Biotechnology and Bioprocess Engineering (2024) 29:520–528
https://doi.org/10.1007/s12257-024-00096-4

RESEARCH PAPER

Establishment of a new promoter trapping vector using 2A peptide

Eun Seon Song[1] · Yun Haeng Lee[1] · Moon Kyoung So[1] · Myeong Uk Kuk[1] · Ji Ho Park[1] · Jee Hee Yoon[1] · Yoo Jin Lee[1] · Duyeol Kim[1] · Byeonghyeon So[1] · Youngjoo Byun[3] · Hyung Wook Kwon[1,2] · Joon Tae Park[1,2]

Received: 16 January 2024 / Revised: 11 March 2024 / Accepted: 19 March 2024 / Published online: 24 April 2024

Abstract
Promoter trapping is a powerful tool for discovering promoters and uses promoter trapping vectors. However, the traditional trapping vector allows expression even if it does not integrate into the host cell genome, and even if it does integrate into the genome, it is more likely to integrate in a region other than the promoter region. In this study, to overcome the shortcomings of traditional trapping vectors, we used the bicistronic 2A system to link GFP and the neomycin resistance gene. Because this vector does not contain a promoter, simultaneous production of GFP and neomycin resistance protein requires integration into the promoter region. In fact, GFP expression was observed in more than 90% of the cell clones that survived in the medium containing antibiotics, confirming that the 2A system operates. The vector insertion location was confirmed through whole genome sequence analysis, and a 1-kb promoter candidate region was selected through promoter motif analysis. In fact, a 1-kb region inserted into a promoterless luciferase expression vector showed strong promoter activity, demonstrating its utility as a tool to find promoters. In summary, we constructed a novel promoter trapping vector using the 2A system and used it to discover the promoter with strong activity. This vector will increase the efficiency of promoter trapping, providing an opportunity to easily discover new promoters in mammalian cells.

Keywords Promoter trapping · Bicistronic 2A system · Trapping vector

1 Introduction

Improving the protein productivity in mammalian cells has been a major research goal in the biopharmaceutical field, and approaches have been attempted from three aspects: process, cell line, and expression vector [1]. However, there are limits to controlling protein productivity simply by improving the process such as pH, media, dissolved oxygen concentration, and stirring speed [2–4]. Moreover, improvement of cell lines through gene editing is progressing slowly because it is not well known which genes play an important role in protein productivity [5, 6]. However, improvements in expression vectors have been more effective in increasing protein productivity than other methods by increasing transcription efficiency [7, 8]. Therefore, there is increasing research interest in improving expression vectors.

The expression vector consists of a coding sequence (CDS), a promoter driving CDS expression, and a transcription terminator for the CDS [7]. In expression vectors, gene expression is controlled by promoters, and protein production efficiency is closely related to the strength of promoter activity [9, 10]. Therefore, the discovery and characterization of promoters with higher activity is important for protein production [11]. Commonly used promoters for protein production are the simian virus 40 (SV40) or cytomegalovirus (CMV) promoters [12]. These virus-derived promoters enable the expression of large amounts of recombinant proteins, but tend to be epigenetically silenced, ultimately reducing protein productivity [13, 14]. These shortcomings have been largely compensated for by using promoters found in Chinese hamster ovary (CHO) cells. One successful example is the promoter of the Chinese hamster elongation

✉ Hyung Wook Kwon
hwkwon@inu.ac.kr

✉ Joon Tae Park
joontae.park@inu.ac.kr

1 Division of Life Sciences, College of Life Sciences and Bioengineering, Incheon National University, Incheon 22012, Korea

2 Convergence Research Center for Insect Vectors, Incheon National University, Incheon 22012, Korea

3 College of Pharmacy, Korea University, Sejong 30019, Korea

factor-1α (CHEF1) gene. The CHEF1 promoter was not epigenetically silenced and was able to produce higher amounts of protein than the CMV promoter [15]. If an endogenous promoter of CHO cells, such as the CHEF1 promoter, is discovered in the future, this promoter could be another alternative to ensure high productivity of the protein.

Promoter trapping is a tool used to discover and characterize new promoters [16, 17]. Although it is an important genetic tool, the probability of successful trapping is not high due to the disadvantages of traditional trapping vectors [16, 18]. For example, traditional vectors contain a marker gene without a promoter (e.g., GFP) and a resistance gene with a promoter (e.g., neomycin resistance gene) [19]. A marker gene is expressed when inserted into the promoter region and serves to identify the promoter. A resistance gene allows cells to survive in media containing antibiotics. However, a resistance gene with its own promoter can be expressed even when the vector is not integrated into the host genome. Therefore, among cell clones that have undergone antibiotic selection, most clones do not have the vector integrated into the genome [20]. Moreover, the integration of a vector into the host genome does not mean that it is integrated into the promoter region. Therefore, the subsequent process of finding clones that incorporated this into the promoter region was time-consuming and labor-intensive. If a trapping vector that can overcome these existing shortcomings is created, efficient promoter trapping might be possible.

2A peptide consists of a sequence of approximately 20 amino acids and serves to link two proteins [21]. Ribosome skipping occurs at the Gly and Pro sequences of the 2A peptide, resulting in releases of two independent peptide molecules from translation of a single mRNA [22]. Therefore, two proteins are produced equally from two linked genes. The usefulness of polycistronic protein production using 2A peptides was further supported by the finding that up to nine genes linked by 2A peptides were co-translated at the same level [23].

In this study, a new trapping vector using 2A peptide was developed to compensate for the shortcomings of existing trapping vectors. These vectors were integrated into the promoter region of the host genome, allowing simultaneous expression of a marker and resistance gene linked to the 2A peptide, greatly simplifying the subsequent identification process. Here, we utilized the developed trapping vector to discover a novel promoter region in CHO cells.

2 Materials and methods

2.1 Cell culture

This study used CHO DG44 cells (A1100001; Thermo Fisher Scientific). The previous method of cultivating cells was employed [24].

2.2 Plasmid design and construction

All plasmids were constructed using standard cloning techniques. The promoter trapping vector consists of a marker gene (GFP) and a resistance gene (neomycin resistance gene) linked by P2A, a 2A peptide. To generate a backbone vector (BV), the pcDNA3.1 vector (V79020; Invitrogen) was modified to contain only promoterless luciferase gene. To ensure that there was promoter activity in the region where the trapping vector was inserted, a 3-kb region (275,654,178–275,657,178 of NC_048596.1 (chr3)) of the insertion region was cloned into BV in forward or reverse orientation. To confirm the promoter analysis, a 1-kb region (275,654,178–275,655,178 of NC_048596.1 (chr3)) was cloned into BV.

2.3 Western blot analysis

Western blot analysis was followed as previously described [25]. Antibodies used in this study included HRP–conjugated GFP antibody (sc-9996 HRP; Santa Cruz Biotechnology, 1:500 dilution in PBS) and HRP–conjugated β–actin (sc-47778; Santa Cruz Biotechnology, 1:1000 dilution in PBS).

2.4 Transfection of the promoter trapping vector

2 μg of promoter trapping vector was transfected into 2×10^6 CHO DG44 cells using Lipofectamine™ 2000 transfection reagent (11668-019; Thermo Fisher Scientific). Using 250 μg/mL G418 disulfate salt solution (G418) (ant-gn-1; Thermo Fisher Scientific) for 2 weeks, transfected cells were chosen.

2.5 Flow cytometric analysis

Using flow cytometry, a FITC setup (530/30 nm bandpass filter with excitation at 488 nm) was used to calculate the proportion of cells expressing GFP. The Cell Quest 3.2 application (Becton Dickinson) was utilized to analyze the results.

2.6 Selection of single cells

Following the selection of antibiotics, 96-well plates (353,072; Falcon) were used to seed single cells into each well. Cells were grown for 21 days. Using a fluorescent microscope (Axiovert 200; Carl Zeiss), single cells expressing GFP were chosen.

2.7 Preparation of genomic DNA

Genomic DNA was isolated from single cells expressing GFP using a genomic DNA prep kit (SGD41-C100; Solgent).

2.8 Library preparation and whole genome sequencing

Whole genome sequencing and library preparation were performed at Theragen Bio Itex. TruSeq Nano DNA Library Prep Kit (FC-121-4001; Illumina) was used for library preparation. A 350-bp insert size was generated through DNA size selection combined with adapters during library construction [26]. Paired-end reads of 2×150 bases were used for the run. Sequencing was then performed using the Illumina Novaseq 6000 platform. Cluster generation was performed on a flow cell using libraries built on cBot hardware (Illumina). After sequencing raw reads, adapter sequences were trimmed using cutadapt v1.10 [27]. Reads selected for assembly scored higher than Q30. A new assembly of high-quality reads was then completed using IDBA-UD [28].

2.9 Measurement of luciferase activity

A kit (E1500; Promega) was used to measure the luciferase activity. Every process was carried out using the guidelines provided in earlier research investigations [29].

2.10 Promoter motif analysis

Promoter motif analysis was conducted using FPROM software (Softberry, Inc.). Threshold for TATA-box less promoters was 0.80.

2.11 Statistical analyses

A statistical software program (SigmaPlot 12.5; Systat Software) was used to conduct the statistical analyses. A Student's *t*-test was employed to ascertain the significance of the difference.

3 Results

3.1 Polycistronic expression using 2A peptide is applicable to CHO cells

To determine whether the two genes linked by the 2A peptide were expressed in CHO cells, a marker gene (GFP) and a resistance gene (neomycin resistance gene) were linked by the 2A peptide (Fig. 1A; a.k.a., GFP-2A-neomycin resistance gene). Before using the GFP-2A-neomycin resistance gene in a trapping vector, we wanted to determine whether it could be expressed in CHO cells. The CMV promoter was placed in front of the GFP-2A-neomycin resistance gene, allowing the expression of GFP protein and neomycin resistance protein (Fig. 1A). A schematic representation of independent protein production from the 2A peptide-containing transgene is shown in Fig. 1A. Vector including CMV-GFP-2A-neomycin resistance gene was transfected into 2×10^6 cells. Transfected cells were selected using a medium containing 250 µg/mL G418 for 2 weeks. Then, the expression of GFP linked to the 2A peptide was examined using a fluorescence microscope. While no GFP was observed in cells transfected with the control vector, more than 90% of cells transfected with the vector expressed GFP (Fig. 1B). These data indicate that GFP and neomycin resistance genes were co-expressed. Western blotting also identified a single protein band corresponding to the molecular mass of the GFP protein in cells transfected with the vector (Fig. 1C). These results indicate that ribosomal skipping occurred on the 2A peptide and two proteins (GFP and neomycin resistance protein) were produced.

3.2 Promoter trapping using promoterless GFP-2A-neomycin resistance gene

Confirmation that the 2A system works in CHO cells led us to investigate whether this new vector could be applied for promoter trapping in CHO cells. Because this vector lacks a promoter, it must enter the endogenous promoter region of the CHO genome for expression (Fig. 2A). A schematic representation of independent protein production after integration into the endogenous promoter region of CHO cells is shown in Fig. 2A. The promoter trapping vector including promoterless GFP-2A-neomycin resistance gene was transfected into 2×10^6 cells. After transfection, a selection process was performed in medium containing 250 µg/mL G418 for 2 weeks. To assess the proportion of cells expressing GFP after antibiotic selection, flow cytometry was performed. GFP expression was detected in 84.12% of surviving cells, whereas GFP was detected in 0.09% of cells transfected with control vector (Fig. 2B).

To select cells expressing only GFP, single cell isolation was performed and 12 single clones expressing GFP were generated. Promoter trapping vectors contain a GFP gene that can be expressed when inserted into the promoter region. Since strong GFP expression indicates that the trapping vector was integrated into the promoter region with strong promoter activity, the single clone showing the strongest GFP expression among 12 single clones was selected using a fluorescence microscope. Then, the

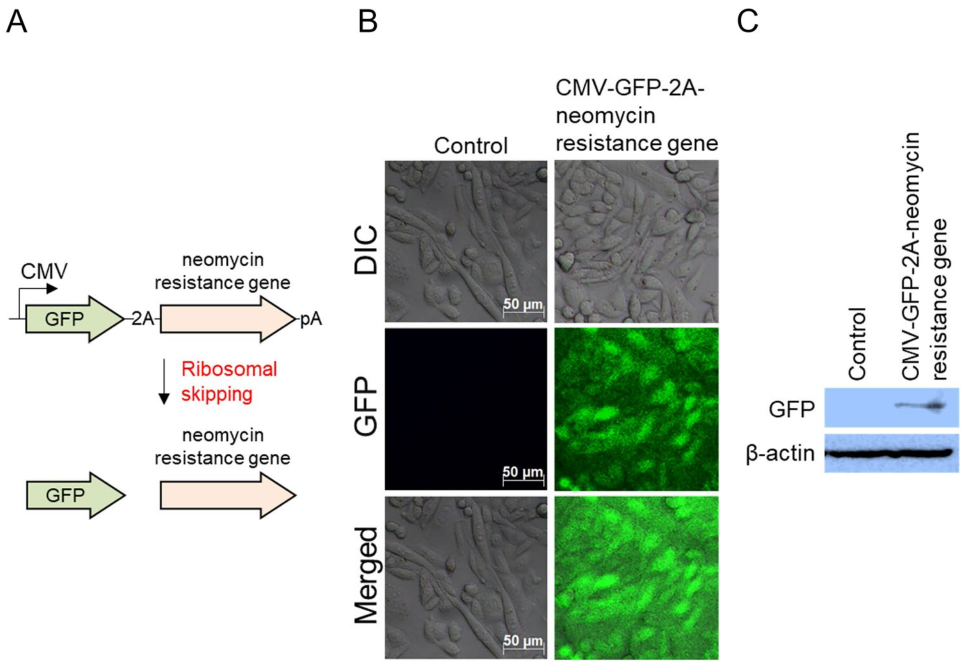

Fig. 1 Polycistronic expression using 2A peptide is applicable to CHO cells. **A** CMV promoter was placed in front of the GFP-2A-neomycin resistance gene. Schematic representation of independent protein production from a 2A-containing transgene is shown. **B** Comparison of differential interference contrast (DIC) and fluorescent micrographs of cells transfected with control vector or vector containing CMV-GFP-2A-neomycin resistance gene. No GFP was observed in cells transfected with control vector, whereas more than 90% of cells expressed GFP in cells transfected with vector containing CMV-GFP-2A-neomycin resistance gene (Scale bar 50 μm). **C** Western blotting identified a single protein band corresponding to the molecular mass of the GFP protein in cells transfected with vector containing CMV-GFP-2A-neomycin resistance gene. *CHO* Chinese hamster ovary, *CMV* cytomegalovirus

proportion of cells expressing GFP in the selected single clone was assessed using flow cytometry. GFP expression was detected in 98.13% of cells (Fig. 2B). These results indicate that the vector integrates the promoter region of CHO cells and that GFP and neomycin resistance genes are co-expressed.

3.3 Identification of promoter trapping vector insertion sites in the CHO genome

Whole genome sequencing was performed to determine which region of the CHO genome trapping vectors were integrated into. Whole genome sequencing is an important tool that can analyze millions of DNA fragments simultaneously to determine where trapping vectors have integrated the CHO genome [30]. Among the soft clipped reads, actual split reads were discovered (Fig. 3, black dotted line). The soft clipping location indicated by the black dashed line indicates where the vector was inserted (Fig. 3). DNA sequence information was shown to the right of the black dotted line (Fig. 3). These DNA sequences were those of the promoter trapping vector, indicating that the vector was integrated into the CHO genome (Fig. 3). The exact location where the vector was inserted was the 275,654,190–275,654,265 region of NC_048596.1 (chr3) (Fig. 3). The insertion site was within the intronic region of the TLC domain containing the 3B (Tlcd3b) gene (Fig. 3). To determine whether there was promoter activity in the region where the vector was inserted, a 3-kb region of the insertion region (275,654,178–275,657,178 of NC_048596.1 (chr3)) was selected in forward or reverse orientation (Fig. 3).

3.4 Identification of the promoter region with promoter activity

To confirm whether the region into which the promoter trapping vector was integrated had promoter activity, the 3-kb region was inserted into a promoterless luciferase expression vector in the forward or reverse direction (Fig. 4A). As a negative control, a blank vector (BV) containing a luciferase gene without a promoter was used (Fig. 4A). The 3-kb reverse region showed no promoter activity, whereas the 3-kb forward region showed a slight increase in promoter activity compared to BV (Fig. 4A). These data indicate that the core region exhibiting promoter activity exists within the 3-kb forward region. Additionally, these data suggest that higher promoter activity could be observed if only the core region was used. Promoter motif analysis using FPROM

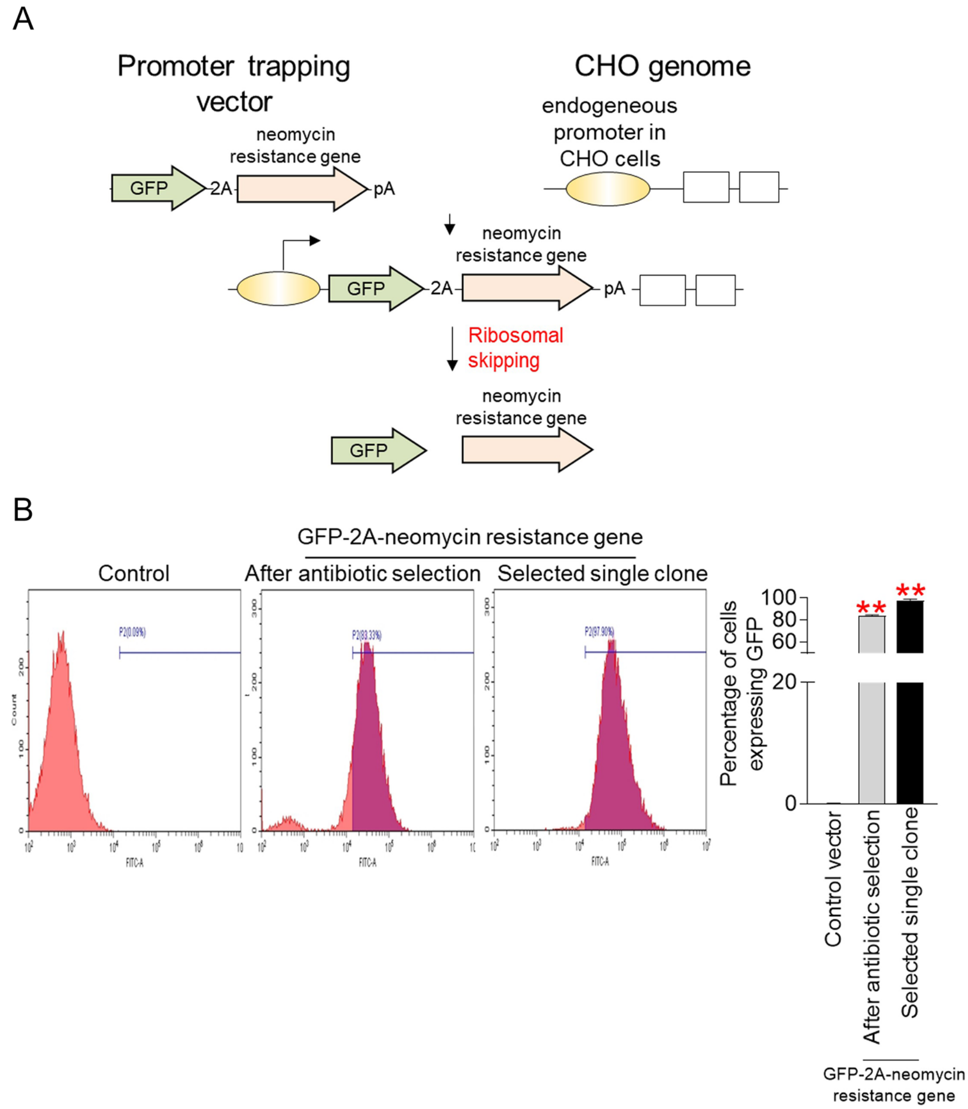

Fig. 2 Promoter trapping using promoterless GFP-2A-neomycin resistance gene. **A** Promoter trapping vector (promoterless GFP-2A-neomycin resistance gene) was transfected to CHO cells. Because this vector lacks a promoter, it must enter the endogenous promoter region of the CHO genome for expression. Schematic representation of independent protein production after integrating in endogenous promoter region in CHO cells is shown. **B** Flow cytometric analysis to assess the proportion of cells expressing GFP. The representative histogram of each group was shown. $^{**}p<0.01$, Student's *t*-test. Means ± SD, N = 3. *CHO* Chinese hamster ovary

software (Softberry, Inc.) was performed to determine which portion of the 3-kb forward region exhibited promoter activity. Promoter motif analysis identified that the TATA box was located at positions 275,654,980–275,654,987 of the 3-kb forward region (Fig. 4B). The 1-kb region where the TATA box is located (275,654,178–275,655,178 of NC_048596.1 (chr3)) was cloned into the BV vector (Fig. 4C). A 1-kb region inserted into a promoterless luciferase expression vector showed a significant increase in promoter activity compared to BV (Fig. 4C).

4 Discussion

Promoters are one of the important elements that allow expression vectors to achieve high and stable protein production [31]. Virus-derived promoters are most commonly used for commercial protein production [32]. However, these promoters do not respond to changes in the internal environment of the host cell, resulting in overexpression of the transgene. Overexpression caused severe stress to cells, resulting in side effects such as early cell death [33]. These promoters also showed diverse activity patterns depending on the cell type. For example, the virus-derived CMV promoter showed 207% activity in human HeLa cells, whereas in other mammalian cell types its activity ranged from 21 to 113% [34]. This variability limits the applicability of this promoter for commercial protein production. To compensate for these shortcomings, an alternative strategy using endogenous promoters derived from each cell has been proposed. Promoter trapping methods have been used to find endogenous promoters [35, 36]. However, fundamental problems with the vectors used for promoter trapping significantly delay the process of finding new promoters. For example, a promoter trap vector consists of a marker gene without a promoter and a resistance gene with a promoter [35, 37–39].

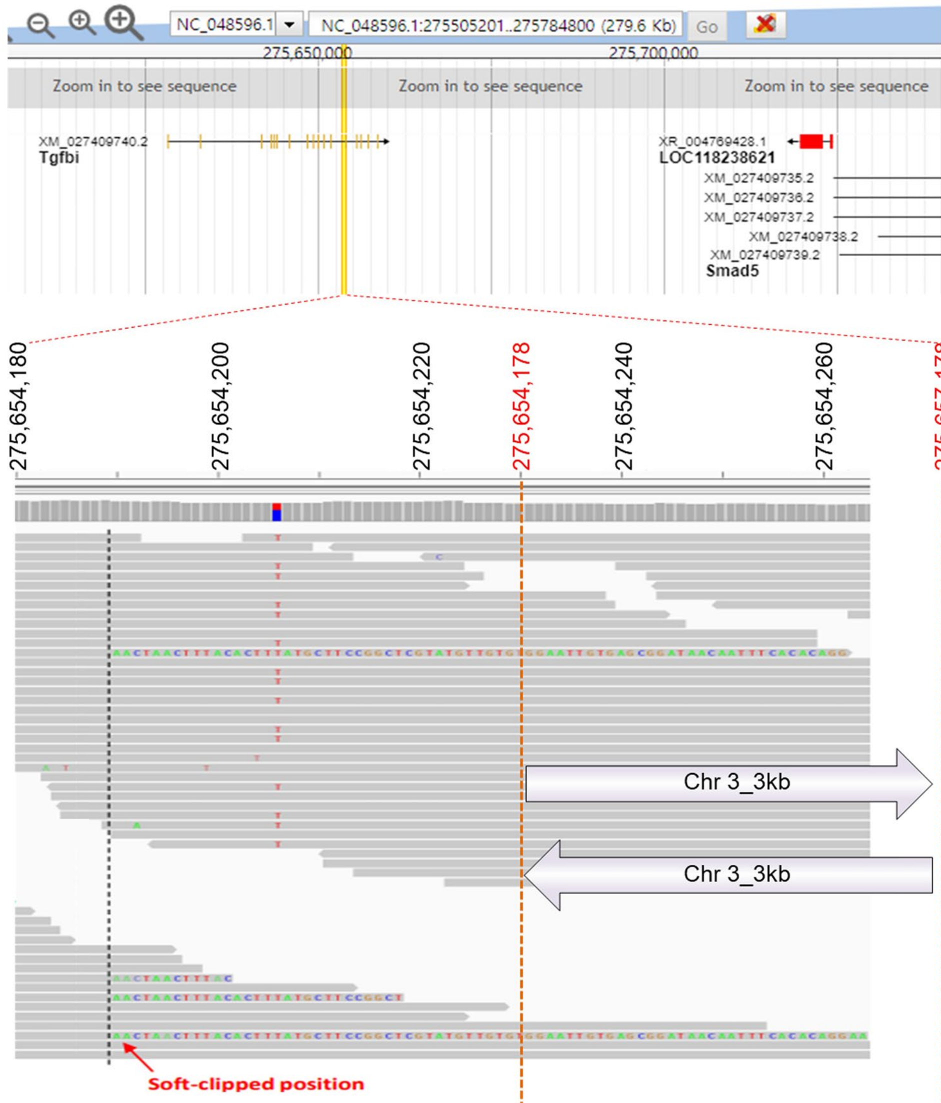

Fig. 3 Identification of promoter trapping vector insertion sites in the CHO genome. Whole genome sequencing was performed to determine which regions of the CHO genome were integrated with the promoter trapping vector. The soft-clipped position indicated by the black dotted line indicates where the vector was inserted. The insertion site of promoter trapping vector into the CHO genome was the 275,654,190–275,654,265 region of NC_048596.1 (chr3). The insertion site was within the intronic region of the TLC domain containing the 3B (Tlcd3b) gene. To ensure that there was promoter activity in the region where the vector was inserted, a 3-kb region (275,654,178–275,657,178 of NC_048596.1 (chr3)) of the insertion region was selected in forward or reverse orientation. *CHO* Chinese hamster ovary

Resistance genes with promoters can be expressed at any time, allowing host cells to survive the antibiotic selection process even if the marker gene is not integrated into the host genome. Additionally, integration of a promoter trapping vector into the host cell genome does not mean that it is integrated into the promoter region. In this study, we developed a promoter trapping vector that can solve the shortcomings of traditional trapping vectors. This vector uses the 2A peptide to enable simultaneous expression of marker and resistance proteins under the influence of one promoter. Indeed, after antibiotic selection, more than 90% of surviving cells were found to express GFP, indicating that the 2A peptide worked well. These findings also indicate that the vector was integrated into the promoter region of the host genome. After obtaining a single clone, whole genome sequencing allowed us to find the region where the vector was integrated. Based on the insertion site, we were able to pinpoint the site with actual promoter activity. To the best of our knowledge, this study is the first to apply the 2A peptide to a promoter trapping vector. In a previous study, the 2A peptide was used to target adeno-associated virus-mediated genes, but was not used to identify the promoter itself [40]. Therefore, this study applying the 2A peptide to a promoter trapping vector will complement the shortcomings of traditional trapping vectors, making it easy to find promoters through promoter trapping.

The most urgent goal of the biopharmaceutical industry is to dramatically increase protein productivity. Improvements in protein productivity were mainly achieved through controlling factors related to process operating conditions [4]. However, there was a limit to the protein production that could be increased through process improvement, and as a result, the high expectations of the biopharmaceutical industry were not met [41]. Other studies have attempted to achieve this goal through cell line improvement. In fact, manipulating anti-apoptotic genes or down-regulation

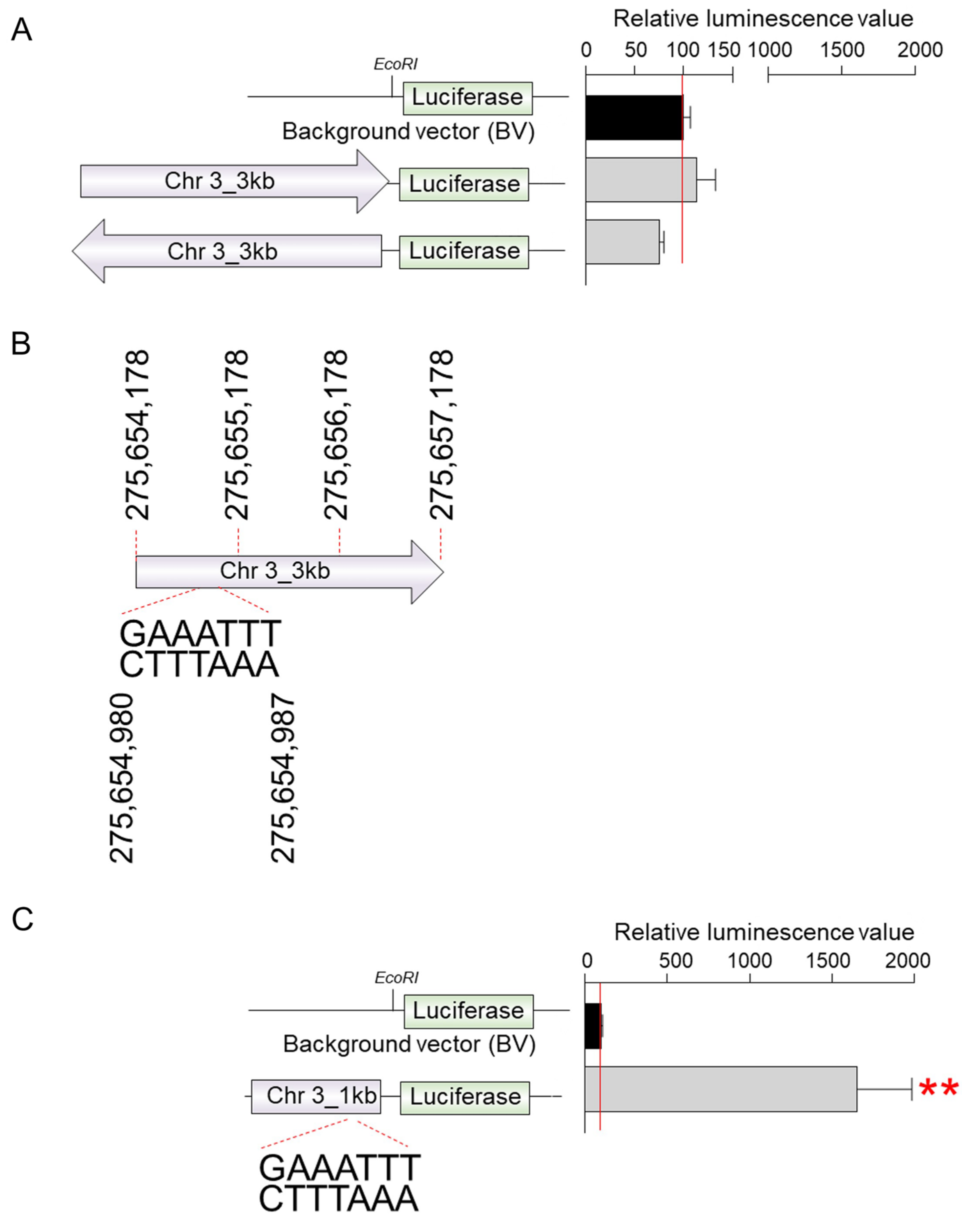

Fig. 4 Identification of the promoter region with promoter activity. **A** Comparison of the luminescence values of a blank vector (BV) containing the luciferase gene without a promoter or a vector containing the luciferase gene with a "3-kb forward region" and "3-kb reverse region" promoters. Mean ± SD, N = 3. **B** Promoter motif analysis using FPROM software (Softberry, Inc.) identified that TATA box located in 275,654,980–275,654,987 location of 3-kb forward region. **C** A 1-kb region (275,654,178–275,655,178 of NC_048596.1 (chr3)) where TATA box locates was re-cloned into the BV vector. A 1-kb region inserted into a promoterless luciferase expression vector showed a significant increase in promoter activity over BV. $^{**}p < 0.01$, Student's *t*-test. Mean ± SD, N = 3

of pro-apoptotic proteins increased protein productivity [42–45]. However, there are limits to improving cell lines through genetic manipulation because it is difficult to know which genes need to be regulated in a short period of time. Recently, studies have reported that protein productivity can be effectively increased through the control of expression vector [46–48]. In particular, by manipulating the promoter constituting the expression vector, transcriptional activity could be controlled, making it possible to produce a cell line with higher protein productivity [49, 50]. In this study, we discovered a novel endogenous promoter in the CHO genome using a novel trapping vector. However, we acknowledge that we have not been able to apply this promoter to commercial protein production to examine how much protein can be produced per unit volume. If further research is conducted on protein productivity using this promoter, the utility of this promoter will be further expanded. Additionally, applying the strategies used to optimize the virus-derived CMV and SV40 promoters to this promoter will likely lead to optimized promoters with consistently high levels of protein productivity.

5 Conclusion

We used a bicistronic 2A system to overcome the shortcomings of conventional trapping vectors. This new system allowed us to effectively discover the new promoter with higher activity in CHO cells. This new trapping vector makes it easy to find endogenous promoters in a variety of cells, which could be a groundbreaking tool for increasing protein productivity in biopharmaceutics.

Acknowledgements This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HP23C0024). This study was also supported by the Basic Research Program (NRF-2021R1A2C1004298) through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Technology and by the Priority Research Centers Program (2020R1A6A1A0304195411) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education.

Author contributions ESS: writing—original draft, writing—review and editing, investigation. MKS: investigation. YHL: investigation. MUK: investigation. JHP: investigation. JHY: investigation. YJL: investigation. DK: investigation. BS: investigation. YB: analysis. HWK: writing—review and editing. JTP: writing—original draft, writing—review and editing.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval Neither ethical approval nor informed consent was required for this study.

References

1. Pontiller J, Gross S, Thaisuchat H et al (2008) Identification of CHO endogenous promoter elements based on a genomic library approach. Mol Biotechnol 39:135–139. https://doi.org/10.1007/s12033-008-9044-9
2. Lu F, Toh PC, Burnett I et al (2013) Automated dynamic fed-batch process and media optimization for high productivity cell culture process development. Biotechnol Bioeng 110:191–205. https://doi.org/10.1002/bit.24602
3. Hashizume T, Ozawa Y, Ying BW (2023) Employing active learning in the optimization of culture medium for mammalian cells. NPJ Syst Biol Appl 9:20. https://doi.org/10.1038/s41540-023-00284-7
4. Radhakrishnan D, Wells EA, Robinson AS (2018) Strategies to enhance productivity and modify product quality in therapeutic proteins. Curr Opin Chem Eng 22:81–88. https://doi.org/10.1016/j.coche.2018.09.005
5. Lim Y, Wong NS, Lee YY et al (2010) Engineering mammalian cells in bioprocessing—current achievements and future perspectives. Biotechnol Appl Biochem 55:175–189. https://doi.org/10.1042/BA20090363
6. Kuystermans D, Krampe B, Swiderek H et al (2007) Using cell engineering and omic tools for the improvement of cell culture processes. Cytotechnology 53:3–22. https://doi.org/10.1007/s10616-007-9055-6
7. Khan KH (2013) Gene expression in Mammalian cells and its applications. Adv Pharm Bull 3:257–263. https://doi.org/10.5681/apb.2013.042
8. Li ZM, Fan ZL, Wang XY et al (2022) Factors affecting the expression of recombinant protein and improvement strategies in Chinese hamster ovary cells. Front Bioeng Biotechnol 10:880155. https://doi.org/10.3389/fbioe.2022.880155
9. Blazeck J, Garg R, Reed B et al (2012) Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. Biotechnol Bioeng 109:2884–2895. https://doi.org/10.1002/bit.24552
10. Juven-Gershon T, Kadonaga JT (2010) Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol 339:225–229. https://doi.org/10.1016/j.ydbio.2009.08.009
11. Baek D, Davis C, Ewing B et al (2007) Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. Genome Res 17:145–155. https://doi.org/10.1101/gr.5872707
12. Wang XY, Zhang JH, Zhang X et al (2016) Impact of different promoters on episomal vectors harbouring characteristic motifs of matrix attachment regions. Sci Rep 6:26446. https://doi.org/10.1038/srep26446
13. Brooks AR, Harkins RN, Wang P et al (2004) Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. J Gene Med 6:395–404. https://doi.org/10.1002/jgm.516
14. 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. https://doi.org/10.1002/bit.23216
15. Ebadat S, Ahmadi S, Ahmadi M et al (2017) Evaluating the efficiency of CHEF and CMV promoter with IRES and Furin/2A linker sequences for monoclonal antibody expression in CHO cells. PLoS ONE 12:e0185967. https://doi.org/10.1371/journal.pone.0185967
16. Blanvillain R, Gallois P (2008) Promoter trapping system to study embryogenesis. Methods Mol Biol 427:121–135. https://doi.org/10.1007/978-1-59745-273-1_10
17. Dunn AK, Handelsman J (1999) A vector for promoter trapping in Bacillus cereus. Gene 226:297–305. https://doi.org/10.1016/s0378-1119(98)00544-7
18. De Palma M, Montini E, Santoni de Sio FR et al (2005) Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells. Blood 105:2307–2315. https://doi.org/10.1182/blood-2004-03-0798
19. Durick K, Mendlein J, Xanthopoulos KG (1999) Hunting with traps: genome-wide strategies for gene discovery and functional analysis. Genome Res 9:1019–1025. https://doi.org/10.1101/gr.9.11.1019
20. Gentile A, D'Alessandro L, Medico E (2003) Gene trapping: a multi-purpose tool for functional genomics. Biotechnol Genet Eng Rev 20:77–100. https://doi.org/10.1080/02648725.2003.10648038
21. Liu Z, Chen O, Wall JBJ et al (2017) Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci Rep 7:2193. https://doi.org/10.1038/s41598-017-02460-2
22. Wang Y, Wang F, Wang R et al (2015) 2A self-cleaving peptide-based multi-gene expression system in the silkworm *Bombyx mori*. Sci Rep 5:16273. https://doi.org/10.1038/srep16273
23. Geier M, Fauland P, Vogl T et al (2015) Compact multi-enzyme pathways in *P. pastoris*. Chem Commun (Camb) 51:1643–1646. https://doi.org/10.1039/c4cc08502g
24. Hwang SY, Lee YH, Kuk MU et al (2021) Improvement of Tol2 transposon system enabling efficient protein production in CHO cells. Biotechnol Bioprocess Eng 26:767–775. https://doi.org/10.1007/s12257-020-0310-4
25. Lee YH, Kim M, Park HJ et al (2023) Chemical screening identifies the anticancer properties of *Polyporous parvovarius*. J Cancer 14:50–60. https://doi.org/10.7150/jca.78302
26. Head SR, Komori HK, LaMere SA et al (2014) Library construction for next-generation sequencing: overviews and challenges. Biotechniques 56:61–64. https://doi.org/10.2144/000114133

27. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200
28. Peng Y, Leung HC, Yiu SM et al (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174
29. Lee YH, Park JY, Song ES et al (2022) Improvement of sleeping beauty transposon system enabling efficient and stable protein production. Biotechnol Bioprocess Eng 27:353–360. https://doi.org/10.1007/s12257-021-0231-x
30. Satam H, Joshi K, Mangrolia U et al (2023) Next-generation sequencing technology: current trends and advancements. Biology (Basel) 12:997. https://doi.org/10.3390/biology12070997
31. Romanova N, Noll T (2018) Engineered and natural promoters and chromatin-modifying elements for recombinant protein expression in CHO cells. Biotechnol J 13:e1700232. https://doi.org/10.1002/biot.201700232
32. Rodova M, Jayini R, Singasani R et al (2013) CMV promoter is repressed by p53 and activated by JNK pathway. Plasmid 69:223–230. https://doi.org/10.1016/j.plasmid.2013.01.004
33. Tossolini I, Gugliotta A, López Díaz F et al (2022) Screening of CHO-K1 endogenous promoters for expressing recombinant proteins in mammalian cell cultures. Plasmid 119–120:102620. https://doi.org/10.1016/j.plasmid.2022.102620
34. Schlabach MR, Hu JK, Li M et al (2010) Synthetic design of strong promoters. Proc Natl Acad Sci USA 107:2538–2543. https://doi.org/10.1073/pnas.0914803107
35. Chen J, Haverty J, Deng L et al (2013) Identification of a novel endogenous regulatory element in Chinese hamster ovary cells by promoter trap. J Biotechnol 167:255–261. https://doi.org/10.1016/j.jbiotec.2013.07.001
36. Yoo YH, Kim YJ, Moon S et al (2020) Fast track to discover novel promoters in rice. Plants (Basel) 9:125. https://doi.org/10.3390/plants9010125
37. von Melchner H, Ruley HE (1989) Identification of cellular promoters by using a retrovirus promoter trap. J Virol 63:3227–3233. https://doi.org/10.1128/JVI.63.8.3227-3233.1989
38. Friedrich G, Soriano P (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev 5:1513–1523. https://doi.org/10.1101/gad.5.9.1513
39. Reddy S, DeGregori JV, von Melchner H et al (1991) Retrovirus promoter-trap vector to induce lacZ gene fusions in mammalian cells. J Virol 65:1507–1515. https://doi.org/10.1128/JVI.65.3.1507-1515.1991
40. Karnan S, Ota A, Konishi Y et al (2016) Improved methods of AAV-mediated gene targeting for human cell lines using ribosome-skipping 2A peptide. Nucleic Acids Res 44:e54. https://doi.org/10.1093/nar/gkv1338
41. Szkodny AC, Lee KH (2022) Biopharmaceutical manufacturing: historical perspectives and future directions. Annu Rev Chem Biomol Eng 13:141–165. https://doi.org/10.1146/annurev-chembioeng-092220-125832
42. Chiang GG, Sisk WP (2005) Bcl-x(L) mediates increased production of humanized monoclonal antibodies in Chinese hamster ovary cells. Biotechnol Bioeng 91:779–792. https://doi.org/10.1002/bit.20551
43. Cost GJ, Freyvert Y, Vafiadis A et al (2010) BAK and BAX deletion using zinc-finger nucleases yields apoptosis-resistant CHO cells. Biotechnol Bioeng 105:330–340. https://doi.org/10.1002/bit.22541
44. Kim NS, Lee GM (2000) Overexpression of bcl-2 inhibits sodium butyrate-induced apoptosis in Chinese hamster ovary cells resulting in enhanced humanized antibody production. Biotechnol Bioeng 71:184–193. https://doi.org/10.1002/1097-0290(2000)71:3%3c184::aid-bit1008%3e3.0.co;2-w
45. Majors BS, Betenbaugh MJ, Pederson NE et al (2009) Mcl-1 overexpression leads to higher viabilities and increased production of humanized monoclonal antibody in Chinese hamster ovary cells. Biotechnol Prog 25:1161–1168. https://doi.org/10.1002/btpr.192
46. Kallnik V, Meyer M, Deppenmeier U et al (2010) Construction of expression vectors for protein production in *Gluconobacter oxydans*. J Biotechnol 150:460–465. https://doi.org/10.1016/j.jbiotec.2010.10.069
47. Zhang K, Su L, Duan X et al (2017) High-level extracellular protein production in Bacillus subtilis using an optimized dual-promoter expression system. Microb Cell Fact 16:32. https://doi.org/10.1186/s12934-017-0649-1
48. Schütz A, Bernhard F, Berrow N et al (2023) A concise guide to choosing suitable gene expression systems for recombinant protein production. STAR Protoc 4:102572. https://doi.org/10.1016/j.xpro.2023.102572
49. Jin LQ, Jin WR, Ma ZC et al (2019) Promoter engineering strategies for the overproduction of valuable metabolites in microbes. Appl Microbiol Biotechnol 103:8725–8736. https://doi.org/10.1007/s00253-019-10172-y
50. Ho SC, Yang Y (2014) Identifying and engineering promoters for high level and sustainable therapeutic recombinant protein production in cultured mammalian cells. Biotechnol Lett 36:1569–1579. https://doi.org/10.1007/s10529-014-1523-4

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.