MASTER THESIS

From Endogenous to Artificial Promoters

Performed at the
Department of Biotechnology
Institute of Applied Microbiology
BOKU
University of Natural Resources and
Applied Life Sciences, Vienna

Submitted by
Michael Meier

Supervisor:
Ao.Univ.Prof. Dipl.-Ing. Dr.nat.techn. Reingard Grabherr
Dipl.-Ing. Dr.nat.techn. Wolfgang Ernst

Assistant Supervisor:
Dipl.-Ing. Martina Baumann Bakk. techn.

Austria, Vienna, June 2011

Felix, qui potuit rerum cognoscere!

Acknowledgments

Writing this master thesis would not have been possible without the support of many people. First of all I would like to express my gratitude to Professor Reingard Grabherr and Dr. Wolfgang Ernst for giving me the opportunity to conduct my studies in their lab. Thank you for always having an open door for my concerns and your encouraging and helpful advises and suggestions.

I appreciated the collegial collaboration with both of them and the BACULO group making this work possible. I especially want to thank my supervisor Martina Baumann for her assistance, the harmonic teamwork and for building the foundation for this research. I am truly grateful to have entered a group like the BACULOs, where a very inspiring atmosphere and great team spirit are present.
I want to thank the group members namely Esther Egger, Stefan Gross, Dr. Stefan Heinl, Miriam Klausberger, Jürgen Mairhofer, Dieter Palmberger, Eva Smolar, Katharina Spath, Philipp Tauber, Christopher Tauer, Peter Vaboschek and Monika Wilde for their discussions and their friendship. It was a pleasure working with you!

I further want to mention Chantal Lucini, Jens Pontiller, David Reinhart, Milica Sevo, Frank Stolz, Christopher Taus and Jakob Wallner. I appreciate all your contributions of time, ideas and experience, supporting me to overcome experimental issues of my work.

Finally I would like to thank family and friends for showing their interests and support and for making my study period unforgettable. Especially I want to thank my parents for their support and believe in me. I am very grateful that they made it possible for me to finish my education.

Abstract

Mammalian expression systems are currently widely used for the production of therapeutic proteins. Among the various production hosts, the Chinese hamster ovary (CHO) cell line has reached a predominant status accounting for more than 70% of therapeutic protein production. To express heterologous genes strong viral derived promoters such as the cytomegalovirus (CMV) major immediate early promoter or the Simian Virus 40 (SV40) immediate early promoter are commonly used. High expression levels achieved with viral derived promoter sequences inflict cellular stress leading to various unwanted side effects. Common stress reactions affect correct assembly and folding of proteins, silencing of introduced heterologous genes or even lead to premature apoptosis of cells. These undesired effects could be avoided by using endogenous gene regulatory elements, which are embedded in a cells regulatory network. Another approach is to analyze existing promoter sequences for their distinctive nucleotide patterns to construct artificial promoters based on that information.

In the master thesis at hand the first part describes the identification of endogenous regulatory elements using Chromatin Immunoprecipitation (ChIP). This method is used to capture *in vivo* protein DNA interactions occurring at promoter and enhancer regions of genes. ChIP derived sequences were analyzed with different promoter prediction programs and mapped to the human reference genome, evaluating the results with the ENCODE database for functional elements in the genome.

In the second part artificial promoter constructs were analyzed using a luciferase reporter system. Different sequence motifs were introduced into the constructs, to study influence on transcriptional activity. Furthermore it was shown that promoter activity of artificial constructs comprising the known response element NFκB can be induced by a stimulus with TNFα.

Zusammenfassung

Die CHO (Chinese Hamster Ovary) Zellinie ist eine der am meist genutzten Produktionssystemen zur Herstellung von rekombinanten Proteinen für therapeutische Anwendungen. Um Gene, die für diese Proteine codieren erfolgreich zu transkribieren werden derzeit relativ starke virale Promotoren genutzt. Diese Sequenzen enthalten den Cytomegalo Virus (CMV) major immediate early Promoter oder den Simian Virus 40 (SV40) immediate early Promoter. Eine ständige hohe Expressionsrate kann jedoch zu Stressreaktionen in der Zelle führen. Diese Reaktionen spiegeln sich in einer nicht korrekten Faltung von Proteinen, Inaktivierung der für das Produkt codierenden Gene, oder einer frühzeitigen Apoptose der Zellen wider.

Um diese unerwünschten Reaktionen zu vermeiden könnten aus den jeweiligen Produktionszellen stammende „endogene" Regulationselemente wie etwa Promotoren eingesetzt werden. Diese Elemente besitzen den Vorteil Teil des zelleigenen Regulationsnetzwerkes zu sein und könnten für längere und stabilere Expressionsraten und somit höherer Produktausbeute sorgen.

Ein anderer Ansatz verfolgt die Idee in hoch transkribierten Genen vorkommende Promotoren auf bestimmte Merkmale ihrer Sequenzen hin zu untersuchen. Die zugrunde liegenden Prinzipien könnten genutzt werden um künstliche Promoter Sequenzen zu erstellen und diese in ihrer Stärke und Induzierbarkeit variabel zu gestalten.

In der vorliegenden Arbeit wurde versucht endogene Promotor Sequenzen mittels Chromatin-Immuno-Präzipitation (ChIP) aus CHO Zellen zu isolieren. Mit ChIP können *in vivo* Interaktionen von DNA und DNA-bindenden Proteinen, die eine wichtige Rolle bei der Regulation der Transkription spielen, ermittelt werden. Da keine Genom Sequenz für den Chinesischen Hamster verfügbar ist, und die isolierten Sequenzen nur geringe *in vitro* Promoteraktivität aufwiesen wurde ChIP auch mit einer humanen Zellline (human embryonal kidney, HEK293) durchgeführt um die Methode zu evaluieren.

Im zweiten Teil wurden künstliche Promoter Sequenzen mit einem Luziferase Reporter gene assay untersucht. Mit dem Ziel Promoter Aktivität in einer vorhersagbaren Weise zu steuern wurden verschiedene Sequenz Motive analysiert. Weiters konnte die Induzierbarkeit eines künstlichen Promoters als Antwort auf einen äußeren Stimulus gezeigt werden.

Contents

1 Introduction

1.1 Recombinant protein production in mammalian cells

Therapeutic proteins currently represent the main drivers for revenue growth in the biopharmaceutical industry and have a demonstrated track record of providing safe and efficacious clinical benefit. In 1986 human tissue plasminogen activator (tPA, Activase; Genentech, S.San Francisco, CA, USA) became the first therapeutic protein from recombinant mammalian cells to obtain market approval. Today about 60-70% of all recombinant protein pharmaceuticals are produced in mammalian cells. In 2009, US market sales for biologicals reached $48.2 billion. Among the top categories listed, monoclonal antibodies (mAbs), growth factors and hormones made up two thirds of the total market sales (Aggarwal, 2010). Like tPA, many of these proteins are expressed in immortalized Chinese hamster ovary (CHO) cells, but other cell lines, such as those derived from mouse myeloma (NS0), baby hamster kidney (BHK), human embryo kidney (HEK-293) and human retinal cells have gained regulatory approval for recombinant protein production (Wurm, 2004).

1.1.1 The CHO expression system

The efficacy of a recombinant protein used for clinical applications is greatly dependent on the production host cell line. For many applications, mammalian cells are superior as compared to other eukaryotic or bacterial production systems because of their capacity for proper protein folding, assembly and post-translational modifications. Especially the oligosaccharide moiety of proteins is often essential for recognition, signalling, and interaction events within and between cells and proteins, and can play an important role in folding and defining the conformation of the protein. CHO cells have proven to be suited for the demand to obtain complex, modified proteins, considering their ability to achieve human like glycosylation patterns (Li and d'Anjou, 2009). Another key issue in regard to the selection of the most suited host system is product safety. Production cell lines must be free from any human pathogenic agents. The capacity to grow in single cell suspension led to the dominating status of CHO cells in industrial scale mass production. The introduction of foreign DNA and

accessibility to genetic modification is also an important factor for high level protein production. Comprising many of these properties, CHO cells are the most widely used mammalian host for the production of therapeutic proteins (Jayapal et al., 2007). After isolation and establishment of the original CHO cell line by Theodore T. Puck of the Department of Medicine at the University of Colorado in 1957 (Puck et al., 1958) it became obvious that these cells were quite resilient and suitable for in vitro cultivation with relatively fast generation times. The next milestone was achieved by Urlaub and Chasin in 1980 (Urlaub and Chasin, 1980) by applying mutagenesis and selection for auxotrophs that exhibited certain nutritional requirements for maintaining growth and viability over long culture periods. The dihydrofolate reductase deficient (DHFR-) CHO cell line, selected from thymidine auxotrophs, is now the most widely used host cell line. DHFR is an important enzyme in the the formation of tetrahydrofolic acid, which is an essential cofactor of the de novo pathways for nuclide acid formation. $DHFR^-$ cells are auxotrophs for glycine, hypoxanthine, and thymidine, and hence these nutrients must be supplemented for growth. Transfection of cells using the heterologous gene combined with a functional copy of the DHFR gene allows a clonal selection when grown in media lacking these supplements. This system also enables the amplification of the introduced gene of interest. By blocking DHFR activity with high levels of methotrexate (MTX), which is a folic acid analog, the cells respond in amplifying the copy number of the DHFR gene. This effect also leads to the co amplification of the gene of interest.

The main drawback of CHO as well as other mammalian cell lines for the production of therapeutic proteins is the relative low volumetric yield of product as compared to bacterial production systems. 10 to 100-fold lower productivity reflects in larger and more expensive production facilities and is a major cost factor. Although major advances have been made over the past few decades in respect of maximisation of therapeutic protein product yield, improvement of specific productivity and viable cell biomass as well as stability of production over extended periods of culture is of utmost importance. The production of proteins from DNA is a complex process that involves several steps, including transcription, posttranscriptional processing, translation, posttranslational processing, and secretion (depending upon whether the particular protein is intracellular or secreted). The level of protein expression from cells is determined by the efficiency of each of these individual points of control. Traditionally,

transcription is considered to be the dominant factor controlling expression. Consequently vector design and engineering are among the various strategies to improve yield (Barnes et al., 2003).

1.1.2 Expression vectors

The essential elements of mammalian expression vectors include a constitutive or inducible promoter capable of robust transcriptional activity, optimized mRNA processing and translational signals that include a Kozak sequence, translational termination codon, mRNA cleavage and polyadenylation signals as well as mRNA splicing signals, a transcription terminator, selection markers for the preparation of stable cell lines and for gene amplification, and prokaryotic origin of replication and selection markers for vector propagation in bacteria (Makrides, 1999).

1.1.3 Promoters used in expression vectors

Promoters can be divided into two classes, those that function constitutively and those that are regulated by induction or repression. Promoters used for high-level production of proteins in mammalian cells should be strong and, preferably, active in a wide range of cell types to permit qualitative and quantitative evaluation of the recombinant protein. Constitutive promoters commonly used in mammalian systems are the simian virus 40 early promoter (SV40), cytomegalovirus immediate-early promoter (CMV), human Ubiquitin C promoter (UBC), human elongation factor 1alpha promoter (EF1A), mouse phosphoglycerate kinase 1 promoter (PGK), and chicken β-Actin promoter coupled with CMV early enhancer (CAGG) (Qin et al., 2010). The drawbacks of viral derived sequences are that they exploit the host transcriptional machinery to a very high extend, which can lead to undesired stress reactions, like the unfolded protein response (UPR) and the endoplasmatic reticulum (ER) stress response and transgene silencing (Brooks et al., 2004). Another issue is the cell cycle dependency of these promoters contributing to the variable expression profiles among the producing cells. Thus, the use of endogenous promoters or completely artificially tailored promoter sequences (discussed later) driving transgene expression are of great interest.

1.1.4 Endogenous promoter/enhancer elements

It has been described that host derived sequences like the 5’ and 3’ flanking regions (~19kb) from the highly expressed housekeeping Chinese hamster EF-1α (CHEF1) gene are able to achieve high level transgene expression exceeding commercial available vectors (including the CMV promoter) even in different cell lines (Deer and Allison, 2004). The analysis of highly expressed genes throughout the exponential and stationary phase of a typical fed batch fermentation led to the identification of the housekeeping ferritin heavy chain (HC) gene locus (Prentice et al., 2007). Like with the CHEF1 sequences, cloning of 5’ and 3’ flanking regions (~13kb) in an expression vector resulted in significantly higher expression levels compared to viral promoters. The length of these regulatory regions show that many more factors besides the core promoter immediately adjacent to the TSS of a gene contribute to stable long term expression. Expression vectors containing only promoters and polyadenylation sequences tend to exhibit wide variations in expression levels due to integration site effects that often lead to repression of transcription (Wilson et al., 1990). Regulatory elements like insulators, enhancers, or matrix attachment regions (MARs) (Kwaks and Otte, 2006), (discussed in greater detail below) found in these sequences, tend to reduce the effect of the integration site dependent host cell chromatin environment.

1.2 Transcriptional Regulation in Eukaryotes

In living organisms only some of the genes are expressed at any given moment. Gene expression profiles underlie considerable changes during the life cycle, among cell types, and in response to fluctuating physiological and environmental conditions. Controlled gene expression is critical for both the homeostasis of cells and the programmed development of multicellular organisms. Protein coding genes are transcribed into mRNA by a large molecular machine termed RNA polymerase II (Pol II). It is common in eukaryotes for genes to be controlled by many (sometimes hundreds) of different regulatory proteins. This complexity is possible because many eukaryotic gene regulatory proteins can act over very large distances (tens of thousands of nucleotide pairs) along the flexible DNA helix, allowing an almost unlimited number of them to influence the expression of a single gene. The packaging of eukaryotic DNA into chromatin provides another level of "epigenetic" regulation adding even more complexity. The time and place that each gene is transcribed, as well as its rates of transcription under different conditions, are determined by the spectrum of gene regulatory proteins (trans-acting) that bind to the regulatory region elements (cis-acting) of the gene.

1.2.1 Focused versus dispersed transcription initiation

Examination of the patterns of transcription initiation reveals two different modes of transcription initiation – focused and dispersed (Fig. 1). In focused initiation, transcription starts at a single nucleotide or within a narrow region of several nucleotides, whereas in dispersed initiation, there are multiple weak start sites over a broad region of about 50 to 100 nucleotides. Some promoters exhibit the combined qualities of both focused and dispersed promoters – for instance, a promoter might have multiple dispersed start sites with one particularly strong start site.

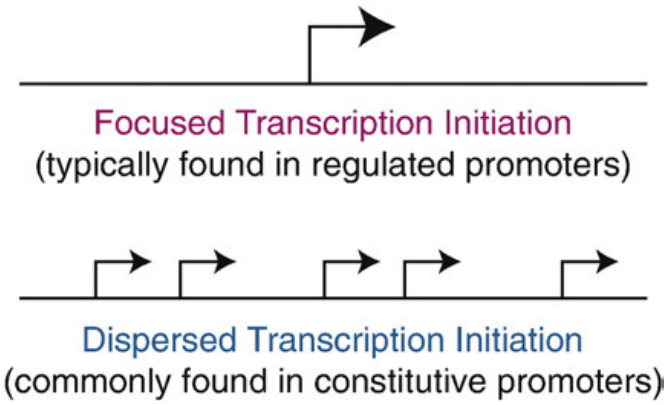


Figure 1 : Focused versus dispersed transcription initiation (Juven-Gershon and Kadonaga, 2010).

Focused transcription initiation occurs in all organisms, and appears to be the predominant or exclusive mode of transcription in simpler organisms. In vertebrates, however, about 70% of genes have dispersed promoters, which are typically found in CpG islands. Generally, it appears that focused promoters are associated with regulated genes, whereas dispersed promoters are used in constitutive genes (Juven-Gershon and Kadonaga, 2010).

1.2.1.1 Focused core promoters and the general transcription factors

The region -35 and +35 base pairs to the transcription start site (TSS) is referred to as the core promoter (reviewed by Smale and Kadonaga, 2003, Kadonaga, 2004). By definition it is the minimal DNA region that is sufficient to direct low levels of activator independent (basal) transcription by Pol II in vitro. Accurate and efficient transcription from the core promoter requires the polymerase along with auxiliary factors that are commonly termed the "basal" or "general" transcription factors, which include transcription factor (TF) IIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH, depicted in Figure 2. With TATA box-dependent core promoters, it has been found that the purified factors can assemble into a transcription preinitiation complex (PIC) in the following order: TFIID, TFIIB, RNA polymerase II-TFIIF complex, TFIIE, and then TFIIH (Butler and Kadonaga, 2002). TFIID is a key factor in basal transcription that is involved in recognition of focused core promoters (Thomas and Chiang, 2006). The multi-subunit complex is composed of the TATA-binding protein (TBP) and 13-14 TBP-associated factors (TAFs) (Albright and Tjian, 2000). They are distinguished by a number, which indicates their apparent or predicted molecular weight, for example (TAF II p250) (Tora, 2002). Binding of TFIID causes a large distortion in

the DNA of the TATA box. This distortion is thought to serve as a physical landmark for the location of an active promoter in the midst of a very large genome, and it brings DNA sequences on both sides of the distortion together to allow for subsequent protein assembly steps. TFIIB is a single polypeptide that interacts with TBP as well as with the DNA upstream of the TATA box the BRE element and also directly modulates the catalytic center of RNA polymerase II in the transcription complex. Moreover, TFIIB has been proposed as a target of transcriptional activator proteins that act to stimulate preinitiation complex assembly (Deng and Roberts, 2007). Recently it has been found, that TFIIB binds also terminator regions and helps to redirect Pol II to the promoter starting the next transcription circle in yeast (Singh and Hampsey, 2007).

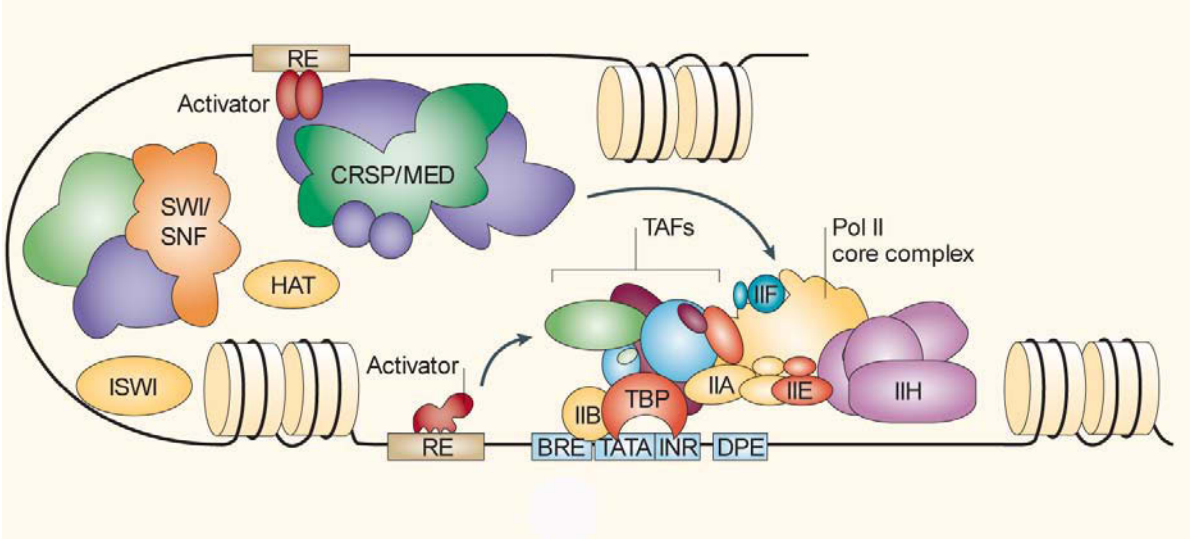


Figure 2: Schematic model of the metazoan pre-initiation complex (PIC), comprising the general Transcriptionfactors (GTFs) TFIIA-H, which mediate binding of Pol II to core promoter motifs (BRE, TATA-box, INR, DPE). ISWI (imitation switch) and histone acetyl transferases (HATs) work to remodel and modify chromatin structure and can be present in numerous multisubunit complexes. SWI/SNF complexes also function to remodel chromatin structure, whereas CRSP/MED is thought to mediate signals between enhancer-bound factors (RE….response element) and the transcriptional machinery (D. J. Taatjes et al., 2004).

TFIIA facilitates DNA binding by TBP in vivo. TFIIA may be regulated by factors that target distinct regions of the protein, but promoters vary significantly in the degree to which they require TFIIA for activation (Liu et al., 1999). TFIIF stabilizes Pol II interaction with TBP and TFIIB and helps attract TFIIE and TFIIH. TFIIE attracts and regulates TFIIH which contains a DNA helicase unwinding the DNA completing the PIC assembly (Deng and Roberts, 2007). TFIID and TFIIB

are the two TFs that directly interact with different cis-acting DNA sequences, termed core promoter motifs adapted from (Dylan J. Taatjes et al., 2004).

1.2.1.1 Core promoter motifs

Core promoters in higher eukaryotes are highly diverse in structure, and each core promoter element is only found in a subset of genes. Several known sequence motifs contribute to core promoter activity: the TATA box, BREu (upstream TFIIB recognition element), Inr (initiator), MTE (motif ten element), DPE (downstream promoter element), DCE (downstream core element), and XCPE1 (X core promoter element 1) (Fig. 2) (Juven-Gershon and Kadonaga, 2010).

The initiator element (Inr) , located around the TSS, is one of the most common core promoter elements (FitzGerald et al., 2006). The Inr consensus sequence was determined to be YYANWYY in humans and TCAKTY in Drosophila. The A nucleotide in the middle of the Inr consensus is often the +1 start site in focused core promoters. Although several factors are reported binders, TAFII250 and TAFII150, subunits of TFIID bind the Inr directly (Chalkley and Verrijzer, 1999). The basal activity of TATA box-containing promoters is dramatically enhanced by the initiator element (INR), which can function in concert with the TATA box in a synergistic manner. The TATA and Inr elements are the only known core promoter elements that, alone, can recruit the PIC and initiate transcription (Sandelin et al., 2007).

The first identified core promoter element was the TATA box (consensus sequence in metazoans TATAWAAR) typically 25 to 30 nucleotides upstream from the TSS (Breathnach and Chambon, 1981). As mentioned above, TATA box binding is mediated by TFIID subunit TATA box binding protein (TBP). Both, the TATA box as well as TBP are conserved from archaebacteria to humans (Hernandez, 1993). Even though TATA box is a well known core promoter motif, it is present in only about 10%-15% of mammalian core promoters (Kim et al., 2005).

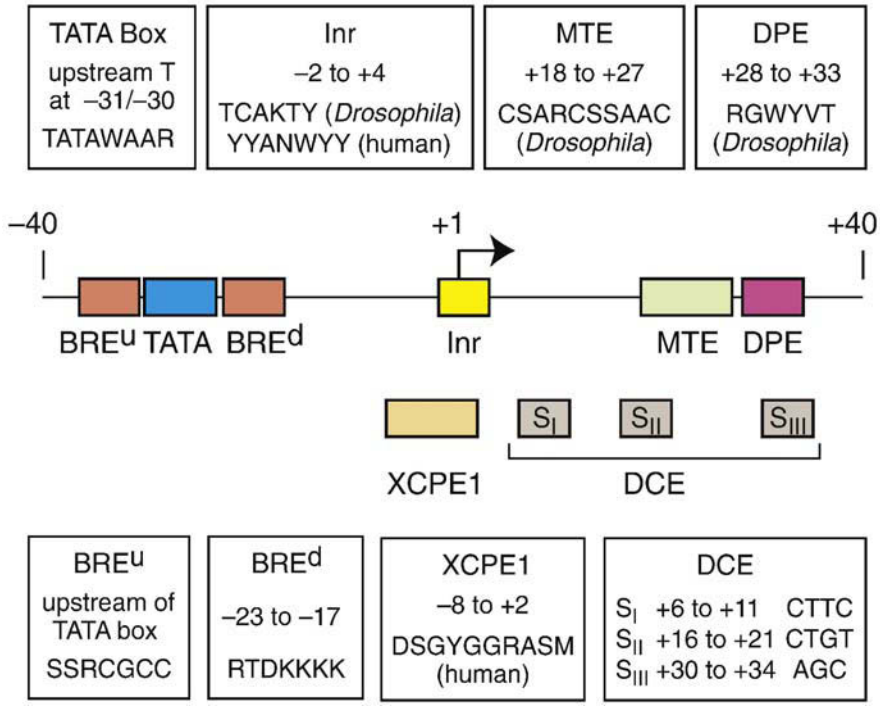


Figure 3: Core promoter motifs typically found in focused promoters. Any specific core promoter may contain some, all, or none of these motifs. The transcription start site is indicated (A+1 in the Inr) (Juven-Gershon and Kadonaga, 2010)

Nucleotide Code:	Base:
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base

Figure 4: IUPAC nucleotide code

Located immediately upstream of a subset (~10%-30%) of TATA box motifs, the upstream TFIIB recognition element (BREu), consensus sequence SSRCGCC, was

identified (Lagrange et al., 1998). Additionally a second TFIIB recognition site, the downstream TFIIB recognition element (BREd), consensus sequence RTDKKKK, was found immediately downstream of the TATA box (Deng and Roberts, 2005). BREu and BREd enhance the formation of the TFIIB-TBP-promoter complex, resulting in higher transcriptional activity. However, BRE also suppresses the basal level of transcription elicited by a core promoter, which increases the amplitude of transcriptional stimulation in the presence of an activator protein (Evans et al., 2001).

Further TFIID recognition sites are the downstream promoter element (DPE) and the motif ten element (MTE) (Burke and Kadonaga, 1996, Ohler et al., 2002). DPE shows sequence conservation from Drosophila to humans (consensus sequence RGWYVT) and functions in concert with the Inr element where spacing between the two elements is critical for transcription. Also the MTE element (consensus sequence CSARCSSAAC) functions cooperatively with the Inr in a spacing dependent manner. The downstream core element (DCE) was discovered in the human ß-globin promoter, and is present in a variety of core promoters containing a TATA motif (Lee et al., 2005). The function of the DCE requires the binding of the TAF subunits to TFIID, whereas TAF1 interacts with the DCE. In recent years two other core promoter elements, the X core promoter element 1 (XCPE1) (Tokusumi et al., 2007) and the X core promoter element 2 (XCPE2) (Anish et al., 2009) have been identified. These elements occur in TATA-less promoters at low frequencies at about 1% of human core promoters.

1.2.1.2 CpG islands (CGIs)

As mentioned above, approximately 60-70% of human genes have dispersed promoters, comprising CpG islands (CGIs). These promoters are mostly associated with both housekeeping and strong tissue-specific genes. These CGIs are regions of DNA, spanning approximately 1000 base pairs, with a high G + C content (greater than 50%) and a high frequency of CpG dinucleotides relative to the bulk genome (Bird, 1986, Gardiner-Garden and Frommer, 1987). These CpG clusters are usually devoid of CpG methylation, whereas the bulk genome is methylated at 70-80% of CpGs (Bird, 1999). 5-Methylcytosine spontaneously deaminates to Thymine (Figure 5). This mutation (T/G mismatch) is not efficiently detected and repaired by DNA repair enzymes like Thymine-DNA-

glycosylase (TGD). As a result there is a depletion of the CpG dinucleotide in the genome over evolutionary time, unless there is a selective pressure to keep this motif.

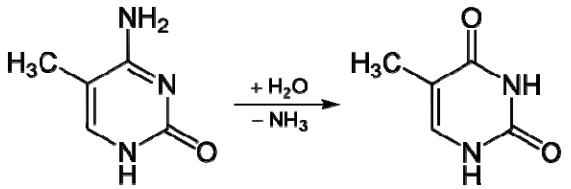


Figure 5: Spontaneous deamination of 5-Methylcytosine to Thymine

Transcriptional activity of genes is often controlled by methylation of cytosines at CpG dinucleotides in their promoter regions. In general, densely methylated promoters are silenced, most likely through the binding of repressor proteins like ubiquitously bound Sp1 (Macleod et al., 1994), whereas unmethylated promoters are largely active. Unmethylated CpG islands also tend to have an unusual, open chromatin organization. A very prominent histone modification coinciding with most promoter CGIs is the trimethylation of lysine 4 of histone H3 (H3K4me3) (Guenther et al., 2007). Recently it was shown that the insertion of a promoterless stretch of CpG-rich DNA into the mouse genome is sufficient to create a novel peak of H3K4me3. The formation of H3K4me3 is mediated by the zinc finger protein Cfp1, which contains a CXXC domain that specifically binds to CpG only when it is unmethylated. Cfp1 is a component of the Set1 complex which subsequently trimethylates histone H3 lysine 4 (Thomson et al., 2010).

1.2.1.3 The chromatin environment

Histone modifications play an important role in transcriptional activation or repression. Nuclear chromatin is a nucleoprotein complex in which DNA is wrapped around octamers of histone proteins called "core histones", which include two copies of each histone type 2A, 2B, 3 and 4. The fundamental repetitive unit of chromatin is called nucleosome, defined as one core histone octamer and the DNA wrapped around it. Transcription of a given gene is affected by the architecture of chromatin in which the gene is embedded. Simplified there are two global chromatin states in a cell: silent heterochromatin and active euchromatin, each of these associated with a distinct set of modifications. Modifications can be divided into those that correlate with activation and those that correlate with repression. Acetylation, methylation, phosphorylation, and ubiquitination have been implicated in activation whereas

methylation, ubiquitination, sumoylation, deimination, and proline isomerization have been implicated in repression (Berger, 2002). Active promoters are marked by acetylation of various residues of histones H3 and H4 and trimethylation of H3K4. Chromatin immunoprecipitation (ChIP) and microarray (ChIP-chip) experiments are able to link chromatin marks with the occurrence of regulatory elements like promoters and enhancers (Heintzman et al., 2007). As silencing is an issue in heterologous gene expression the cross talk between cis-acting DNA elements and the subsequent establishment of a transcriptional active chromatin environment are of interest. A recent review outlines various examples of genetic variants altering epigenetic marks, elucidating how genetic and epigenetic mechanisms co-participate in regulating transcription (Zaina et al., 2010). Genetic variants are sometimes associated with changes in DNA methylation patterns or histone marks, which in turn can provoke rearrangements of chromatin architecture that may have long-range transcriptional effects. As chromosomal rearrangements are common in CHO production cell lines (Ruiz and Wahl, 1990), it is of interest to keep the chromatin environment of the transgene accessible for sustained protein production in stable cell lines. For transient assays performed in the thesis at hand, these effects do not play a role.

1.2.2 Distal regulatory elements

A distinguishing mark of eukaryotic genomes is the long distance (often several hundred kilobases) over which distant cis-acting elements such as enhancers, insulators, locus control regions, and silencing elements co-regulate transcription. The long distances are overcome by a DNA looping mechanism, bringing core promoter and distant regulatory elements in close proximity (reviewed by Vilar and Saiz, 2005).

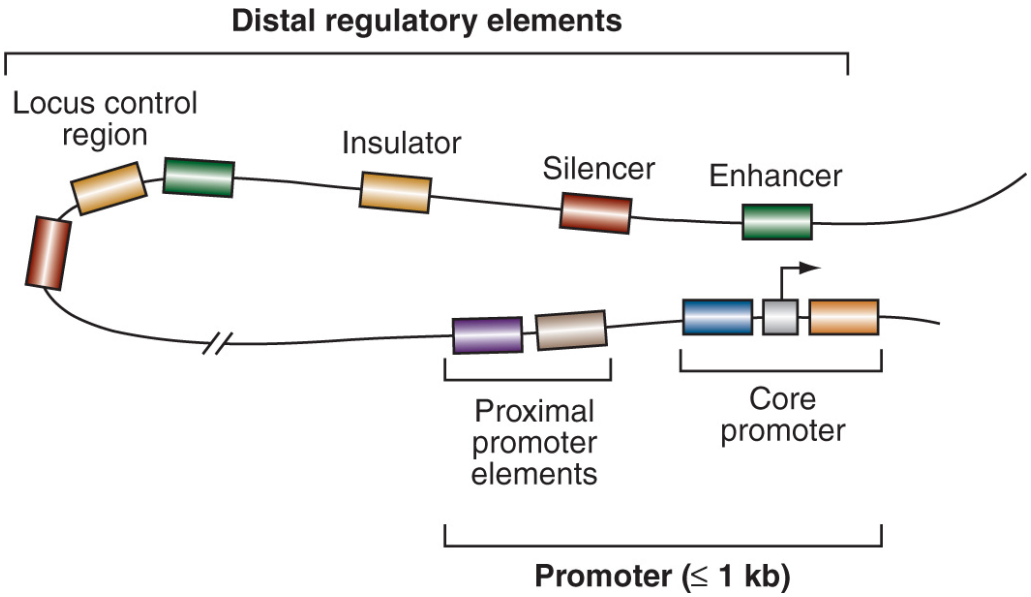


Figure 6: Schematic illustration of a typical gene regulatory region. The promoter, which is composed of a core promoter and proximal promoter elements, typically spans less than 1 kb pairs. Distal (upstream) regulatory elements, which can include enhancers, silencers, insulators, and locus control regions, can be located up to 1 Mb pairs from the promoter. These distal elements may contact the core promoter or proximal promoter through a mechanism that involves looping out the intervening DNA (Maston, Evans et al. 2006)

1.2.2.1 Enhancers

Enhancers can be defined as DNA sequences that serve to recruit transcription factors which promote the decondensation of repressed chromatin and/or facilitate the assembly of the transcription machinery at gene promoters (Maston et al., 2006). There are three known mechanisms by which enhancer bound TFs act to enhance transcription at target promoters via protein cofactors. The recruitment of the SWI/SNF complex modifies the chromatin structure in an ATP dependent fashion resulting in nucleosome modeling (Peterson, 1995). Another class of cofactors remodels chromatin structure by covalently modifying the N-terminal tails of histones either directly decondensing packed nucleosomes, or establishing platforms for the recruitment of additional remodeling factors (Zhang and Reinberg, 2001). The third mechanism involves the Mediator, a transcriptional coactivator, which forms a complex with cohesin, which can form rings that connect two DNA segments via DNA looping (Kagey et al., 2010). Enhancers show more cell type specificity, compared to that of promoters and insulators and are therefore thought to play an important role in cell type specific gene expression (Visel et al., 2009). New evidence is indicating that long ncRNAs

are associated with enhancer regions and that such non-coding transcription correlates with the increased activity of the neighboring genes. Additional experiments suggest that enhancer-function can be mediated through a transcribed long ncRNA and that this might be a common function for long ncRNAs (Orom and Shiekhattar, 2011).

1.2.2.2 Insulators

Insulators are elements that restrict transcriptional enhancers from activation of unrelated promoters or act as a barrier preventing the expansion of repressive heterochromatin (Burgess-Beusse et al., 2002). Enhancer blocking in vertebrates requires the association of the protein CCCTC binding factor (CTCF), a large protein with 11 zinc finger motifs in its DNA binding domain (Bell et al., 1999). Like with enhancers, there is again a close association between CTCF and the cohesin complex, which is stabilizing long range interactions (Wendt et al., 2008).

1.2.2.3 Silencers

Silencers are structural related to enhancers, but confer a negative effect on the transcription of a target gene. They provide binding sites for repressors which are thought to block binding of a nearby activator (Nakabayashi et al., 1991), recruit histone modifying enzymes establishing a more repressive chromatin state (Lynch and Rusche, 2009).

1.2.2.4 Nuclear scaffold/matrix attachment regions

Regulation of transcription also occurs through scaffold/matrix attachment regions (S/MARs), the chromatin regions that bind the nuclear matrix. The nuclear laminar consists of a network of Ribonucleoproteins RNPs and other nonhistone proteins, that serves as a scaffold for loops of chromatin (reviewed by Bode et al., 2000). Common for S/MARs are AT-rich sequences, which are responsible for hairpin structures and a high unwinding potential. They are frequently located close to enhancers but also in introns or at borders of transcription units.

1.2.2.5 Locus control regions (LCRs)

If multiple of the above described cis-acting elements are involved in regulating an entire locus or gene cluster they are referred to as Locus control regions

(LCRs). The identification of LCRs in the mammalian ß-globin locus and the characterization of LCRs in other loci have proofed the looping model for long range physical interactions (Noordermeer et al., 2008).

1.3 The CHO-Genome and Transcriptome

Despite the extensive use of CHO cells as a production host, little sequence information is available for molecular based research. Bacterial artificial chromosome (BAC) libraries derived from CHO cells provide information for physical mapping of the CHO genome (Omasa et al., 2009). The Isolation and sequencing of expressed sequence tags (ESTs) from CHO cell cDNA libraries has generated databases containing more than 60,000 sequences and allowed for the development of CHO-specific DNA microarrays (Kantardjieff et al., 2006). Transcriptome studies by means of microarray technology are currently used to examine differential expression of high-producing cell lines and to identify gene candidates for host cell engineering (Nissom et al., 2006, Yee et al., 2009). Such studies are limited in that only a subset of genes expressed at sufficiently high levels is captured for sequence analysis, completely lacking the information of regulatory elements of non transcribed parts of the genome. As next-generation sequencing technologies including 454 and Illumina are currently evolving, rapid acquisition of genomic sequence from CHO cell lines is possible (Birzele et al., 2010). The limitation of NGS methods is the low coverage (1x) of the CHO genome, insufficient for *de novo* assembly, requiring mapping to annotated mouse and rat genomes which show a high degree of DNA sequence homology with the Chinese hamster. As regulatory elements are less conserved among different species than coding regions, a comparative genomic approach with the mentioned methods above might not be successful. Isolation and identification of regulatory elements applying methods, not requiring a reference genome, provides an alternative to the above mentioned approaches.

1.4 Artificial promoters

Having discussed the drawbacks of viral derived promoter sequences in chapter 1.1.3, a more desirable solution are "artificial" promoters, i.e. "made-up"

sequences that are not found in living organisms, that can be engineered to the required behavior and expression levels. The more predictable these sequences, the easier it is to optimize a system for recombinant protein production.

Designing artificial sequences that attract transcription factors and initiate transcription requires determining the features that capture the interactions between these proteins and the DNA. Highly active promoters exhibit nucleotide patterns that are different from the majority of the genome, most notably the abundance of GC and the di-nucleotide CpG. Since methylated CpG is subject to spontaneous deamination, it is the most infrequent dinucleotide genome-wide, unless there is a selective pressure to keep it. CpG-containing motifs have been reported to be both necessary and sufficient to bind Pol II abundantly in more than one tissue to transcribe both housekeeping genes and genes with tissue-specific expression in multiple cell types. Grabherr et al. analyzed the genome-wide frequencies of short motifs showing that infrequent occurring nucleotide compositions are a distinctive attribute of promoters (Grabherr et al., 2011). In the study a score (the "αscore") was computed for 12 consecutive base pairs ("12-mer"), based on the frequencies of di-nucleotides, tri-nucleotides etc. compared to the genome-wide expectation.

The α score: Let N denote the number of k-mers (k consecutive nucleotides, k = 2, 3…12) in the genome, and φ_k^i the genome wide occurrence count of the k-mer starting at position i (using zero-based counting) in the 12 base pair (bp) long sequence, then the score α_k is

$$\alpha_k = -\ln\frac{{\varphi_k}^0}{N} - \sum_{i=0}^{12-k} \ln\frac{{\varphi_k}^i}{{\varphi_{k-1}}^i}$$

and the composite score for all k

$$\alpha = \sum_{k=2}^{12} \alpha_k$$

This measure (α_k) tracks with the G/C and CpG content of highly active promoters, but also incorporates contributions from other nucleotides, and can be computed for very short sequences so that these can be quantified as "promoter-like" on a sliding scale.

To create longer, contiguous artificial promoter sequences, the *a* score for each possible 12-mer, including those not present in the human genome was calculated, subsequently overlapping uncommon motifs to build a "promoter-like" 160,000 nucleotide (nt) long contiguous sequence, which is very rich in G/C (60%) and CpG (22%), and contains exact instances of the consensus of known binding sites, such as the TFIIB Recognition Element (BRE), TATA-Box, CAAT-Box, and Inr. The sequence has no homology to any human (or any other sequenced) genome over more than 18 base pairs.

Using *in silico* promoter prediction programs based on fruit fly (Reese, 2001) different promoter constructs using different criteria and of different lengths (50, 110, 200, 232 and 300 base pairs) were selected from the concatomer. All constructs except a 50 nt long sequence gave rise to strong *in vitro* promoter activity in a reporter gene assay in four mammalian cell lines: CHO (hamster ovary); P19 (mouse embryo); Vero (monkey kidney); and HEK293 (human kidney). The 232 (bp) base pair long sequence Ars 232 is made of contiguous sequences out of the concatomer containing 2 high-scores, and was used in the thesis at hand to test if artificial sequences might comprise a scaffold for the incorporation of desired gene regulatory elements and transcription factor binding sites that meet the requirement of the particular production system.

1.5 Identification of gene regulatory elements

In eukaryotes ranging from nematodes to flies to the mammals, the number of coding genes is similar and it is now thought that an increased number of transcriptional regulatory elements are contributing to the observed species variety and cell differentiation (Levine and Tjian, 2003). The identification of these elements controlling gene expression, and characterization of their interaction with the respective TFs is therefore of great interest. Although computational approaches to identify regulatory elements are applied to the ever growing number of sequenced genomes (Ohler et al., 2002), due to the lack of an existing reference genome for the Chinese Hamster different experimental strategies are available to overcome this limitation. In recent years one of the most applied method to identify regulatory sequences bound by TFs was

Chromatin Immunoprecipitation (ChIP) (reviewed by Collas, 2010). Basically ChIP begins with chemically crosslinking the cells proteins and DNA within 1 Å, delivering a snapshot of the cells current DNA-protein interactions. The complexes are then sheared by supersonic and the protein of interest is precipitated with specific antibodies against it. Reverse crosslinking elutes the DNA sequences bound by the protein. The derived sequences can further be amplified, labelled, and hybridized to a DNA microarray. This method is referred to as ChIP-Chip and can provide a genome wide analysis of Protein-DNA interactions. As massive parallel sequencing methods evolved in recent years, ChiP-seq can be used to directly sequence the obtained fragments instead of hybridising them. ChIP-seq results in higher resolution, fewer artefacts, greater coverage and a larger dynamic range than ChIP-chip (Park, 2009). The main drawback of massive parallel sequencing is that they are dependent on a reference genome for mapping the reads. Although expression profiling without genome sequence information in CHO cells by next generation sequencing methods was successful (Birzele et al., 2010), regulatory elements comprise less conserved and shorter sequences to be mapped to genomes of related organisms like Mus musculus. Other methods like rapid amplification of cDNA ends (RACE) (Frohman et al., 1988), cap analysis of gene expression (CAGE) (Shiraki et al., 2003) or gene identification signature (GIS) (Ng et al., 2005) can be applied to search for cis-regulatory sequences upstream of the 5’ end of highly transcribed genes. As with ChIP-seq mapping the derived sequences to a reference genome is currently not possible in the case of the Chinese hamster and therefore the major limitation.

1.6 The nuclear factor kappa B (NF-κB)

The eukaryotic transcription factor NF-κB was identified as a protein that bound to the intronic enhancer of the immunoglobulin kappa light chain in mature B- and plasma cells but not pre B-cells (Sen and Baltimore, 1986). In mammals, the NFkB family is composed of five related transcription factors: p50, p52, RelA (p65), c-Rel and RelB, and they form a variety of homodimers and heterodimers, each of which activates its own characteristic set of genes (Hoffmann et al., 2006).These transcription factors are related through an N-terminal, 300 amino acid, DNA binding/dimerization domain, called the Rel

homology domain (RHD), through which they can form homodimers and heterodimers that bind to 9-10 base pair DNA sequence (the commonest form of which is GGGACTTTCC), known as kB sites, in the promoters and enhancer regions of genes, thereby modulating gene expression. Members of the nuclear factor kappa B (NF-κB) family of dimeric transcription factors (TFs) regulate expression of a large number of genes involved in immune responses, inflammation, cell survival. Moreover, pathological dysregulation of NFkB is linked to inflammatory and autoimmune diseases as well as cancer (reviewed by May, 2006).

1.6.1 NfkB signal transduction

Various cell-surface receptors activate the NFkB signaling pathway in animal cells. Toll receptors in Drosophila and Toll-like receptors in vertebrates, for example, recognize pathogens and activate this pathway in triggering innate immune responses. The receptors for tumor necrosis factor a (TNFα) and interleukin-1 (IL1), which are vertebrate cytokines especially important in inducing inflammatory responses, also activate this pathway. The Toll, Toll-like, and IL1 receptors belong to the same family of proteins, whereas TNF receptors belong to a different family; all of them, however, act in similar ways to activate NFkB. When activated, they trigger a multiprotein ubiquitylation and phosphorylation cascade that releases NFkB from an inhibitory protein complex, so that it can translocate to the nucleus and turn on the transcription of hundreds of genes that participate in inflammatory and innate immune responses. Further stimulants activating NfkB are bacterial like lipopoly-saccharide LPS and viral products, UV irradiation and oxidative stress (Gilmore, 2006). TNF was first identified as a soluble cytokine that is produced upon activation by the immune system, able to exert significant cytotoxicity on many tumor cell lines and to cause tumor necrosis in certain animal model systems (Gray et al., 1984). Research during the past two decades has shown the existence of a superfamily of TNF proteins consisting of 19 members that signal through 29 receptors (example given in Figure 7) These ligands while regulating normal functions such as immune responses (via the NfκB-pathway), haematopoiesis and morphogenesis, have also been implicated in tumorigenesis, transplant rejection, septic shock, viral replication, bone resorption, rheumatoid arthritis and diabetes (Aggarwal, 2003).

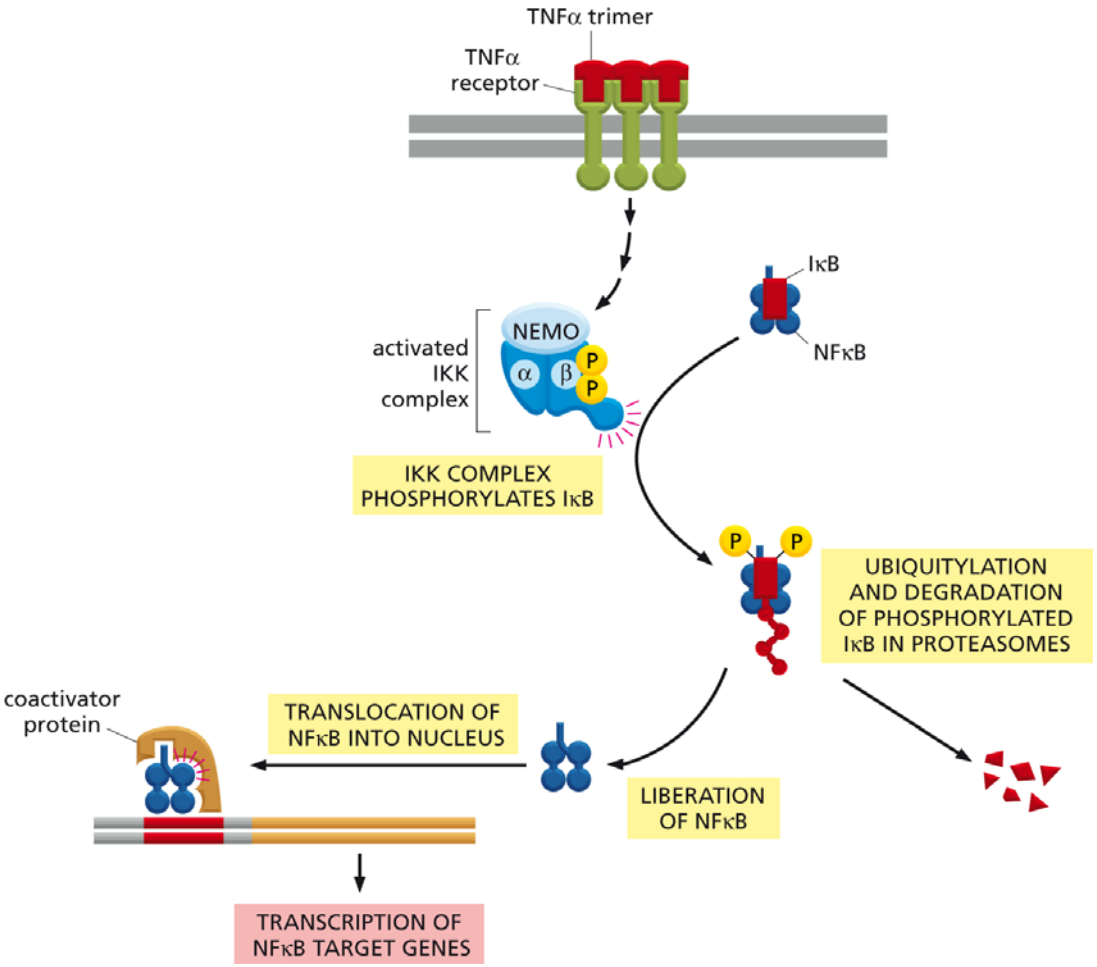


Figure 7: Example for NfκB activation through the binding of TNFα, which causes a rearrangement of the clustered cytosolic tails of the receptors, followed by the recruitment of various signaling proteins, resulting in the activation of a serine/threonine kinase that phosphorylates and activates IκB kinase (IKK). IKK consists of catalytic kinase subunits (IKKa and/or IKKb) and a scaffold, sensing protein called NF-κB essential modulator (NEMO). As such, activation of NF-kB dimers is the result of IKK-mediated, phosphorylation-induced degradation of the IκB inhibitor, which enables the NF-κB dimers to enter the nucleus and activate specifc target gene expression (Hoffmann et al., 2006, Alberts, 2008, 953).

2 Objectives

The scope of this master thesis was in the first part evaluation of ChIP-cloning to identify CHO endogenous promoter elements. This was based on the work of Martina Baumann (Baumann, 2008). ChIP-cloning derived sequences from HEK293 cells are mapped to the human reference genome and compared with data from the ENCyclopedia Of DNA Elements (ENCODE) database (Birney et al., 2007) to evaluate the method and subsequently reduce the high rate of non functional elements derived with this method in CHO cells.

In the second part a different approach implying artificial promoter sequences to establish an inducible promoter based on previous work of Jens Pontiller on artificial promoters (Grabherr et al., 2011). Both approaches have the ultimate goal to find sequences that could drive heterologous gene expression in mammalian cells. While the endogenous approach is exploiting the host machinery resulting in enhanced transgene expression, the scope of artificial promoters is to establish a promoter platform, which is able to carry various known regulatory motifs (TATA-box, BRE, NfκB) to custom design promoters that are suitable for biotechnological and medical applications. One of these applications could be a reporter system containing specific response elements like the NfkB element in an artificial promoter that is active in a wide variety of cell types. The aim was to proof that an artificial promoter is able to work with an inducible element (NfκB) resulting in a specific response upon a stimuli activating the NfκB-pathway.

3 Material and Methods

3.1 Bacteria and mammalian cells

3.1.1 Bacterial strains

- *Escherichia coli (E. coli)* strains DH10B, JM109 (lab stocks) were used for cloning purpose.
- One Shot®TOP10 Chemically Competent *E. coli* (Invitrogen, USA) were used for ChIP-cloning

3.1.2 Mammalian cell lines

- CHO dhfr$^-$ cell line (American Type Culture Collection, ATCC, USA)
- HEK293 cell line (American Type Culture Collection, ATCC, USA)
- U937 cell line (American Type Culture Collection, ATCC, USA)

3.2 Plasmids

3.2.1 pGL3 luciferase reporter vectors

The pGL3 luciferase reporter vectors (Promega, USA) were used for the quantitative analysis of fragments that potentially regulate CHO gene expression. The backbone of these vectors contains a modified coding region for firefly *(Photinus pyralis*) luciferase that has been optimized for monitoring transcriptional activity in transfected eukaryotic cells. The vectors contain the ampicillin resistance gene for selection in *E. coli.*

3.2.2 pGL3-Basic vector

The pGL3-Basic vector lacks eukaryotic promoter and enhancer elements. Any expression of luciferase activity in transfected cells depends on the inserted DNA fragment into the MCS (multiple cloning site) upstream from the firefly luciferase gene. Beside for testing transcriptional activity for obtained promoter candidates and artificial promoter sequences cloned into the vector, the initial pGL3- Basic vector was used as negative control.

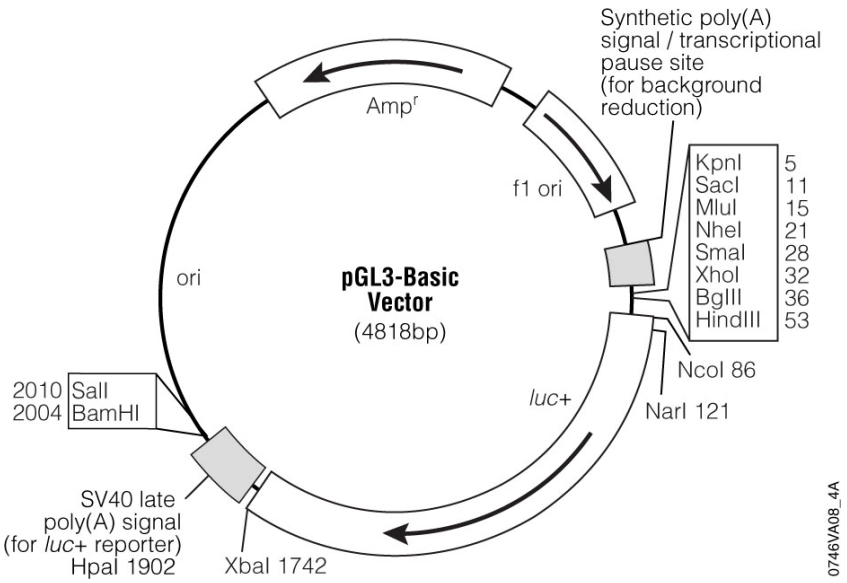


Figure 8:pGL3-Basic vector circle map. *luc*+: cDNA encoding the modified firefly luciferase; Amp^r: ampicillin resistance gene; f1 ori: origin of replication derived from filamentous phage; ori: origin of replication in *E.coli*, Expression of firefly luciferase depends on insertion of a functional promoter upstream from *luc*+(PROMEGA, 2007a).

3.2.3 pGL3-Promoter vector

The pGL3-Promoter vector contains the SV40 promoter upstream of the luciferase gene. The vector was used as a positive control.

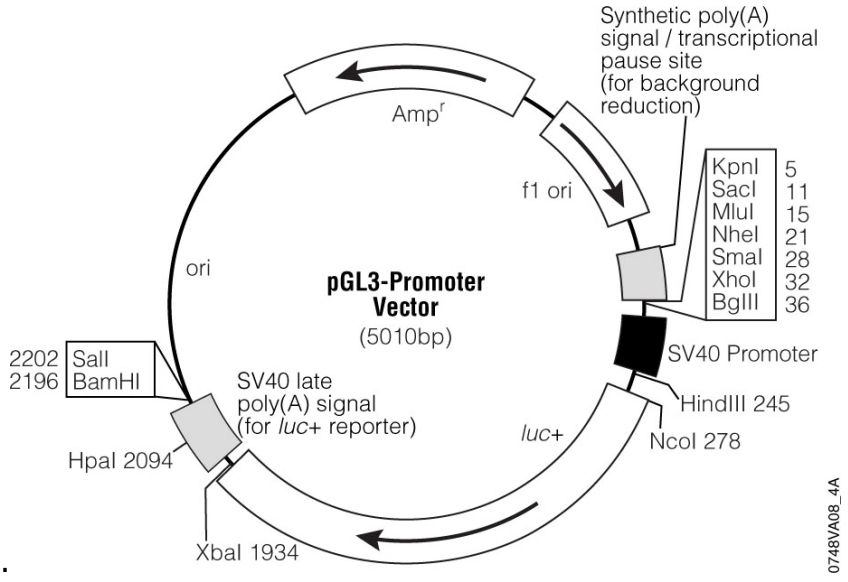


Figure 9: pGL3-Promoter vector circle map *luc*+: cDNA encoding the modified firefly luciferase; Amp^r: ampicillin resistance gene; fl ori: origin of replication derived from filamentous phage; ori: origin of replication in *E. coli*, The expression of the firefly luciferase is driven by a SV40 promoter (PROMEGA, 2007a)

3.2.4 pRL-SV40 vector

The pRL-SV40 vector (Promega, USA) is an internal control reporter and was used in combination with the pGL3 luciferase reporter vectors to co-transfect CHO cells. The pRL-SV40 vector contains a cDNA encoding the *Renilla* luciferase, which was originally cloned from the marine organism *Renilla reniformis.* Furthermore, the vector contains the SV40 enhancer and early promoter elements driving the *Renilla* luciferase gene as well as the ampicillin resistance gene for selection in *E. coli.*

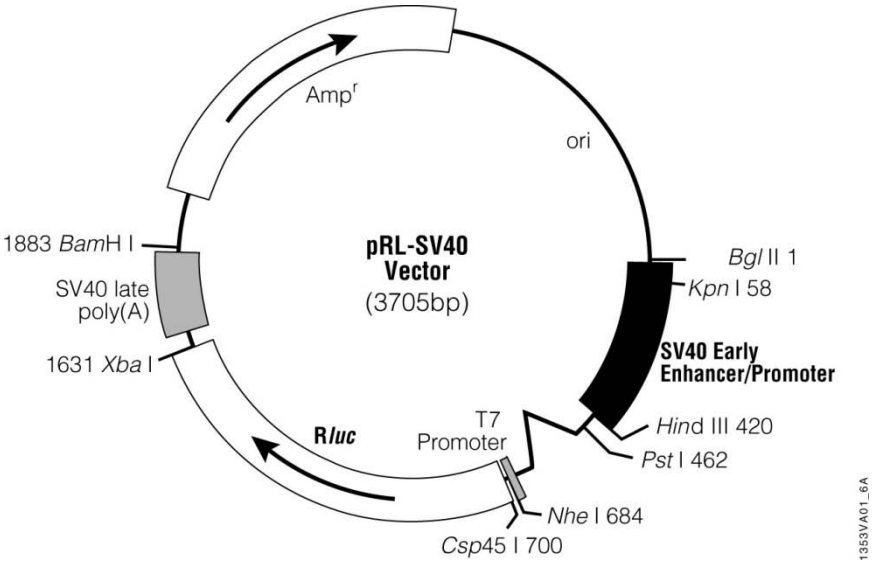
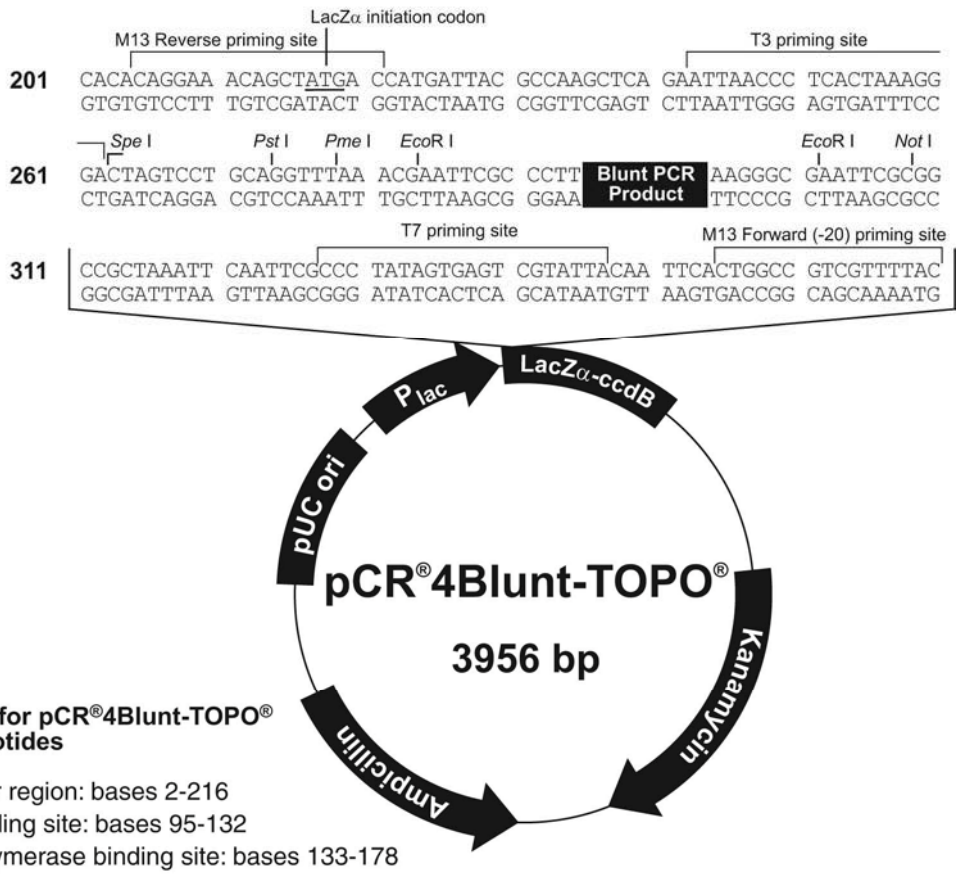


Figure 10: pRL-SV40 vector circle map. R*luc*: cDNA encoding the Renilla luciferase; Amp^r: ampicillin resistance gene; ori: origin of replication in *E. coli*; Expression of Renilla is driven by a SV40 early enhancer/promoter (PROMEGA, 2007b).

3.2.5 pCR®4Blunt-TOPO

The sequences derived from ChIP experiments were cloned into the Zero Blunt®TOPO® PCR Cloning vector according to the protocol from Invitrogen® (Invitrogen, 2004). See Blunt end ligation with pCR®4Blunt-TOPO chapter 3.3.14 for details. Besides the standard components (pUC ori, Amp^r) for selection and propagation in *E.coli*, the pCR®4Blunt-TOPO® allows direct selection of recombinants via disruption of the lethal E.coli gene, ccdB. The vector contains the ccdB gene fused to the C-terminus of the LacZα fragment. Ligation of a blunt-end PCR product disrupts expression of the lacZα-ccdB gene fusion permitting growth of only positive recombinants upon transformation into E.coli. Cells that contain non-recombinant vector are killed upon plating.



Comments for pCR®4Blunt-TOPO®
3956 nucleotides

lac promoter region: bases 2-216
CAP binding site: bases 95-132
RNA polymerase binding site: bases 133-178
Lac repressor binding site: bases 179-199
Start of transcription: base 179
M13 Reverse priming site: bases 205-221
LacZα-*ccdB* gene fusion: bases 217-810
LacZα portion of fusion: bases 217-497
*ccd*B portion of fusion: bases 508-810
T3 priming site: bases 243-262
TOPO® Cloning site: bases 294-295
T7 priming site: bases 328-347
M13 Forward (-20) priming site: bases 355-370
Kanamycin promoter: bases 1021-1070
Kanamycin resistance gene: bases 1159-1953
Ampicillin (*bla*) resistance gene: bases 2203-3063 (c)
Ampicillin (*bla*) promoter: bases 3064-3160 (c)
pUC origin: bases 3161-3834

(c) = complementary strand

Figure 11: pCR®4Blunt-TOPO vector (Invitrogen, 2004).

3.3 Molecular biology methods

All cloning procedures were performed as described by Sambrook and Russell (Sambrook and Russell, 2006) or as recommended by the manufacturer of used kits. Sterile double distilled water (ddH_2O) was used for all reactions and preparations of buffers and solutions.

3.3.1 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) is one of the most important and applied methods in molecular biology. This technique facilitates the enzymatic *in vitro* amplification of specific DNA regions. The principle of PCR is based on the cyclic repetition of three steps.

Denaturation: The double-stranded DNA template is heated to 92 - 98°C depending on the DNA polymerase used. This causes separation of DNA template by disrupting the hydrogen bonds between complementary bases**.**

Primer Annealing: The temperature is reduced allowing the annealing of the two specific primers to the single-stranded DNA template (50 - 69°C). The annealing temperature primarily depends on the melting temperature (T_m) of the primer and should be similar for the pair of primers used.

Elongation (Extension): The temperature is increased to the optimum activity temperature of the heat-stable DNA polymerase used (commonly 72°C). The polymerase synthesizes a new DNA strand that is complementary to the DNA template strand by adding deoxynucleoside triphosphates (dNTPs) in 5' to 3' direction.

All PCR experiments were conducted using a Biometra™ T3 Thermocycler, a Biometra" Professional Thermocycler, or a BIO-RAD CI000™ Thermal Cycler and 0.2 ml reaction tubes.

3.3.2 Common PCR

Conventional PCR was carried out in a 50 pi reaction volume. For cloning purpose the Phusion™ high-fidelity DNA polymerase (Finnzymes Oy, Espoo, Finland), creating blunt- end fragments, was used. This enzyme has a 3' - 5' exonuclease activity for proof-reading, resulting into strongly reduced error rates during the

PCR process. Alternatively, KOD DNA Polymerase (Merck KGaA, Darmstadt, Germany) or Biotools DNA Polymerase (Biotools B&M Labs, S.A.) were used.
The components for one PCR assay are listed inTable 1, exemplified for the Phusion DNA Polymerase.

components	**volume / 50 µl**
ddH20	36.5 µl
5 x Buffer HF	10 µl
dNTPs (10 mM)	1 µl
Phusion DNA Polymerase (2 U/ µl)	0.5 µl
primer for (10 pmol/ µl)	0.5 µl
primer back (10 pmol/ µl)	0.5 µl
template (1 pg/ µl - 10 ng/ µl)	1 µl

Table 1: Pipetting instructions using Phusion DNA Polymerase

cycle step	**temperature**	**time**	**number of cycles**
Initial denaturation	98°C	30 s	1
Denaturation	98°C	7 s	30-35
Annealing	T_m + 3°C	20 s	
Extension	72°C	15-30 s/kb	
Final Extension	72°C	5 min	1

Table 2: Cycling instructions using Phusion DNA Polymerase

3.3.3 Colony PCR - Screening for positive transformants

Colony PCR was used to directly amplify specific DNA regions out of bacterial cells. It allowed a control of the presence of a defined plasmid segment in bacterial clones. For PCR screenings the Biotools DNA Polymerase (Biotools B&M Labs, Spain) without a 3' - 5' exonuclease activity was used.

components	volume/ 100 µl
10 x Mg free buffer	10 µl
$MgCl_2$ (50 mM)	4 µl
dNTPs (10 mM)	1 µl
primer for (10 pg/ µl)	1 µl
primer back (10 pg/ µl)	1 µl
ddH_20	82.5 µl
BIOTOOLS DNA Polymerase (5 U/ µl)	0.5 µl

Table 3: Pipetting instructions using Biotools DNA Polymerase

cycle step	temperature	time	number of cycles
Initial denaturation	95°C	I min	1
Denaturation	94°C	30 s	30
Annealing	55°C - 72°C	30 s	
Extension	72°C	1 min/kb	
Final Extension	72°C	2 min	1

Table 4: Cycling instructions using Biotools DNA Polymerase

For PCR-Screening a single colony from the transformation plate was picked up with a sterile pipette tip and whisked in 30 µl mastermix. With the same tip an LB-agar masterplate, containing the appropriate antibiotic, was inoculated. The masterplate was necessary for subsequent identification and amplification of positive clones.

3.3.4 Primer list

Primers used for amplifications of DNA fragments for cloning purpose	Name
5'-ACTACTGCTAGCAGACCACAACTGGAAATGTCCTTTG-3'	Candidate K5 for1_NheI
5'-CTACTACCCGGGCACCATTTGTCTCATTCAGAGAGAAGAATC3'	Candidate K5 back1_XmaI
5'-ACTACTCCCGGGTCAGAGAGAAGAATCCTGGGGTAAG-3'	Candidate K5 back2_XmaI
5'-ACTACTCCCGGGGCAATTGTTGAAGAAAGCACAGCTAC-3'	Candidate F6 reverse back1_XmaI
5'-ACTACTCCCGGGGAAGAACGCATTGTGATTTTGTACATAGC-3'	Candidate F6 reverse back2_XmaI
5'-AGTAGTGCTAGCGCCATCACAAAACAGAGAGCC-3'	Candidate F6 reverse for1_NheI
5'-AGTAGTGCTAGCGCAGAGCCTTCTAGTGCTG-3'	Candidate F6 reverse for2_NheI

Screening primers for pGL3	
5´-CTAGCAAAATAGGCTGTCCC-3´	RVprimer3
5´-GACGATAGTCATGCCCCGCG-3´	RVprimer4

Screening primers for pCR®4Blunt-TOPO	
5'-CAGGAAACAGCTATGAC-3'	M13 Rev
5'-GTAAAACGACGGCCAG-3'	M13 For

Primers used for modifiying artificial promoter elements	
5'-TATCCGAATCGTCCGATTAC-3'	ArS232-dT12_del_5NFkB_1Inh_as
5'-TATCCAAAGTATTACTTTACTTATG-3'	ArS232-dT12_del_5NFkB_1Inh_ss
5'-TTATACGCGCGTCGAGGTAC-3'	ArS232-dT2_5NFkB_as
5'-CACGCGAGCGGTTCGAAC-3'	ArS232-dT2_5NFkB_ss

Oligonucleotides used for linker mediated PCR ChIP-Protocol	
5'- GCGGTGACCCGGGAGATCTGAATTC- 3'	JW102
5'- GAATTCAGATC- 3'	JW103

Table 5: All primers, oligonucleotides used during practical work are listed, classified by the application.

3.3.5 Agarose gel electrophoresis

Agarose gel electrophoresis is used for analytical as well as for preparative purposes. The principle of this method is based on the anionic nature of DNA in solution and on the migration of negatively charged molecules towards the positive pole in an electric field. The polymerization degree of the agarose gel as well as the molecule size determine the velocity of migration. For visualization under ultraviolet light, gels have to contain a fluorescent DNA stain.The most common method to make DNA visible is using the dye ethidium bromide which fluoresces under UV light when intercalated into DNA. Size and concentration of

the DNA fragments can be estimated by comparison to an appropriate DNA marker.

Agarose gel electrophoresis was used for analytical purposes such as evaluation of PCRs or restriction digests as well as for preparation of a specific DNA fragment or plasmid. For all gel electrophoreses performed, 1% agarose gels were used. For this purpose, agarose was melted in the appropriate volume of 1 x TAE buffer using a microwave oven.

Quantity	Component
3.6 g	Agarose
7.2 g	50x TAE Buffer
349.2 g	ddH_20

Table 6: Composition of 1% agarose gel

After cooling down to about 60°C, ethidium bromide (18 pi for 360 g gel) was added. Then the liquid gel was poured into gel preparation trays assembled with the appropriate combs and allowed to solidify at room temperature. The gel was put into an electrophoresis chamber and covered with 1 x TAE running buffer.

DNA samples were mixed with a suitable amount of loading buffer and applied to the gel. As loading dye a 6x BX buffer was used containing the two tracking dyes bromophenol blue and xylene cyanol FF. In 1% agarose gels bromophenol blue co-migrates with -300 bp DNA, while xylene cyanol FF co-migrates with -4000 bp DNA. One volume of loading buffer was added to five volumes of DNA sample. Slots of an analytical gel were loaded with up to 30 µl and slots for preparative gels with up to 60 µl. An analytical gel electrophoresis was performed at 130 V whereas 90 V was used for a preparative gel. For DNA detection a molecular imager (Gel Doc™ XR System) containing an UV-transilluminator and a digital camera was used. For preparation purpose, bands of the desired size were cut out using a bench UV-transilluminator (TPB-M/WL).

3.3.6 Solutions for Agarose Gel Electrophoresis

50x TAE Buffer

Quantity	Component
242 g/1	TRIS base
57.1 ml/l	Glacial acetic acid
100 ml/1	0.5 M EDTA (pH 8)

Table 7: Composition of 50 x TAE buffer

The compounds were mixed together, filled up with ddH_20 water and sterilised by autoclaving at 121°C for 20 min.

Running Buffer

Quantity	Component
20 ml/1	50 x TAE
30 µl/1	Ethidium bromide

Table 8: Composition of running buffer

6 x Gel-loading Buffer (BX)

Quantity (w/v)	Component
0.25%	Bromphenol blue
0.25%	Xylen cyanol FF
30%	Glycerol in water

Table 9: Composition of 6 x loading buffer

3.3.7 DNA extraction and purification from agarose gel

DNA was purified from excised gel slices with NucleoSpin[8] Extrakt II, the illustra™ GFX™ PCR DNA and Gel Band Purification Kit as recommend by the manufacturer. With all kits the gel slice is completely dissolved and loaded to a column containing a silica membrane which binds DNA in presents of chaotropic salts added by the binding buffer. Residual contaminations like salts and soluble macromolecular components are removed by washing with an ethanolic buffer. Pure DNA is finally eluted with nuclease-free ddH_20.

3.3.8 DNA quantification

The quantity of DNA in a solution was determined using a spectrophotometer (NanoPhotometer™ or NanoDrop 1000) for measuring the absorbance at 260 nm and calculating the corresponded concentration. Additionally, these spectrophotometers determine the ratio of the absorbance at 260 nm to the absorbance at 280 nm which gives an indication of protein contamination in the solution

3.3.9 Plasmid preparation - Mini / Midi

Plasmid mini preparation was performed with the E.Z.N.A.® Plasmid Miniprep II DNA Kit (Peqlab, Erlangen, Germany) or with the Wizard Plus SV Miniprep DNA

Purification System (Promega, Madison, WI, USA) according to purchaser's manuals.

Midi preparation of plasmid-DNA was achieved using the NucleoBond® Xtra Midi EF (Macherey-Nagel, Germany). EF stands for endotoxin free. Endotoxins are released from cells in small amounts during cell growth and in very large quantities upon cell death and lysis and thus also during plasmid purification. Like intact cells the free LPS molecules induce inflammatory reactions of the mammalian immune system. Therefore they have to be removed quantitatively from plasmid preparations to guarantee high transfection rates and high viability of transfected cells. Due to their amphiphilic nature and their negative charge endotoxins behave like DNA and are co-purified with most common plasmid purification systems. NucleoBond® Xtra Midi EF kit yields low levels (< 0.05 EU/ µg plasmid DNA) of endotoxin in plasmid preps.

The kits used for isolation of plasmid-DNA out of *E. coli* cells are based on a modified alkaline lysis. The bacteria are cultured over night in LB-medium containing the appropriate antibiotic, pelleted and lysed by addition of solutions containing the anionic detergent SDS. After neutralisation, the preparation is clarified by pelleting of the bacterial debris and the lysate is loaded on a column. There plasmid-DNA binds to a silica membrane and can be purified from contaminations by washing with an ethanolic buffer.

After drying of the column plasmid-DNA is eluted with nuclease-free water and stored at -20°C.

3.3.10 Restriction digest

Restriction enzymes are prokaryotic endonucleases which recognize and cleave specific double stranded DNA sequences generating sticky ends (3' or 5' overhangs) or blunt ends. Restriction digests were used for the purpose of preparing inserts and plasmids for cloning the ChIP derived sequences into reporter vector pGL3-basic. All restriction endonucleases and appropriate buffers were purchased from New England Biolabs(NEB). For complete digestions, a ratio of 2 - 10 U of restriction endonuclease to 1 pg DNA, the appropriate NEBuffer, bovine serum albumin (BSA) if required, and ddH_2O were used and incubated over night at the optimal temperature as recommended by the manufacturer.

Inserts and plasmids for cloning were isolated by preparative agarose gel electrophoresis following DNA extraction and purification.

3.3.11 Dephosphorylation

Phosphatases catalyze the hydrolysis of 5 ' phosphate groups of DNA and are often used for the dephosphorylation of restriction digested cloning vectors in order to prevent self-ligation. Dephosphorylation was performed after the digest of a plasmid vector and thermal inactivation of the restriction endonuclease using the Antarctic Phosphatase. For this purpose 1 U of the enzyme per 1 pg DNA and the appropriate amount of Antarctic Phosphatase Reaction Buffer were directly added to the digest mixture and incubated for 1 h at 37°C following inactivation at 65°C for 5 min and immediate purification of the plasmid DNA by preparative gel electrophoresis and DNA extraction and purification.

3.3.12 Phosphorylation

For the preparation of inserts generated by PCR for cloning, the addition of 5' phosphates to the ends of the PCR product is required.
The phosphorylation reaction was performed using 10 U of T4 Polynucleotide Kinase, the appropriate amount of T4 DNA Ligase Buffer, ddFkO, and the purified PCR product resulting in a 20 µl reaction volume. After incubation at 37°C for 45 min, T4 Polynucleotide Kinase was temperature inactivated at 70°C for 1 h. Phosporylated DNA was directly used for subsequent ligations.

3.3.13 Ligation

For DNA ligation, the T4 DNA Ligase was used which catalyzes the formation of a phosphodiester bond between juxtaposed 5' phosphate and 3' hydroxyl termini in double stranded DNA. For sticky end ligation 100 – 200 ng linearised vector and a 4 -5 fold molar surplus of insert was used. The assays for sticky-end ligation containing digested vector and insert with compatible cohesive ends, the appropriate amount of T4 Ligase Buffer, 1 µl T4 DNA Ligase (400 U), and ddH_20 were incubated at room temperature for 1 h or at 16°C over night. Isopropanol or ethanol precipitation was used to purify the ligated DNA in order to remove interfering salts for subsequent transformation into *E. coli* via electroporation.

3.3.14 Blunt end ligation with pCR®4Blunt-TOPO

The sequences derived from ChIP experiments were cloned into the Zero Blunt®TOPO® PCR Cloning vector according to the protocol from Invitrogen® (Invitrogen, 2004). The TOPO®Cloning reaction uses a plasmid vector (pCR®4Blunt-TOPO®), which is supplied linearized with Vaccinia virus topoisomerase I covalently bound to the 3´ ends (referred to as "activated" vector). Topoisomerase I from Vaccinia virus binds to duplex DNA at specific sites and cleaves the phosphodiester backbone after 5'-CCCTT in one strand. The energy from the broken phosphodiester backbone is conserved by formation of a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue (Tyr-274) of topoisomerase I. The phospho-tyrosyl bond between the DNA and enzyme can subsequently be attacked by the 5' hydroxyl of the original cleaved strand, reversing the reaction and releasing topoisomerase. TOPO®Cloning exploits this reaction to efficiently clone PCR products.

For the TOPO® reaction fresh PCR product was mixed with the reagents listed in Table 10 and incubated for 30 min at room temperature (22-23°C). The reaction was then placed on ice, and transformation in either chemically competent *E. coli* or electrocompetent *E. coli* was performed.

Reagent	**Chemically Competent *E. coli***	**Electrocompetent *E. coli***
PCR product	4 µl	4 µl
Salt solution	1 µl	
Diluted salt solution (1:4)		1 µl
ddH_2O	add to 5µl	add to 5µl
pCR®4Blunt-TOPO®	1 µl	1 µl
Final volume	6 µl	6 µl

Table 10: TOPO® reaction mix for either chemically competent *E. coli* or electrocompetent *E. coli*

3.3.15 DNA precipitation

Alcohol precipitation was used to concentrate DNA solutions and to remove unwanted salts (e.g. after a DNA ligation, TOPO® cloning reaction, during ChIP experiments). If salts are present when performing electroporation, the conductivity is altered and DNA uptake could be hindered.

3.3.16 Isopropanol precipitation

Isopropanol precipitation was carried out by adding an equal volume of isopropanol and a 0.1-fold volume of 3 M sodium acetate (pH 5.2) to the DNA solution. After incubation at room temperature for 10 min and subsequent centrifugation at 15°C and 16,000 g for 30 min, the obtained DNA pellet was washed with 70% ethanol. The pellet was completely air dried to remove residual ethanol and dissolved in at least 10 µl ddH_20.

3.3.17 Ethanol precipitation

For ethanol precipitation a 2.5-fold volume of 96% ethanol and a 0.1-fold volume of 3 M sodium acetate (pH 5.2) was added to the DNA solution. After incubation for 30 min at -20°C, centrifugation was performed at 4°C and 16,000 g for 30 min. The pellet was washed with 70% ethanol, air dried, and resuspended in at least 10 pi ddH_20.

3.3.18 Transformation of chemically competent One Shot® TOP10 cells

- 2 µl of the TOPO® Cloning reaction was added to a vial of One Shot® Chemically Competent *E. coli* and mixed gently. Cells were incubated for 30 min on ice.
- The cells were heat-shocked for 30 seconds at 42°C without shaking and immediately transferred to ice.
- 250 µl of room temperature S.O.C. medium was added and vials were incubated at 200 rpm at 37°C for 1 hour.
- 10-50 µl from each transformation were put on a pre-warmed LB plate with 25 mg/ml Ampicillin and incubated overnight at 37°C

3.3.19 Transformation of electrocompetent cells

The most important physical method for introducing plasmid-DNA into bacteria as well as mammalian cells is electroporation. The permeability of the cell membrane is increased dramatically caused by an externally applied electrical field. Exposure of membranes to high-voltage leads to their breakdown and the formation of pores that allow macromolecules to enter or leave the cell. The time

constant and the field strength applied to the sample are critical for successful uptake of DNA. The time constant depends on the total resistance of the sample and the capacitance of the pulse circuit for the electroporation device. As total resistance is the reciprocal of conductivity, the concentration of salt in the cell suspension influences the value of the time constant; this is why excess salt is removed during the plasmid DNA purification. The field strength is determined by the initial voltage applied to the cell suspension and the distance between the electroporation cuvette electrodes.

3.3.20 Preparation of electrocompetent *E. coli* cells

For preparation of electrocompetent cells *Escherichia coli* DH10B, DH5a or NEB10ß were incubated o/n in 20 ml LB-medium without antibiotics at 37°C on a temperated shaker. Two 2000 ml flasks were filled with 400 ml LB-medium and inoculated 1:100 with the o/n culture. After incubation until an OD_{600} of 0.6 to 0.8 the cells were harvested by centrifugation at 4000 rpm for 8 min at 4°C. The supernatant was decanted and the pellet was resuspended in 500 ml ice-cold 1 mM HEPES. The washing step was repeated three more times by resuspending the pellet successively in 250 ml, 100 ml 1 mM HEPES and 60 ml 10 % glycerine. Finally the cells were resuspended in 6 ml 10 % glycerine, aliquoted (50 µl per 1.5 ml Eppendorf tube), shock frozen in liquid nitrogen and stored at -80°C. All centrifugation steps were performed with a Beckmann centrifuge (Coulter Avanti J-20 XP) with a JLA 10.500 rotor (TM J-LITE Series ROTOR).
To prove the suitability of the competent cells the transformation efficiency was tested by electroporation with 20 pg pUC19 plasmid and plating appropriate dilutions on LB-agar containing Ampicillin.

3.3.21 Transformation of electrocompetent E. *coli* cells

Prior electroporation 10 x 2 x 45 mm Sarstedt cuvettes (Nümbrecht, Germany) had to be pasted up with stripes of agglutinated conductive aluminium foil. 50 µl aliquots of electrocompetent *E. coli* were thawed on ice, mixed with plasmid-DNA to be amplified and transferred to pre-chilled cuvettes. The Bio-Rad (Hercules, CA, USA) Gene Pulser apparatus was set to program Ec2 for bacteria (2.5 kV, 1000 Ω, 25 µF) and the cells were electroporated with one pulse (time constant < 6.0 milliseconds). Immediately after transformation the bacteria were revitalised in 900 µl prepared SOC-medium at 37°C for at least half an hour

under gentle shaking. Then the cells were plated onto pre-dried LB-agar plates with a diameter of 90 mm, containing the appropriate selection marker and incubated o/n at 37°C.

3.4 Growth media

3.4.1 LB-Medium and LB-Agar (1.5 % w/v)

LB-medium was made up of the ingredients listed in Table 11.

Quantity	Component
10 g/l	Peptone from casein
5 g/l	Yeast extract
10 g/l	Sodium chloride, NaCl

Table 11: Composition of LB-medium

The components listed in Table 11 were dissolved in double distilled water (ddH_20), portioned into 500 ml flasks and autoclaved at 121°C for 20 min. The sterile media was stored at 16°C and required antibiotics were added before usage. Stock solutions of antibiotics were diluted thousand fold to obtain working concentrations. Ampicillin (100 mg/ml) and kanamycin (50 mg/ml) stock solutions were prepared by dissolving the antibiotics in ddH_20, followed by sterile filtration using 0.22 µm filters and stored in 500 µl aliquots at -20°C.

3.4.2 SOC-Medium

Quantity	Component	Concentration
10 g/l	Peptone from casein	
5 g/l	Yeast extract	
0.5844 g/l	Sodium chloride (NaCl)	10 mM
0.2237 g/l	Potassium chloride (KCl)	3 mM
2.0330 g/l	$MgCl_2$.6 H20	10 mM $MgCl_2$
2.4648 g/l	$MgS0_4$.7 H20	10 mM $MgS0_4$
3.6032 g/l	Glucose	20 mM

Table 12: Composition of SOC-medium

All compounds except glucose were dissolved in ddH_20 water and autoclaved at 121 °C for 20 min. Glucose was sterilised separately in order to prevent

unwanted reactions and reunited with the rest of the medium prior portioning to 10 ml aliquots. The SOC-medium was stored at 4°C.

3.4.3 Cell Culture Media and Buffers

3.4.3.1 CHO dhfr– Growth Medium

CHO dhfr$^-$ suspension cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) and Roswell Park Memorial Institute 1640 (RPMI 1640) Medium (Biochrom AG) mixed in a 1:1 ratio. The medium was supplemented with 4 mM L-glutamine (Sigma Aldrich, St. Louis, USA), 0.25 % v/v Soy Peptone (HyPep 1510, Sheffiled Pharma, Norwich, UK), 0.1 % Pluronic F68 and a protein free supplement (by courtesy of Polymun GmbH, Vienna, Austria) and added with 1 x HT supplement (Biochrom AG) leading to a final concentration of 100 µM Hypoxanthine and 16 µM Thymidine.

3.4.3.2 HEK293/U937 Growth Medium

HEK293 cells were grown in Roswell Park Memorial Institute 1640 (RPMI 1640) Medium (Biochrom AG) supplemented with 4 mM L-glutamine (PAA, Austria) and 10% fetal bovine serum (GIBCO).

3.4.3.3 Phosphate buffered saline (PBS) buffer

Concentration	Components
10 g/l	KH_2PO_4
12 g/l	$Na_2HP0_4 \bullet 2 H_20$
2 g/l	KCl
80 g/l	NaCl

Table 13: Composition of PBS 10x Concentration

PBS (pH 7.4) was used as washing solution for CHO dhfr$^-$ and HEK cells.

3.5 Analytical Methods

3.5.1 SDS-Polyacrylamid Gel Electrophoresis - SDS-PAGE

Electrically charged biomolecules can move in an electric field. The mobility of the ions depends on charge, size and shape of the molecules. These differences are the basis of electrophoresis (Judd, 2002).

Polyacrylamide gel possesses excellent features for the separation of proteins. The pore size can be controlled by choosing various concentrations of acrylamide and methylene- bis-acrylamide (a cross-linking reagent) at the time of polymerization. Furthermore using this gel for separation of proteins with similar charges but different shapes and sizes can be achieved with high reproducibility.
Sodium-dodecyl-sulfate (SDS) is an anionic detergent, which binds predominantly to proteins causing denaturation. This kind of polyacrylamide-gel-electrophoresis is one of the most commonly used methods to separate complex protein mixtures and also to determine the relative molecule masses of the proteins.
Under saturated conditions (about 1.4 g SDS is bound per gram protein) all proteins are loaded with a certain amount of negative charges per unit of mass. Therefore, in most cases the mobility of the protein-SDS-complex through the molecular sieve of the gel is proportional to the logarithm of the relative mass (M_r).
Two different systems of SDS-PAGE were used, i.e self- prepared electrophoresis gels on BioRad Mini-Protean II Dual Slab Cells and Invitrogen® NuPAGE® Novex® Tris-Acetat Mini Gels on XCell SureLock Mini-Cells, respectively.

3.5.2 Preparation of electrophoresis gels

The mixtures for two separation and collection gels with polymerization ratio of 1:37.5 are shown in Table 14.

	Separation gel	**Stacking gel**
Monomer concentration	10%	5%
Distilled water	4 ml	5.65 ml
1.5 M Tris-HCl	2.50 ml	
0.5 M Tris-HCl	-	2.5ml
Acrylamide/Bis	3.3 ml	1650 µl
10 % (^{w}/v) SDS	100 µl	100µl
10 % APS	100 µl	100 µl
TEMED	4 µl	4 µl

Table 14: Polyacrylamide gel for SDS-PAGE.

The separation gel was cast and covered with 2-butanol to get a plain surface. After polymerization of the gel for 30 min the butanol was discarded. The gel was rinsed with water and the surface dried using filter paper. The collection gel (height ~ 1 cm) was cast and the comb with 10 slots was set in place. After 15 minutes of polymerization the comb was removed and the gels were put into the electrophoresis chamber filled with 1x running buffer. The samples were diluted 1:2 in 2x sample buffer and heated for 95°C for 4 min. Samples (15 - 20 µl) and 5 µl PageRuler™ (Fermentas)Unstained Protein Ladder were loaded on the gel using a Hamilton syringe. The gel was run at constant 200 V and max. 70 mA.

Acrylamide/Bis	**1x Running buffer (pH 8.3)**
(30%T;2.67%C; 1:37.5)	160 ml 5x Running buffer (pH 8.3)
146.0 g acrylamide	640 ml distilled water
4.0 g Bis	
Made up to 500 ml with AD	**2x Sample buffer stock solution**
(filtrated and stored at 4°C)	2.0 ml Tris-HCl (pH 6.8)
	1.6 ml glycerol (20%)
Separation gel	3.2 ml 15% SDS (6%)
1.5 M Tris/HCl (pH 8.8)	0.4 ml 0.5% ($^{w}/_{v}$) bromine phenol blue
54.45 g Tris base	
120 ml distilled H_2O	**2x Sample buffer**
added to pH 8.8 with 6 N HCl	900 µl stock-2x sample buffer
Made up to 300 ml with AD	100 µl ß-mercaptoethanol
Collection gel	10% ($^{w}/_{v}$) Ammoniumpersulfate (APS)
0.5 M Tris/HCl (pH 6.8)	50 mg ammoniumpersulfate
6.0 g Tris base	500 µl distilled water
60 ml distilled H2O	
added to pH 6.8 with 6 N HCl	**10% ($^{w}/_{v}$) Sodium-dodecyl-sulfate (SDS)**
Made up to 100 ml with AD	1.0 g SDS
	10 ml distilled water
5x Running buffer (pH 8.3)	
15.15 g Tris base	**N,N,N',N'-tetramethylethylendiamine**
72.0 g glycine	(TEMED)
5.0 g SDS	

Made up to 1000 ml with AD (stored at 4°C)	**PageRuler™ Prestained Protein Ladder**

Table 15: Components of SDS-PAGE.

3.5.3 NuPAGE- system

For the analysis with TAF II p250 (L-20) sc-17134 antibody against TAFIIp250, NuPAGE®Novex® Tris-Acetate precast gels (3-8% gradient), respectively Bis-Tris gels (4-12% gradient) were used. 15 µl of the sample were mixed with 4xSDS sample buffer. The inner buffer chamber was filled with 200 ml and the outer buffer chamber was filled with 600 ml 1xNuPAGE SDS running buffer. 15 µl of the sample mix were loaded on the gel which was then run at constant 150V and 70 mA for 1h.

3.5.4 SimplyBlue™ SafeStain

SimplyBlue™ SafeStain is a ready-to-use colloidal Coomassie® G-250 staining solution used to visualize protein bands in polyacrylamide gels.

- The gel was placed in 100 ml of ultrapure water in a loosely covered container and microwave on 950 watts for 1 minute until the solution almost boiled.
- Two washing steps were performed with 20 ml of dH_2O for respectively for 1 min in a microwave.
- The gel was put on an orbital shaker for 1 minute.
- After the last wash, 20 ml of SimplyBlue™ SafeStain was added and boiled for 1 min
- The gel was put back on orbital shaker for 5 minutes until bands developed
- 20 ml of 20% NaCl were added for 10 minutes and gel was analyzed.

3.5.5 Western electroblotting

In order to make the proteins in the SDS-PAGE gel accessible to antibody detection, proteins are moved from within the gel onto a membrane made of nitrocellulose or polyvinylidene difluoride (PVDF), where they are bound and immobilized. Semi-dry blotting is performed with plate electrodes in a horizontal configuration, sandwiching a gel and membrane between sheets of buffer-soaked

filter paper that function as the ion reservoir. During electrophoretic transfer, negatively charged molecules migrate out of the gel and move towards the positive electrode, where they are deposited on the membrane. The plate electrodes, separated only by the gel and filter paper stack, provide high field strength (V/cm) across the gel, allowing for very efficient, rapid transfers.
Briefly, a Roti®-PVDF membrane (ROTH) and 6 pieces of blotting filter paper was cut out appropriately to fit the gel. The membrane was activated for 30 sec in 100% methanol and rinsed with dH_2O. The membrane and the filter papers were soaked in the TOWBIN transfer buffer for a few minutes in a shallow tray.
3 pieces of pre-soaked blotting filter paper were put on the anode plate of the Novex® Semi-Dry Blotter. Next the gel was put upon, followed by the membrane and 3 additional filter papers. Air bubbles were removed, the plates were screwed together and the transfer was conducted at 20V 170 mA for 50 min. To monitor transfer of proteins to the membrane, 5 ml 0,1% Ponceau S solution (Sigma-Aldrich) was applied to the membrane and analyzed after 5 min incubation at room temperature.

TOWBIN (Transfer buffer)	
25 mM Tris	3.03 g
129 mM Glycin	14.4 g
20% Methanol	200 ml
	add to 1l with H_2O

Table 16: TOWBIN buffer composition

3.5.6 Immunodetection

After blotting the membrane was put into blocking solution (supplemented with 3% milk powder) for at least 2 h or overnight.

(TPBS) Blocking solution	
10x PBS	100 ml
Tween 20	1 ml
ddH2O	900 ml

Table 17: Blocking solution

The membrane was washed with TPBS. 10 µg of the specific antibody was diluted with 10 ml TPBS supplemented with 1% (0.1 g) milk powder and applied on the

membrane in a small tray on a rotator at room temperature (20-23°C) for 1 h shaking.

The membrane was washed 3x with TPBS and incubated with secondary antibody (dilution 1:2000) on a rotator at room temperature (20-23°C) for 1 h shaking. Following 3x washing with TPBS for 5 min the membrane was incubated with the detection substrate, which either was the Super Signal West Pico Chemiluminescent Substrate for horseradish peroxidase (HRP) (Thermo Scientific), or with NBT (nitro-blue tetrazolium chloride) and BCIP (5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt) for alkaline phosphatase conjugated antibody following manufacturer's instructions. Chemiluminescent Signals were analyzed with a Fusion Fx7 detection device (2011 VILBER LOURMAT Deutschland GmbH).

Company	**Primary antibody**	**Target**	**Produced in**
Santa Cruz, CA, USA	(SI-1): sc-274	TFIIB (35kDa)	rabbit
	(L-20): sc-17134	TAFIIp250 (250kDa)	goat

Company	**Secondary antibody**	**Product number**
Sigma-Aldrich®	Anti-Rabbit IgG-Peroxidase	A6154
	Anti-Goat IgG-Alkaline Phosphatase	A4187

Table 18: Antibodies used for western blot analysis

3.6 Isolation of endogenous promoter sequences with ChIP-cloning

3.6.1 Chromatin immunoprecipitation

Chromatin Immunoprecipitation (ChIP) is used to investigate the interaction between proteins and DNA in the cell. It aims to determine whether specific proteins are associated with specific genomic regions, such as transcription factors on promoters or other DNA binding sites. ChIP was performed with HEK293 cells based on a mammalian ChIP-on-chip protocol (Agilent, 2010). For detailed discussion of experimental setup see results section in chapter 4.1.2.

3.6.2 Crosslinking of cells

Between 5x10^7 and 1x10^8 cells (HEK293) were used for each immunoprecipitation. Cells were crosslinked with 3 ml crosslinking solution added directly to culture media for 10 min shaking 100rpm at room temperature. 1.5 ml 2.5 M glycine was added to flasks to quench the formaldehyde. The cells were rinsed with 5 ml ice cold PBS into 50 ml collection tube followed by centrifugation at 1,400 g at 4°C for 10 min. Cell pellet was resuspended in 10 ml PBS and centrifuged at 1,400 g at 4°C for 10 min. Cells were immediately frozen with liquid nitrogen and stored at -80°C.

3.6.3 Cell Lysis and nuclear fractionation

Frozen cells were thawed on ice. All buffers were supplemented with protease inhibitors prior to use. The cell pellet was resuspended in 5 ml of lysis buffer 1, rocked for 10 min at 4 °C, followed by centrifugation for 5 min (4°C).
Supernatant was discarded and 5 ml lysis buffer 2 was added for 10 min at 4°C on a rotator, followed by centrifugation for 5 min (4°C). The supernatant was discarded.
The cells were resuspended in 3 ml lysis buffer 3 and sonicated using a SONOPULS HD 2070 sonicator. A program consistent of 25 sec pulsing followed by 60 sec pause with output power 4 was repeated 10x keeping the lysate on ice. 200 µl 10% Triton X-100 were added to the sonicated lysate, followed by centrifugation at 20,000 g for 10 min at 4°C. Supernatants were pooled and stored at -20°C.

3.6.4 Magnetic bead preparation

50 µl Dynal (Invitrogen) magnetic beads were washed with blocking solution 2x (using a magnetic device) and resuspended in 500 µl blocking solution. 250 µl of beads in blocking solution were used to pre-clear lysates from unspecific DNA binding to beads.

3.6.5 Immunoprecipitation of chromatin

The specific antibody (seeTable 18) was incubated with cell lysate on a rotator over night at 4°C. 250 µl pre-cleared beads were added and incubated for 1 h rotating at 4°C. Supernatant was discarded and beads were washed with 1 ml

RIPA buffer 1 (4x), 1 ml RIPA buffer 2 (4x), LiCl buffer (2x) and finally TE buffer (5x).

3.6.6 Elution and Reversing Cross-Linking

Beads were resuspended in 210 µl of elution buffer and heated to 65°C for 15 min while shaking. After centrifugation at 16.000 g for 1 min at room temperature 200 µl were put into a new Eppendorf tube and incubated at 65°C overnight.

3.6.7 Cellular protein and RNA digestion

200 µl of TE buffer was added and supplemented with RNaseA (0.2 mg/ml) and incubated for 2 h at 37°C. 300 mM $CaCl_2$ was added with proteinase K (0.2 mg/ml) and incubated for 30 min at 55°C. 400 µl of phenol:chloroform:isoamyl alcohol was added and samples were centrifuged with 14,000 g for 30 sec. The aqueous layer was transformed into a new 1.5 ml microfuge tube and ethanol preticipated and the DNA concentration was measured.

3.6.8 Preparation of Linkers

Linkers described in Table 5 were used to amplify ChIP derived sequences. Oligos were annealed using a temperature gradient. 375 µl of each linker (40 mM stock) was supplemented with 250 µl tris-HCl (1M) pH 7.9 and heated to 95°C for 5 min followed by 70°C for 1 min and cooling to 4°C with a rate of 0.4°C/min. Reaction was performed using BIO-RAD CI000™ Thermal Cycler and 0.2 ml reaction tubes.

3.6.9 DNA blunt ending and linker ligation

55 µl of sample was added to 55 µl of blunting mix according to Table 19 and

Blunting Mix		
Component	**final concentration**	**55 µl Mix**
10x Ne Buffer 2	1x	11.0 µl
10 µg/µl BSA (NEB)	5 µg	0.5 µl
10 mM dNTPs	100 µM	1.1 µl
3 U/µl T4 DNA polymerase (NEB)	1.5 U	0.5 µl
ddH2O		41.9 µl
total		55 µl

Table 19: Blunting Mix

incubated for 20 min at 12°C. Samples then were precipitated with Isopropanol. Blunt end ligation with linkers, ethanol precipitation, PCR and agarose gel electrophoresis of amplified products was performed as described above chapter 3.3. The PCR product was used directly for TOPO® cloning reaction described in chapter 3.3.14 and 3.3.18.

3.6.10 Nuclear fractionation for western blot analysis

To obtain extracts for immune detection a nuclear fractionation protocol was used (all solutions listed in Table 20) (Abcam, 1998-2011). Briefly 1 ml of buffer A was supplemented with protease inhibitor cocktail and 500 µl were added to a cell pellet of approximately 5x10^7 cells and incubated on ice for 10 min. After centrifugation at 3000 rpm for 10 min at 4°C, supernatant (containing everything except except large plasma membrane pieces, DNA, nucleoli) was removed and frozen at -20°C. Cells were resuspended in 374 µl of buffer B supplemented with 26 µl of 4.6 M NaCl to give a final concentration of 300 mM NaCl. After 30 min of incubation on ice, extracts were centrifuged at 24,000 g for 20 min at 4°C. Supernatant containing nuclear proteins was frozen at -70°C.

Cross-linking solution				
	End concentration	Stock solution	100	ml
Formaldehyde	11%			
NaCl	0.1 M		0.584333	g
EDTA	1 mM	50 mM	2	ml
EGTA	0.5 mM	50 mM	1	ml
HEPES pH 8.0	50 mM	1 M	5	ml

Lysis buffer 1				
	End concentration	Stock solution	100	ml
NaCl	0.14 M		0.818	g
EDTA	1 mM	50 mM	2	ml
Glycerine	10 %	100 %	10	ml
NP-40	0.5 %	70 %	0.713333	ml
Triton X-100	0.25 %	100%	0.25	ml
HEPES pH 7.5	0.05 M	1 M	5	ml

Lysis buffer 2				
	End	Stock solution	100	ml

	concentration			
NaCl	0.2 M		1.168667	g
EDTA	1 mM	50 mM	2	ml
EGTA	0.5 mM	50 mM	1	ml
Tris pH 8.0	10 mM	1 M	1	ml

Lysis buffer 3				
	End concentration	Stock solution	100	ml
EDTA	1 mM	50 mM	2	ml
EGTA	0.5 mM	50 mM	1	ml
Tris pH 8.0	10 mM	1 M	1	ml

RIPA buffer 1 pH 8				
	End concentration	Stock solution	100	ml
HEPES pH 7.5	50 mM	1 M	5	ml
EDTA	1 mM	50 mM	2	ml
Sodium deoxycholate	0.7 %	0.2	3.5	ml
NP-40	1 %	20%	5	ml
LiCl	0.5 M	M=42.39g/mol	2.11	g

RIPA buffer 2 pH 8				
	End concentration	Stock solution	100	ml
Tris-HCl, pH 8	50 mM	1 M	5	ml
NaCl	150 mM	58.44g/mol	0.8766	g
Sodium deoxycholate	0.7 %	0.2	3.5	ml
NP-40	1 %	20%	5	ml
SDS	0.001	0.1	1	ml

Elution buffer pH 8				
	End concentration	Stock solution	100	ml
Tris pH 8.0	50 mM	1M	2	ml
EDTA	10 mM	50 mM	20	ml
SDS	1 %	10%	10	ml

LiCl buffer pH 8				

	End concentration	Stock solution	100	ml
Tris pH 8.0	10mM	1M	1	ml
LiCl	250mM	42.39g/mol	1.06	g
NP-40	0.5	0.2	2.5	ml
Sodium deoxycholate	0.5	0.2	2.5	ml
EDTA	1mM	50mM	2	ml

Block Solution				
	End concentration	Stock solution	50	ml
PBS	1x	1x	50	ml
BSA	0.005		250	mg

Nuclear Fractionation Protocol

buffer A pH 7,9				
	End concentration	Stock solution	50	ml
HEPES pH 7.9	10mM	1 M	500	µl
KCl	10mM	2.5M	200	µl
EDTA pH=8	0.1mM	0.05M	100	µl
EGTA pH=8	0.1mM	0.2M	25	µl
dH2O			49.175	ml
Protease inhibitor cocktail	DTT-Stock: 0.1M	Für 1ml	10	µl
	PMSF-Stock:		5	µl

buffer B pH 7,9				
	End concentration	Stock solution	50	ml
HEPES pH 7.9	20mM	1 M	500	µl
KCl	400mM	2.5M	200	µl
EDTA pH=8	1mM	0.05M	100	µl
EGTA pH=8	1mM	0.2M	25	µl
Glycerin	0.1	0.4	3.75	ml
dH2O			45.425	ml
Protease inhibitor cocktail	DTT-Stock: 0.1M	Für 125µl	1.25	µl
	PMSF-Stock:		0.7	µl

Table 20: Solutions and chemicals used in ChIP and nuclear fractionation protocol

3.6.11 Sequencing of Plasmids

Plasmids were sequenced by Eurofins MWG Operon (Germany) or Agowa (Germany) using Sanger sequencing technique according to the manufacturer

guidelines. 15 pM of primers (M13 Rev, M13 For) were added to 1000 ng plasmid DNA in a 15 µl volume.

3.6.12 Mapping and in silico analyses

Sequences obtained by ChIP-cloning were aligned using the nucleotide blast of the Basic Local Alignment Search Tool (BLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequences identified as human were further analyzed with the UCSC Genome Browser (GRCh37/hg19 Assembly) (Kent et al., 2002). The browser provides a display of the sequence query, together with dozens of aligned annotation tracks (known genes, predicted genes, ESTs, mRNAs, CpG islands, assembly gaps and coverage, chromosomal bands, mouse homologies, transcription factor binding sites, histone modification patterns, DNaseI hypersensitivity regions). These tracks are obtained from the Encyclopedia of DNA Elements (ENCODE) Consortium. For each experiment type (ChIP-seq, DNase-seq, etc.), the ENCODE investigators conduct multiple experiments, using different cell lines, tissue samples and other variables for the experiment type to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active (Rosenbloom et al., 2010).

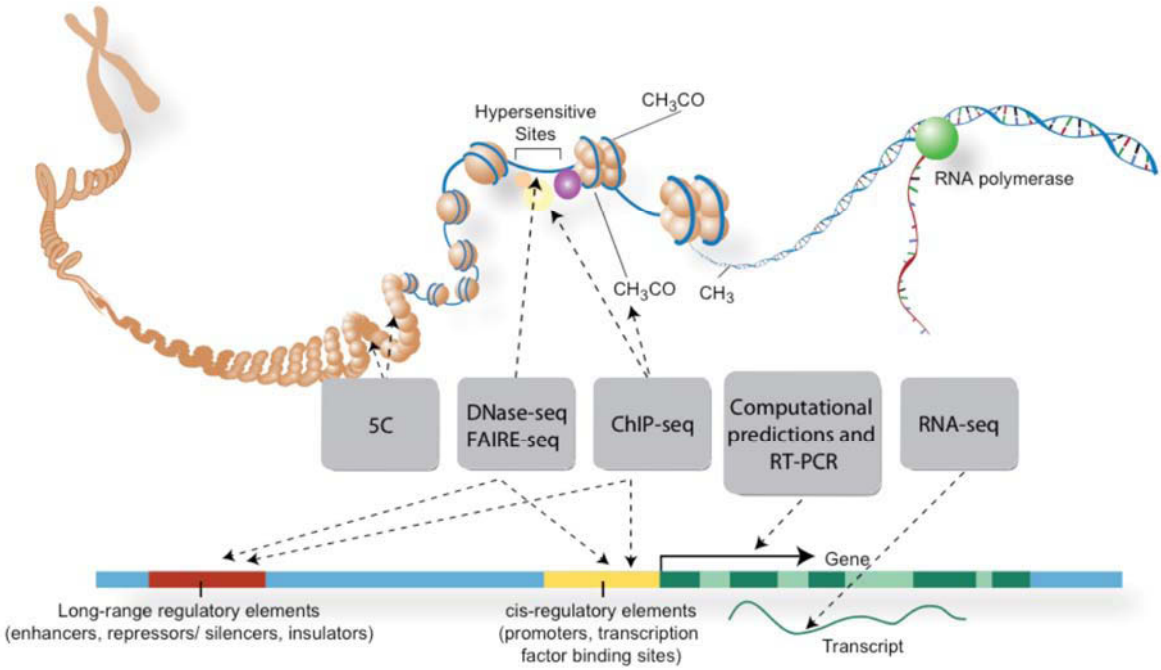


Figure 12: The ENCODE database comprises information from a variety of assays and methods to identify functional elements (5C, DNase-seq, FAIRE-seq, ChIP-seq, RNA-seq and computional predictions and RT-PCR) represented on an idealized model of mammalian chromatin and a mammalian gene (Myers et al., 2011).

Furthermore, potential transcription factor binding sites like TATA boxes, Spl binding sites, or NF-kB binding sites were identified using the web-based transcription factor binding sites (TFBSs) prediction tool ConSite (http://www.phylofoot.org/consite) (Sandelin et al., 2004) or the promoter prediction program NNPP 2.2 (neural network promoter prediction 2.2; http://www.fruitfly.org/seq_tools/promoter.html) (Reese, 2001).

3.7 Cell culture methods

3.7.1 Subculture of HEK293 cells

HEK293 adherent cells were cultivated in 75 cm^2 or 175 cm^2 cell culture flasks (BD Biosciences, NJ USA) and subcultured twice a week. Old media was discarded and cells were washed with 5-10 ml 1xPBS. Cells were detached from flask with 2 ml of Trypsin EDTA (1:250) (10x) (PAA, Austria), resuspended in 15 – 30 ml RPMI 1640 Media, passaged in a ratio of 1:5 to 1:7 into new flasks and incubated at 37°C and 7% $C0_2$.

3.7.2 Subculture of CHO dhfr$^-$ cells

CHO dhfr$^-$ cells were cultivated at 37°C and 7% $C0_2$ in 75 cm^2 culture flasks (BD Biosciences, NJ USA) and were passaged twice a week. Cell number was determined using a hemocytometer. The passage ratio was adapted to seed 1x10^5 cells/ml.

3.7.3 Subculture of U937 cells

U937 suspension cells were cultivated at 37°C and 7% $C0_2$ in 75 cm^2 culture flasks (BD Biosciences, NJ USA) and were passaged twice a week. Cell number was determined using a hemocytometer. The passage ratio was adapted to seed 1x10^5 cells/ml.

3.7.4 Hemocytometer (Neubauer counting chamber)

Determination of cell number and viability was performed with a Neubauer counting chamber. Therefore, 1 ml cell suspension was supplied with 200 µl of 0.5 % trypan blue (dilution factor: 1.2), which is capable of penetrating the cell membrane of dead cells, whereas viable cells remain unstained.

The Neubauer counting chamber consists of 9 big squares with an area of 1 mm^2 each and a depth of 0.1 mm resulting into a volume of 0.1 µl. Total cell number and viability were calculated according to following formulae:

total cell number / ml = number of cells / big square * 10000 * 1.2
viability (%) = (living cell number/ ml : total cell number/ ml) * 100

3.7.5 Transfection of mammalian cells

4x10^6 cells were transfected with 10 µg of the cloned reporter vectors using the electroporation system Amaxa® Cell Line Nucleofector® Kit V (Lonza, Switzerland). Transfection was carried out as described in the manufacturer's instruction using program H-14 (CHOdhfr$^-$), A-23 (HEK293) or U-13 (U937).

Cells were centrifuged at 200 g for 10 min. Then the supernatant was removed completely and the cell pellet was resuspended with the nucleofection mixture containing the 10 µg plasmid DNA and 100 µl of the transfection medium composed of 82 µl Nucleofector Solution V and 18 µl Supplement. Immediately after transfection, the cell suspension was transferred into 6-well plates containing 4 ml preheated culture medium and incubated at 37°C and 7% $C0_2$.

The pGL3 Luciferase reporter vectors containing the firefly luciferase (*Photinus pyralis*) gene were used as transfection plasmids. The pGL3-Basic vector containing no promoter served as negative control whereas the pGL3-Promoter vector containing the SV40 promoter was used as positive control. The promoter activity of this viral promoter was set to 100%. All other measurements refer to this value within the same cell-line.

To generate consistent results, normalization to account for transfection efficiency and cell number variability was required. For this purpose, 1 µg of pRL-SV40 containing the *Renilla* luciferase was co-transfected.

3.7.6 Measurement of Bioluminescence

Bioluminescence was measured 6h or 48 h post transfection using the Bright-Glo™ Luciferase Assay System or the Dual-Glo™ Luciferase Assay System.

For general testing of luciferase activity, the Bright-Glo™ Luciferase Assay System was used. 100 µl of the cell suspension was transferred into black 96 well plates and 100 µl of Bright-Glo™ Reagent was added. After incubation for 10 min, bioluminescence was measured using the infinite® MI000 microplate reader (i-control™ microplate reader software).

For transfections normalized by co-transfection with pRL-SV40, detection of bioluminescence was conducted using the Dual-Glo™ Luciferase Assay System. To measure firefly luciferase activity, 50 µl of the cell suspension was transferred into black 96 well plates and 50 µl of Dual- Glo® Luciferase Reagent was added. After incubation for 10 min, bioluminescence was measured using the (i-control™ microplate reader software). For subsequent measurement of *Renilla* luciferase activity, 50 µl of Dual-Glo® Stop & Glo® Reagent was added and bioluminescence was measured after incubating for another 10 min. The normalized promoter activity was determined from the ratio of firefly luciferase activity to *Renilla* luciferase activity.

The bioluminescence reaction catalyzed by firefly and *Renilla* luciferases is illustrated in Figure 13.

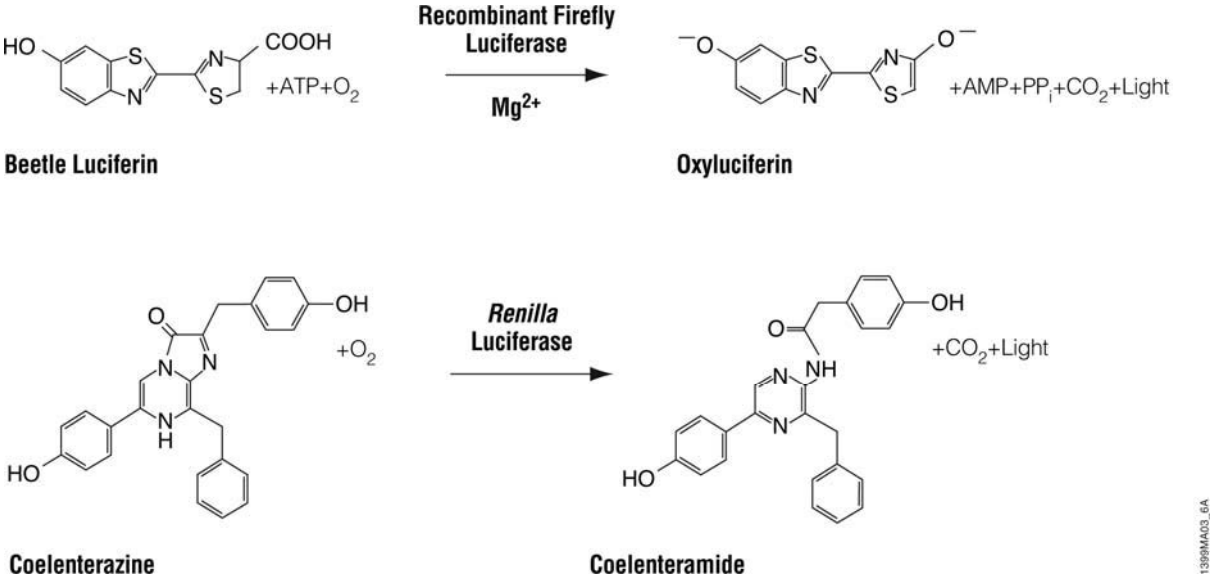


Figure 13: Bioluminescent reactions catalyzed by firefly and *Renilla* luciferases Mono-oxygenation of beetle luciferin is catalyzed by firefly luciferase in the presence of Mg^{2+}, ATP, and molecular oxygen. Unlike beetle luciferin, coelenterazine undergoes mono-oxygenation catalyzed by *Renilla* luciferase but requires only molecular oxygen (PROMEGA, 2011).

3.7.7 LPS and TNFα-Induction

Induction experiments were conducted with Lipopolysaccharide from *Escherichia coli* 055:B5 (Sigma) and recombinant human TNFα (Invitrogen). Cells were transfected with specific plasmid described above. Adherent cells were split immediately, suspension cells after 12 h of revitalization in same amount of media (4 ml). LPS and TNFα was added to the media and bioluminescence was measured 6 h post induction.

3.7.8 Construction of synthetic promoter elements

The construct ArS 232 (232 base pairs in length) was selected from the 160,000 nucleotide (nt) long promoter-like concatomer (Grabherr et al., 2011). All inserts have been assembled either by oligo synthesis (Sigma Aldrich, Austria) followed by annealing and PCR or by gene synthesis (Geneart, Germany) and cloning into the reporter vector pGL3 Basic (Promega, Madison, Wisconsin) upstream of a firefly luciferase gene, which was done by Jens Pontiller. Insertion of NfκB and TATA box elements was done by oligo synthesis (Sigma Aldrich, Austria) followed by annealing and PCR described in molecular methods section. For sequences see appendix.

4 Results and discussion

4.1 Part I: Evaluation of ChIP-cloning to isolate gene regulatory sequences

Mammalian production systems for therapeutic proteins are currently dependent on strong viral promoter sequences driving expression of the gene of interest. CHO cells are the predominant host among eukaryotic production systems for therapeutic proteins. Efficient transcription of a heterologous gene is currently dependent on strong viral promoter sequences, yielding high expression levels at the expense of cellular stress reactions affecting product quality and viability of the cells. The isolation and identification of CHO endogenous gene regulatory elements could yield a promoter which is part of the cells regulatory network avoiding the drawbacks of currently used viral promoters.

4.1.1 ChIP-cloning with CHO cells

Due to the lack of public available genomic sequences of the Chinese hamster *Cricetulus griseus*, a method independent of a reference genome was adapted to identify gene regulatory sequences. Chromatin Immunoprecipitation (ChIP) was used to isolate regulatory sequences bound by transcription factors *in vivo* (Agilent Technologies, 2011). Therefore cells are treated with a formaldehyde, crosslinking the cells proteins and DNA within 1 Å, delivering a snapshot of the cells current DNA-protein interactions. Shearing of the DNA with ultrasound produced short 200 - 1000 (bp) base pair fragments, which were precipitated using a specific antibody against transcription factor TFIIB, respectively the TFIID subunit TAFp250, which bind the core promoter elements BRE (TFIIB recognition element) as well as the TATA box binding protein (TBP) associated TAF1. The antibody-protein-DNA complexes are isolated with magnetic beads coated with protein G. Several washing steps to reduce unspecific binding to the bead surface were performed, followed by reverse crosslinking eluting the bound DNA. Isolated sequences were blunt end ligated and transformed into competent *E.coli*. Colony PCR was conducted to analyze the insert size using agarose gel electrophoresis. Clones positive for an insert were sequenced using Sanger

sequencing by Eurofins MWG Operon (Germany). To get information about potential promoter activity of the isolated sequences, *in silico* analyses with promoter prediction program ConSite (Sandelin et al., 2004) was conducted. Promising candidates, comprising many binding sites, were cloned into the reporter vector pGL3-basic directly upstream the firefly luciferase gene and relative luciferase activity was measured 48 hours post transfection of CHO dhfr^{-} cells.

Figure 14 depicts an example of 2 isolated CHO sequences (F6, K6) which were analyzed revealing multiple TF binding sites.

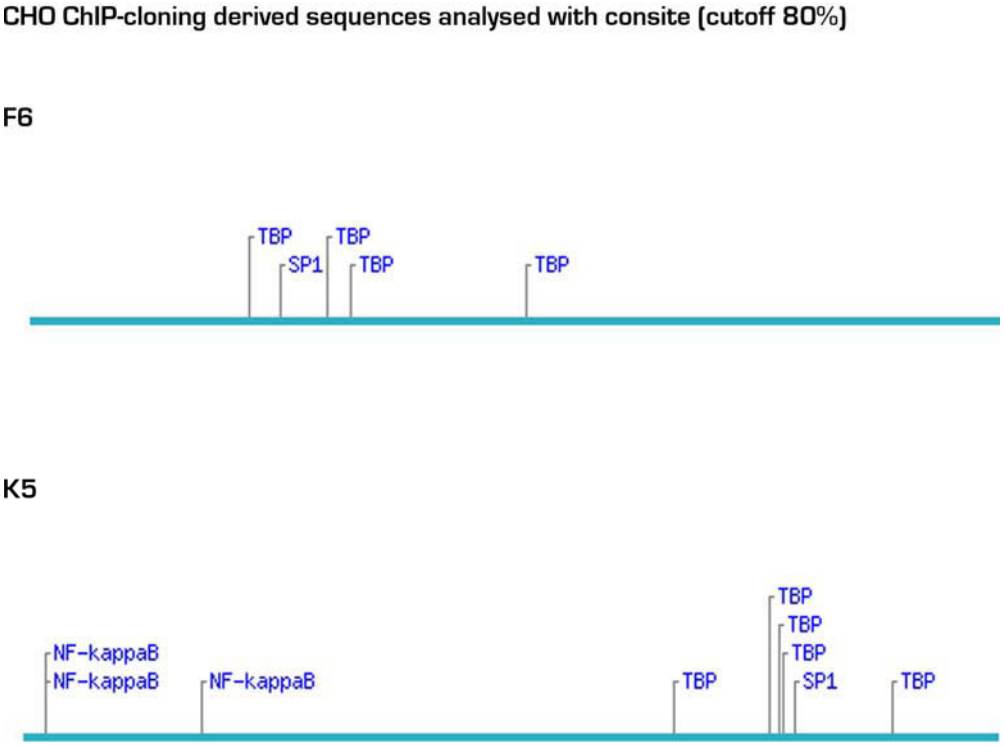


Figure 14: Example of Consite output file analyzing CHO derived sequences from ChIP-cloning experiments. Sequences were analyzed for TBP, SP1, NfκB.

The sequences of F6 and K5, both contained an ATG start codon, which could interfere with promoter activity of the construct. To analyze the influence of the ATG start codon, truncated fragments of F6 and K5 were constructed by by oligo synthesis (Sigma Aldrich, Austria) followed by annealing and PCR described in molecular methods section. The luminescence values obtained did not report any significant reporter activity compared with the SV40 positive control (Figure 15). The results of K5 and F6 reflect the outcome of multiple ChIP-cloning experiments conducted by Martina Baumann leading to an enrichment of non regulatory sequences.

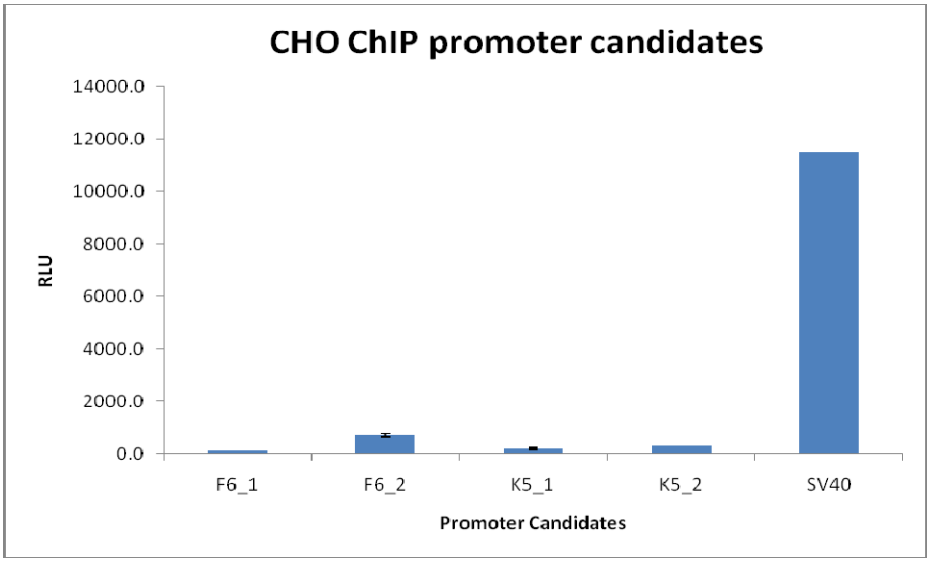


Figure 15: Measurement of relative luciferase activity 48 hours post transfection of CHO dhfr$^-$ cells with potential promoter candidates F6_1, K5_1 and truncated fragments F6_2, K5_2 cloned into pGL3-basic.

4.1.2 ChIP-cloning with HEK293 cells

To evaluate the method of ChIP-cloning to identify regulatory elements, human embryonal kidney cells (HEK293) cells were used instead of CHO cells. Sequences obtained by ChIP-cloning were aligned using the nucleotide blast of the Basic Local Alignment Search Tool (BLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequences identified as human were further analyzed with the UCSC Genome Browser (GRCh37/hg19 Assembly) (Kent et al., 2002). The browser provides a display of the sequence query, together with dozens of aligned annotation tracks (known genes, predicted genes, ESTs, mRNAs, CpG islands, assembly gaps and coverage, chromosomal bands, mouse homologies, transcription factor binding sites, histone modification patterns, DNaseI hypersensitivity regions) (Rosenbloom et al., 2010). These tracks are obtained from the Encyclopedia of DNA Elements (ENCODE) Consortium. in the ENCODE genome browser to identify functional elements.

As described above and in the methods section HEK293 cells were treated with formaldehyde and ChIP was performed. To amplify eluted DNA fragments short linkers were ligated and linker mediated PCR was conducted. Linker mediated PCR products were loaded on a 1% agarose gel (Figure 16).

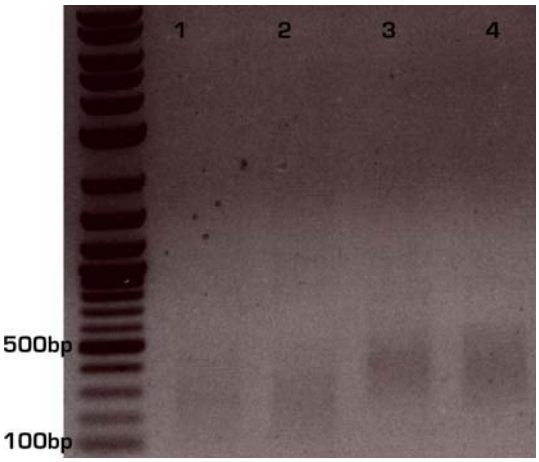


Figure 16: Linker mediated PCR. Lane 1– 2 PCR product of the first round. Lane 3 – 4 PCR product of second amplification step. 2-log DNA ladder.

After blunt end ligation into pCR®4Blunt-TOPO and transformation into chemically competent *E.coli* using the TOPO reaction (Invitrogen-Corporation, 1999–2004), colony PCR showed a number of differently sized fragments in the range of 300 - 700bp (Figure 17).

Figure 17: Colony-PCR showing size distribution of ChIP derived sequences. 1-17 amplified inserts of pCR®4Blunt-TOPO. L: 2-log DNA ladder.

Sequence	length bp	ConSite TFBS	NNPP TSS	Enhancer	DnaseI
A2	1020	13	3		
B1	956	6	1		
B3	1004	5	1		+
B5	700	7	0		
C1	960	8	2		
C4	884	13	1		
D5	442	8	1		
E1	367	0	1	+	+

E2	1017	0	0		
E4	352	0	0		
A5	568	3	0		
A6	425	6	0		
A8	406	3	0		
B5	359	1	0		
B8	498	0	0		
B9	567	3	0		
B10	494	0	0		
C3	461	11	1		

Table 21: ChIP-cloning sequences were analyzed with ConSite for transcription factor binding sites (NfkB, Sp1, TBP) and NNPP identifying possible TSS. Enhancer and DNaseI hypersensitivity regions from database of Encyclopedia of DNA Elements (ENCODE) were found only in two cases.

Table 21 reflects the great discrepancy of in silico analysis and mapping results, limiting this approach for identification of endogenous promoter elements in CHO cells with ChIP. Although multiple binding sites are predicted, only two sequences showed annotated enhancer activity or were sensitive to DNaseI, indicating promoter activity. Having an excess of non specifically precipitated DNA fragments followed by a non high-throughput method like in ChIP-cloning drastically raises the screening effort to identify regulatory elements. What influences binding of unspecific sequences?

Successful ChIP is strongly dependent on antibody specificity, which was analyzed by western blot.

4.1.3 Testing antibody specificty in westernblot analyses

Nuclear extracts of CHO and HEK293 cells were analyzed by SDS-PAGE followed by western blot analysis. The two antibodies used in ChIP, TFIIB (SI-1): sc-274 and TAF II p250 (L-20): sc-17134 (Santa Cruz, CA, USA) target the general transcription factors TFIIB (35kDa) and the large subunit of TFIID, TAF1 which is a TATA box binding protein (TBP)-associated factor (250kDa).

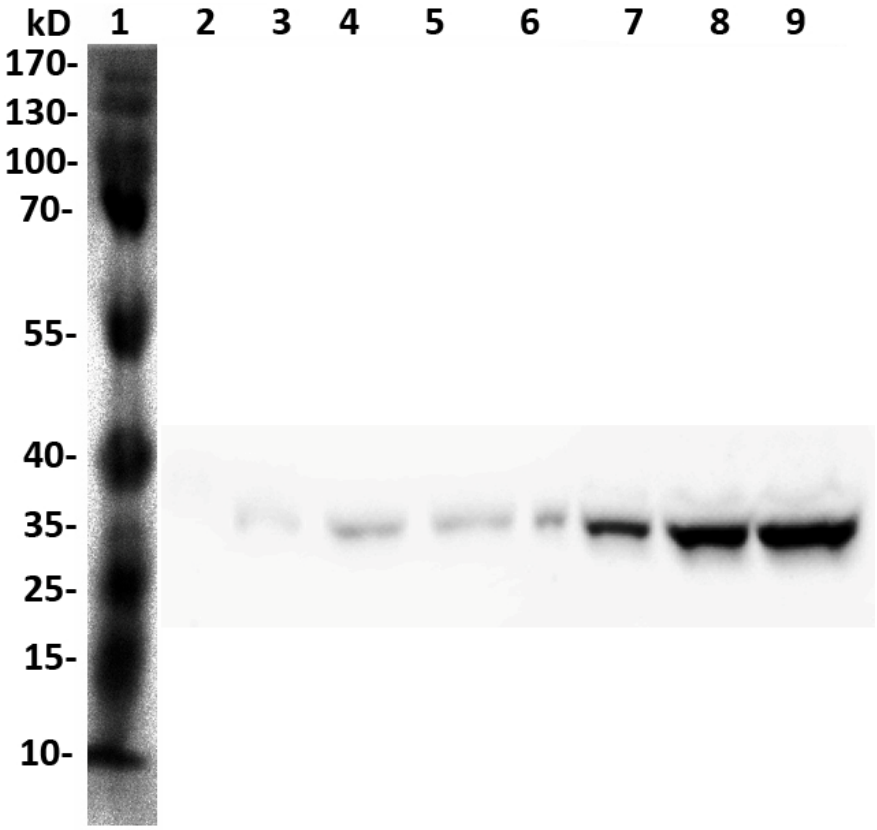


Figure 18: Western blot of 10% tris-glycine gel SDS-page against TFIIB (SI-1): sc-274 (Santa Cruz). Lane 2-5, CHO nuclear fractionation extract (1µg, 5µg, 10µg, 15µg). Lane 6-9, HEK nuclear fractionation extract (1µg, 5µg, 10µg, 15µg). Lane 1 PageRuler™ (Fermentas). Detection with anti-rabbit IgG (whole molecule) peroxidase conjugate super signal west pico chemiluminescent substrate.

The antibody (SI-1): sc-274 was able to detect TFIIB (35kDa) both in HEK and CHO nuclear extracts, showing high specificity resulting in a single band at the 35 kDa marker position (Figure 18). Detection was lower in CHO extracts. This could be explained by the fact that the antibody was developed against human TFIIB, leading to a lower binding activity for Chinese hamster TFIIB. Most antibodies on the market are designed against human or mouse derived proteins. No activity could be observed with TAF II p250 (L-20): sc-17134 against TAFIIp250, although transfer of nuclear proteins to the PVDF membrane was determined with Poinceau S staining (Figure 19). This antibody was developed against human TAFIIp250, is polyclonal and should therefore bind multiple targets on this large protein, but failed in multiple attempts detecting its target in nuclear extracts of both cell lines.

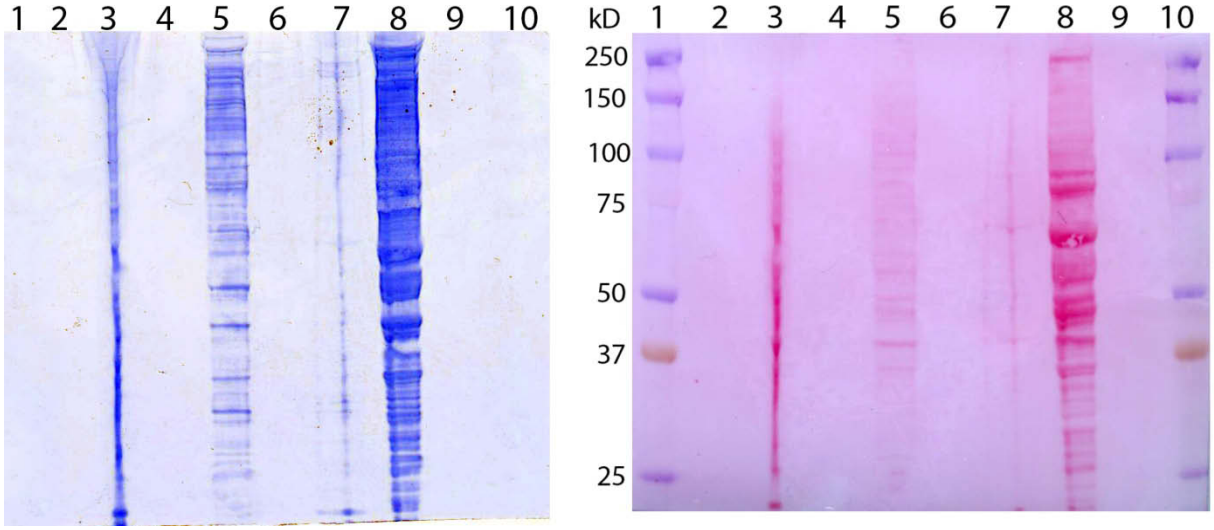


Figure 19: Left: Reducing SDS-PAGE 4-12% Bis-Tris (NuPAGE, Invitrogen) of CHO and HEK nuclear fractionation samples. CHO Pellet lane 3, CHO lysate lane 5. HEK Pellet lane 7, HEK lysate lane 8. Lane 1 and 10 Protein Standard. Gel was stained with Simply Blue SafeStain (Invitrogen) after blotting. Right: PVDF membrane stained with 0,1% Poinceau S solution.

4.1.4 Limitations of ChIP-cloning and further considerations

Unfortunately even when using the western blot validated antibody for ChIP experiments, the high rate of unspecific binding drastically raised screening efforts to possibly identify gene regulatory sequences. Non specific binding could have several reasons:

Crosslinking of cells to covalently link protein/DNA interactions has an effect on antibody specificity. Although the antibody was tested in western blots using nuclear extracts, formaldehyde could alter the antibodies target protein or other proteins leading to decreased binding and/or cross reactivity with other proteins bound to non regulatory sequences.

Recently a number of genome wide ChIP studies revealed, that transcription factors vary greatly in their number of genomic binding sites, and that binding events can significantly exceed the number of known or possible direct gene targets. It has therefore been proposed, that these excess binding sites account for low level gene expression or induce chromatin looping, changing chromatin and nuclear structure, (MacQuarrie et al., 2011).

A valuable tool to further evaluate specificity of the antibody used is to conduct an RNA interference knockdown experiment, where the absence of the target

protein should also result in the absence of the specific band in Western blot analysis, validating the antibody target recognition and possible cross reactivity reactions (Yvonne B. Sullivan, 2008).

As next generation sequencing technologies such as 454 and Illumina are currently evolving, rapid acquisition of short reads is possible. Usually these technologies are used in genome wide ChIP experiments where ChIP derived DNA fragments are directly sequenced and mapped to a reference genome (Park, 2009). It was shown that expression profiling of CHO cells was possible by mapping the sequence reads to a scaffold of a related organism (mouse) (Birzele et al., 2010). It might be possible to exploit the same approach for ChIP experiments, conducted with one specific and one unspecific antibody, mapping the sequence reads to the mouse reference genome and subsequently identify regulatory hotspots and gene regulatory elements on a genome wide basis.

4.2 Part II: Artificial promoters

Having discussed the drawbacks of viral derived promoter sequences in chapter 1.1.3, as well as the often finite possibilities to isolate endogenous promoter sequences a feasible alternative solution are "artificial" promoters, i.e. "made-up" sequences that are not found in living organisms. Designing artificial sequences that attract transcription factors and initiate transcription requires determining the features that capture the interactions between these proteins and the DNA. Highly active promoters exhibit nucleotide patterns that are different from the majority of the genome, most notably the abundance of GC and the di-nucleotide CpG. Since methylated CpG is subject to spontaneous deamination, it is the most infrequent dinucleotide genome-wide, unless there is a selective pressure to keep it. CpG-containing motifs have been reported to be both necessary and sufficient to bind Pol II abundantly in more than one tissue to transcribe both housekeeping genes and genes with tissue-specific expression in multiple cell types. In the work of Grabherr et al. (Grabherr et al., 2011), genome-wide frequencies of short motifs were analyzed showing that infrequent occurring nucleotide compositions are a distinctive attribute of highly active promoters. In the study a score (the "α score") was computed for 12 consecutive base pairs ("12-mer"), based on the frequencies of di-nucleotides, tri-nucleotides etc. compared to the genome-wide expectation (for details see chapter 1.4).

To create longer, contiguous artificial promoter sequences, the *a* score for each possible 12-mer, including those not present in the human genome was calculated, subsequently overlapping uncommon motifs to build a "promoter-like" 160,000 nucleotide (nt) long contiguous sequence, which is very rich in G/C (60%) and CpG (22%), and contains exact instances of the consensus of known binding sites, such as the TFIIB Recognition Element (BRE), TATA-Box, CAAT-Box, and the Initiator element (Inr). The sequence has no homology to any human (or any other sequenced) genome over more than 18 base pairs.

Using *in silico* promoter prediction programs based on fruit fly (Reese, 2001) different promoter constructs using different criteria and of different lengths (50, 110, 200, 232 and 300 base pairs) were selected from the concatomer. All constructs except a 50 nt long sequence gave rise to strong *in vitro* promoter

activity in a reporter gene assay in four mammalian cell lines: CHO (hamster ovary); P19 (mouse embryo); Vero (monkey kidney); and HEK293 (human kidney). The 232 (bp) base pair long sequence ArS232 is made of contiguous sequences out of the concatomer containing 2 high-scores, and was used in the thesis at hand to test if artificial sequences might comprise a scaffold for the incorporation of desired gene regulatory elements and transcription factor binding sites that meet the requirement of the particular production system. ArS232 contains one perfect and 12 imperfect (1 mismatch allowed) instances of TFIIB Recognition Element (BRE) (SSRCGCC), TATA-Box (TATAWAAR), CAAT-Box (GGCCAATCT), and the Initiator element (Inr) (YYANWYY).

ArS232

TCGACGCGCGTATAACACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTAC
GCCCGTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCA
AGCGCCGATCGGTAATCGGACGATTCGGATACGCGAGTTCGGACGTACGAGCGTGATA
CGGCGCGTAACGGTGCGCGTTAAAACGCCGACGCGTCATAACCGCGACTCGTCGACGC

4.2.1 Modifying ArS232 promoter activity in a predictable manner by introducing various sequence motifs

The aim was to adjust promoter strength by changing the nucleotide composition as well as to introduce known motifs. Making artificial sequences more predictable could in the long term lead to optimized systems for recombinant protein production.

Following Ars 232 modifications were made:

- Introduction of Inhibitory sequences composed of nucleotides occurring more frequently in the genome is expect to decrease promoter activity

Ars232-dT12-1I

TCGACGCGCGCACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTACGCCC
GTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCAAGCG
CCGATCGGTAATCGGACGATTCGGATATATCCAAAGTATTACTTTACTTATGTACGGCGC
GTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT12_3I

TCGACGCATGAAGATTCTTAACATCTTCGTTGGCGCGCTAACGCGAGAATCACTTAACA
AAGGGCTTATCAATCGCGCGACTTGTGCGCGACAAGATGTGAAACTTATGAAGGCGCC
GATCGGTAATCGGACGATTCGGATACGCGAGTTCGGACGTACGAGCGTGATACGGCGC
GTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT12_4I

TCGACGCATGAAGATTCTTAACATCTTCGTTGGCGCGCTAACGCGAGAATCACTTAACA
AAGGGCTTATCAATCGCGCGACTTGTGCGCGACAAGATGTGAAACTTATGAAGGCGCC
GATCGGTAATCGGACGATTCGGATATATCCAAAGTATTACTTTACTTATGTACGGCGCGT
AACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

- The nuclear factor kappa light chain B (NFκB) response element was introduced to gain more information about the interplay of known sequence motifs acting in concert with artificial sequences. 5 copies of NFκB were analyzed together with 1-4 copies of overrepresented inhibitory sequences

ArS232-dT12_5NF-kB

TCGACGCGCGCACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTACGCCC
GTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCAAGCG
CCGATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCGCTGG
GGACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCCGCGAGTTCGGACGTACGAGC
GTGATACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACG
C
ArS232-dT12_5NF-kB_1I

TCGACGCGCGCACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTACGCCC
GTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCAAGCG
CCGATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCGCTGG
GGACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCTATCCAAAGTATTACTTTACTTA
TGTACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT12_5NF-kB_3I

TCGACGCATGAAGATTCTTAACATCTTCGTTGGCGCGCTAACGCGAGAATCACTTAACA
AAGGGCTTATCAATCGCGCGACTTGTGCGCGACAAGATGTGAAACTTATGAAGGCGCC
GATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCGCTGGGG
ACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCCGCGAGTTCGGACGTACGAGCGT
GATACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT12_5NF-kB_4I

TCGACGCATGAAGATTCTTAACATCTTCGTTGGCGCGCTAACGCGAGAATCACTTAACA
AAGGGCTTATCAATCGCGCGACTTGTGCGCGACAAGATGTGAAACTTATGAAGGCGCC
GATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCGCTGGGG
ACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCTATCCAAAGTATTACTTTACTTATG
TACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

Promoter constructs were cloned into the pGL3-basic vector and transfected into HEK293 cells. 48h post transfection bioluminescence was measured as depicted in Figure 20.

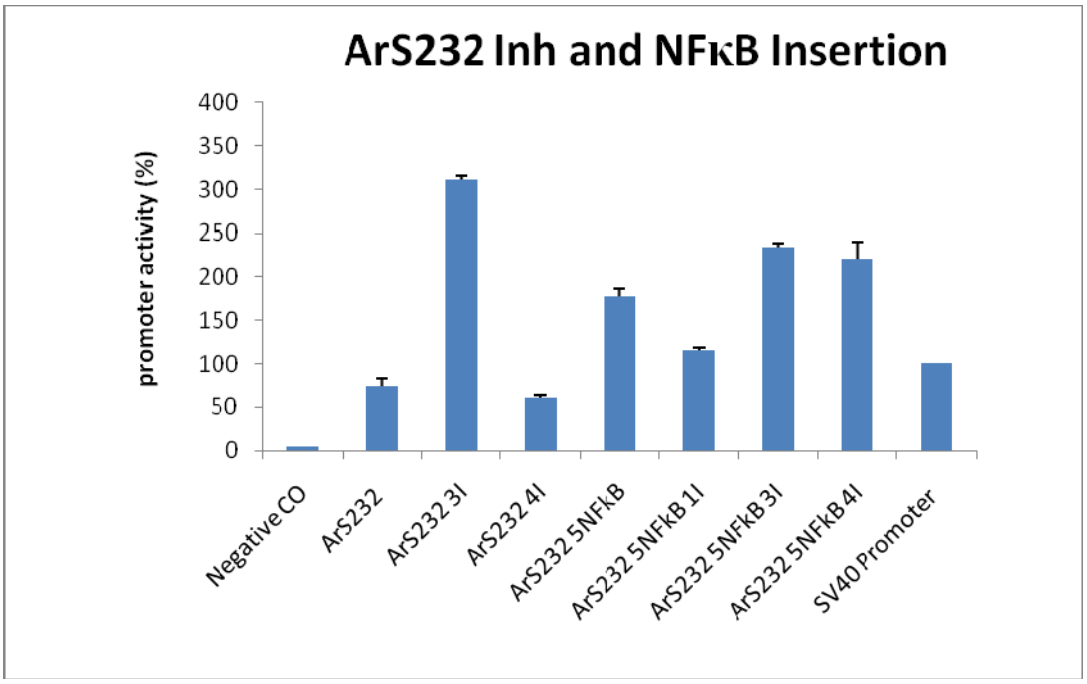


Figure 20: Measurement of bioluminescence 48 h post transfection of HEK293 cells with ArS232 promoter constructs ArS232, ArS232 3I, ArS232 4I, ArS232 5NFκB, ArS232 5NFκB 1I, ArS232 5NFκB 3I, ArS232 5NFκB 4I, SV40 Promoter upstream a firefly luciferase reporter gene in the pGL3 vector and a negative control comprising the pGL3-basic vector (data was internally normalized by renilla luciferase).

Inhibitory elements should not attract TFs and therefore lower the activity of the promoter construct. Surprisingly this is not the case when three inhibitory elements are present in ArS232 3I where activity increases up to 300% compared with the SV40 promoter. Inserting 4 inhibitory motifs in ArS232 4I the activity is reduced again to the level of ArS232. Inhibitory motifs containing AT-rich sequences which are known to possess a higher propensity for bending, with longer stretches being more prone to bending, alleviating the binding of TFs to neighboring binding sites (Huang et al., 2005). Introducing a too long stretch of AT-rich sequence in the ArS232 4I construct seems to negatively influence promoter activity. Introduction of five NfκB sites also elevates activity (ArS232 5NFκB), but the effect of introducing inhibitory motifs is somewhat different than observed with ArS232. As mentioned in chapter 1.6.1(NfkB signal transduction), in not induced cells NfκB is localized in the cytosol, therefore binding sites should not be occupied. This seems to have a positive effect on promoter activity, which

can even be enhanced further by the introduction of inhibitory sequences (ArS232 5NfκB 3I and 4I). This result shows that promoter activity can be adjusted multifold by introducing either known motifs, as well as changing nucleotide composition of artificial sequences. Further experiments are needed to investigate why certain compositions of motifs e. g. ArS 3I show enhanced promoter activity contradicting the prediction. To decipher the interplay of different motifs step by step, the prominent motif TATA-box was introduced.

4.2.2 TATA-box insertion has a minor effect on promoter activity

To narrow down the influence of one specific element on artificial promoter activity, one TATA-box motif was inserted in ArS232 1I, ArS232 5NFκB and ArS232 5NFκB 1I.

Ars232-dT2-1I

TCGACGCGCGTATAACACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTAC
GCCCGTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCA
AGCGCCGATCGGTAATCGGACGATTCGGATATATATCCAAAGTATTACTTTACTTATGTACG
GCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT2_5NF-kB

TCGACGCGCGTATAACACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTAC
GCCCGTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCA
AGCGCCGATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCG
CTGGGGACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCCGCGAGTTCGGACGTAC
GAGCGTGATACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTC
GACGC

ArS232-dT2_5NF-kB_1I

TCGACGCGCGTATAACACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTAC
GCCCGTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCA
AGCGCCGATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCG
CTGGGGACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCTATCCAAAGTATTACTTT
ACTTATGTACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGA
CGC

Bioluminescence values are shown in Figure 21. Introduction of a TATA-box elevated promoter activity in all three constructs, but the effect was weaker than the one observed with the introduction of inhibitory elements in the elements ArS232 3I, depicted in Figure 20. It seems that promoter activity is rather influenced by the introduction of long stretches of multiple motifs like 5 NFκB or 3 inhibitory sequences than by altering a short motif like a TATA-box. Nevertheless the prediction that insertion of a TATA-box would result in higher promoter activity was confirmed.

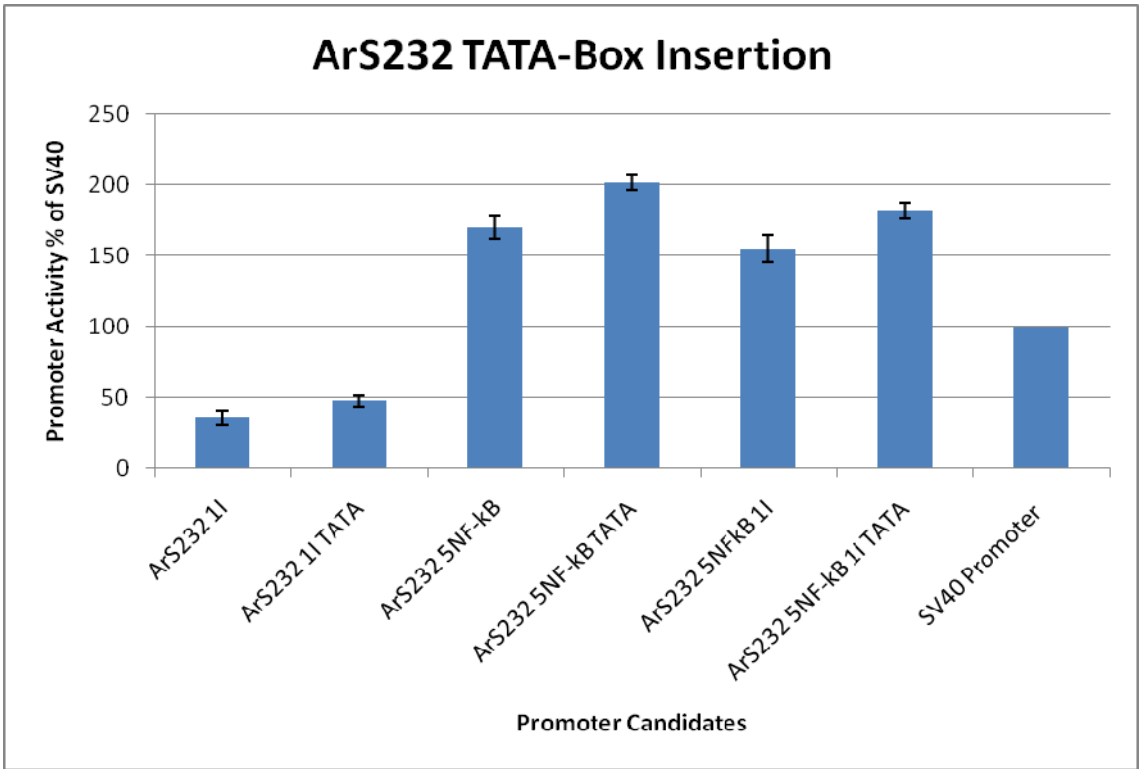


Figure 21: Measurement of bioluminescence 48h post transfection of HEK293 cells with ArS232 1I /-TATA, ArS232 5NFκB /-TATA and ArS232 5NFκB /-TATA upstream a firefly luciferase reporter gene in the pGL3 vector (data was internally normalized by renilla luciferase).

4.2.3 Stimulating ArS232 5NFκB promoter activity with TNFα

The ArS232 5NFκB construct was used to test if an inducible promoter element like the NFκB response element could lead to a higher *in vitro* activity when induced by stimuli, like LPS or TNFα, known for activating the NFκB-pathway through receptor mediated proteolysis (see introduction). HEK293 cells where transfected with ArS232 respectively ArS232 5NFκB and incubated for 12 h for recovery. 20 ng/ml TNFα was added and bioluminescence was measured 6 h post

induction. Figure 22 depicts a 2.9 fold increase in promoter activity when inducing the ArS232 5NFκB construct with 20 ng/ml TNFα for 6h, whereas the ArS232 activity is reduced by 2.7 fold upon induction. The result shows that the 5 NFκB response elements embedded in the ArS232 backbone are able to drive reporter gene expression upon specific induction with NFκB. An intriguing feature of signaling by TNFα is its ability to stimulate both death and survival in cells through directly regulating the expression of antiapoptotic genes such as cIAP1/2 and Bcl-XL (Kucharczak et al., 2003). If NF-κB signaling is blocked, then exposure to TNFα induces rapid apoptosis in most cell types. Internal normalization should compensate for an apoptotic effect of TNFα on the ArS232 transfected cells. The reduction of 2.7 fold could therefore not be explained with a different viability after stimulation with TNFα. ArS232 is able to drive reporter gene expression in different mammalian cell lines (CHO, P19, Vero, HEK293) (Grabherr et al., 2011), and appears to be less variable and species specifically regulated than the viral promoter SV40. With the ArS232 5NFκB construct an artificial inducible test promoter was constructed, applicable for reporter gene studies on inhibitory molecules targeting dysregulated pathways of NFκB, or theoretically any other inducible element working in the environment of the ArS232 backbone.

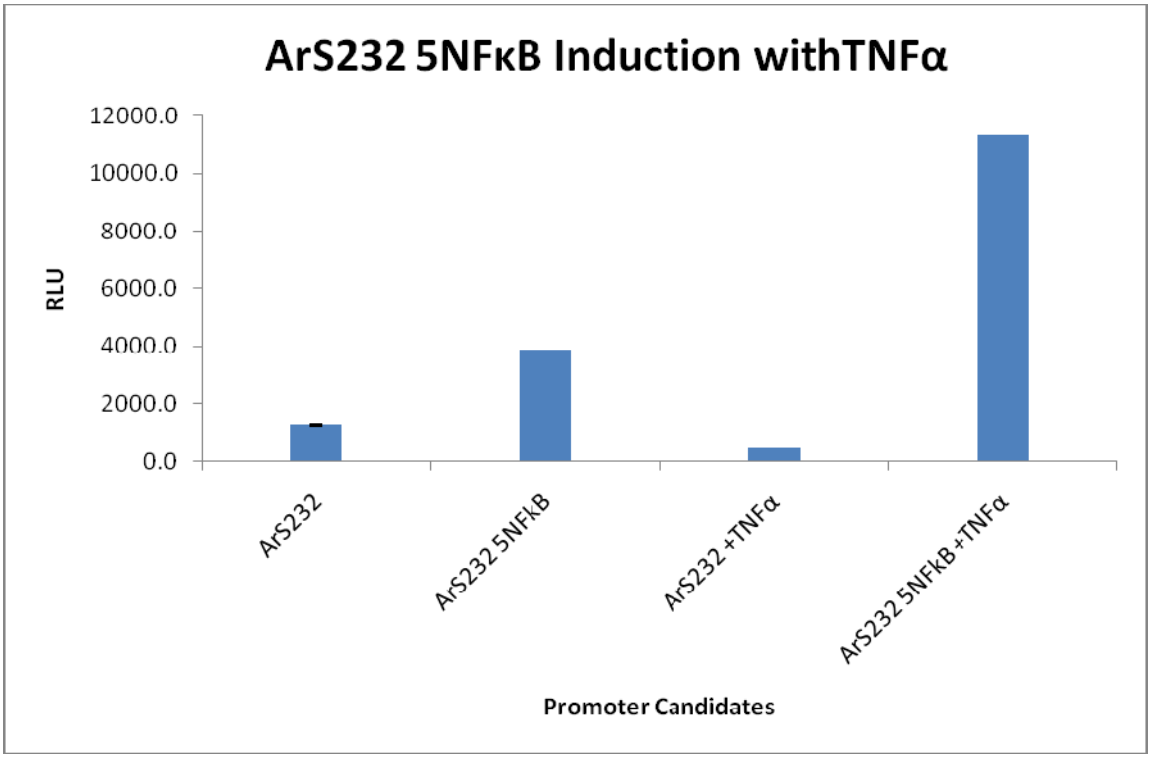


Figure 22: Measurement of bioluminescence of HEK293 cells transfected with ArS232 respectively ArS232 5NFκB upstream a firefly luciferase reporter gene in the pGL3 vector (data was internally normalized by renilla luciferase) 6 h post induction with 20 ng/ml TNFα.

4.2.4 U937 cell line and LPS induction

To further investigate, if the ArS232 5NFκB element could act as a live cell reporter upon stimulation of the various NFκB pathways, U937 cells were chosen. U937 cells, established from a diffuse histiocytic lymphoma cell line, mature and differentiate in response to a number of soluble stimuli, adopting the morphology and characteristics of mature macrophages. HEK293 cells show no reaction upon stimulation with lipopolysaccharide (LPS), because they lack the toll-like-receptor 4 (TLR4)) responsible for binding LPS mediated by cofactors LPS binding protein LBP, MyD88 and CD14 leading to signal transduction via the NFκB pathway active in the innate immune response. Although stably transfected reporter cell lines (HEK293) expressing various TLRs exist (for example HEK-Blue™ TLR2 cells) (InvivoGen, 2010), the ArS232 5NFκB element could act as a transient reporter element. The biggest issue with U937 cells is to find a sufficiently efficient transfection protocol. Although Martinet et al. (Martinet et al., 2003) reported 80% positive transfectants using the Nucleofactor® Kit V (Amaxa®), these values could not be achieved. U937 cells were transfected with the ArS232 5NFκB respectively the ArS232 element, after 12 h stimulation with 1µg/ml LPS was applied. Measurement of bioluminescence 8 h post induction resulted in poor overall promoter activity and no induction (data not shown). Therefore the method for transfecting U937 cells has to be improved to test if the ArS232 construct is able to drive reporter gene expression in this cell line.

5 Conclusion

Mammalian cell lines, especially CHO cells, are widely used for the production of recombinant therapeutic proteins, which currently represent the main drivers for revenue growth in market sales, reflecting also the clinical benefit of "biologicals". The capacity for proper folding, assembly and post translational modifications made mammalian hosts superior compared to other eukaryotic or bacterial production systems. One of the main drawbacks and limitations are the relatively low volumetric production rate, reflecting in larger and more expensive production facilities, as well as the time consuming and laborious development of a production cell line. Different efforts improving mammalian production systems, mainly focusing on process, media, and cell line improvements, are currently underway. Focusing on the transcriptional regulation of heterologous gene expression, and subsequently higher production titers is an approach pursued by the study at hand. The undesired side effects of currently used viral derived sequences are e. g. transgene silencing, heterogeneity in production clones, premature activation of apoptotic pathways. The use of endogenous regulatory elements for trans gene expression avoids these and other undesired effects.

In this study Chromatin Immunoprecipitation (ChIP) was used to isolate genomic fragments bound by the genral transcription factors, which assemble and attract Pol II at promoter regions of genes. As the genomic sequence of the Chinese hamster is currently not available, methods which are independent of a reference genome are in favor for identifying endogenous elements in CHO cells. To evaluate the specificity of ChIP, isolates from HEK293 cells were mapped to the human reference genome and also analyzed with promoter prediction programs. The results showed a great discrepancy between in silico analyses, which yielded various binding sites as well as transcription start sites indicating promoter activity, and mapping results from ENCODE database of regulatory elements. It seems that both methods are limited in their specificity and yield a high number of false positives. Western blot analysis of nuclear extracts with antibodies used in the ChIP experiments revealed that only the smaller 35 kDa TFIIB protein could be detected, whereas the larger 250 kDa subunit of TFIID TAFp250 was

not detectable. Overall, the ChIP-cloning was not specific enough demanding laborious and time consuming screening efforts to identify regulatory elements.

In the second part, artificial promoter constructs were modified with known regulatory motifs (TATA-box, BRE, NfκB) to custom design promoters that are suitable for biotechnological and medical applications. Interestingly these artificial sequences could drive reporter gene expression at levels comparable or even exceeding that of the SV40 viral promoter in transient reporter gene expression assays. The artificial promoter comprises underrepresented genomic sequences which are lost in genomes over evolutionary time due to the high mutation rate of the hypermethylated CpG dinucleotide, unless there is a selective pressure to keep these sequences, e. g. at promoter regions of genes. Transient reporter gene assays with artificial promoter constructs, comprising the above mentioned known motifs, as well as overrepresented AT rich sequences, were conducted. This experiment revealed that overrepresented elements, which should actually not attract TFs, elevated activity up to 300% compared with the SV40 promoter. As discussed, AT-rich sequences are known to possess a higher propensity for bending, which could alleviate the binding of TFs to neighboring binding sites. This result shows that more experiments have to be carried out, to investigate proper spacing of elements, to make promoter activities more predictable. To investigate if the Ars 232 backbone could work with an inducible response element (NfκB), induction with TNFα was performed, resulting in a 2.9 fold induction of reporter gene expression. This result could be interpreted as an expansion of tools modifying promoter strength in a less empirical and more straightforward approach. An application of the inducible promoter construct could be a reporter system containing specific response elements like the NfkB element in an artificial promoter that is active in a wide variety of cell types.

6 References

ABCAM (1998-2011): Nuclear fractionation protocol.

AGGARWAL, B. B. (2003): Signalling pathways of the TNF superfamily: a double-edged sword. Nat Rev Immunol 3: 745-756.

AGGARWAL, S. (2010): What's fueling the biotech engine[mdash]2009-2010. Nat Biotech 28: 1165-1171.

AGILENT (2010): Mammalian ChIP-on-chip Protocol.

AGILENT TECHNOLOGIES, I. (2011): Mammalian ChIP-on-chip Protocol 10.2. Agilent Technologies, Inc.5301 Stevens Creek Blvd. Santa Clara, CA 95051.

ALBERTS, B. (2008): Molecular biology of the cell. 5. ed., reference ed., New York, NY [u.a.]: Garland Science.

ALBRIGHT, S. R.; TJIAN, R. (2000): TAFs revisited: more data reveal new twists and confirm old ideas. Gene 242: 1-13.

ANISH, R. et al. (2009): Characterization of Transcription from TATA-Less Promoters: Identification of a New Core Promoter Element XCPE2 and Analysis of Factor Requirements. PLoS ONE 4: e5103.

BARNES, L. M.; BENTLEY, C. M.; DICKSON, A. J. (2003): Stability of protein production from recombinant mammalian cells. Biotechnology and Bioengineering 81: 631-639.

BAUMANN, M. (2008): Master Thesis: Idendification and Characterisation of potential CHO promoter sequences. Vienna, Austria, University of Natural Resources and Applied Life Sciences.

BELL, A. C.; WEST, A. G.; FELSENFELD, G. (1999): The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98: 387-396.

BERGER, S. L. (2002): Histone modifications in transcriptional regulation. Current Opinion in Genetics & Development 12: 142-148.

BIRD, A. (1999): Molecular biology - DNA methylation de novo. Science 286: 2287-2288.

BIRD, A. P. (1986): CpG-rich islands and the function of DNA methylation. Nature 321: 209-213.

BIRNEY, E. et al. (2007): Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799-816.

BIRZELE, F. et al. (2010): Into the unknown: expression profiling without genome sequence information in CHO by next generation sequencing. Nucleic Acids Research 38: 3999-4010.

BODE, J. et al. (2000): Transcriptional augmentation: Modulation of gene expression by Scaffold/Matrix-Attached regions (S/MAR elements). Critical Reviews in Eukaryotic Gene Expression 10: 73-90.

BREATHNACH, R.; CHAMBON, P. (1981): Organization and Expression of Eucaryotic Split Genes Coding for Proteins. Annual Review of Biochemistry 50: 349-383.

BROOKS, A. R. et al. (2004): Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. The Journal of Gene Medicine 6: 395-404.

BURGESS-BEUSSE, B. et al. (2002): The insulation of genes from external enhancers and silencing chromatin. Proceedings of the National Academy of Sciences of the United States of America 99: 16433-16437.

BURKE, T. W.; KADONAGA, J. T. (1996): Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes & Development 10: 711-724.

BUTLER, J. E. F.; KADONAGA, J. T. (2002): The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & Development 16: 2583-2592.

CHALKLEY, G. E.; VERRIJZER, C. P. (1999): DNA binding site selection by RNA polymerase II TAFs: a TAFII250-TAFII150 complex recognizes the Initiator. EMBO J 18: 4835-4845.

COLLAS, P. (2010): The Current State of Chromatin Immunoprecipitation. Molecular Biotechnology 45: 87-100-100.

DEER, J. R.; ALLISON, D. S. (2004): High-Level Expression of Proteins in Mammalian Cells Using Transcription Regulatory Sequences from the Chinese Hamster EF-1α Gene. Biotechnology Progress 20: 880-889.

DENG, W.; ROBERTS, S. (2007): TFIIB and the regulation of transcription by RNA polymerase II. Chromosoma 116: 417-429.

DENG, W.; ROBERTS, S. G. (2005): A core promoter element downstream of the TATA box that is recognized by TFIIB. Genes Dev 19: 2418.

EVANS, R.; FAIRLEY, J. A.; ROBERTS, S. G. E. (2001): Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. Genes & Development 15: 2945-2949.

FITZGERALD, P. C. et al. (2006): Comparative genomics of Drosophila and human core promoters. Genome Biology 7: 22.

FROHMAN, M. A.; DUSH, M. K.; MARTIN, G. R. (1988): Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proceedings of the National Academy of Sciences 85: 8998-9002.

GARDINER-GARDEN, M.; FROMMER, M. (1987): CpG Islands in vertebrate genomes. Journal of Molecular Biology 196: 261-282.

GILMORE, T. D. (2006): Introduction to NF-kappa B: players, pathways, perspectives. Oncogene 25: 6680-6684.

GRABHERR, M. G. et al. (2011): Exploiting Nucleotide Composition to Engineer Promoters. PLoS ONE 6: e20136.

GRAY, P. W. et al. (1984): Cloning and expression of cDNA for human lymphotoxin, a lymphokine with tumour necrosis activity. Nature 312: 721-724.

GUENTHER, M. G. et al. (2007): A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. Cell 130: 77-88.

HEINTZMAN, N. D. et al. (2007): Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39: 311-318.

HERNANDEZ, N. (1993): TBP, a universal eukaryotic transcription factor? Genes & Development 7: 1291-1308.

HOFFMANN, A.; NATOLI, G.; GHOSH, G. (2006): Transcriptional regulation via the NF-kappa B signaling module. Oncogene 25: 6706-6716.

HUANG, D.-B. et al. (2005): Crystal Structure of a Free [kappa]B DNA: Insights into DNA Recognition by Transcription Factor NF-[kappa]B. Journal of Molecular Biology 346: 147-160.

INVITROGEN (2004): Zero Blunt® TOPO® PCR Cloning Kit for Sequencing

INVITROGEN-CORPORATION (1999–2004).

INVIVOGEN (2010): HEK-Blue™-hTLR2 Cells, SEAP Reporter 293 cells expressing the human TLR2 gene; Technical data sheet.

JAYAPAL, K. et al. (2007): Recombinant Protein Therapeutics from Cho Cells - 20 Years and Counting. CHO Consortium: SBE Special Edition 40-47.

JUDD, R. C. (2002): Peptide Mapping by Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis. IN WALKER, J. M. (Ed.) The Protein Protocols Handbook. Humana Press.

JUVEN-GERSHON, T.; KADONAGA, J. T. (2010): Regulation of gene expression via the core promoter and the basal transcriptional machinery. Developmental Biology 339: 225-229.

KADONAGA, J. T. (2004): Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors. Cell 116: 247-257.

KAGEY, M. H. et al. (2010): Mediator and cohesin connect gene expression and chromatin architecture. Nature 467: 430-435.

KANTARDJIEFF, A. et al. (2006): Developing genomic platforms for Chinese hamster ovary cells. Biotechnology Advances 27: 1028-1035.

KENT, W. J. et al. (2002): The Human Genome Browser at UCSC. Genome Research 12: 996-1006.

KIM, T. H. et al. (2005): A high-resolution map of active promoters in the human genome. Nature 436: 876-880.

KUCHARCZAK, J. et al. (2003): To be, or not to be: NF-[kappa]B is the answer - role of Rel//NF-[kappa]B in the regulation of apoptosis. Oncogene 22: 8961-8982.

KWAKS, T. H. J.; OTTE, A. P. (2006): Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. Trends in Biotechnology 24: 137-142.

LAGRANGE, T. et al. (1998): New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes & Development 12: 34-44.

LEE, D.-H. et al. (2005): Functional Characterization of Core Promoter Elements: the Downstream Core Element Is Recognized by TAF1. Molecular and Cellular Biology 25: 9674-9686.

LEVINE, M.; TJIAN, R. (2003): Transcription regulation and animal diversity. Nature 424: 147-151.

LI, H.; D'ANJOU, M. (2009): Pharmacological significance of glycosylation in therapeutic proteins. Current Opinion in Biotechnology 20: 678-684.

LIU, Q. et al. (1999): Analysis of TFIIA Function In Vivo: Evidence for a Role in TATA-Binding Protein Recruitment and Gene-Specific Activation. Molecular and Cellular Biology 19: 8673-8685.

LYNCH, P. J.; RUSCHE, L. N. (2009): A Silencer Promotes the Assembly of Silenced Chromatin Independently of Recruitment. Mol. Cell. Biol. 29: 43-56.

MACLEOD, D. et al. (1994): Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes & Development 8: 2282-2292.

MACQUARRIE, K. L. et al. (2011): Genome-wide transcription factor binding: beyond direct target regulation. Trends in Genetics 27: 141-148.

MAKRIDES, S. C. (1999): Components of Vectors for Gene Transfer and Expression in Mammalian Cells. Protein Expression and Purification 17: 183-202.

MARTINET, W.; SCHRIJVERS, D. M.; KOCKX, M. M. (2003): Nucleofection as an efficient nonviral transfection method for human monocytic cells. Biotechnology Letters 25: 1025-1029-1029.

MASTON, G. A.; EVANS, S. K.; GREEN, M. R. (2006): Transcriptional regulatory elements in the human genome. Annual Review of Genomics and Human Genetics 7: 29-59.

MAY, M. J. (2006): A Nuclear Factor in B Cells and Beyond. The Journal of Immunology 177: 7483-7484.

MYERS, R. M. et al. (2011): A User's Guide to the Encyclopedia of DNA Elements (ENCODE). Plos Biology 9.

NAKABAYASHI, H. et al. (1991): A position-dependent silencer plays a major role in repressing alpha-fetoprotein expression in human hepatoma. Mol. Cell. Biol. 11: 5885-5893.

NG, P. et al. (2005): Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. Nat Meth 2: 105-111.

NISSOM, P. M. et al. (2006): Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. Molecular Biotechnology 34: 125-140.

NOORDERMEER, D. et al. (2008): Transcription and Chromatin Organization of a Housekeeping Gene Cluster Containing an Integrated β-Globin Locus Control Region. PLoS Genet 4: e1000016.

OHLER, U. et al. (2002): Computational analysis of core promoters in the Drosophila genome. Genome Biology 3: research0087.1-0087.12.

OMASA, T. et al. (2009): Bacterial artificial chromosome library for genome-wide analysis of Chinese hamster ovary cells. Biotechnology and Bioengineering 104: 986-994.

OROM, U. A.; SHIEKHATTAR, R. (2011): Long non-coding RNAs and enhancers. Current Opinion in Genetics & Development 21: 194-198.

PARK, P. J. (2009): ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 10: 669-680.

PETERSON, C. L. (1995): The SWI/SNF Protein Machine: Helping Transcription Factors Contend With Chromatin-Mediated Repression. IN ALAN, P. W. (Ed.) The Nucleosome. Amsterdam, Elsevier.

PRENTICE, H. L. et al. (2007): High level expression of proteins using sequences from the ferritin heavy chain gene locus. Journal of Biotechnology 128: 50-60.

PROMEGA (2007a): pGL3 Luciferase Reporter Vectors Technical Manual.

PROMEGA (2007b): Technical Bulletin: pRL-SV40 Vector (2007).

PROMEGA (2011): Dual-Glo® Luciferase Assay System Technical Manual.

PUCK, T. T.; CIECIURA, S. J.; ROBINSON, A. (1958): GENETICS OF SOMATIC MAMMALIAN CELLS : III. LONG-TERM CULTIVATION OF EUPLOID CELLS FROM HUMAN AND ANIMAL SUBJECTS. J Exp Med 108: 945-56.

QIN, J. Y. et al. (2010): Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. PLoS ONE 5: e10611.

REESE, M. G. (2001): Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Computers & Chemistry 26: 51-56.

ROSENBLOOM, K. R. et al. (2010): ENCODE whole-genome data in the UCSC Genome Browser. Nucleic Acids Research 38: D620-D625.

RUIZ, J. C.; WAHL, G. M. (1990): Chromosomal destabilization during gene amplification. Molecular and Cellular Biology 10: 3056-3066.

SAMBROOK, J.; RUSSELL, D. W. (2006): The condensed protocols from Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

SANDELIN, A. et al. (2007): Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet 8: 424-436.

SANDELIN, A.; WASSERMAN, W. W.; LENHARD, B. (2004): ConSite: web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Research 32: W249-W252.

SEN, R.; BALTIMORE, D. (1986): Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell 46: 705-716.

SHIRAKI, T. et al. (2003): Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proceedings of the National Academy of Sciences 100: 15776-15781.

SINGH, B. N.; HAMPSEY, M. (2007): A Transcription-Independent Role for TFIIB in Gene Looping. Molecular Cell 27: 806-816.

SMALE, S. T.; KADONAGA, J. T. (2003): THE RNA POLYMERASE II CORE PROMOTER. Annual Review of Biochemistry 72: 449-479.

TAATJES, D. J.; MARR, M. T.; TJIAN, R. (2004): Opinion - Regulatory diversity among metazoan co-activator complexes. Nature Reviews Molecular Cell Biology 5: 403-410.

TAATJES, D. J.; MARR, M. T.; TJIAN, R. (2004): Regulatory diversity among metazoan co-activator complexes. Nat Rev Mol Cell Biol 5: 403-410.

THOMAS, M. C.; CHIANG, C.-M. (2006): The General Transcription Machinery and General Cofactors. Critical Reviews in Biochemistry and Molecular Biology 41: 105-178.

THOMSON, J. P. et al. (2010): CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464: 1082-1086.

TOKUSUMI, Y. et al. (2007): A new core promoter element XCPE1 (X core promoter element 1) directs activator-, mediator-, and TBP-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters. Molecular and Cellular Biology MCB.01363-06.

TORA, L. (2002): A unified nomenclature for TATA box binding protein (TBP)-associated factors (TAFs) involved in RNA polymerase II transcription. Genes & Development 16: 673-675.

URLAUB, G.; CHASIN, L. A. (1980): Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proceedings of the National Academy of Sciences 77: 4216-4220.

VILAR, J. M. G.; SAIZ, L. (2005): DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. Current Opinion in Genetics & Development 15: 136-144.

VISEL, A. et al. (2009): ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457: 854-858.

WENDT, K. S. et al. (2008): Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451: 796-U3.

WILSON, C.; BELLEN, H. J.; GEHRING, W. J. (1990): Position Effects on Eukaryotic Gene Expression. Annual Review of Cell Biology 6: 679-714.

WURM, F. M. (2004): Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotech 22: 1393-1398.

YEE, J. C.; GERDTZEN, Z. P.; HU, W.-S. (2009): Comparative transcriptome analysis to unveil genes affecting recombinant protein productivity in mammalian cells. Biotechnology and Bioengineering 102: 246-263.

YVONNE B. SULLIVAN, J. N., DOUGLAS E. HUGHES AND BRIAN L. WEBB (2008): Validation of antibody specificity using RNA Interference -- Sullivan et al. 22 (1): 1070.25 -- The FASEB Journal.

ZAINA, S.; PEREZ-LUQUE, E. L.; LUND, G. (2010): Genetics Talks to Epigenetics? The Interplay Between Sequence Variants and Chromatin Structure. Current Genomics 11: 359-367.

ZHANG, Y.; REINBERG, D. (2001): Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev 15: 2343-60.

7 Appendix

7.1 Ars Sequences

Ars 232-dT12

TCGACGCGCGCACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTACGCCC
GTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCAAGCG
CCGATCGGTAATCGGACGATTCGGATACGCGAGTTCGGACGTACGAGCGTGATACGGC
GCGTAACGGTGCGCGTTAAAACGCCGACGCGTCATAACCGCGACTCGTCGACGC

Ars232-dT12-1I

TCGACGCGCGCACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTACGCCC
GTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCAAGCG
CCGATCGGTAATCGGACGATTCGGATATATCCAAAGTATTACTTTACTTATGTACGGCGC
GTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT12_3I

TCGACGCATGAAGATTCTTAACATCTTCGTTGGCGCGCTAACGCGAGAATCACTTAACA
AAGGGCTTATCAATCGCGCGACTTGTGCGCGACAAGATGTGAAACTTATGAAGGCGCC
GATCGGTAATCGGACGATTCGGATACGCGAGTTCGGACGTACGAGCGTGATACGGCGC
GTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT12_4I

TCGACGCATGAAGATTCTTAACATCTTCGTTGGCGCGCTAACGCGAGAATCACTTAACA
AAGGGCTTATCAATCGCGCGACTTGTGCGCGACAAGATGTGAAACTTATGAAGGCGCC
GATCGGTAATCGGACGATTCGGATATATCCAAAGTATTACTTTACTTATGTACGGCGCGT
AACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT12_5NF-kB

TCGACGCGCGCACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTACGCCC
GTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCAAGCG
CCGATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCGCTGG
GGACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCCGCGAGTTCGGACGTACGAGC

GTGATACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACG
C
ArS232-dT12_5NF-kB_1I

TCGACGCGCGCACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTACGCCC
GTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCAAGCG
CCGATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCGCTGG
GGACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCTATCCAAAGTATTACTTTACTTA
TGTACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT12_5NF-kB_3I

TCGACGCATGAAGATTCTTAACATCTTCGTTGGCGCGCTAACGCGAGAATCACTTAACA
AAGGGCTTATCAATCGCGCGACTTGTGCGCGACAAGATGTGAAACTTATGAAGGCGCC
GATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCGCTGGGG
ACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCCGCGAGTTCGGACGTACGAGCGT
GATACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT12_5NF-kB_4I

TCGACGCATGAAGATTCTTAACATCTTCGTTGGCGCGCTAACGCGAGAATCACTTAACA
AAGGGCTTATCAATCGCGCGACTTGTGCGCGACAAGATGTGAAACTTATGAAGGCGCC
GATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCGCTGGGG
ACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCTATCCAAAGTATTACTTTACTTATG
TACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

Ars232-dT2-1I

TCGACGCGCGTATAACACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTAC
GCCCGTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCA
AGCGCCGATCGGTAATCGGACGATTCGGATATATCCAAAGTATTACTTTACTTATGTACG
GCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGACGC

ArS232-dT2_5NF-kB

TCGACGCGCGTATAACACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTAC
GCCCGTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCA
AGCGCCGATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCG
CTGGGGACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCCGCGAGTTCGGACGTAC
GAGCGTGATACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTC
GACGC

ArS232-dT2_5NF-kB_1I

TCGACGCGCGTATAACACGCGAGCGGTTCGAACGTTGGCGCGCTAACGCGAGTCGTAC
GCCCGTCAACGCGGATCAATCGCGCGACTTGTGCGCGACGTTAGACCGCCGATCGTCA
AGCGCCGATCGGTAATCGGACGATTCGGATAGGGGACTTTCCGCTTGGGGACTTTCCG
CTGGGGACTTTCCGCTGGGGACTTTCCGCTGGGGACTTTCCTATCCAAAGTATTACTTT
ACTTATGTACGGCGCGTAACGGTGCGCGCGCCGACGCGTCATAACCGCGACTCGTCGA
CGC

7.2 CHO Sequences

F6_1

GAAGAACGCATTGTGATTTTGTACATAGCAATTGTTGAAGAAAGCACAGCTACTCTGAAT
ATTTTCTCAGATAGTTTTCAAGCCCTAAATATTTGCAAGTGGTGCATCTTAAAATTTCTTT
GCACTTTTTTTATGTCCTCCCAGCCGATTCCACCTCCATCTCCTATATGAAAGACTATGA
AAGAGAAAGAAGGGTTGATTCTGAAAGATGTGGGACAGTACAGGCTTACATCCACTTCC
ACTTGGAATTGCAGCACTAGAAGGCTCTGCTTAGCTACTTTTCTAGCATGAGTAAAGGA
ATTAAGGTGCAGAGGAATTCAGTAACTTAATATTAAGTTCACAGAGCTGAACAGCTACA
GAGTAGAGGCTGCAGCATCACTTCTAACAACCAGTGGCCATAAACTCCTTCCCATAAGT
GACAAGTTATTCCTTTGTTCTACAAAGATGTCTTCTGAGAAGGCTTTGAAAGTCCAAAAT
AATGGCTTGTCTTCTAGAAAATCTGTTGGTGTAGGGCTCTCT GTTTTGTGATGGC

F6_2

GAAGAACGCATTGTGATTTTGTACATAGCAATTGTTGAAGAAAGCACAGCTACTCTGAAT
ATTTTCTCAGATAGTTTTCAAGCCCTAAATATTTGCAAGTGGTGCATCTTAAAATTTCTTT
GCACTTTTTTTATGTCCTCCCAGCCGATTCCACCTCCATCTCCTATATGAAAGACTATGA
AAGAGAAAGAAGGGTTGATTCTGAAAGATGTGGGACAGTACAGGCTTACATCCACTTCC
ACTTGGAATTGCAGCACTAGAAGGCTCTGCTTAGCTACTTTTCTAGCATGAGTAAAGGA
ATTAAGGTGCAGAGGAATTCAGTAACTTAATATTAAGTTCACAGAGCTGAACAGCTACA
GAGTAGAGGCTGCAGCATCACTTCTAACAACCAGTGGCCATAAACTCCTTCCCATAAGT
GACAAGTTATTCCTTTGTTCTACAAAGATGTCTTCTGAGAAGGCTTTGAAAGTCCAAAAT
A

K5_1

AGACCACAACTGGAAATGTCCTTTGCAAGCAGAACTGAAAAGCAATGACAATAGTAACA
GTAACACCTGTGAAGTACTGGCAATGCAAAAGGAGAGTTTCCCTGTAGCTACAATAGGG
CCTTTGCTAAGTATACACTCAGCACTGAGGCTCACTGTGAGGGTGCTGACCAGGACCTC
CTGATCTCAGATGAAGTCAAAAGTAGGGAATCTTACAATTTCACAACTGACACCAGAGA
TGTTACTAGGAACTCCTTTTTGAGACATGTATCACTAGAACACACATTTACTAGAAGAAA
AAAGTGATCCGTGTTGAACAGGGTTCTGAGACAGGTCCATAAAAAGGAACAAGTACAGA
TAGGTTTTCTGCCTTTCCACATTACATACTCCCTTATAAAGCACACCTCTGTTACTGACTG
CATTACTGCTCGGCCTCTGTGAGTGAAAGCTCTTCATTTACACTTACCCCAGGATTCTTC
TCTCTGAATGAGACAAATGGTG

K5_2

AGACCACAACTGGAAATGTCCTTTGCAAGCAGAACTGAAAAGCAATGACAATAGTAACA
GTAACACCTGTGAAGTACTGGCAATGCAAAAGGAGAGTTTCCCTGTAGCTACAATAGGG
CCTTTGCTAAGTATACACTCAGCACTGAGGCTCACTGTGAGGGTGCTGACCAGGACCTC
CTGATCTCAGATGAAGTCAAAAGTAGGGAATCTTACAATTTCACAACTGACACCAGAGA
TGTTACTAGGAACTCCTTTTTGAGACATGTATCACTAGAACACACATTTACTAGAAGAAA
AAAGTGATCCGTGTTGAACAGGGTTCTGAGACAGGTCCATAAAAAGGAACAAGTACAGA
TAGGTTTTCTGCCTTTCCACATTACATACTCCCTTATAAAGCACACCTCTGTTACTGACTG
CATTACTGCTCGGCCTCTGTGAGTGAAAGCTCTTCATTTACACTTACCCCAGGATTCTTC
TCTCTGA

7.3 Abbreviations

AMP	adenosine 5'-monophosphate
Amp^r	ampicillin resistance gene
AS	antisense primer
ATP	adenosine 5'-triphosphate
BAC	bacterial artificial chromosome
bcl-2	B-cell lymphoma 2 gene
BHK	baby hamster kidney
BLAST	Basic Local Alignment Search Tool
bp	base pair(s)
BRE	TFIIB recognition element
BREd	downstream BRE
BREU	upstream BRE
BSA	bovine serum albumin
CAGE	cap analysis of gene expression
CAGG	chicken β-Actin promoter coupled with CMV early enhancer
CAT	chloramphenicol acetyltransferase
cDNA	complementary DNA
CGI	CpG island
CHEF1	Chinese hamster EF-1 a
ChIP	chromatin immunoprecipitation
CHO	Chinese hamster ovary
CMV	cytomegalovirus
CTCF	CCCTC binding factor
DCE	downstream core element
ddH20	double distilled water
DHFR	dihydrofolate reductase
$dhfr^-$	dihydrofolate reductase deficient
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphates
DPE	downstream promoter element
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic
EF1A	human elongation factor 1alpha promoter
EF-la	elongation factor-la
ER	endoplasmatic reticulum
EST	expressed sequence tag
g	gravity
Gb	giga base pairs
gDNA	genomic DNA
GFP	green fluorescent protein
GIS	gene identification signature
GPDH	glycerol-3 -phosphate dehydrogenase
GTF	general transcription factor

HBV	hepatitis B virus
HC	ferritin heavy chain gene locus
HEK-293	human embryo kidney 293
HEK-293	human embryo kidney
HEPES	4-(2-hydroxyethyl)-1 -piperazineethanesulfonic acid
HIV	human immunodeficiency virus
Inr	initiator
Jundl	jun proto-oncogene related gene dl
kb	kilo base pairs
LB medium	Luria-Bertani medium
LCR	locus control region
LDH-A	lactate dehydrogenase A
LTR	long tenninal repeat
lue	luciferase gene
M	molar [moll-1]
mAB	monoclonal antibody
MC S	multiple cloning site
mRNA	messenger RNA
MTE	motif ten element
MTX	methotrexate
NF-kB	nuclear factor kappa-light-chain-enhancer of activated B cells
NL\R	matrix attachment region
NNPP	neural network promoter prediction
NS0	mouse myeloma
OD	optical density
ori	origin of replication
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PET	paired-end ditag
PGK	mouse phosphoglycerate kinase 1 promoter
PIC	preinitiation complex
Pol II	RNA polymerase II
polyA	polyadenylation
PPi	pyrophosphate
PPP	promoter prediction program
PWM	position weight matrix
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNAi	RNA interference
rpm	revolutions per minute
RSV	Rous sarcoma virus
SAGE	serial analysis of gene expression
SAR	scaffold attachment region
SDS	sodium dodecyl sulfate
SOC medium	Super Optimal Catabolite medium
SS	sense primer

SV40	Simian virus 40
TAE	TRIS-acetate-EDTA
TAF	TBP-associated factor
TBA	TATA-binding protein
TF	transcription factor
TFBS	transcription factor binding site
TFIIA	transcription factor for RNA polymerase IIA
T_m	melting temperature
tPA	tissue plasminogen activator
TRIS	Tris(hydroxymethyl)-aminomethan
TSS	transcription start site
U	unit
UBC	human Ubiquitin C promoter
UPR	unfolded protein response
UTR	untranslated region
UV	ultraviolet
V	volt
w/v	weight/volume
XCPE1	X core promoter element 1
XCPE2	X core promoter element 2
TNFα	Tumor necrosis factor alpha
RHD	Rel homology domain