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

Citation for published version (APA):
Chapter 6

General Discussion
What have we achieved?

Identification of STAR elements, a novel class of genetic elements

In chapter one, the general introduction, we describe the problems that are encountered with the commercial production of recombinant therapeutic proteins. In essence, low predictability of creating protein-producing clones, low protein expression levels or yield and finally, instability of expression over time. In this thesis all of these problems have been addressed. First we will discuss what we have achieved in trying to solve these problems.

Our approach was novel in the sense that we envisioned that epigenetic silencing of gene expression plays an important role in the production problems. Our first action was to identify genomic DNA elements that are able to counteract these epigenetic chromatin-associated gene repression mechanisms and apply these elements to improve the expression of a transgene. This resulted in the isolation of multiple anti-repressor elements or STAR elements. In what aspects are STAR elements different from the other DNA elements that have been employed to enhance gene expression? Virtually all these DNA elements have been identified with a fundamentally different underlying idea in mind. For instance, MARs were initially identified as stretches of DNA that bind to the nuclear matrix, hence the name Matrix Attachment Regions. Only later these elements were tested for potentially beneficial effects on gene expression. In contrast with MARs, improving the stability of transgene expression was the goal and rationale to search for Ubiquitous Chromatin Opening Elements (UCOEs), which consist of divergently oriented promoter regions of ubiquitously expressed house keeping genes. Finally, the cHS4 element was identified by virtue of its ability to block the interaction between an enhancer and a promoter and hence is called an insulator, a rather broadly defined class of elements. From the start, the cHS4 element was claimed to convey position-independent expression of a transgene. This property of the HS4 element is thought to result in the induction of more transgene-expressing colonies. We have coined the term predictability for this property. In fact, there are only scant indications in the literature that the cHS4 induces higher gene expression levels as well. Our results (see also below), indicating that the cHS4 does not enhance gene expression, are therefore not surprising, when the original claims for the cHS4 insulator are closely examined.

One obvious question concerning STAR elements is whether they share sequence motifs with the other mentioned DNA elements. The answer is in essence no. The smallest DNA
element we compared is the cHS4 insulator, which encompasses an important recognition sequence for the transcription factor CTCF. The binding of CTCF to the cHS4 insulator is essential for the cHS4 to operate properly\textsuperscript{13,14}. No CTCF binding sites have been found in STAR elements. Another example is UCOEs, which encompass divergently oriented housekeeping gene promoters. No putative promoter sequences have been noted in STAR elements. Finally, MARs are defined by a list of properties, such as topoisomerase sites and long AT rich stretches\textsuperscript{2}. These properties have been incorporated in ‘search and identify’ programs that are able to recognize and even predict a MAR sequence. When STAR elements are run through such programs, they are not recognized as MAR. This short list indicates that STAR elements belong to a different class of genetic elements, at least at the level of the primary sequence.

Although the above emphasizes the differences between the DNA elements, they should not be overstated. As shown in chapter two, STAR elements, for instance act as powerful enhancer blockers as well. This indicates that the different elements share functional properties and that the differences between the elements are more artificial than we currently known. In this thesis the emphasis lies on the functional, operational effects of STAR elements on gene expression. As will be discussed below, there may, however, be a common underlying molecular mechanism involving chromatin structure. Changes in patterns of histone modifications may be of importance for the action of at least both the cHS4 insulator and STAR elements.

\textit{Predictability of expression}

Predictability of expression relates to one of the main problems of producing recombinant therapeutic proteins: Selection with traditional systems often results in many clones that do not express the recombinant protein at all. In case of the often-used co-transfection selection system this can be tentatively explained. In the first place, the selection gene is often coupled to a strong (viral) promoter. This results in high expression levels of the selection gene. However, most selection genes encode for enzymes that inactivate a toxic precursor in the medium. Even trace quantities of such enzymes are sufficient to assure cell survival. Driving the expression of these enzymes by strong promoters is therefore in fact very contra-productive! An example of such an enzyme is the often-used Neomycin resistance protein, which is considered to be a low-stringency selection marker (see below).
Secondly, the plasmid on which the resistance gene resides is often different from the plasmid that harbours the gene of interest (GOI). These two plasmids are therefore co-transfected, which in general results in lowering the selection stringency of the system. Why is that? The selection gene containing plasmid is commonly co-transfected in low molar ratios compared to the plasmid with the GOI (eg. selection gene: GOI 1:10). The reason for using these ratios is because it is taken for granted that this co-transfection warrants co-integration of both plasmids into the host genome. Overabundance of the GOI would not only guarantee the presence of this gene, but also result in higher expression levels of the GOI in comparison to the selection marker. There is no need for the cell to have more than one or few copies of the selection gene, as explained above. These few copies are sufficient to provide all the necessary selection protein to overcome the selective pressure. However, it is still possible, and actually quite common, that a cell incorporates only the resistance gene and no GOI, resulting in resistant clones that do not produce any protein of interest. In practice, there will usually be relatively low copy numbers of the GOI and ultimately low expression levels. Therefore, although still most commonly employed, co-transfection of the GOI with an enzymatic selection marker is not very effective.

The use of a non-enzymatic selection marker is already a major improvement. The Zeocin resistance protein, for instance, forms a dimer that stoichiometrically binds one Zeocin molecule. To effectively neutralize all Zeocin molecules, many more Zeocin resistance protein molecules have to be produced than if it were an enzyme such as the Neomycin resistance protein. This more stringent selection will favour integration of more copies or at better genomic locations also leading to more copies of the GOI.

Another improvement is not to co-transfect the plasmids encompassing respectively the selection marker gene and the GOI. Instead of using two separate plasmids to express the GOI and the selection marker, a single plasmid expressing both genes can be used. Obviously, such plasmids are more difficult to make, because they are larger. Although being an improvement over the co-transfection system, this system still does not warrant a high degree of co-expression of both genes. This is because the GOI and the selection marker still form two independently regulated expression cassettes.

The only way to guarantee co-expression of both genes is to express both proteins using a single bicistronic mRNA. This mRNA results from coupling the gene of interest to the selection marker gene through an Internal Ribosome Entry Site (IRES) sequence. Now the selection marker is not independently regulated from the gene of interest. A warning is appropriate though. The use of the IRES can result in high expression levels of for instance
the Neomycin resistance protein, something one would like to avoid, as explained above. Therefore a so-called attenuated IRES is often used. This IRES still uses the benefits of a bicistronic mRNA, but it employs a lower translation initiation frequency of the second gene, in this case the selection marker gene. Particularly the use of an attenuated IRES, in combination with the stringent Zeocin selection marker, is very effective and results in the creation of a very stringent selection system. However, companies like Invitrogen still sell the Neomycin resistance gene, placed behind a strong IRES sequence.

We elaborated on these insights in chapter two, where we initially incorporated the STAR elements in plasmids that were co-transfected with a plasmid that contains the Neomycin resistance gene. Since we observed many more colonies, due to the incorporation of the STAR elements, this allowed a more stringent selection to obtain high-producing clones and simultaneously eliminate the low-producing clones. The change from the co-transfected Neomycin resistance gene to the Zeocin resistance gene, coupled through an attenuated IRES sequence greatly improved the effectiveness of the system. The vast majority of non- and low-producing colonies were automatically omitted from the screens and subsequent analysis. Apparently, these low-producing colonies did not produce enough Zeocin selection marker protein to eliminate the Zeocin selection agent effectively. The colonies that survived indeed displayed higher protein expression levels than with the co-transfection selection system (see chapter two). Following up on these insights and results, we developed the STAR-Select system. As described in Chapter four, we achieved that only a very limited number of colonies now survive, usually about tenfold less than with the attenuated IRES-Zeocin system. This is not only a significant improvement over the already stringent IRES-Zeocin selection system; it is a huge improvement when compared with the co-transfection methodology that is still often used.

We are not aware of comparable selection systems that convey similar high selection stringencies. Probably closest in stringency are the GS-CHO selection system\textsuperscript{15,16}, developed by the biotechnology company Lonza, and the dhfr intron-selection system\textsuperscript{17}, developed by Genentech. Both systems employ metabolic markers for selection. The GS-system was developed in NS0 cells that miss the enzyme to synthesize the amino acid glutamine. Only when a plasmid is transfected that harbours the glutamine synthetase (GS), the cell can survive. This system has been claimed to be very stringent. One drawback is, however, that only NS0 cells miss this enzyme. All other mammalian cells have at least one intact allele of the GS gene and the resulting GS protein needs to be repressed to be useful as selection marker.
The dhfr intron-system utilizes differential splicing to produce two proteins from a single bicistronic primary transcript. Most of the RNA is being spliced and translated into the GOI, a small percentage of the primary transcript however is not being processed and translated into the dhfr protein. This system was developed in DUXB11 cells (dhfr-) and in close analogy to the GS-system it requires inhibition of endogenous dhfr to operate in other (dhfr+) cell lines.

In summary, many selection systems are available to create mammalian cell lines that harbour stably integrated transgenes. Although it might appear that the use of stringent selection systems encompasses all the advantages of having to screen less clones to find higher producers, use of the least stringent selection system of all, co-transfection of the Neomycin selection marker, is still common. Also ‘vector’ companies such as Invitrogen still support and sell many selection systems that are suboptimal, to say the least. Below we will elaborate on the advantages of using the highest selection stringency possible.

Selection stringency, an important key in solutions for production problems
As explained above, the most apparent, early noted effect of STAR elements was the induction of up to tenfold more surviving colonies. This introduced the problem that to find THE highest producing clone, screening of even more clones than without STAR elements was required. Initially we solved this problem by creating a bicistronic mRNA in which the gene of interest was coupled to the selection marker through an attenuated IRES sequence (chapter two and the discussion above). This created a more stringent selection system and resulted in killing of the majority of emerging clones.

In theory, the selective pressure of this system can be increased unlimited. One could just raise the concentration of selection agent in the culture medium. Particular with the stoichiometrically operating Zeocin selection system, adding more Zeocin molecules to the medium would just outnumber the Zeocin resistance protein molecules. However, probably due to toxicity of the Zeocin selection agent, this approach has its practical limits. Often we have observed that cells appear to revert to become resistant to Zeocin, without producing any protein of interest at all (unpublished observations). What lies at the basis of this effect is unknown, but it certainly sets a limit to endless raising concentrations of Zeocin, or any other selection agent for that matter.

Compared to the traditional co-transfection approach, with the IRES Zeo selection system we observed a significant reduction in the number of emerging colonies and a marked
increase in protein expression levels. To elaborate on these findings, we designed a novel selection system, called STAR-Select. The STAR-Select system employs translation initiation principles that have never been used in such applications before. This makes the STAR-Select system a novel concept. Although in chapter four only the use of the Zeocin selection marker is described, the STAR-Select system has also been successfully used with other, more commonly used markers such as Neomycin and Blasticidin (unpublished results). This makes the STAR-Select system versatile and broadly applicable. One application, for instance, is to place two different selection markers upstream of the light and heavy chain of a monoclonal antibody. Such selection on two independently regulated genes, allows the isolation of even fewer clones that display higher antibody expression levels (unpublished results). One further advantage of the novel STAR-Select system is that the stringency can easily be modulated. This is important because every cell line has different sensitivities towards specific selection agents. In fact this complicates easy expression of recombinant therapeutic proteins in the diverse range of different cell lines that are in use by biotechnology companies. It is unavoidable that every different cell line needs to be titrated for a particular selection agent. In case a cell line is too sensitive for a specific selection agent to allow efficient formation of enough clones for analysis, this is an inherent problem of this cell line. In the STAR-Select system, however, the selection marker can be changed itself. Instead of the described TTG Zeo modification in the STAR-Select system (see chapter four), now for instance a less stringent GTG Zeo modification (the GTG triplet has a higher translation initiation efficiency then the TTG triplet) can be used. This in turn allows the use of higher Zeocin concentrations in the culture medium, which can be useful to find an optimum selection regime for the particular cell line.

Beside the practical utility of the novel STAR-Select system, its most important fundamental implication is that it shows the power of STAR elements at their best. The initial observations after application of STAR elements were the increased number of colonies. Even in equal numbers of colonies picked and analyzed, we found small but consistent increases in protein expression levels. Without performing large-scale screens and analysis of thousands of colonies it would have been impossible to determine the factor by which STAR elements can enhance protein expression levels. By eliminating the vast majority of low-producing colonies, we kept only a small population of colonies that indeed showed large increases in protein expression levels. Hence we conclude that in the context of the STAR-Select system the positive effects of the STAR elements on gene expression become most clear. This statement can even be extended to MAR elements. Although these
elements have been around for a long time, reports on beneficial effects on protein expression levels have been conflicting\textsuperscript{5,6}. We show (chapter five) that MARs indeed have a large, positive effect on protein expression levels. As with STAR elements, these effects become, however, only apparent in the context of the very stringent STAR-Select system.

\textit{Yield improvement.}

Where do we stand in respect to increasing protein expression levels? The easiest answer is to show a comparison with and without STAR elements in the context of several selection systems. In Figure 1, d2EGFP expression levels in stably transfected clones are shown. Take for example the two extremes: co-transfection without STAR elements as compared to STAR-Select with STAR elements. The average d2EGFP fluorescence level in colonies emerged after co-transfection without STAR elements is 11. The average d2EGFP fluorescence level in colonies emerged after transfection with the STAR-Select system, with STAR elements is 971, and increase of a factor 88. Although these effects are enormous, care should be taken with its interpretation. For one, d2EGFP fluorescence data cannot be simply extrapolated to real, secreted recombinant therapeutic proteins. For instance, as discussed in chapter five, marked increases in EPO expression levels were achieved but it is apparent that these are no ‘factor 88’ increases. Obviously, other parameters such as proper protein folding and/or secretion may limit the production of secreted proteins. This said, however, our reported EPO values are still high, in comparison with values that have been reported in the literature\textsuperscript{18-20}. Specifically, the EPO expression levels after dhfr-mediated gene amplification are still lower than what we report. It is significant that our values are achieved after one round of transfection and selection, whereas dhfr-mediated EPO production required multiple rounds of amplification and subcloning, as outlined in the introduction. In conclusion, even when compared to the best systems available up to now, GS- and dhfr-mediated gene amplification, STAR-Select stands out positively. Also expression levels of a monoclonal antibody in CHO-K1 cells are high and amongst the highest levels that have been reported in the literature (unpublished results).

One potential explanation for the high protein expression levels was that application of the stringent STAR-Select system might lead to these high expression levels because of the introduction of many more gene copies, as is the case with dhfr-mediated gene amplification. One surprise was that this does not seem to be the case. It appears that overall the numbers of genes that are stably incorporated remain well under 50. This is low
in comparison with dhfr-mediated gene amplification, which can involve thousands of gene copies. It implies that the expression level per gene copy is higher than that per gene copy after dhfr-mediated gene amplification. In fact, even in the high-producing clones we observe copy number dependent gene expression. This result also provides strong evidence

Fig 1. Frequency histograms of GFP expression in stable clones minus and plus STAR-elements in various selection systems.
Up to 100 stable clones minus and plus STAR-elements are analyzed for GFP expression. Selection was performed by either co-transfection of the selection marker (A), IRES coupled selection (B) or the novel STAR-select system (C). Significantly higher expression levels were seen in all three selection systems when STAR-elements were present compared to control constructs without STAR-elements. Increased stringency of selection leads to higher expression levels, with STAR-select being the most stringent selection system.
for the positive effects of STAR elements on gene expression levels. If similar gene copy numbers had been found as with dhfr-mediated gene amplification, no positive effects by the STAR elements on the promoter would have been required. Higher gene copy numbers alone would have explained the higher expression levels. Instead, STAR elements in the STAR-Select system strongly enhance promoter activity to allow the introduction of much less gene copies than with dhfr-mediated gene amplification. This enhanced gene activity results in high enough TTG Zeo protein levels for the cell to survive.

A final remark is in place when it comes to the combination of STAR elements and the use of chromatin-opening enzymes such as the histone acetyltransferase (HAT) p300. We noted an additive effect of STAR elements and targeted LexA-p300HAT to the promoter. These experiments were performed in the IRES Zeo selection system. It would be interesting to see whether combining targeted p300HAT also has an additive effect in the context of the STAR-Select system. One could argue that the STAR-Select system is so stringent, that only optimal-context clones can survive the selection. The effects of the opener enzymes described in chapter three could well be limited to those clones with a ‘mediocre’ site of integration: a non-optimal chromatin context around site of integration that still warrants high enough protein expression to survive the already stringent IRES-Zeo selection.

**Effects of several DNA elements on protein expression: a comparison**

In chapter one, we outlined the claimed effects on gene expression of several DNA elements that have been identified and employed in recent years. In table 1 we repeat this and add the results with STAR elements described in this thesis. STAR elements fall in the same functional category as MARs and UCOEs, in the sense that they induce increased promoter activity, and stability of expression, in a cell type and promoter independent

<table>
<thead>
<tr>
<th>Element</th>
<th>Size</th>
<th>Increased expression</th>
<th>Stability of expression</th>
<th>Cell type specific</th>
<th>Promoter independent</th>
<th>Copy number dependency</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCR$^{21}$</td>
<td>16 kb</td>
<td>?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>cHS4$^{22,23}$</td>
<td>1.2-2.4 kb</td>
<td>?</td>
<td>Yes</td>
<td>?</td>
<td>?</td>
<td>No</td>
</tr>
<tr>
<td>UCOE$^{8,9}$</td>
<td>2.5-8 kb</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>MARs$^{5,6}$</td>
<td>~3 kb</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>STAR$^{1,24}$</td>
<td>0.5-2 kb</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

(?: Unknown or inconsistent results) From: Trends in Biotechnology 2006 Mar;24(3):137-42
fashion. In contrast to MARs and UCOEs, we describe in chapter 2, 3 and 4 that STAR elements also convey copy number dependent gene expression. Hence it appears that STAR elements form a different class of genetic elements than MARs and UCOEs. This point is extensively discussed in chapter five, where we directly compare several DNA elements with STAR elements. For instance, it appears that DNA elements such as the cHS4 insulator, UCOE and MARs are no anti-repressor or STAR elements, as defined by the anti-repressor assay we described in chapter two. This indicates that these elements operate in a different manner. STAR elements were identified and isolated by their virtue to survive in a selection system that uses chromatin-associated repression. In contrast, Matrix Attachment Regions (MARs) are DNA elements that bind to the nuclear matrix and are thought to be involved in creating higher order structuring of chromatin in topologically independent loops. As such MARs may have a predominant structural function in gene organization. Ubiquitous Chromatin Opening Elements (UCOEs) consist of divergently oriented promoters of ubiquitously expressed housekeeping genes. It can therefore be envisioned that potentially direct effects of UCOEs on promoters differ mechanistically from more indirect effects of MARs on gene regulation through a structural, gene-organizing manner. From chapter five it also becomes clear that the MARs and STAR elements have both profound positive effects on gene expression levels. In our hands neither the cHS4 nor the UCOE had such effects. It is in particular clear that the novel STAR-Select system is beneficial for both MARs and STAR elements. As outlined in chapter five, the system the biotechnology company Selexis uses is a complex co-transfection protocol in which MARs are co-transfected and super-transfected in trans. Although we could repeat part of these results, this protocol only works properly in the context of this low-stringent co-transfection system. Another way of representing the data in chapter five is to draw a direct comparison between what other groups and Selexis claim, and what we achieved in the context of STAR-Select. This again reinforces that the best operating DNA elements are MARs and STAR elements. However, both should be used in the context of STAR-Select to gain the largest increases in gene expression.

Finally, how could one graphically present the effects of STAR elements, either when used alone or in the context of STAR-Select? Although we currently do not know the underlying molecular basis of the positive action of STAR elements, we can say a lot more about what they operationally do. As shown in the upper panels of Figure 3, protein expression levels are limited due to the stringency of the selection system. Raising the selection stringency through the addition of more selection agent to the culture medium eventually has its limits.
and usually results in cell death (upper, left panel). Application of STAR elements somehow changes the form of the curve (lower, left panel). With similar selection stringencies, protein expression levels are now raised from 1 to ~2.4 (arbitrary units). With STAR-Select, however, the selection stringency is raised, and now protein expression levels of ~5 can be reached.

---

Fig 2. Comparison of various expression augmenting DNA elements. Up to 24 stable clones were tested for each construct.

(A) Modest effects of the expression augmenting DNA elements are seen in a low stringency selection system (co-transfection). (B) Some, but not all, DNA elements show marked increases in GFP expression levels when used in the STAR-select system. The dotted line compares the levels of expression with and without STAR and STAR-select technology. Compared to other expression augmenting DNA-elements, STARs operate best in the classical co-transfection system. Both MARs and STARs operate very well in the STAR-select system, increasing expression levels up to tenfold compared to the control.
In a similar fashion, one can analyze colony numbers. It is a well-established fact, which is also demonstrated several times in this thesis that lowering selection stringency will result in the emergence of more colonies (upper, right panel). Here again the form of the curve is changed by introduction of STAR elements (lower, right panel). With similar selection stringencies, now more colonies emerge (arbitrarily from 1.4 to 2.2), as has been described in this thesis. However, raising the selection stringency with the STAR-Select system lowers the relative number of emerging colonies to 1.

*Fig 3. Schematic action of the STARs and the STAR-Select system.*
Both the level of expression (A) and the number of colonies (B) is dependent on the stringency of selection. Using the same stringency, shielding the transgene with STAR-elements increases both the levels of expression (C) and the number of colonies (D). Increasing stringency of selection will eventually lead to cell death (A,B, C and D). In the novel STAR-select system, unprecedented levels of stringency can be achieved which will lower the number of colonies but heighten the expression levels (C and D, Red arrows).
**What has not been solved?**

Until now we have not discussed the important issue of stability of gene expression in relation to the application of STAR elements. As shown in chapters two and three, we observed increased stability of gene expression when STAR elements were applied, either alone or in combination with targeted LexA-p300HAT protein to the promoter of the reporter gene. However, we faced a fundamentally different situation when we tested stability in the context of the STAR-Select system. The vast majority (>95%) of the tested colonies showed loss of the extremely high expression levels (Figure 4C). However, the expression levels did not drop back to zero, as others and we have often noted when no STAR elements were incorporated in the construct at all. Instead, when the selection agent (Zeocin) was removed from the culture medium, the average expression levels in the clones went back to ~30-45% of the original expression levels. These expression levels remained stable over prolonged periods of time. Interestingly, these expression levels became comparable with those that we achieved with STAR elements in the context of the IRES Zo selection system described in chapters 2 and 3. This could point towards a mechanism in which the levels of stable expression in the absence of selective pressure are restricted to an upper limit, for instance to ensure fast enough cell growth. This situation could well be compared with a steady state level of expression, albeit without the absolute need for the cell to produce the exogenous proteins. There were, however, some puzzling observations in this regard. When only low concentrations of selection agent Zeocin was present in the culture medium, the expression levels did remain stable over a long period of time (Figure 4B). The Zeocin concentration could even be lowered to a point that wild type cells would no longer die due to the toxicity of the Zeocin. It is striking that the potent performance of the STAR-Select selection system, can be maintained by sub-lethal doses of selection agent. Furthermore, when Zeocin was added to clones that had lowered or lost their expression upon removal of selective pressure, we observed a significant amount of cell-death. However, a large proportion of the cells survived and regained exactly the same high expression levels as before the selective pressure was removed. This points to a certain flexibility in the system. If, for instance, all cells would harbour selection genes that were silenced by heterochromatization after removal of the selection pressure, re-activation of these genes might be very difficult, if not impossible. This would inevitably lead to death of all cells. This notion was reinforced by other observations. When the integrated copy numbers of the constructs were monitored, loss of expression after removal of selection pressure did on average not result in loss of gene copies. Instead, the expression per copy
became lower. When selection pressure was reintroduced, the copy number did on average not increase, but the expression per copy was increased (unpublished observations).

Taken together, these observations indicate that instability of expression in most clones is not due to loss of gene copies or permanent silencing. Instead the system shows a remarkable flexibility in the sense that the promoter activity is modulated by the presence of absence of selection pressure. Since only a low activity of the promoter is required to produce enough Zeocin resistance protein for cell survival, it appears that the only requirement for long-term stability is a low, but constant level of transcription through the expression units. There are precedents for such a scenario. It was shown that intergenic transcription in the β-globin gene cluster is required for chromatin remodelling and β-globin gene expression\textsuperscript{25}. Induction of this intergenic transcription in non-erythroid cells led to the expression of the erythroid specific β-globin genes\textsuperscript{26}.

**Fig 4. Improved stability using sub-lethal doses of selection.**
The average expression of 24 stable STAR-select clones (containing STARs 7 and 67) was followed for 105 days after transfection. Initial selection took place under normal conditions (150 Zeo). Three weeks after the transfection, expression levels were measured by FACS analysis and the clones were splitted and grown on three selection conditions. Complete removal of selective pressure led to a marked drop in average expression levels. A sub-lethal dose of selection (25 Zeo) is enough for expression levels to remain at the same high level as under normal selection conditions (150 Zeo), expression levels remain stable for at least 3 months.
How does it work?

Throughout this thesis the important topic of how STAR elements work has been treated as an enigma. What we and others have shown is that STAR elements are functionally highly conserved. These human DNA elements function as anti-repressors in a human cell-line, they improve transgene expression in both human and hamster cell-lines and they protect transgenic constructs from silencing in transgenic mice\textsuperscript{27}. Some of the STAR elements are also highly conserved on the DNA sequence level. For instance, the human STAR 40 (chapter two) shares a 75\% overall identity with mouse STAR 40. This is in particular high since the DNA is non-coding genomic DNA. The mouse homologue was purely identified by its sequence homology with human STAR 40, however it has proven to be a bonafide STAR element. The genomic location of this STAR element is also conserved, in both the human and mouse genome STAR 40 is located in close proximity of the Interleukine 17 receptor gene. All these features are even more outspoken with another STAR element, STAR 18 (Unpublished observations). This STAR element is located in the human HOX D cluster on chromosome 2q31.2 and is 79\% identical to the corresponding mouse STAR18. What is of more interest though, is its precise location. During evolution the homeotic Antennaepedia and Ultrabithorax gene clusters in insects such as Drosophila melanogaster have been fused together in the single, large mammalian HOX cluster. Later on in evolution this cluster has been duplicated twice, resulting in 4 HOX clusters. It is precisely at the junction of the ancestral Antennaepedia and Ultrabithorax gene clusters where this fusion has taken place that STAR 18 is located. We have not found any STAR18 homologues on the HOX A,B, and C clusters, nor in lower animals such as Drosophila melanogaster indicating that STAR 18 has emerged relatively late in the evolution of the HOX clusters Since the HOX clusters are primary targets for the chromatin-associated Polycomb group repressors, it is tempting to speculate that this STAR element has an important function in delimiting the action of Polycomb group proteins in (negatively) regulating the HOX cluster. Another example is found directly upstream of the Hox D4 gene. STAR 1 encompasses a previously identified retinoic acid response element of the Hox D4 gene. These examples may point towards an important, evolutionary conserved function of STAR elements in Hox gene regulation, but they tell nothing about potential working mechanisms. What can those be?
Fig 5. Schematic overview of the human Hox clusters and the location of STAR18.
(A) The human genome contains four clusters of homeobox genes (Hox genes). The order of the genes reflects the regions in which they are expressed along the anterior-posterior axis, with group 1 genes being the most anterior and group 13 genes the most posterior. The arrows indicate the respective BAC-clones used to make a genomic Hox library. This library was used to screen for the presence of STAR elements. The Hox genes are aligned with the respective fly homologues. The *Drosophila* genome contains two homeotic clusters, the bithorax and antennapedia complexes. In vertebrates, these two ancestral loci have fused into one complex. In mammals, this joined complex has duplicated twice, resulting in four Hox clusters. Indicated in red, STAR18 is located at the junction of these two ancestral loci on the Hox D cluster.
(B) A high degree of homology is seen when the human STAR18 sequence is aligned with the respective mouse sequence. Red boxes on the mouse element represent regions of >80% sequence identity. The unusual high degree of conservation continues for about 500bp upstream of the identified STAR18 element, indicated by the grey boxes.
(C) Both human and mouse elements showed potent anti-repressor activity when tested in the STAR screen. The relative strength of the fragments (+ or ++) was based on the number of surviving colonies and their relative growth rate.
The most obvious mode of action may be that STAR elements constitute a specialized form of chromatin structure. This property then allows a promoter to work better. This one property appears to be the only certain property of STAR elements. In chapter two we claimed that STAR elements counteract chromatin-associated repression. In the employed selection screen chromatin-associated proteins were targeted to the promoter driving the Zeocin resistance gene. Only colonies survived that harboured a STAR element, placed between the targeted chromatin-associated repressors and the promoter of the Zeocin resistance gene. However, we also observed that if the SV40 promoter did not drive the Zeocin resistance gene, but the more powerful CMV promoter, the targeted chromatin-associated repressors were not able to sufficiently repress the Zeocin resistance gene. Apparently, the cell is then still able to produce sufficient Zeocin resistance protein to survive.

This allows an explanation for the action of STAR elements other than that they only counteract chromatin-associated repression. In this explanation STAR elements enhance the SV40 promoter to become approximately as strong as the CMV promoter, with the operational effect that the Zeocin resistance gene does not become sufficiently repressed. This eventually results in cell survival. This explanation is entirely feasible if STAR elements turned out to be pleiotropic enhancers that activate the SV40 promoter. In chapter two we already tested and excluded this possibility. STAR elements do not operate as an enhancer element in the classical sense. Enhancers operate in a transient fashion and in transient assays STAR element definitely did not act as enhancers.

This leaves, however, open the possibility that STAR elements do act as sort of enhancer, but only in a setting that involves proper chromatin structure. This can consist of STAR containing constructs that are stably integrated in the chromatin of the host cell. To our knowledge no such enhancer-like elements have been identified that can exert their action only in the context of stable integrants. So even if STAR elements exert a direct effect on the promoter and do not directly counteract chromatin-associated repression, this effect is novel. It again points to a special chromatin structure of the STAR elements.

Different levels of chromatin structure can be discerned here. The most elegant explanation would be that STAR elements contain DNA that forms chromatin, or more precise, nucleosomes that do not allow particular histone modifications to occur. These modifications could be methylated H3K27, a docking site for the Polycomb protein, or methylated H3K9, a docking site for HP1. If this were the case, this would explain that STAR elements do not allow formation of ‘heterochromatin’ like structure encompassing
Polycomb or HP1 proteins. This would in turn facilitate a promoter to become or remain more active in a chromatin surrounding full of Polycomb- or HP1-associated chromatin stretches. We do, however, not know whether chromatin stretches exist at all with the property that specific histone modifications do not occur. We do know from very recent work that stably integrated STAR elements in transfected constructs indeed have elevated levels of histone acetylation. Vice verse, STAR elements are relatively poor in methylated histone H3K27 and H3K9 methylation (unpublished observations). These results indicate that significant modulation takes place at the level of histone modifications in the STAR elements. The histone acetylation patterns that we observe in stably integrated STAR elements are not unlike the histone acetylation patterns that are observed in and around the stably integrated HS4 insulator. Mutskov et al. have shown an enrichment of histone H3 and H4 acetylation levels on a transgenic construct harbouring the HS4 insulator compared to a HS4-less control. Furthermore, they show a marked decrease in both histone H3 methylation and DNA methylation when the transgene is insulated by the HS4 element. In the presence of Trichostatin A, a histone deacetylase inhibitor, also uninsulated transgenes remain relatively stable upon removal of selective pressure. Confirming that deacetylation of histones is the trigger for silencing. The authors conclude that the HS4 insulator prevents transgene silencing by inhibiting histone deacytelases or and/or by recruiting histone acetylases. Other potential explanations for STAR action are localization models. It has been shown that inclusion of MARs or insulators such as gypsy in a construct results in a re-localization of the entire expression cassette to the periphery of the cell nucleus. This has led to the hypothesis that such DNA elements are involved in targeting expression cassettes to compartments in the nucleus that are more favourable for gene expression. While this is possible, we have no indication whatsoever that STAR elements operate in such a manner.

Finally, what are the possibilities that STAR action is defined through the binding of specific proteins? The action of the HS4 insulator, the scs/scs' boundary elements and the gypsy insulator element all require specific DNA binding proteins for at least part of their action. The enhancer blocking activity of all of these elements was shown to be dependent on binding specific proteins (CTCF, Zw5, Beaf-32 and Su(Hw), respectively). Omission of the specific protein binding sites or mutation of the insulator binding proteins through genetics destroys the enhancer blocking activity of these insulators. Since STAR elements also display enhancer blocking activity, one might expect this function to be dependent on
protein binding. There are, however, indications that plead against the notion that STAR elements operate on the basis of specific DNA binding proteins. As explained above, we did not find any conserved CTCF binding sites in any STAR element, or any other insulator-binding protein binding sites for that matter. As shown in chapter 2, several deletion mutants of STAR 40 were tested in the original anti-repressor screen and several parts of STAR40 were shown to be bonafide STAR elements by themselves. This would at least argue against a single protein-binding site. The main reason is, however, that we could not determine any significant sequence homologies between the 65 human STAR elements that we isolated. If the action of the STAR elements were to be mediated by specific DNA binding proteins, either known or novel proteins, one might expect at least a common set of DNA binding proteins. This would in turn imply some common sequence motifs. Since these, to date, have not been identified, we have no indications that specific DNA binding proteins are pivotal for STAR action. The explained options of specialized chromatin structures that either counteract chromatin-associated gene repression or have a direct effect on promoter activity are in our opinion the more plausible explanations.

These speculations point at a serious drawback of the described approaches involving DNA elements to augment expression levels of recombinant proteins. Many of these approaches are operationally defined; they work or work not. Very little is known about the molecular mechanisms underlying any of the DNA elements discussed. For instance, the best investigated mammalian insulator, HS4 has only a small ‘core’ element (250 bp) that is responsible for insulator activity and beneficial effects on transgene expression. As outlined above, an essential part of the core HS4 element is the CTCF protein that binds to it. Loss of CTCF binding site results in abolishment of enhancer blocking activity. Still, a recent report shows that the CTCF binding site is not at all important in, or necessary for, protection against position effect. Abolishment of the CTCF binding site therefore does NOT alter the long-term stability of expression. It is telling how complex these DNA elements are when even within a stretch of only 250 bp two such different activities can reside. The point of this example is that there is a long way to go to understand why and how epigenetic gene regulatory elements are beneficial for therapeutic protein expression.
How to proceed?

Improvements of the described technology can be envisioned at different levels. One is to further increase protein expression levels. This could, for instance, be achieved by introducing more gene copies. Adding complementary and proven technology such as the dhfr amplification system could be a simple way forward. Whereas this is a tempting option, it would, however, result in mixing of old and new technology. The introduction of large arrays of amplified genes has its drawbacks and has been proven unstable when the dhfr inhibitor methotrexate is removed from the culture medium.

A more feasible path to follow is probably extensive medium optimization. As explained in the introduction, claims are that established increases in protein expression levels over the last decades can largely be accounted for by media optimization. While this is not our area of expertise, this certainly is a lead that should be followed up. Combining our STAR-related approach with optimization of the production process may in fact represent a next major step forward.

Another area where advantages can be booked is suspension growth. Our methodology has only been tested extensively in the context of small-scale adherent cell cultures. These by definition also contain serum in the culture medium. Most companies however use serum-free suspension cultures for their final production runs. It is always a challenge to change cells to medium without serum in order to grow them in suspension. Not only does this lengthen the time-lines to get to a producer cell-line, it is also quite common that cells lose their expression during this adaptation phase. It would be of great advantage, both in terms of predictability and time, to work in suspension from the start. Transfection and selection (especially in case of the stringent STAR-select system) in suspension, in the absence of serum, is a technology in itself. In order to move the technology into a commercially viable direction, these steps need to be taken.

With respect to the functional analysis of STAR elements, the just started approach to analyze chromatin structure of stably integrated STAR elements appears to be a viable option. Whereas preliminary results indeed point towards histone modification patterns that are distinct from the surrounding chromatin, this analysis can easily be extended towards other parameters of chromatin structure. It will also be of interest to extend the analysis to situations in which selection pressure is removed and the high expression levels in the STAR-Select system are lost. While we have observed a remarkable flexibility in the
regulation of the CMV promoter, it would be of considerable interest to investigate whether changes in chromatin structure also play a role.

**Concluding remarks.**

Over the years, many approaches were followed to improve the expression of exogenous proteins in mammalian cells, particularly of therapeutic proteins. Several groups and companies have implemented either already known or novel DNA elements to augment expression. Various gene amplification strategies were used to increase gene-copy numbers. Finally, improvements were made by optimizing cell culture medium and several new cell-lines were generated to accommodate the ever-growing market for therapeutic proteins. Even though all of these approaches have led to significant improvements, no holy grail has been found as of yet to solve all expression problems. It is questionable whether such a holy grail really exists. It is more likely that the production of each therapeutic protein requires a somewhat different approach.

In our own niche of vector optimization, we achieved remarkable improvements over the last six years. These improvements were all based on (theoretical) knowledge of epigenetic gene regulation mechanisms in mammalian cells, and were empirically tested in our protein expression platform. While the focus of our research over the last years has been to implement the STAR technology in an industrial protein production platform, the academic community has found new applications for this technology as well. Several transgenic mouse models for instance have successfully been generated containing our STAR elements to improve transgene expression.

Two major remaining challenges for the next few years are to get more fundamental insight into the working mechanisms of STAR elements and to put the technology into industrial use. More insight on the working mechanism is interesting from a scientific point of view, but also critical for the acceptance of the technology both in the academic world and in industry. With new insights on how STAR elements operate, new opportunities for improvement and additional add-ons to the technology arise.

Will we see, in some years time, products in clinical use made possible or affordable by using this technology? The future is written in the STARs.
References


