**RESEARCH**

# Identification of CHO Endogenous Gene Regulatory Elements

Jens Pontiller · Andreas Maccani · Martina Baumann ·
Ingo Klancnik · Wolfgang Ernst

**Abstract** The objective of this approach was to identify new CHO endogenous gene regulatory elements that are capable of regulating foreign gene expression in recombinant CHO host cells. The standard technology for the production of many biopharmaceutical products is frequently based on expression vectors that utilize strong mammalian viral promoters like SV40 or CMV which allow for very high expression rates but this may lead to constitutive over-expression resulting in a permanent stress for the cell. In addition, some heterologous promoters are cell-cycle dependent and can be subject to gene silencing generating heterogeneity within the cell population. Here, we describe the construction of a genomic CHO library and the subsequent identification and isolation of selected target sequences that are believed to be responsible for high level expression of the associated genes. The method that was used to isolate these regions of interest relies on gene specific amplification with primer pairs binding on different genes and the vector sequence. Flanking regions of these fragments were identified through Inverse PCR from fragmented and self-ligated genomic DNA. Expression levels of both the initially derived and the mapped fragments were determined through a luciferase reporter assay.

**Keywords** Endogenous promoter · Chinese hamster ovary cells · Recombinant protein expression · Library construction · DNA fragmentation

J. Pontiller (✉) · A. Maccani · M. Baumann · W. Ernst
Department of Biotechnology, Austrian Center of
Biopharmaceutical Technology, University of Natural Resources
and Applied Life Sciences, Vienna, Austria
e-mail: jens.pontiller@boku.ac.at

I. Klancnik
POLYMUN Scientific Immunbiologische Forschung GmbH,
Vienna, Austria

## Introduction

In recent years, mammalian cells have become the host system of choice for the production of recombinant proteins used for therapeutic applications. Using a mammalian expression system is advantegous concerning product secretion and post-translational modifications such as correct protein folding and glycosylation. Product quality, drug safety, and process-economy play a key role in the evaluation of expression systems to be used for the production of biopharmaceuticals [1]. One of the most widely used mammalian expression systems is the Chinese hamster ovary (CHO) cell system.

Efforts to optimize this system are mainly focused on process and cell line improvement but the transfected vector construct has significant impact on the overall performance. To achieve efficient transcription of the recombinant gene the expression plasmid is frequently engineered to contain regulatory sequences. A well-designed mammalian active expression cassette allows efficient cellular transcription and translation and ensures the production of large amounts of stable messenger RNA. For that purpose, commonly strong viral promoters are used, which guarantee high yields, but at the expense of premature activation of cellular apoptotic pathways. There are also several other drawbacks related to viral promoters such as the cytomegalovirus (CMV) immediate early promoter, the simian virus 40 (SV40) immediate early promoter or the Rous sarcoma virus (RSV) long-terminal-repeat promoter [2], since they can lead to permanent over-expression of the recombinant protein resulting in a number of stress reactions like the unfolded protein response (UPR) or the endoplasmatic reticulum (ER) stress response. Furthermore, these viral promoters show a cell-cycle dependence with the greatest transcriptional activity in the S-phase and the possibility of

gene silencing in stable cell lines leading to a heterogeneity between transfectants [3].

Here, we describe the identification of CHO-specific promoters directly from genomic DNA. For that purpose a CHO genomic library was constructed containing fragments of various lengths which were derived after enzymatic and mechanical fragmentation. This method is based on a gene specific amplification with primer pairs binding on different genes and the vector sequence. Flanking regions of these fragments were identified through Inverse PCR from fragmented and self-ligated genomic DNA. Expression levels were determined through a luciferase reporter assay.

## Materials and Methods

### Molecular Cloning and PCR

All cloning procedures were performed as described by Sambrook and Russel [4]. Restriction endonucleases and other modifying enzymes were purchased from New England Biolabs (USA) and used as recommended by the manufacturer. For all PCR amplifications, the Phusion High-Fidelity DNA Polymerase (Finnzymes, Finland) was used according to the manufacturer's instructions to achieve high accuracy of the products.

Genomic DNA was isolated from $2 \times 10^7$ CHO cells using the DNeasy tissue kit (Qiagen, Germany) according to the manufacturer's instructions. The purified DNA was fragmented using restriction endonucleases *Stu* I, *Ssp* I, *Msc* I and *Sca* I and mechanical shearing using a nebulizer (Invitrogen, USA) to ensure maximum heterogeneity of the library.

### Library Construction and PCR Amplifications

The obtained genomic DNA fragments were inserted via blunt-end coning into the pMACS 4.1 vector (Miltenyi Biotec, Germany) containing the ampicillin resistance gene. In order to avoid re-ligation without an insert, PCR amplified vector copies were used, which drastically reduced background. After transformation of the ligation mixture into electrocompetent *E. coli* (MegaX DH10B, Invitrogen, USA), colonies on LB/ampicillin agar plates were rinsed with 10 ml of LB/ampicillin medium and cultured in 50 ml LB/ampicillin medium over night. Subsequent plasmid preparation, using the PureYield Plasmid MidiPrep System (Promega, USA), resulted in a genomic DNA library.

Nested PCR reactions were performed using 6 ng of the constructed library per 50 μl reaction volume as template with primer pairs binding on different genes and the vector sequence, whereas an adequate amount of PCR product of the first round of amplification was directly used as template for a second amplification run without purification.

To generate templates for Inverse PCR, genomic CHO DNA was digested using a single restriction endonuclease for each preparation. The term "inverse" refers to a pair of primers that extends divergently from the known to the unknown region. Different restriction endonucleases (*Apa*L I, *Apo* I, *Bam*H I, *Bgl* II, *Bsa*H I, *Bsa*W I, *Bsp*H I, *Bsr*F I, *Eae* I, *Eco*R I, *Hae* II, *Hind* III, *Kpn* I, *Nhe* I, *Pci* I, *Sac* I, *Sph* I, *Xba* I, *Xho* I) have been used. The obtained genomic DNA fragments were purified using NucleoSpin Extrakt II (Macherey–Nagel, Germany) and self-ligated in order to generate a pool of circular DNA fragments followed by isopropanol precipitation. Two sets of Inverse PCR primers oriented in reverse direction of the usual orientation were designed on the obtained DNA sequence of the fragments derived from Library PCR in order to perform a nested PCR. The first PCR reaction was conducted using 200 ng of the self-ligated template per 50 μl reaction volume. The resulting PCR products were separated via agarose gel electrophoresis, bands were cut out and purified using NucleoSpin Extrakt II (Macherey–Nagel, Germany).

### Fragment Preparation and Reporter Vector Cloning

Potential transcription factor binding sites like TATA boxes or Sp1 binding sites located on the DNA fragments obtained by Inverse PCR were identified using the online transcription factor prediction program Consite (http://asp.ii.uib.no:8090/cgi-bin/CONSITE/consite/) or the online promoter prediction program Neural Network Promoter Prediction (http://www.fruitfly.org/seq_tools/promoter.html). Based on these data, fragments of different length were generated in order to identify motifs essential for promoter activity. Therefore, several primers were designed based on the sequenced fragments derived from Inverse PCR to generate different constructs of various length directly from chromosomal DNA by PCR amplification using 200 ng of genomic CHO DNA as template per 50 μl reaction volume.

Products from PCR libraries and amplified fragments were separated via agarose gel electrophoresis and cut out bands were purified using NucleoSpin Extrakt II (Macherey–Nagel, Germany). The obtained fragments were cloned into the multiple cloning region of the pGL3-Basic reporter vector (Promega, USA) containing the firefly (*Photinus pyralis*) luciferase gene.

### CHO Cell Line Cultures and Transfections

The CHO dhfr⁻ cell line was purchased from the American Type Culture Collection (ATCC, USA). These cells were cultured in Dulbecco's modified eagle medium (DMEM) and Ham's F-12 medium mixed in a 1:1 ratio (Invitrogen, USA) supplemented with 4 mM L-glutamine, 0.1% (w/v) Pluronic, 0.25% (w/v) soya peptone,

13.60 mg/l of hypoxanthine and 3.88 mg/l of thymidine, at 37°C and 7% $CO_2$.

$4 \times 10^6$ CHO dhfr$^-$ cells were transfected with 10 µg of the cloned reporter vectors and 1 µg of the pRL-SV40 vector (Promega, USA) containing the *Renilla* luciferase gene using the Amaxa Cell Line Nucleofector Kit V (Lonza, Switzerland) electroporation system. The pGL3-Basic vector (Promega, USA) containing no promoter served as negative control whereas the pGL3-Promoter vector (Promega, USA) containing the SV40 promoter was used as positive control. The appropriate amount of plasmid DNA and pRL-SV40 vector was diluted into 100 µl transfection medium and processed according to the manufacturer's instruction.

Reporter Assay

For evaluation of promoter activity of obtained CHO DNA fragments the Dual-Glo Luciferase Assay System (Promega, USA) was used. To measure firefly luciferase activity, 50 µl of Dual-Glo luciferase reagent was added to 50 µl of the cell suspension 48 h post transfection. After incubation for 10 min, bioluminescence was measured using the Synergy 2 Multi-Mode Microplate Reader (Biotek, USA). 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.

## Results and Discussion

Genomic CHO DNA Library

The construction of a genomic library requires the generation of a random pool of genomic DNA. For this purpose, many different methods are described in the literature including mechanical techniques like passage through a large gauge needle of a syringe [5–7], nebulization [8], sonication [9], stirring in a blender [10], or enzymatical treatments like digestion by restriction endonucleases [4] or DNase I [11]. For this experiment, the chromosomal CHO DNA was sheared by nebulization as well as cut by single digestion using the blunt-end generating restriction endonucleases *Msc* I, *Sca* I, *Ssp* I, and *Stu* I in order to ensure maximum heterogeneity. These endonucleases were chosen from a panel of different restriction endonucleases tested with regard to high efficiency and absence of methylation sensitivity. Resulting fragments were blunt-end cloned into a promoterless vector resulting in a genomic CHO DNA library. PCR amplification of the
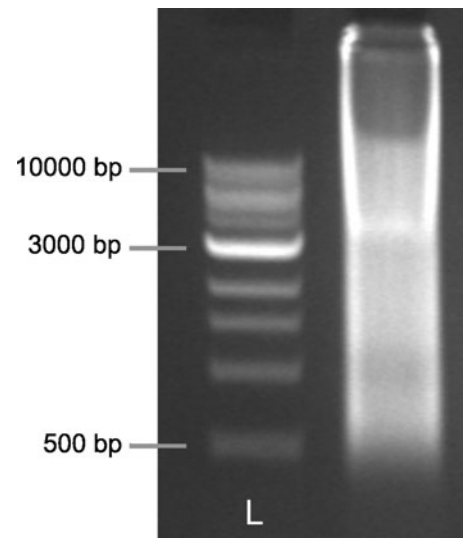


**Fig. 1** Agarose gel electrophoresis, size distribution of the genomic CHO DNA library determined by PCR amplification of the library pool; *Lane L*: 1 kb DNA Ladder (New England Biolabs, USA)

pooled genomic DNA library showed an even size distribution over a wide range (Fig. 1).

Identification of 5′ Flanking Regions of Known CHO Genes

Transcription regulatory sequences are generally located upstream of the transcription start site [12, 13]. Our intension was the identification of the 5′ flanking region of highly abundant CHO genes like those coding for ribosomal proteins. Therefore, known cDNA sequences derived from the Consortium for Chinese Hamster Ovary Cell Genomics [14] of highly expressed CHO genes were aligned against the fully annotated genome of the house mouse (*Mus musculus*) in order to identify exon 1 of the respective genes. Two primers specifically binding to exon 1 were designed to perform a nested PCR. Additionally, two other sets of primers were constructed, annealing to the backbone of the vector upstream and downstream of the inserted CHO DNA sequence, in order to achieve amplification of the specific genomic fragment, which was inserted in random orientation into the plasmid via blunt-end cloning (Fig. 2). Applying this technique, several 5′ flanking regions of known genes could be obtained by amplification via nested PCR (Fig. 3). Obtained fragments were cloned into the reporter vector pGL3 Basic upstream of the firefly luciferase gene in order to analyze putative promoter activity of these constructs.

In order to further characterize identified sequences and to potentially increase promoter or enhancer activity, the flanking regions have been investigated by Inverse PCR, a
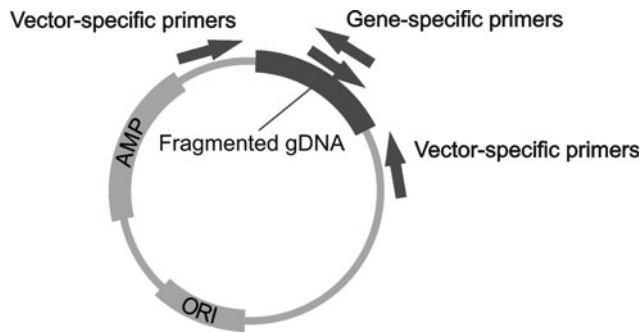
**Fig. 2** Primer design for Library PCR; *AMP* ampicillin resistance gene, *ORI* origin of replication for propagation in *E. coli*

method that enables the rapid in vitro amplification of unknown DNA sequences that flank a region of a known sequence. This technique uses the common PCR, but has the primers oriented in the reverse direction of the usual orientation. The template for Inverse PCR is a restriction fragment that has been self-ligated to form a circular DNA molecule (Fig. 4) [15]. Applying this technique, 5′ and 3′ flanking regions up to 3,500 bp of the three fragments
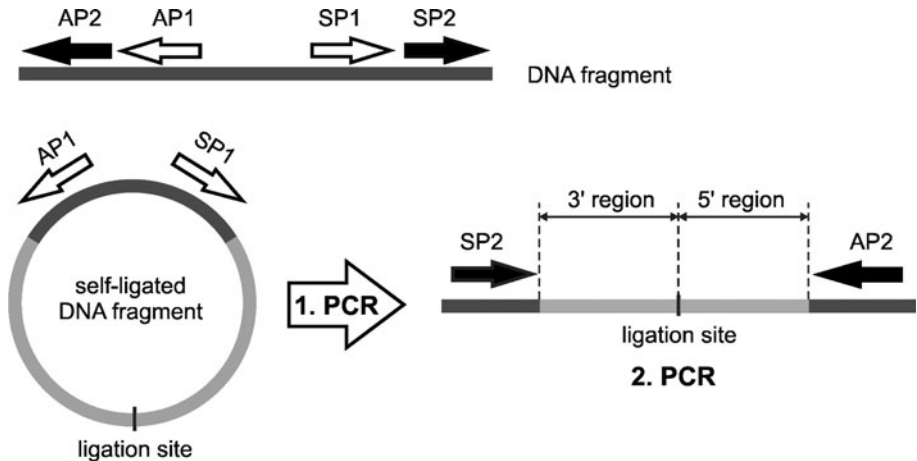
showing promoter activity could be identified (Fig. 5). These fragments were fully sequenced and the self-ligation sites (Gene A: *Eae* I-site, Gene B: *Bgl* II-site and Gene C: *Eco*R I-site) could be rediscovered and thus enabled the exact identification of the 5′ and 3′ flanking region of the initial sequences. Several PCR amplifications were performed directly from genomic DNA in order to get fragments of various length using primers specific to the newly discovered flanking regions containing the restriction sites *Nhe* I and *Xma* I for direct cloning into the reporter vector's multiple cloning region. All plasmids were transfected into CHO cells and after 48 h the resulting bioluminescence was measured. Promoter activity was determined as percentage of the value for the SV40 promoter which was used as positive control. Values were further normalized to *Renilla* luciferase measurements, expressed from a co-transfected plasmid, to account for transfection efficiency and cell number variabilities.

Figure 6 shows the map of all different constructs and Figure 7 corresponding promoter activity as percentage of the value for the SV40 promoter. While the original

**Fig. 3** Agarose gel electrophoresis, Identified 5′ flanking regions of three genes by PCR amplification using a genomic CHO DNA library as template; *Lane L1*: 2-log DNA Ladder (New England Biolabs, USA); *Lane L2*: FastRuler DNA Ladder Low Range (Fermentas, Canada)
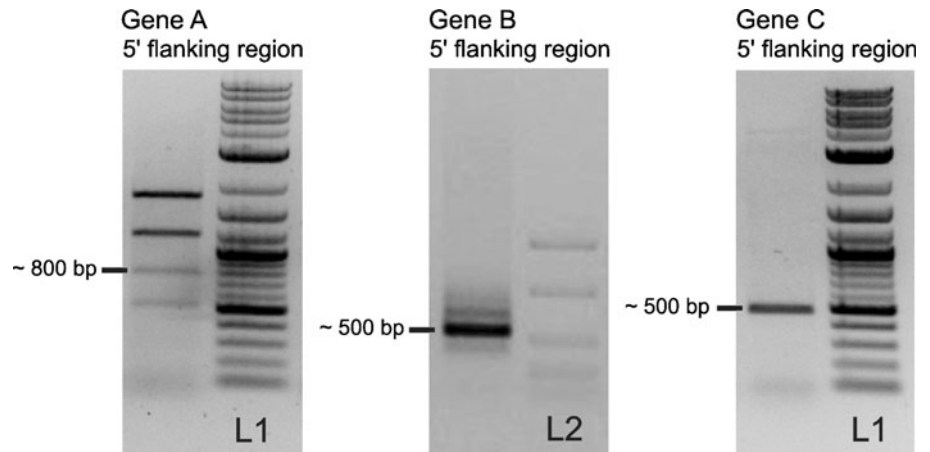


**Fig. 4** Inverse PCR schema; *SP1* sense primer 1, *AP1* antisense primer 1, *SP2* sense primer 2, *AP2* antisense primer 2; First PCR was performed using the self-ligated CHO DNA fragments as template and the primers AP1 and SP1. The second PCR using the product of the first round of PCR amplification as template and SP2 and AP2 as primers led to a fragment containing the 3′ and 5′ flanking region of the initial DNA fragment

**Fig. 5** Agarose gel electrophoresis, Identified 5′ and 3′ flanking regions of the three fragments showing promoter activity by Inverse PCR; *Lane L*: 2-log DNA Ladder (New England Biolabs, USA); Remaining lanes: various restriction endonucleases used for generating self-ligated templates for Inverse PCR led to fragments of different sizes
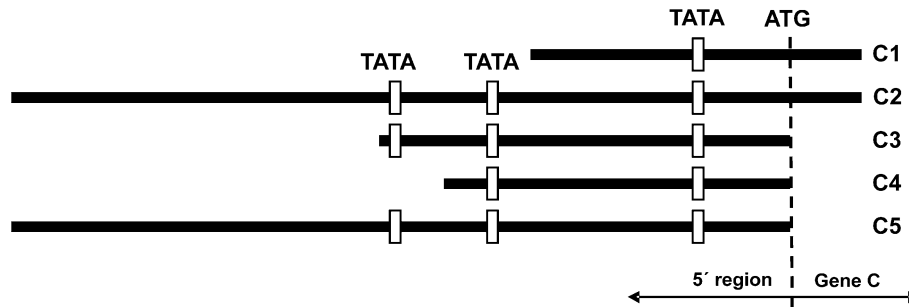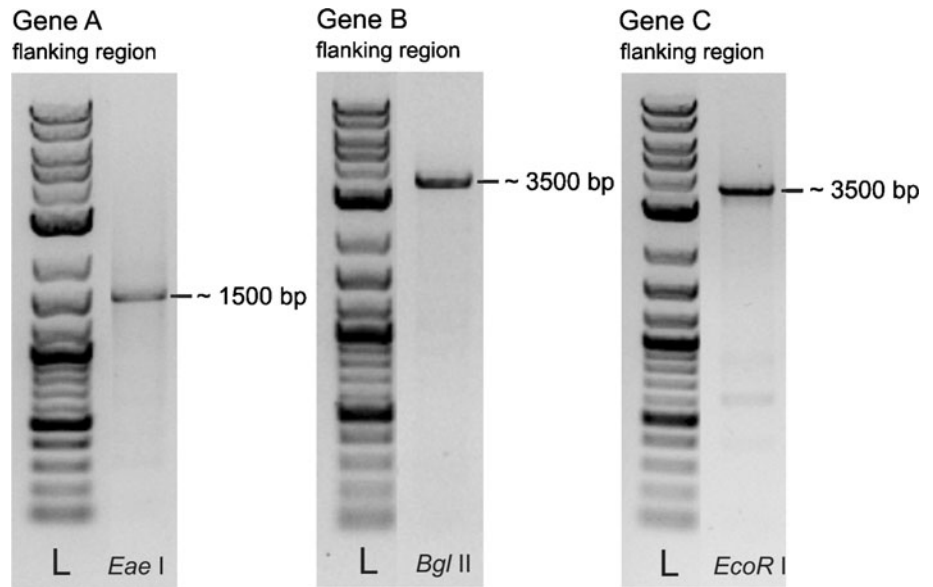
**Fig. 6** 5′ noncoding region of Gene C: schematic representation of the original and truncated fragments used in transfection experiments, TATA boxes predicted by the online promoter prediction program

Neural Network Promoter Prediction (http://www. fruitfly.org/seq_tools/promoter.html) and the start codon ATG

**Fig. 7** Reporter activity assay; Constructs transfected in Chinese hamster ovary cells and their promoter activity as percentage of the value for the SV40 promoter; (−): negative control, promoterless luciferase reporter vector pGL3-Basic; A: 5′ region of Gene A; B: 5′ region of Gene B; (+): positive control, firefly luciferase reporter vector pGL3-Promoter containing the SV40 promoter; C1–C5: constructs of various length of Gene C 5′region
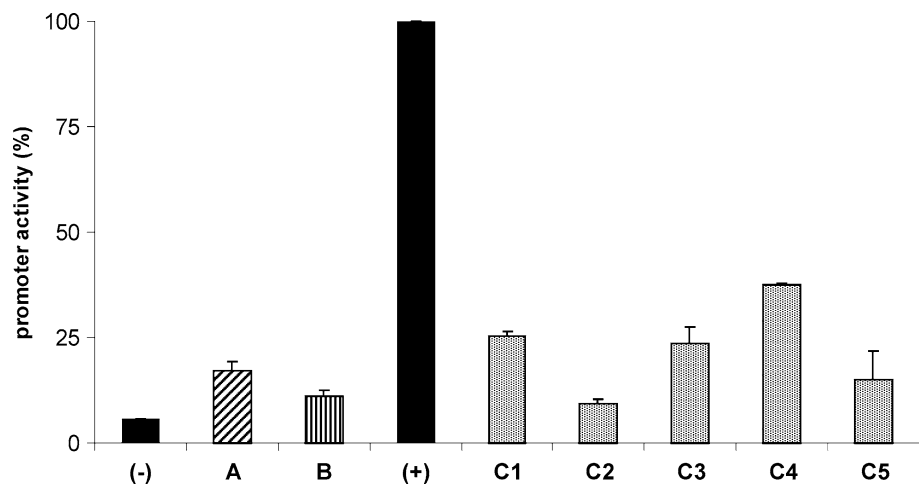
construct C1 derived via PCR amplification from the genomic DNA library shows 25% of the activity of the SV40 promoter, for construct C4 derived through Inverse

PCR nearly 40% of the SV40 control could be detected. Compared to construct C1, which contains only one predicted TATA box, C4 comprises two predicted TATA

boxes and has a shortened 3′-end starting from the start codon ATG. Constructs C2, C3, and C5, which include a third predicted TATA box (and a prolonged 3′-end in case of construct C2) did not show an improved luciferase activity in the reporter system.

These constructs have to be further characterized concerning their function in stable cell clones expressing an industrially relevant product. Another attractive goal could be to screen for constructs encoding inducible promoter sequences suitable for the expression of, e.g., cell toxic products. Such conditionally inducible gene expression can be particularly advantageous in order to optimize cell growth and achieve high cell densities during the initial phase. In a second production phase the promoter for the target protein can be activated by a specific stimulus to increase and prolong transgene expression. This kind of integrated process optimization allows adjusting gene expression characteristics and process conditions.

Overall, this method offers a very comfortable tool for the discovery of novel endogenous promoters even though only regulatory elements of known gene sequences can be identified using this approach. However, it is possible to directly target only highly expressed genes in order to increase the chance to find strong regulatory elements.

## References

1. Werner, R., Noé, W., Kopp, K., & Schlüter, M. (1998). Appropriate mammalian expression systems for biopharmaceuticals. *Arzneimittelforschung, 48*, 870–880.
2. Spenger, A., Ernst, W., Condreay, J., Kost, T., & Grabherr, R. (2004). Influence of promoter choice and trichostatin A treatment on expression of baculovirus delivered genes in mammalian cells. *Protein Expression and Purification, 38*, 17–23.
3. Dale, L. (2006). Mammalian expression cassette engineering for high-level protein production. *BioProcess International, 4*, 14–22.
4. Sambrook, J., & Russel, D. W. (2001). *Molecular cloning, a laboratory manual* (3rd ed.). Cold Spring Habor, NY: Cold Spring Harbor Laboratory Press.
5. Davison, P. (1959). The effect of hydrodynamic shears on the deoxyribonucleic acid from T2 and T4 bacteriophages. *Proceedings of the National Academy of Sciences USA, 45*, 1560–1568.
6. Davison, P. (1960). Sedimentation of deoxyribonucleic acid isolated under low hydrodynamic shear. *Nature, 185*, 918–920.
7. Schriefer, L., Gebauer, B., Qui, L., Waterston, R., & Wilson, R. (1990). Low pressure DNA shearing: A method for random DNA sequence analysis. *Nucleic Acids Research, 18*, 7455–7456.
8. Okpodu, C., Robertson, D., Boss, W., Togasaki, R., & Surzycki, S. (1994). Rapid isolation of nuclei from carrot suspension culture cells using a BioNebulizer. *Biotechniques, 16*, 154–159.
9. Deininger, P. (1983). Random subcloning of sonicated DNA: Application to shotgun DNA sequence analysis. *Analytical Biochemistry, 129*, 216–223.
10. Hershey, A. D. (1960). Molecular homogeneity of the deoxyribonucleic acid of phage T2. *BioProcess International, 2*, 143–152.
11. Anderson, S. (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Research, 9*, 3015–3027.
12. Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., & Hume, D. (2007). Mammalian RNA polymerase II core promoters: Insights from genome-wide studies. *Nature Reviews Genetics, 8*, 424–436.
13. Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature, 424*, 147–151.
14. Consortium for Chinese Hamster Ovary Cell Genomics (2009) http://hugroup.cems.umn.edu/CHO/cho_index.html.
15. Ochman, H., Gerber, A., & Hartl, D. (1988). Genetic applications of an inverse polymerase chain reaction. *Genetics, 120*, 621–623.