Employing epigenetics to augment protein expression in mammalian cells

Kwaks, T.H.J.

Publication date
2008

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.
Chapter 1

General Introduction
The production of recombinant therapeutic proteins

Therapeutic proteins

Traditionally, the public identifies medicines as a pill or a fluid. These are produced by pharmaceutical industries and usually contain so-called ‘small molecules’. These active compounds have been identified through the execution of large-scale screens, performed with random small molecule libraries that may contain up to millions of substances. However, during the last decades, many proteins with potent biological activities were discovered. Examples are growth factors and enzymes such as blood clotting factors. The potent activity of even small amounts of these proteins led to the realization that they might be useful as therapeutic agents. One potential benefit could be that ‘body-own’ proteins have less pleiotropic side effects than small molecules that are bound to interfere with multiple cellular targets. Also, when a missing protein is causal for a disease, ‘simple’ administration of these proteins could cure a patient. Well-known examples are the hormone insulin that is lacking in diabetes patients, or certain clotting factors that lack in the blood of hemophilia patients. Small molecules can hardly be expected to fulfill a simple scenario of replacing a missing or dysfunctional protein.

A major problem of isolating and applying potential therapeutic, ‘body-own’ proteins is, however, finding a suitable source. Application of purified therapeutic proteins is therefore restricted to classes of proteins that allow isolation from readily available human or animal sources. For instance, the insulin hormone was isolated from pig pancreases. Although this pig-derived insulin is biologically active and has been successfully used for the treatment of diabetes, adverse immunological side effects due to its animal origin were also noted. The porcine insulin differs from its human counterpart in only one single amino acid, indicating the degree of difficulty in finding the proper source. Another example was the isolation of human growth hormone from human placentas. This hormone is present in very small quantities; fortunately however it is very potent; hence the isolation of only small quantities, but from many placentas was still lucrative.

One of the world's largest biotechnology companies Genzyme became this successful, because its core business was the isolation of glucocerebrosidase from human placentas. This isolated enzyme replaces the glucocerebrosidase enzyme missing in patients that have Gaucher disease. Gaucher is an otherwise fatal childhood disease characterized by bone destruction and enlargement of the liver and the spleen. 20,000 human placentas provided
only a year’s supply of the enzyme for a single patient, at a cost of $400,000 annually. The treatment was considered to be very effective. In 1994 a recombinant form of the enzyme, called Imiglucerase, was approved by the FDA. Since the introduction of the recombinant product the costs per patient have substantially reduced to around $160,000 a year.

The dangers of isolated and injected ‘replacement’ proteins became apparent in the 1980’s. Purified human growth hormone isolated from the pituitary gland of human cadavers and provided by governmental programs like the National Pituitary Agency (NPA), was connected to the development of Creutzfeldt-Jacob disease in about 100 patients worldwide. These patients all received growth hormone replacement treatment in the 1960’s, long before anything was known about the infectious particles causing this transmissible spongiform encephalopathy. After this ‘incident’ the need for human growth hormone from other sources than purified from human pituitary glands became much more apparent. In 1985, the FDA rapidly approved the first recombinant human growth hormone Genentech had been working on since 1981.

A final example is clotting factor VIII, which is isolated from human blood. As of today, this is common practice and human serum-derived Factor VIII is still the main source of Factor VIII, worldwide. As with the above examples, about glucocerebrosidase and human growth hormone, also this example shows the limitations of isolation of therapeutic proteins from organs. In the first place, the source must be readily available, and in large quantities. In case of human blood, this is probably an almost perfect source, both in terms of availability and quantities. There is, however, also a major risk involved: Viral contamination. Screening for deadly viruses such as HIV and hepatitis C is therefore routine in blood banks. This strict quality control makes the isolation of blood clotting factors much more expensive. Moreover, while there are good screening methodologies for viruses such as HIV and hepatitis C, there are no good screens for prions, the causal pathogen of the deadly Creutzfeldt-Jacob disease. In particular the threat of this last disease, as indicated above, has led to an increased call for different approaches to produce therapeutic proteins.

**DNA technology and the first recombinant therapeutic proteins**

The advent of molecular biology has led to the identification of genes that encode proteins with therapeutic potential. Beyond that, the technological breakthrough in 1973 of molecular cloning and the ability to produce proteins in the laboratory has opened
entirely novel perspectives. In 1977 the first human recombinant protein, somatostatin, was reported to be produced in a micro-organism. Insulin was the first human therapeutic protein to be produced in bacteria at an industrial scale. Being approved by the FDA in 1982, it is still in use and is a major selling product of Ely Lilly in Indianapolis, USA. Even after 24 years, it still creates in excess of 2 billion $ in revenues, each year. Human growth hormone was the first recombinant therapeutic product approved by the FDA (Federal Food and Drug Administration) in 1985 to be completely manufactured and marketed by a biotechnology company (Genentech). Both Insulin and Human growth hormone are relatively simple, small proteins though, with hardly any complex post-translational modifications such as glycosylation or selenation. This feature allows production in bacteria, which themselves are not able to execute complex post-translational modifications. For more complex therapeutic proteins that require extensive post-translational modifications to be biologically active, however, production has to take place in eukaryotic cells that do have the capacity for adding proper post-translational modifications. Today, about 70% of all therapeutic proteins are therefore produced in eukaryotic cell systems.

Which classes of therapeutic proteins have emerged in the last decades? tPA (tissue plasminogen activator) was the first biotechnologically engineered therapeutic protein produced in recombinant mammalian cells to be approved by the FDA in 1987, tPA is used to prevent blood-clotting. It is a secreted serine protease which converts the proenzyme plasminogen to plasmin, a fibrinolytic enzyme. The biological activity and especially the plasma clearance rate of tPA are highly dependent on the degree of glycosylation. The manufacturer of tPA, San Francisco based Genentech, was amongst the very first companies that was founded to employ genes and biotechnology for the production of therapeutic proteins, the first biotech company to market it’s own products, the first company to use mammalian cells as a production platform and the company is still at the forefront of biotechnological innovations. Production of tPA took place in CHO cells. This CHO, or Chinese Hamster Ovary cell line has been the industrial workhorse amongst eukaryotic production cells for decades.

Another important biotechnologically engineered therapeutic protein is erythropoietin or EPO. EPO is a hormone produced by peritubular cells in the kidney and it stimulates the formation of red blood cells. As such it is an important therapeutic agent for the treatment of blood amnesia that can be a consequence of renal failure. The stability, activity and
secretion of this hormone are all dependent on the degree of N-linked glycosylation\textsuperscript{13,14}. The plasma half-life of recombinant human EPO is highly dependent on the level of sialylation, normal half-life is 5-6 hours as opposed to <2 minutes in the case of desialylated EPO\textsuperscript{15}. Despite all these difficulties, EPO is an immensely successful product. Production of EPO on CHO cells generates more than 10 billion $ a year. The major share of these revenues is for Amgen, who hold the US patent on EPO production. Unlike most blood clotting factors, which can be readily produced from blood sources, neither tPA nor EPO have tissue-derived equivalents that are used as therapeutic agent. However, while most clotting factors are still being purified from blood, both Factor VII and VIII are also produced by biotechnological means. But due to production difficulties with these biotechnologically engineered proteins, their share in the market is still limited.

\textbf{Recombinant therapeutic antibodies}

All above-mentioned therapeutic proteins are single chain polypeptides, but in recent years other classes of more complex proteins have reached the market. One important class is formed by the therapeutic monoclonal antibodies. Already in the beginning of the 20th century, scientists envisioned that antibodies might have potent beneficial effects in the treatment of human diseases. The term ‘magic bullet’ was coined by Paul Ehrlich (who shared the 1908 Nobel prize for medicine with Metchnikoff for their work on the immune system) and not without reason. It was realized that the specificity of an antibody for its target is very specific, probably more specific than can be achieved ever with any small molecule. The specific binding of an antibody to its target might therefore be used to block or stimulate the action of this target. Since antibodies are ‘body-own’, there might be less adverse reactions in terms of immunological side effects. In short, the promise that therapeutic antibodies hold is that they might mimic the way our body works. The discovery of the hybridoma technology\textsuperscript{16} in 1975 by Kohler and Milstein (later honored with the Nobel prize) opened the way to test these presumptions. Several monoclonal antibodies against a variety of human cancer cells were raised and tested. The initial results were, however, somewhat of a disappointment since the immunological side effects were much more severe then expected and the antibodies often proved rather ineffective as therapeutic agent. One major reason appeared to be that the antibodies that were raised in mice evoked an immunological response against the mouse protein portions of the antibody\textsuperscript{17}. From an immunological point of view, the antibodies were largely ineffective because they failed to invoke the immunocascades expected in a normal human
immunoresponse\textsuperscript{18-21}. Over the years these problems were overcome, however, by developing humanized antibodies. At the genetic level the murine parts of the antibody were removed and swapped for the corresponding human parts of the antibody\textsuperscript{22}. The first humanized monoclonal antibody approved in 1997 by the FDA was Zenapax, produced by the Swiss company Roche. This antibody combats organ rejection after transplantation by binding to and inhibiting a receptor on activated white blood cells. These humanized antibodies proved much less immunogenic whereas they are much more biologically active. Due to these developments, several humanized monoclonal antibodies have now reached the market and many more are currently being tested in clinical trials\textsuperscript{23}. Some of these humanized antibodies have been incorporated in the standard treatment of several human cancers. Two examples are trastuzumab (Herceptin) and bevacizumab (Avastin), both developed and marketed by Genentech. Herceptin blocks the interaction between the epidermal growth factor (EGF) and a mutated EGF receptor HER2/neu that is overexpressed in many metastatic breast tumors\textsuperscript{24}. Avastin targets and inhibits the binding of vascular endothelial growth factor (VEGF) to its receptors\textsuperscript{25}. Tumors often produce both VEGF and its receptor, which results in the promotion of blood vessels formation. This in turn promotes tumor growth, thus creating an autocrine loop. Herceptin has become a standard therapeutic agent to treat subclasses of metastatic breast cancer. Avastin is now being applied as treatment against an increasing range of solid tumors. Although these treatments can be considered as successful, it should be pointed out that Ehrlich’s early vision of a ‘magic bullet’ has not been fulfilled as yet with these types of antibody. Application of none of the present generation of therapeutic antibodies results in cure of human cancers. Instead, the addition of Herceptin or Avastin to the treatment results in prolonged life expectancy, only when the antibodies are delivered in combination with other, more traditional therapies\textsuperscript{26}. Furthermore, although antibodies demonstrate high target specificity, side effects are still observed. From the above example, the obvious targeting of for instance VEGF and its receptor during normal blood vessel formation is a problematic side-effect\textsuperscript{27} which could lead to increased bleeding. This said, however, physicians consider the introduction of monoclonal antibodies to the repertoire of anti-cancer agents as a major step forward. The hope and expectation is that novel generation antibodies will be more specific for tumor-related targets and have more limited side effects.

A final example of the use of monoclonal antibodies is the monoclonal antibody Enbrel, marketed by biotech companies Amgen and Wyeth. Enbrel is an artificially engineered
antibody that binds and deactivates Tumor Necrosis Factor (TNF)\textsuperscript{28,29}. TNF\textsuperscript{28,29} causes Rheumatoid Arthritis when present at too high levels. Enbrel is highly effective, but very hard to produce. The main reason for that is that the antibody is made from the combination of two naturally occurring soluble human 75-kilodalton TNF\textsuperscript{28,29} receptors linked to an Fc portion of an IgG1, a very complex molecule to produce for any cell. This example demonstrates the next level of problems that have been encountered in recent years: even if an antibody is proven to be very effective, it can be so hard to produce sufficient quantities that the antibody may never reach the market.

**Problems with the production of recombinant therapeutic proteins**

Production problems consist of two related problems. One is that during the development of a novel antibody it is desirable to test different variants of the antibody at the ‘lowest’ level of a clinical trial I. Typically, the potential toxicity of the antibody or parameters such as clearance from the blood after injection are established in clinical trials I. Hence this phase involves rather elaborate research, not involving human beings as test objects. However, if sufficient material for a clinical trial phase I can not be produced within a rather restricted period of time, the lack of enough interesting ‘lead variants’ may result in early termination of an entire program.

A second production problem is decisive during the production phase, which involves large-scale production in 10,000 to 20,000 liter vessels. If the specific productivity of a protein, stated as pg/ cell/ day is too low, also the volumetric productivity, stated as gram/liter will be too low. This may imply that the so-called cost-of-goods will become too high for a particular protein ever to be profitable. Eventually such factors are decisive for a company to produce a protein or not. To illustrate this, the biotechnological production of the above mentioned factor VIII does not exceed expression levels of 2-3 pg/ cell/ day. To be able to produce enough factor VIII, an airplane hanger full of culture medium is required, continuously. Although this may sound to be hardly competitive with blood-purified factor VIII, the 1.5 grams produced yearly of recombinant factor VIII is worth >$2 billion (Bayer’s Kogenate\textsuperscript{®} and Baxter’s Recombinate\textsuperscript{®} are the main rFVIII products on the market). And as mentioned above, the ever-strict rules of the FDA concerning virus contamination, recombinant factor VIII is still considered to be profitable.

Given these large-scale production problems it may not come as a surprise that there is a worldwide lack of production facilities. Current production facilities are fully booked years in advance. After the FDA approved Avastin in 2003, Genentech decided to build a 100,000
A dedicated biopharmaceutical liter production plant, estimated at $400 million, has been built to produce Avastin alone, which has resulted in exceeding costs of therapeutic proteins. Early 2006, several news items showed complaints from physicians in the Netherlands that hospitals could no longer afford prescribing anti-cancer drugs like Avastin and other anti-cancer monoclonal antibodies (fig 1). This not only exemplifies the exceeding costs of novel therapeutic drugs, but also the high level of public interest in these matters. One can safely conclude that easing the production of therapeutic proteins will have a major impact.

**Figure 1. Dutch patients do not get new anti-cancerous drugs.**
Newspaper clipping of “The Volkskrant”. Cancer patients in the Netherlands receive less new anti-cancerous drugs than patients in other European countries because of the costs of these drugs.
Potential solutions for the production problems

What are the options? A bacterium is still the fastest multiplying organism with its generation time of about 20 minutes. It can grow to very high densities and it can produce a large part of its body weight as exogenous protein. However, as outlined above, bacteria are often not suitable hosts to produce complex, eukaryotic proteins, as they lack the ability to add proper post-translational modifications often seen on therapeutic human proteins. The first suitable simple eukaryotic cell is yeast. Yeast has the ability to add post-translational modifications to a protein, but these are yeast-specific. Efforts are being made to replace the yeast post-translational machinery with the human variant. This would allow producing a therapeutic protein with the proper post-translational characteristics, made by a fast and also to high density growing microorganism. However, even if successful in the end, it will still take many years for these novel developments to be implemented in large-scale industrial production processes and to replace, inefficient, but proven current production systems.

The same story also holds for plants. Potentially, producing therapeutic proteins in plants has huge advantages. In general, plants require relatively little care. Another advantage, particular for the third world, is that growing and eating (modified) plants that are already part of the normal diet could avoid often-difficult distribution problems of therapeutic proteins. However, designing plants that possess a humanized post-translational machinery is in its infancy. Furthermore, because of the high development costs, hardly any transgenic crops will be sold to 3rd world countries, unless they are highly subsidized.

Transgenic animals are another option to create large quantities of secreted therapeutic protein. Potentially, a herd of transgenic goats in a meadow can out-produce an entire production plant culturing mammalian cells. Still the transgenic animal option is not a success story. Technically, it is far more challenging to create a transgenic animal producing a recombinant protein than to create a producer cell-line. More importantly however, the public opinion remains very much opposed. Despite this all, however, in 2007, Leiden-based Pharming for instance is on the verge of introducing a transgenic animal-produced therapeutic protein. The therapeutic protein in question is recombinant human C1 inhibitor for the treatment of Hereditary Angioedema.

With these options being either in early stages of development or remaining controversial, production of recombinant therapeutic proteins in mammalian cells remains a highly competitive alternative. The technological basis is relatively low key and the technology is accepted by public opinion. Above all, it is proven technology and several cell lines are still
the real workhorses of the biotechnology industry. This thesis is concerned with production problems and potential solutions, which could open the way to improved production of recombinant proteins in mammalian cells.

**Current approaches to solve these production problems**

The development of manufacturing processes for recombinant therapeutic proteins in mammalian cells follows a well-established scheme (fig 2). Initially, the recombinant gene is transferred to the cells. In addition, a second gene, either on the same plasmid or on a separate plasmid, is transferred that provides recipient cells with a selective advantage. During the selection phase, only those cells expressing the selector gene will survive. Depending on the method of selection, a percentage of the surviving cells will also express the recombinant therapeutic protein. Individual clones are evaluated for recombinant protein expression and after multiple rounds of sub-cloning a limited number of candidates is chosen. Eventually, one cell line with the appropriate growth and productivity characteristics is chosen for the production of the recombinant protein. The whole development process, from DNA to producer cell-line takes at least 12 months and an incredible amount of work. Shortening these time-lines, or reducing the amount of cell-lines to be screened, would greatly help to reduce the costs of such therapeutic proteins.

*Production problems of recombinant therapeutic proteins in mammalian cells can be distinguished in several levels.*

Predictability: DNA constructs harboring the gene of interest are introduced in the cells and stable integration into the genome of these cells takes place. Integration of DNA is thought to be random and is an inefficient process. It is, therefore, not easy to obtain a clone with stably integrated DNA that produces much protein. Usually, many clones need to be screened before a stable cell line that produces protein is established.

Yield or protein expression levels: After identifying a clone that produces protein, it turns out that the protein expression levels in such clones are often very low. Finding a high producing clone is very time-consuming, resulting in long delays of the onset of clinical trails.
Stability over time: Even when these two hurdles are overcome and a high-producing clone has been established, these high expression levels are often not maintained after an often-short period of time.

What are the approaches that the biotechnology industry in general takes to overcome these production problems?

**Predictability**

The answer to this question can be very short. Usually, the predictability problem is simply tackled by brute force. Either thousands of clones are isolated by hand, or they are being ‘picked’ by a robot. Obviously, only large companies can afford the use of multi-million $ high throughput robotics.

*Figure 2. Cell-line development scheme.*

The development of a production cell-line follows a well defined scheme. From gene transfer to cell-line screening to process development, this whole process takes a minimum of 12 months.
Yield

Over the years, the main approach in industry to achieve higher protein production levels has been by employing gene amplification\textsuperscript{12,36}. Placing of either the \textit{dhfr} (dihydrofolate reductase) gene or the \textit{GS} (glutamine synthetase) gene on a plasmid that also harbors the gene of interest makes an amplification strategy possible. The \textit{dhfr} protein is an enzyme that is part of the folate pathway and it converts folate into 5,6,7,8 tetrahydrofolate, a methyl group shuttle required for the de novo synthesis of purines (Hypoxanthine), thymidylic acid (Thymidine), and certain amino acids (Glycine). CHO-DG44 cells lack the \textit{dhfr} gene and these cells therefore need glycine, hypoxanthine and thymidine in the culture medium to survive\textsuperscript{37}. If, however, the \textit{dhfr} gene is present on the expression cassette, the cell can convert the non-toxic substance folate into 5,6,7,8 tetrahydrofolate, bypassing the need for exogenous glycine, hypoxanthine and thymidine in the culture medium. The \textit{GS} gene works in a similar fashion, it catalyzes the synthesis of glutamine from glutamate and ammonia and provides the only pathway for synthesis of glutamine. These principles have been used for many years as selection methodology to create stably transfected mammalian cell lines.

Most cell lines, such as the widely used CHO-K1, possess at least one functional \textit{dhfr} allele. Usage of the \textit{dhfr} system is therefore largely restricted to CHO-DG44 cells (and derivatives) which are homozygous \textit{dhfr}-negative. Some reports claim the use of the \textit{dhfr} system in \textit{dhfr}-positive cells as well. This can be accomplished by inhibition of the endogenous \textit{dhfr} gene using the specific inhibitor methotrexate.

The \textit{GS} system operates best in functionally \textit{GS} deficient cell lines like NS0\textsuperscript{38,39}. These cells do not produce sufficient levels of \textit{GS} to survive and so require an exogenous source of glutamine. However, there are also several reports claiming that cells with sufficient levels of endogenous \textit{GS}, like CHO-K1, can be used in combination with the specific \textit{GS} inhibitor methionine sulphoximine\textsuperscript{40-43}.

By an unknown molecular mechanism amplification of both the \textit{dhfr} (or \textit{GS}) gene and the gene of interest occurs when the \textit{dhfr} gene is inhibited by methotrexate (methionine sulphoximine in the case of \textit{GS}). The whole locus encompassing the \textit{dhfr} gene is multiplied to ensure high enough expression of the \textit{dhfr} protein for the cell to survive. The resulting higher copy number of the gene of interest often (but not always) results in higher protein of interest expression levels. There are, however, multiple problems with this methodology. For one, the increased gene copy numbers do usually not result in a linear increase of protein expression levels\textsuperscript{44}. Furthermore, in many cases the number of genes is amplified to
thousands and this often results in genomic instability of the cells. When methotrexate is removed from the culture medium, the amplified gene loci are often removed from the host DNA. Not surprisingly, the protein expression levels then drop, which is a serious problem. To keep the protein expression levels high, continuous selection pressure by methotrexate is needed. However, the FDA does not accept this, since methotrexate is highly toxic and teratogenic, therefore the methotrexate has to be removed from the final product, the therapeutic protein. One often used trick is to grow cells up to a volume of ~1000 liter. Then the cells are changed to culture medium without methotrexate and transferred to a culture batch of 10-20,000 liters. In this way, the toxic substance is removed and the decrease in expression levels due to unstable expression is limited to the final production run only.

To obtain sufficient gene amplification, often multiple rounds of sub-cloning with increasing amounts of methotrexate are required. During this process, an original cell clone is re-suspended in such a way that culture wells receive statistically less than one cell, which then grows to a daughter- or subclone. All of these subclones must subsequently be analyzed for protein expression levels. Only those subclones that have elevated protein expression levels in comparison to the original parental clone are taken for another, similar round of subcloning. This entire procedure can take up to five months\textsuperscript{44}, a long time when multiple variants of a single therapeutic protein are needed fast for testing in a clinical trial phase I. However, although gene amplification is a long, inefficient and costly process, it is THE approach industries have taken for decades. Many products on the market have been produced in this way. Once a production clone has been established, through whatever numbers analyzed or how many rounds of gene amplification, the usual follow-up consists of medium optimization. Some opinions state that media optimization is the key to it all\textsuperscript{12,36}. Establishing growth characteristics, feeding strategies for the cultures, metabolic patterns, all vary between individual cell cultures. This is a science in itself. Claims are, however, that the major increases in production levels that have been achieved over the decades are solely due to developments in culture optimization.

Finally, a novel approach to get enough protein for clinical phase I material fast has been developed recently\textsuperscript{45-47}. This involves large-scale transient expression of the gene of interest. If the transfection efficiency is high enough and if it is possible to scale the entire transfection procedure up to ~100 liter, chances are that sufficient material is produced for a clinical trial phase I. Such an approach circumvents establishing stable clones with
integrated DNA in the host genome. Obviously, this may solve the problem of obtaining enough material for clinical trial phase I, but it does not solve the problems in the final large-scale production phase.

**Stability**

No simple solutions for the lack of stability of protein expression over time have yet been found. As indicated above, the removal of methotrexate from the culture medium often results in a fast and dramatic drop in protein expression levels\(^{43,44}\). The way this problem is circumvented is usually to leave the protein producing cells under continuous selection pressure. However, there are reports that even under continuous selection pressure the stability of expression levels is not warranted\(^{44,48,49}\). The presence of selection in production runs of therapeutic proteins is not favored. Selection agents are often very expensive, which adds up when 10-20,000 liter vessels have to be filled. Furthermore, the FDA requires that such agents are removed form the final product, the therapeutic protein. This implies costly purification procedures as well as quality control. Therefore, even if the presence of the selection agent is required for maintaining reasonable expression levels, the agent is usually present only until the production batch is in a volume of ~10 liter. After that the selection agent is omitted from the medium before transfer to larger production vessels. The hope is then that during the final phases of the production run (still up to 4 weeks) expression levels remain reasonably stable.

**Employing epigenetic gene regulation mechanisms to resolve protein production problems**

In this thesis we describe approaches to solve protein production problems that stem from knowledge of epigenetic gene regulation mechanisms. What are epigenetics or epigenetic regulation of gene expression?

When a transgene is transfected into a cell and integration into the host-genome occurs, efficient expression of the transgene is highly dependent on the place of its integration\(^{50-52}\). In the genome, genes reside as chromatin, the complex structure of DNA and associated proteins, in particular histones. Basically, two forms of chromatin, euchromatin and
heterochromatin can be distinguished. In general, endogenous genes located in heterochromatin are repressed or silenced and genes that reside in euchromatin are active. The faithful transmission of the repressed status of genes is part of a so-called epigenetic memory system. When a transgene integrates into the genome, integration into or close to heterochromatin will result in a silenced or repressed state of the transgene, whereas integration into actively transcribed euchromatin usually leads to transgene expression. In other words, the expression level of a transgene is dependent on the site of genomic integration. One goal of several approaches we will describe is to create a system in which expression is independent of the site of genomic insertion. Since a large proportion of the genome is in the form of heterochromatin, the chance that a transgene integrates in or close to heterochromatin and consequently is silenced or repressed is also high. Importantly, once genes obtain a repressed heterochromatic state, this status is faithfully transmitted to daughter cells. An integrated transgene adapts the chromatin state of the surrounding area. At the molecular level, the DNA double helix is wrapped around histones and this constitutes the basic building block of chromatin: the nucleosome (see figures 3 and 4). During the last decade a lot has become known about the role of histone modifications on (epigenetic) gene regulation. Acetylation, methylation, phosphorylation and ubiquitination of lysines and serines of amino-terminal histone tails have enormous impact on whether a gene is active or inactive. Transcriptionally silent chromatin for instance can be marked by enrichment of di- and tri-methylated H3K9 or H3K27, together with overall histone hypo-acetylation (Fig 3B and 4). In contrast, transcriptionally active chromatin is marked by enrichment of tri-methylated H3K4 and overall hyper-acetylation (Fig 3A). Particular the combinatorial nature of these modifications or ‘Histone Code’ adds a complex layer of epigenetic gene regulation to existing modes of gene regulation by promoters and enhancers. The specific

Figure 3. Events involved in heterochromatinization by the HP1 repressive complex.
(A) Addition of positively charged acetyl groups to lysines in histone tails, such as lysine K9 in histone 4, renders the nucleosome more open for transcription. In general, histone acetylation, mediated by a histone acetyltransferase (HAT) has a positive influence on gene expression. (B) When acetyl groups are removed from the histone tails by histone deacetylases (HDAC), genes become repressed, particularly when also methyl groups are added to the same lysine by a histone methyl transferase (HMT). (C) The methylated lysine K9 is a docking place for the heterochromatin protein HP1, a potent chromatin-associated repressor. Subsequently, more and more nucleosomes become deacetylated, methylated and HP1 saturated and hence gene repression becomes more and more effective. (D) DNA methyltransferases (DNMT) recruited by the HP1 repressive complex add a final methylation mark to the DNA to assure permanent silencing. Adapted from: Trends in Biotechnology 2006 Mar;24(3):137-42
histone modifications overlap with chromatin states that have broadly been defined as the active euchromatin or the inactive heterochromatin states. Heterochromatin has further been sub-divided into constitutive, or permanently inactive heterochromatin, such as centromeres and facultative heterochromatin such as the inactivated female X chromosome. Only one of the two female X chromosomes becomes inactivated and this process falls together with temporary enrichment of tri-methylation of H3K27 (Fig 4). This specific mark serves as docking site for the Polycomb repressive complex and the Polycomb repressive complex is therefore transiently associated with the inactivated X chromosome. On the other hand, methylated lysine 9 on histone H3 serves as docking site for the HP1 repressive complex (see below and Fig 3C), and this complex is more often associated with constitutive heterochromatin, such as centromeres.

At least three overlapping epigenetic phenomena are involved in regulating gene expression: histone modifications or the occurrence of so-called histone variants, repressive protein complexes such as Polycomb and HP1, and DNA methylation. During the last few years it has become apparent that these regulatory mechanisms are interlinked. These complex relations are visualized in Figures 3 and 4. For instance, DNA methylation has long been recognized as being associated with repressed chromatin states and inhibition of gene expression. There are two general mechanisms by which DNA methylation inhibits gene expression: it can prevent the binding of transcription activating proteins or it can attract methyl-CpG binding proteins. These last proteins exert their repressive action by recruiting so-called co-repressors, such as histone deacetylases and histone methyltransferases. Thereby DNA methylation is mechanistically linked to histone modifications.

Another complex link between all three epigenetic regulatory phenomena involves Polycomb-mediated gene repression (Fig 4). The Polycomb group protein Enhancer of Zeste (EZH2) is a histone methyltransferase. EZH2 can methylate H3K27. The methylated H3K27 mark serves as docking site for the Polycomb protein itself. Furthermore, the EZH2 protein also interacts with DNA methyltransferases, thereby linking histone modifications, Polycomb repressive complexes and DNA methylation.

What is the importance of these epigenetic gene regulatory mechanisms?
An example of large scale epigenetic gene silencing is imprinting, which causes a difference in behavior between the alleles inherited from each parent. About 80 human genes are known to be imprinted, mostly lying in large gene clusters. One of the best-
studied examples of imprinting is the Igf2/H19 locus. The Igf2 gene is always silenced on the paternal allele, whereas the H19 gene is always silenced on the maternal allele.64 One of the most dramatic examples of long-term gene silencing is the already mentioned X-chromosome inactivation in females. Early in development, on one of the two X-chromosomes in the female embryo the vast majority of genes are completely silenced, a process affecting more than a thousand genes.65 Early on in the X-inactivation process, the Polycomb group proteins EED and EZH2 play an important role, specifically since the EZH2 histone methyltransferase induces the H3K27 methylation marks.65,66. Also in X-inactivation DNA methylation is the final epigenetic mark that appears after the histone modifications and Polycomb-mediated repression and is thought to be associated with long-term maintenance.

Figure 4. Events involved in heterochromatinization by the Polycomb repressive complex (PRC1).

Presence of negatively charged acetyl groups to lysines in histone tails, such as lysine K27 in histone 4, renders the nucleosome more open for transcription. When acetyl groups are removed from the histone tails by histone deacetylases (HDAC), genes become repressed, particularly when also methyl groups are added to the same lysine by the histone methyl transferase EZH2. The methylated lysine K27 is a docking place for the Polycomb repressive complex 1 (PRC1), a potent chromatin-associated repressive complex. Subsequently, more and more nucleosomes become deacetylated, methylated and PRC1 saturated and hence gene repression becomes more and more effective. Finally, DNA methyltransferases (DNMT) recruited by the Polycomb repressive complex add a final methylation mark to the DNA to assure permanent silencing.

inactivation DNA methylation is the final epigenetic mark that appears after the histone modifications and Polycomb-mediated repression and is thought to be associated with long-term maintenance.

The pivotal role of epigenetic gene regulation became also clear when epigenetic abnormalities were found to be the cause of human cancers. The examples of disruption of DNA methylation patterns in the promoter regions of tumor-suppressor genes in human
cancers are numerous. Methylation of these genes may occur early in tumor progression and lead to dysfunction of important cellular pathways, including those controlling cell cycle, apoptosis and cell signaling. A number of the Polycomb group proteins have also been implicated in tumor-development. Overexpression of the Polycomb group gene BMI1 in transgenic mice was shown to result in suppression of the p16INK4A/p19ARF cell-cycle regulators. These 2 genes normally function as part of a fail-safe mechanism preventing uncontrolled cell proliferation. They act on both the Rb and p53 pathways, inducing cellular senescence or apoptosis. Suppression of these genes leads to desensitization of cells to apoptosis and facilitates cell proliferation and tumorigenesis. In humans, overexpression of BMI1 has been linked to various tumor-types. Another example of a Polycomb group protein involved in human cancer is EZH2. Overexpression of the EZH2 protein was shown to be involved in metastatic prostate cancer and aggressive forms of breast cancer.

What is the connection between epigenetics and transgene silencing?
As mentioned before, integration of a transgene is thought to be a random process. About 98% of the mammalian genome consists of transcriptionally inactive chromatin, either heterochromatin or inactive euchromatin. Therefore, the chance that a transgene will integrate into or close to an unfavorable epigenetic context is very large. Even if the transgene has integrated into an active region of the genome and is expressed at high enough levels, it is common that, after several rounds of cell division, the expression levels drop. Since there are no changes at the DNA level in the cell, epigenetic gene regulatory mechanisms must play a pivotal role in this silencing. Since the new epigenetic status of the transgene is part of the overall epigenetic memory of the cell, the silenced state will be faithfully transmitted to all daughter cells.

In summary, it is thought that gene repression associated with heterochromatin involves a shift in the balance between histone acetylation, histone methylation and DNA methylation. We will use the events described above and depicted in figure 3 to identify the points at which can be interfered to counteract epigenetic repression. We describe and discuss several strategies that emerged over recent years to counteract epigenetic silencing to improve the expression of transgenic proteins.
Histone modification

One of the oldest strategies to create and maintain a favorable chromatin conformation is based upon the difference in histone modifications between active and repressed chromatin. As shown in Fig 3A and 3B, an open or active chromatin domain is generally associated with histone acetylation and repression involves removal of the acetyl group. Deacetylation of the histone tails is performed by histone deacetylases (HDACs). Blocking the action of HDACs by adding a HDAC inhibitor to the culture medium may therefore keep histones in a higher acetylation state, which enables transcription. Indeed, the HDAC inhibitor butyrate represents one of the oldest means that has been used to enhance gene expression.75 However, since all nucleosomes are being influenced in this way, this method is by de-fold crude and non-specific.76 A disadvantageous side effect of sodium butyrate is also inhibition of cell growth and induction of cellular apoptosis.77 Still there are many reports that claim higher expression levels of therapeutic proteins in the presence of butyrate. Thrombopoietin, erythropoietin, interferon-6, and Factor VIII are some of the proteins of which the expression levels in CHO cells have been increased by adding sodium butyrate to the culture medium. However, to our knowledge no use in a large scale, industrial setting has been reported.

An alternative method to influence the acetylation status of a transgene is to target a histone acetyltransferase (HAT) to the promoter of a transgene. This warrants that the promoter of the transgene remains continuously in an acetylated or active state. Targeting HAT activity has the advantage that hyper-acetylation only occurs at the site where it is wanted, this being the promoter of the transgene, and not genome wide, as with butyrate. This approach will be discussed in chapter three of this thesis.

DNA approaches other than gene amplification

Gene amplification through the dhfr gene is a tedious and inefficient process, as outlined above. Other approaches, however, have been tried to achieve better integration of the gene of interest into the host DNA and higher gene/protein expression levels. Some of the currently used approaches will be discussed.
**Targeting of zinc finger proteins to a promoter.**

Pabo and co-workers have developed this elegant approach\(^2\) in 1997 and the San Diego based biotechnology company Sangamo Biosciences Inc is propagating this technology. Basically, they make a fusion protein between a zinc finger DNA binding protein and an activator of gene expression. Each particular zinc finger binding domain recognizes a 9 or 10 bp DNA stretch, combinations of various zinc finger domains can be made to increase the specificity of the binding\(^3\). A library of 17 million novel polydactyl proteins was created binding the 5’-(GNN)6-3’ family of 18-bp DNA sites. To specifically control the expression of a gene, an 18bp sequence present in the 5’-UTR of the gene of interest, this being either an endogenous gene or a transgene is targeted. When the zinc finger binding protein has been isolated, it is engineered to form a fusion protein with a module that is a potent gene activator. The gene encoding this novel fusion protein is then introduced into the host cell. In principle the zinc finger binding protein will target only the gene of interest, which will be activated by the activator module of the fusion protein\(^4,5\). Although the approach is very elegant, the fusion protein still needs to be transfected to a host cell (the therapeutic protein producing cell line) that has been established with much effort. Furthermore, the fusion protein needs to be expressed at sufficient levels and be stable over time. These requirements form precisely the protein production problem; hence in way it only replaces the basic problem with another set of problems.

**Incorporation of the transgene in artificial chromosomes**

The recurring problem of instability of gene expression over prolonged periods of time is at least in part caused by epigenetic gene silencing mechanisms that reside in the chromatin of the host cell. One way to circumvent these problems is to create a chromatin environment that warrants that the gene will not become silenced. This can be achieved by introducing an artificial chromosome\(^6,7\) and directly target the gene of interest to multiple known sites on this chromosome. The Vancouver-based biotechnology company Chromos Molecular Systems Inc. has developed this approach (ACE System\(^8\)). Introduction of the artificial chromosome into the cell nucleus at least circumvents that this chromosome will be incorporated in the host genome: the artificial chromosome is recognized as such and replicates normally as a host chromosome during cell division\(^9\). The challenge with this system is to choose DNA sequences to build the 60 Mb artificial chromosome that are not prone to gene silencing. Obviously, it requires very specialized personnel to physically isolate the engineered chromosome and inject them manually into the nuclei of cells.
However, a stable cell-line containing the extra chromosome can be used for multiple products. These cells can then be targeted and propagated for establishing a cell line that produces the recombinant therapeutic protein.

**Targeting to hot spots.**
A different approach to improve transgene expression is to utilize so-called hot spots: endogenous sites in the genome where overall, high gene expression occurs\textsuperscript{90,91}. If one could target a transgene to such hot spots, the transgene would likely be expressed at a high level as well. One way to achieve this is to use recombination tools such as Cre-lox or flp-recombinase\textsuperscript{92-95} (Fig 5). Epigenetic gene silencing is circumvented, because the endogenous sites of open chromatin and high gene expression have already been established during evolution. The concept is tempting and this technology can in principle be used in different cell lines. However, there are limitations to the technology. In principle, only a very limited number of copies can be targeted to a hot spot, because of the flanking, recombining sequences. Activation of the recombinase, which is essential in the process, will result in excising multiple gene copy concatamers that contain LoxP sites, until only one or two copies remain. This will inevitably result in less expression of the protein of interest than when copy-number dependency is achieved with cis-acting DNA elements. Another disadvantage of the technology is that the genome of the host must be quite stable. It has been well established that much-used industrial cells such as CHO-K1 are genetically very unstable. This may imply that the hot spot does not remain at the same chromosomal location and this could result in events that eventually lead to reduction of protein expression.

**Expression augmenting DNA elements**
A widely used strategy to improve transgene expression is to incorporate cis-acting DNA elements in the transgene constructs. The DNA elements that we will describe have in common that they appear to have no effect on transient gene-expression levels, this in contrast to enhancers. Instead, these elements all need one or several rounds of replication to function, indicating a mechanism involving chromatin structure. Hence they fall into the
category of epigenetic gene regulators. Ideally such DNA elements fulfill a number of criteria. They should:

- be <4-6 kb long, since else constructs become unstable or too large for cloning
- convey higher expression levels to a transgenic protein
- convey stability of expression over time
- be universally applicable, that is in any cell line, and with any promoter
- convey copy-number dependent gene expression

In table 1, we tentatively rank how the described DNA elements fulfill these criteria.
**LCR**

One of the oldest applications of an epigenetic regulatory DNA element is the human \( \beta \)-globin Locus Control Region. This complex locus contains five DNaseI hypersensitive (HS) sites spanning about 16 kbs of genomic DNA that regulate the proper expression of fetal, embryonic and adult \( \beta \)-globin genes. Placing the \( \beta \)-globin LCR in front of a transgene leads to high and stable transgene expression over time in transgenic mice\(^51\). The LCR also conveys copy number dependent expression of gene activity. This is an important characteristic, nowadays considered the hallmark of an LCR\(^96\). Often the expression of a transgene is copy-number independent. So, even if many hundreds of gene copies are integrated into the genome, the expression levels do not increase significantly as compared with cells that harbor much less gene copies. Therefore, when a DNA element conveys copy-number dependency, integration of more gene copies also results in increased transgene expression.

A drawback of the LCR is, however, that it is too large for efficient use in the simple set-up of over expressing a protein in mammalian cell lines. Furthermore, because of the tissue specificity of the \( \beta \)-globin LCR (HS sites 1 to 4 are present only in erythroid cells) it works only in the context of cells derived from erythropoietic cell lineages and globin-related promoters. The same is true for other LCRs like the \( \lambda \).5-VpreB1 LCR\(^97\) and the human growth hormone LCR\(^98\). This limits the usefulness and simple applicability of LCRs in mammalian cell lines. Therefore, use of LCRs has largely been restricted to the creation of transgenic animals.

**Insulators: cHS4.**

The best-studied mammalian insulator is the 1.2 kb long, most 5’ located DNaseI HS site of the chicken \( \beta \)-globin LCR\(^50,99\). The term insulator means that the element is able to block the action between an enhancer and a promoter\(^100,101\). In other words, the DNA element

---

**Table 1. Reported characteristics of the described epigenetic regulatory DNA elements.**

<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(^51)</td>
<td>16 kb</td>
<td>?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>cHS4(^102,110)</td>
<td>1.2-2.4 kb</td>
<td>?</td>
<td>Yes</td>
<td>?</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>UCOE(^112,113)</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(^117,118)</td>
<td>~3 kb</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

(?: Unknown or inconsistent results) Adapted from: Trends in Biotechnology 2006 Mar;24(3):137-42
insulates the enhancer from a promoter. Incorporation of the cHS4 insulator into a transgene construct conveys stability of transgene expression over time\textsuperscript{102}. The cHS4 insulator has not only been claimed to increase the stability of transgene expression in mammalian cell lines, beneficial effects on its use in gene therapy vectors have also been reported\textsuperscript{103-106}. The cHS4 insulator induces only modest increases of transgene expression in CHO-K1 cells\textsuperscript{107}, particularly when compared to reported increased expression levels by other elements (see below). Probably this is due to the fact that the cHS4 insulator is derived from the β-globin LCR, whose action is by definition erythropoietic cell line specific. Therefore, beneficial effects of the cHS4 might be largely restricted to the context of erythropoietic cells and β-globin-related promoters. This possibly implies a limited usefulness in industrial settings. Scientifically the cHS4 element is very interesting though, since a lot is known about the underlying molecular mechanisms. It probably attracts HAT activity\textsuperscript{102,108,109}, (Fig 3A), while methylation- and acetylation-patterns\textsuperscript{108,109} in and outside domains point towards real barrier function against the events shown in Fig 3B and 3C. When ranked according to the criteria set above, the element is small and transgene expression becomes more stable. However, it does not appear to be universally applicable and it does not convey copy number dependent gene expression\textsuperscript{110}.

**UCOEs**

One way to circumvent the tissue-specificity of a LCR is to use promoter regions of ubiquitously expressed housekeeping genes. The basic idea behind this is that housekeeping genes are always active at a constant and sometimes very high level. Regulatory elements in the vicinity of these promoters should therefore be useful to activate a transgene to high expression levels\textsuperscript{111}. Especially the dual, divergently expressed promoters of the HNRPA2B1 and CBX3 genes were shown to be able to maintain a favorable chromatin conformation and to protect a transgene from silencing\textsuperscript{112,113}. This and similar elements have been called Ubiquitous Chromatin Opening Elements. The elements are relatively big (~16kb\textsuperscript{112}) as compared to the other DNA elements. However, a recent paper describes a much shorter version (~2.5 kb) that was still able to convey higher transgene expression levels in a number of mammalian cell lines. Importantly, this UCOE appears to act also on the often used, strong viral hCMV promoter\textsuperscript{113,114}. Expression levels, increased up to sixteen-fold by the presence of the UCOE, are stable over prolonged periods of time. No copy-number dependency of expression has been reported. The UK based company Cobra Therapeutics Ltd. developed this technology.
**MARs**

Matrix Associated Regions (MARs) are DNA elements that bind to the nuclear matrix, a proteinous structure in the cell nucleus. MARs are thought to be involved in creating higher order structuring of chromatin in topologically independent loops\(^\text{115,116}\). The hypothesis is that the expression of genes that reside in such loops are coordinately regulated. MARs were identified in the beginning of the 1980s. There are many reports on the influence of different MARs on transgene expression, but they differ in the assessment of their functionality\(^\text{117,118}\). In general, the chicken lysozyme MAR and the human \(\beta\)-globin MAR have been shown to have a positive effect on stable transgene expression in CHO cells whereas the chicken pi \(\alpha\)-type MAR has a negative effect. Not much is known about how MARs work at the molecular level and why they have beneficial effects on transgene expression.

Several reports state that MARs operate cell line and promoter independent\(^\text{117,118}\). They are relatively small (~3 kb) and can be applied easily. However, they do not convey copy-number dependent gene expression\(^\text{117}\). A recent report shows that the effect of placing a MAR on the transgene construct (in cis) can be enhanced by co-transfection of additional MAR elements in trans. A 5- to 10-fold expression increase has been reported\(^\text{119}\). The Swiss-based company Selexis uses MAR technology to elevate expression levels of recombinant therapeutic proteins.

**Our approach: anti-repressor elements**

Fig 3C shows that heterochromatin-mediated gene repression can involve ‘spreading’ of the HP1 complex over at least multiple nucleosomes. We designed a genetic screen that allows cells only to survive when a DNA element is present that blocks chromatin-associated repressors. We indeed isolated such DNA elements and applied them in transgene constructs, to obtain more protein expressing clones and higher protein expression levels. This approach is the topic of chapter two of this thesis.
Outline of thesis

In this chapter, the advent of therapeutic proteins as promising classes of modern medicines is described. The severe problems with the (large-scale) production of these recombinant proteins result in high ‘costs of good’ and exceeding high prices. This threatens the effective introduction of many promising recombinant therapeutic proteins. Approaches that industries traditionally follow to solve these problems are described, as well as novel, more recent developments. Our approach is to employ insights into epigenetic regulation of gene expression in order to solve production problems of recombinant, therapeutic proteins. In chapter two a novel genetic screen is described to isolate human DNA elements that are able to counteract chromatin-associated gene repression. Application of these anti-repressor or STAR elements in plasmids that harbor a gene of interest results in the establishment of many more cell clones that express protein, as well as higher protein expression levels. The STAR elements are remarkably conserved during evolution, both functionally and on the DNA level.

In chapter three enzymes that are involved in the ‘opening’ of chromatin structure, such as histone acetyltransferases (HATs) are targeted to the promoter of an introduced gene of interest. This results in higher proteins expression levels and increased stability of expression.

In chapter four a novel selection strategy is described. This system is developed because incorporation of STAR elements leads to a large increase in the number of colonies after transfection. Application of this very stringent selection system kills the vast majority of established colonies that are not ‘top-level’ protein producers. Together with the application of STAR elements both the low predictability of protein expression and low yields are addressed.

In chapter five several DNA elements that are described in this introductory chapter are tested side by side, in the context of traditional selection systems and the novel developed selection system.

In chapter six the presented results and insights are discussed. What has and what has not been achieved is evaluated, along with suggestions for future research.
References


