factor 1α gene promoter)	pBudCE4.1	Life Technologies	
	pFUSE2-CLIg-hk pFUSE-CHIg-hG1	InvivoGen (www.invivogen.com)	Vectors specifically designed for Ig expression with light chain and heavy chain constant regions on the vector.
CHOEF1α (CHO elongation factor 1α gene promoter	pDEF38	CMC biologics (www.cmcbio.com)	5’ and 3’ flanking regions and the first intron of CHO EF-1α are included.
FerH, FerL (human ferritin heavy and light chain gene promoters)	pVITRO2-MCS	Invivogen	5’UTR is modified to reduce iron regulation of promoters.
CAG (CMV enhancer/ß-actin promoter)	pCAGG	Gene Bridges (www.genebridges.com)	

Fig. 1 "Trapping" of novel promoters and enhancers. Plasmid vectors that are either lacking a promoter or carrying a weak SV40 core promoter can be used as a promoter trap or enhancer trap respectively. The trap vectors can be used in two ways: (1) Directly transfecting the trap vector into the host cell. In this method, the green fluorescence protein (GFP) reporter or selection marker (SM) would only be expressed when the vector is integrated downstream of an endogenous promoter or enhancer, (2) Shotgun cloning DNA fragments from the target host into the trap vector before transfection. The fragments integrated would need to be a promoter or enhancer for the GFP or SM to be expressed. GFP screening or drug selection is subsequently performed to isolate positive cells. Sequence of the segments upstream of the trap vector or the ligated fragment can be identified by methods like nested PCR or genomic library screens. The novel promoter or enhancer would be further evaluated and engineered before use. Pro: Promoter; Enh: Enhancer; RDS: Random DNA segments.

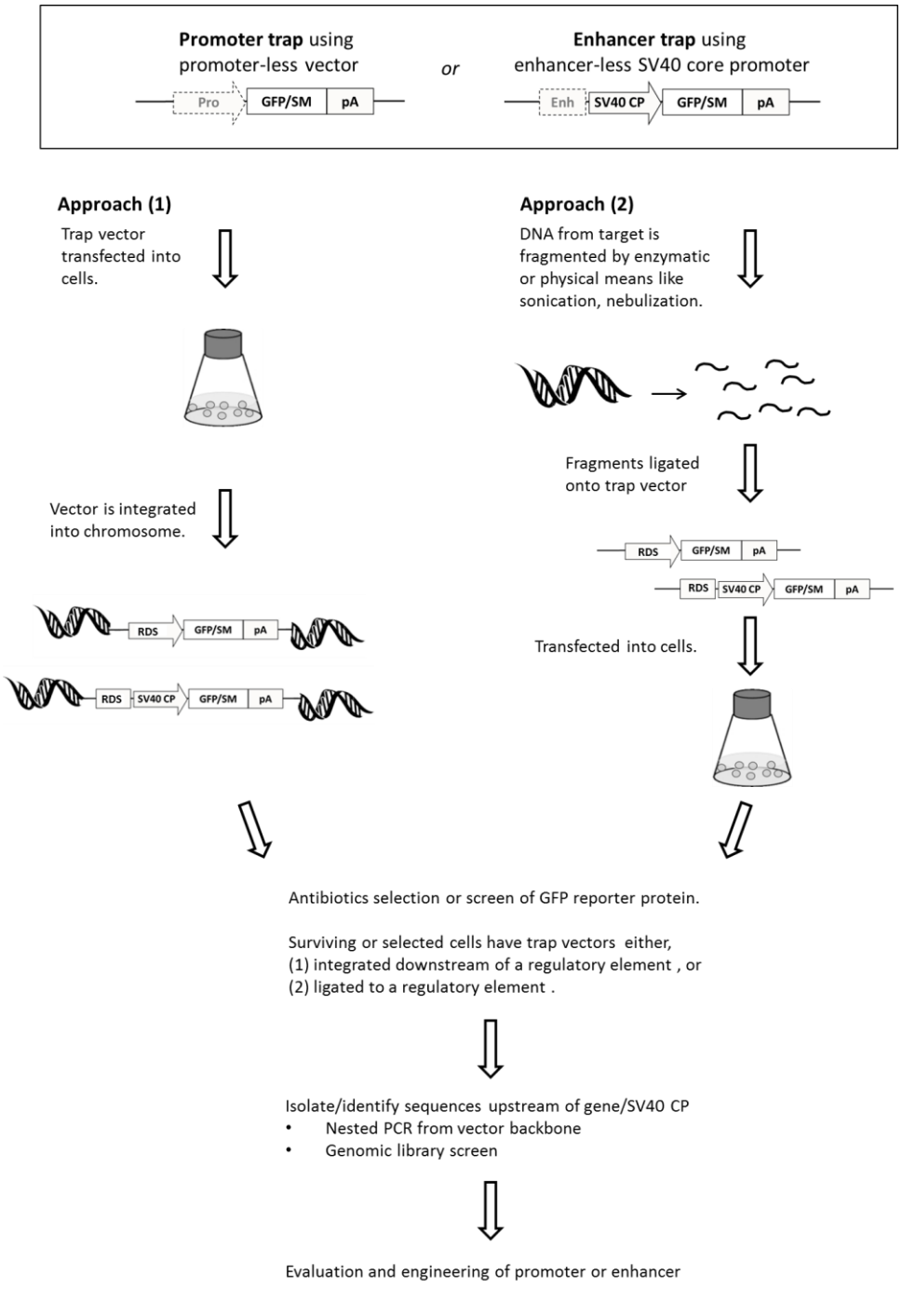

Promoter trap using promoter-less vector
Pro
GFP/SM
pA
or
Enhancer trap using enhancer-less SV40 core promoter
Enh
SV40 CP
GFP/SM
pA
Approach (1)
Trap vector transfected into cells.
Vector is integrated into chromosome.
RDS
GFP/SM
pA
RDS
SV40 CP
GFP/SM
pA
Approach (2)
DNA from target is fragmented by enzymatic or physical means like sonication, nebulization.
Fragments ligated onto trap vector
RDS
GFP/SM
pA
RDS
SV40 CP
GFP/SM
pA
Transfected into cells.
Antibiotics selection or screen of GFP reporter protein.
Surviving or selected cells have trap vectors either,
(1) integrated downstream of a regulatory element , or
(2) ligated to a regulatory element .
Isolate/identify sequences upstream of gene/SV40 CP
• Nested PCR from vector backbone
• Genomic library screen
Evaluation and engineering of promoter or enhancer

Fig. 1

Attachment to manuscript

Click here to download Attachment to manuscript: Response to reviewer comments_2014-03-12.docx

AUTHOR'S QUICK CHECKLIST FOR PREPARING MANUSCRIPTS FOR SUBMISSION

A copy of this, appropriately checked and signed, should accompany each submitted paper.

Your paper must follow the advice given in our **Instructions for Authors** and **Guidance Notes** which are available online. If you are uncertain about the relevancy of your paper for publication in this Journal, please read our **Aims and Scope** also available online. If you are unsure about anything please contact the Editor-in-Chief. Failure to comply with these instructions will delay assessment of your papers. Authors whose first language is not English, and may be inexperienced in writing for an international journal, are strongly urged to have their manuscript thoroughly checked before submission. We normally allow incorrectly presented papers to be resubmitted on one occasion only. The use of a professional paper writing agency is strongly recommended for those authors who are doubtful of their command of the English language.

Please do not tick the following boxes until you have complied with the instructions indicated.

☑ All authors have seen a copy of the paper AND have approved its submission.

☑ Authors must indicate in which Section of the Journal they would like their papers to appear.

☑ All text, **including** Abstract, tables, references, footnotes and figure legends, is double-spaced printed (**NOT $1^1/_2$ spacing)** and uses a full-page width (15 cm).

☑ All pages (including references, tables, figure legends, figures and any Supplementary Data) are numbered sequentially. (Tables precede figures.) Lines on each page are numbered either sequentially throughout the paper or on each individual page.

☑ The title should be informative and clear. (A running title is not used.) Do not use unspecified, non-standard abbreviations in the title. Do not use expressions such as "The effect of…".

☑ Key words or phrases (but not abbreviations) are given in **alphabetical order** after the Abstract. Up to 7 can be used.

☑ Abstract (140 words maximum) is clear and concise and gives quantitative values of key results. (Comparative values are of little use and should not be used.)

☑ Sub-headings in all sections are clearly indicated but are **NOT** numbered.

☑ All non-standard abbreviations should be kept to a minimum but need to be defined at first mention in each section including tables and figure legends.

☑ Organisms must be named according to the latest available lists of plants and microorganisms. If appropriate, a synonym may be used.

☑ Ensure you follow the Journal style and give correct abbreviations of standard units (for example: ml, l, g, kg, h, min, s, kDa etc. Concentrations are given as 10 g glucose l^{-1} **NOT** 10 g l^{-1} glucose. Centrifugal force (*g*) is used and not rpm.

☑ Tables and figures are to be self-explanatory with key details of procedures given in the legend or footnotes. 100% relative values **MUST** be given as the actual (absolute) values in the legend or footnote. It is recommended that procedures used only once should be detailed in the legends to figures or footnotes to tables rather than being placed in the **Methods** section.

- ☑ Tables, each on a separate sheet, are suitable either for single (7.5 cm) or double column (16 cm) printing. Titles of tables are included with the table and appear above the table itself.

- ☑ Values in tables and text **MUST NOT** exaggerate the accuracy of the methods used.

- ☑ Figures, each on a separate sheet, are suitable for printing at 7.5 cm width; exceptionally 16 cm. All symbols are clear, all lines are of adequate thickness and equal blackness, labelling is accurate and clear. Blank spaces should be avoided. Figures are numbered on the page.

- ☑ Legends to figures are grouped together on a separate sheet(s) which precedes the figures themselves.

- ☑ Position for each table and figure should be indicated in the text by a marginal note or a clear note between paragraphs in the text.

- ☑ Please ensure that photographs will print satisfactorily at the indicated column width with clear resolution of detail.

- ☑ Micrographs have a clear scale bar on the photograph with its size given in the legend.

- ☑ References agree between text and list. References in the text and list conform to the Journal style.

- ☑ Non-standard abbreviations are defined at first use but are also kept to a minimum.

- ☑ A list of mathematical notations, if used, appears before the Introduction.

July 2013