Selection markers and expression enhancing elements: novel strategies to create mammalian production cell lines

Hoeksema, F.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
GENERAL DISCUSSION
In the general introduction we described the problems that are being encountered with the commercial production of biopharmaceuticals and in particular recombinant therapeutic proteins and monoclonal antibodies (MAbs). In essence, low protein expression levels, a high variability in expression levels, long time-lines to isolate appropriate clonal cell lines as well as instability of expression over time are major problems that still hamper the efficient industrial production of biopharmaceuticals. In this thesis we address a number of aspects that accompany these problems. Here, we will first summarize these problems, as well as discuss past attempts of our group to solve them. Then we will discuss how we currently envision solutions, in light of the work presented in this PhD thesis. Finally, we will discuss whether or not we have achieved our objectives and what we see as potential, future directions.

1 Problems with the Production of Biopharmaceuticals

The biopharmaceuticals that have reached the market are in general very expensive. This is mainly due to three reasons: 1) the initial identification or design of a proper biopharmaceutical that has therapeutic potential, 2) production costs and 3) clinical testing costs. In particular the trajectory of clinical phases is extremely expensive, since it involves the participation of many people and disciplines. But neither the identification of biopharmaceuticals nor their testing costs are subject of this thesis. Here we are concerned with the production of biopharmaceuticals. Also here we can discern two main critical phases: the establishment of a good producing production platform and the post-production phase. The last concerns purification of the biopharmaceutical and extensive testing for contaminants.

The establishment of a good production platform is the topic of this PhD thesis. It ranges from the choice of host, the selection of clonal cell lines that produce decent to high levels of the biopharmaceutical and the stability of expression over time. The most popular hosts for the production of biopharmaceuticals remain mammalian cell lines, in particular CHO cells. These CHO cell lines have been chosen for a number of reasons. The CHO cell line is known for its high stability of chromosomally integrated heterologous transgenes, its relative ease of cultivation at large scale, and for its relatively simple adaptation to adherence-independent growth in serum- and protein-free media\textsuperscript{1-3}. CHO cells are fast growing, can be readily transfected and can perform complex post-translational modifications required for biological activity of the therapeutic protein. CHO cells have been used safely for the production of recombinant therapeutics with excellent therapeutic characteristics for many years and as such are a trustworthy system. A last reason for the popularity of CHO-derived cell lines as production platform is that some of these CHO cell lines are dhfr-negative, CHO-DG44 being the best example\textsuperscript{4,5}. This allows the employment of the DHFR gene as a marker for gene amplification when present on the expression cassette as means to elevate protein expression levels\textsuperscript{6-9}. In the biotechnology industry, this is currently the most widely used means to achieve high protein expression levels.

The statement that dhfr-negative CHO cells are the most commonly used production cell lines is in fact indicative for the problems that face the industry. The establishment of a production cell line through gene amplification is a tedious, laborious, time-consuming and therefore very costly procedure. First of all, many stably transfected colonies have to be isolated to begin with. Each of these colonies has to undergo gene amplification, upon the addition of the competitive inhibitor methotrexate (see Chapter 7 for a detailed explanation). Since the outcome of the gene amplification procedure is highly unpredictable and variable\textsuperscript{10}, each of these amplified pools of clones has to be subcloned to obtain an independently established clone that grows well and that also shows increased protein expression levels due
to the increased number of amplified gene copies. Then the entire procedure of gene amplification, subcloning and testing has to be repeated, until in the end a few cell lines can be selected that show high protein expression levels. This entire procedure lasts more than six months, a long period, since this is only the beginning. After production of a limited amount of biopharmaceutical, the testing of the biopharmaceutical can proceed in preclinical studies and succeeding clinical trials. Only when the outcome is positive, up-scaling of the cell lines can begin. This is another challenge since it is not automatically that a cell line that grows well in a small vessel, will behave the same in a 10-20,000 liter bioreactor. Furthermore, the stability of selected clones over long-term culture is a critical parameter for commercial production. Expression levels are usually stable in the presence of the toxic methotrexate. In the production phase, however, such substances are preferentially avoided. Gradual loss of the gene copy number and reduction of protein expression levels are often the result\textsuperscript{11,12}.

The above sketched problems of long timelines, large screening efforts and instability of expression in the absence of methotrexate have evoked the search for solutions. These come from many angles. Robots have been developed to isolate and analyze very large numbers of stably transfected clones. This brute force approach is not unsuccessful, but is affordable only by large companies. Another solution is to individually test the best culture conditions for any promising cell line. Also this approach is successful, but the need to initially screen many clones is not avoided and it often requires a very specialized service company to execute this elaborate task. Finally, many attempts have been made to achieve higher gene expression levels, without the need for the time-consuming gene amplification process. Typically, heterologous DNA elements have been identified that are able to elevate gene expression levels, when flanking the gene expression cassettes. An overview of these gene expression augmenting DNA elements is given in Kwaks and Otte (2006)\textsuperscript{13}. Our group, for instance, isolated so-called STAR elements, genomic DNA sequences that were identified by virtue of their ability to counteract chromatin-mediated repression. When employed, these STAR elements strongly enhanced gene activity in a number of experimental set-ups\textsuperscript{14}.

Another problem that accompanies industrial production of biopharmaceuticals concerns the stability of protein expression over time. The isolation of any protein expressing cell line requires that transfected cells have to be selected from a pool of cells that in majority contains non-transfected cells. Usually this involves the use of toxic antibiotics that kill the non-transfected cells. This results in the creation of cell lines that are stably transfected, i.e. the plasmids that contain the gene expression cassettes are stably integrated in the genome of the cell. It is common procedure to analyze the resulting clones for their protein expression levels when they are still growing in the presence of the selecting antibiotics (or in the presence of toxic methotrexate). However, when these antibiotics are omitted from the cell culture medium, the protein expression levels will often decline to very low levels within weeks. This is a problem since cell line up-scaling and the subsequent production run usually last more than such a period of time. An alternative is to keep the antibiotics or methotrexate in the culture medium and this is often the only solution. It is, unfortunately, not a favored solution. Antibiotics can be very expensive and keeping the cells under selection pressure in a 10-20,000 liter bioreactor is very costly. Furthermore, the presence of antibiotics and methotrexate impede purification of the biopharmaceuticals during downstream processing and this is costly as well. This situation is a recurring theme. Whether the cell line has been produced after extensive gene amplification, or by antibiotic selection, stability of protein expression levels over time has remained a largely unresolved issue. An alternative approach is to screen many high-producing cell lines for their stability of protein expression in the absence of methotrexate and/or antibiotics and select the most favorable cell line for up-scaling for a production run. Obviously, this is a laborious and time-consuming screen.
2 **PREVIOUS WORK TO SOLVE THE PROBLEMS**

As indicated above, we previously identified a class of genomic DNA elements that are able to augment gene expression levels. These STAR elements stand for STimulatory and Anti-Repressor elements. They were identified in a genetic screen that employed chromatin-associated repressor proteins such as Polycomb and heterochromatin protein HP1. STAR elements were found by virtue of their ability to counteract the negative repression mediated by proteins such as HP1 and Polycomb. When STAR elements were subsequently utilized to flank gene expression cassettes, it was noted that gene expression levels were elevated. Also, they induced many more clones than plasmids without STAR elements. This constituted a problem, because again an extensive screening effort was needed to identify a clone that displayed high protein expression levels. The solution of this problem was to develop and employ a very stringent selection system. This implies that selection conditions on a cell are so harsh, that only cells that exhibit very high recombinant gene expression levels are able to survive these conditions. We called this system STAR-Select. In essence, a selection marker is placed upstream of the gene of interest, resulting in a bicistronic mRNA. The selection gene contains a startcodon with attenuated translation initiation frequency, whereas the gene of interest contains a startcodon with optimal translation initiation. In order to obtain sufficient selection marker protein in the context of the startcodon with attenuated translation initiation frequency, high mRNA levels have to be produced by the transfected clones to survive. Since the selection marker is part of a bicistronic mRNA, the protein of interest will simultaneously be translated to a high level, this from the startcodon with optimal translation initiation frequency. By flanking these stringent constructs with STAR elements we are routinely able to create a limited number of clones that express the protein of interest at high levels.

Unfortunately, it turned out that there are two major problems with the STAR-Select system. In the first place, the only modified startcodon for the STAR-Select system that proved to be useful, was a TTG instead of the optimal ATG. A much less stringent modified startcodon, such as GTG did not convey much advantage in terms of selection stringency, as compared with available selection systems. On the other hand, the TTG modification turned out to be too stringent for use in almost all tested cell lines, except in CHO-K1 and CHO-DG44. In suspension growing CHO-derivatives, the TTG Zeocin selection marker proved in fact to be too stringent, which is a problem for industrial application. Unfortunately, no translation initiation codons that might provide intermediate selection stringencies (between TTG and GTG) are available. As a consequence, the STAR-Select selection system turned out to be very inflexible and of limited use when it came to practical applicability.

A second problem accompanying the STAR-Select system was the stability of protein expression levels over time. As with other systems, protein expression levels dropped to low levels when the respective antibiotic was omitted from the culture medium. These were reasons that we started to search for other means to create a more flexible selection system that also provides inherent stability of protein expression levels.

3 **OUR CURRENT APPROACH**

3.1 **A more flexible selection system**

As indicated above, the STAR-Select system has several drawbacks for practical use, one being that the TTG Zeo selection marker is in practice the only translation initiation codon that can be used. We therefore sought to develop a selection system that is more flexible than the STAR-Select system. We describe two lines along which we approached the problem (Chapter
In short, we used small peptides that are placed immediately upstream of the Zeocin selection marker. Such small peptides are between 8 and 131 amino acids long and start with an optimal ATG, which is positioned out of frame of the selection marker. The small peptide is terminated by a stop codon, which is placed adjacent to the translation initiation codon of the downstream selection marker. The underlying concept is that the translation initiation machinery will start to translate the small peptide and will normally be terminated at the stop codon that is placed at the end of the peptide. However, when the peptide is not too long, the translation machinery will commence at the translation initiation codon of the downstream selection marker. The efficiency with which translation is reinitiated is dependent on the length of the preceding small peptide. When the small peptide is only a few amino acids long, the frequency of re-initiation of translation at the ATG of the downstream gene will still be high. When the upstream peptides become progressively longer, the frequency with which reinitiation of translation of the downstream gene will, however, decrease. When the downstream gene is a selection marker gene, this decreased frequency of translation implies less available selection marker protein. This in turn signifies a higher selection stringency of the system. When the small peptide approach is employed in the context of a bicistronic mRNA, this results in a higher availability of mRNA of a gene of interest, and consequently, higher expression levels of the protein of interest. This is a similar principle as the STAR-Select system. High bicistronic mRNA levels are created to warrant sufficient selection marker protein expression levels for cell survival, concomitantly with high expression levels of the protein of interest. Principally different from the (TTG Zeo) STAR-Select system though is that now a scale of stringencies can be employed. Simply increasing or decreasing the length of the small peptide respectively increases or decreases the selection stringency of the system. This provides much more flexibility than the previously designed STAR-Select system. The merits and drawbacks of this novel system will be discussed below.

3.2 Stability of gene expression: metabolic selection markers

One approach to obtain inherent stability of protein expression over time is the use of so-called metabolic markers. A metabolic selection marker protein is an enzyme that is involved in one or more essential steps in a metabolic pathway. With essential is meant that the cell is not able to synthesize such metabolic building blocks itself, which means that the building block has to be present in the culture medium in order to allow the cell to survive. Essential amino acids are well-known examples of such essential metabolic building blocks. Since mammals like us are auxotrophic and therefore cannot synthesize these amino acids, they need to be present in our food.

The enzymes that are involved in the synthesis of essential metabolic building blocks can be used as selection marker. The gene encoding this enzyme has to be present on the construct that contains the gene of interest. Provided that the essential metabolic building block is absent from the culture medium and a precursor of the building block is present, the enzyme that is now available to the cell will allow synthesis of the essential metabolic building block. Importantly, in principle no other selection agents need to be used. Omitting the essential metabolic building block from the culture medium is often not a problem, because many modern culture media are custom-made designed and manufactured by specialized companies. Addition of the precursor molecule can be more of a problem. It must be readily available commercially for ease of use. Secondly, it must be able to enter the cell. Finally, by preference, it must be non-toxic, else the same situation emerges as with antibiotics in that the toxic precursor must be carefully removed during downstream processing. However, if these conditions are met, the usefulness of metabolic markers is evident. No toxic agents are required in order to keep the cell under constant selection pressure. After all, the essential metabolic building block must be produced at all times, hence the selection marker protein
must also be produced at all times. When the marker is coupled to the gene of interest, for instance through a bicistronic mRNA, this warrants constant high expression levels of the protein of interest and thus stability of protein expression over time.

Not many enzymes fulfill all these criteria to be employed as a selection marker in CHO cells. For instance, the use of the enzyme that converts L-histidinol to L-histidine has been described as a metabolic selection marker. Unfortunately, the precursor L-histidinol is toxic, which reduces the suitability of this marker for industrial applications\textsuperscript{16}. On the other hand, the enzyme that converts indole to L-tryptophan has also been described as metabolic selection marker. The precursor indole is non-toxic and the selection system appears to be useful\textsuperscript{3,16}. Still, for reasons we do not understand, not many follow-up studies indicate that this metabolic marker has been successfully implemented by companies.

In this PhD thesis we present three examples of enzymes that are involved in the synthesis of essential metabolic building blocks in for instance CHO cell lines. We extensively tested the enzymes involved in the folate pathway (DHFR), L-threonine (thrB/thrC) and L-cysteine (CLase) synthesis for their usefulness as metabolic selection marker.

The DHFR enzyme is a well-know example of a metabolic marker. The DHFR enzyme is part of the folate pathway that converts dihydrofolate into tetrahydrofolate, which is required for the synthesis of purines (hypoxanthine), thymidyl acid (thymidine), and the amino acid glycine. As indicated above, CHO-DG44 cells lack the \textit{Dhfr} gene and they therefore need glycine, hypoxanthine and thymidine in the culture medium to survive. If, however, the \textit{Dhfr} gene is present on the expression cassette, the cell can convert dihydrofolate into tetrahydrofolate, provided that the metabolic end-products are absent from the culture medium and that the precursor dihydrofolate is present in the culture medium\textsuperscript{5}. As explained in Chapter 7, we used DHFR as selection marker in a different experimental set-up than the one that is usually applied. We did not induce gene amplification by methotrexate, to avoid using this toxic substance and the time and labor intensive rounds of gene amplification and subsequent subcloning. In Chapter 7, we show that when cells are initially selected by antibiotics and then switched to DHFR selection, this ‘passive’ use of the metabolic marker forces cells to maintain long-term transgene expression. However, one drawback that we encountered was the inability to isolate transfected cells by direct DHFR selection. Furthermore, we observed increased protein expression coinciding with reduced cell growth when cells with GTG or TTG \textit{Dhfr} variants were switched to DHFR selection. We concluded that we needed a means to reduce the selection stringency in order to apply direct DHFR selection and to obtain a better balance between stable protein expression levels and cell growth. Meanwhile, we isolated IRES mutants (Chapter 8) and developed the small peptide approach (Chapter 5), which might provide proper selection stringencies. Hereby, we may be able to improve our set-up of DHFR selection.

The enzymes that convert L-homoserine to the essential amino acid L-threonine were also tested for their applicability as metabolic selection marker. Mammalian cells are auxotrophic for these enzymes and they can thus potentially be used as selection marker. Also with this marker we encountered that direct selection of stably transfected clones in culture medium without L-threonine is very difficult (Chapter 6). With the enzymes that are involved in the synthesis of L-threonine, initial selection with antibiotics is necessary to obtain subclones. Then switching to media lacking L-threonine facilitates the maintenance of stable protein expression in the absence of antibiotics. Alternatively, switching to medium containing high concentrations of the precursor L-homoserine in the presence of L-threonine is another option to maintain stable expression. However, the necessity of initial selection with antibiotics limits the usefulness L-threonine as metabolic selection marker.

We reached another, more positive conclusion with the enzyme that is part of the metabolic pathway that converts L-cystathionine into the amino acid L-cysteine. The enzyme,
Cystathionine γ-lyase (CLase) is lacking from CHO cells and is thus a metabolic marker candidate (Chapter 8). We found that use of CLase as selection marker allows for the direct selection of stably transfected CHO cell lines. Below we will discuss the merits and drawbacks of this enzyme for use as metabolic marker in mammalian cells.

3.3 Gene expression augmenting DNA elements that induce more colonies than STAR elements

As we indicated above, one problem with the STAR-Select system that employs the TTG Zeo selection marker is that the stringency of this selection system results in the creation of only very few stably selected colonies or cell lines. This is not much of a problem when adherently growing cell lines are used, because non-transfected, dying cells are easily washed away. But it is a major problem when suspension cell lines are used. Few adherent colonies can easily be spotted on the culture plates. It is very hard, however, to recover the few surviving cells in a suspension pool of dying cells. Furthermore, the debris of dying cells is the opposite of an ideal environment for the surviving cells and this debris cannot easily be removed from a suspension cell culture. Therefore, it is has turned out that the STAR-Select system has been difficult to implement under industrial conditions which involve suspension cell lines. We therefore searched for means to improve the STAR-Select system. Our goal was to identify gene expression augmenting DNA elements that would induce more surviving cells that displayed at least equally high protein expression levels, as compared to STAR elements. This last aspect implies that the stringency of the selection system must be as high as the stringency of the STAR-Select system. Therefore we used the STAR-Select system as screening method to identify such potential elements. This was possible because the stringency of the STAR-Select system is that high, that without the inclusion of STAR elements in the constructs no cells survive at all. This therefore provided a rather clean background of the screens we undertook. As described in Chapters 3 and 4, we identified two types of DNA elements that fulfill the two criteria of a larger number of surviving cells that display high protein expression levels. In the next section we will discuss the merits and drawbacks of these elements for use as gene expression augmenting DNA elements in mammalian cells.

4 HAVE WE ACHIEVED OUR OBJECTIVES?

4.1 A more flexible selection system

As described above, we faced a problem with the previously developed STAR-Select system in the sense that it is very inflexible when it comes to applicability. The problem turns around the very stringent TTG translation initiation codon, which provides too high selection stringency, particularly when used in suspension cell cultures and other mammalian cell lines than CHO. As described in Chapter 5, we developed a novel selection system, based on the use of small peptides that are placed upstream of the selection marker gene. As described, this novel principle allows a larger range of stringencies, which can be simply ‘titrated’ by increasing or decreasing the length of the small peptide. There is, however, a limit to this system as well. Enlarging the small peptide to longer than ~100 amino acids does not further result in increasing selection stringencies. This was noted because the number of established colonies did not decrease anymore (a sign of increased selection stringencies) upon incorporation of longer small peptides (Chapter 5). Neither did the protein expression levels increase upon incorporation of increased peptide lengths. We therefore had to resort to adding an additional variation: Zeocin marker proteins with diminished functionality. We devised a genetic screen in bacteria to find Zeocin marker mutants that have a decreased ability to neutralize the available Zeocin in the culture medium. The screen, which was based on an Error Prone PCR (EPP)
protocol, resulted in the isolation of several Zeocin marker mutants that displayed higher selection stringencies in CHO-DG44 cell lines. When combined with small peptides, we were able to design a Zeocin-based selection system that showed a number of intermediate stringencies around the TTG Zeo selection stringency. In this sense, our efforts were successful and we achieved our goal of designing a far more flexible selection system than the STAR-Select system, with a range of stringencies. The current system allows testing and adjusting the selection stringency in multiple cell lines and under different culture conditions.

There is, however, a remark that has to be made. The inclusion of increasingly longer peptides encountered restrictions. The system is not that simple to apply that there is a linear range of selection stringencies, simply based on the length of the peptide. Above ~100 amino acids, the increasing lengths of the peptides have no more effect on the selection stringency. In that sense our search for Zeocin marker mutants with inherently higher selection stringencies is a weakness of the system.

Another question is why using the small peptides at all when a simple, Error Prone PCR protocol is sufficient to create numerous Zeocin marker mutants that also display a range of selection stringencies? The answer to this is not simple. The main reason we prefer not to use EPP Zeocin mutants alone is the consequence this has for the amount of protein the cell has to make. In case of the EPP Zeocin marker protein, simply more mutant Zeocin protein has to be made by the cell in order to obtain sufficient ability to neutralize the Zeocin agent in the culture medium. Generally, it has been observed that high producer cell lines have slower growth rates due to the metabolic burden thrust on them. In multiple experiments we noted that cells tended to grow slower when more severe Zeocin mutants were employed as selection marker, which consequently force more mutant Zeocin protein to be made by the cell (data not shown). In contrast, with the small peptide approach, ‘normal’ wild type Zeocin marker protein is being made. To achieve sufficient Zeocin marker protein levels, merely transcription is increased to achieve sufficient translation of the Zeocin marker, and this appears to be less of a metabolic burden on the cell. To minimize the amount of heterologous protein for the cell to make, we therefore chose to combine moderate mutant Zeocin marker proteins with very small peptides (8 to 14 amino acids long). We indeed found that when a moderately stringent EPP mutant Zeocin marker protein was combined with a very small peptide, little or no negative effects on cell growth were observed. This is the reason that we rather avoid using very stringent EPP Zeocin marker proteins alone and rather resort to a combination of such moderate mutant proteins and small peptides.

A final means to titrate expression of a selection marker is placing this marker behind a mutated IRES, as described in Chapter 8 for CI.ase selection. We already successfully combined the so-called IRES\textsuperscript{T} with a small peptide of 14 amino acids to express very little amounts of protein (data not shown). In this way, the combination of different IRES mutants, small peptides and mutant selection markers, creates an enormous variation in available selection stringencies.

Although we have not used this so far, the use of small peptides placed upstream of a gene of interest is not limited to selection markers. In principle, this approach can be used to modulate the expression levels of any protein of interest. This is in particular useful when applied to regulatory proteins of which only very small amounts are needed by the cell. The small peptide approach allows (within the noted limits) the titration of such proteins.

4.2 Stability of protein expression

As shown in Chapters 6 to 8, multiple enzymes that are involved in the synthesis of essential metabolic building blocks are candidates for use as metabolic selection marker. As has been discussed in the respective chapters, the main difficulties with these potential selection markers are 1) the killing curves of wild-type cells in culture medium devoid of a specific metabolic
building block and 2) obtaining the right selection stringency of the markers. Specifically with L-threonine, the killing rate of non-transfected cells in culture medium devoid of L-threonine was very low. That is, for instance, as compared to the killing rate of cells in the presence of the Zeocin agent. A direct result was that it was impossible to use threonine selection for direct selection of transfected cells. With this we mean that cells were selected in culture medium devoid of L-threonine, without additional selection agent like Zeocin in the medium. In this set-up, we were not able to select stably transfected colonies. This signifies that these enzymes were thus not practical for use as selection marker for direct selection. Also with DHFR as metabolic marker, we were not able to directly select transfected cells. However, as explained above, proper selection stringencies might improve DHFR selection.

Only when cells were first selected by means of Zeocin selection, the L-threonine and DHFR metabolic markers could be used as so-called ‘maintenance’ markers. This implies that after initial selection with Zeocin, established colonies could subsequently be switched to a culture medium devoid of Zeocin agent and also lacking the metabolic end products. The benefit of this procedure was that the protein expression levels remained very stable in the absence of toxic Zeocin in the culture medium. In that sense we can claim that we achieved the objective of stable protein expression, without the need for toxic and expensive antibiotics. However, this procedure is not very elegant and straightforward. Direct selection, without the need for changes in culture medium would be preferable. In that sense, neither the DHFR enzyme, nor the enzymes involved in the synthesis of L-threonine fulfilled this criterion in our experimental set-up.

Our best option remains the CLase enzyme that is involved in the synthesis of the amino acid L-cysteine. With this enzyme, direct selection of transfected CHO cells was possible. This is probably due to the fact that wild-type cells die rather rapidly in culture medium lacking L-cysteine. Since the killing rate is fast, stably transfected colonies clearly emerge from the background of untransfected cells, enabling the isolation of the transfected cells. Not only was direct selection possible, the protein expression levels in the established cell lines were high and remained stable over time. These characteristics indicate that CLase can be a useful enzyme to be employed as metabolic selection marker. Also the non-toxic precursor L-cystathionine is readily available and apparently is able to enter the cell.

Are there drawbacks to the use of CLase as selection marker? One problem we encountered was that the expression levels of the CLase protein had to be tightly regulated itself. When coupled to the modified translation initiation codons TTG or GTG we noted that these modifications were too stringent and no CHO-DG44 cells survived at all or cell growth was severely hampered. Only with an appropriate IRES mutant we were able to select stably transfected CHO-DG44 cells that combined high expression of the protein of interest with good cell growth. Our overall conclusion is therefore that tight regulation of the CLase expression levels is required to make this enzyme a useful metabolic selection marker protein.

A final remark concerning the use of CLase as metabolic selection marker is that L-cysteine is nutritionally considered to be a non-essential amino acid, because it can be synthesized in the livers of mammals. For cell culture applications, however, L-cysteine is considered an essential amino acid that must be supplied from an exogenous source. However, upon addition of the precursor L-cystathionine, various cell lines were able to grow in L-cystine deficient medium, indicating enzyme activity for the conversion of L-cystathionine to L-cystine18,19. This was not the case for Chinese hamster cell lines and certain continuous lymphoid and myeloid tumor cell lines of human and rodent origin18-21. This makes CLase very useful as metabolic selection marker for CHO-K1 cell lines. For the application of CLase as metabolic marker for other cell lines, one has to keep in mind that particular mammalian cells might be prototrophic and are able to produce L-cysteine, disabling the use of CLase as metabolic marker in those cell lines.
4.3 Preferred promoters

Initially, we undertook a rather systematic search to define the most optimal promoter to be used to produce proteins in CHO cell lines. Our analysis included a panel of 11 cellular promoters that we selected through literature searches. Apart from some structural aspects of the promoters, the main reason to include a specific promoter in our analysis was reported endogenous mRNA expression levels associated with these promoters\textsuperscript{22}. As shown in Chapter 2, we found that in most aspects that we tested, the human $\beta$-actin promoter performed best in CHO-K1 and CHO-DG44 cells. This was in terms of induced numbers of colonies and protein expression levels. Furthermore, we put a selected number of promoters to the test by subjecting the cells to harsh conditions, such as overgrowing culture conditions and serum free cell culture conditions. Under these conditions, the human $\beta$-actin promoter behaved best. Although a number of studies have evaluated the strengths of promoters in various cellular contexts\textsuperscript{23-31}, it is rather surprising that no such analysis in CHO-K1 and CHO-DG44 cells has been put forward in the literature and that the human $\beta$-actin promoter remains a relatively unknown in the biotechnology industry. The viral CMV promoter is still the most commonly used promoter for production purposes and is considered as a reliable promoter that confers high protein expression levels\textsuperscript{32}, although the strength of the CMV promoter can vary considerably in different cell types and under different conditions\textsuperscript{31,33-36}. Until now, we have not encountered such problems with the human $\beta$-actin promoter in our studies, and therefore this remains our preferred promoter.

4.4 The Rb1E/F genomic fragment as gene expression augmenting element

We used the $\beta$-actin promoter in the context of the stringent TTG Zeocin STAR-Select system to search for gene expression augmenting elements. This study was undertaken because of the noted difficulties with the TTG Zeo STAR-Select system to induce sufficient transfected cells, particularly when employed in serum free suspension cell cultures. To obtain stably transfected pools of cells that display high protein expression levels, more transfected cells surviving selection are required than there are induced by STAR elements. We therefore undertook the described screen to identify potential elements that 1) are able to induce more colonies in the context of the stringent TTG Zeo selection system than STAR elements, and 2) these colonies must also display high protein expression levels. We achieved our objectives by the identification of the Rb1 genomic fragments (Chapter 4) and the RPL32 promoter as second promoter to be placed upstream of the ‘primary’ $\beta$-actin promoter (Chapter 3).

The genomic fragments that we called Rb1E or Rb1E/F are able to induce at least two times more stably transfected colonies than STAR elements. We found this to be the case in the context of multiple reporter genes, different promoters and culture conditions. In Chapter 4 we showed that in the context of the stringent ttgZeo selection marker, incorporation of the Rb1E element resulted in increased numbers of stably transfected cell lines expressing d2EGFP or the therapeutic protein hEPO. Importantly, also with an expression cassette encoding for two different MAbs, the Rb1E element induced more stable cell lines as compared to STAR elements. The average MAb expression levels were equally high as the expression levels induced by the STAR configuration (Figure 1).

Most importantly, the Rb1E/F fragment induced more stably transfected suspension cells. In this sense, our objective to identify a gene expression augmenting fragment with more desired characteristics than STAR elements has been successful. What remains though is that we have hardly ideas how these genomic elements operate mechanistically. We excluded a number of possibilities. As a logical first step we tested whether we might coincidently be dealing with a STAR element. As we showed in Chapter 4, this appeared not to be the case. The Rb1E/F elements were neither enhancers in the classical sense, nor promoters. Another
In the context of the stringent TTG Zeo selection system, Rb1E induces more stable MAb-expressing CHO-DG44 clones than STARs.

CHO-DG44 cells were transfected with DNA of constructs as shown, either containing anti-RhD or anti-EpCAM. Constructs were flanked either with STARs or with Rb1E and control constructs lacked such DNA elements. The specific productivities are shown in pg/cell/day. Approximately 2 weeks after transfection, the numbers of stably established colonies were counted. The respective colony numbers are shown above the values of antibody expression.

Various studies revealed that most genomes are highly transcribed and a large amount of transcription occurs outside mapped protein-coding genes. The number of noncoding RNAs (ncRNAs) is much higher than previously predicted\(^{37-39}\). Some ncRNAs are processed into smaller products and are able to negatively regulate gene expression, like microRNAs\(^{40,41}\). However, there are also thousands of longer transcripts whose functions are still unknown\(^{42-47}\). These long ncRNAs include interlacing and overlapping sense and antisense transcripts and those derived from introns or intergenic regions and are most likely involved in various levels of gene regulation. Most of the well-characterized examples are long ncRNAs transcribed on the sense or antisense strand that interfere with transcription of a neighboring gene\(^{48}\). However, numerous studies also revealed examples of long ncRNAs involved in distinct processes such
as ncRNAs that facilitate chromatin modification, ncRNAs that modulate transcriptional activator or co-activator function and ncRNAs that control the formation or activity of transcription complexes. There are various examples of ncRNAs positively influencing gene expression. For instance, it seems that noncoding transcription is involved in the opening and maintenance of the active state of HOX clusters. Opening of the cluster at sites of activation of intergenic transcripts is accompanied by changes in histone modifications and a loss of interaction with Polycomb group repressive complexes. The transcription of ncRNA itself, not the actual transcripts, was proposed to be involved in the regulation of protein-coding RNA expression.

Non-coding transcripts have also been detected across the human β-globin locus control region (LCR), including the intergenic regions of the locus. Another study suggested that intergenic transcriptional elongation of the enhancer-assembled transcription machinery from the LCR into the cis-linked globin promoter is important for globin gene expression. In addition, a correlation was observed between intergenic transcription at different stages of developing human erythroid cells and acetylated chromatin domains. Since these acetylated domains are thought to be transcriptionally active in a developmentally regulated manner, these data also suggest a link between intergenic transcription and developmentally regulated opening of chromatin domains. Finally, deletion of an endogenous intergenic transcript between the γ- and δ-globin genes affects the endogenous transcription of the downstream β-globin gene. Additional examples of non-coding transcription correlating with (and causing) locus activation were described for the switch of Polycomb group response elements from a repressed to an activated state in Drosophila, in the LCR of the major histocompatibility complex II locus, in the T cell receptor locus, and upstream of the lysozyme gene in activated macrophages.

The above described examples of ncRNAs that are involved in regulation of gene expression, may shed light on our observations that the Rb1E/F fragment augments heterologous gene expression levels in transfected cells. This is also suggested by our finding that the endogenous Rb1E/F fragment might be a region of intergenic activity. This may play a role in enhancing the accessibility of the endogenous Rb1 chromatin locus and activation of the Rb1 gene. If this creates a particular chromatin configuration of the Rb1E/F genomic region, this in turn might give an explanation for the gene expression augmenting effects of Rb1E/F. Whether this is actually the case and how this intergenic transcription would function in augmenting heterologous gene expression could be a fruitful research topic.

4.5 The RPL32 promoter as gene expression augmenting element

The fact that we identified a promoter that was placed adjacent to another ‘primary’ promoter to have a beneficial effect on both induced colony numbers and gene expression levels, was a surprise. In many described cases, the existence of two adjacent promoters, either convergent, tandem or overlapping, has an overall negative effect on gene expression levels. Most commonly the term ‘transcriptional interference’ is attached to this effect, which refers to the direct negative impact of one transcriptional activity on a second transcriptional activity in cis. Reasons such as occupation of one promoter by RNA polymerase II (RNA pol II) precluding its occupation of the second promoter, competition for the same enhancer, dislodging of the RNA pol II complex by the arrival of an elongation complex and termination of the transcription progress because of converging elongation complexes have all been put forward for negative transcription interference. For example, transcription of a ncRNA across the promoter region of a downstream protein-coding gene can directly interfere with transcription factor binding, and thus prevent the protein-coding gene from being expressed. During growth of S. cerevisiae in a rich medium, the upstream SRG1 promoter is active and its activity
Figure 2. Rb1E, RPL32 and the combination thereof induce many anti-RhD expressing cell lines in the context of the pp8ZeoEpp5 selection marker. 

CHO-DG44 cells were transfected with DNA of the shown constructs, containing anti-RhD as the gene of interest. Stringency of selection is created by a small peptide of 8 amino acids and a ZeocinEpp5 marker mutant. For comparison, also the STAR-select construct was transfected. The specific anti-RhD productivities are shown in pg/cell/day. Approximately 2 weeks after transfection, the numbers of stably established colonies were counted. The respective colony numbers are shown above the values of antibody expression.

interferes with the tandem downstream promoter of the serine biosynthetic gene SER3, thereby blocking the unnecessary biosynthesis of serine59.

However, as described above in some examples of ncRNAs, the transcription process itself can be able to mediate gene activating effects. For instance, the very act of transcription itself is thought to be involved in the opening and maintenance of the active state of HOX clusters59. Another example of the transcription process being responsible for gene activation is shown in yeast, where intergenic transcription extending in the PHO5 promoter is required for nucleosome eviction and gene activation. Inactivation of RNA Pol II itself abolishes eviction completely. In contrast, increasing the level of the ncRNA generated in this region did not affect gene activation, indicating that the transcription process itself is the cause60. Mechanistically, transcriptional elongation causes a broad spectrum of effects to the underlying chromatin template, including chromatin remodeling, nucleosome eviction, and changes in the acetylation and methylation state of histone tails61-63, effects that are all due to the association of multiple enzymatic activities with the elongating RNA Pol II complex64-68.
The effects of transcriptional elongation are also proposed as possible explanations for the ‘chromatin opening’ functioning of ubiquitous chromatin opening elements (UCOEs), which are CpG island fragments containing two divergent housekeeping gene promoters. For instance, a UCOE placed upstream of the CMV promoter results in increased transgene expression. The authors suggest that chromatin remodeling to an open configuration during transcriptional elongation may explain the positive effects of UCOEs.

Can transcription from the RPL32 promoter result in an open chromatin structure, enabling the creation of many recombinant cell lines with high-level protein expression? What mechanistic basis underlies the ‘cooperative’ transcription we find when the RPL32 promoter is placed upstream of the primary promoter, is currently beyond our understanding. Our finding, however, clearly shows the beneficial effects on multiple ‘primary’ promoters we tested. Not only the β-actin promoter was positively influenced by addition of the RPL32 promoter, also the EF1α and viral SV40 and CMV promoters were influenced positively. The effects of RPL32 were completely abolished when RPL32 orientation was reversed or when the ‘core promoter’ of RPL32 was removed, indicating that active transcription in the direction of the ‘primary’ promoter is an absolute requirement for RPL32 to function as a gene expression augmenting element.

Not only did we obtain more stably transfected colonies with the RPL32--primary promoter configuration than with added STAR elements, also the protein expression levels in these colonies were high. This includes the expression of therapeutic MAbs (Figure 2). Hereby, our main objective to find gene expression augmenting elements that also allow more efficient transfection in suspension cell cultures was achieved.

5 WHAT COULD BE POSSIBLE FUTURE DIRECTIONS FOR THE DESCRIBED RESEARCH?

5.1 Incorporation of all described components into one integrated gene expression platform

The currently described components that are all parts of gene expression systems currently more or less stand on their own. Ideally, these components need to be integrated into one gene expression platform. What would this look like? The ‘primary’ promoter driving the gene expression cassette would be the β-actin promoter. We identified this promoter as being the most efficient promoter for the establishment of an optimal number of stably transfected cell lines with high protein expression levels. Selection would take place by means of the CLase enzyme, its expression level being titrated and regulated by the inclusion of a small peptide. The entire gene expression cassette would be flanked by a combination of the Rb1E/F fragment and the RPL32 promoter. How far are we from this ‘ideal’ gene expression platform?

Currently we use the β-actin promoter as favorable promoter. We also use the Rb1E/F fragment instead of the STAR elements as gene expression augmenting elements. Furthermore, we incorporated the RPL32 promoter in the configuration: Rb1E/F-RPL32--gene of interest--Rb1E/F. As shown in Figure 2, it appears that the combination of Rb1E/F and RPL32 delivers some advantage in terms of more efficient cell line creation and productivity of those cell lines. With the β-actin promoter driving the expression cassette with the therapeutic MAb anti-RhD and a moderate Zeocin mutant (EPP5) combined with a small peptide (of 8 amino acids) as selection marker, many recombinant DG44 cell lines were created when Rb1E/F, RPL32 or a combination thereof were employed. When no DNA elements were incorporated, only few stable cell lines were created (Figure 2). When both colony numbers and productivity are considered, the combination of Rb1E/F and RPL32 seems preferable. Our new selection
system combined with Rb1E/F and RPL32 delivers an advantage both in terms of colony number and productivity when compared to the ttgZeo selection marker combined with STAR elements (Figure 2). Specifically in the context of suspension cells, the combination of Rb1E/F and RPL32 delivers an advantage in the creation of stably transfected cell lines (data not shown).

One important goal of this research has been and still is to improve the difficult production of biopharmaceuticals, and in particular therapeutic MAbs. As indicated above, we have indeed been successful in creating stable cell lines that produce high levels of therapeutic MAbs, when either the Rb1E/F fragment or the RPL32 promoter or a combination thereof was used as gene expression augmenting element.

Currently, we use Zeocin as selection marker, specifically the described EPP5, EPP7 and EPP14 Zeocin marker mutants, combined with small peptides of 8 or 14 amino acids. These Zeocin variants turn out to be very useful in the context of suspension cell lines (data not shown). Our goal is, however, to replace the Zeocin selection marker with CLase as metabolic marker for direct selection in CHO-DG44 suspension cells.

5.2 Inducible gene expression systems

Throughout our studies we have encountered difficulties with cell growth, specifically when a difficult to produce MAb is used. For instance, it is feasible to achieve high specific productivities with MAbs, that is, per cell, a rather high amount of MAb is produced. Unfortunately, we noted that in some cases the cells do not grow well. In terms of volumetric productivities (the total protein production per volume) this is a very unwanted side effect. Whereas the cell may produce a lot of protein of interest, when growth is impaired, a 10,000 liter vessel will not be filled up with producing cells and this means that in effect only limited quantities of the protein are made. What is the reason for this and can we circumvent this problem?

The production of a complex protein such as a MAb or a fusion protein by a cell that has not been designed by nature for that purpose is a load for that cell. The entire secretion machinery of a CHO cell has not been made for the massive secretion of proteins. Furthermore, the production of high levels of heterologous protein is thought to constitute a metabolic burden for the mammalian cell. One consequence of this can be that the cell, which is forced by high selection pressure, will direct most of its energy toward the production of the recombinant proteins, away from putting its energy in cell growth. Although described for bacteria, this remains to be studied in more detail in eukaryotic cells.

It is, however, often observed that a mammalian cell culture that is characterized by high specific protein expression levels tends to grow much slower. Whereas wild-type immortalized mammalian cells may divide every 20 hours, highly productive cells may divide only once every 60 hours. This apparent inverse relationship between growth and productivity is reversed in controlled proliferation technology, which consists of conditionally arresting the cell cycle of production cell lines so that they can devote all of their metabolic energy typically used for biomass creation to production of protein pharmaceuticals. A transfected mammalian cell culture that was characterized by low specific protein expression levels was forced to stop growing. This was accomplished by the induction of cell-cycle arrest genes. Stopping cell growth simultaneously resulted in an increased cell specific expression level of secreted protein up to 15-fold. Several other studies confirmed that after cell cycle arrest the productivity of the cell increased.

This inverse relationship of growth and productivity indicates an intrinsic property of mammalian cells. And as such, this is not unique. This characteristic has been fully employed for protein production in bacteria, by using two components of the endogenous E.coli lac operon. A lac operator (lacO) is inserted within the promoter region that drives the gene of
interest. The Lac repressor (LacI) can bind to this operator and this results in repression of transcription of the gene of interest. Upon addition of IPTG, an allolactose analogue that binds to LacI, the affinity of LacI for lacO is lowered and transcription can proceed. It is common procedure in the laboratory to inoculate a bacterial cell culture in the absence of the inducer IPTG. No protein will be produced, but the bacterial cell culture can grow to high densities. Upon addition of IPTG to the culture, growth stops and the bacteria start to produce the protein of interest at a high production rate.

We envision that such an expression platform could be developed for mammalian cells as well. Such an expression platform differs from the above described proliferation technique in that it uses the protein production induced growth decline. Thus, in the first phase, protein production is repressed and cells will be able to grow normally. Since we use high-stringency selection systems, upon induction, high-level protein production almost by default results in reduced growth of the cells. The underlying idea is that an ideal production process would allow expansion of the cells until they have reached their optimal production phase, which is then preserved by the induction of a sustained growth arrest, induced by the metabolic burden of high-level protein expression. Especially when biopharmaceuticals are involved which are unstable, growth-impairing or cytotoxic, such an inducible system is indispensable.

It remains a major goal, however, to create an inducible expression system that 1) achieves high protein expression levels, 2) uses clinically inert inducer molecules, 3) uses a compact genetic design of an expression system that can easily be implemented in various cell lines, and 4) allows the distinction into phases of rapid cell growth and low protein expression, followed by a phase of slow cell growth but high protein expression levels. This phase with high productivity must also be sustained for a prolonged period during the production run.

6 CONCLUDING REMARKS

The research presented in this thesis aims to contribute to a more efficient creation of mammalian biopharmaceutical-expressing cell lines. We show that the strong human β-actin promoter, the Rb1E/F element and the RPL32 ‘secondary’ promoter enable the creation of many recombinant cell lines with high-level protein expression, this in the context of a stringent selection system. This is particularly convenient in case of recombinant suspension cell lines, which are difficult to create under conditions of a stringent selection system. Our novel selection system in fact allows the titration of the selection stringency, ensuring a suitable stringency for the creation of high-level expressing cell lines. Again, this is a feature that facilitates the creation of recombinant suspension cell lines, meanwhile ensuring high-level expression of the protein of interest.

An important advantage of the protein expression platform we developed is that cell lines expressing high levels of the protein of interest are created without the need for time-consuming rounds of gene amplification with subsequent laborious subclone screening. This is specifically useful for new biopharmaceuticals that rapidly need to be tested in (pre)clinical trials. Many protocols that are currently in use in the industry require the presence of toxic substance in order to maintain stable protein expression levels over prolonged periods of time. Here, we present an elaborate search for metabolic selection markers that ensure stable protein expression, without the need for toxic products. In particular the use of the CLase enzyme is very promising. Performing production runs in the absence of toxic compounds will help to reduce costs of downstream processing. In addition, the established stability of protein expression levels should result in increased product yield.
In conclusion, the research in this thesis potentially presents efficient means to create biopharmaceutical-producing cell lines, allowing for more economical biopharmaceutical production and thereby contributing to meet the current high demand on biopharmaceuticals.

Currently, all our experiments are performed at the level of the laboratory bench. This is inherent to establishing novel technology through academic laboratory research. The next logical step will be, however, to perform similar experiments in cell cultures that are scaled to volumes commonly employed in industrial bioreactors. To achieve success at this next level, joint efforts with an industrial partner will be needed. We believe this is worthwhile the effort, since our novel protein expression platform has the promise to reduce the currently very high prices of existing biopharmaceuticals. When successful, the current strong demand from society to keep the costs of health care acceptable, may be met.

REFERENCES


68. Balkrishnan, L. and Milavetz, B. Histone hyperacetylation in the coding region of chromatin undergoing transcription in SV40 minichromosomes is a dynamic process regulated directly by the presence of RNA polymerase II. *J Mol Biol* 365, 18-30 (2007).


