Employing epigenetics to augment protein expression in mammalian cells
Kwaks, T.H.J.

Citation for published version (APA):
A novel, high stringency selection system allows screening of few clones for high protein expression


1 These authors contributed equally to this work.

a ChromaGenics, Kruislaan 406, 1098 SM Amsterdam, The Netherlands
b Crucell, Archimedesweg 4, 2333 CN Leiden, The Netherlands
c Swammerdam Institute for Life Sciences, University of Amsterdam, Kruislaan 406, 1098 SM Amsterdam, The Netherlands
* Corresponding author at: Swammerdam Institute for Life Sciences, University of Amsterdam, Kruislaan 406, 1098 SM Amsterdam, The Netherlands. Tel.: +31 20 525 5115; fax: +31 20 525 5090.
E-mail address: arie.otte@science.uva.nl (A.P. Otte).

Abstract

To obtain highly productive mammalian cell lines, often large numbers of clones need to be screened. This is largely due to low selection stringencies, creating many, but low protein producing clones. To remedy this problem, a novel, very stringent selection system was designed, to create few, but high protein producing clones. In essence, a selection marker with a start codon that confers attenuated translation initiation frequency was placed upstream of the gene of interest with a start codon that confers optimal translation initiation. From the transcribed bicistronic mRNA, the selection marker is translated at a low frequency, and the protein of interest at a high frequency. This selection system is so stringent that clones form only rarely. However, application of anti-repressor elements, which increase promoter activity, did induce the formation of clones that expressed proteins at high levels. When combined with anti-repressor elements, this novel selection system can be a valuable tool to rapidly create few, but highly productive mammalian cell lines.

**Keywords:** Selection system; Mammalian cell-lines; Anti-repressor elements
Introduction

A major effort is often required to create mammalian cell lines that express a therapeutic protein at high level\(^1\). An important reason is that upon transfection, large numbers of clones are generated. However, the expression of the therapeutic protein in the majority of the resulting cell lines is often low\(^2,3\). Therefore, many clones need to be screened to obtain a highly productive producing clone. We encountered this problem when we identified and applied genetic elements that counteract chromatin-mediated gene repression. These anti-repressor elements induced higher protein expression levels, but they also induced up to 10-fold more clones\(^4\). Raising the stringency of selection often helps to reduce the number of generated clones and thereby increases the chance to identify a high protein-producing clone\(^3\). This can be achieved by increasing the concentration of selection agent in the culture medium, but due to toxicity of the selection agent, this approach has its limits. Another approach is to modify the selection system itself. We previously raised the stringency of the selection system by placing an attenuated IRES (Internal Ribosome Entry Site) sequence between the expression unit of the gene of interest and the Zeocin resistance gene\(^4,5\). This indeed reduced the number of clones as well as increased the protein expression levels in the resulting cell lines\(^4,5\). To elaborate on this finding, we designed a novel selection system. Our goal was to further reduce the number of generated clones as well as to obtain higher protein expression levels than we previously achieved with the selection system based on an IRES sequence.

Results

**Design of a novel selection system**

According to the scanning model of translation initiator in eukaryotes the small ribosomal subunit enters the mRNA at the 5’ end and scans the mRNA in the 5–3’ direction until it encounters a start codon\(^6\). We employed this mechanism by placing the Zeocin resistance gene upstream of a gene of interest (d2EGFP), which results in the generator of a bicistronic mRNA (Fig. 1A, upper panel). d2EGFP was cloned immediately downstream of the stop codon of the Zeocin gene and in a different open reading frame to avoid the putative formation of a Zeocin-d2EGFP fusion protein. To avoid potential internal
translation initiation in the coding Zeocin resistance mRNA, all internal AUG translation initiator codons were removed. To create an ATG-free Zeocin resistance gene, the only internal ATG coding for an internal methionine (at amino acid position 84) was replaced by a TTG, which replaced the methionine by leucine. This mutation did not change the activity of the Zeocin resistance protein (data not shown). In addition, we placed multiple stop codons behind the Zeocin gene to stop translation in all three reading frames (TGATTGATTGA; bold indicates the three stop codons, in three reading frames). This precludes that any unwanted initiation of translation within the Zeocin open reading frame results in a fusion product.

This approach resulted in a CMV promoter – ATG Zeo – ATG d2EGFP cassette. Translation is expected to initiate mainly at the start codon AUG (accAUGg in an optimal configuration) of the Zeocin resistance mRNA (accATGg in the corresponding Zeocin gene). Furthermore, translation is expected almost never to start at the AUG translation initiator codon of the downstream d2EGFP. As a result, high levels of Zeocin resistance protein will be translated, which will result in the formation of many clones. Since no translation of d2EGFP mRNA is expected to occur (Fig. 1A, upper panel), these cell lines will not express d2EGFP protein. Upon transfection of this construct, we indeed observed that many clones formed, but these clones did not express d2EGFP (Fig. 1B).

Next, we modified the Zeocin resistance gene by replacing the start codon AUG by a modified start codon UUG (accTTGg in the corresponding DNA) that conveys attenuated resistance.

Figure 1. The mechanistic basis and the application of the novel selection system.

(A) The Zeocin resistance gene was placed upstream of the d2EGFP gene. In the upper bar, both genes had an optimal translation initiation codon (accATGg), which becomes accAUGg in the correspondent mRNA. Translation (as indicated by the small and large ribosomal subunits) will start at the accAUGg of the Zeocin mRNA, theoretically riesling in high Zeocin resistance protein and many stably transfected cell clones that do not display d2EGFP expression (predicted in the right panel). In the lower bar, the translation initiation codon of the Zeocin resistance gene was changed to accTTGg (accUUGg in the corresponding mRNA). (B) As predicted in the right panel in (A), many clones were formed with AUG Zeo–d2EGFP, but these did not express d2EGFP. In contrast, only few clones were formed with UUG Zeo–d2EGFP, but these did express d2EGFP. (C) The upper panels are examples of d2EGFP fluorescence in clones that resulted from transfection with the a control TTG-Zeo-d2EGFP construct and the construct flanked with anti-repressor elements 7 and 67 upstream and anti-repressor element 7 downstream of the expression cassette. The phase contrast pictures of the same clones are shown in the middle panels. The lower panels show d2EGFP fluorescence histograms of cell lines that emerged from these clones. A line is arbitrarily drawn at the same d2EGFP signal level to facilitate comparisons. (D) CHO-K1 cells were transfected with the indicated d2EGFP construct and selected on Zeocin. The d2EGFP value in the individual clones and the mean d2EGFP fluorescence values in these clones are indicated. A comparison is made between clones resulting from transfection with d2EGFP constructs harboring the novel TTG Zeo-d2EGFP selection system, or the d2EGFP-IRES-Zeo selection system, with and without anti-repressor elements (represented by the blue/black ovals).
A novel, high stringency selection system
translation initiation frequency\textsuperscript{8,9}. This modified Zeocin resistance gene was again placed upstream of d2EGFP, which contains the optimal start codon accAUGg. These changes resulted in a CMV promoter–TTG Zeo–ATG d2EGFP cassette. Translation of the resulting bicistronic mRNA can be expected to only rarely initiate at the attenuated start codon accUUGg of Zeocin\textsuperscript{8,9} leading to low levels of Zeocin resistance protein (Fig. 1A, lower panel). With equal concentrations of Zeocin selection agent in the culture medium, low Zeocin resistance protein levels will result in massive cell death and the formation of only a few clones (Fig. 1A, lower panel). We indeed observed that transfection of the TTG Zeo-d2EGFP construct resulted in the formation of only a few clones (Fig. 1B). Furthermore, due to inefficient translation initiator at accUUGg of the Zeocin mRNA, the small ribosomal unit should most frequently skip this accUUGg. Instead the small ribosomal unit will scan until an optimal translation initiation codon is encountered, in this case the accAUGg codon of d2EGFP. If translation at the modified accUUGg codon of the Zeocin mRNA would be reduced to $x\%$, this would result in $100-x\%$ translation at the optimal accAUGg translation initiator codon of d2EGFP mRNA (Fig. 1A). We indeed observed that the obtained clones expressed d2EGFP protein (Fig. 1B).

Together these results show that placing a selection marker with an attenuated translation initiator codon upstream of a gene of interest with an optimal translation initiation codon, creates a selection system for generating stably transfected mammalian cell lines.

\textit{In the context of anti-repressor elements, the novel selection system induces high protein expression levels}

Multiple transfection experiments taught that the stringency of the novel selection system was so high that it prevented efficient formation of clones. The resulting cell lines expressed d2EGFP, but they grew poorly and the expression of d2EGFP was not uniformly distributed in the clone (Fig. 1C, left panels). The poor growth of these cell lines could result from too low expression levels of the Zeocin resistance protein to effectively neutralize the Zeocin in the culture medium. Previously we showed that incorporation of anti-repressor elements in a construct results in higher protein expression levels\textsuperscript{4}. We therefore incorporated anti-repressor elements in the construct. This in turn might result in higher levels of Zeocin resistance protein thereby increasing the chance of cell survival and thus the formation of more colonies. We placed two anti-repressor elements upstream of the CMV promoter–TTG Zeo–d2EGFP cassette and one element downstream of the cassette (Fig. 1 C and D). We used the previously described anti-repressor element 7\textsuperscript{4} and a novel
element 67. Incorporation of the combination of these two anti-repressor element resulted in higher protein expression levels than with a single anti-repressor element (data not shown). Transfection of the anti-repressor shielded TTG Zeo-d2EGFP construct and selection with similar Zeocin concentrations indeed resulted in an increase of the number of clones to 73, as compared to 14 clones obtained with the control TTG Zeo-d2EGFP construct (Table 1). Analysis of 24 randomly isolated clones with the anti-repressor element shielded TTG Zeo-d2EGFP construct also showed that the mean d2EGFP fluorescence signal increased to 889, as compared to 168 with the TTG Zeo-d2EGFP control construct (all 14 clones analyzed). The peak fluorescence signal increased from 300 to 1878, respectively (Fig. 1D). Furthermore, in contrast with the TTG Zeo-d2EGFP control clones, the anti-repressor element shielded TTG Zeo-d2EGFP construct clones grew well and displayed a uniform distribution of d2EGFP expression, resulting in rather sharp fluorescence profiles in the cell lines (Fig. 1C).

The results show that in the context of anti-repressor elements the novel selection system induces efficient establishment of cell lines that express high protein levels.

The novel selection system is very stringent

Our previous findings\textsuperscript{4,5} that application of anti-repressor elements results in an increased number of clones as well as higher protein expression levels were obtained in the context of a very high stringency selection system. In this system, a selection marker is placed downstream of a gene of interest, coupled through an attenuated IRES sequence\textsuperscript{4,5}. We tested the differences between the d2EGFP-IRES-Zeo and the novel TTG Zeo-d2EGFP configurations. We monitored the number of generated clones as well as protein expression levels in the resulting cell lines. Equal amounts of DNA were transfected and selection took place with equal Zeocin concentrations in the medium. We found that the anti-repressor element shielded TTG Zeo-d2EGFP construct generated fewer clones (73), as compared to the anti-repressor element shielded d2EGFP-IRES-Zeo construct (400) (Table 1). Fewer clones were also generated with the unshielded TTG Zeo-d2EGFP construct (14), as compared to the unshielded d2EGFP-IRES-Zeo construct (150) (Table 1). Importantly, the d2EGFP expression levels were greatly increased, from a mean d2EGFP fluorescence signal of 151 with the anti-repressor element shielded d2EGFP-IRES-Zeo construct to a mean signal of 889 with the anti-repressor element shielded TTG Zeo-d2EGFP construct (Fig. 1D). The peak fluorescence signal increased from 330 to 1878, respectively (Fig. 1D). These results show that the novel TTG Zeo selection system is indeed more stringent than
The IRES-Zeo selection system. It also indicates that using the novel selection system in the context of anti-repressor elements, only few clones need to be generated and analyzed.

The novel selection system allows low copy numbers and high expression per copy

The observed higher d2EGFP expression levels in clones generated with the novel selection system could be due to incorporation of more gene copies. Hence, we analyzed the number of the integrated d2EGFP gene copies in relation to d2EGFP protein expression levels. For this analysis, we used 11 of the above-described TTG Zeo-d2EGFP clones (three very low expressing clones died in course of the analysis) and the 24 anti-repressor element shielded TTG Zeo-d2EGFP clones (Fig. 1D). We found that the copy number in case of the anti-repressor shielded clones was less than 40, which is fairly low when compared to for instance dhfr-mediated gene amplification\(^{10,11}\), which can result in many hundreds to thousands of copies. A good correlation was found between the copy number of the anti-repressor element shielded TTG Zeo-d2EGFP construct and d2EGFP fluorescence signal, as indicated by the high coefficient of determination (\(R^2 = 0.799\)). In contrast, a poor correlation was found between the copy number of the TTG Zeo-d2EGFP control construct and the d2EGFP fluorescence signal, as illustrated by the coefficient of determination (\(R^2 = 0.202\)) (Fig. 2A). Furthermore, the d2EGFP fluorescence signal per copy increased from \(11.4 \pm 4.3\) with the TTG Zeo-d2EGFP control construct to \(51.6 \pm 4.8\) with the anti-repressor element shielded TTG Zeo-d2EGFP construct (Fig. 2B).
Fig. 2. Anti-repressor elements convey higher expression per integrated gene copy, in a copy number dependent fashion.

Genomic DNA was isolated from indicated colonies that were stably transfected with d2EGFP constructs with or without anti-repressor element (represented by the blue/black ovals). The copy number of the d2EGFP construct was determined using real-time PCR. The copy number was plotted against the d2EGFP expression levels and the coefficients of determination were added (A). The lower panel (B) shows that the average d2EGFP expression per copy (indicated ±S.E.M.) was significantly ($P < 0.0001$) increased in clones that were transfected with anti-repressor-shielded constructs.
In conclusion, the d2EGFP expression observed with the novel anti-repressor element shielded TTG Zeo-d2EGFP selection system is not due to the introduction of many more gene copies. Instead, incorporation of anti-repressor elements in the TTG Zeo-d2EGFP construct results in more d2EGFP expression per copy as well as copy number dependent d2EGFP expression.

**Fig. 3. Application of the novel selection system induces clones with high hEPO expression.**

CHO-K1 cells were transfected with the indicated EPO constructs and selected on Zeocin (blue/black ovals represent anti-repressor elements). Up to 24 independent clones were isolated and propagated before secreted EPO expression levels were determined. The EPO values in the individual clones (indicated in pg/cell/day) and the mean EPO expression levels in these clones are indicated. A comparison is made between clones transfected with the novel TTG Zeo-EPO selection constructs and the EPO-IRES-Zeo constructs, in either case with and without anti-repressor elements. The EPO expression levels with the anti-repressor-shielded TTG Zeo-EPO construct were statistically higher ($P < 0.0001$) than the EPO expression levels with any of the other constructs. Also the EPO expression levels with the anti-repressor-shielded EPO-IRES-Zeo construct were statistically higher ($P < 0.0001$) than the EPO expression levels with the non-shielded EPO-IRES-Zeo construct.
The novel selection system operates well with a secreted, therapeutic protein

To more directly assess the value of the novel selection system we extended our analysis to a well-known commercially important protein. We cloned the human EPO (erythropoietin) gene downstream of the TTG Zeo selection marker. We found that the anti-repressor shielded EPO-IRES-Zeo construct generated more clones than the control EPO-IRES-Zeo construct (300 versus 164, respectively) (Table 1). The anti-repressor element shielded TTG Zeo EPO construct generated more clones than the control TTG Zeo EPO construct (41 versus 5, respectively) (Table 1). Indicating the higher selection stringency, the anti-repressor shielded TTG Zeo-EPO construct generated fewer clones than the anti-repressor shielded EPO-IRES-Zeo construct (41 versus 300, respectively) (Table 1).

Mean EPO expression levels increased from 3.3 pg/cell/day with the TTG Zeo-EPO control construct, to 17.7 pg/cell/day with the anti-repressor element shielded TTG Zeo-EPO construct. The peak EPO expression level increased from 5.5 to 28.3 pg/cell/day, respectively (Fig. 3). In comparison, with the IRES selection system, the mean EPO expression levels increased from 5.8 pg/cell/day with the control EPO-IRES-Zeo, to 9.0 pg/cell/day with the anti-repressor element shielded EPO-IRES-Zeo construct (Fig. 3). These obtained EPO expression levels are without gene amplification. In comparison, reported EPO values range from 12 to 47 pg/cell/day, but these values were all achieved after multiple rounds of dhfr-mediated gene amplification. Our result shows that the novel selection system can readily be applied for the production of important therapeutic proteins, such as EPO.

Discussion

We report a novel, high stringency selection system for the creation of stably transfected mammalian cell lines. In short, a selection marker with a start codon that confers attenuated translation initiation is placed upstream of the gene of interest with a start codon that confers optimal translation initiation. Thereby this system utilizes well-established rules of translation initiation. We show that this selection system is more stringent than another, available stringent selection system in which the selection marker is coupled through an IRES sequence. We compared these systems using both the d2EGFP protein and a secreted, commercially relevant protein, EPO.

A novel, high stringency selection system
Our novel selection system is so stringent that only few clones are formed, unless previously identified anti-repressor elements that increase protein expression levels are applied\(^4\). In this manner, the selection system optimally exploits the expression augmenting characteristics of these anti-repressor elements. One of the previously observed characteristics of the anti-repressor elements was that they convey copy number dependent gene expression. This characteristic appears to be retained with the novel selection system. Whereas some clones harbor only few (\(< 4\)) copies (Fig. 2A), the majority of the clones harbor between 20 and 30 copies. This may indicate that even with the application of the anti-repressor elements, the expression of the bicistronic mRNA, derived from only a few copies is too low to produce sufficient Zeocin resistance protein for survival of the cell. Although the \(d2\)EGFP expression per copy is elevated in clones that contain anti-repressor element shielded constructs (Fig. 2B), there is apparently a limit to the maximum \(d2\)EGFP expression levels that can be reached in low copy clones. Whether this indicates that the constructs are integrated in transcriptionally rich or ‘open’ regions of the genome cannot be concluded from these results. Since the ranges in expression per copy are small (Fig. 2B), one might argue that integration is rather independent of the position in the genome. On the other hand, if the anti-repressor elements target the plasmids to a transcriptionally rich region, this would also result in higher expression per copy. Answering these questions clearly requires more fundamental insight into the molecular mechanisms underlying anti-repressor element action.

The strategy to modify selection markers is not restricted to the Zeocin resistance gene, but can also be extended to other selection markers, such as blasticidin, neomycin and puromycin (data not shown). It should be pointed out that also with these selection markers all the internal ATGs have to be removed from the genes; four, fifteen and three for blasticidin, neomycin and puromycin, respectively. Furthermore, the selection system operates well in other cell lines than CHO-K1, such as CHO-DG44, with other promoters, such as the SV40, EF1\(\alpha\) and actin promoters, and we have successfully applied the system to express more complex proteins like monoclonal IgG antibodies (data not shown).

Several other strategies have been followed to achieve high selection stringencies. Reff changed the context around the ATG startcodon of the neomycin resistance gene\(^15\). The mutated neo gene was placed in a separate expression unit, downstream of an independent promoter, unlike in our system where the modified selection gene is under control of the same promoter as the gene of interest. An alternative system described by Lucas et al. places a selection marker in an intron\(^16\). Efficient splicing of this intron warrants that high
expression levels of the RNA are required to obtain sufficient selection marker protein. Finally, the open reading frame of the neomycin selection gene has been mutated to obtain a less functional protein, requiring higher expression levels of such mutated neo protein. This also results in higher expression levels of the protein of interest\textsuperscript{17,18}. Here we compared our novel selection system with the system that places a selection marker downstream of a gene of interest, coupled through an attenuated IRES sequence. We have shown that the selection system we report is more stringent than the ‘IRES’ system. We do not know how our system compares with the other above mentioned systems, in terms of selection stringency. However, given that we hardly get any colony formation in the absence of anti-repressor elements indicates that our novel system is very stringent. All above-mentioned selection systems were designed to operate without application of anti-repressor elements, hence this might indicate that these systems are less stringent than our novel selection system.

At present, many hundreds to thousands of stably transfected clones need to be screened to establish a few high protein-producing cell lines. The novel selection system greatly reduces this number of clones, without the need to add more selection agent to the culture medium, which could result in unwanted side effects of toxicity. Instead the system employs relatively simple modifications of the selection markers themselves. Combining the novel selection system with anti-repressor elements thus allows rapid generation of only a limited number of clones. Subsequently, cell lines can be established that produce therapeutic proteins at higher levels than can be achieved with current selection systems. Particularly in an industry where speed to create for instance phase I clinical trial material is of utmost importance, the reported selection system can make a difference. We are aware that the EPO results are obtained with adherent cells and serum containing culture medium. However, preliminary results indicate that adaptation to serum-free suspension medium does not result in significant changes in EPO expression levels. We therefore envision that this selection system, in combination with anti-repressor elements can become a valuable tool for the biotechnology industry.
Experimental protocol

Vector constructions
The TTG Zeo selection vector was constructed by using the primer TTGGATCGGATCCACCTTGGCCAAGTTGACCAGTGCCGTTC) to create the accUUGg translation initiation codon. Zeocin contains one ATG, which codes for methionine (aa 84) and to avoid internal initiation of translation at the AUG in the resulting mRNA, it was replaced by TTG, which codes for leucine. This was done by using the primer AGGGCCCCGCCCCACCGGCTGCTCGCCGATCTCGGTCAAGGCCGGC (this is a reverse primer, so CAA specifies the Leu) and PCR. The resulting Zeocin gene contained no in frame or out frame ATGs. The Zeocin and d2EGFP transcription units were placed behind the CMV promoter. Potentially, GTGs or CTGs in an optimal translation initiation context (ANNGTGN or GNNGTGG; ANNCTGN or GNNCTTG, respectively) could potentially serve as translation initiation startpoint. Since the Zeocin open reading frame contains three GTGs and five CTGs in an optimal translation initiation context, we mutated these sites to create a non-optimal translational initiation context. This will prevent putative premature translation initiation, upstream of the optimal ATG of the d2EGFP gene. We found, however, no differences in d2EGFP levels whether plasmids did, or did not contain GTG and CTG triplets with an optimal translation initiation context. We concluded that aberrant translation initiation within the Zeocin ORF does not affect the level of d2EGFP expression.

Transfection and analysis of clones
CHO-K1 cells were grown in HAM F12 medium (Invitrogen), supplemented with 9%FBS. 6×10^6 cells were transfected using Lipofectamine™ 2000 (Invitrogen) according to the manufactures protocol. We used non-linearized plasmid DNA, since we did not observe differences on the number of emerging colonies or on the d2EGFP expression levels when we compared linearized versus non-linearized DNA. One day after the transfection, cells were seeded to petri dishes and 150 μg/ml Zeocin selection agent was added to the culture medium. The cells are cultured for ~2 weeks, until stably transfected cell clones were formed. Clones were picked, transferred to six wells plates and expression of d2EGFP fluorescence levels were analyzed using flowcytometry (Beckman Coulter, EpicsXL). Four weeks after transfection, the Zeocin concentration was lowered to 25 μg/ml and protein expression levels remained stable over a period exceeding 3 months. This Zeocin
concentration is too low to kill wild-type CHO-K1 cells. However, in the absence of Zeocin, we observed a slow decrease in d2EGFP expression in the course of a 2 month period. For the present analysis we therefore kept the cells under a ‘maintenance’ concentration of 25 µg/ml Zeocin. With the EPO constructs we followed the same procedure. Clones were selected in the presence of 150 g/ml Zeocin selection agent, and ~4 weeks after transfection the Zeocin concentration was lowered to 25 µg/ml Zeocin. All reported d2EGFP and EPO expression data were obtained ~6 weeks after transfection.

**Determination of copy numbers**
Stable CHO-K1 clones transfected with TTG Zeo and anti-repressor element shielded TTG Zeo vectors were isolated and assayed for d2EGFP fluorescence levels. Genomic DNA was purified and the copy number of the \(d2EGFP\) gene was determined using realtime PCR. The ratio between the \(d2EGFP\) and \(actin\) signal was taken as the relative copy number. A serial dilution with a plasmid containing the \(d2EGFP\) gene in a background of genomic CHO-K1 DNA was used to determine the absolute copy number.

**Epo measurements**
Equal number of cells (100,000) were seeded and cultured in six-well dishes for 2 days before cells were counted and the medium was collected. The amount of secreted human recombinant erythropoietin was determined using an ELISA-kit (R&D systems).

**Statistical analysis**
Statistical analysis was performed using Microsoft Excel 2002 SP3 for Windows and Graphpad Prism 4 for Windows.

**Acknowledgements**
We thank Jaap Goudsmit, Menzo Havenga, Richard Verhage and David Satijn for critically reading the manuscript.
References
