

# Use of MAR Elements to Increase the Production of Recombinant Proteins

Cori Gorman, Salina Arope, Mélanie Grandjean,  
Pierre-Alain Girod, and Nicolas Mermod

**Abstract** The biopharmaceutical industry continues to face the challenge of producing large amount of recombinant proteins for use as therapeutics, and eighty percent of protein therapeutics in clinical development are produced in mammalian cell systems. Approaches to increase production addressing growth conditions, such as the improvement of media composition and process control, or transcription of the recombinant gene via the use of strong promoters/enhancers and amplification of gene copy number, have increased the yields obtained from mammalian cells considerably over the past decades. However these processes remain laborious, and extensive screening of clones is often required, as stable cell line and/or protein production is not always obtained. Unstable or variable expression is linked to the location of transgene integration site, the regulation of gene expression, the silencing of genes, and the loss of gene copies. Genetic elements that may remodel chromatin to maintain the transgene in an active configuration are now being employed increasingly to improve protein production using mammalian cells. Here we will review how one type of such elements, the MARs, may increase transgene integration into the cell genome and decrease silencing effects to reduce expression variability. We also illustrate how inclusion of these elements in expression vectors leads to increased specific productivities ranging from 20 to 100 picograms per cell and per day (p/c/d), resulting in protein titers above 5 g/l.

---

C. Gorman and P.-A. Girod  
Selexis SA, 18 chemin des Aulx, 1228 Plan-les-Ouates, Switzerland

C. Gorman  
DNA Gateway International, Inc. 55 New Montgomery St. Ste 605  
San Francisco CA, 94105, USA

S. Arope, M. Grandjean, and N. Mermod (✉)  
Laboratory of Molecular Biotechnology, University of Lausanne, 1015, Lausanne, Switzerland  
e.mail: nicolas.mermod@unil.ch

## 1 Introduction to Epigenetic Silencing Issues in the Generation of Production Cell Lines

During the past two decades, cultured mammalian cells have become the most widely used platform for producing recombinant therapeutic proteins. Improving yield and stability of protein expression are, therefore, of considerable value to the industry. Previously, such improvements have mainly originated from optimizing downstream production processes and media development. Despite recent advances in the field of cell line generation, expression levels in mammalian cells are relatively low and often unstable over time; these events result in high development and production costs for therapeutic proteins. The average reported yields in mammalian cells (usually 0.5–2 g/l) are several fold lower than the yields from bacteria and yeast systems, while even the highest reproducible yields, 5 g/l, remain at least three fold lower than obtained in more simple production systems (Kwaks and Otte, 2006).

Despite significant effort, the current schemes for cell line development remain, to a large extent, empirical. There is a considerable degree of variability, and our understanding of the sources of variability in the mammalian cell line development process remains limited. This process is laborious, and extensive screening of clones, often spanning over several months, is still widely practiced in industry. Obtaining cell lines that maintain stable protein production is of utmost importance, particularly for industrial use where the goal is to commercialize the protein being produced. Problems with stability can impact the time and effort required to generate working and master cell banks. The loss of productivity between the initial cell isolates and the end-of-production cells can compromise regulatory approval and, in the worst-case scenario, may result in rejection of a particular cell line after months of development efforts. Unfortunately, the method of selecting cell lines relies heavily on a degree of chance. In part, this high degree of variability is due to the effect that random transgene integration into host cell chromosome exerts on transgene transcription often resulting in silencing. Another cause of variability results from the integration of a varying number of transgene copies from one clone to the next. Finally, a last source of variability is variegation, a phenomenon that results in the cycling of cells between productive and non-productive phases, which again may affect differently distinct cell clones.

The insertion of genes into certain areas of chromatin can lead to the so-called “position effect” (Wilson et al., 1990). Silencing of transfected genes in mammalian cells is a fundamental problem that probably involves the relatively inaccessible status of the DNA when it is imbedded in chromatin. Transgene integration, when it occurs through a random process, can either occur in highly condensed, silenced region of the chromatin, heterochromatin, or in more open and active chromatin, euchromatin (Eissenberg, 1989; Eissenberg et al., 1992; Zahn-Zabal et al., 2001). Integration into heterochromatin may result in minimal or no transgene expression. Because a large proportion of the genome is in the form of heterochromatin, the chance that a transgene integrates in, or close to, heterochromatin, and consequently is silenced or repressed, is high.

Other regions of the chromosome may be subjected to slow silencing effects (Pankiewicz et al., 2005). Because this slow silencing may not be readily apparent,

it can lead to a gradual loss of productivity that suddenly appears after an initial phase of seemingly stable expression during clone isolation and characterization (see chapter by Alan Dickson in this book). A related process, known as position effect variegation (PEV), is thought to result from the stochastic spread or retreat of heterochromatin towards or away from the gene location (Eissenberg et al., 1992; Volfson et al., 2006). PEV typically leads to clones that possess heterogeneous levels of expression when comparing distinct cells in the monoclonal population. Often, this heterogeneous expression is not apparent when determining the titer of secreted proteins such as immunoglobulins in the cell supernatant, but nevertheless may limit the yield. The combined effects of random transgene incorporation with these chromatin-mediated epigenetic effects collectively result in only a small percentage (less than 1%) of the initially isolated cells being capable of producing high amounts of the desired protein (Girod and Mermod, 2003). Therefore, lengthy selection and screening procedures are often required to select and identify those cells with the proper growth, transgene expression, long-term stability, and protein quality properties required for large-scale production.

For several years, Barnes et al. (2001, 2003, 2004, 2006, 2007; Barnes and Dickson, 2006) have examined the process of loss of protein production in mammalian cell lines. Despite repeated rounds of cloning, the cell lines they derived showed a wide variation in terms of maximum obtainable cell densities, rates of growth, and accumulation of secreted recombinant protein (Barnes et al., 2001). Several lines of data suggest that rapid phenotypic drift may be occurring during culture and, therefore, the cells derived from a single cell, as a result of cloning, soon diverge to become a mixed population. In this context, the term 'stable cell lines' refers to cell populations that retain stability of expression during prolonged culture (Barnes et al., 2001).

Though a variety of mammalian cells have been used for recombinant protein production, including mouse myeloma derived (NS0), human embryonic kidney (HEK-293), baby hamster kidney (BHK) and more recently, the human retina derived (PerC6) cells, the most commonly used host cell lines remain the CHO cells. The popularity of CHO based expression is largely due to the ability to use DNA amplification techniques in these cells to increase transgene copy number. However, both the DHFR and GS methods of amplifying transgenes can result in genetic instability. CHO cells typically undergo genomic rearrangements and amplifications of the locus of DNA integration resulting in increased copy numbers for both DHFR and the protein of interest. Often, clones containing several hundred copies of the vector construct can be found following amplification. This high copy number does not however lead to uniform high expression or to stable production. It is commonly reported that recombinant protein production can drop significantly within the 2 months following high-producing clone selection, particularly when the selection pressure is removed (Fann et al., 2000; Jun et al., 2006; Kim et al., 1998a; Strutzenberger et al., 1999). For instance, Strutzenberger et al. (1999) have shown that when dhfr amplification with methotrexate was used, over 75% of the integrated transgenes were lost once the drug was removed. In the absence of selective pressure, expression is lost just weeks following selection. In one study, the relative decrease in specific productivity varied among subclones, ranging from 30% to 80% (Kim et al., 1998b; Kim et al.,

2001). Southern and Northern blot analyses showed that this decreased productivity resulted mainly from the loss of amplified antibody gene copies and their respective cytoplasmic mRNAs (Barnes et al., 2006).

Overall, it is clear from the examples presented above that there is a commercial need within the biotechnology industry to understand the problem of instability of protein production associated with recombinant mammalian cell lines. Loss of recombinant gene copy number and the overgrowth of non-producing populations of cells may result in low production; however, there are several other factors that may affect expression levels and stability of production. The possibility of improving cell-line stability and decreasing the variability associated with the generation of a cell line is now possible through the use of techniques that assure that the transgenes are actively transcribed following integration and that the chromatin surrounding the transgene remains capable of active transcription. The potential of these techniques for saving time as well as human and financial resources is extremely promising. This will also open up avenues for more rapid and effective use of additional types of mammalian cells, beyond CHO cells, as expression hosts (Barnes et al., 2004, 2007).

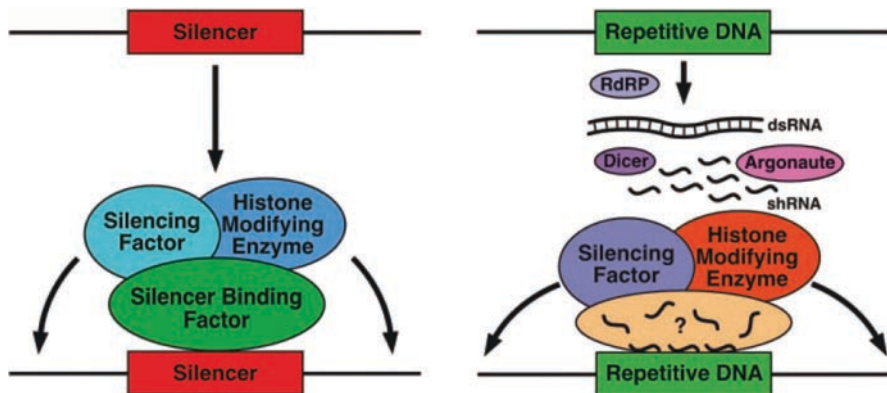
## 2 Causes of Instability During Mammalian Cell Production

The consistency of growth, productivity, or product characteristics with each successive generation of the cell line defines cell line stability and these factors contribute to the overall process consistency. Some of the issues that lead to instability of protein production in mammalian cells include gene amplification, loss of genetic material, methylation and the location of integration. Gene amplification occurs through the mechanism of chromosomal rearrangement, which involves chromosomal breaks (Andrulis et al., 1983; Melton et al., 1982). Amplification can result in decreased stability of transgene expression due to such breaks and rearrangements (Flintoff et al., 1984; Yoshikawa et al., 2000). CHO cells are known to have an unstable karyotype, with chromosome rearrangements arising from translocations and recombination events, as in the amplification procedures (Yoshikawa et al., 2000). As discussed above, the predominant use of CHO cells has been paralleled with gene amplification selection methods; however, loss of protein production following amplification has been reported for several proteins including interferon, tissue plasminogen activator, and antibodies. In some cases, production levels can reach a stable value after an initial decrease during the first 30 to 50 days of culture (Kim et al., 1998a). Even in the presence of selective pressure, Jun et al. (2006) found that the stability of antibody producing subclones was very poor. Furthermore, the specific secretion rate decreased by 50% after 100 passages even with selective pressure. This might be explained by the selective silencing of the transgenes at chromosomal sites that are prone to epigenetic regulation (position effect), as the amplified gene arrays are often scattered at multiple loci in the host genome. Thus instability may be a concern in the development of CHO cell

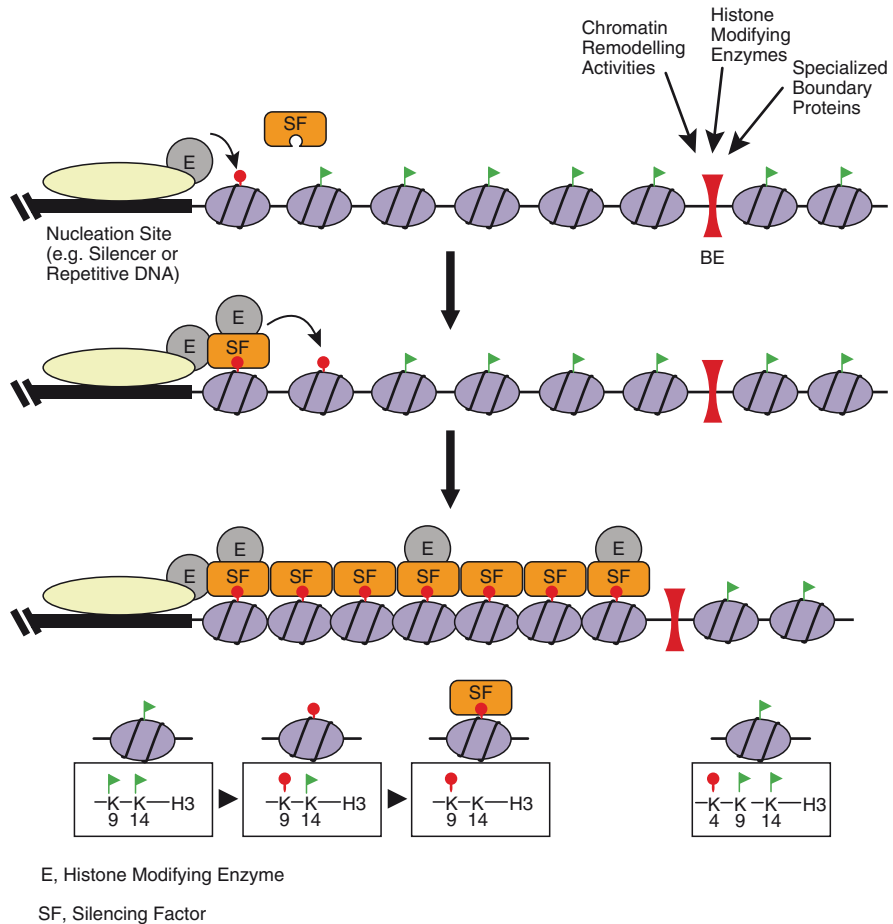
lines with DHFR and GS-mediated gene amplification (Kim et al., 1998a; Kim and Lee, 1999; Fann et al., 2000; Jun et al., 2006).

In addition to the potential loss of transgenes following amplification, the large number of repetitive gene sequences that results from this process may induce methylation of DNA sequences, thus preventing transcription (Fouremana et al., 1998). More recently, RNAi-mediated chromatin remodeling linked to the occurrence of repetitive sequences, such as the hundreds of copies that result from amplification, has been shown to contribute to gene silencing (Almeida and Robin, 2005; Morris, 2008). These observations provide a likely explanation for the fact that expression following amplification does not increase proportionally with the transgene copy number. Mechanisms by which repeated sequences such as inverse repeat transgene arrays and RNAi may trigger silent chromatin assembly include physical pairing of homologous sequences and/or DNA–RNA or RNA–RNA interactions (Selker, 1999; Matzke et al., 2001). The connection between RNAi and heterochromatin assembly has suggested a model for the RNA-mediated epigenetic structuring of the eukaryotic genomes. Double-stranded RNA is processed into small RNAs, which in turn provide specificity for targeting histone-modifying activities and epigenetic modification of the genome through homology recognition (Fig. 1; Grewal and Moazed, 2003).

Instability of expression can be due to the regulation of gene expression, the silencing of genes, and the loss of gene copies. However, it must be stressed that these mechanisms are not mutually exclusive, and often the regulation of gene expression and the occurrence of instability of expression involve interplay between different mechanisms. DNA is highly condensed into the chromatin structures, and this condensation often hinders the accessibility of DNA to transcription complexes (Felsenfeld, 1992, 1996; Woodcock and Dimitrov, 2001). Activation of transcription



**Fig. 1** Mechanisms for the initiation of heterochromatin. Heterochromatic structures can be nucleated by specific cis-acting sequences, called silencers, which are recognized by DNA binding proteins (*left*). Transcripts generated by repetitive DNA are processed into siRNAs by a mechanism requiring components of the RNAi machinery (from Grewal and Moazed, 2003, reprinted with permission from AAAS)



**Fig. 2** Model for formation of silenced chromatin domains. After the recruitment to a specific heterochromatin nucleation site by proteins that directly bind DNA or are targeted by way of RNAs, histone modifying enzymes (E) such as deacetylases and methyltransferases modify histone tails to create a binding site for silencing factors (SF). Spreading of silencing complexes is blocked by the presence of boundary elements (BE). The modifications associated with the amino terminus of histone H3 in fission yeast heterochromatin (*bottom left*) and euchromatin (*bottom right*) are illustrated as an example (from Grewal and Moazed, 2003, reprinted with permission from AAAS)

requires the rearrangement of chromatin structure or chromatin remodeling. Chromatin remodeling, which is performed by a range of remodeling complexes, is a loose term used to define any event that alters the nuclease sensitivity of a DNA region (West et al., 2002; Grewal and Moazed, 2003) (Fig. 2).

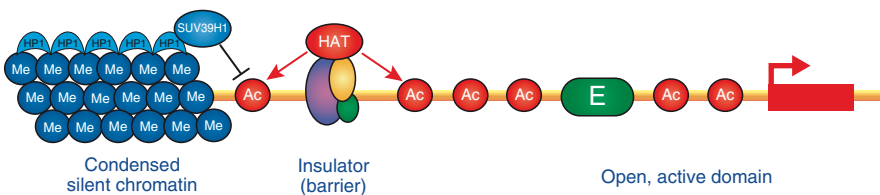
The highly condensed heterochromatin domains are interspersed along with relatively decondensed euchromatic regions (Fig. 2; Grewal and Moazed, 2003). Given that heterochromatin structures, once nucleated, can spread in *cis*, resulting

in epigenetic silencing of adjacent genes, cells have evolved antagonistic mechanisms that protect active regions from the repressive effects of nearby heterochromatin.

Chromatin and proteins important for the control of transcription can undergo a variety of modifications, such as methylation, acetylation, phosphorylation, and ubiquitinylation. It has been reported that DNA methylation of transfected DNA can play a major role in the regulation of expression. Reports have suggested that methylation causes the repression of gene expression and hypomethylated DNA around the promoter region of genes is often associated with elevated transcriptional activity. Acetylation is also an important step in transcriptional control. As a general rule, transcriptionally active genes usually exhibit acetylation, whereas transcriptionally inactive genes do not. Finally, ubiquitination has been suggested to lead to transcriptionally active DNA by disrupting higher order chromatin structures, hindering internucleosomal interactions, and/or by disrupting the association of linker histones with nucleosomes (Esteller, 2008; Feinberg, 2008).

### 3 Use of MAR Elements to Boost and Stabilize Expression

Typically, the stability of recombinant cell lines is determined by monitoring cell growth and protein production for several months. For some cell lines, however, protein productivity diminishes over time, usually as a result of changes in the regulation of transgene expression (Strutzenberger et al., 1999). Regulation of higher order chromatin structure is directly coupled with regulation of the expression and integrity of the genetic information of eukaryotes and is likely to be a major force in the origin and evolution of genes, chromosomes, genomes, and organisms. Some of these problems are caused by gene silencing at the level of chromatin – so-called epigenetic gene silencing. Specialized DNA elements known as boundary elements have been shown to mark the borders between adjacent chromatin domains and to serve as barriers against the effects of silencers and enhancers from the neighboring regions (West et al., 2002; Labrador and Corces, 2002; Fig. 3). Perhaps boundaries



**Fig. 3** A model for the barrier activity of insulators. A schematic diagram based on the example of the upstream boundary of the chicken beta-*globin* locus. Insulator proteins constitutively recruit histone acetyltransferases that acetylate flanking nucleosomes (*red spheres*). Acetylation serves to inhibit histone modifications required for the propagation of transcriptionally silent condensed chromatin (*packed bluespheres*). Barriers act to terminate the chain of repressive chromatin by competing in the histone-modification process (West et al., 2002)

delimit structural domains by interacting with each other or with some other nuclear structure (Labrador and Corces, 2002). When these human genetic elements are included in the expression vectors, the chromatin structure flanking the transgene is maintained in an active configuration. Here, we describe the inclusion of a specific type of DNA element that has been used during the past years to interfere with epigenetic gene silencing, with the aim of enhancing and stabilizing transgene expression.

One method to overcome positional dependent inactivation is the use of vectors that include a matrix or scaffold attachment region (MAR or SAR) that repress silencing. The anti-silencing effect observed in the presence of MAR may be mediated by chromatin modifications such as histone hyperacetylation at the site of chromosomal transgene integration locus (Recillas-Targa et al., 2002; Yasui et al., 2002) or changes in a specific subnuclear localization (Bode et al., 2003; Hart and Laemml, 1998). In addition, the general increase in transgene expression can be explained in several ways. For example, transcription of transgenes can be improved, either directly or indirectly, by an activation of the transgene promoter or enhancer by MARs. MARs may also favor integration in a permissive locus within the chromosome, or they may increase the number of integrated transgene copies.

The MAR element associated with the lysozyme gene in chicken is one of the most studied elements (Zahn-Zabal et al., 2001; Girod and Mermoud, 2003; Girod et al., 2005). The chicken lysozyme locus contains a 3 kb regulatory region known as the A element. This element was originally used as a MAR in a series of experiments on the effect of MARs on gene expression (Stief et al., 1989). These experiments were exciting because transgenes flanked by the A element exhibited expression that was proportional to gene copy number ('copy number-dependent'), suggesting that the element had been able to insulate transgene expression from gene silencing or position effects. The intact element has been shown to contain both enhancer and matrix-binding activities. When the intact element was divided into 1.32 and 1.45 kb pieces, both were able to confer copy number-dependent transgene expression. However, when smaller fragments were tested, the portion of the A element that bound to the nuclear matrix no longer conferred copy number dependence (Phi-Van and Strätling, 1996), and the possibility must be considered that at least some of the original effects were attributable to the enhancer portion of the element rather than the matrix-binding portion (Allen et al., 1996).

In the Zahn-Zabal et al. (2001) study, the chicken lysozyme MAR was compared with other chromatin elements with respect to the ability of these elements to augment expression. Single chromatin elements, as well as combinations of elements, were tested for their capacity to increase stable transgene expression in industrially relevant CHO cells. The chicken lysozyme 5' MAR was the only element to significantly enhance reporter expression in pools of stable clones. While increased expression in pools of stable clones is indicative of an overall positive effect of the chicken lysozyme MAR on transgene expression, it does not provide information as to the probability of isolating a high producer clone. In order to address this issue, individual colonies were isolated and the level of expression of the transgene was measured. CHO cells were transfected with luciferase expression vectors containing no, one, or two MARs, and 15 individual colonies were randomly



isolated and analyzed for each construct. Consistent with the results obtained with pools of stable clones, the average expression level of the clones analyzed increases with the number of MARs present on the construct. The use of the MAR elements also increases the proportion of high-producing clones, thus reducing the number of clones that need to be screened. Thus, MARs have been used to improve the expression of transgenes, in cells cultured *in vitro* and *in vivo* (Girod et al., 2005; Gutierrez-Adan and Pintado, 2000; Zahn-Zabal et al., 2001). Furthermore, the expression level of the most productive clones was found to be higher for constructs bearing MARs; therefore, fewer clones needed to be picked and analyzed to identify a high-level production clone when MARs were present on the expression plasmid.

Other types of epigenetic regulatory elements have also been studied, such as chromatin sequences associated with the  $\beta$ -globin gene (Forrester et al., 1989). The  $\beta$ -globin gene locus control region is comprised of five DNase I hypersensitive sites (Ostermeier et al., 2003). Expression vectors containing four of the DNase I hypersensitive motifs have been shown to increase  $\beta$ -globin mRNA levels 8- to 13-fold following transfection into mouse erythroleukemia cells, while vectors containing just two motif sites increased globin expression to a lesser extent. These first  $\beta$ -globin sequences were seen to display cell-type specificity in that no effect was seen when the constructs were assayed in 3 T3 fibroblasts. More recent studies have also characterized the human  $\beta$ -globin MAR element. Kim et al. (2004) showed that the human  $\beta$ -globin MAR improves transgene expression in CHO cells. They constructed various deletion constructs with different orientations and examined their effects on the frequency of  $\beta$ -Gal positive colonies and on transgene expression levels. The enhancing effects of the human  $\beta$ -globin MAR depended on the integrity of the full-length fragment (regardless of the orientation) as all of the deletion constructs were much less active. Furthermore, there was no effect of the MAR on transient expression (Kim et al., 2004).

Two groups have studied the MAR/SAR elements associated with the human  $\beta$ -interferon gene. Klehr et al. (1991) transfected DNA corresponding to the complete chromatin domain of human  $\beta$ -interferon gene into mouse L cells. When the transgene is flanked by SARs, the gene's transcription was enhanced 20–30-fold with respect to DNAs containing only the immediate regulatory elements. To elucidate the role of SAR elements in the transcriptional enhancement, the position of the genomic element was varied relative to several artificial promoter-gene combinations. The data showed that SARs enhance general promoter functions in an orientation- and partially distance-independent manner; the effect of these elements is restricted to the integrated state of transfected templates. Similar to the results seen by Kim et al. (2004), when studying the  $\beta$ -globin MAR, the SAR elements studied by Klehr et al. (1991) were generally found to have an antagonizing effect during transient expression. Kim et al. (2005a) analyzed the frequency of positive colonies by *in situ*  $\beta$ -galactosidase staining when the human  $\beta$ -interferon SAR element is included in the vector. Two copies of the human  $\beta$ -interferon SAR element enhanced the frequency of positive colonies only by nearly 40% versus that obtained using one copy of human  $\beta$ -interferon SAR element, although the gene expression was enhanced twofold. The frequencies of positive colonies obtained from two copies of human

interferon-beta SAR element and from one copy of human  $\beta$ -globin MAR element are about the same, although the expression of  $\beta$ -galactosidase gene with two copies of human  $\beta$ -interferon SAR element was about 50% greater than with one copy of the human  $\beta$ -globin MAR element.

These data suggest that the additional copy of the human  $\beta$ -interferon SAR element at the flanking region of the  $\beta$ -galactosidase expression unit affects transgene expression more than the frequency of positive colonies. In the case of the expression of recombinant genes in CHO cells, applications of MAR/SAR elements have been reported for the chicken lysozyme MAR element (Zahn-Zabal et al., 2001) and for the human  $\beta$ -globin MAR element (Kim et al., 2004). In a previous study Kim et al. (2004) demonstrated that the human  $\beta$ -interferon SAR element is less effective than the human  $\beta$ -globin MAR element; however, in the 2005 study, the human  $\beta$ -interferon SAR element was more effective than the human  $\beta$ -globin MAR element when two flanking human  $\beta$ -interferon SAR element were used (Kim et al., 2005). Interestingly, the chicken lysozyme MAR element was most effective when the two flanking MAR elements were used, but this was not case for the human  $\beta$ -globin MAR element (Zahn-Zabal et al., 2001; Kim et al., 2004). Therefore, it appears that the enhancing effects of MAR/SAR elements on the expression of recombinant genes require their proper configurations.

#### **4 Identification of MAR DNA Sequences that Mediate Increased Expression**

Association of the MARs in the chromosomal DNA with the nuclear matrix organizes the higher order structure of the genome, forming looped structures that are likely to be equivalent to active chromatin domains in terms of transcription as well as replication. The nuclear matrix was originally described as a framework of the nucleus that remains insoluble after selective extraction of histones and DNA in the chromatin loops (Sjakste and Sjakste, 2001; Girod and Mermod, 2003). The MAR sequences are generally AT-rich at 70% and possess potential of DNA bending (Yamasaki et al., 2007). These AT rich regions of MARs, which are composed of either a tract of homopolymeric adenine (dA) or a stretch of adenine.thymine dinucleotides (dA.dT), are thought to play a significant role in MARs functions. However, it has not been clearly elucidated how these unique DNA sequences regulate MAR activities. The binding of regulatory proteins to these A + T sequence motifs as well as the structural features of the A + T rich regions, which include curved DNA configuration (Homburger, 1989), a strong potential for strand separation (Bode et al., 1992), narrow minor groove width (dictated by oligo d.A tracts) may altogether mediate the functional activity of MARs in chromatin remodeling and gene expression.

The A + T rich elements have been shown to have transcriptional activation capacity in stable transformants of both plant and animal cells (Nowak et al., 2001; Bode et al., 1992). They have regions where base pairs tend to break under an unwinding stress (base unpairing region: BUR), centered at a sequence ATATAT that are referred

to as BUR nucleation sequence. The tendency of base unpairing in the MAR DNA was shown to be essential in binding to the nuclear matrix and enhancing the promoter activity (Yamasaki et al., 2007).

The correlation between DNA curvature and transcriptional activation has been demonstrated by Ohyama and colleagues (Nishikawa et al., 2003), whereby a 36 bp left hand curved DNA segment activated transcription from the herpes simplex virus thymidine promoter (HSV tk) in transiently transfected COS-7 cells. The curved DNA segment was referred as T4, containing four tracts of 4 oligo dA (5' GTGAAAAACATGGAAAAACATGAAAAACATGAAAAAC-3'), designed to have a specific left hand rotation. The T4 left hand curved DNA is predicted to have high affinity for histone core and indeed it was shown to associate with nucleosomes. This led the authors to conclude that T4 activated transcription by forming part of the nucleosome, thus, arranging the TATA box of the promoter outwards and therefore, facilitating the initiation of transcription (Nishikawa et al., 2003). Several years later, Ohyama and workers further tested the effects of longer left hand curved DNA segment comprised of 2 to 40 tandem repeats of the T4 segment on transgene activation in COS-7 and HeLa cells (Sumida et al., 2006). All the left hand curved T4 tandem repeats activated HSV tk promoter in transient assays in COS cells. The effect of right hand curved DNA was also tested but had very little effect on promoter activity, at least in the two cell lines tested. The degree of transcriptional activation correlated with the length of the curved DNA. In particular, the T32 segment was the most effective curved DNA segment, activating HSV tk promoter 150-fold relative to control construct with straight DNA fragment.

The effect of curved DNA on transcription was also tested in the context of genomic chromatin in HeLa cells. The T20 segment was shown to activate transcription of reporter gene regardless whether the construct was integrated in intergenic or coding region of a gene. By contrast, the transgene expression was extinguished in the control HeLa cell lines in which the curved DNA was removed from the reporter gene. The results of this study were important in demonstrating that this left hand curved DNA segment minimizes silencing and increases transcription of reporter gene regardless of the locus of integration. The transcriptional activation by T20 functions perhaps in a similar mechanism to that accounted for T4 curved segment but having a more dramatic effect since the T20 segment is longer with higher density of histones. The ability of T20 to “capture” and reposition histones may facilitate the accessibility of the promoter to transcriptional machinery (Kamiya et al., 2007). Nevertheless, it is possible that other regulatory proteins, such as high mobility group (HMG)-1 non-histone chromosomal protein (Landsman and Bustin, 1993) and SATB1 (Bode et al., 1992), both bind to curved DNA structures and further change DNA conformation upon binding.

Naturally occurring A+T sequence motifs derived from the MAR at the 3' end of immunoglobulin heavy chain enhancer and 5' upstream of the human  $\beta$ -interferon have also been investigated for their ability to activate transcription. Multimerization of synthetic oligonucleotides containing an AATATATTT sequence motif derived from the two MARs mentioned above, were demonstrated to be potent in increasing SV40 promoter activity in stably transformed mouse L cells, almost comparable to

the transcriptional activation levels by the full 2.2 kb hu  $\beta$ -interferon gene MAR (Bode et al., 1992). However, whether the AATATAAT sequence motif may form a complex with nucleosomes is unknown. The transcriptional activity of the multimerized A+T sequence motif may be also related to the structural features of this DNA element and to its ability to bind to the nuclear scaffold. Indeed, the AATATAAT sequence motif was shown to have high potential for base-unpairing and nuclear matrix binding. There seems to be a correlation between the unwinding potential of this sequence and the potency of nuclear-matrix binding as well as transcriptional activation. Mutations of this DNA motif resulted in loss of unwinding property of the MAR, reduced affinity to the nuclear scaffold and loss of capability to enhance the transcriptional activity (Bode et al., 1992). Work by others have also shown that the decrease in the thermodynamic stability of MARs is correlated with enhanced strength in binding to the nuclear scaffold *in vitro* as well as in the ability to activate transcription *in vivo* (Allen et al., 1996; Schubeler et al., 1996). Specific A+T sequence motif from these two MARs may favor transcriptionally active complex by keeping the transcriptional domain in an “open” and “relaxed” conformation since MARs are thought to separate chromatin into strained loop domains. This nucleation site for DNA unwinding may also accept released histones (Clark and Felsenfeld, 1991) from the region and recruit topoisomerases (Gilmour et al., 1986) to prevent condensation of the transcriptional domain (Bode et al., 1996). However, it must be noted that the nuclear scaffold binding strength does not necessarily correlate to the potency of MAR to activate gene expression (Girod et al., 2007). A summary of the ways that the A+T sequence motif of a MAR element may modulate expression is given in Table 1.

Although there are experimental results outlining a role of A+T rich elements in several processes, we still do not understand the underlying mechanisms related to

**Table 1** Summary of the mechanisms by which the A+T rich element of MARs may exert their transcriptional activation effect

Feature of A+T Motif	Mechanism	Reference
Maintaining a local “open” chromatin domain	Due to inherent curvature of the DNA	Bode et al. (1992, 1996, 2006)
Organizing chromatin structure into loop domains	By attaching to the nuclear scaffold	Bode et al. (1992, 2006)
Facilitating the unwinding and base-unpairing of DNA sequences	Due to inherent curvature of the DNA	Bode et al. (1992, 2006)
Increasing the accessibility of TATA box and <i>cis</i> -DNA elements of the promoter	Functioning as a “trap” for histones	Girod et al. (2007); Kamiya et al. (2007)
Increasing concentration of functional proteins	Serving as target sequences for transcription factors and chromatin modifiers	Gilmour et al. (1986)
Facilitating chromatin remodeling and transactivation	Increasing the concentration of chromatin proteins in a particular region	Gilmour et al. (1986)

the transcriptional initiation and the anti-silencing effects by the A + T rich elements. Further investigations of the role of MAR-associated A + T rich elements are required and may lead to useful applications of this DNA element in “chromatin engineering” (Kamiya et al., 2007). Sequences derived from MARs or synthetically designed oligomers with A + T rich sequence motifs can be a useful tool to increase and maintain high transgene expression for research applications, recombinant protein productions and gene-based therapy. Furthermore, due to their small size, they can be practical when constructing viral or non-viral vectors where the size could be a constraint.

## **5 Identification of the MAR–Binding Proteins as Mediators of Increased Expression**

MAR/SAR activity is unlikely to arise uniquely from the intrinsic properties of its DNA motifs. Rather, the ability to protect against position effect and to regulate transcription may depend on the contribution of the protein factors that bind these motifs (Liebich et al., 2002). Transcription factors binding-sites found in the nuclear matrix are extremely diverse. This is not surprising, as MARs constitute important regulatory elements of genes, involved in DNA replication, transcription, repair, and recombination. Below, we will discuss three known MAR transcription factors: SATB1, CTCF and HMGA family of proteins.

### ***5.1 Special AT-Rich Binding Protein (SATB1)***

SATB1 binds to AT rich base-unpairing sequences (Kohwi-Shigematsu et al., 1998) (referred as ‘ATC sequence context’), where one strand consists of adenine, thymidine, cytosine but not guanine (Dickinson et al., 1992). SATB1 controls gene expression by anchoring DNA sequences to the nuclear scaffold resulting in the formation of “cage-like structures” that separates heterochromatin from euchromatin (Cai et al., 2003), in a cell-type specific manner, as it is predominately expressed in thymocytes. In addition, SATB1, serves as a docking site for recruiting chromatin remodeling proteins such as ACF, ISWI, HAT and HDAC, and these chromatin modifiers were suggested to activate or suppress gene expression through nucleosome remodeling histone acetylation or deacetylation at SATB1 bound MARs (Yasui et al., 2002; Kumar et al., 2005).

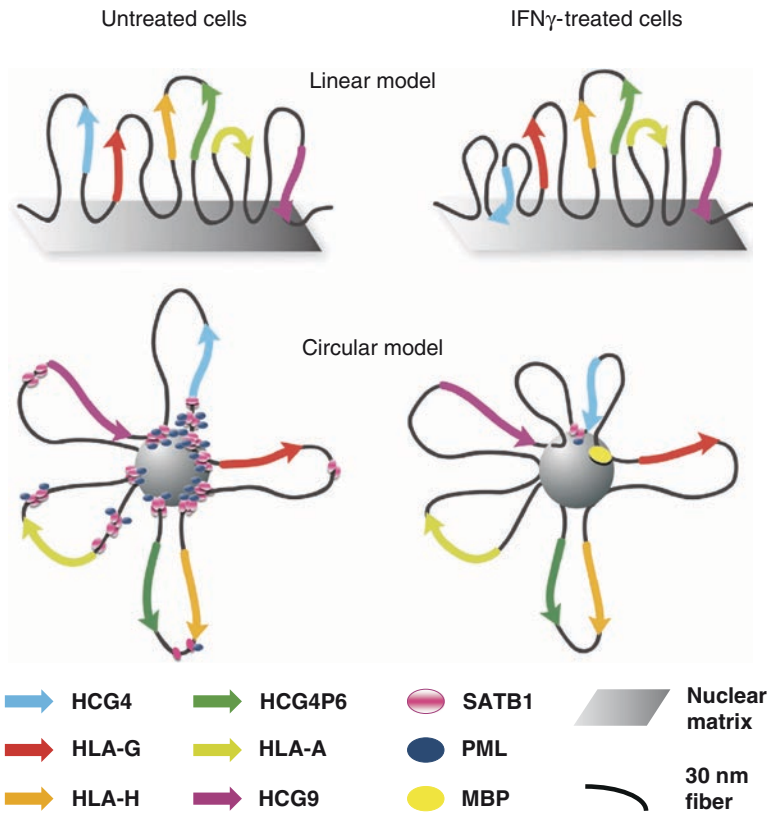
The ability of SATB1 to recruit either HAT (coactivator) or HDAC (corepressor) appears to be mediated by the phosphorylation state of SATB1 (Kumar et al., 2006). A study conducted in T cells demonstrated that the phosphorylation of SATB1 by protein kinase C (PKC) was followed by the recruitment of HDAC1 to the IL-2 promoter, resulting in repression of IL-2 transcription. Dephosphorylation of

SATB1 exerted the opposing effects, whereby; the interaction of SATB1 with PCAF causes the derepression of the IL-2 gene. Kumar et al. suggest that a similar mechanism involving the posttranslational modification of SATB1 may be involved in the global regulation of gene expression.

An elegant study by Cai et al. (2006) demonstrated how SATB1 regulates long-range intrachromosomal interactions by changing the chromatin loop landscape to coordinate the expression of several genes (Il4, Il5, Il13) in T-helper 2 (TH-2) cells. Two essential methods were used in their study: (1) chromatin conformation capture (3C) assay to determine whether two remote genomic sequences interact, and (2) CHIP-loop assay to determine chromatin loops that are attached at their bases with a specific protein. Their results showed that upon TH-2 cells activation, SATB1 expression was induced to assemble a transcriptionally active chromatin structure at the cytokine locus. The “cage-like structure” was made up of numerous loops, all attached to SATB1 at their bases. In addition, histone H3 acetylated at Lys9 and Lys 14, c-maf (a transcription factor in TH-2 cells important for Il13 expression), chromatin remodeling enzyme Brg1 and RNA polymerase II are all bound within this 200-kb region. When SATB1 expression was reduced using RNAi, the TH-2 cells did not form a dense loop structure (same structure as found in inactivated TH-2 cells) and as a result, Il4, Il5, Il13 were not expressed. Therefore, this study has provided an insight to how SATB1 may coordinate the expression of multiple genes in a cluster by bringing them to closer proximity. This would allow for a more efficient interaction between the promoter of these genes and transcriptional regulatory factors.

In another study, Kumar et al. (2007) described how SATB1 organizes the gene rich region of MHC-1 class locus into several chromatin loops by anchoring the MARs to the nucleus at specific distances to ensure proper expression of genes within the locus. Promyelocytic leukemia (PML) oncoprotein, a protein associated with the nuclear body was identified as a SATB1 interacting protein. Together, SATB1 and PML formed a functional complex with putative MARs at the base of the chromatin loops. They mapped five chromatin loops within the MHC-I locus in Jurkat cells (Kumar et al., 2007) spanning 300 kb of the MHC-I locus containing HCG-9 and HLA-F genes in the presence and absence of SATB1. SATB1 appears to be involved in the activation of HCG-9 and repression of the expression of most other genes on the MHC-I locus (see Fig. 4). Using SATB1 RNAi, the chromatin architecture of MHC-1 locus underwent a reorganization taking chromatin structure similar to that of  $\gamma$ IFN treated Jurkat cells, leading to upregulation of HCG4, HCG4P6 as well as HCG-9. The expression of HCG-9 was enhanced when the gene became part of the giant loop.

Although SATB1 is predominantly expressed in thymocytes, its expression in other cell-types also affects chromatin organization and the regulation of many genes. Recently, Kohwi-Shigematsu and workers (Han et al., 2008) showed that the SATB1 expression in breast cancer cells is associated with these cells to become metastatic by reprogramming the chromatin organization, thus resulting in the upregulation of metastasis-associated gene. When SATB1 expression was abolished by RNA-interference in highly aggressive cancer cells (MDA-MB-231),



**Fig. 4** Schematic representation of the chromatin loop structure of MHC-I locus in control cells and  $\gamma$ IFN treated cells as determined by ChIP-loop assay (details of the experiment and results are described in Kumar et al., 2006). Diagrams on *top* depict the loop landscape in linear fashion while those at the *bottom* indicate the same in a circular manner. On the *lower left*, the non-random distribution of SATB1 and PML across the MHC-I locus is exhibited by depicting the occupancy of the two proteins deduced by chromatin immunoprecipitation assay. On *lower right side*, the diagram depicts only the major changes upon  $\gamma$ IFN treatment, notably the chromatin loop containing HCG-9 that becomes larger and extends out from the core of the chromosome, resulting in enhanced HCG-9 gene expression as well as the replacement of SATB1 by another MAR-binding protein (MBP, depicted by a *yellow ellipse*) (modified from Galande et al., 2007)

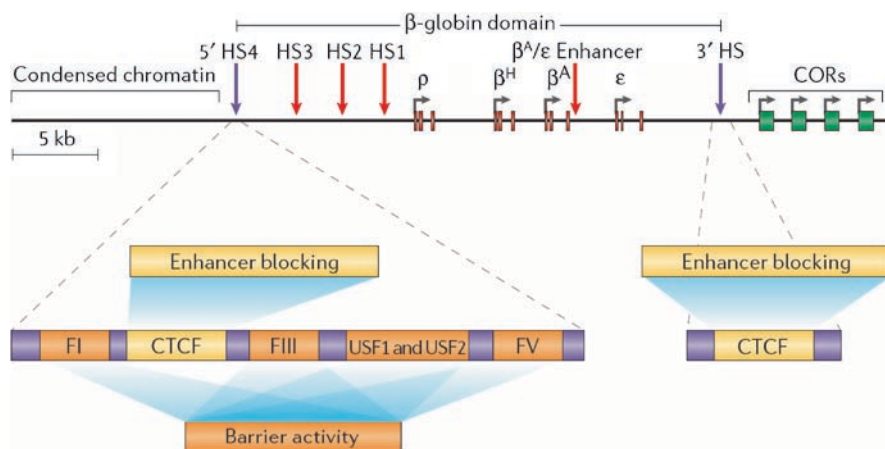
expression of more than 1,000 genes were altered and reversed the tumor growth and metastasis *in vivo*. SATB1 expression in non-aggressive (SKBR3) cells induced the expression of many genes that are associated with aggressive-tumor phenotypes, causing these cells to acquire the ability to metastasis *in vivo*. HMGA, another MAR associated transcription factor that binds to the minor groove of AT rich regions, is also implicated in breast cancer cell progression (Reeves et al., 2001) and may cooperate with SATB1 to promote cell growth and differentiation (Han et al., 2008). The authors propose that SATB1 may be used as a “molecular indicator” to predict

the progression of breast tumors and future studies should investigate SATB1 as a therapeutic target for metastatic breast diseases.

## 5.2 CCCTC-Binding Factor (CTCF)

CTCF is a ubiquitously expressed nuclear protein with 11-zinc finger DNA binding domain (Filippova et al., 1996; Klenova et al., 1993). It is known to have enhancer blocking activity, preventing the action of an enhancer on a promoter when placed in between the two. CTCF also possesses barrier and/or insulator activity, as it may protect transgenes from position effect variegation or heritable silencing through the spread of heterochromatin (Gaszner and Felsenfeld, 2006). The chicken cHS4  $\beta$ -globin insulator, located at the 5' extremity of the chicken  $\beta$ -globin locus, is the first vertebrate enhancer-blocking insulator to be identified (Chung et al., 1993). While, the enhancer blocking activity of HS4 is mediated by CTCF, the barrier activity results from the combined effects of USF1, USF2, FI-, FIII- and FV-binding proteins (Fig. 5).

To date, there have been no reports showing that CTCF is directly involved in the protection of a locus from heterochromatin mediated silencing. However, this is a likely possibility, as CTCF has been shown to bind to nucleophosmin (also



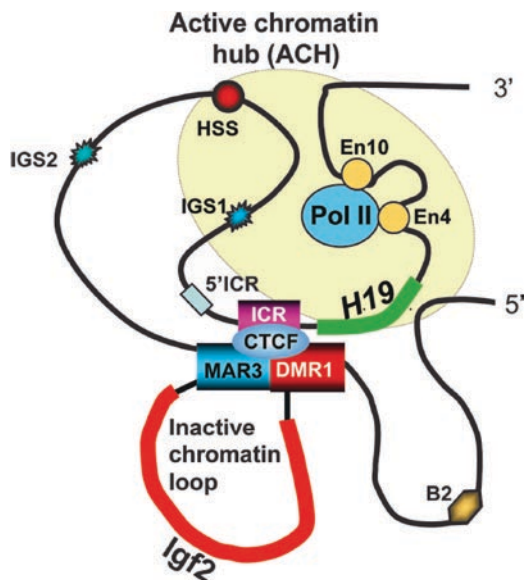
**Fig. 5** In the chicken  $\beta$ -globin gene locus, the 5'HS4 and 3'HS insulator elements define the limits of a chromatin domain that encompasses the developmentally regulated beta-globin gene cluster and its locus-control region (LCR), which is comprised of the HS1–3 and  $\beta$  A/epsilon enhancers. The HS4 element possesses both enhancer blocking and barrier activity, presumably to prevent the LCR from inappropriately activating genes outside the domain and at the same time protecting the globin cluster against silencing that emanates from the flanking condensed-chromatin region. Enhancer blocking is mediated by CTCF, whereas barrier activity results from the combined effect of USF1 and USF2 and the as yet uncharacterized FI-, FIII- and FV-binding proteins. 3' HS binds CTCF and functions only as an enhancer-blocking insulator (reprinted by permission from Macmillan Publishers Ltd: Nat. Rev. Genet., Gaszner and Felsenfeld, 2006, copyright 2006)



known as B32), suggesting that CTCF may anchor to the nuclear matrix, creating an independent “loop” domain that would shield the transgene from silencing. Work by Rincon-Arano et al. (2007) demonstrated that the chicken cHS4  $\beta$ -globin insulator protected transgene from silencing by the telomeric heterochromatin. In this study, EGFP transgene integration was targeted to the telomere of the chicken cell line HD3 with or without cHS4  $\beta$ -globin insulator. The cHS4  $\beta$ -globin insulator sustained transgene expression of a single-copy integrant for over 100 days. By contrast, the un-insulated single copy clones showed a rapid extinction of the transgene expression. RNAi-mediated knockdown of USF1 did not alter cHS4 protection of the transgene from telomeric silencing, demonstrating that cHS4 insulation of the transgene is not dependent on USF1. There was no direct evidence for the role of CTCF in the protective effect. Recruitment of CTCF as a fusion to the GAL4 DNA binding domain did not protect from telomeric silencing (Esnault et al., 2009), suggesting that other nuclear factors must be recruited to the cHS4 to play a role in the protection of the transgene against telomeric position effect (TPE).

The role of CTCF as an enhancer-blocking insulator in the regulation of gene imprinting and monoallelic gene expression (Fedoriw et al., 2004; Ling et al., 2006) is well characterized in the imprinted IGF-2 (insulin-like growth factor 2)-H19 locus. IGF-2 is only expressed from the paternal allele, whereas H19 is expressed from the maternal allele, both genes sharing the same enhancers. The imprinting control region (ICR) is located at the 5' flank of H19 gene and its deletion results in biallelic expression of both IGF-2 and H19, suggesting the role of H19-ICR to repress the maternal IGF-2 allele (Thorvaldsen et al., 1998). CTCF binding to the H19-ICR is thought to be required for the IGF-2 repression (Kaffer et al., 2000). Work by Kurukuti et al. (2006) demonstrated that the IGF-2 promoter on the paternal chromosome interacts with the enhancers. By contrast, on the maternal allele, this interaction is prevented by CTCF binding to the maternal ICR. CTCF binding to the ICR regulates its interaction with matrix attachment region 3 (MAR3) and differentially methylated region (DMR) 1 at IGF-2 gene, forming a condensed loop around the maternal IGF-2 locus. As a result, the interactions between of IGF-2 promoter and the H19 enhancers are prevented, leading to the silencing of IGF-2 expression (Fig. 6).

Since the initial discovery of CTCF, there has been great interest in identifying potential binding sites for CTCF in the eukaryotic genome as this knowledge is essential to understand how *cis*-regulatory elements coordinate expression of target genes. Kim et al. (2007) identified over 13,000 novel putative CTCF-binding sequences as well as confirmed CTCF binding sites in the human genome, using chromatin immunoprecipitation followed by genome-tiling microarrays methodology (Kim et al., 2005b). They found that most of the putative CTCF binding sites are located far from the transcriptional start sites and that their distribution is strongly correlated with genes. Interestingly, CTCF localization appears to be similar in different cell types, as determined by their analysis on the primary human fibroblast cells and hematopoietic progenitor cell line U937. In some cases, CTCF binding sites were located at the boundaries of distinct chromatin structures, but this was not a general phenomenon, as many other binding sites did not coincide with boundaries. Again this evidence points to the fact that if CTCF contributes to the



**Fig. 6** Model showing contacts established within the maternal allele at the IGF-2/H19 region in neonatal liver. The model suggests a mechanism of how CTCF controls the repression of maternal IGF2 gene (located within inactive chromatin loop). CTCF binding to the H19 ICR, regulates its interaction with the matrix attachment region (MAR)3 and differentially methylated region (DMR)1, forming a condensed loop around the IGF-2 gene, restricting the H19 enhancers (en4 and en10) access to the IGF-2 promoter. This model is based on results from neonatal liver only and may not apply to other tissues (Kurukuti et al., 2006; copyright 2006 National Academy of Sciences, USA)

establishment of chromatin boundaries by elements such as MARs, other as yet unidentified activities must also contribute to the boundary effect.

### 5.3 High Mobility Group (HMGA)

The high mobility group A family of proteins comprises of 3 proteins HMGA1a, HMGA1b, and HMGA2 (previously known as HMGI, HMGY, and HMGI-C respectively) (Sgarra et al., 2004). These proteins contain three positively highly charged regions called the AT-hook since they bind to the minor groove of AT rich sequences of the promoter regions and MARs. HMGA1 protein has been found to co-localize with the enzyme topoisomerase II and histone H1 (Saitoh and Laemmli, 1993; Saitoh and Laemmli, 1994) suggesting that it acts as a regulator of gene transcription by controlling the structure of chromatin. It has been demonstrated that HMGA proteins can serve as transcriptional activators in the context of chromatin by displacing histones H1 from MAR sequences (Zhao et al., 1993).

Earlier footprinting studies showed that HMGA proteins preferentially bind to a stretch of five or six AT base pairs (Solomon et al., 1986). However, more recent

studies have shown that HMGA proteins have more sequence specificity, requiring two or three appropriately spaced AT rich sequences as a single multivalent binding site (Maher and Nathans, 1996). For example, HMGA proteins simultaneously bind to two or three runs of AT base pairs in the regulatory regions on human  $\beta$ -interferon enhancer (Thanos and Maniatis, 1992), and the promoter regions of interleukin-2 (Baldassarre et al., 2001) and interleukin-2 receptor  $\alpha$ -chain gene (John et al., 1995). Using a PCR-base systematic evolution of ligands by exponential enrichment (SELEX), Cui and Leng (2007) identified two consensus sequences for HMGA2: 5'-ATATTTCGCGAWWATT-3' and 5'-ATATTGCGCAWWATT-3', where W represents A or T. These sequences can be divided into three segments: the first segment has five base pairs that is AT rich, the middle segment has four base pairs that GC-rich and the last segment has six base pairs that is AT-rich. All three segments are required for HMGA2 binding.

Indirect evidence for the role of HMGA in MAR functional activity was provided by studies performed using aggressive breast carcinoma cell lines, showing elevated expressions of HMGA1a and HMGA1b as compared to non-metastatic cells (Liu et al., 1999). South-western blot analysis using whole protein extracts from these tumor cells exhibited strong binding of these HMGA proteins to a synthetic MAR probe composed of multimer containing the 25-bp sequence derived from a MAR 3' of the IgH enhancer. This 25-bp sequence of MAR is a base-unpairing region (BUR) and it binds to the nuclear matrix with high affinity. Western blot and protein sequencing analysis confirmed that these BUR-binding proteins were indeed HMGA proteins. By contrast, the HMGA proteins were shown to bind poorly to a mutated MAR probe that is still AT-rich but has lost the unwinding propensity. Therefore, HMGA proteins appeared to strictly bind to base-unpairing sequences, one of the key structural element of MARs, and they may participate in gene regulation to trigger metastatic phenotype in breast cancer cells. Similarly to SATB1, HMGA proteins may thus be used as a biomarker for tumor progression. Whether the implication of these proteins in cancer progression may result from their proposed contribution to MAR activity is an interesting but as yet unestablished possibility.

Other MAR transcription factors include B-cell specific protein called BRIGHT (Herrscher et al., 1995), NMP4 proteins known to bind to minor groove of homopolymeric (dA:dT) sites in the core unwinding regions of MARs (Torrunguang et al., 2002) and scaffold attachment factor-A (SAF-A), a multifunctional matrix specific factor that recognizes AT-rich DNA sequences (Romig et al., 1992). These proteins may also contribute to mediating some of the conformational and/or chromatin structure effects of the MAR, but their specific contribution(s) to these effects remain to be identified.

## **6 Effects of MARs on the Copy Number of Integrated Transgenes**

In the examples given above, it is clear that MAR elements can enhance and maintain long-term expression by acting on the structure of chromosomes and of chromatin, and that these effects can lead to increased transcription of the transgenes.

In some studies, the MAR elements also appear to reduce the variability within a polyclonal cell population. Thus, MARs may provide more consistent and elevated transcription to each integrated transgene copy. However, other mechanisms such as those that relate to the transgene copy number may also concur to increased expression. In this section, we discuss the role that MAR elements may take to augment transgene integration in the host genome, thereby yielding increased transgene copy number and overall expression.

Several studies demonstrated that MAR elements increase the number of integrated transgene copies in transfected plants and mammalian cells. For instance, it was found that the inclusion of the human MAR 1-68 in transfected plasmids significantly enhances the number of copies integrated in the host genome, as compared to cells transfected without MAR. Indeed, quantitative PCR assays performed either on stable cell populations or clones confirmed a 3–4-fold higher transgene copy number in cells transfected with the MAR 1-68 (Girod et al., 2007; Galbete et al., 2009), in agreement with previous observations (Kim et al., 2004; Girod et al., 2005). In Kim et al. (2004), the authors observed higher transgene copy numbers when these genes are co-transfected with the human  $\beta$ -globin MAR in CHO cells. Girod et al. (2005) achieved similar results in CHO cell clones transfected with the chicken lysozyme MAR. Furthermore, fluorescent *in situ* hybridization analysis of metaphase chromosomes of stable polyclonal populations showed generally much greater intensity of a fluorescent probe in cells transfected with MARs, therefore confirming the increase of transgene integration (Girod et al., 2007).

Similarly, many examples showed that MAR elements enhanced expression in a copy-number dependent manner. For example, transgenic mice carrying multiple copies of a reporter gene flanked by the chicken lysozyme MAR expressed the gene at levels proportional to copy number, indicating that a complete gene locus, as defined by its chromatin structure, functions as an independent regulatory unit when introduced into a heterologous genome (Bonifer et al., 1990, 1994). In addition, the presence of a MAR from the chicken lysozyme locus reduced variability and conferred a copy number-dependent increase in transgene expression in transgenic rice plants (Oh et al., 2005). However, Park and Kay (2001) observed that the chicken lysozyme MAR did not improve the number of proviral DNA copies integrated in mouse hepatocytes whereas the immunoglobulin-kappa MAR exhibited a 2.5-fold augmentation.

In contrast, a study by Wang et al. (2007) revealed that the expression of the CAT enzyme in stably transformed lines of the microalgae *Dunaliella salina* was not significantly proportional to the gene copy numbers, suggesting that the effects of MARs on transgene expression may not be through increasing transgene copies. In addition, it was shown that in preimplantation mouse embryos, flanking SARs stimulated transgene expression in a copy-dependent manner. But in the differentiated tissues of newborn and adult mice, correlation with copy number was lost (Thompson et al., 1994). Furthermore, Baur et al. (2004) demonstrated that even a single gene copy might also result in a variegated expression, as show by the spontaneous changes of expression of a luciferase reporter gene integrated near HeLa cell telomeric heterochromatin. Thus, there is a clear benefit in including MAR elements in the transfection vector to increase transgene expression.

However, the ability of a MAR to confer a copy number-dependent expression to the transgene by insulating them to prevent gene silencing or position effect is less clear. Differences in the effect noted by various experimenters, in some occasions working with the same MAR element, may result from other factors influencing expression, such as the promoters and vector backbones as well as the cell lines and transfection methods used.

MAR elements appear to be able to counteract silencing effects, as exemplified when comparing stable cell populations transfected with or without MAR. Transgene copy number and cell fluorescence levels were shown to correlate well in the presence of MAR, indicating that the increase of transgene expression results from a similar increase in transgene integration (M. Grandjean and N. Mermoud, unpublished data). In contrast, the normalization of EGFP mRNA levels relative to the gene copy number from stable cell clones indicated that the MAR increased gene expression by twofold on average (Galbete et al., 2009). Thus, increased transgene expression observed with MAR is likely to result both from the integration of more transgene copies in the genome of cells and from MAR-mediated inhibition of epigenetic silencing events that are associated with the integration of tandem gene copies.

It is known that mammalian cells in culture contain the enzymatic machinery required to mediate recombination between newly introduced plasmid DNA molecules and that the frequency of homologous recombination or non-homologous end-joining between co-injected plasmid molecules to form concatemers is extremely high, approaching 100% (Folger et al., 1985). However, integration of one of these concatemers into one of the chromosome is a relatively rare event in mammalian cells (Folger et al., 1985). As a result, multiple copies of the transfected gene are not scattered throughout the host genome, but they co-integrate as concatemer at a single locus in the host chromosome, usually in tandem head-to-tail orientation (Folger et al., 1982), the integration site being different in independent transformants (Robins et al., 1981). However, recombination between the newly introduced DNA from transfection and its homologous chromosomal sequence occurs exceedingly rarely in mammalian cells, at a frequency of 1:1,000 cells receiving DNA (Thomas et al., 1986).

The high copy number of transgenes integrated in the genome of the cells with a MAR does not result from a more efficient plasmid import into the nucleus during transfection, or from the occurrence of multiple chromosomal integration events (Girod et al., 2007; Grandjean et al., personal communication). This effect might rather be linked to an effect of MAR on increased DNA concatemerization and/or facilitated transgene integration. Indeed, MARs may play a role as DNA recombination signals. It was previously shown that MAR elements could regulate recombination processes such as immunoglobulin gene rearrangement (Xu et al., 1996). Breakpoints of recurrent deletions and translocations in leukemia were found to occur at MARs, thus facilitating their illegitimate recombination at the nuclear matrix (Iarovaia et al., 2004). Finally, retroviruses showed a strong preference for integration in the vicinity of MARs (Johnson and Levy, 2005).

How MARs may increase transgene integration is currently unknown. Because they mediate a permissive chromatin structure, MARs could improve homologous

recombination between transfected plasmids, thus allowing the formation of larger concatemers, yielding the observed increased number of gene copies that integrate within the genome of cultured cells without leading to multiple integration sites within a transfected cell. Alternatively, but not exclusively, MAR may interact with proteins of the repair machinery that are known to contribute also to homologous recombination and non-homologous end-joining events.

## 7 Effects of MARs on Transgene Expression Variegation

In addition to the effects of MARs on transgene silencing and integration, an additional role of these elements in preventing variegation is currently being uncovered. The high variability among independent transformants in stable expression is thought to depend on the site of transgene integration in the chromosome (Kalos and Fournier, 1995; Recillas-Targa et al., 2002). Indeed, transgene integration may be influenced by the fortuitous presence of regulatory elements at the random integration locus in the host genome. In addition, transgene expression is thought to reflect particular chromatin structure coming from adjacent chromosomal domains (Robertson et al., 1995; Henikoff, 1996; Wakimoto, 1998). However, variability of expression can also be noted from distinct cells within a monoclonal population that have the transgene integrated at the same chromosomal locus. This effect, described as variegation, is most clearly seen when individual cells express transgenes with easily detectable products such as short half-life fluorescent proteins. However, the extent of this effect will itself vary when assessing individual clones, and some integration sites may thus be more prone to variegation than others.

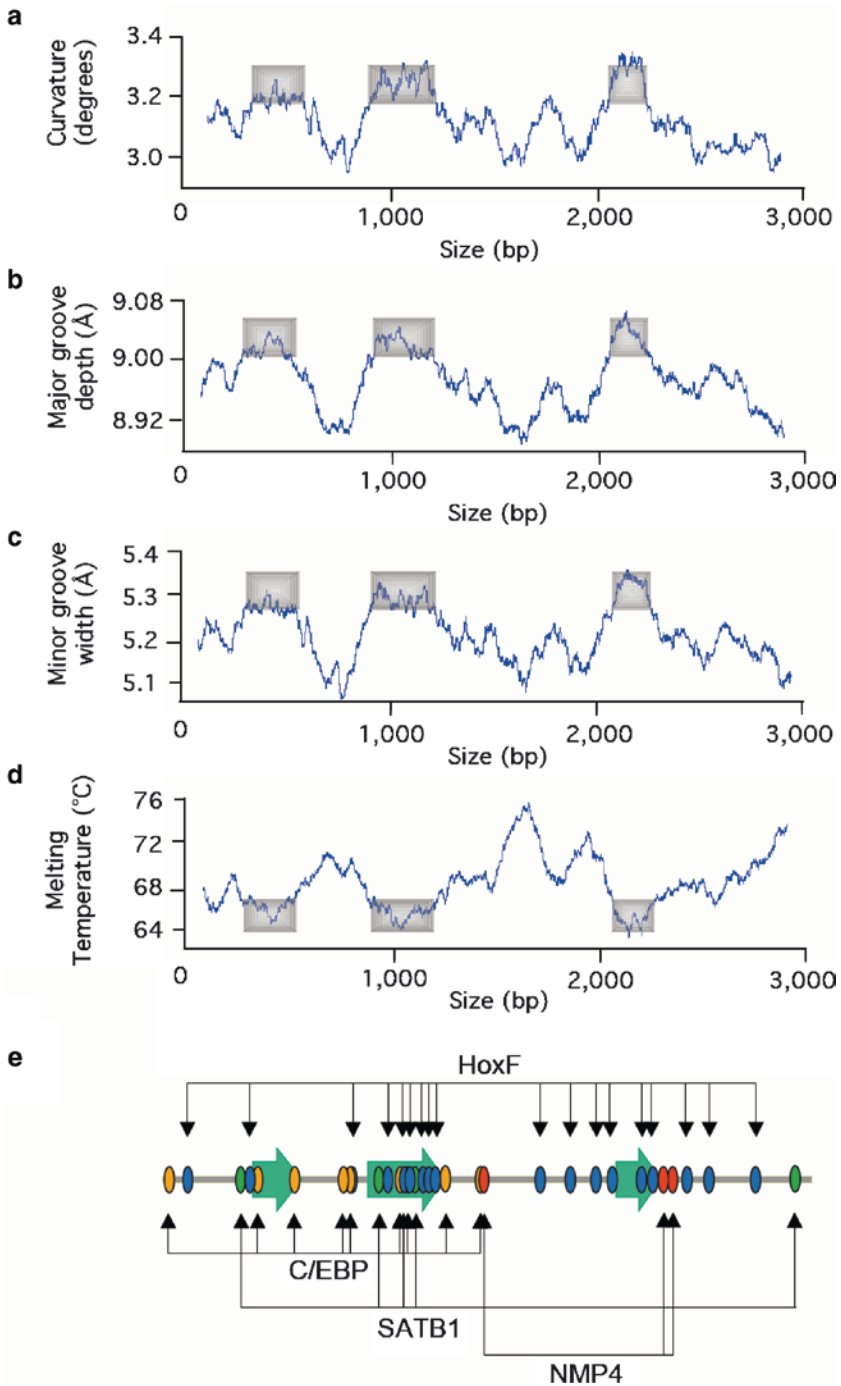
The human MAR 1-68 was found to decrease variegation in addition to its effect to improve transgene expression, as cells within individual colonies showed similar levels of expression (Girod et al., 2007). The localization of transgene integration sites, as assessed by fluorescent *in situ* hybridization, did not show any multiple integration events and transgenes did not appear to be targeted to any specific chromosomal sites or particular chromosomal structures in cells transfected with the MAR 1-68 (Derouazi et al., 2006; Girod et al., 2007). Time-lapse microscopy of GFP expression in single cells indicated that MAR 1-68 mediated constant transgene expression, while cells generated without this element would cycle between states of high expression and silent states within a time frame of hours and days (Galbete et al., 2009). Thus, in addition to their long-term effects on the inhibition of heterochromatin formation, MAR can also act positively to mediate constant gene transcription, as opposed to expression cycling usually obtained from transgenes devoid of these epigenetic regulators. The MAR effect on expression variegation was discovered recently, and the molecular mechanisms that oppose a variegated expression pattern remain uncharacterized, but it may conceivably be linked to the action of MARs on chromatin structure and/or on the assembly or firing of transcription initiation complexes at promoters.

## 8 Isolation of Potent MAR Elements via Bioinformatics to Generate Producer Cell Lines

In the above section, we have discussed MAR elements that have been identified associated with specific genes: chicken lysozyme,  $\beta$ -globin and  $\beta$ -interferon, for example. Here we discuss the isolation of MAR elements on a genome-wide approach using bioinformatics. Since no unique consensus sequence for MARs has been found (Boulikas, 1993; Kramer and Krawetz, 1995), identifying nuclear matrix attachment DNA regions *in silico* based on high A+T percentage has proven feasible (Evans et al., 2007; Girod et al., 2007). Girod et al. (2007) designed a computational method to predict MARs from human genomic sequences based on the specific characteristics of the A+T rich region such as the low melting temperature, high curvature, deep major groove depth and wide minor groove width of the DNA, as well as from the occurrence of binding sites for particular transcription factors (Girod et al., 2007). They identified 1,566 high-scoring putative MAR sequences when the algorithm was set with very stringent parameters. Out of these, they selected seven putative MAR elements for further analysis based on the presence of A+T rich core region and putative binding sites of transcription factors. All the selected seven potential MARs contain a long stretch of DNA (200 bp to 1.5 kb in length) made up of approximately 70–85% AT dinucleotides, almost devoid of any guanine and cytosine nucleotides. They assessed the ability of each MAR to activate transgene expression and found that all but one of the seven newly identified MAR elements augmented substantially EGFP transgene expression in stably transfected CHO cells. One of these MARs significantly increased IgG production in CHO cells and maintained high expressions of erythropoietin transgene from an inducible doxycycline promoter in mice. Whether or not the AT rich regions of these MARs play a significant role in minimizing silencing and activating transgene transcription in CHO or animal model is to be further investigated.

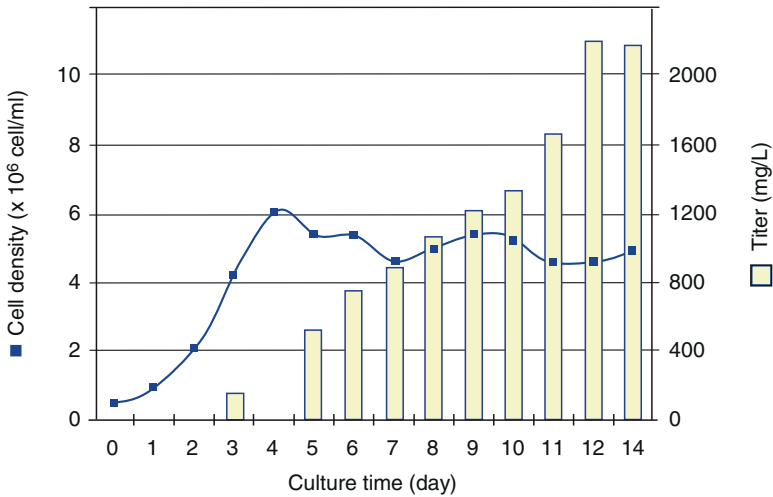
Computational analysis on the 3-kilo base pair chicken lysozyme 5' MAR elements showed three regions within this MAR that contain potentially curved DNA structures, a deep major groove and low DNA melting temperatures (Fig. 7). Within these regions there are short A+T rich sequence motifs composed of stretches of 6 to 10 oligo dA predicted to mediate nucleosome positioning and curved DNA configuration. There appear to be a correlation between the distribution of nucleosome positioning motifs and the sequences that increase EGFP expression levels. However, it is not clear if the A+T rich elements alone confer most of the MAR effects in sustaining EGFP transgene expression in CHO.

These novel human elements overcome the need for amplification, and assure that all copies of the transgenes are actively expressed. Girod et al. (2007) assessed the effect of MAR 1-68 on antibody expression in CHO cells. In this study a comparison of the MAR 1-68 and the chicken lysozyme MAR was made with respect with the ability of each element to augment protein production. The highest antibody expression occurred using MAR 1-68, with one clone secreting over 70 picograms of antibody per cell and per day (p/c/d). This compares favorably with



**Fig. 7** Computational (SMARScan) analysis of chicken lysozyme MAR. (a) Double helix bending angle, (b) major groove depth, (c) minor groove width, (d) DNA melting temperature, and (e) schematic diagram of chicken lysozyme MAR with putative binding sites for transcription factors C/EBP, Hox F, NMP4 and SATB1 (marked by *colored ellipses*) (from the authors own work, Girod et al., 2007)





**Fig. 8** Cell densities and IgG titers are from 1-liter bench-top bioreactors seeded at a target seed density of  $0.5 \times 10^6$  cells/ml with a CHO cell clone producing an immunoglobulin gamma (IgG) under the control of a human MAR element. Cell line identification was performed without transgene amplification during a 15 week period, and titers and cell densities were determined before process or media optimization (from the authors own work, Varghese et al., 2008)

the levels achieved with the chicken lysozyme MAR elements, which peaked at 30 p/c/d. Approximately one clone in 30 shows a productivity of 30 p/c/d or higher with the human MAR, whereas isolation and screening of more than 300 clones was required using the chicken lysozyme MAR. Clones secreting large amounts of immunoglobulin were adapted to growth in suspension in serum-free synthetic medium, and they maintained high and stable expression without selection pressure as long as tested, during several months.

These human MARs, and new sequences derived from these elements, are currently being used to generate cell lines for the commercial production of pharmaceuticals as well as for diagnostic kits. An example of the productivities that are routinely obtained is illustrated in Fig. 8. The main benefits of the incorporation of such elements in expression vectors are reduced time, efforts and costs for the generation, screening and characterization of cell lines, often coupled to a gain in productivity, because stable and very high producer clones can be obtained from the screening of few cell lines.

## 9 Conclusions

MAR elements have been linked to a bewildering array of activities, including the formation of higher order chromosomal loops and their positioning to sub-nuclear compartments enriched in proteins mediating DNA transcription and RNA maturation,

the recruitment of proteins mediating chromatin modifications that decrease silencing effects, the reduction of variegation effects that limits transgene expression, and increased transgene integration into the cell genome. All of these effects are likely to contribute to very high expression levels of adjacent genes, yielding elevated production of recombinant proteins by cells such as the CHO or HEK293 lines. Specific productivities ranging from 20 to 100 picogram p/c/d have been reported during the development of commercial cell lines, and titers above 5 g/l have been achieved.

In addition to the use of these elements to generate producer cell lines, progresses have been made to identify the fundamental constituents of MARs, despite a wide variety of sequences and activities. MARs appear to act as scaffolds that combine DNA and protein elements working cooperatively to control chromatin structure. For instance, particular DNA sequences may act to position nucleosomes, whereas other sequences act as docking sites for proteins that mediate modifications of the histones and a gene expression-permissive chromatin structure. At present, however, a detailed molecular understanding of the contribution of these elements to the action of MARs on gene expression or DNA recombination is missing, which has precluded the assembly of totally synthetic MARs from individual optimized building blocks. Nevertheless, we speculate that these goals that will be found worthy of further research efforts, and that these efforts in turn will yield even simpler procedures to construct mammalian cell lines that produce high titers of recombinant proteins.

## References

- Allen G, Hall GJ, Michalowski S, Newman W, Spiker S, Weissinger A, Thompson W (1996) High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco. *Plant Cell* 8:899–913
- Almeida R, Robin C (2005) RNA silencing and genome regulation. *Trends Cell Biol* 15:251–258
- Andrulis IL, Duff C, Evans-Blackler S, Worton R, Siminovitch L (1983) Chromosomal alterations associated with overproduction of asparagine synthetase in albizzin-resistant Chinese hamster ovary cells. *Mol Cell Biol* 3:391–398
- Baldassarre G, Fedele M, Battista S, Vecchione A, Klein-Szanto AJ, Santoro M, Waldmann TA, Azimi N, Croce CM, Fusco A (2001) Onset of natural killer cell lymphomas in transgenic mice carrying a truncated HMG1-C gene by the chronic stimulation of the IL-2 and IL-15 pathway. *Proc Natl Acad Sci USA* 98:7970–7975
- Barnes LM, Dickson AJ (2006) Mammalian cell factories for efficient and stable protein expression. *Curr Opin Biotech* 17:381–386
- Barnes LM, Bentley CM, Dickson AJ (2001) Characterization of the stability of recombinant protein production in the GS-NS0 expression system. *Biotechnol Bioeng* 73:261–270
- Barnes LM, Bentley CM, Dickson AJ (2003) Stability of protein production from recombinant mammalian cells. *Biotechnol Bioeng* 81:631–639
- Barnes LM, Bentley CM, Dickson AJ (2004) Molecular definition of predictive indicators of stable protein expression in recombinant NS0 myeloma cells. *Biotechnol Bioeng* 85:115–121
- Barnes LM, Moy N, Dickson AJ (2006b) Phenotypic variation during cloning procedures: analysis of the growth behavior of clonal cell lines. *Biotechnol Bioeng* 94:530–537
- Barnes LM, Bentley CM, Moy N, Dickson AJ (2007) Molecular analysis of successful cell line selection in transfected GS-NS0 myeloma cells. *Biotechnol Bioeng* 96:337–348

- Baur JA, Shay JW, Wright WE (2004) Spontaneous reactivation of a silent telomeric transgene in a human cell line. *Chromosoma* 112:240–246
- Bode J, Kohwi Y, Dickinson L, Joh T, Klehr D, Mielke C, Kohwi-Shigematsu T (1992) Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science* 255:195–197
- Bode J, Stengert-Iber M, Kay V, Schlake T, Dietz-Pfeilstetter A (1996) Scaffold/matrix-attached regions: topological switches with multiple regulatory functions. *Crit Rev Eukaryot Gene Expr* 6:115–138
- Bode J, Goetze S, Heng H, Krawetz SA, Benham C (2003) From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. *Chromosome Res* 11:435–445
- Bode J, Winkelmann S, Gotze S, Spiker S, Tsutsui K, Bi C, A KP, Benham C (2006) Correlations between scaffold/matrix attachment region (S/MAR) binding activity and DNA duplex destabilization energy. *J Mol Biol* 358:597–613
- Bonifer C, Vidal M, Grosveld F, Sippel AE (1990) Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *EMBO J* 9:2843–2848
- Bonifer C, Yannoutsos N, Kruger G, Grosveld F, Sippel AE (1994) Dissection of the locus control function located on the chicken lysozyme gene domain in transgenic mice. *Nucl Acids Res* 22:4202–4210
- Boulikas T (1993) Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. *J Cell Biochem* 52:14–22
- Cai S, Han HJ, Kohwi-Shigematsu T (2003) Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet* 34:42–51
- Cai S, Lee CC, Kohwi-Shigematsu T (2006) SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet* 38:1278–1288
- Chung JH, Whiteley M, Felsenfeld G (1993) A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* 74:505–514
- Clark DJ, Felsenfeld G (1991) Formation of nucleosomes on positively supercoiled DNA. *EMBO J* 10:387–395
- Cui T, Leng F (2007) Specific recognition of AT-rich DNA sequences by the mammalian high mobility group protein AT-hook 2: a SELEX study. *Biochemistry* 46:13059–13066
- Derouazi M, Martinet D, Besuchet Schmutz N, Flaction R, Wicht M, Bertschinger M, Hacker DL, Beckmann JS, Wurm FM (2006) Genetic characterization of CHO production host DG44 and derivative recombinant cell lines. *Biochem Biophys Res Commun* 340:1069–1077
- Dickinson LA, Joh T, Kohwi Y, Kohwi-Shigematsu T (1992) A tissue-specific MAR/SAR DNA-binding protein with unusual binding site recognition. *Cell* 70:631–645
- Eissenberg JC (1989) Position effect variegation in *Drosophila*: Towards genetics of chromatin assembly. *Bioessays* 11:14–17
- Eissenberg JC, Morris GD, Reuter G, Hartnett T (1992) The heterochromatin-associated protein HP-1 is an essential protein in *Drosophila* with dosage-dependent effects on position-effect variegation. *Genetics* 131:345–352
- Esnault G, Majocchi S, Martinet D, Besuchet-Schmutz N, Beckmann JS, Mermod N (2009) Transcription factor CTFA acts as a chromatin domain boundary that shields human telomeric genes from silencing. *Mol Cell Biol* 29:2409–2418
- Esteller M (2008) Epigenetics in cancer. *N Engl J Med* 358:1148–1159
- Evans K, Ott S, Hansen A, Koentges G, Wernisch L (2007) A comparative study of S/MAR prediction tools. *BMC Bioinformatics* 8:71–100
- Fann CH, Guirgis F, Chen G, Lao MS, Oiret JM (2000) Limitations to the amplification and stability of human tissue-type plasminogen activator expression by Chinese hamster ovary cells. *Biotechnol Bioeng* 69:204–12
- Fedoriw AM, Engel NI, Bartolomei MS (2004) Genomic imprinting: antagonistic mechanisms in the germ line and early embryo. *Cold Spring Harb Symp Quant Biol* 69:39–45
- Feinberg AP (2008) Epigenetics at the epicenter of modern medicine. *JAMA* 299:1345–1350

- Felsenfeld G (1992) Chromatin as an essential part of the transcriptional mechanism. *Nature* 355:219–224
- Felsenfeld G (1996) Chromatin unfolds. *Cell* 86:13–19
- Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV (1996) An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol* 16:2802–2813
- Flintoff WF, Livingston E, Duff C, Worton RG (1984) Moderate-level gene amplification in methotrexate-resistant Chinese hamster ovary cells is accompanied by chromosomal translations at or near the site of amplified DHFR gene. *Mol Cell Biol* 4:69–76
- Folger KR, Wong EA, Wahl G, Capecchi MR (1982) Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Mol Cell Biol* 2:1372–1387
- Folger KR, Thomas K, Capecchi MR (1985) Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. *Mol Cell Biol* 5:59–69
- Forrester WC, Novak U, Gelinas R, Groudine M (1989) Molecular analysis of the human beta-globin locus activation region. *Proc Natl Acad Sci USA* 86:5439–5443
- Fouremana P, Winfield JA, Peter J, Hahn PJ (1998) Chromosome breakpoints near CpG islands in double minutes. *Gene* 218:121–128
- Galande S, Purbey PK, Notani D, Kumar PP (2007) The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr Opin Genet Dev* 17:408–414
- Galbete JL, Buceta M, Mermod N (2009) MAR elements regulate the probability of epigenetic switching between active and inactive gene expression. *Mol Biosyst* 5:143–150
- Gaszner M, Felsenfeld G (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* 7:703–713
- Gilmour DS, Plugfelder G, Wang JC, Lis JT (1986) Topoisomerase I interacts with transcribed regions in *Drosophila* cells. *Cell* 44:401–407
- Girod P-A and Mermod N (2003) Use of scaffold/matrix attachment regions for protein production. *Gene Transfer and Expression in Mammalian cells*. In S.C. Makrides (ed.), Elsevier Science pp. 359–379
- Girod P-A, Zahn-Zabal M, Mermod N (2005) Use of the chicken lysozyme 5' matrix attachment region to generate high producer CHO cell lines. *Biotechnol Bioeng* 91:1–11
- Girod P-A, Nguyen DQ, Calabrese D, Puttini S, Grandjean M, Martinet D, Regamey A, Saugy D, Beckmann JS, Bucher P, Mermod N (2007) Genome-wide prediction of matrix attachment regions that increase gene expression in mammalian cells. *Nat Methods* 4:747–753
- Grewal SIS, Moazed D (2003) Heterochromatin and epigenetic control of gene expression. *Science* 301:798–802
- Gutierrez-Adan A, Pintado B (2000) Effect of flanking matrix attachment regions on the expression of microinjected transgenes during preimplantation development of mouse embryos. *Transgenic Res* 9:81–89
- Han HJ, Russo J, Kohwi Y, Kohwi-Shigematsu T (2008) SATB1 reprograms gene expression to promote breast tumor growth and metastasis. *Nature* 452:187–193
- Hart CM, Laemmli UK (1998) Facilitation of chromatin dynamics by SARs. *Curr Opin Genet Dev* 8:519–525
- Henikoff S (1996) Dosage-dependent modification of position-effect variegation in *Drosophila*. *Bioessays* 18:401–409
- Herrscher RF, Kaplan MH, Lelsz DL, Das C, Scheuermann R, Tucker PW (1995) The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes Dev* 9:3067–3082
- Homberger HP (1989) Bent DNA is a structural feature of scaffold-attached regions in *Drosophila melanogaster* interphase nuclei. *Chromosoma* 98:99–104
- Iarovaia OV, Shkumatov P, Razin SV (2004) Breakpoint cluster regions of the AML-1 and ETO genes contain MAR elements and are preferentially associated with the nuclear matrix in proliferating HEL cells. *J Cell Sci* 117:4583–4590

- John S, Reeves RB, Lin JX, Child R, Leiden JM, Thompson CB, Leonard WJ (1995) Regulation of cell-type-specific interleukin-2 receptor alpha-chain gene expression: potential role of physical interactions between Elf-1, HMG-I(Y), and NF-kappa B family proteins. *Mol Cell Biol* 15:1786–1796
- Johnson CN, Levy LS (2005) Matrix attachment regions as targets for retroviral integration. *Virology* 2:68–77
- Jun SC, Kim MS, Hong HJ, Lee GM (2006) Limitations to the development of humanized antibody producing Chinese hamster ovary cells using glutamine synthetase-mediated gene amplification. *Biotechnol Prog* 22:770–780
- Kaffer CR, Srivastava M, Park KY, Ives E, Hsieh S, Battle J, Grinberg A, Huang SP, Pfeifer K (2000) A transcriptional insulator at the imprinted H19/Igf2 locus. *Genes Dev* 14:1908–1919
- Kalos M, Fournier REK (1995) Position-independent transgene expression mediated by boundary elements from the apolipoprotein B chromatin domain. *Mol Cell Biol* 15:198–207
- Kamiya H, Fukunaga S, Ohyama T, Harashima H (2007) The location of the left-handedly curved DNA sequence affects exogenous DNA expression in vivo. *Arch Biochem Biophys* 461:7–12
- Kim SJ, Lee GM (1999) Cytogenetic analysis of chimeric antibody-producing CHO cell sin the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. *Biotechnol Bioeng* 64:741–749
- Kim SJ, Chun JR, Hong HJ, Lee GM (1998) Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. *Biotechnol Bioeng* 58:73–84
- Kim NS, Kim SJ, Lee GM (1998) Clonal variation within dihydrofolate reductase-mediate gene amplified Chinese hamster ovary cells: stability in the absence of selective pressure. *Biotech Bioeng* 60:679–88
- Kim NS, Byun TH, Lee GM (2001) Key determinants in the occurrence of clonal variation in humanized antibody expression of CHO cells during dihydrofolate reductase mediated gene amplification. *Biotechnol Prog* 17:69–75
- Kim JM, Kim JS, Park DH, Kang HS, Yoon J, Baek K, Yoon Y (2004) Improved recombinant gene expression in CHO cells using matrix attachment regions. *Biotechnol J* 107:95–105
- Kim JD, Yoon Y, Hwang HY, Park JS, SYu J Lee, Baek K, Yoon J (2005a) Efficient selection of stable Chinese hamster ovary (CHO) cell lines for expression of recombinant proteins by using human  $\beta$ -interferon SAR element. *Biotechnol Prog* 21:933–937
- Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B (2005b) A high-resolution map of active promoters in the human genome. *Nature* 436:876–880
- Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanekov VV, Ren B (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128:1231–1245
- Klehr D, Maass K, Bode J (1991) Scaffold-attached regions from the human interferon beta domain can be used to enhance the stable expression of genes under the control of various promoters? *Biochemistry* 30:1264–1270
- Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, Neiman PE, Lobanekov VV (1993) CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol* 13:7612–7624
- Kohwi-Shigematsu T, deBelle I, Dickinson LA, Galande S, Kohwi Y (1998) Identification of base-unpairing region-binding proteins and characterization of their in vivo binding sequences. *Methods Cell Biol* 53:323–354
- Kramer JA, Krawetz SA (1995) Matrix-associated regions in haploid expressed domains. *Mamm Genome* 6:677–679
- Kumar PP, Purbey PK, Ravi DS, Mitra D, Galande S (2005) Displacement of SATB1-bound histone deacetylase 1 corepressor by the human immunodeficiency virus type 1 transactivator induces expression of interleukin-2 and its receptor in T cells. *Mol Cell Biol* 25:1620–1633

- Kumar PP, Purbey PK, Sinha CK, Notani D, Limaye A, Jayani RS, Galande S (2006) Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity *in vivo*. *Mol Cell* 22:231–243
- Kumar PP, Bischof O, Purbey PK, Notani D, Urlaub H, Dejean A, Galande S (2007) Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat Cell Biol* 9:45–56
- Kurukuti S, Tiwari VK, Tavosoidana G, Pugacheva E, Murrell A, Zhao Z, Lobanenkov V, Reik W, Ohlsson R (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci USA* 103:10684–10689
- Kwaks THJ, Otte AP (2006) Employing epigenetics to augment the expression of therapeutic proteins in mammalian cells. *Trends Biotechnol* 24:127–142
- Labrador M, Corces VG (2002) Setting the boundaries of chromatin domains and nuclear organization. *Cell* 111:151–154
- Landsman D, Bustin M (1993) A signature for the HMG-1 box DNA-binding proteins. *Bioessays* 15:539–546
- Liebich I, Bode J, Reuter I, Wingender E (2002) Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). *Nucleic Acids Res* 30:3433–3442
- Ling JQ, Li T, Hu JF, Vu TH, Chen HL, Qiu XW, Cherry AM, Hoffman AR (2006) CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* 312:269–272
- Liu WM, Guerra-Vladusic FK, Kurakata S, Lupu R, Kohwi-Shigematsu T (1999) HMG-I(Y) recognizes base-unpairing regions of matrix attachment sequences and its increased expression is directly linked to metastatic breast cancer phenotype. *Cancer Res* 59:5695–5703
- Maher JF, Nathans D (1996) Multivalent DNA-binding properties of the HMG-1 proteins. *Proc Natl Acad Sci USA* 93:6716–6720
- Matzke M, Matzke A, Kooter J (2001) RNA: guiding gene silencing. *Science* 293:1080–1083
- Melton DW, Brennand J, Ledbetter DH, Konecki DS, Chinault AC, Caskey CT (1982) Phenotypic reversion at the hprt locus as a consequence of gene amplification. In: Schimke RT (ed) *Gene Amplification*. Cold Spring Harbor Laboratory, New York, pp 59–65
- Morris K (2008) RNA-mediated transcriptional gene silencing in human cells. *Curr Top Microbiol Immunol* 320:211–214
- Nishikawa J, Amano M, Fukue Y, Tanaka S, Kishi H, Hirota Y, Yoda K, Ohyama T (2003) Left-handedly curved DNA regulates accessibility to cis-DNA elements in chromatin. *Nucleic Acids Res* 31:6651–6662
- Nowak W, Gawlowska M, Jarmolowski A, Augustyniak J (2001) Effect of nuclear matrix attachment regions on transgene expression in tobacco plants. *Acta Biochim Pol* 48:637–646
- Oh SJ, Jeong JS, Kim EH, Yi NR, Yi SI, Jang IC, Kim YS, Suh SC, Nahm BH, Kim JK (2005) Matrix attachment region from the chicken lysozyme locus reduces variability in transgene expression and confers copy number-dependence in transgenic rice plants. *Plant Cell Rep* 24:145–154
- Ostermeier GC, Liu Z et al (2003) Nuclear matrix association of the human beta-globin locus utilizing a novel approach to quantitative real-time PCR. *Nucleic Acids Res* 31(12):3257–3266
- Pankiewicz R, Karlen Y, Imhof M, Mermod N (2005) Reversal of the silencing of tetracycline-controlled genes requires the coordinate action of distinctly-acting transcription factors. *J Gene Med* 7:117–132
- Park F, Kay MA (2001) Modified HIV-1 based lentiviral vectors have an effect on viral transduction efficiency and gene expression *in vitro* and *in vivo*. *Mol Ther* 4:164–173
- Phi-Van L, Strätling WH (1996) Dissection of the ability of the chicken lysozyme gene 50 matrix attachment region to stimulate transgene expression and to dampen position effects. *Biochemistry* 35:10735–10742
- Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, Gaszner M, Felsenfeld G (2002) Position-effect protection and enhancer blocking by the chicken  $\beta$ -globin insulator are separable activities. *Proc Natl Acad Sci USA* 99:6883–6888

- Reeves R, Edberg DD, Li Y (2001) Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells. *Mol Cell Biol* 21:575–594
- Rincon-Arano H, Furlan-Magaril M, Recillas-Targa F (2007) Protection against telomeric position effects by the chicken cHS4 beta-globin insulator. *Proc Natl Acad Sci USA* 104:14044–14049
- Robertson G, Garrick D, Wu W, Kearns M, Martin D, Whitelaw E (1995) Position-dependent variegation of globin transgene expression in mice. *Proc Natl Acad Sci USA* 92:5371–5375
- Robins DM, Ripley S, Henderson AS, Axel R (1981) Transforming DNA integrates into the host chromosome. *Cell* 23:29–39
- Romig H, Fackelmayer FO, Renz A, Ramsperger U, Richter A (1992) Characterization of SAF-A, a novel nuclear DNA binding protein from HeLa cells with high affinity for nuclear matrix/scaffold attachment DNA elements. *EMBO J* 11:3431–3440
- Saitoh Y, Laemmli UK (1993) From the chromosomal loops and the scaffold to the classic bands of metaphase chromosomes. *Cold Spring Harb Symp Quant Biol* 58:755–765
- Saitoh Y, Laemmli UK (1994) Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell* 76:609–622
- Schubeler D, Mielke C, Maass K, Bode J (1996) Scaffold/matrix-attached regions act upon transcription in a context-dependent manner. *Biochemistry* 35:11160–11169
- Selker EU (1999) Gene silencing: repeats that count. *Cell* 97:157–160
- Sgarra R, Rustighi A, Tessari MA, Di Bernardo J, Altamura S, Fusco A, Manfioletti G, Giancotti V (2004) Nuclear phosphoproteins HMGA and their relationship with chromatin structure and cancer. *FEBS Lett* 574:1–8
- Sjakste NR, Sjakste TG (2001) Use of scaffold/matrix-attachment regions for protein production. *Mol Biol* 35:627–635
- Solomon MJ, Strauss F, Varshavsky A (1986) A mammalian high mobility group protein recognizes any stretch of six A.T base pairs in duplex DNA. *Proc Natl Acad Sci USA* 83:1276–1280
- Stief A, Winter DM, Strätling WH, Sippel AE (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature* 341:343–345
- Strutzenberger K, Borth N, Kunert R, Steinfeldner W, Katinger H (1999) Changes during subclone development, ageing of human antibody-producing recombinant CHO cells. *J Biotechnol* 69:215–226
- Sumida N, Nishikawa J, Kishi H, Amano M, Furuya T, Sonobe H, Ohyama T (2006) A designed curved DNA segment that is a remarkable activator of eukaryotic transcription. *FEBS J* 273:5691–5702
- Thanos D, Maniatis T (1992) The high mobility group protein HMG I(Y) is required for NF-kappa B-dependent virus induction of the human IFN-beta gene. *Cell* 71:777–789
- Thomas KR, Folger KR, Capecchi MR (1986) High frequency targeting of genes to specific sites in the mammalian genome. *Cell* 44:419–428
- Thompson EM, Christians E, Stinnakre MG, Renard JP (1994) Scaffold attachment regions stimulate HSP70.1 expression in mouse preimplantation embryos but not in differentiated tissues. *Mol Cell Biol* 14:4694–4703
- Thorvaldsen JL, Duran KL, Bartolomei MS (1998) Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev* 12:3693–3702
- Torrunguang K, Alvarez M, Shah R, Onyia JE, Rhodes SJ, Bidwell JP (2002) DNA binding and gene activation properties of the Nmp4 nuclear matrix transcription factors. *J Biol Chem* 277:16153–16159
- Varghese J, Alves W, Brill BJ, Wallace M, Calabrese D, Regamey A, Girod PA. (2008) Rapid development of high-performance, stable mammalian cell lines for improved clinical development. *Bioprocess J* 7:30–36
- Volfson D et al (2006) Origins of extrinsic variability in eukaryotic gene expression. *Nature* 439:861–864
- Wakimoto BT (1998) Beyond the nucleosome: epigenetic aspects of position-effect variegation in *Drosophila*. *Cell* 93:321–324
- Wang T, Xue L, Hou W, Yang B, Chai Y, Ji X, Wang Y (2007) Increased expression of transgene in stably transformed cells of *Dunaliella salina* by matrix attachment regions. *Appl Microbiol Biotechnol* 76:651–657

- West A, Gaszner M, Felsenfeld G (2002) Insulators: many functions, many mechanisms. *Genes Dev* 16:271–288
- Wilson C, Bellen HJ, Gehring WJ (1990) Position effects on eukaryotic gene expression. *Annu Rev Cell Biol* 6:679–714
- Woodcock CL, Dimitrov S (2001) Higher-order structure of chromatin and chromosomes. *Curr Opin Genet Dev* 11:130–5
- Xu Y, Davidson L, Alt FW, Baltimore D (1996) Deletion of the Ig kappa light chain intronic enhancer/matrix attachment region impairs but does not abolish V kappa J kappa rearrangement. *Immunity* 4:377–385
- Yamasaki K, Akiba T, Yamasaki T, Harata K (2007) Structural basis for recognition of the matrix attachment region of DNA by transcription factor SATB1. *Nucleic Acids Res* 35:5073–5084
- Yasui D, Miyano M, Cai S, Varga-Weisz P, Kohwi-Shigematsu T (2002) SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* 419:641–645
- Yoshikawa T, Nakanishi F, Ogura Y, Oi D, Omasa T, Katakura Y, Kishimoto M, Suga K (2000) Amplified gene location in chromosomal DNA affected recombinant protein production and stability of amplified genes. *Biotechnol Progr* 16:710–715
- Zahn-Zabal M, Kober M, Girod PA, Imhof M, Chatellard P, de Jesus M, Wurm F, Mermod N (2001) Development of stable cell lines for production or regulated expression using matrix attachment regions. *J Biotechnol* 87:29–42
- Zhao K, Kas E, Gonzalez E, Laemmli UK (1993) SAR-dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG-I/Y is enriched in H1-depleted chromatin. *Embo J* 12:3237–3247