Methods to create a stringent selection system for mammalian cell lines

**H. J. M. Van Blokland · F. Hoeksema ·
M. Siep · A. P. Otte · J. A. Verhees**

Abstract The efficient establishment of high protein producing recombinant mammalian cell lines is facilitated by the use of a stringent selection system. Here, we describe two methods to create a stringent selection system based on the Zeocin resistance marker. First, we cloned increasingly longer stretches of DNA, encoding a range of 8–131 amino acids immediately upstream of the Zeocin selection marker gene. The DNA stretches were separated from the open reading frame of the selection marker gene by a stopcodon. The idea behind this was that the translation machinery will first translate the small peptide, stop and then restart at the AUG of the Zeocin marker. This process, however, will become less efficient with increasingly longer stretches of DNA upstream of the Zeocin marker that has to be translated first. This would result in lower levels of the Zeocin selection marker protein and thus a higher selection stringency of the system. Secondly, we performed a genetic screen to identify PCR induced mutations in the Zeocin selection protein that functionally impair the selection marker protein. Both the insertion of increasingly longer peptides and several Zeocin selection protein mutants resulted in a decreasing number of stably transfected colonies that concomitantly displayed higher protein expression levels. When the Zeocin mutants were combined with very short small peptides (8–14 amino acids long), this created a flexible, high stringency selection system. The system allows the rapid establishment of few, but high protein producing mammalian cell lines.

Introduction

To create stably transfected mammalian cell lines an efficient selection system is required to (a) kill untransfected cells and (b) achieve high protein expression levels. In general, with a more stringent selection system low-expressing cells will be killed more efficiently, and the expression levels of the protein of interest will be higher. Also, more stringent selection systems usually create fewer colonies after transfection, but these colonies will display higher protein expression values.

Previously, we described a very high stringency selection system in which a selection marker was modified with startcodons that confer attenuated translation initiation frequency (Kozak 2001; Van Blokland et al. 2007). These codons are used for translation of the selection marker that is placed

H. J. M. Van Blokland · F. Hoeksema ·
M. Siep · A. P. Otte · J. A. Verhees (✉)
Swammerdam Institute for Life Sciences, University
of Amsterdam, Science Park 904, 1098 XH Amsterdam,
The Netherlands
e-mail: J.A.Verhees@uva.nl

upstream of a gene of interest that has a startcodon with optimal translation initiation properties. The modified startcodon can be a GTG or TTG instead of an ATG, resulting in selection markers such as GTG Zeocin (GTG Zeo), TTG Zeocin (TTG Zeo) or TTG Neomycin (TTG Neo). The selection marker is translated at a low frequency from the transcribed bicistronic mRNA due to the attenuated translation initiation codon. Thus, in order to produce enough selection marker protein for the cell to survive, high amounts of bicistronic mRNA levels have to be expressed. Since the bicistronic mRNA also encompasses the gene of interest, which has an optimal translation initiation codon, concomitantly high levels of the protein of interest will be produced.

One drawback of this selection system is that, in practice, there are in fact only two suitable modified translation initiation codons available, the GTG and TTG codons. Of these two, the use of the GTG codon in combination with the Zeocin selection marker created not a very stringent selection system (Van Blokland et al. 2007). In contrast, the TTG codon in combination with the Zeocin marker provided a selection stringency that operates very well in cell lines such as CHO-K1 or CHO-DG44, but with other cell lines, the TTG Zeocin variant proved to be too stringent. To circumvent this problem we attempted to create a different selection system that has more flexibility than only one available modified translation initiation codon. Here we describe a novel selection system that allows a gradual modulation of the selection stringency of the Zeocin marker and the establishment of a few cell lines that display high protein expression levels.

Materials and methods

Creation of the small open reading frames

ppLUC 8 and ppLUC 14 were made by annealing the following complementary primers:

    ppLUC8-I; ctagc**atg**ggtcttatgattatgtccggt**taa**g

    ppLUC8-II: gatccttaaccggacataatcataagacccatg

    ppLUC14-I: ctagc**atg**ggcgaactgtgtgtgagaggtcctatg attatgtccggt**taa**g

    ppLUC14-II: gatccttaaccggacataatcataggacctctcacacacagttcgcccatg

    ppLUC 23, 75, 91 and 131 were made by amplifying a stretch of DNA from the luciferase gene (accession number pGL3 Basic E1751, Promega) by PCR using the following primer combinations:

    pp23-forward: gctagc**atg**gggaaaacgctgggcgttaatc

    pp74-forward: aggcgctagc**atg**gccaagaggttccatctg

    pp91-forward: aggcgctagc**atg**ggaaattgcttctggtggcgctc

    pp131 forward: aggcgctagc**atg**gaattattttttgaggtcgttgc

    luc stop-reverse: aggcactagt**tta**accggacataatcataggac

The resulting DNA fragments were cloned into a NheI site upstream of the d2EGFP reporter gene. For the cyclin kinase inhibitor p30 Kip1 derived peptides, the same procedure was applied as for the LUC peptides using annealing primers for ppKIP8 and 14 and amplifying a stretch of DNA in case longer peptide by PCR using a cDNA clone as a template (accession number U49649.1).

Creation of the Error Prone PCR Zeocin mutants

We created mutations in the Zeocin resistance marker gene by PCR amplification of the gene in such a way that random mutations are introduced. More manganese and magnesium ions in the reaction mix, as well as an adjusted mix of nucleotides induce random mutations in the PCR product (Bloom et al. 2005). 50 ng of the pCMV/ZEO plasmid (Invitrogen, V50120) that harbors the Zeocin gene was amplified with 0.75 μM each of the following primers:

    Zeo forward: attaggatccacc**atg**gccaagttgaccagtgccg

    Zeo reverse: accggaattc**tca**gtcctgctcctcggccacg

The reaction was performed in 7 mM $MgCl_2$, 75 μM $MnCl_2$, 0.2 mM dATP, 0.2 mM dGTP, 0.5 mM dTTP, 0.5 mM dCTP, 1× GOTaq buffer (Promega) with 5 u GOTaq polymerase (Promega). Amplification was for 10–40 cycles for 1 min at 95 °C, 1 min at 50 °C, and 1 min at 72 °C. The resulting Zeocin fragments derived from the PCR reaction were cut with BamH1 and EcoRI, and cloned behind the EM7 promoter in a pBluescript derived plasmid, which also harbors the Ampicillin resistance gene. E.coli (XL10) colonies were selected on LB agar plates with 100 μg/mL ampicillin. Since the Ampicillin resistance gene is not affected by the PCR procedure on the Zeocin resistance gene, equal numbers of ampicillin-resistant colonies are to be expected, even if the Zeocin resistance gene is functionally totally destroyed by mutations.

Randomly chosen ampicillin-resistant recombinants were then plated on LB agar plates containing both 100 µg/mL ampicillin and 50 µg/mL Zeocin. The growth of the recombinants was then compared to the growth of E.coli XL10 transformed with the wild type Zeocin gene. This results in a lower ratio of Zeo/Amp resistant colonies. Colonies showing impaired growth on Zeocin were then further characterized by plating them on various Zeocin concentrations and genotypically indentified by sequencing (BaseClear, Leiden, The Netherlands).

Cell culture, transfection, and analysis of clones

CHO-DG44 (Urlaub et al. 1983) cells were grown in DMEM F12 medium, supplemented with 4.6% fetal bovine serum (FBS), 2 mM glutamine, 100 U/mL penicillin,100 µg/mL streptomycin, 100 µM sodium hypoxanthine, 16 µM thymidine (Invitrogen), and 10 mM $MgCl_2$ at 37 °C/5% CO_2.

For transfections, 0.4×10^6 CHO-DG44 cells were seeded in 6-well culture plates 24 h prior to transfection. Cells were transfected with 3 µg of plasmid DNA using LipofectamineTM 2000 (Invitrogen) as described by the manufacturer. In brief, LipofectamineTM 2000 was combined with plasmid DNA at 4 µL of lipofectamine/µg of pDNA. The mixture was added to the cells, which had grown to 70–90% confluence. After 5 h, the transfection mixture was replaced by fresh medium. The following day, cells were seeded in serial dilutions into medium containing Zeocin (Invitrogen) at a concentration of 400 µg/mL. Approximately 12 days after transfection, individual colonies became visible and these were isolated and propagated in 24-well plates in medium containing Zeocin. When grown to ∼70% confluence, cells were transferred to 6-well plates. Cells were continued to grow in 6-well plates for another 1–2 weeks before FACS analysis was performed. The d2EGFP expression levels were determined on an Epics XL Beckman Coulter flow-cytometer. In case of transient transfections, d2EGFP expression levels were determined 24 h following transfection.

RT PCR protocol

Zeocin RNA levels were quantified in a selected number of clones, by using qRT-PCR. Individual clones were grown in 6-well plates from which the total RNA was extracted using the RNeasy extraction kit (Qiagen) according to the manufacturer's protocol. Per sample 0.2 µg of RNA was treated with DNAseI followed by cDNA synthesis using random hexamer primers (SuperScript III, Invitrogen). The cDNA was diluted (1:100) followed by a real-time PCR (40 cycles in a two-step protocol) using the Cybr-Green Taq mix (Bio-Rad). Primer-pairs with amplicons of 100–200 bp length were used that were specific for coding region of the Zeocin selection marker and the endogenous β-actin gene, which was used to relatively quantify the RNA levels per sample. For each primer pair a minus RT control sample was included to quantify the background signal.

Western blot analyses

Immunodetection of the different mutated Zeocin proteins was performed preparing total cell lysates of 1.6×10^6 CHO-DG44 cells, 24 h post-transfection in 100 µL 2× SDS sample buffer (100 mM dithiothreitol, 2% SDS, 50 mM Tris, pH 6.8, 10% glycerol and 0.1% bromophenol blue), at approximately 2×10^6–1×10^7 cells/mL. Per sample 10 µL was subjected to electrophoresis on a 15% SDS–polyacrylamide gel and transferred to a nitrocellulose membrane by electroblotting. A 1:3,000 dilution of an anti-Sh ble (Zeocin resistance protein) rabbit antiserum (Cayla) was incubated with the membrane followed by a secondary incubation with a 1:5,000 dilution of goat anti-rabbit antibody conjugated to alkaline phosphatase (Jackson Immuno Research Lab).

Results

Experimental set up

Normally, translation of the wild type Zeocin selection marker starts with an ATG. We preserved this configuration, but we placed small open reading frames (ORFs) of different lengths upstream of this AUG. Each stretch started with an optimal AUG to provide a translation initiation codon, and a TAA stop codon, which was placed upstream of the AUG of the Zeocin selection marker. The idea underlying these Zeocin gene configurations is that translation will be

initiated at the AUG of the short ORF, and terminated at the stop codon of the RNA stretches, thus creating small non-functional peptides. However, since the peptides are relatively small, the translation machinery is likely to re-initiate translation when it encounters a second AUG present on the same messenger RNA; in this configuration at the start of the coding region for Zeocin (Kozak 2001). This ability of the translation machinery will diminish however, when the upstream open reading frames become longer. Thus, a Zeocin selection marker open reading frame (ORF) coupled to a small upstream ORF encoding 8 amino acids will be less efficiently translated than when no extra RNA stretch is present (wild type), but will still be more efficiently translated than a Zeocin selection marker coupled to an upstream ORF encoding 91 amino acids. As a result, the last mentioned Zeocin marker (with a peptide of 91 amino acids) will functionally become a more stringent selection marker than the Zeocin selection marker preceded by the 8 amino acids long small peptide, which in turn will be more stringent than the wild type Zeocin marker. The translation efficiency of the selection marker and thereby the selection stringency in this configuration depends on the length of the ORF in front of the Zeocin selection marker.

To test this experimental set up we created a series of plasmids in which short stretches of DNA were placed upstream of a reported gene. As reporter gene we took the d2EGFP gene, which was driven by the CMV promoter (Fig. 1a). The ORFs were taken arbitrarily from the coding region of the luciferase gene (Fig. 1b) or p30 Kip1 gene (Fig. 1c). Different lengths were taken encoding, respectively peptides of 7, 13, 22, 73, 90 and 131 amino acids. Each ORF started with an optimal AUG to provide a translation initiation codon, thus creating peptides of 8, 14, 23, 74, 91 and 131 amino acids in total (as such indicated in Fig. 1b, c). All ORFs further contained a TAA stop codon, which was placed upstream of the AUG of the d2EGFP gene, with 19 nucleotides between the stopcodon of the luciferase ORF and the AUG of the d2EGFP gene. Seven different constructs were thus created, containing no peptide (pp0), or a small peptide (pp for 'petit peptide'), 8, 14, 23, 74, 91 or 131 amino acids long (Fig. 1d). These constructs were transfected to CHO-DG44 cells and 24 h after transfection cells were analyzed for d2EGFP protein expression by flowcytometry. The fluorescence signal
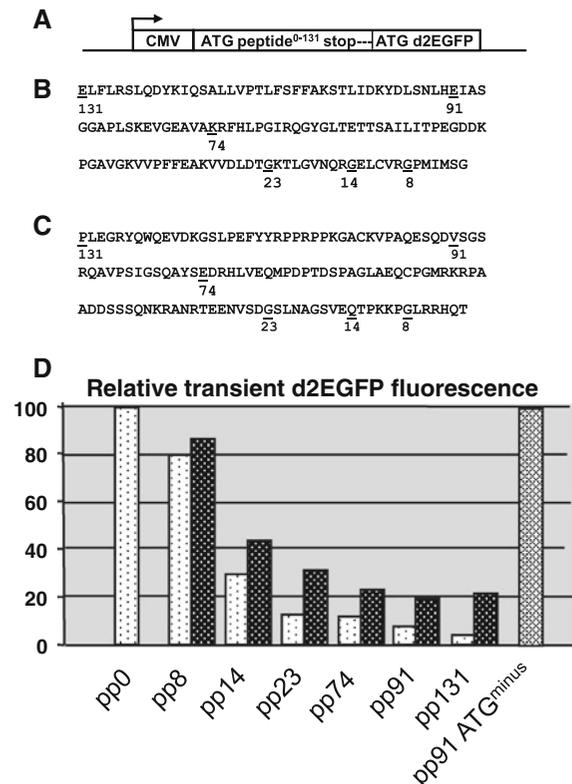

Fig. 1 Influence of the length of an upstream small peptide on transient expression. a Stretches of DNA of 6 different lengths were cloned immediately upstream of ATG of the d2EGFP reporter gene. Each DNA stretch contained a 5′ ATG and was terminated by a TAA stop codon. The CMV promoter drove expression. Different constructs were thus created, containing no peptide, or a small peptide (pp for 'petit peptide'). The constructs are a control construct, containing no peptide (pp0), pp8, 14, 23, 74, 91 and 131. Furthermore, a control stretch of DNA, derived from the luciferase gene, containing no internal ATGs and no ATG translation initiation codon was placed upstream from the d2EGFP gene (called pp91 ATG⁻). b Stretches of DNA that were taken from the luciferase gene. The protein sequence and position of the above indicated stretches are shown. c Stretches of DNA that were taken from the p30 Kip1 gene. The protein sequence and position of the above indicated stretches are shown. d The above described constructs were transiently transfected and 24 h after transfection cell were analyzed for d2EGFP protein expression by flowcytometry. The d2EGFP values with the luciferase (light bars) and p30 Kip1 (black bars) are shown and their respective numbers indicate the various small peptides

derived from d2EGFP (destabilized) is linear with the amount of available d2EGFP protein in a cell, and is thus a reliable indicator of the d2EGFP expression levels in the cell. As is shown in Fig. 1d, placing a short ORF upstream of the d2EGFP reporter gene had

a profound and progressive interfering effect on the d2EGFP expression levels. Inclusion of an 8 and 14 amino acid long luciferase peptide (Fig. 1b) in the construct resulted in a decrease of d2EGFP expression to 80 and 30%, respectively, of the d2EGFP expression level of the control construct (put at 100%). The d2EGFP expression levels decreased further when longer peptides were included, down to <10% of the control d2EGFP expression levels when a 91 or 131 amino acid peptide was placed upstream of the d2EGFP gene. With increasingly longer peptides derived from the p30 Kip1 gene (Fig. 1c) we also noted lower d2EGFP values, but these were less outspoken than with the peptides derived from the luciferase gene (Fig. 1d). Based on this result we decided to perform the further experiments with the small peptides derived from the luciferase gene (Fig. 1b) only.

A basic assumption in these experiments was that the decline in d2EGFP signal correlates with translation of the small peptide. In a control experiment we placed a different DNA stretch upstream of the d2EGFP gene. This stretch of DNA encoded 91 amino acids, but lacked an optimal AUG at the start to provide a translation initiation codon, and contained no internal ATGs. When we transfected this construct, we noted no decrease in d2EGFP expression levels at all (Fig. 1d). This suggests that this DNA stretch, that potentially encodes 91 amino acids, is not translated and simply functions as a 5′ untranslated leader. Although this is indirect evidence, this result indicates that the decreases in d2EGFP values we observe upon incorporation of increasingly longer upstream located peptides, are indeed due to translation of these peptides.

Placing increasingly longer peptides upstream of the Zeocin selection marker results in a decreased number of colonies

We next placed the increasingly longer luciferase ORFs immediately upstream of a gene encoding a selection marker. We did this to test whether this would create more stringent selection marker, due to the increasingly diminished protein expression levels of such selection marker. In this example we chose the Zeocin selection marker protein. We placed the described pp8, 14, 23, 74, 91 and 131 DNA stretches of the luciferase gene immediately upstream of the

Zeocin gene (no reporter genes were included in these constructs). In order to compare the novel Zeocin selection stringencies with the STAR-Select system, we compared the constructs with Zeocin selection markers that were modified at the translation initiation codon, e.g., the ATG, GTG and TTG translation codons. We previously observed that inclusion of STAR elements in STAR-Select constructs have a positive effect on the number of induced colonies. We therefore tested all constructs in the absence and presence of a combination of STAR 7 and 67 upstream of the expression cassettes and STAR 7 downstream of the expression cassettes. This configuration has been reported to provide a favorable context for highly elevated protein expression levels in multiple cell lines and with different promoters (Van Blokland et al. 2007).

The same amount of DNA of all constructs was transfected to CHO-DG44 cells and selection was performed with 400 µg/mL Zeocin in the culture medium. After ∼2 weeks the number of stably established colonies was counted. As shown in Fig. 2, transfection of the STAR-Select constructs containing the ATG Zeo selection marker created most stable colonies (∼1,750) in the absence of STAR elements and >2,500 in the presence of STAR elements. The inclusion of the GTG Zeo and TTG Zeo selection markers resulted in significantly less stable CHO-DG44 colonies. GTG Zeo induced ∼150 colonies in the absence and ∼750 in the presence of STAR elements (Fig. 2). TTG Zeo induced 6 colonies in the absence and 45 in the presence of STAR elements (Fig. 2).

In comparison, inclusion of pp8 Zeo in the construct induced ∼1,050 colonies in the absence and >2,000 in the presence of STAR elements (Fig. 2). With the exception of the pp14 (∼500 colonies), the inclusion of longer peptides resulted in the formation of only a few stable colonies when no STAR elements were included in the construct. However, when STAR elements were added to flank the expression cassette, pp14, 23, 74, 91 and 131 still gave ∼2,000, ∼1,400, ∼475, ∼250 and ∼225 stable colonies, respectively.

Thus, inclusion of progressively longer peptides resulted in the establishment of a decreasing number of stably transfected colonies. This indicates that with increasingly longer peptides the resulting Zeocin selection marker system becomes more stringent.
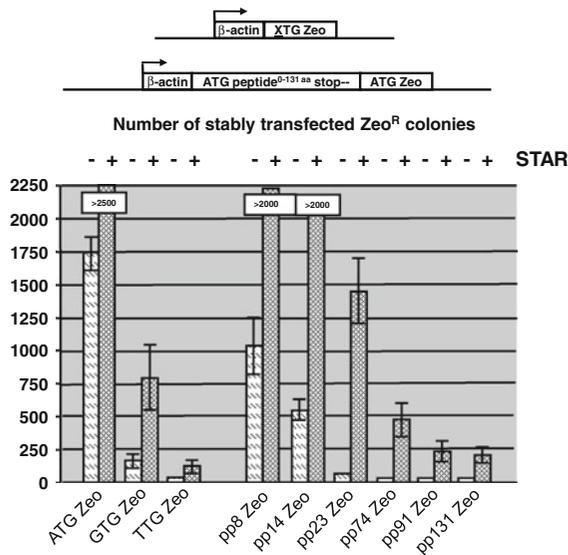
Fig. 2 Influence of the length of an upstream peptide on colony formation in the absence or presence of STAR elements. The same pp encoding luciferase stretches as used in Fig. 1b (pp8, 14, 23, 74, 91 and 131) were placed immediately upstream of a gene encoding the Zeocin selection marker. The novel pp-zeocin selection stringencies were compared with known Zeocin STAR-Select selection markers, e.g., the ATG/GTG/TTG Zeo configurations. All constructs were tested in the presence and absence of flanking STARs 7 and 67 encompassing the human β-actin promoter driven expression cassettes and STAR 7 downstream of the expression cassettes. The various constructs are schematically depicted and the *bars* indicate the average number from three experiments of stably transfected Zeocin resistant colonies, obtained with the various constructs as indicated. The SEM is indicated with *error bars*

However, inclusion of the pp131 peptide did not result in significantly less colonies than inclusion of the pp91 peptide. Apparently there is an upper limit to the length of the peptide in relation to the decrease in colony numbers.

Placing increasingly longer peptides upstream of the Zeocin selection marker results in increased expression levels of a reporter protein

We next tested the effects of increasingly longer small peptides on protein expression levels. To include the Zeocin selection marker variants in the constructs, we placed these downstream of an internal ribosome entry site (IRES) sequence (Fig. 3). The d2EGFP gene was cloned upstream of the IRES sequence. We also compared the constructs with the ATG/GTG/TTG Zeo-d2EGFP STAR-Select
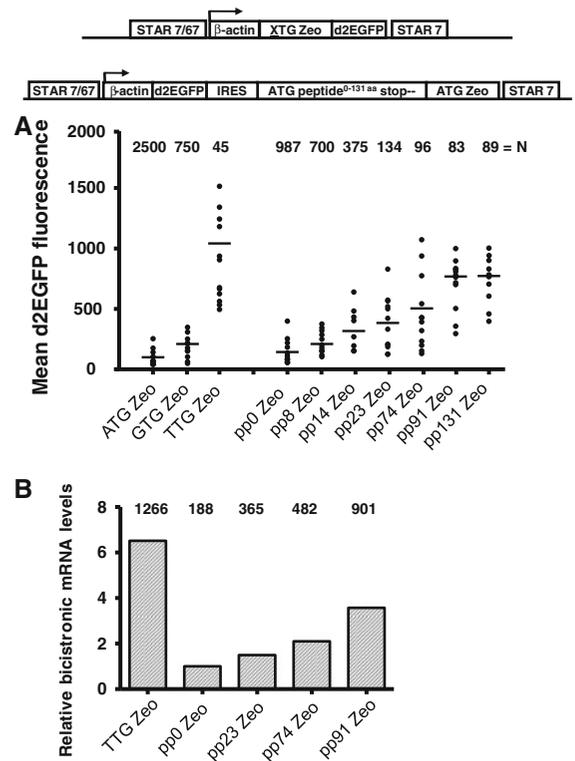

Fig. 3 The use of small peptides creates a high stringency selection system that can be used to achieve high protein expression levels. **a** The novel Zeocin selection markers with the same pp encoding luciferase stretches as used in Fig. 1 (pp8, 14, 23, 74, 91 and 131) were placed behind an internal ribosome entry site (IRES). The Zeocin genes and IRES were placed downstream of the d2EGFP reporter gene, to determine the expression levels after selecting stably transfected clones. As controls the ATG Zeo gene (or pp0) behind the IRES sequence was used and for comparison the constructs with the ATG/GTG/TTG Zeo STAR-Select configuration were included. The various constructs are schematically depicted and the mean d2EGFP expression levels in the Zeocin resistant colonies are indicated (*bars*). **b** Real-time RT–PCR analysis of stably transfected CHO-DG44 clones with TTG Zeocin and pp0, 23, 74 or 91 ATG Zeocin placed behind the IRES. Total RNA was extracted from 3 clones per construct. First strand cDNA was used as a template using Zeocin and endogenous β-actin specific primers. The Ct values were normalized per sample for the expression of β-actin and the fold induction was calculated relative to the average Ct values in the pp0 ZEO clones. The numbers (*top*) indicate the mean d2EGFP expression levels in the three clones

configurations. All constructs were flanked with STAR elements 7 and 67.

The same amount of DNA of all constructs was transfected to CHO-DG44 cells and selection was performed with 400 µg/mL Zeocin in the culture medium. Also with these constructs we noticed a

declining number of colonies when increasingly longer peptides were placed upstream of the Zeocin selection marker (Fig. 3). It is important to note that the set up in this experiment is different form the experiment described in Fig. 2. In Fig. 2, the Zeocin selection marker was placed immediately downstream of the β-actin promoter, whereas in Fig. 3 the Zeocin markers were placed downstream of an IRES sequence. Translation of the Zeocin marker downstream of an IRES sequence in a bicistronic mRNA is expected to be less efficient. The consequently lower Zeocin protein levels are expected to raise the stringency of the selection system, with fewer colonies as result. This is exactly what we observed: the overall number of colonies shown in Fig. 3 is reduced in comparison with the numbers shown in Fig. 2.

Next, up to 12 independent colonies were propagated before flow cytometric analysis (EPIXS-XL, Beckman-Coulter), 3–4 weeks after transfection. In a single FACS analysis, fluorescence signals from a sample that contain up to 4,000 cells are analyzed. One such sample of cells is taken from an independent, stably transfected cell colony. Since the signal will vary amongst the individual cells in the colony, the mean fluorescence level of the $\sim 4,000$ cells in the sample is taken as a measure for the d2EGFP expression level in the stably transfected cell colony. As shown in Fig. 3a, incorporation of increasingly longer peptides upstream of the Zeocin selection marker, gave significantly higher d2EGFP expression levels, as compared to the control construct with the ATG Zeo (pp0) marker. The average d2EGFP expression level in the independent colonies rose from ~ 100 to ~ 700 with the ATG pp0 Zeo and pp131 Zeo marker, respectively. However, no further increase in d2EGFP fluorescence values were observed with the pp131, in comparison to the pp91 peptide, suggesting there is an upper limit to the increase of selection stringencies, conveyed with these peptides. Furthermore, also the colony number did not decline when the peptide became longer than 91 amino acids (Fig. 3a). This is in agreement with the data shown in Fig. 2, in the context of a slightly different experimental set up.

It is also important to note that incorporation of the TTG Zeo selection marker in this same experiment resulted in an average d2EGFP expression level of 1,150, which is still higher than the average d2EGFP

values reached with the pp91 and 131 small peptides. It is possible that selection of increasingly higher protein expressing pp91 or pp131 clones might be compromised due to an increasing metabolic load in high protein expressing cells. However, we believe that this is not the explanation for the upper limit in protein expression levels with the pp91 and 131 peptides that we observe. If these expression levels were too high, also colonies induced by the TTG Zeo construct, which display higher d2EGFP values, would have been compromised, and this is not the case. Also, we did for instance not note a decline in the growth rate of the highest expressing TTG Zeo clones. This indicates that high d2EGFP expression values in these colonies are not yet in such a high range that they start to negatively influence selection of these colonies.

One possibility is that the increased d2EGFP expression levels are not due to altered mRNA levels in the different mRNAs, but that inclusion of the DNA stretches encoding the small peptides influence the translation rate of the d2EGFP reporter gene. We therefore analyzed the mRNA levels by real time PCR in clones that were induced by the TTG ZEO, pp0, 23, 74 and 91 constructs as shown in Fig. 3b. We found that Zeocin mRNA levels in the clones induced by the different constructs increased upon inclusion of a longer peptide, coinciding with higher d2EGFP expression values (Fig. 3b). A same trend was found for colonies induced by constructs containing TTG Zeo marker proteins (Fig. 3b). This indicates that the differences we find in d2EGFP fluorescence levels are due to differential mRNA levels.

We also determined whether the distance between the stopcodon of the small ORF and the AUG of the d2EGFP gene was of influence on the expression levels of d2EGFP. The 19 nucleotides we use in the constructs described in Figs. 1, 2 and 3 were based on the length of the PCR primer that contained suitable restriction sites. We did, however, also vary the distance between the stopcodon of the small peptide and the AUG of the d2EGFP gene, up to 200 nucleotides. We found no differences in d2EGFP values (*data not shown*) and therefore decided to use the 19 nucleotides in all further experiments.

To assess the broader applicability of the small peptides to modulate the stringency of a selection marker we performed two experiments. We transfected the constructs described in Fig. 3 to CHO-K1

cells and found that also in this cell line less colonies were induced by constructs that contained increasingly longer peptides (*data not shown*) and further that these colonies displayed higher d2EGFP values (*data not shown*). The result was very similar to the described results with CHO-DG44 cells and shows that these results were not cell type dependent. We also included the same range of small peptide encoding DNA stretches upstream of the Neomycin resistance gene and transfected these to CHO-DG44 cells. Although the overall number of emerging colonies was higher with the Neomycin resistance gene than with the Zeocin resistance gene, we observed the same trend: inclusion of longer peptides resulted in fewer colonies, displaying higher d2EGFP expression levels (*data not shown*). This result indicates the potential for a broader applicability of the use of small peptides to modulate the stringency of selection markers.

Overall we therefore, conclude that the inclusion of small DNA stretches that translate to small peptides upstream of a selection marker can be used to increase the selection stringency of selection markers.

The creation of Zeocin mutant proteins with attenuated ability to neutralize Zeocin

Inclusion of subsequently longer ORFs upstream of the Zeocin selection marker has a limit, both in terms of colony numbers and protein expression levels. In comparison (Figs. 2, 3), the TTG Zeocin variant induces significantly fewer colonies and these colonies display higher protein expression levels. In order to obtain similar colony numbers and protein expression levels as the TTG Zeocin variant we sought other ways to increase the selection stringency of the Zeocin marker, in the context of the small peptide concept. One way of doing this is impairing protein function by the introduction of mutations. We created mutations in the Zeocin resistance marker gene by PCR amplifying the gene in such a way that random mutations were introduced. More manganese and magnesium ions in the PCR reaction mix, as well as an adjusted mix of nucleotides induce random mutations in the PCR product of the Zeocin resistance gene (Bloom et al. 2005). The resulting Zeocin fragments derived from the PCR reaction were cloned behind the EM7 promoter. The plasmid also harboured the Ampicillin resistance gene (driven by its

natural, beta lactamase promoter) and colonies were selected on 100 μg/mL Ampicillin. Randomly chosen Ampicillin-resistant recombinants were then plated on agar plates containing both 100 μg/mL Ampicillin and 50 μg/mL Zeocin. The growth of the recombinants was compared to the growth of *E. coli* XL10 transformed with the wild type Zeocin gene. Since the Ampicillin resistance gene is not affected by the PCR procedure on the Zeocin resistance gene, equal numbers of Ampicillin-resistant colonies are to be expected, even if the Zeocin resistance gene is functionally totally destroyed by mutations. Thus, a functional impaired Zeocin gene would result in a lower ratio of Zeo/Amp resistant colonies. We indeed found (Fig. 4a) that increasing the number of PCR cycles resulted in a decreasing number of Zeocin resistant transformants. Consequently, the ratio of Ampicillin resistant transformants that were also Zeocin resistant decreased. However, inclusion of Zeocin fragments in the construct that had undergone 40 PCR cycles hardly delivered colonies that were both Ampicillin and Zeocin resistant. Apparently, resulting Zeocin fragments were not anymore able to produce a selection protein with enough functionality to confer Zeocin resistance. We therefore chose to concentrate on Zeocin mutation screens that resulted from 15 PCR cycles.

A number of Zeocin marker mutants were plated on different Zeocin concentrations, ranging from 0 to 100 μg/mL Zeocin, as indicated in Fig. 4b. Note that all constructs containing a mutated Zeocin marker still grew efficiently on ampicillin alone (*top panel*). As control Zeocin marker gene we included the wild type Zeocin protein (Fig. 4b). The novel Zeocin mutants are described as Zeocin EPP, which stands for *Error Prone PCR*. These mutants were in varying degree affected by increasing Zeocin concentrations. For instance, the Zeo^{EPP66} mutant was less affected by increasing concentrations of Zeocin than the Zeo^{EPP7} mutant (Fig. 4b). In Fig. 5 the amino acid positions of several EPP Zeocin mutations are shown. For instance, the EPP 66, 7 and 14 mutants harbored 1, 2 and 3 mutations, respectively, at different amino acid positions.

Error Prone PCR created Zeocin mutants confer different selection stringencies

Next we tested whether the described Zeocin mutants could be used as high stringency Zeocin selection
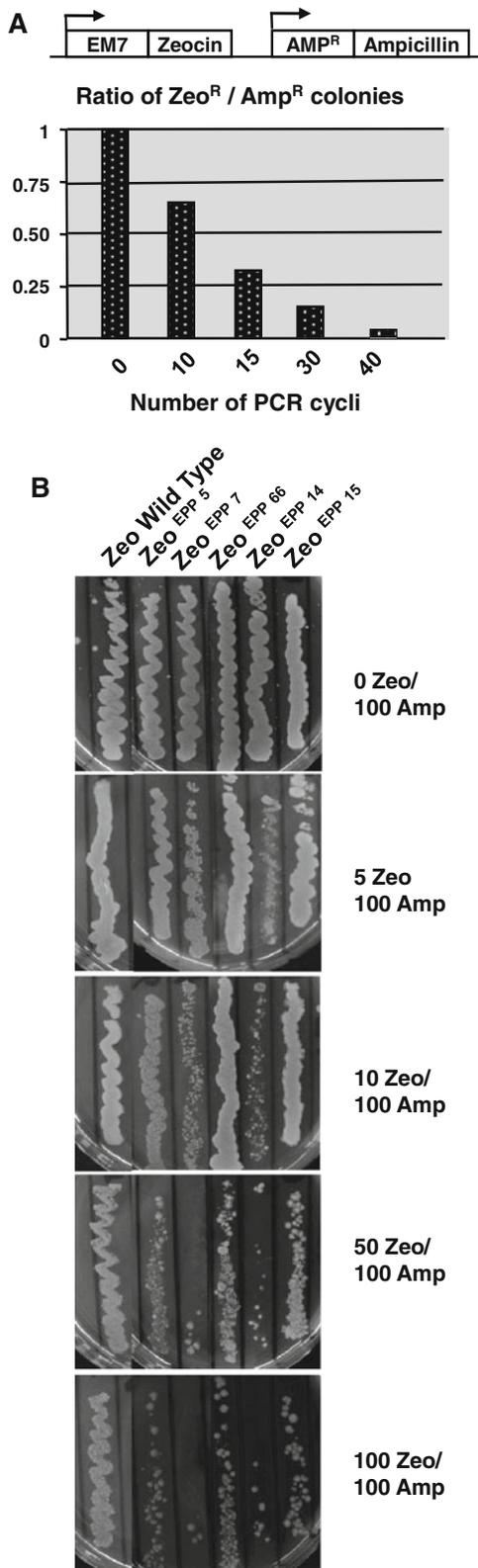
A

markers in mammalian CHO-DG44 cells. The different Zeocin EPP mutants were cloned in an expression cassette, encompassing the human β-actin promoter that drove the d2EGFP gene, followed by an IRES sequence and the Zeocin EPP mutants. STAR 7 and 67 elements flanked the expression cassettes, as shown in Fig. 6. As comparison we included three STAR-Select constructs in the same experiment. The ATG/GTG/TTG Zeo configurations induced >2,500, >2,000 and 68 colonies, respectively (Fig. 6). Selection was performed with 400 µg/mL Zeocin concentration for all constructs. The decreasing numbers of formed colonies signify the increasing selection stringencies conferred by the ATG, GTG and TTG translation initiation codons. In comparison, the wild type Zeocin gene placed downstream of the IRES sequence also resulted in the induction of >2,000 colonies.

The different EPP Zeocin mutants induced varying colonies under the same selection conditions. The Zeocin EPP5, 7, 66, 14 and 15 mutants were able to induce 215, 67, 435, 64 and 113 stably transfected CHO-DG44 colonies, respectively (Fig. 6). This result shows that the EPP mutations create Zeocin resistance marker proteins with different selection stringencies.

When we analyzed the average d2EGFP fluorescence values in the respective clones, we found varying values with the different Zeocin mutants (Fig. 7a). Of the control, STAR-Select constructs, the TTG Zeo STAR-Select configuration gave the highest d2EGFP values (average 1,235). The ATG and GTG STAR-Select configurations gave average d2EGFP fluorescence values of 93 and 147, respectively (Fig. 7a). In comparison, the Zeocin EPP5, 7, 66, 14 and 15 mutants induced an average d2EGFP values of 525, 973, 187, 1,016 and 674, respectively (Fig. 7a). This indicates that the EPP5, 7, 14, and 15 mutations induce a selection stringency in the Zeocin selection marker protein that lies between the selection stringency of the GTG and TTG Zeo STAR-Select configuration.

The higher selection stringencies conveyed by the described EPP Zeocin mutants could be due to

B

Fig. 5 Amino acid
substitutions in various
Zeocin mutants with
reduced activity

```
WT        makl+tsavpv ltardvagav efwtdrlgfs rdfveddfag        40
EPP 5     ---------- ---------- G--------- ----------
EPP 7     ---------- ---------- -------R-- ----------
EPP 66    ---------- ---------- ---------- ----------
EPP 14    --------A- -A-------- ---------- ----G-----
EPP 15    ---------- ---P------ ---------- ----------

WT        vvrddvtlfi savqdqvvpd ntlawvwvrg ldelyaewse        80
EPP 5     ---------- ---------- ---------- -----T----
EPP 7     ---------- ---------- ---------- --G-------
EPP 66    ---------- ---------- ---------- ----------
EPP 14    ---------- ---------- ---------- ----------
EPP 15    ---------- ---------- ---------- --K-------

WT        vvstnfrdas gpamteigeq pwgrefalrd pagncvhfva eeqd 124
EPP 5     ---------- ---------- ---------- ---------- ----
EPP 7     ---------- ---------- ---------- ---------- ----
EPP 66    -E-------- ---------- ---------- ---------- ----
EPP 14    ---------- ---------- ---------- ---------- ----
EPP 15    -------G-- ---------- ---------- ---------- ----
```
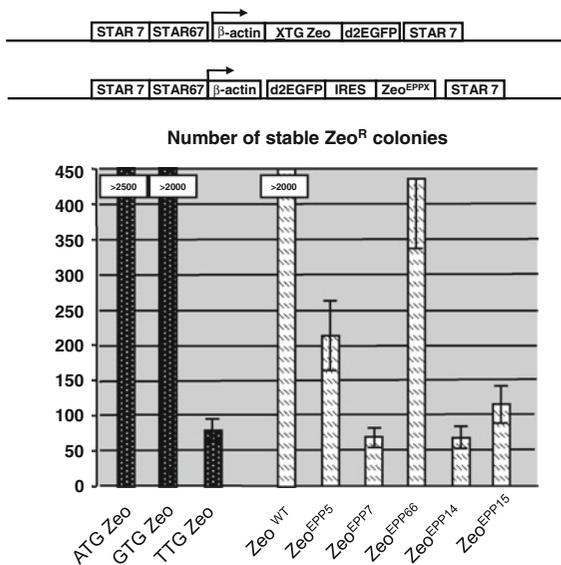
Fig. 6 Colony formation induced by Zeocin EPP mutations.
Zeocin mutants as indicated were cloned in an expression
cassette, encompassing the human β-actin promoter that drove
the d2EGFP gene, followed by an IRES sequence and the
Zeocin EPP mutants. STAR 7 and 67 elements flanked the
expression cassettes. In the same experiment, known Zeocin
STAR-Select selection markers (ATG/GTG/TTG Zeo) were
included for comparison. The various constructs are schemat-
ically depicted and the *bars* indicate the average number from
3 experiments of stably transfected Zeocin resistant colonies
obtained with the constructs as indicated. The SEM is indicated
with *errror bars*

impaired functionality of the mutant marker protein.
Alternatively, these mutants could be less stable than
the wild type Zeocin resistance protein. To address
these possibilities, we analyzed whether the expres-
sion levels of these different Zeocin mutants were
similar to the expression of the wild type Zeocin
gene. We transfected the different EPP Zeo mutants
transiently to CHO-DG44 cells and analyzed the
expression levels of the Zeocin resistance protein by
using a specific antibody against the Zeocin or Sh ble
protein. As shown in Fig. 7b, we found that mutated
Zeocin proteins had a lower expression level than the
wild type Zeocin resistance protein, except mutant
EPP66. The least stringent EPP66 mutant showed the
highest Zeocin marker protein levels, whereas one of
the most stringent mutants EPP7 was hardly detect-
able. The Zeocin protein levels with EPP5, 14 and 15,
lies between EPP66 and 7. This result indicates that a
decreased stability of the resistance protein is at least
partly responsible for the increased selection strin-
gency conveyed by the mutants.

We conclude that Error Prone PCR can be used to
create mutant Zeocin marker proteins that convey
high selection stringency in mammalian cells. The
error prone PCR strategy appears to be successful to
create Zeocin mutants that display a range of
selection stringencies. However, the gaps between
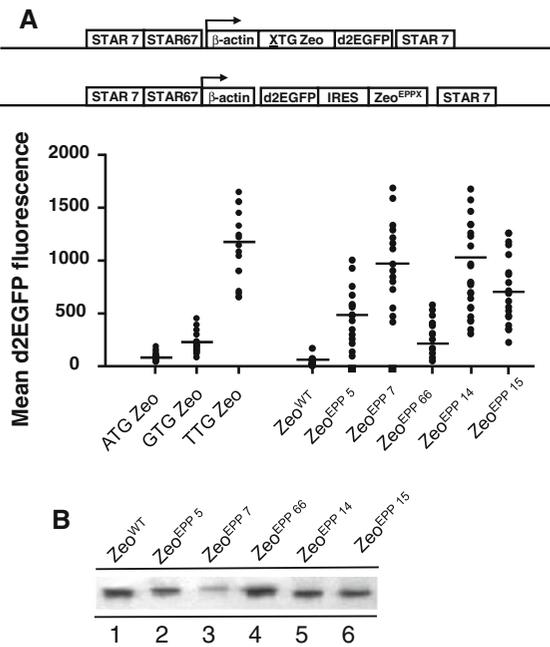these stringencies are rather large and resemble the

Fig. 7 Influence of Zeocin mutations on reporter protein expression levels. **a** Zeocin mutants as indicated were placed behind an internal ribosome entry site (IRES). The EPP Zeocin genes and IRES were placed downstream of the d2EGFP reporter gene, to determine the expression levels after selecting stably transfected clones. As controls the wild type Zeocin gene (Zeo WT) behind the IRES sequence was used and for comparison the constructs with the ATG/GTG/TTG Zeo STAR-Select configuration were included. The various constructs are schematically depicted and the mean d2GFP expression levels in the Zeocin resistant colonies are indicated (*bars*). **b** Western blot analysis showing the transient CMV driven expression of the wild type Zeocin resistance marker and the Zeocin mutants in CHO-DG44 cells 24 h post-transfection
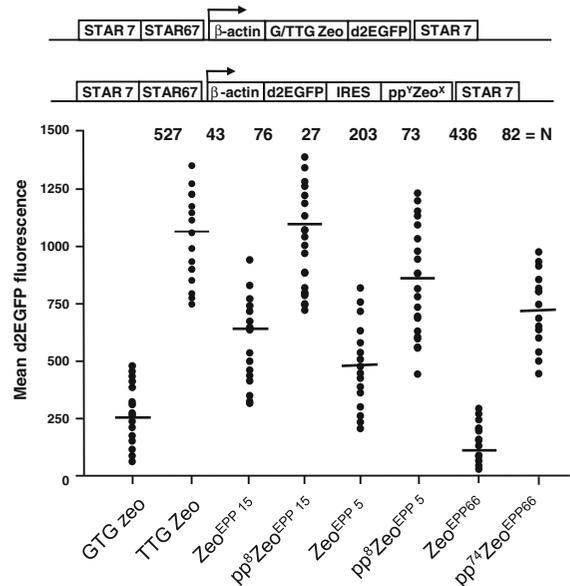
Fig. 8 Influence of the combination of Zeocin mutations with the use of small peptides on colony formation. Zeocin mutants EEP 5 and 15 with or without small peptide pp8 were placed behind an internal ribosome entry site (IRES). The Zeocin genes and IRES were placed downstream of the d2EGFP reporter gene. As control the construct with the TTG Zeo STAR-Select configuration was included. The constructs are schematically depicted and the *bars* indicate the mean d2EGFP expression levels in the stably transfected Zeocin resistant colonies

above-explained differences in selection stringencies between the GTG and TTG Zeo configurations. In order to create more 'in-between' selection stringencies, we combined several Zeocin EPP marker mutants with small peptides.

Small peptides, combined with Error Prone PCR created Zeocin mutants, modulate the selection stringencies of these Zeocin mutants

We created the Zeocin mutants harboring small peptides to more subtly modify the selection stringency of the Zeocin marker than with EPP mutations. We tested several combinations of small peptides and EPP Zeocin mutants. In Fig. 8 we show three

combinations: a pp8 small peptide placed upstream of the ATG of the Zeo^{EPP15} and Zeo^{EPP5} mutants and a pp73 peptide upstream of the Zeo^{EPP66} mutant. These new Zeocin configurations were placed downstream of the IRES sequence and d2EGFP gene (Fig. 8). Introducing the small peptides upstream of the ATG of the Zeo^{EPP15}, Zeo^{EPP5} and Zeo EPP^{66} mutants resulted in a progressive decrease in the number of stably transfected CHO-DG44 colonies (Fig. 8). In either case, the number of stably transfected clones decreased by more than 50% when a small peptide was added to the Zeo mutant (Fig. 8). We further found that the expression levels of the d2EGFP protein strongly increased upon the inclusion of the pp8 or 74 small peptide (Fig. 8). The inclusion of the rather long small peptide (pp74) in the Zeo^{EPP66} containing construct resulted in a selection stringency that resembled the number of colonies and d2EGFP expression values that were induced by the Zeo^{EPP15} mutant (Fig. 8). However, the Zeo^{EPP15} and Zeo^{EPP5} mutants, combined with

pp8 induced a number of colonies that displayed average d2EGFP fluorescence values in the same range as those induced by the TTG Zeo STAR-Select configuration (Fig. 8). Inclusion of a pp14 small peptide upstream of the Zeo^{EPP15} and Zeo^{EPP5} mutants resulted in an even further decrease of the induced number of colonies and in average d2EGFP values that exceeded values induced by TTG Zeo (data not shown). The pp8 Zeo^{EPP15} configuration resembled most closely the TTG Zeo configuration, whereas the pp8 Zeo^{EPP5} configuration provided an intermediate selection stringency between the GTG Zeo and TTG Zeo configurations (Fig. 8).

These results indicate that with the inclusion of a small peptide is indeed possible to modify the selection stringencies of the EPP Zeocin mutants to a similar range as the STAR-Select configurations.

Discussion

Various stringent selection systems for the establishment of stable mammalian cell lines have been described. The need for high stringency selection comes from the following consideration. With low stringency selection systems it often takes a considerable effort to screen vast numbers of transfected colonies in order to establish a cell line that produces the protein of interest at acceptable high expression level. Therefore, quite some efforts have been made to create high stringency selection systems for mammalian cell lines. In all these cases, high selection stringency involves either low expression levels of the selection marker protein, or in case of normal expression levels, impaired functionality of the selection marker protein.

There are multiple ways to achieve a high selection stringency. For instance, the selection marker protein, such as the Neomycin resistance protein has been mutated to become less functional (Sautter and Enenkel 2005). Weak promoters (Niwa et al. 1991) and destabilising sequences (Ng et al. 2007) have been employed to induce low expression levels of the selection marker. To increase the selection stringency of the system, methotrexate that inhibits the activity of the dhfr selection marker protein has been added and as a result the entire locus encompassing the dhfr gene is amplified (Kaufman and Sharp 1982; Alt et al. 1978; Kaufman et al.

1985). These 4 examples have in common what all stringent selection systems want to achieve. By lowering the availability and/or functionality of the selection marker protein a cell has to increase the transcription of the transfected selection genes in order to raise the level of functional selection marker protein. If the cell fails to do that, it will die, due to the presence of the selection agent/inhibitor in the culture medium. Underlying all these strategies is the thought that concomitant with raising the expression levels of the selection marker genes, also higher levels of the protein of interest are achieved. However, when the selection marker and the protein of interest are not on the same messenger RNA, this is not automatically the case. Therefore, the best way to achieve this goal is when the genes are coupled to each other. For instance, an IRES sequence physically couples the gene of interest to the selection marker gene. Raising the expression level of the selection marker automatically results in an increased transcription of the bicistronic mRNA as well in the expression of the gene of interest. Previously, we took a novel step in raising the selection stringency to a higher level than with the use of an IRES sequence. The stringent selection system we described was based on attenuation of the translation of the selection marker protein. Although described for the Zeocin selection marker, that system could in principle be applied to a variety of selection marker proteins. However, two major difficulties arose. As explained in the introduction, in practice, only the TTG translation initiation codon was suitable for use in the stringent selection system. Whereas the GTG translation initiation codon provided insufficient selection stringency, the TTG codon, on the other hand, created too high selection stringency for cell lines other than CHO derivatives. A second problem arose from the fact that in this experimental set up the selection marker is placed upstream of the gene of interest. The idea behind this concept was that an attenuated translation initiation codon would result in scanning of the mRNA by the translation machinery until it would meet a bona fide AUG codon. In the system, this would be the AUG of the gene of interest. However, to achieve that situation, it had to be avoided that the translation machinery met an internal AUG in the selection marker mRNA. In the presence of internal AUGs, translation would initiate, resulting in either a fusion protein between part of the

selection marker and the protein of interest, or in merely a truncated selection marker protein. This could be circumvented by replacement of all internal AUGs in the selection marker gene. In some cases this was easy. Zeocin only possesses one AUG. However, the hygromycin selection marker already contains 24 AUGs, of which 8 are in frame, making it virtually impossible to replace them all.

Here, we describe two ways to modify the Zeocin selection marker to create a stringent selection system for mammalian cell lines. First, by merely placing a small ORF upstream of the gene of interest, separated by a stop codon, we circumvent the need to replace internal AUGs in the selection gene. Second, the range in length of the potential peptides creates much more flexibility than the use of the TTG translation initiation codon only. However, even the use of a peptide with substantial length, proved to be insufficient to create similar selection stringency as the TTG Zeocin marker. This was compensated though by introducing one (or at most 3) mutations in the Zeocin selection marker protein that impair the function of the Zeocin marker protein. By combining these two approaches we created a versatile selection system that allows the establishment of few mammalian cell lines that display high protein expression levels. Importantly, we can create selection stringencies that are intermediate in relation to the respectively high TTG Zeocin and low GTG Zeocin selection stringencies.

Another aspect that needs to be discussed is whether the small peptides are bioactive and require purification when applied to enhance the expression of therapeutic proteins in an industrial setting. We randomly chose the sequence from the luciferase gene and have no indication that the resulting small peptides would have any function at all. Furthermore, with no secretion signal present, it appears unlikely that the peptides will be secreted at all. However, since we can not exclude the possibility completely, we prefer to use the shortest peptides possible, pp8 and 14. Finally, even in case the small peptides would be released into the culture medium, these peptides are so small that it is unlikely that they would co-purify with the usually very large commercial therapeutic proteins.

The molecular mechanism underlying the action of the upstream placed short ORF on the regulation of the Zeocin selection marker protein is not entirely clear. The working hypothesis, on which basis we designed the experiments, was that it involves initiation of translation at the small peptides and re-initiation of translation at the AUG of the gene of interest by ribosomes that have resumed scanning despite the translation termination signal of the initial small peptide. We can not prove this assumption directly. But in a control experiment we showed that when no ATG translation initiation codon was present in the DNA stretch potentially encoding the small peptide and with further no internal ATGs present, no decreases in d2EGFP levels were observed (Fig. 1d). This result suggests that the observed actions of the small peptides on colony numbers and d2EGFP values is indeed related or due to the translation of the small peptides. This is not so far fetched, since there is precedence for such a scenario in nature. The HER-2 receptor contains a short upstream ORF in the HER-2 mRNA which represses the downstream translation of the HER-2 protein in a variety of cell types (Child et al. 1999). It appears that this inhibitory mechanism is used as a control mechanism for the expression levels of this oncoprotein. Results presented here and other studies in mammalian systems show that the efficiency of re-initiation is reduced as the length of the ORFs is increased (Kozak 2001). Given this fact and the result we show in this paper, it is therefore possible to extend the use of small peptides, other than in selection systems. When for instance biologically active proteins need to be expressed at an attenuated level, modifying the translation efficiency of the protein can be achieved by placing an ORF upstream of the gene of interest. This may therefore provide a simple but effective tool to modulate the expression levels of any protein of interest, not only for modulating the expression of a selection marker but also to modulate multi-subunit proteins and multi-gene engineering.

References

Alt FW, Kellems RE, Bertino JF, Schimke RT (1978) Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells. J Biol Chem 253:1357–1370

Bloom JD, Silberg JJ, Wilke CO, Drummond DA, Adami C, Arnold FH (2005) Thermodynamic prediction of protein neutrality. Proc Natl Acad Sci USA 102:606–611

Child SJ, Miller MK, Geballe AP (1999) Translational control by an upstream open reading frame in the HER-2/neu transcript. J Biol Chem 274:24335–24341

Kaufman RJ, Sharp PA (1982) Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase cDNA gene. J Mol Biol 159:601–621

Kaufman RJ, Wasley LC, Spiliotes AJ, Gossels SD, Latt SA, Larsen GR, Kay RM (1985) Co amplification and coexpression of human tissue-type plasminogen activator and murine dihydrofolate reductase sequences in Chinese hamster ovary cells. Mol Cell Biol 5:1750–1759

Kozak M (2001) Constraints on reinitiation of translation in mammals. Nucleic Acids Res 29:5226–5232

Ng SK, Wang DIC, Yap MGS (2007) Application of destabilizing sequences on selection marker for improved recombinant protein productivity in CHO-DG44. Metab Eng 3:304–316

Niwa H, Yamamura K, Miyazaki J (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene 108:193–199

Sautter K, Enenkel B (2005) Selection of high-producing CHO cells using NPT selection marker with reduced enzyme activity. Biotechnol Bioeng 89:530–538

Urlaub G, Käs E, Carothers AM, Chasin LA (1983) Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. Cell 33:405–412

Van Blokland HJM, Kwaks THJ, Sewalt RGAB, Verhees JA, Klaren VNA, Siersma TK, Korse JWM, Teunissen NC, Botschuijver S, Van Mer C, Man SY, Otte AP (2007) A novel, high stringency selection system allows screening of few clones for high protein expression. J Biotechnol 128:237–245