Efficient mAb production in CHO cells with optimized signal peptide, codon, and UTR

Min You^{1,2} · Yi Yang^{1,2} · Chuanqi Zhong³ · Fentian Chen^{1,2} · Xin Wang^{1,2} · Tianrong Jia^{1,2} · Yuanzhi Chen^{1,2} · Bing Zhou^{1,2} · Qingyu Mi^{1,2} · Qinjian Zhao^{1,2} · Zhiqiang An^{2,4} · Wenxin Luo^{1,2} · Ningshao Xia^{1,2}

Received: 29 December 2017 / Revised: 2 April 2018 / Accepted: 5 April 2018 / Published online: 8 May 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Antibody drugs have been used to treat a number of diseases successfully. Producing antibodies with high yield and quality is necessary for clinical applications of antibodies. For a candidate molecule, optimization of a vector to produce sufficient yield and an accurate primary structure is indispensable in the early stage of the production process development. It is especially important to maintain the fidelity of N-terminal sequence. In order to produce antibodies with a high yield and accurate N-terminal, the expression vector was systematically optimized in this study. First, the heavy chain and light chain were co-expressed in Chinese hamster ovary (CHO) cells with different signal peptides. Mass spectrometry (MS) revealed that signal peptides Esp-K, Bsp-H, and 8Hsp-H were accurately deleted from mature antibodies. Further, the yield was doubled by codon optimization and increased by 50% with the presence of untranslated regions (UTR). The combination of UTR with optimal codon and signal peptide to form an expression vector resulted in yield improvement of 150% and correct N-terminal sequences. Moreover, the main product peak was above 98% as assessed by size-exclusion chromatography (SEC). Additionally, the bioactivity of products made from optimized transient gene expression (TGE) was almost identical to the standard sample. The production efficiency and product quality from the identified TGE optimization strategy was further demonstrated through application to two other antibodies. The expression level of SGE (stable gene expression) can also be improved effectively with this optimization strategy. In conclusion, vector optimization via combination of optimized signal peptide, codon, and UTR is an alternative approach for efficient antibody production with high fidelity N-terminal sequence in CHO cells.

Keywords Antibody · Expression · N-terminal sequence · Signal peptide · Codon · Untranslated regions

Introduction

The unique combination of high affinity, remarkable specificity, and long half-life make monoclonal antibodies (MAbs) successful in the treatment of diseases including bacterial infections, autoimmune disorders, and cancer (Elgundi et al. 2016; Strohl 2017). The majority of these antibodies are produced by recombinant DNA technology in mammalian cells (Birch and Racher 2006; Kunert and Reinhart 2016). The process of therapeutic drug production must meet productivity, quality, and safety criteria (Gaughan 2016; Li et al. 2010).

MAB production in mammalian cells can be performed through either transient gene expression (TGE) or stable gene expression (SGE). Transient expression allows quick production of small amounts of antibodies for testing during the early stages of drug discovery (Ho et al. 2013; Zhang et al. 2017). Stable expression process is more widely used in large-scale industrial production (Ho et al. 2013). Currently, the highest

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00253-018-8986-5>) contains supplementary material, which is available to authorized users.

✉ Wenxin Luo
wxluo@xmu.edu.cn

¹ National Institute of Diagnostics and Vaccine Development in Infectious Diseases, Xiamen University, Xiamen 361102, China

² State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, Xiamen University, Xiamen 361102, China

³ State Key Laboratory of Cellular Stress Biology, Xiamen University, Xiamen 361102, China

⁴ Texas Therapeutics Institute, The Brown Foundation of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA

yield for transient expression is around 1 g/L, but 12 g/L can be obtained by stable expression (Frenzel et al. 2013; Kim et al. 2012).

The productivity of antibody expression involves multiple factors. For transient expression, the cell culture media, gene delivery reagents, host cell line, and expression vector are the decisive factors for yield (You et al. 2013). For high-yield expression in a stable and high-producing cell line, the most important factors are appropriate host cell line, an efficient vector, and a good selection protocol (Jostock and Knopf 2012).

Host cell line and expression vector are important for both transient expression and stable expression. The CHO cell line, due to its good performance in protein folding, secretion, and post-translational modification, is the most commonly used host cell line for both transient and stable expression. To achieve high levels of gene expression, vectors such as pTT5 usually have strong promoters (e.g., the CMV promoter) to drive expression of the antibody genes (Durocher et al. 2002). The pTT5 vector has also included favorable RNA processing signals such as polyA tail, presence of an intron to encourage export from the nucleus, and a splice site to remove this intron (Birch and Racher 2006). A high mRNA level, obtained by strong promoters and polyA tail, is important for high-yield expression. In addition to strong promoters and a polyA tail on the plasmid, 3' untranslated regions (UTR) are included at the 3' end of the coding sequence to stabilize the mRNA (Kuersten and Goodwin 2003).

In translation, mRNA produced by transcription is decoded by a ribosome complex to produce a specific amino acid chain, or polypeptide. Optimization of codon usage for CHO (Chung et al. 2013) and using 5' UTR (Araujo et al. 2012) effectively improved yield by speeding up translation processing. The choice of synonymous codons in many species is strongly biased. There is a correlation between high expression and the use of selective codons in a given organism (Brandis and Hughes 2016). Efficient expression of the codon-optimized gene can be attributed to two factors: first, the abundance of isoacceptor tRNAs and modified nucleotides at the anticodon wobble position available in a host; second, the formation of a secondary structure of the transcripts favorable for translation (Kanaya et al. 2001; Nakamura et al. 2000). In eukaryotic cells, translation of cellular mRNA is generally mediated by a "cap" at the 5' end of mRNA upstream of the coding sequence in UTR which is responsible for interaction of the messenger RNA with the ribosome (Lopez-Lastra et al. 2005).

In addition to mRNA levels and translation, the translocation of secretory proteins into the lumen of the endoplasmic reticulum (ER) is a limiting step. It has been demonstrated by many labs that alternative signal peptides can lead to increased protein secretion (Knappskog et al. 2007; Kober et al. 2013; Zhang et al. 2005).

Besides productivity, high quality is required for safety. First and foremost, the primary structure of protein must be accurately maintained. Excess amino acid residues in the antibody may be harmful for the patient. Appropriate signal sequences are used to target the heavy and light chain polypeptides to the correct part of the secretory pathway. Further, use of these appropriate sequences results in the correct cleavage to maintain an accurate N-terminal (Gibson et al. 2017; Nielsen et al. 1997a b). Optimization of signal peptides is also essential in establishing the manufacturing process for antibody production (Gibson et al. 2017; Kober et al. 2013).

The humanized antibody, B45 hAb against hepatitis B virus (HBV), showed therapeutic efficacy in our previous experiments in animal models (Luo et al. 2016). In order to construct a vector for efficient production of B45 hAb, the coding sequence, including appropriate codon and signal peptide, were optimized separately and then combined. Subsequently, a UTR was added to the ends of the coding sequence to form an optimal vector. The optimized construct architecture for vector resulted in production enhancement and accurate N-terminal sequence caused by correct cleavage. In order to validate the applicability of this strategy, the strategy was applied to optimize expression of two other antibodies, 37 hAb against hemagglutinin of H5N1 virus (Luo et al. 2008) and Herceptin against the breast cancer receptor HER2. Similar results were obtained.

Materials and methods

Vector construction and plasmid DNA preparation

The primary DNA fragments of variable regions of B45 and H37 were isolated using a phage display library (Luo et al. 2008, 2016). The cDNA of Herceptin from NCBI was made by Genewiz. The cDNA of the human IgG 1 constant regions were obtained from a human cDNA library. All sequences of tested antibodies are listed in Table 1. All DNA fragments post codon optimization were synthesized by Genewiz. The variable regions were fused to the constant regions of human IgG1 HC or kappa LC by overlap extension PCR. Each antibody HC and LC was fused with different signal peptides and inserted into pTT5 at *EcoR* I/*Bam*H I. The signal peptides used are shown in Table 1. Plasmid DNA was amplified in *DH5α* (TianGen) and purified according to a protocol described in a previous study (You et al. 2013).

Signal peptide optimization

To analyze the secretion efficiency and cleavage, signal peptides listed in Table 2 were fused at the 5' primer of coding sequence. The plasmid with chimeric signal peptide was transfected into CHO cells. Culture media was sampled each

Table 1 Overview of all tested sequences of antibodies

Sequence	Antibody	Description	GenBank accession number
45VK	B45, anti-HBV antibody	The wild-type variable region of kappa chain	MG815636
45VKCO		The codon-optimized variable kappa chain	MG815637
45VH		The wild-type variable region of heavy chain	MG815638
45VHCO		The codon-optimized variable heavy chain	MG815639
37VK	37, anti-H5N1 antibody	The wild-type variable region of kappa chain	MG815640
37VKCO		The codon-optimized variable kappa chain	MG815641
37VH		The wild-type variable region of heavy chain	MG815642
37VHCO		The codon-optimized variable heavy chain	MG815643
Herceptin-VK	Herceptin, anti-Her2 antibody	The wild-type variable region of kappa chain	GM685463.1
Herceptin-VKCO		The codon-optimized variable kappa chain	MG815644
Herceptin-VH		The wild-type variable region of heavy chain	GM685465.1
Herceptin-VHCO		The codon-optimized variable heavy chain	MG815645
CH	Human IgG1 constant regions	The wild-type constant region of heavy chain	MG815646
CHCO		The codon-optimized constant heavy chain	MG815647
CK		The wild-type constant region of kappa chain	MG815648
CKCO		The codon-optimized constant kappa chain	MG815649

day for protein quantification. Three replicates of each tests were performed to reduce systematic errors.

Antibody splice sites were predicted by the Signal P 4.1 Server, a web server that predicts the location of signal peptide cleavage sites in amino acid sequences from different organisms (<http://www.cbs.dtu.dk/services/SignalP/>) (Petersen et al. 2011).

Cell culture and transfection

CHO-S cells were obtained from Gibco and maintained in 125-mL shake flasks (Corning) with SFM 4 (HyClone), which were supplemented with 4 mM L-glutamine, 0.1 μ M HT (hypoxanthine and thymidine, Sigma), and 1 g/L PF68 (Sigma). Cultures were shaken at 110 rpm on an orbital shaker with a rotational diameter of 2.5 cm (Kuhner AG, Birsfelden, Switzerland) in the presence of 37 °C, 5% CO₂, and 75% humidity. 15 mL cells (2×10^6 cells/mL) were transfected with 75 μ g plasmid DNA as described previously (You et al. 2013).

Reduced capillary electrophoresis-SDS assay (CE-SDS)

The purity of antibodies was determined by PA800 plus pharmaceutical analysis systems with commercial CE-SDS kit (Cat# 390953). Briefly, A 50 μ M I.D. bare fused-silica capillary with 30.2 cm total length (20.0 cm effective length) was used in this assay. The signals were collected by a PDA detection module. For the preparation of samples, SDS buffer was used to dilute the samples to nearly 1 mg/mL, then added internal standard and β -mercaptoethanol to the sample solution and water bath at 100 °C for 3 min. The default method was used to run the samples in PA800 plus.

MS determination of the cleavage of the signal peptide in antibodies

To determine cleavage sites, antibodies with different signal peptides were purified by protein A. Antibodies suspended in a loading buffer (50 mM Tris pH 6.8, 2% SDS, 5% 2-mercaptoethanol, 0.01% bromophenol blue, 8% glycerol)

Table 2 Overview of all tested signal peptides

Name	Sequence	Protein	Organism	GenBank accession number
B	MKWVTFISLLFLFASAYS	Serum albumin preproprotein	<i>Homo sapiens</i>	NP_000468
E	MTRLTVLALLAGLLASSRA	Azurocidin preproprotein	<i>Homo sapiens</i>	NP_001691
8Hsp	MGRLTSSFLLLIVPAYVLS	8C11 Heavy chain preproprotein	<i>Mus musculus</i>	MG815651
8Ksp	MSVPTQVLGLLLLWLTDARC	8C11 kappa chain preproprotein	<i>Mus musculus</i>	MG815650

were electrophoresed in a 12% SDS-PAGE gel. The light chain and heavy chain were collected separately from the gel.

The bands corresponding to target proteins were excised and subjected to in-gel digestion. The gel pieces were first washed with 50 mM ammonium bicarbonate (pH 8.0)/50% ACN and dried. The dried gel pieces were incubated with reduction buffer containing 10 mM DTT in 50 mM ammonium bicarbonate (pH 8.0) at 56 °C for 30 min, followed by alkylation with the buffer containing 50 mM iodoacetamide at room temperature in the dark for 30 min. Next, the gel pieces were washed with 100% ACN and dried. Trypsin in 50 mM ammonium bicarbonate (pH 8.0) was added into gel pieces at the ratio of protein/trypsin (100:1). Gel pieces were then incubated at 4 °C for 1 h followed by 37 °C for 12–18 h. The tryptic peptides were extracted from the gel pieces using the buffer containing 66.7% ACN/1.6% formic acid in H₂O and desalted with C18 STAGEtips. The tryptic peptides were analyzed on mass spectrometry AB Sciex TripleTOF 5600. The wiff files were searched for target protein sequences associated with common contaminant proteins using the ProteinPilot (V4.5 beta) software with the search mode “Rapid.”

Codon optimization

The variable regions of antibodies and the constant regions of human IgG1 were optimized with the Codon OptimWiz (Genewiz, China) expert software. The candidate sequences are assessed taking into account codon usage in *Cricetulus griseus*, the secondary structure of mRNA, GC content, species-specific sequence motifs, removal of restriction enzyme cutting sites and cleavage sites, and repeat sequences. The DNA fragments with optimized codon sequences were obtained by synthesis.

Vector construction with UTR

5'UTR(GCCGCCACC) and 3'UTR(AACGGGCTGATGCTGCACCAACTGTATCCA, GeneBank:M35669) were introduced to the 5' and 3' ends of antibody coding sequences by PCR. These fragments were cloned to pTT5 by *EcoR* I and *Xba* I.

Cells counting and antibody yield evaluation

The culture were mixed with trypan blue and assayed by an automatic analyzer (Countstar, China) for cell density and viability determination. The antibody yield of the transfectants was determined by Cedex Bio (Roche, Germany) with human IgG1 HB module.

The Cedex Bio uses absorption photometry for determining the amount of absorbance in a fluid. The absorbance

which is linear correlation to the antibody concentration is used to calculate the antibody titer in the solution. Before assay, calibrations and QC were performed, resulting in a standard curve for antibody yield calculating. In brief, the culture supernatant was put into the sample area manually and sampled and mixed with the reaction reagent by the equipment automatically. About 8 min later, the absorbance value was read and the concentration was calculated.

Size-exclusion chromatography-high-performance liquid chromatography (SEC-HPLC)

The product from the optimized TGE was purified by protein A and analyzed using a TSK-Gel column G3000 (Tosoh Biosep) on a Waters HPLC (Waters). The mobile phase is a 100-mM phosphate buffer at pH 7.0, and the regenerative solution is 150 mM NaCl. The purified protein at 100 µg/mL was loaded at a flow rate of 0.5 mL/min. The detection wavelength is 280 nm.

Binding activity assay

The purified antibody made from optimized TGE and standard samples were diluted in a gradient using a workstation (Beckman FX, USA) beginning at 1000 ng/mL. The diluted samples were added to the 96-w plates coated with antigens and investigated the binding activity by chemical luminescence assay (CLA). Experiments in all wells were repeated at three times. The CLA values were submitted to Prism 7 for EC50 calculation.

Stable cell pool development and yield evaluation

The genes of the antibody were inserted into the dual CMV promoter expression vector with gs marker pGS2 (Fig. S-1), and the plasmid was linearized and transfected into suspended CHO-k1 cell. The cells were plated into 96-w plates at 5000 c/w with 25µM MSX at 48 h post transfection. After 15 days of cultivation, the 30 mini pools with the highest yield of each transfection were selected to be expanded by HTRF (Cisbio).

For the top 30 mini pools, batch or fed-batch culture was carried out with initial seeding density at 5E5c/mL. For the batch culture, the cells were inoculated in CD CHO (Gibco) + HT medium and cultured for 7 days at 37 °C. For the fed-batch culture, cells were seeded in CD Forti (Gibco) and fed with 10% Feed C medium (Gibco) at day 3, day 6, day 8, day 10, and day 12. The concentration of glucose was maintained above 5 g/L, and the cultivation temperature shifted to 32 °C at day 6. The whole fed-batch process lasts 14 days.

Results

Selection of appropriate signal peptide for antibody expression

Signal peptides play an important role in the secretion of recombinant proteins expressed in mammalian cells (Trosse et al. 2007; Young and Rance 2008). Many studies have shown that optimization of signal peptides could improve the production of recombinant proteins (Haryadi et al. 2015; Kober et al. 2013). Based on these observations, four promising signal peptides (listed in Table 2) previously reported to increase the antibody production (Kober et al. 2013; Luo et al. 2007), were tested in this study.

Suspension culture CHO-S cells were co-transfected with pTT5 expression plasmid containing genes of the heavy chain and the light chain of B45 hAb fused with the signal peptides in Table 1. Samples were taken from cultures every day for productivity testing and cell counting. Maximum productivity for each condition is shown in Fig. 1a. We found that co-expression of 8Hsp-H and 8Ksp-K led to the lowest expression, 30 mg/L. A comparable yield, 40 mg/L was obtained by co-expression of B-H mix with B-K (B-H + B-K) or E-H with E-K (E-H + E-K). However, there were no significant differences in the cell viability and density (Fig. 1b, c).

The cleavage of signal peptide accompanied by the protein secretion is even more important for industrial production than productivity. To predict the cleavage of the signal peptide, the DNA sequences of the kappa chains or heavy chains with signal peptide at N-terminal were submitted to the Signal P 4.1 Server. As seen in Fig. S-2, all signal peptides were

predicted to result in a correct cleavage, and all signal peptides could be cut off completely.

In order to verify the cleavage site, the products were purified using protein A and analyzed by reduced CE-SDS separation. For all products (shown in the left column of the Fig. 2), the area of each K-chain was about 34%, and the H-chain was about 65%. As a result, the K/H for each product was 1:1 (mol/mol). The retention time of the signal peak was almost coincident. Subsequently, the light chain and heavy chain were assayed by MS. As shown in the right column of Fig. 2, the K-chain with E signal peptide (Fig. 2a), H-chain with B signal peptide (Fig. 2b), and H-chain with 8Hsp (Fig. 2c) resulted in correct N terminals. For other fusions, there were unwanted amino acid residues from the signal peptide on the N-terminal. For safety, additional amino acid residues are undesirable. Therefore, K-chain with E signal peptide (E-K), H-chain with B signal peptide (B-H), and H-chain with 8Hsp (8Hsp-H) were selected as candidates.

To explore antibody productivity with optimal signal peptide, the E-K were mixed with B-H or 8Hsp-H and were co-expressed in CHO cells separately. The 8Hsp-H and E-K combination led to a higher titer, 45 mg/L, than the combination of B-H and E-K ($P = 0.005$, Fig. 3a). However, there was no significant difference in the viable cell density (Fig. 3b) and viability (Fig. 3c).

To investigate the splicing site of the signal peptide, the products were analyzed by MS. The peptide spectrum from MS confirmed that all chains possessed the correct N-terminal peptide without any additional or missed amino acids (data not shown). Taking into account the productivity and signal peptide cleavage, the combination of K-chain fused with E

Fig. 1 The yield of antibody with different signal peptides. Different vector combinations containing light and heavy chains with different signal peptides were transfected into CHO cells. The titers were assayed by Cedex (a), and the cultures were taken every day for the counting of cell density (b) and viability (c) by Countstar. The data analysis was performed with SPSS 20.0. All reported P values are two-sided with an

alpha value of 0.05. B-H, B signal peptide as the leader sequence of H-chain; B-K, B signal peptide as the leader sequence of K-chain; E-H, E signal peptide fused to H-chain; E-K, E signal peptide fused to K-chain; 8Hsp-H, 8Hsp as the leader sequence of H-chain; 8Ksp-K, 8Ksp as the leader sequence of K-chain

Fig. 2 Determination of the antibody N-terminal peptides by MS. The reduced CE-SDS and MS assays for antibodies with signal peptides B (a), E (b) and 8Sp (c). The antibodies, with different signal peptides, were analyzed by reduced CE-SDS (left). The N-terminal peptide was detected

by MS (right) for K-chain and H-chain separately. The amino acids underlined in red are the remaining of the signal peptide. The amino acid residues in gray are the residues which are undetectable by MS

and H-chain fused with 8Hsp were selected for antibody expression.

Enhance the expression with codon optimization and UTR

The constant region of human IgG 1 was from the human cDNA library, and the codons are biased to human. The variable regions of B45 hAb were from phage library (Luo et al. 2016), and there was no obvious codon bias in the library. A

web server, Codon OptimWiz (Genewiz, China), was used to optimize the codon of the variable region and constant region to bias to CHO cell. All sequences are listed in Fig. S-3. The information of sequence-optimized genes is summarized in Table S-1. Compared to the sequence before codon optimization, the codon adaptation index (CAI)s were reached more than 0.95, and the GC content was adjusted to about 60%.

To investigate the productivity enhancement after codon optimization, different vector combinations containing light and heavy chains with default signal peptide (8Ksp and

Fig. 3 The TGE yield of antibody with optimal signal peptide. The plasmids with the candidate signal peptide were co-transfected into CHO cells. The antibody titer (a), the viable cell density (b), and viability (c) were detected

8Hsp) and optimal codon were co-transfected into CHO cells. As expected, the yield was increased by 100% with codon optimization, and the 45KOC made a greater contribution to the yield improvement because co-transfected 45KOC with H led to an equivalent yield to co-transfected 45KOC and 45HOC (Fig. 4a). There were no significant differences in cell density and viability for all tests (data not shown).

Kozak sequence, a typical 5' UTR, is reported to play a role in efficient initiation of the translation process. The 3' UTR of murine Ig kappa chain v-j2-c (GeneBank: M35669, 30 bp) was used to stabilize mRNA in numerous studies (Li and Chen 2011). In order to improve the productivity further, a 5' UTR was added to the 5' end of the sequence with optimal signal peptide and codon, and a 3' UTR was added to the 3' end. A 50% increase was observed from transient expression of the vector with UTR (Fig. 4b), and a correct N-terminal sequence was detected from the product examined by MS (data not shown). So far, an optimal vector with an appropriate signal peptide, codon, and UTR has been constructed and

could express antibody with the correct N-terminal sequence and high yield.

The SEC-HPLC and binding activity assay for products from TGE

The TGE culture with vector containing optimized signal peptide, codon, and UTR was purified using protein A and subjected to SEC-HPLC assay. The result is shown in Fig. 5a. There was only one sharp peak with retention time at 15.8 min. After analysis by Empower 3, it was found that the percentage of HMW was 1.2%, the LMW was 0%, and the main peak was 98.8%. This result indicates that the antibody expressed with the optimized TGE method was of a good homogeneity.

To explore the biological activity of the product with optimized TGE, the purified product and standard reference material were tested by CLA for binding activity with HBsAg. The reaction curves of the product with optimized TGE and

Fig. 4 Improved TGE of antibody by codon optimization and UTR. **a** The expression vectors of light and heavy chains with optimal codon were co-transfected into CHO cells. The cultures were taken everyday for the determination of antibody titers by Cedex, and the peak yields for each culture were shown. **b** The plasmids with optimal codon, appropriate signal peptide, and UTR were co-transfected into CHO cells. The highest productivities for each test are shown

Fig. 5 The SEC-HPLC and binding activity for the product. **a** The SEC-HPLC assay for the purified product with optimized TGE. **b** Binding activity of product with optimized TGE

standard material are coincident (Fig. 5b). Subsequently, CLA data were entered into Prism for calculation of EC₅₀. EC₅₀ of standard material was 0.002108, and EC₅₀ of the product with optimized TGE was 0.001981. These results demonstrate that the optimized TGE method does not affect the biological activity of the product.

Optimization of expression for other antibodies

In order to investigate how application of the TGE optimized strategy could be applied generally, the TGE method was applied for expression of two other antibodies: Herceptin, a humanized antibody against Her2, and 37 hAb, a humanized antibody against H5N1.

The VK and VH sequences of Herceptin were taken from NCBI (Hoerr et al. 2016). The cDNAs of 37 hAb were taken from a phage display library (Luo et al. 2008). The signal peptides in Table 2 were fused to the N-terminal of the variable region for TGE. We found that the highest TGE was achieved by B signal peptide for Herceptin and antibody 37 hAb. The highest yield of Herceptin was 83 mg/L ($P < 0.05$) and the highest yield of 37 hAb was 64 mg/L ($P < 0.05$) (Fig. 6a, b). The products were purified and assayed by MS. Fortunately, we found all signal peptides had the correct N terminals (Fig. 6c, d). With the highest yield, B signal peptide was chosen for further study.

Next, the codons of the variable regions of 37 hAb and Herceptin were optimized by Codon OptimWiz. Codon optimization is shown in Fig. S-4. The parameters of optimized codons are summarized in Table S-2. After codon optimization, the codon adaptation index (CAI) increased to more than 0.96, and the GC content was adjusted to about 61%.

Finally, a UTR sequence was added to the 5' and 3' of the antibody ORF, and the cDNAs of 37 hAb and Herceptin with

UTR were separately inserted into Ptt5 at multiple cloning sites to generate expression vectors. The plasmids of light chain and heavy chain were co-transfected into CHO cells. The yield of TGE for Herceptin post optimization was 2.2-fold higher than that of control, and the yield for 37 hAb was 2.3-fold higher (Fig. 7). The yield of Herceptin was increased from 51 to 109.9 mg/L, while the productivity of 37 hAb was raised from 30 to 69 mg/L after optimization (Fig. 7). There was no significant change in the binding activity of the products (data not shown). These results indicate that TGE with optimized signal peptide, codon, and UTR can improve the expression of antibody, and this strategy could be replicated in expression of other antibodies.

Optimization the expression of stable cell pool

Transient expression may be affected by a variety of factors, and the yield may be varied from lot to lot. To verify if this optimization strategy is applicable in stable gene expression, the genes of the antibody were cloned into the GS expression vector pGS2 (Fig. S-1) to generate stable expression vector. The CHO-K1 were transfected with the stable expression vectors and screened for stable mini pools with MSX. The top 30 mini pool in titer of each transfection were subjected to batch culture. It is exhibited in Fig. 8a that a higher frequency and higher yields were gained post optimization with codon and UTR. The maximum yield of control group (8 Hsp-H + 8 Ksp-K) is 73 mg/L, and the maximum yield of the group with signal peptide optimization (8 Hsp-H + E-K) is 104 mg/L. The highest yield of the mini pools with codon optimization (8 Hsp-HOC + E-KOC) is achieved to 203 mg/L. Furtherly, a yield of 277 mg/L was reached after adding UTR. Figure 8b shows the expression level of top 10 mini pools with fed-batch culture, the mean titer of the control group (8 Hsp-H + E-K) is

Fig. 6 Screening the best signal peptides for Herceptin and 37 hAb. **a** The yield of TGE for Herceptin with different signal peptide. **b** The yield of TGE for 37 hAb with different signal peptide. **c** The N-terminal of Herceptin with different signal peptide by MS. **d** The N-terminal of 37 hAb with different signal peptide by MS

557 mg/L, and the average yield of the group with signal peptide, codon, and UTR optimization (UTR + 8Hsp-HOC + E-KOC) is up to 1770 mg/L. The mean values in Fig. 8b is consistent with the values in Fig. 8a. It proves that the optimization of signal peptide, codon, and UTR can also effectively improve the level of stable expression.

Discussion

The clinical and commercial success of therapeutic antibodies has led to the need for large-scale production in mammalian cell culture and rapid expansion of global manufacturing capacity (Birch and Racher 2006). In addition to high productivity, high quality is indispensable for therapeutic drugs. In the development of a production process, every parameter is optimized to achieve high productivity and accurately maintain the antibody molecule. In the early stage of production process development, TGE always is used to predict the yield and quality in SGE. In terms of quality, an accurate N-terminal sequence is indispensable. Vector design is essential for success in both TGE and SGE. In this study, we optimized a TGE strategy with an appropriate signal peptide, codon, and UTR

to increase expression by 150% and obtain an antibody with an accurate N-terminal sequence. And we verified that this optimization method is also effective for stable expression.

An expression cassette is made up of, at least, strong promoters, an open reading frame (ORF) and a polyA tail.

Fig. 7 Improved expression of Herceptin and 37 hAb by TGE optimization. Herceptin and 37 hAb were expressed by optimized TGE with optimized signal peptide, codon, and UTR. The titers were measured

Fig. 8 Optimization of stable gene expression. **a** The genes of antibody were inserted into plasmid pGS2, resulting in stable expression vectors. The vectors were transfected into suspension CHO-K1 and the stable cell pools were screened with 25 μ M MSX. The top 30 mini pools were subjected to batch culture, and the expression levels were determined. **b** The top 10 mini pools of each transfection were proceeded in fed-batch culture. At the end of culture, the yields were tested by Cedex

Sometimes an additional element, such as an enhancer or intron, can improve the yield. The UTR at 3' and 5' terminals of the ORF can enhance translation and stabilize mRNA (Araujo et al. 2012; Kuersten and Goodwin 2003; Lopez-Lastra et al. 2005; Mazumder et al. 2003). The most common 5' UTR is the Kozak sequence. Many studies have shown that the expression levels were elevated through translation enhanced by the Kozak sequence (Lopez-Lastra et al. 2005). The 3' UTR has been used to stabilize mRNA (Erson-Bensan 2016). In this work, we enhance the yield by 50% through UTR addition. It is possible to extend the half-life of mRNA, resulting in a higher mRNA level, as assessed by northern blot (Fig. S-5).

Normally, the ORF region of secreted proteins includes a signal peptide for secretion. The signal peptide is cut in the secretion stage. On one hand, the signal peptide determines the efficiency of protein secretion and the yield. We have enhanced the antibody expression using signal peptides, which is in agreement with previous reports (Knappskog et al. 2007). In Lars Kober's study (Kober et al. 2013), vector with B signal peptide or E signal peptide for both K-chain and H-chain led to a higher yield in CHO-K1. In this work, we found the yield from the B signal peptide fused both chain led the highest yield.

On the other hand, the splice site of signal peptide is important for the accuracy of the molecules, because unnecessary amino acid residues from signal peptide are undesirable, especially for therapeutic antibodies. Splice site is a necessary consideration for signal peptide selection. Different signal peptides may result in different splice sites (Kotia and Raghani 2010). Currently, there are many software programs that can predict the signal peptide cleavage site. But the use of software predictions has not been approved by the FDA; the FDA has only approved the use of the sequences from N-terminal sequenced by MS or other sequencing methods. Predictions may not match the sequencing results, because

cleavage sites can alternate (Kotia and Raghani 2010). In this work, all signal peptides tested could lead to splice sites corresponding to software predictions. However, as evaluated by MS sequencing, only the co-expression of E signal fused B45-K-chain and 8Hsp fused B45-H-chain led to the correct splice sites. In order to meet FDA requirements, co-expression of E-K and 8Hsp-H was selected to produce B45 hAb, although this combination is not the best in terms of yield. For the other two antibodies, all signal peptides are accurately removed in the products.

Previous studies show that codon bias in expression of exogenous genes is a common phenomenon. In non-mammalian expression systems (Burgess-Brown et al. 2008; Han et al. 2010) or mammalian expression systems (Fath et al. 2011), codon optimization is an effective method to improve yield. But the range of improvement is not guaranteed; some special features of proteins could result in lower expression (Fath et al. 2011). In this study, the productivity was enhanced by 100% through codon optimization. Increased expression resulting from codon optimization has been attributed to translational enhancement (Kudla et al. 2006; Robinson et al. 2008). Many publications suggest that codon optimization predominantly raises mRNA levels (Anson and Dunning 2005; Kudla et al. 2006, 2009). Other publications show that the GC content affects the stability of the mRNA directly and affects the mRNA level indirectly (Graf et al. 2004). In this study, increased mRNA level was detected from the culture transfected with the vector after codon optimization with increased GC content (Fig. S-4). We suppose that, together with a higher mRNA level from increased GC content, enhancement of transcription resulted in increased protein expression.

Product quality results, SEC-HPLC, and the binding activity for antibody product generated from optimized TGE were

also comparable to the control product. We applied this strategy to optimize the TGE of two other antibodies and obtained similar results, indicating replicability of this method.

In this study, we have demonstrated that optimization with an appropriate codon, signal peptide, and UTR can enhance expression yield and produce antibodies with accurate N-terminal sequence in TGE experiments. While codon optimization, signal peptide optimization, and using UTR has been used extensively in enhancing expression for stable production and TGE individually, this is the first example of vector engineering combining appropriate codon, signal peptide, and UTR to increase production of a therapeutic protein with accurate molecular by TGE. Such a production process development begins at stable cell line construction, which contains vector design, transfection, antibiotic pressure, cloning, and scale-up culture. Vector engineering is one of many possible strategies that affect the physiological state of production with significant benefit to yields for TGE, SGE, and industrial production. As the demand for faster development and expression of therapeutic protein grows in the biotechnology industry, the need for TGE is certain to increase in importance (Majors et al. 2008a, b). Adapting vector engineering to the available TGE technologies is likely to help accelerate the development of production process in the coming years.

Acknowledgements We thank Dr. Georgina Salazar at the University of Texas Health Science Center at Houston for her careful and critical reading of the manuscript.

Funding This study was supported by the NSFC(31670927) and Natural Science Foundation of Fujian Province (2017J01066).

Compliance with ethical standards

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Anson DS, Dunning KR (2005) Codon-optimized reading frames facilitate high-level expression of the HIV-1 minor proteins. *Mol Biotechnol* 31(1):85–88. <https://doi.org/10.1385/MB:31:1:085>
- Araujo PR, Yoon K, Ko D, Smith AD, Qiao M, Suresh U, Burns SC, Penalva LO (2012) Before it gets started: regulating translation at the 5' UTR. *Comp Funct Genomics* 2012:475731. <https://doi.org/10.1155/2012/475731>
- Birch JR, Racher AJ (2006) Antibody production. *Adv Drug Deliv Rev* 58(5–6):671–685. <https://doi.org/10.1016/j.addr.2005.12.006>
- Brandis G, Hughes D (2016) The selective advantage of synonymous codon usage bias in *Salmonella*. *PLoS Genet* 12(3):e1005926. <https://doi.org/10.1371/journal.pgen.1005926>
- Burgess-Brown NA, Sharma S, Sobott F, Loenarz C, Oppermann U, Gileadi O (2008) Codon optimization can improve expression of human genes in *Escherichia coli*: a multi-gene study. *Protein Expr Purif* 59(1):94–102. <https://doi.org/10.1016/j.pep.2008.01.008>
- Chung BK, Yusufi FN, Mariati, Yang Y, Lee DY (2013) Enhanced expression of codon optimized interferon gamma in CHO cells. *J Biotechnol* 167(3):326–333. <https://doi.org/10.1016/j.jbiotec.2013.07.011>
- Durocher Y, Perret S, Kamen A (2002) High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res* 30(2):E9
- Elgundi Z, Reslan M, Cruz E, Sifniotis V, Kayser V (2016) The state-of-play and future of antibody therapeutics. *Adv Drug Deliv Rev* 122: 2–19. <https://doi.org/10.1016/j.addr.2016.11.004>
- Ersen-Bensan AE (2016) Alternative polyadenylation and RNA-binding proteins. *J Mol Endocrinol* 57(2):F29–F34. <https://doi.org/10.1530/JME-16-0070>
- Fath S, Bauer AP, Liss M, Spriestersbach A, Maertens B, Hahn P, Ludwig C, Schafer F, Graf M, Wagner R (2011) Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. *PLoS One* 6(3):e17596. <https://doi.org/10.1371/journal.pone.0017596>
- Frenzel A, Hust M, Schirrmann T (2013) Expression of recombinant antibodies. *Front Immunol* 4:217. <https://doi.org/10.3389/fimmu.2013.00217>
- Gaughan CL (2016) The present state of the art in expression, production and characterization of monoclonal antibodies. *Mol Duvers* 20(1): 255–270. <https://doi.org/10.1007/s11030-015-9625-z>
- Gibson SJ, Bond NJ, Milne S, Lewis A, Sheriff A, Pettman G, Pradhan R, Higazi DR, Hatton D (2017) N-terminal or signal peptide sequence engineering prevents truncation of human monoclonal antibody light chains. *Biotechnol Bioeng* 114(9):1970–1977. <https://doi.org/10.1002/bit.26301>
- Graf M, Deml L, Wagner R (2004) Codon-optimized genes that enable increased heterologous expression in mammalian cells and elicit efficient immune responses in mice after vaccination of naked DNA. *Methods Mol Med* 94:197–210
- Han JH, Choi YS, Kim WJ, Jeon YH, Lee SK, Lee BJ, Ryu KS (2010) Codon optimization enhances protein expression of human peptide deformylase in *E. coli*. *Protein Expr Purif* 70(2):224–230. <https://doi.org/10.1016/j.pep.2009.10.005>
- Haryadi R, Ho S, Kok YJ, Pu HX, Zheng L, Pereira NA, Li B, Bi X, Goh LT, Yang Y, Song Z (2015) Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. *PLoS One* 10(2):e0116878. <https://doi.org/10.1371/journal.pone.0116878>
- Ho SC, Tong Y, Yang Y (2013) Generation of monoclonal antibody producing mammalian cell lines. *Pharm Bioproc* 1(1):71–87. <https://doi.org/10.4155/pbp.13.8>
- Hoerr I, Probst J, Pascolo S (2016) RNA coded antibody, patent, US20160185640A1
- Jostock T, Knopf HP (2012) Mammalian stable expression of biotherapeutics. *Methods Mol Biol* 899:227–238. https://doi.org/10.1007/978-1-61779-921-1_15
- Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T (2001) Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol* 53(4–5):290–298. <https://doi.org/10.1007/s002390010219>
- Kim JY, Kim YG, Lee GM (2012) CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl Microbiol Biotechnol* 93(3):917–930. <https://doi.org/10.1007/s00253-011-3758-5>
- Knappskog S, Ravneberg H, Gjerdrum C, Trosse C, Stern B, Pryme IF (2007) The level of synthesis and secretion of *Gaussia princeps* luciferase in transfected CHO cells is heavily dependent on the

- choice of signal peptide. *J Biotechnol* 128(4):705–715. <https://doi.org/10.1016/j.jbiotec.2006.11.026>
- Kober L, Zehe C, Bode J (2013) Optimized signal peptides for the development of high expressing CHO cell lines. *Biotechnol Bioeng* 110(4):1164–1173. <https://doi.org/10.1002/bit.24776>
- Kotia RB, Raghani AR (2010) Analysis of monoclonal antibody product heterogeneity resulting from alternate cleavage sites of signal peptide. *Anal Biochem* 399(2):190–195. <https://doi.org/10.1016/j.ab.2010.01.008>
- Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M (2006) High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol* 4(6):e180. <https://doi.org/10.1371/journal.pbio.0040180>
- Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* 324(5924):255–258. <https://doi.org/10.1126/science.1170160>
- Kuersten S, Goodwin EB (2003) The power of the 3' UTR: translational control and development. *Nat Rev Genet* 4(8):626–637. <https://doi.org/10.1038/nrg1125>
- Kunert R, Reinhart D (2016) Advances in recombinant antibody manufacturing. *Appl Microbiol Biotechnol* 100(8):3451–3461. <https://doi.org/10.1007/s00253-016-7388-9>
- Li Q, Chen X (2011) Mamalian cell expression vectors and utilization, patent, US20110097798A1
- Li F, Vijayasankaran N, Shen AY, Kiss R, Amanullah A (2010) Cell culture processes for monoclonal antibody production. *MAbs* 2(5):466–479
- Lopez-Lastra M, Rivas A, Barria MI (2005) Protein synthesis in eukaryotes: the growing biological relevance of cap-independent translation initiation. *Biol Res* 38(2–3):121–146
- Luo W, Chen Y, Li L, Xu C, Miao J, Shih JW, Zhang J, Xia N (2007) Construction and characterization of the chimeric antibody 8C11 to the hepatitis E virus. *FEMS Immunol Med Microbiol* 51(1):18–25. <https://doi.org/10.1111/j.1574-695X.2007.00253.x>
- Luo W, Chen Y, Zhen Z, Li G, Chen H, Zhang J, Xia N (2008) Humanized antibodies against H5 subtype avian influenza virus hemagglutinin protein and uses thereof, patent, CN101851289A
- Luo W, Zhou B, Zhang T, Yuan Q, Xia N (2016) Antibodies against hepatitis B surface antigen and uses thereof, patent, CN201610879693.5
- Majors BS, Arden N, Oyler GA, Chiang GG, Pederson NE, Betenbaugh MJ (2008a) E2F-1 overexpression increases viable cell density in batch cultures of Chinese hamster ovary cells. *J Biotechnol* 138(3–4):103–106. <https://doi.org/10.1016/j.jbiotec.2008.08.003>
- Majors BS, Betenbaugh MJ, Pederson NE, Chiang GG (2008b) Enhancement of transient gene expression and culture viability using Chinese hamster ovary cells overexpressing Bcl-x(L). *Biotechnol Bioeng* 101(3):567–578. <https://doi.org/10.1002/bit.21917>
- Mazumder B, Seshadri V, Fox PL (2003) Translational control by the 3'-UTR: the ends specify the means. *Trends Biochem Sci* 28(2):91–98. [https://doi.org/10.1016/S0968-0004\(03\)00002-1](https://doi.org/10.1016/S0968-0004(03)00002-1)
- Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res* 28(1):292
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997a) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 10(1):1–6
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997b) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst* 8(5–6):581–599
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786. <https://doi.org/10.1038/nmeth.1701>
- Robinson F, Jackson RJ, Smith CW (2008) Expression of human nPTB is limited by extreme suboptimal codon content. *PLoS One* 3(3):e1801. <https://doi.org/10.1371/journal.pone.0001801>
- Strohl WR (2017) Current progress in innovative engineered antibodies. *Protein Cell* 9:86–120. <https://doi.org/10.1007/s13238-017-0457-8>
- Trosse C, Ravneberg H, Stern B, Pryme IF (2007) Vectors encoding seven oikodin signal peptides transfected into CHO cells differ greatly in mediating *Gaussia* luciferase and human endostatin production although mRNA levels are largely unaffected. *Gene Regul Syst Biol* 1:303–312
- You M, Liu Y, Chen Y, Guo J, Wu J, Fu Y, Shen R, Qi R, Luo W, Xia N (2013) Maximizing antibody production in suspension-cultured mammalian cells by the customized transient gene expression method. *Biosci Biotechnol Biochem* 77(6):1207–1213. <https://doi.org/10.1271/bbb.120968>
- Young R, Rance J (2008) Mammalian expression vector with a highly efficient secretory signal sequence, patent, WO 2008/148519 A2
- Zhang L, Leng Q, Mixson AJ (2005) Alteration in the IL-2 signal peptide affects secretion of proteins *in vitro* and *in vivo*. *J Gene Med* 7(3):354–365. <https://doi.org/10.1002/jgm.677>
- Zhang G, Liu J, Fan W, Chen Q, Shi B (2017) An efficient transient expression system for enhancing the generation of monoclonal antibodies in 293 suspension cells. *Curr Pharm Biotechnol* 18(4):351–357. <https://doi.org/10.2174/1389201018666170320110545>