Research Article

A novel regulatory element (E77) isolated from CHO-K1 genomic DNA enhances stable gene expression in Chinese hamster ovary cells

Shin-Young Kang[1], Yeon-Gu Kim[1,2], Seunghee Kang[2], Hong W. Lee[1,2] and Eun G. Lee[1,2]

[1] Department of Bioprocess Engineering, Korea University of Science and Technology (UST), Daejeon, Republic of Korea
[2] Biotechnology Process Engineering Center, KRIBB, Ochang, Republic of Korea

Vectors flanked by regulatory DNA elements have been used to generate stable cell lines with high productivity and transgene stability; however, regulatory elements in Chinese hamster ovary (CHO) cells, which are the most widely used mammalian cells in biopharmaceutical production, are still poorly understood. We isolated a novel gene regulatory element from CHO-K1 cells, designated E77, which was found to enhance the stable expression of a transgene. A genomic library was constructed by combining CHO-K1 genomic DNA fragments with a CMV promoter-driven GFP expression vector, and the E77 element was isolated by screening. The incorporation of the E77 regulatory element resulted in the generation of an increased number of clones with high expression, thereby enhancing the expression level of the transgene in the stable transfectant cell pool. Interestingly, the E77 element was found to consist of two distinct fragments derived from different locations in the CHO genome shotgun sequence. High and stable transgene expression was obtained in transfected CHO cells by combining these fragments. Additionally, the function of E77 was found to be dependent on its site of insertion and specific orientation in the vector construct. Our findings demonstrate that stable gene expression mediated by the CMV promoter in CHO cells may be improved by the isolated novel gene regulatory element E77 identified in the present study.

Supporting information available online

1 Introduction

Chinese hamster ovary (CHO) cells are the most widely utilized mammalian cell line for biopharmaceutical production because of their capacity for post-translational modifications, which are essential to the biological activities of the therapeutic product [1]. The establishment of high-producing CHO cells and optimization of the cell culture process have led to significant improvement in therapeutic recombinant protein production at titers greater than 10 g/L [2]. Furthermore, the recent publication of the genome sequence of CHO cells is expected to enable further optimization of these cells for applications in biotechnology [3]. Therefore, it is not surprising that great emphasis is placed on CHO cells as the major platform for therapeutic protein production.

Correspondence: Dr. Eun Gyo Lee, Biotechnology Process Engineering Center, KRIBB, 30 Yeongudanji-ro, Ochang-eup, Cheongju-si, Chungcheongbuk-do 363-883, Republic of Korea
Email: eglee@kribb.re.kr

Abbreviations: AP-1, activator protein 1; **BLAST**, basic local alignment search tool; **CHO**, Chinese hamster ovary; **CMV**, cytomegalovirus; **EASE**, expression augmenting sequence element; **E-box**, enhancer box; **FACS**, fluorescence-activated cell sorting; **GFP**, green fluorescent protein; **IRES**, internal ribosome entry site; **MAR**, matrix attachment region; **STAR**, stabilizing and anti-repressor element; **TFRE**, transcription factor regulatory element; **UCOE**, ubiquitous chromatin opening element; **Zeo^R**, zeocin resistance gene

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2016, 11, 633–641

BIO tec
visions
www.biotecvisions.com

The development of suitable CHO cell lines for therapeutic recombinant protein production typically involves vector engineering, gene amplification, and clone screening technologies [4]. It is necessary to construct an optimized expression vector that includes a strong promoter, an appropriate signal peptide, selected introns, and codon optimization [5]. Vector delivery into mammalian cells results in the integration of the plasmid DNA into the host genomic DNA and generates significant clonal variation in terms of the level and stability of expression due to position effect variegation and position-dependent silencing [6]. In order to overcome these issues, gene regulatory DNA elements, such as matrix attachment regions (MARs), insulators, ubiquitous chromatin opening elements (UCOEs), stabilizing and anti-repressor elements (STARs), and expression augmenting sequence element (EASE), have been identified and utilized in vector engineering, resulting in the generation of stable mammalian cell lines with high expression [7]. These gene regulatory elements have been evaluated for their enhancing and stabilizing effects on transgene expression driven by the human cytomegalovirus (CMV) promoter in CHO cells.

Several approaches have been used to identify novel regulatory DNA elements in genomic DNA. The human MARs were identified by genome-wide prediction using computational analysis of the chicken lysozyme MAR element, which is a well-known gene regulatory DNA element [8, 9]. A human genomic library was constructed to screen for STARs that block chromatin-associated repressors related to heterochromatin formation [10]. EASE was obtained from genomic DNA in CHO cells with high expression of the recombinant dimeric tumor necrosis factor receptor driven by the CMV promoter [11]. The genomic library approach has also been used to search for novel regulatory DNA elements in CHO cells [12, 13], and these studies have focused on the identification of CHO endogenous promoter elements to replace the usage of viral promoters such as CMV or SV40. Therefore, although CHO cells are the preferred mammalian cell line for the production of recombinant therapeutic proteins, gene regulatory elements from CHO cells that may be utilized in combination with strong viral promoters are not as well studied as those from cells of human origin.

In the present study, we identified novel gene regulatory elements in CHO-K1 genomic DNA and demonstrated that these elements enhance recombinant protein expression driven by the human CMV promoter in CHO cells. After the generation of stable cells with the genomic DNA library, cells expressing the highest levels of green fluorescent protein (GFP) were isolated by fluorescence-activated cell sorting (FACS). A genomic DNA fragment designated E77, which consisted of two distinct fragments derived from different locations in the CHO genome shotgun sequence, was obtained from a clone with the highest fluorescence intensity. The E77 fragment was reinserted into a GFP expression vector and its effectiveness and characteristics as a gene regulatory DNA element were verified in CHO cells.

2 Materials and methods

2.1 Vector construction

A pcDNA/GFP/Zeo plasmid vector, modified from the pcD-NA3.1(+)/Zeo vector (Invitrogen, Carlsbad, CA, USA), was used to construct the CHO-K1 genomic DNA library. The GFP-coding sequence was obtained from the pIRES2-EGFP vector (Clontech, Mountain View, CA, USA) and cloned into *Nhe*I and *Xba*I sites of the pcDNA3.1(+)/Zeo vector. A *Bam*HI restriction site was introduced 231 bp downstream of the 5' end of the pcDNA3.1(+)/Zeo vector, and placed upstream of the CMV promoter.

pGIZ was constructed as a bicistronic vector encoding a single transcription unit, the GFP and zeocin resistance gene (Zeo^R), with an internal ribosome entry site (IRES) obtained from the pIRES2-EGFP vector (Clontech). A GFP-IRES- Zeo^R coding sequence was cloned into the modified pcDNA3.1(+)/Zeo vector treated with *Pvu*II (New England Biolabs, UK) to remove the F1 origin of replication and the zeocin selection marker gene. Plasmid vectors based on pGIZ were constructed, and E77, or partial sequences of this fragment, were inserted into these vectors. E77, the isolated genomic DNA fragment, was inserted into the *Bam*HI site located upstream of the CMV promoter (depicted as E77), the *Pvu*II site located downstream of the polyadenylation signal (depicted as 3' E77), or within both the 5'- and 3'-flanking regions (depicted as dual E77). As the E77 element consisted of two distinct fragments derived from different locations in the CHO genome shotgun sequence, a reverse-oriented E77 sequence (E77r), a forward sequence of E77 (E77-t1, 1–1537 bp of E77), and a backward sequence of E77 (E77-t2, 1533–2894 bp of E77) were inserted into the *Bam*HI site located upstream of the CMV promoter. Truncated sequences (1–1148 bp, 1–857 bp, and 857–1537 bp) and the extended sequence of E77-t1 (-359–1537 bp) were also cloned into the pGIZ vector.

2.2 Cell culture and transfection

CHO-K1 cells (CRL-61; American Type Culture Collection (ATCC), USA) were cultured in Roswell Park Memorial Institute (RPMI; Gibco, Carlsbad, CA, USA) medium supplemented with 10% fetal bovine serum (FBS; Gibco) for adherent culture, or in SFM4CHO (HyClone, Logan, Utah, USA) medium supplemented with 4 mM glutamine (HyClone) for suspension culture at 37°C with 5% CO$_2$. The cells were transfected with plasmid vectors using Lipofectamine^{TM} 2000 (Invitrogen) according to the manufacturer's instructions. After two days, the cells were cultured in the presence of 500 µg/mL zeocin (Invitrogen) for two weeks to generate stable cell pools.

GFP-expressing individual cells were generated by transfection of E77-containing and non-E77 control vectors into the CHO-K1 cells. Then, transfected cells were inoculated (50 cells per well) in 96-well plates under zeocin

selection pressure. Single clones were grown in 96-well plates for three weeks and expanded for further analysis to determine fluorescence intensity.

2.3 Library construction and DNA fragment recovery

The genomic DNA library was constructed by *Sau*3AI digestion of CHO-K1 genomic DNA and the pcDNA/GFP/Zeo vector. Genomic DNA was isolated from the CHO-K1 cell line using the DNeasy® Blood & Tissue Kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions, and fractionated by *Sau*3AI digestion at various enzyme concentrations, times, and temperatures. Genomic DNA was treated with a serial dilution of *Sau*3AI from 2 units/µg of genomic DNA, and incubated at 37°C for 0.5, 1, 3, and 6 h, at 25°C for 1, 3, 6 and 12 h, at 16°C for 6, 12 and 24 h, and at 4°C for 12 and 24 h. After digestion, fragments were concentrated by ethanol precipitation and loaded onto 1% agarose gel to extract the size-fractionated genomic DNA. The 1–4 kilobase (kb)-sized DNA fragments were extracted from the gel using QIAquick Gel Extraction kit (Qiagen) and ligated into the *Bam*HI site of the pcDNA/GFP/Zeo vector, which was treated with calf intestinal alkaline phosphatase (Takara, Kyoto, Japan). After transformation of the ligation mixture into *Escherichia coli* (DH5α), about 200 000 colonies were generated on LB/ampicillin agar plates. Colonies on the plates were collected with LB/ampicillin medium and cultured for 3 h at 37°C with shaking. The genomic library was obtained by subsequent plasmid preparation using the EndoFree® Plasmid Purification kit (Qiagen).

The CHO-K1 cells were transfected with the genomic DNA library and cultured under zeocin selection pressure, and CHO-K1 cells transfected with the pcDNA/GFP/Zeo vector were used as controls. The selected cells were maintained and analyzed by FACSAria™ (BD Biosciences, San Jose, CA, USA) to compare GFP expression levels of the control and library groups. The top 0.1% of the cells in the library groups exhibited higher fluorescence intensity than cells in the control group. Approximately 8000 cells were sorted to obtain individual cells. The isolated cells were cultivated in 96-well plates by seeding one cell per well, and 220 single clones were expanded in 24-well plates. Finally, 97 clones were selected by measuring the fluorescence intensity and cell density and subjected to further analysis.

The inserted genomic DNA fragments in the selected cells were isolated by genomic DNA extraction and PCR amplification using the following primers: 5'-CAATT-GCATGAAGAATCTGC-3' and 5'-CCGTCATTGACGT-CAATAGG-3'. The primer binding sites were located at 161–180 bases and 501–520 bases in the pcDNA3.1(+)/Zeo vector in order to distinguish between self-ligation and insertions. The amplified sequences were analyzed by gel electrophoresis and sequencing. The isolated

genomic DNA fragments were cloned into the *Bam*HI site of the pcDNA/GFP/Zeo vector using the In-Fusion® HD cloning kit (Clontech) according to the manufacturer's instructions.

2.4 Quantification of transgene expression

The pGIZ constructs with the inserted sequences were transfected into CHO-K1 cells and cultured under zeocin selection pressure. Stable cell pools were maintained every three or four days, and the fluorescence intensity of cells was analyzed with the Guava easyCyte™ HT system (Millipore, Darmstadt, Germany) using the green fluorescence channel. After eight weeks, cells were grown in culture medium without zeocin selection pressure to investigate the stability of the transgene. GFP expression levels were determined by measuring the median values of fluorescence intensity. For statistical significance, three independent biological samples in each group were analyzed.

3 Results

3.1 A novel regulatory DNA element was identified in CHO-K1 genomic DNA

The genomic library was constructed by enzymatic fragmentation of genomic DNA and the pcDNA/GFP/Zeo vector containing a CMV promoter-driven GFP reporter gene. Genomic DNA isolated from CHO-K1 cells was fragmented by *Sau*3AI under varying reaction times, temperatures, and enzyme concentrations. A *Bam*HI restriction site was placed upstream of the CMV promoter to enable insertion of the genomic DNA fragments into the vector, exploiting the similar sticky ends (GATC) of *Bam*HI and *Sau*3AI. The genomic DNA fragments and vector were ligated and transformed into *E. coli* to generate the CHO-K1 genomic DNA library. CHO-K1 cells were transfected with the constructed library and cultured under zeocin selection pressure to obtain stable cells (Fig. 1A). After selection, the control and library cell pools were analyzed by FACS to determine GFP expression. The top 0.1% of cells harboring genomic DNA showed higher fluorescence intensity compared with the control. Approximately 8000 cells were isolated by FACS and inoculated into 96-well plates to obtain individual clones. Subsequent selection was carried out by measuring the fluorescence intensity and cell density. Finally, 97 clones were selected for isolation of novel regulatory elements. Flow cytometry was performed in order to identify a clone with high fluorescence intensity and unimodal distribution in the selected cell population. Among the 97 clones, 17 clones were found to exhibit a broad range or bimodal distribution of GFP expression and were therefore excluded from further analysis. Genomic DNA was then isolated
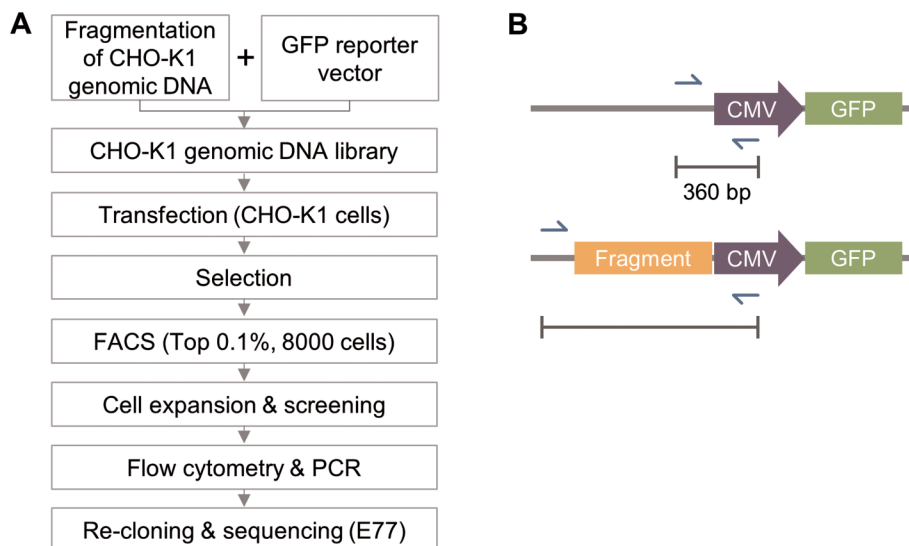
Figure 1. Isolation of a novel regulatory element from CHO-K1 cells. (**A**) A schematic view of the screening procedure for identification of a novel regulatory DNA element in CHO-K1 genomic DNA. (**B**) Recovery of integrated DNA fragments. "CMV" and "GFP" represent the human CMV promoter and a GFP-coding region in pcDNA/GFP/Zeo vector, respectively. Blue arrows indicate the binding sites of primers used for PCR to recover DNA fragments of vectors integrated into cells, and "360 bp" represents the PCR amplicon size for PCR performed on genomic DNA from cells containing the control vector pcDNA/GFP/Zeo. The PCR product obtained by amplification of genomic DNA from cells harboring the E77-containing vector is larger than that derived from cells with the control vector when the former template contains the unknown DNA fragment.

and PCR was performed to obtain the integrated DNA fragment from the selected clones. PCR primers were designed to distinguish between the fragment-inserted and self-ligated vectors (Fig. 1B). Self-ligated vectors and clones with more than two PCR products were excluded from further analysis. Forty-seven clones with an amplicon size of 3 kb were found to have a similar DNA sequence, and flow cytometry data revealed that most of these exhibited a high GFP expression level and unimodal distribution. Finally, a single clone (designated clone 77) that harbored the 3 kb DNA fragment and exhibited the highest GFP expression level was selected.

3.2 E77-containing vectors enhanced transgene expression levels in stably transfected cells

E77, the sequence isolated from clone 77, was incorporated into the pcDNA/GFP/Zeo vector to investigate its effect on transgene expression levels. The E77-containing and non-E77 control vectors were transfected into CHO-K1 cells, and the cells were inoculated into 96-well plates to obtain stably transfected individual clones under zeocin selection pressure. After selection, single clones were expanded and analyzed to determine GFP expression levels. As shown in Fig. 2, the incorporation of E77 resulted in the generation of an increased number of GFP-positive clones with higher fluorescence intensity. The average median fluorescence intensity was found to increase by two-fold in all selected clones and three-fold in the top 50 clones with E77. The number of clones exhibiting increased GFP expression was two-fold higher in the

E77 group than in the control group. The number of cells in the stable polyclonal cell population with higher GFP fluorescence levels was found to increase following the incorporation of E77 into the GFP-IRES-Zeo^R sequence-containing pGIZ vector (Supporting information, Fig. S1). An E77-containing antibody expression vector confirmed the results obtained with the GFP expression vector (Supporting information, Fig. S2). The polyclonal cell
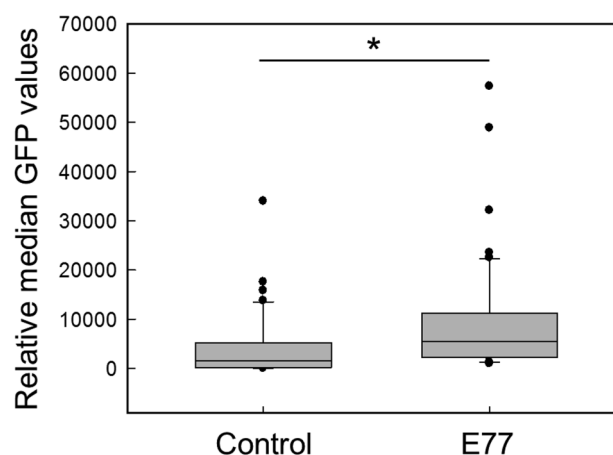

Figure 2. The relative change in GFP expression levels of individual clones. GFP-expressing cells were generated by transfection of E77-containing and non-E77 control vectors into CHO-K1 cells under zeocin selection pressure. Selected clones were then analyzed to evaluate GFP expression levels. The box plot represents the median fluorescence intensity and quartile distributions of 50 clones with high expression in each group. *p < 0.0001 (Student's t-test).

Biotechnology
Journal
www.biotechnology-journal.com

Biotechnol. J. 2016, 11, 633–641

BIO tec
visions
www.biotecvisions.com

population consisting of stable transfectants exhibited a five-fold increase in specific productivity and nine-fold in the percentage of cells expressing antibodies. Therefore, the use of E77-containing vectors was found to enhance transgene expression levels in stably transfected cells by generating an increased number of clones with high expression of the transgene.

3.3 Sequence analysis of E77

In order to identify the E77 regulatory element, the pcDNA/GFP/Zeo vector containing E77 was sequenced and analyzed using the basic local alignment search tool (BLAST; http://blast.ncbi.nlm.nih.gov) and CHO-blast (http://www.chogenome.org). The E77 element was found to be a 2894-bp fragment, originating from two distinct genes located at different positions in the CHO-K1 whole-genome shotgun sequence. As shown in Fig. 3, E77 was found to be composed of a forward sequence (1–1537 bp, designated E77-t1) and a backward sequence (1533–2894 bp, designated E77-t2), as a result of *Sau*3AI being located at 1532 bp of E77. The two separate *Sau*3AI-digested fragments may have combined to form a single fragment during ligation with the vector. Alignment of E77 sequence with the CHO-K1 reference sequence assembly (GCF_000223135.1) indicated a 95% match with NW_003619835.1 (scaffold16408) and NW_003615552.1 (scaffold6389), and 99% match with NW_003613979.1 (scaffold1936). The shotgun sequences were found to be unplaced genomic scaffolds and the regions displaying similarity with E77 were not annotated with functional features.

According to the NCBI nucleotide database, several regions of the forward sequence (2–1210 bp and 97–1158 bp) of E77 shared approximately 71% similarity with the beta-globin gene cluster of *Peromyscus* (deer mice) (EU559333.1, EU204642.1), and the regions displaying sequence similarity contained non-coding regions. The beta-globin gene cluster is known to contain regulatory DNA elements, which enhance transgene expression levels and inhibit gene silencing [14]. Although

the forward sequence of E77 was not found to be part of the beta-globin gene cluster of CHO-K1 cells, the extended sequence of E77-t1 (-359–1537), which was similar to AFTD01141250 (a contig of NW_003615552.1) and EU204642.1, increased the GFP expression level by 45% (Supporting information, Fig. S3). Additionally, we found that E77 contains cis-regulatory elements; in particular, three types of transcription factor regulatory elements (TFREs) including activator protein 1 (AP-1), enhancer box (E-box), and GATA-1 (Supporting information, Table S1) were identified.

3.4 E77 and its derivative sequences enhanced stable gene expression

MARs or insulators may be added to a vector in cis, either at the 5' end of the promoter or at the 3' end of the polyadenylation signal sequence, or both. The orientation of MAR has been investigated in order to elucidate its molecular characteristics [15, 16]. In order to examine the characteristic features of E77, the seven pGIZ constructs were transfected into CHO-K1 cells to investigate the effects of the element on GFP expression. Three independent stable transfectants were generated for each construct under selection pressure and the fluorescence intensity was measured. As shown in Fig. 4, cells transfected with E77-containing vectors showed statistically significant differences in GFP expression compared with the control cells transfected with the vector containing the CMV promoter only. The average median fluorescence intensity of GFP showed statistically significant increases of 50, 83, and 46% in cells harboring E77, 3' E77, and dual E77, respectively ($p < 0.05$). Although the GFP expression levels in the cells harboring E77r, E77-t1, and E77-t2 were not as high as in cells with E77, the average GFP median values of the former were also increased by 23, 34, and 19%, respectively.

Vectors harboring the UCOE and CMV promoter elicited substantial improvements in the GFP expression level in CHO-K1 cell lines, and the transgene expression levels remained high in stable cell pools for more than 100
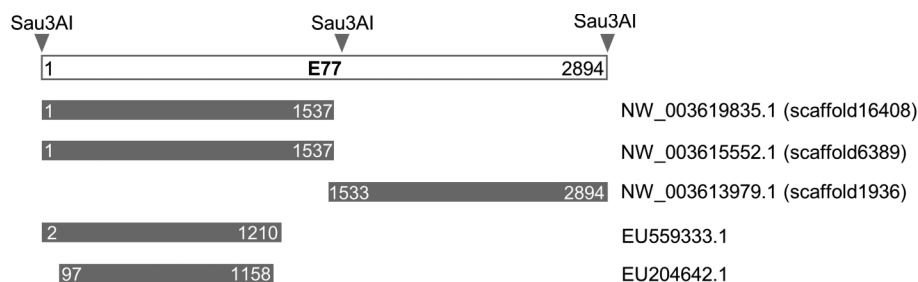

Figure 3. Sequence analysis of the E77 element. E77 was sequenced and analyzed using BLASTN 2.2.29+ [26]. "*Sau*3AI" represents the recognition sites for restriction enzyme *Sau*3AI. Numbers in boxes indicate the base pairs of E77. NW_003619835.1 (scaffold16408), NW_003615552.1 (scaffold6389), and NW_003613979.1 (scaffold1936) are the accession numbers from the CHOblast search (http://blast.chogenome.org/index.php). EU559333.1 and EU204642.1 are the accession numbers from nucleotide BLAST in NCBI (http://blast.ncbi.nlm.nih.gov).
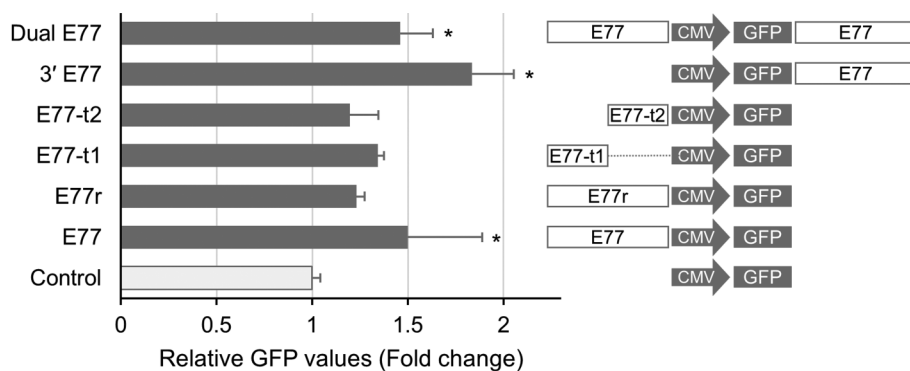
Figure 4. Characterization of E77. The relative GFP values represent the average fold change of median fluorescence intensity in stable cell populations at three weeks after selected cells were cultured under zeocin selection pressure. The right panel shows the vector constructs based on pGIZ, containing various insertions of E77 or its partial sequences. "GFP" represents the GFP-IRES-Zeo^R-polyadenylation signal sequence. Control, CMV promoter only; E77, the 5'-flanking region; E77r, the reverse orientation of E77; E77-t1 and E77-t2, partial sequences of E77; 3' E77, the 3'-flanking region; dual E77, both the 5'- and 3'-flanking regions. Bars represent the mean + standard deviation of three independent transfections. *$p < 0.05$ (One-way analysis of variance using the Holm–Sidak method).

generations [17]. In order to evaluate transgene stability over a prolonged period, stable cell populations at seven weeks were maintained for 13 additional weeks in the presence or absence of zeocin. As shown in Fig. 5A, when stable transfectants were generated with the vector containing the elements, the percentages of GFP-positive cells within the population was maintained, except for cells with E77-t1 which show a significant decrease in the percentage of GFP-positive cells on removal of selection pressure. Histograms obtained from flow cytometric analysis (Fig. 5B) clearly demonstrate the effects on transgene stability. The populations of GFP-positive cells were comparable, regardless of selection pressure, after 20 weeks when E77 was incorporated into the expression vector. E77-t2 was also capable of sustaining transgene expression in the absence of selection pressure. These results demonstrate that it is possible to enhance stable gene expression in CHO cells by transfecting these cells with vectors harboring E77 and its derivative forms.
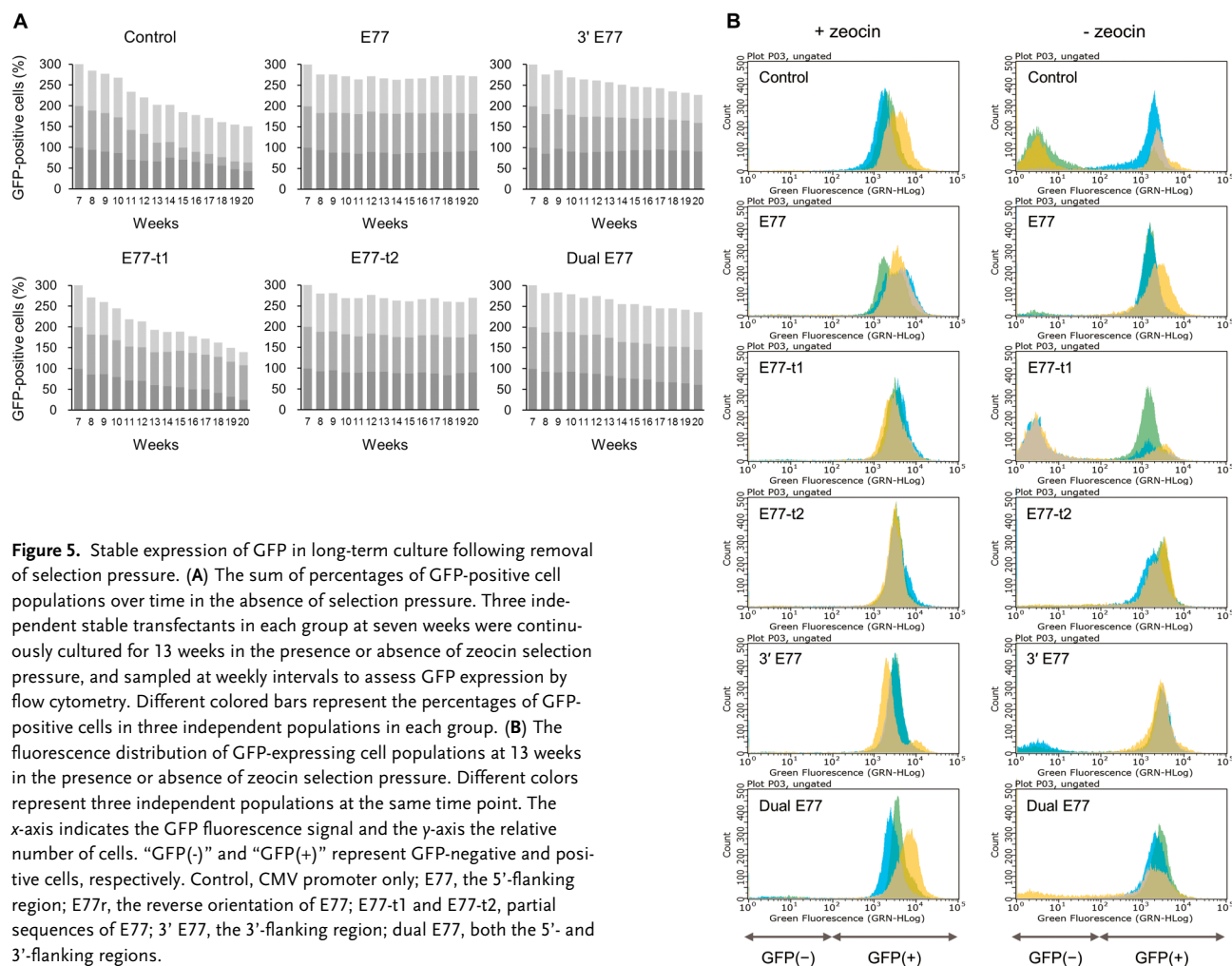
4 Discussion

4.1 Enhancement of stable transgene expression by regulatory DNA elements

In this study, we identified a novel gene regulatory element that combines with the CMV promoter to improve the stable expression of recombinant proteins in CHO cells. Numerous regulatory DNA elements of 1.5–4 kb in size, such as human beta-globin MAR (3 kb, [18]), chicken lysozyme MAR (2960 bp, [8]), human MAR X-29 (3492 bp, [9]), UCOE (minimal 1.5 kb, [17]), and EASE (3.6 kb, [11]), have been identified using GFP reporter plasmid systems with viral promoters, such as those of CMV and SV40. Such GFP-based vector systems enable simple and facile technique identification of regulatory DNA elements

[8–10, 17]. Accordingly, DNA fragments of 1–4 kb in size derived from CHO-K1 genomic DNA were incorporated into a GFP reporter vector harboring the CMV promoter to construct a genomic library for screening. Stably transfected cells harboring the genomic library were found to display higher GFP expression levels compared with cells containing the control vector. A 3-kb DNA fragment was obtained from the clone with the greatest fluorescence intensity by screening.

When this DNA fragment was used to flank a transgene, we observed a substantial increase in stable transgene expression compared with that in a control plasmid harboring the CMV promoter only. Incorporating E77 into the GFP expression vector resulted in enhancement of the median expression level by approximately two-fold compared with the control and an increase in the number of individual clones exhibiting higher GFP expression. The increased number of individual clones with higher expression improved the median expression level of the selected cell population. It was previously demonstrated that the frequency of individual clones with higher expression levels of beta-galactosidase or GFP was increased by incorporating human beta-globin MAR or UCOE into the vector, respectively, thereby increasing the expression levels of the entire cell population [17, 18]. The pGIZ vector containing E77 also conferred an increase in the GFP expression in the polyclonal stable cell population. The pGIZ vector which includes the GFP-IRES-Zeo^R sequence, was constructed to examine the effect of E77 on GFP expression levels in stable polyclonal cell populations more accurately. The specific productivity of monoclonal antibodies was five-fold higher in stable transfectants carrying the vector containing E77 than in those harboring the control. The generation of highly productive mammalian cell lines is usually based on stable expression systems in which target gene expression is affected by the chromosomal integration site in the host

Biotechnology
Journal
www.biotechnology-journal.com

Biotechnol. J. 2016, 11, 633–641

BIO tec
visions
www.biotecvisions.com

Figure 5. Stable expression of GFP in long-term culture following removal of selection pressure. (**A**) The sum of percentages of GFP-positive cell populations over time in the absence of selection pressure. Three independent stable transfectants in each group at seven weeks were continuously cultured for 13 weeks in the presence or absence of zeocin selection pressure, and sampled at weekly intervals to assess GFP expression by flow cytometry. Different colored bars represent the percentages of GFP-positive cells in three independent populations in each group. (**B**) The fluorescence distribution of GFP-expressing cell populations at 13 weeks in the presence or absence of zeocin selection pressure. Different colors represent three independent populations at the same time point. The x-axis indicates the GFP fluorescence signal and the y-axis the relative number of cells. "GFP(-)" and "GFP(+)" represent GFP-negative and positive cells, respectively. Control, CMV promoter only; E77, the 5'-flanking region; E77r, the reverse orientation of E77; E77-t1 and E77-t2, partial sequences of E77; 3' E77, the 3'-flanking region; dual E77, both the 5'- and 3'-flanking regions.

cell line. The utilization of viral promoters for cell line development is associated with transgene instability due to epigenetic events such as DNA methylation as well as genetic events such as the loss of copy numbers [19]. Therefore, various cis-regulatory elements have been identified that protect the transgene from position effects and improve its expression levels [20]. In our study, E77 was utilized in conjunction with the CMV promoter to confer transgene stability in long-term culture in the absence of selection pressure.

4.2 Characteristics of E77

E77 was found to be composed of two distinct sequences within different genomic scaffolds (Fig. 3), and to contain TFREs, such as AP-1, E-box, and GATA-1, which are present in viral promoters and known to be active in CHO cells [21]. Despite the presence of these elements, the sequences of E77 have not been annotated even 10 kb upstream or downstream of the mapped location. Interestingly, the forward sequence

of E77, E77-t1, shared approximately 70% sequence identity with the non-coding regions of the beta-globin gene cluster of *Peromyscus*, which belongs to the family Cricetidae of which the Chinese hamster is also a member. The expression level of GFP was slightly higher in the transfected cells harboring the E77-t1-containing vector, and significantly higher in cells with the vector containing the extended E77-t1 sequence, which is conserved in deer mice (Supporting information, Fig. S3). However, the rate of transgene silencing was varied in the subset of cells transfected with plasmids containing E77-t1. Although we have no direct evidence as to the basis of this phenomenon, we speculate that E77-t2 may contribute to the maintenance of transgene stability in the E77 element. E77-t1 and E77-t2 were found to differ in terms of their base composition and TFRE content. The MAR element contains AT-rich sequence motifs and the topoisomerase consensus sequence [14]. The AT contents of E77-t1 and E77-t2 were 59.73 and 67.03%, respectively. E77-t2 was found to contain tandem repeats

Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2016, 11, 633–641

BIOtec
visions
www.biotecvisions.com

of GATA and three binding sites of GATA-1, which is a transcriptional activator with essential functions in chromatin remodeling and histone modification at DNase I-hypersensitive sites of the beta-globin locus control region [22]. It is therefore likely that E77 confers transgene stability due to the base composition of this element and the presence of specific TFREs in E77-t2. However, the enhanced rate of expression due to the two partial sequences was lower than that due to E77. This finding suggests that combining both fragments results in greater efficiency in terms of expression level and long-term transgene stability in the absence of selection pressure.

Regulatory elements, such as MARs, enhancers, and insulators, are known to be independent of orientation in expression vectors. Additionally, MARs may be cloned upstream or downstream of the expression cassette [16]. However, the level of expression of the gene of interest is dependent on the orientation of certain elements such as the UCOE [23], the ribosomal protein L32 element [24], and post-transcriptional regulatory elements [25]. The increased level of GFP expression was two-fold lower in E77r than in E77, suggesting that the function of E77 may be affected by its orientation. The 3' E77 showed higher expression in stable cell pools, suggesting that the incorporation of E77 into the expression cassette may be affected by its insertion site. However, the effectiveness of E77 and 3' E77 was comparable in terms of transgene expression level and stability.

In conclusion, a novel gene regulatory DNA element designated E77 was isolated from the CHO-K1 cell line by genomic library-based screening. The E77 element was found to confer enhanced stable gene expression in conjunction with the CMV promoter. The effectiveness of E77 was due to the combination of two distinct genomic sequences located in different CHO genomic scaffolds. The characteristic features of E77 are related to its constitution, orientation, and insertion sites in the vector construct. The E77 regulatory element is expected to be utilized to generate stable CHO cell lines with high expression of recombinant proteins for potential application in the production of biopharmaceutical compounds.

5 References

[1] Walsh, G., Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* 2014, *32*, 992–1000.

[2] Kim, J. Y., Kim, Y. G., Lee, G. M., CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Appl. Microbiol. Biotechnol.* 2012, *93*, 917–930.

[3] Kildegaard, H. F., Baycin-Hizal, D., Lewis, N. E., Betenbaugh, M. J., The emerging CHO systems biology era: Harnessing the 'omics revolution for biotechnology. *Curr. Opin. Biotechnol.* 2013, *24*, 1102–1107.

[4] Lai, T., Yang, Y., Ng, S. K., Advances in mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals (Basel)* 2013, *6*, 579–603.

[5] Zhu, J., Mammalian cell protein expression for biopharmaceutical production. *Biotechnol. Adv.* 2012, *30*, 1158–1170.

[6] Huang, Y., Li, Y., Wang, Y. G., Gu, X. et al., An efficient and targeted gene integration system for high-level antibody expression. *J. Immunol. Methods* 2007, *322*, 28–39.

[7] Cacciatore, J. J., Chasin, L. A., Leonard, E. F., Gene amplification and vector engineering to achieve rapid and high-level therapeutic protein production using the Dhfr-based CHO cell selection system. *Biotechnol. Adv.* 2010, *28*, 673–681.

[8] Girod, P. A., Zahn-Zabal, M., Mermod, N., Use of the chicken lysozyme 5' matrix attachment region to generate high producer CHO cell lines. *Biotechnol. Bioeng.* 2005, *91*, 1–11.

[9] Girod, P. A., Nguyen, D. Q., Calabrese, D., Puttini, S. et al., Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nat. Methods* 2007, *4*, 747–753.

[10] Kwaks, T. H., Barnett, P., Hemrika, W., Siersma, T. et al., Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. *Nat. Biotechnol.* 2003, *21*, 553–558.

[11] Aldrich, T. L., Viaje, A., Morris, A. E., EASE vectors for rapid stable expression of recombinant antibodies. *Biotechnol. Progr.* 2003, *19*, 1433–1438.

[12] Pontiller, J., Gross, S., Thaisuchat, H., Hesse, F., Ernst, W., Identification of CHO endogenous promoter elements based on a genomic library approach. *Mol. Biotechnol.* 2008, *39*, 135–139.

[13] Pontiller, J., Maccani, A., Baumann, M., Klancnik, I., Ernst, W., Identification of CHO endogenous gene regulatory elements. *Mol. Biotechnol.* 2010, *45*, 235–240.

[14] Harraghy, N., Gaussin, A., Mermod, M., Sustained transgene expression using MAR elements. *Curr. Gene Ther.* 2008, *8*, 353–366.

[15] Arope, S., Harraghy, N., Pjanic, M., Mermod, N., Molecular characterization of a human matrix attachment region epigenetic regulator. *PLoS One* 2013, *8*, e79262.

[16] Harraghy, N., Buceta, M., Regamey, A., Girod, P. A., Mermod, N., Using matrix attachment regions to improve recombinant protein production. *Methods Mol. Biol.* 2012, *801*, 93–110.

[17] Williams, S., Mustoe, T., Mulcahy, T., Griffiths, M. et al., CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. *BMC Biotechnol.* 2005, *5*, 17.

[18] Kim, J.-M., Kim, J.-S., Park, D.-H., Kang, H. S. et al., Improved recombinant gene expression in CHO cells using matrix attachment regions. *J. Biotechnol.* 2004, *107*, 95–105.

[19] Kim, M., O'Callaghan, P. M., Droms, K. A., James, D. C., A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol. Bioeng.* 2011, *108*, 2434–2446.

[20] Palazzoli, F., Bire, S., Bigot, Y., Bonnin-Rouleux, F., Landscape of chromatin control element patents: Positioning effects in pharmaceutical bioproduction. *Nat. Biotechnol.* 2011, *29*, 593–597.

Biotechnology
Journal
www.biotechnology-journal.com

Biotechnol. J. 2016, 11, 633–641

BIO tec
visions
www.biotecvisions.com

[21] Brown, A. J., Sweeney, B., Mainwaring, D. O., James, D. C., Synthetic promoters for CHO cell engineering. *Biotechnol. Bioeng.* 2014, *111*, 1638–1647.

[22] Cho, Y., Song, S. H., Lee, J. J., Choi, N. et al., The role of transcriptional activator GATA-1 at human beta-globin HS2. *Nucleic Acids Res.* 2008, *36*, 4521–4528.

[23] Zhang, F., Frost, A. R., Blundell, M. P., Bales, O. et al., A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors. *Mol. Ther.* 2010, *18*, 1640–1649.

[24] Hoeksema, F., Hamer, K., Siep, M., Verhees, J. A., Otte, A. P., Placing the RPL32 promoter upstream of a second promoter results in a strongly increased number of stably transfected mammalian cell lines that display high protein expression levels. *Biotechnol. Res. Int.* 2011, *2011*, 492875.

[25] Mariati, Ho, S. C., Yap, M. G., Yang, Y., Post-transcriptional regulatory elements for enhancing transient gene expression levels in mammalian cells. *Methods Mol. Biol.* 2012, *801*, 125–135.

[26] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J. et al., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 1997, *25*, 3389–3402.
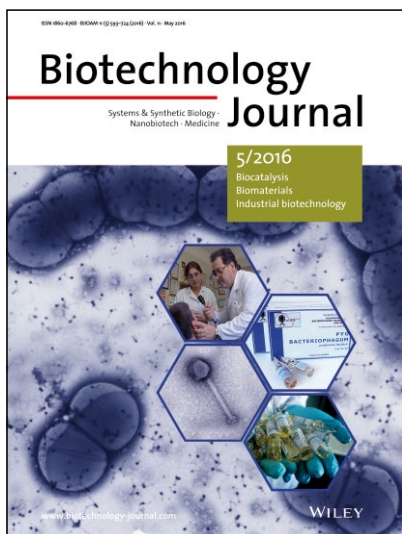
Biotechnology Journal – list of articles published in the May 2016 issue.