Chromosome Research **11**: 435–445, 2003.

From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics

J. Bode[1], S. Goetze[1], H. Heng[2,3], S. A. Krawetz[2,4] & C. Benham[5]
[1]*GBF–German Research Center for Biotechnology/Epigenetic Regulation, Mascheroder Weg 1, D-38124 Braunschweig; Tel: +49 531 6181 251; Fax: +49 531 6181 251; E-mail: jbo@gbf.de;* [2]*Center for Molecular Medicine and Genetics;* [3]*Department of Pathology, Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, MI, USA;* [4]*Obstetrics & Gynecology, Institute for Scientific Computing;* [5]*UC Davis Genome Center, Davis, CA, USA*

Key words: chromatin, DNA structure, fluorescence, gene expression, *in-situ* hybridization, lamins, nuclear matrix, nucleosome

Abbreviations: BUR, base-unpairing region; CUE, core-unpairing element; FISH, fluorescence *in-situ* hybridization; IFN, interferon; LCR, locus-control region; SBS, SATB1-binding sequences; SIDD, stress-induced duplex destabilization; S/MAR, scaffold-matrix attachment region; UE, unpairing elements

Abstract

Eukaryotic genomes are functionally compartmentalized into chromatin domains by their attachment to a supporting structure that has traditionally been termed the nuclear matrix. Present evidence indicates the dynamics of this entity, which requires particular properties of the elements that mediate this kind of interaction. Above all, this is enabled by the so-called 'mass binding phenomenon' by which scaffold/matrix-attachment regions (S/MARs) reversibly associate with ubiquitous factors. Recent investigations and novel techniques have shown that these contacts can be altered by modulators as well as by specific interactions with the components of enhancers and locus control regions.

Introduction

In eukaryotes the regulation of transcription is determined, to a large extent, by an interplay of chromatin structure and nuclear position. This becomes particularly evident in cases where transgenes are introduced, for the efficient production of relevant proteins. During recent years, the chromatin field has matured to a level where it contributes major improvements to eukaryotic expression systems. Understanding how genes are differentially expressed in the nuclear context requires in-depth knowledge of the functional compartmentalization of chromosomes. We will concentrate on distinct principles of nuclear architecture and describe the cellular tools that target parts of the genome to nuclear compartments so as to become less mobile than others. Particular attention is paid to the elements that are used to tether genes to nuclear structures, to their characterization, molecular properties and performance.

Architectural principles of nuclei

According to a traditional but still controversial model of nuclear organization, eukaryotic genomes

are functionally compartmentalized into chromatin domains by their attachment to a supporting structure that consists of protein and RNA. This structure has been termed the nuclear skeleton, nuclear scaffold or nuclear matrix. Since there is increasing evidence for the existence of such an entity but at the same time observations underlining its dynamic nature, we prefer 'matrix' (a dimensional field of variables) over the alternative terms that imply rigidity.

A nuclear matrix is thought to provide the structural basis for intranuclear order. By all available techniques, it consists of identical major components: the nuclear lamina, an inner network comprising core filaments and a more diffuse portion (Wan *et al*. 1999). Whether the entity is seen as a stable or a transient transcription-dependent structure depends – to a large extent – on the method applied for its characterization. What is known with certainty is the fact that the genome itself is segmented at 5–200-kb intervals by elements called 'scaffold-attachment' or 'matrix-associated regions' (S/MARs). We will describe the context-dependent properties of these elements as the tools mediating a permanent/constitutive or a facultative (activity-dependent) attachment. These features are dependent on binding mode, binding strength and genomic context. They will finally be amalgamated into a model that considers all modes of association (Figure 2 below).

Dominant signals in the eukaryotic genome

Eukaryotic chromosomes are organized in a hierarchy of levels. Among these, the nucleosome is one of the first. Recent observations suggest that, in addition to sequence-dependent features and occupied factor binding sites, there is yet another determinant that localizes the nucleosomes along genomic DNA: DNA bend sites occur once per four nucleosomes on average (Kiyama & Trifonov 2002). These repetitive sites not only play a key role in nucleosome positioning but also in the subsequent orders of chromatin folding. The periodicity may be locally disrupted at sequences with regulatory potential, for instance at sites of increased DNAse I accessibility (hypersensitive sites). These regions may lead to the omission of a bend, placing them into a region of weak phasing activity, concomitant with increased transcription efficiencies. Conversely, silencers frequently overlap these bent regions. It is remarkable that these findings are compatible with the genome segmentation rules that describe eukaryotic genomes as blocks of 350 bp (two nucleosomes or 120 amino acids) as a consequence of ancient combinatorial fusions (Trifonov 1995, Trifonov *et al*. 2001). The fusion points are usually marked by a clustering of Met codons indicative of primordial unit size genes. This general design has been maintained by transposons which tend to profit from it in various ways (see below).

By combining *in-vitro* assays with a biomathematical algorithm, we have recently obtained evidence for yet another structural level that may rank next in this hierarchy (Goetze *et al*. 2003). In extended non-coding regions, we detected restricted DNA segments for which the DNA duplex is destabilized. These peaks, which have been designated 'UEs' (unpairing elements; Bode *et al*. 1992), occur at a phase of about 3000 bp. Some of these UEs correspond to the points where DNA is attached to the nucleoskeleton which in turn may give rise to nuclease-hypersensitive or fragile genomic sites (Bode *et al*. 2000). It is generally accepted that strand separation can occur under the conditions of superhelical strain and in the absence of dsDNA binding proteins.

Using the Stress-Induced-Duplex-Destabilization (SIDD) algorithm, UEs can now routinely be localized by computational analysis. The corresponding SIDD profiles are calculated using a statistical mechanical procedure in which superhelical deformation is partitioned between strand separation, twisting within denatured regions and residual superhelicity (Goetze *et al*. 2003).

Scaffold/matrix attachment regions

S/MARs of more than 1 kb have been considered as chromatin domain borders, playing a critical role in nuclear architecture and function. They are thought to organize the 25 million nucleosomes in a mammalian nucleus into approximately 60 000 chromatin loops with an average of 70 kb. Active genes tend to be part of looped domains as small as 4 kb whereas inactive regions are accomodated as units of up to 200 kb (Bode *et al*. 2000). A number of *in-vitro* assays have been developed to search for

S/MARs. Significantly, all these assays identify the same fundamental class of anchorage sequences.

From the UE to a BUR and beyond

In essentially all cases examined to date, S/MARs coincide with SIDD regions of eukaryotic DNA that are predicted to be extensively destabilized (Goetze *et al*. 2003). Typically, these 'base-unpairing regions' (BURs) consist of multiple UEs spaced according to distinct criteria (Figure 1B). This facilitates the multisite attachment of the responsible proteins or protein complexes (see scheme in Figure 1A). The *in-vivo* and *in-vitro* properties of an S/MAR from the upstream border of the human interferon-β gene (AC G000308 SM S/MAR: SM0000002; HS$IF NB1-ESM S/MAR: SM0000005; HS$IFNB1-GSM S/MAR: SM0000006; HS$IFNB1-K, SM S/MAR: SM0000008; HS$IFNB1-I SM S/MAR: SM0000009; HS$IFNB1-H,SM S/MAR: SM0000010; HS$IFNB1-D) have been extensively studied. The element comprises a pronounced 2.2-kb base-unpairing region (BUR) with a well-defined nucleation center (core-unpairing element, CUE) and a series of evenly spaced additional UEs (Goetze *et al*. 2003 and references therein).

These base-unpairing regions were initially detected by applying a variety of single-strand specific agents to supercoiled plasmid DNA or to DNA in living cells which contains supercoiled portions as well. Among these, haloacetaldehydes show a preferred reactivity at the sites of an S/MAR (Bode *et al*. 1995). Related results have been described with osmium tetroxide (OsO_4) which footprints certain AT-rich S/MARs (Paul & Ferl 1993). Potassium permanganate ($KMnO_4$) is a still more vigorous agent which, in contrast to the former ones, marks regions with a pronounced strand separation potential but shows little discrimination for the state of DNA *in vivo* and *in vitro*. We could in fact show that the pattern of $KMnO_4$ reactive sites closely resembles the pattern of UEs in a SIDD profile (Knopp 2001). Before knowledge about these properties emerged, S/MARs have been characterized by a variety of

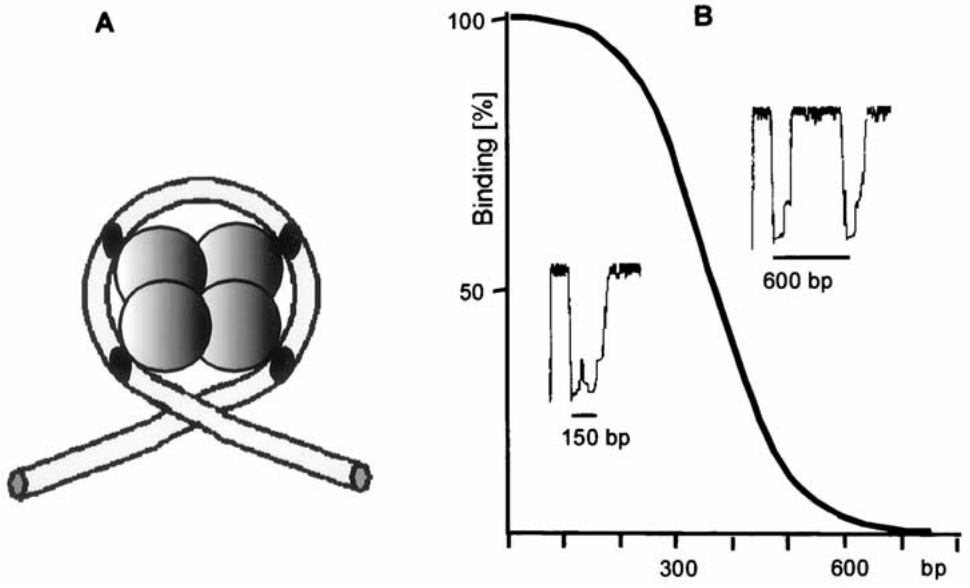

Figure 1. S/MAR-proteins interactions: Prerequisites for mass-binding. (**A**) Binding of SAF-A complexes has been suggested to involve regularly-spaced AT patches which are thought to represent unpairing elements (UEs). Each protein contains a single SAF-box which mediates weak binding. Only an ordered protein complex can mediate strong, though dynamic, binding in the sense that individual contacts can be broken and reformed after the passage of a tracking protein like RNA polymerase. (**B**) Communication between individual UEs is required to create activity. Dimerization of a sub-S/MAR size fragment (130 bp) from the immunoglobulin kappa S/MAR creates an element associating with the nuclear scaffold. S/MAR activity is gradually lost as spacers between the unpairing regions are extended (Tsutsui 1998, Bode *et al*. 2003).

classical procedures that extract histones and other soluble factors from nuclei. Using a combined *in-silico/in-vitro* approach, we are now in the position to narrow down the responsible sequences and to derive the rules that govern reassociation strength (Kramer & Krawetz 1996, Kramer *et al.* 1998a, Goetze *et al.* 2003). Binding affinity appears to be an important predictor for the biological activity of an S/MAR in a number of functional tests. These assays revealed S/MAR-associated properties that are in full agreement with the postulated roles of an enhancer-supporting structure, a boundary element, an insulator and an ARS-like sequence in higher eukaryotic cells (Bode *et al.* 2003).

The particular mode of S/MAR–matrix interaction indicates that binding is not directly correlated to the primary sequence but rather to the structure-forming potential of DNA (Bode *et al.* 1998). Already, the first bioinformatic approach ever devoted to S/MAR prediction, the 'MAR-finder' algorithm (Singh *et al.* 1997) was based on the statistical occurrence of motifs that are related to DNA secondary structure. Among these are spacing criteria for AT-patches similar to those depicted in Figure 1 (Tsutsui 1998), sequences that are abundant in 3′UTRs, curvature/kink-inducing motifs, topoisomerase II recognition sites and motifs found at origins of replication (ORIs) indicating the scope of activities in which S/MARs may have a function.

In the special case of ORIs, it has been known for some time that a DNA unwinding element is required for its function. Frequently, these elements comprise inverted repeats which have found their perfection in viral systems where they are easily torsionally strained to form cruciform structures due to the action of a tracking protein. This process can be supported by retrieving the energy stored in a nearby base-unpaired region. As a whole, the relation of S/MARs and ORIs is remarkable. S/MARs have been mapped adjacent to ORIs where they mediate function (Pemov *et al.* 1998) and an episomally replicating vector has been described, the performance of which depends on an S/MAR element (Lipps *et al.* 2002).

Structural networks: The mass binding mode

S/MARs are usually associated with proteins whose sequence recognition requirements are rather relaxed as a consequence of their secondary structure-forming potential. We have described the ways that known interacting partners recognize and interact with BUR-related features. These either directly involve the strand-separation phenomenon or rely on BUR-related structures (Goetze *et al.* 2003). In the present context, we will only consider the most ubiquitous representatives, i.e. the lamins and the most prominent scaffold-attachment factor, SAF-A for which new information has emerged. In addition, we will discuss one cell-type-specific factor, SATB1, as an example of ongoing studies illuminating the way that constitutive activities of S/MARs can be modulated. Common to all these proteins is the 'mass-binding mode' according to which low-affinity binding of many individual protein molecules results in strong and specific interactions that can serve specific regulatory roles.

Lamins

The nuclear lamina, a protein meshwork underlying the nuclear membrane, is primarily comprised of the intermediate filament proteins, lamins A, B and C. It came as a surprise when Hozak *et al.* (1995) showed that these components not only form the lamina but also part of a diffuse skeleton, called a 'veil' that branches throughout the interior of the nucleus. Unlike their peripheral counterparts, these internal lamins are buried in dense patches of chromatin and are thereby inaccessible to antibodies. It is only after the removal of chromatin, that the knobs and nodes on an internal skeleton can be immunolabelled with lamin A specific antibodies (Neri *et al.* 1999).

B-type lamins were actually the first members of the group to be identified as a binding component of the S/MAR that is dependent on their polymerization status (Ludérus *et al.* 1992). The polymerized alpha-helical rod contributes the specific binding whereas the unpolymerized domain does not bind. A-type lamins and the structurally related proteins desmin and NuMA also specifically associate with S/MARs *in vitro* and form internal networks at least under conditions of overexpression (Harborth *et al.* 1999).

Competition studies revealed the existence of two different types of interactions reflecting different structural features of S/MARs, i.e. the minor groove of double-stranded DNA or single-stranded

regions. It appears likely that the interaction of nuclear matrix proteins with the single-stranded S/MAR regions serves to stabilize the transcriptionally active state of chromatin (Ludérus *et al*. 1992, 1994). In the living cell, the peripheral nuclear lamina consists of lamin filaments and integral membrane proteins, including the lamin B receptor and lamina-associated polypeptides (LAPs) that in turn interact with chromatin proteins. Structural links between the peripheral lamina and the internal nuclear matrix are thought to be mediated by LAP2 family members, filament protein Tpr and nucleoporin Nup153.

The dynamic nature of lamins during interphase has been derived from gfp-lamin fusions (Goldman *et al*. 2002). During the initial stages of G1, gfp-lamin B1 fluorescence fully recovers 10 min after photobleaching while requiring 60 min after nuclei have assembled. In the case of lamin A, an intense nucleoplasmic and lamina fluorescence is observed in early G1. However, these signals show vastly different recovery rates with half lives of a few seconds or 90 min, respectively. This indicates that lamin A exists in two very different states of assembly. Later in G1, the gfp-lamin A and B proteins form a stable nucleoplasmic structure (veil). Studies with dominant-negative mutants lacking their aminoterminal domains show that both DNA synthesis and transcription by RNA polymerase II (but not of polymerases I and III) become severely impaired. This indicates that lamins play an important role in the functional organization of interphase chromatin. Transcriptionally active genes and transcription factors have been identified in lamin-enriched nuclear matrix preparations suggesting that lamins are parts of a matrix which provides for the assembly or stabilization of active transcription complexes.

SAF-A

In addition to lamins, the internal nuclear matrix contains hnRNP proteins associated with transcription, transport, and processing of hnRNA. This is consistent with the view that the internal nuclear matrix supports or even results from these processes. The most prominent S/MAR-binding member of this group, scaffold-attachment factor A (SAF-A), also known as SP120 (Tsutsui 1998) or hnRNP-U (Goehring & Fackelmayer 1997) associates with multiple S/MAR elements. UV-crosslinking has shown that this RNA binder is likewise associated with DNA *in vivo*. S/MAR binding is optimal when 'AT patches' (short tracts of consecutive As and Ts which most likely correspond to the unwinding elements discussed above) are distributed according to the rules specified by Tsutsui (1998). Interestingly, it is not the AT-richness *per se* but the geometry by which these AT-tracts/UEs are distributed, that influences binding: SAF-A shows a pronounced propensity to self-polymerize on S/MAR-DNA and this state explains the positive cooperativity of the association process (Figure 1).

The protein's primary structure reflects its dual role. There are two independent nucleic acid binding domains: (1) a C-terminal RNA/ssDNA binding domain (RGG box), and (2) an S/MAR specific N-terminal domain, called 'SAF-box'. The latter has recently been localized to 31 amino acids that reside in a part of the protein cleaved off during apoptosis (Kipp *et al*. 2000, Goehring & Fackelmayer 1997). The SAF-box recurs in several eukaryotic proteins and represents the first characterized protein domain specifically recognizing S/MARs.

A characteristic association of a cloned SAF-box with S/MARs can be demonstrated in pull-down experiments but only when a critical protein density is reached on the surface of Sepharose-beads (Kipp *et al*. 2000). According to the 'mass-binding' principle, each individual domain interacts just weakly with an AT-patch/UE and it is only the simultaneous binding of multiple SAF-boxes that confers a strong specific interaction. This is consistent with the observation that there are few naturally occurring S/MARs below a critical length of 250 bp. The failure of ssDNA to compete for the interaction of S/MARs with SAF-boxes shows that many but not all criteria of scaffold–S/MAR interactions can be explained by SAF-box proteins. The 50% competition limit found for the complete scaffold may therefore reflect the contribution of the lamins (Bode *et al*. 1998).

What is the organization of SAF-A within the cell? Recent experiments have shown that fluorescently labeled constructs of SAF-A decorate an intranuclear network, both in living cells and in permeabilized nuclei, quite similar to the nuclear matrix network visualized by electron microscopy. In addition, SAF-A can be used as a marker protein

to visualize a filamentous salt-stable entity in *in-vitro* reconstituted nuclei, suggesting that the protein may at least be a major component of a nuclear matrix (Frank Fackelmayer, personal communication).

Martens *et al*. (2002) have observed a strong interaction between a histone acetyltransferase (p300) and SAF-A. This provides some of the first evidence for how SAF-A may mediate some of the positive effects on transcriptional rate and long-term stability that have been ascribed to S/MARs. While p300 does interact with SAF-A in association with an S/MAR, it does not if it is associated with hnRNP particles. The proposed model states that p300, by forming a bridge between SAF-A at S/MAR elements and the transcription factors, enables a quick and complete histone acetylation once transcription is to be initiated. It should be noted that this configuration requires an already potentiated domain as described by Kramer *et al*. (1998b).

Functional networks: The SATB1 paradigm

S/MARs either colocalize with the boundaries of nuclease-sensitive chromatin domains or with transcriptional control sequences and may thus serve multiple functions. A very prominent model is the immunoglobulin heavy chain locus, which is regulated by a transcribed intronic enhancer that is flanked by S/MAR elements. Interestingly, these S/MARs comprise binding sites for negative regulatory factors which are thought to repress the enhancer, by interfering with S/MAR functions, in cell types where this locus is not active. These enhancer-associated attachment regions are essential for μ-gene transcription in transgenic mice that is accompanied by the S/MAR-dependent formation of a DNAse I-sensitive domain extending from the hypersensitive enhancer (Forrester *et al*. 1999). This function could be the simple consequence of S/MARs being a sink of negative superhelicity that can assist unwinding adjacent genomic districts or be due to a 'domain opening' function in the framework of the model developed by Laemmli and colleagues (Zhao *et al*. 1993).

At these loop-internal positions, S/MAR functions are mediated by a variety of factors that affect transcription in either a positive or negative manner. In many cell types, they are occupied by a negative regulatory factor (NF-μNR; Scheuermann & Chen 1989), which in late-stage B cells is displaced by Bright, a transcriptional activator (Herrscher *et al*. 1995). Later studies have demonstrated that only in conjunction with the S/MARs is the enhancer able to mediate the accessibility and activation of distal promoters which was ascribed to effects both on DNA demethylation and histone acetylation (Fernandez *et al*. 2001). In T-cells, the same site is occupied by a special AT-binding protein (SATB1). In the following, we will briefly illuminate the regulatory network which is triggered by this factor.

Although associating with BURs, i.e. stretches with strand-separating potential, SATB1 has been described as a protein that binds to dsDNA. In mouse thymocyte nuclei, the protein forms a 3D network that regulates genes by folding chromatin into loop domains by tethering specialized DNA elements (Yasui *et al*. 2002). Binding specificity is mediated by an atypical homeodomain and two Cut-like repeats within a specialized ATC context that readily becomes base-unpaired in supercoiled DNA. Using the ChIP-assay, 16 SATB1-binding sequences (SBS) of 100 bp to 1.1 kbp (SBS1-16; all of them S/MARs by the common criterion) were isolated from T-lymphoid cells and shown to conform to the ATC rule (deBelle *et al*. 1998).

SATB1 was initially described as a protein interfering with the transcription of S/MAR reporter constructs (Kohwi-Shigematsu *et al*. 1997). However, this is clearly more complex. In $SATB1^{-/-}$ thymocytes, the SATB1-bound genomic loci are detached from the nuclear matrix. In comparison, T cells in SATB1 knockout mice are arrested at the 'double positive' ($CD4^{+}CD8^{+}$) stage compromising their immune system which leads to lethality at week 3. The defect is amplified by the temporal and spatial mis-expression of numerous genes including the normally silent IL-2 receptor gene (Alvarez *et al*. 2000). The results of this and other studies suggest that the action of SATB1 depends on the site of action where it can assemble either a transcriptionally active complex or a silencing complex depending on the proteins by which it is recruited. When the above-mentioned SBS sequences were used as probes to localize SATB1 in nuclear matrices, which had been depleted of endogenous DNA, a typical probe produced only two signals (Cai & Kohwi-Shige-

matsu 1999) indicating that a specific subset of SATB1 is labeled by each SBS probe. Nevertheless, SATB1 appears to function as a 'landing platform' for several ATP-dependent chromatin-remodeling complexes where it sets the stage for either activation or repression, e.g. HDAC (histone-deacetylases; Yasui *et al*. 2002).

Nuclear matrix and chromatin domains

Evidence for the formation of chromatin loops by the attachment to a supporting structure is multifold. Evidence emerged from an *in-vivo* crosslinking strategy with cis-dichlorodiammine-platinum (cis-DDP). This reagent has a high specificity for S/MAR matrix interactions capturing the above-mentioned proteins in addition to a number of cell-type specific factors (Ferraro *et al*. 1996). Second, the use of topoisomerase II inhibitors introduces cuts at S/MARs which are the prototype sites of the enzyme's action. This leads to the excision of complete chromatin domains (Yarovaya & Razin 1998). Third, novel halo-FISH procedures have been developed, permitting the localization of S/MARs relative to the matrix (Schmid *et al*. 2001, Heng *et al*. 2001 and submitted). These techniques enable the visualization of specified chromatin loops and show that S/MARs are in fact the elements delimiting a domain, although in a dynamic manner.

Each chromatin domain is considered as an autonomously regulated unit in the eukaryotic genome. Thereby, an enhancer situated within the domain acts on the promoter of the contained gene but not beyond the domain's boundaries (insulation phenomenon). It has been demonstrated that the borders of 'prototype domains' are regions of decreasing DNAse I accessibility which in turn coincide with S/MARs (Kramer & Krawetz 1996, Antes *et al*. 2001). Some of these S/MARs are in fact capable of restricting the action of an enhancer and, by definition, are insulators, although there are other examples where a specialized insulator element is required. Insulator–insulator interactions can eliminate such an activity (West *et al*. 2002), whereas S/MARs permit regulatory elements to approach each other on the nuclear matrix increasing the efficiency of their interaction (Antes *et al*. 2001).

Whether a given domain remains in the potentiated-open (active, histone hyperacetylated) or closed (inactive, heterochromatic, histone H3 and DNA-methylated) state is determined by a central switch, the locus-control region. LCRs can be composed of multiple constituents, including enhancers, insulators and tissue-specific boxes, that can be clustered together as in the prototype human domain of β-globin genes or scattered throughout the domain as exemplified by the chicken lysozyme locus. More than 30 units have been described for mammals that appear to meet the criteria of an LCR. Each of its components differentially affects gene expression and it is their interaction and collective action that functionally defines an LCR. In their studies on the enhancer-mediated control of the immunoglobulin μ-gene, Fernandez *et al*. (2001) have demonstrated that the combination of enhancer and S/MARs constitutes a functional LCR suggesting that S/MARs may be one class of the non-enhancer elements that are implicated in LCR function.

Dynamic aspects: A model

At the outset of this review, we have discussed dominant DNA structures in the genome. Recent data suggest that, evolutionarily, dominant DNA structures in the genome and especially regions with base-unpairing propensity may have been one of the first prerequisites of gene expression. For the yeast CUP1 promoter, Leblanc *et al*. (2000) have demonstrated that RNA polymerase II can initiate transcription if the template is simply negatively supercoiled. In this case, the CUP1 initiation element is intrinsically unstable, permitting strand separation and the entry of the polymerase. Thus, transcription occurs independent of a TATA box or transcription factors. Parallel studies on the strand separation potential clearly show the CUP1 initiation element to be intrinsically unstable, permitting the entry of the polymerase. These results, together with a prior analysis of SIDD characteristics of regions surrounding several yeast genes (Benham 1996), supports the idea that primitive promoters are stress points within the DNA double helix. One can then pose that transcription factors were introduced to regulate promoter melting and thus gene expression. Such a refined process may involve binding to remote UEs and a

subsequent transfer of the unwound state to the critical promoter sequences.

Prokaryotes use gyrases to induce negative superhelicity but lack of such an enzyme in eukaryotes requires different actions. Regarding the importance that has been recently assigned to histone acetylation (Struhl 1998), one of the first proposals for the way eukaryotes generate superhelicity merits particular attention: the finding by Norton *et al*. (1989) that, with increasing levels of acetylation, there is a decrease of the negative linking number change from -1.04 to -0.82. In this manner, previously constrained negative supercoils are released. These processes may be supported by ATP-dependent chromatin-remodeling machines, which convert the energy from ATP hydrolysis into superhelical stress leading to the dissociation of a segment of DNA from the histone surface (Laengst & Becker 2001).

For plasmids, superhelicity is constrained by their covalently closed nature whereas higher eukaryotic genomes have to use the looped domain structure for this purpose. Failure to effect superhelicity may be the main reason that S/MARs do not exert a transcriptional effect in transient expression experiments. This is in contrast to the transcriptional augmentation that is clearly observed in the context of a circular episome (Hoertnagel *et al*. 1995). Although this is a common S/MAR effect, even this influence depends on the context (Schuebeler *et al*. 1996): In a series of model experiments using the IFN-β S/MAR, transcriptional initiation was supported by an S/MAR 4 kb downstream from the promoter, whereas, at distances below 2.5 kb, transcription was essentially shut off. The S/MAR itself was transcribed, i.e. there was no pausing or premature termination. Thus, it has been suggested that the protein contacts can only be broken at a certain superhelical tension and that binding partners change according to the topological status. The ability of RNA polymerase to read through an S/MAR is essential for S/MARs localized within introns. It has been hypothesized that long transcription units must be provided with the capability of dynamically forming loops as the polymerase

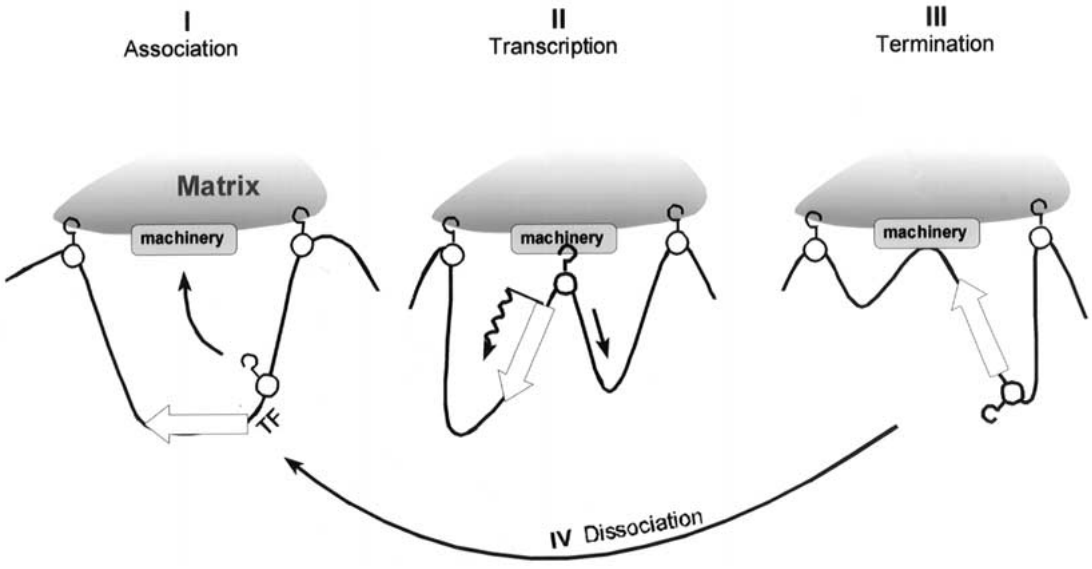

Figure 2. S/MAR-functions: constitutive and facultative. A chromatin domain with constitutive S/MARs at its termini (**I**). When functional demands require the specific translocation of the constituent gene to the matrix, the facultative S/MAR responds to topological changes which are initiated by the association of transcription factors and supported by histone hyperacetylation. Topological changes are propagated once the gene is pulled through the transcriptional machinery (**II**). This model is consistent with previous observations on the induction of the human IFN-β gene which is accompanied by an increased binding of an S/MAR immediately downstream from the gene (Bode *et al*. 1995). At the same time, the coding region becomes DNase I sensitive and the flanking S/MARs become more tightly bound and thus more resistant to DNase I digestion. Transcription is terminated by dissociation (**IV**), i.e. restoration of the silent but transcriptionally poised chromatin state (**I**).

proceeds (Mielke *et al*. 2002). By using facultative S/MARs, even simple genes have the capability to reduce loop size at times of transcriptional activity (Bode *et al*. 1995; Figure 2).

For a better understanding of S/MAR-dynamics, we have subjected clones which differ in the number of transfected IFN-β copies to halo-FISH analyses and visualized aspects of their interaction with the nuclear matrix *in situ*. To this end, interphase nuclei were gently fixed on a microscopic slide, extracted and stained with DAPI to reveal a circular internal portion (remnants of the nuclear structure harboring the nuclear matrix) and a brightly fluorescing halo (loops emanating between functional S/MAR–matrix contacts). DNA probes specific for the insert (red fluorescence) and, in addition, for the endogenous murine IFN-β (green fluorescence) were used to visualize the respective location of the genes. This procedure has led to the following conclusions (H. Heng *et al*. submitted):

- The proportion of matrix-associated DNA depends both on S/MAR-binding strength and copy numbers;
- Although the association of S/MARs with the matrix is a dynamic process, low copy number transgenes (1–10) can be accommodated at the matrix most of the time;
- At higher copy number, the majority of genes partition into the halo portion whereby they lose transcriptional competence. Heterochromatization and methylation are well known consequences of high copy number constructs that lead to repeat-induced silencing (i.e. the cosuppression phenomena).

These analyses resolve a long-standing contention why S/MARs meet only one out of two criteria that have been used to define an 'insulator' function. They clearly reduce clone-to-clone variation of expression while they do not mediate a copy-number-dependent expression when co-integrated at a single genomic site (which is the consequence of standard transfection and transgenic protocols; Bode *et al*. 2003).

In this article we have highlighted some aspects of nuclear organization that can be exploited for the predictable expression of foreign genes. In accord with the concepts of Trifonov (1995) vectors with a transposon-like structure will have a better chance of being expressed since their architecture complies with the rules of structural harmony and - last but not least - apparently also due to the preference of these elements for targets with secondary structures that are otherwise typical for scaffold/matrix attachment regions (Tikhonov *et al*. 2001).

References

Alvarez JD, Yasui DH, Niida H, Joh T, Loh DY, KohwiShigematsu T (2000) The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Gene Dev* **14**: 521–535.

Antes TJ, Namicu SJ, Fournier REK, Levy-Wilson B (2001) The 5′ boundary of the human apolipoprotein B chromatin domain in intestinal cells. *Biochemistry* **23**: 6731–6742.

Benham CJ (1996) Duplex destabilization in superhelical DNA is predicted to occur at specific transcriptional regulatory regions. *J Mol Biol* **255**: 425–434.

Benham C, Kohwi-Shigematsu T, Bode J (1997) Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions. *J Mol Biol* **274**: 181–196.

Bode J, Kohwi Y, Dickinson L *et al*. (1992) Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science* **255**: 195–197.

Bode J, Schlake T, Rios-Ramirez M *et al.* (1995) Scaffold matrix-attached regions: Structural properties creating transcriptionally active loci. *Int Rev Cytol* **162A**: 389–454.

Bode J, Bartsch J, Boulikas T *et al.* (1998) Textbook of gene therapy and molecular biology. Transcription-promoting sites in mammalia: Their elucidation and architectural principles. *Gene Ther Mol Biol* **1**: 551–580.

Bode J, Benham C, Ernst E *et al.* (2000) Fatal connections: When DNA ends meet on the nuclear matrix. *J Cell Biochem* **35** Suppl.: 3–22.

Bode J, Goetze S, Ernst E *et al.* (2003) *New Comprehensive Biochemistry – Gene Transfer and Expression in Mammalian Cells Architecture and Utilization of Highly-expressed Genomic sites*. Amsterdam: Elsevier, (in press).

Cai ST, Kohwi-Shigematsu T (1999) Intranuclear relocalization of matrix binding sites during T cell activation detected by amplified fluorescence in situ hybridization. *Methods* **19**: 394–402.

deBelle I, Cai ST, Kohwi-Shigematsu T (1998) The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) *in vivo* in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops. *J Cell Biol* **141**: 335–348.

Fernandez LA, Winkler M, Grosschedl R (2001) Matrix attachment region-dependent function of the immunoglobulin mu. *Mol Cell Biol* **21**: 196–208.

Ferraro A, Cervoni L, Eufemi M, Altieri F, Turano C (1996) Comparison of DNA–protein interactions in intact nuclei from avian liver and erythrocytes: A cross-linking study. *J Cell Biochem* **62**: 495–505.

Forrester WC, Fernandez LA, Grosschedl R (1999) Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions. *Gene Dev* **13**: 3003–3014.

Goehring F, Fackelmayer FO (1997) The scaffold/matrix attachment region binding protein hnRNP-U (SAF-A) is directly bound to chromosomal DNA in vivo: A chemical cross-linking study. *Biochemistry* **36**: 8276–8283.

Goetze S, Gluch A, Benham C, Bode J (2003) Computational and in vitro analysis of destabilized DNA regions in the interferon gene cluster: the potential of predicting functional gene domains. *Biochemistry* USA **42**: 154–166.

Goldman RD, Gruenbaum Y, Moir RD, Spann DK, Shumker TP (2002) Nuclear lamins: building blocks of nuclear architecture. *Genes Dev* **16**: 533–547.

Harborth J, Wang J, GuethHallonet C, Weber K, Osborn N (1999) Self assembly of NuMA: multiarm oligomers as structural units of a nuclear lattice. *Embo J* **18**: 1689–1700.

Heng HH, Krawetz SA, Lu W, Breer S, Liu G, Ye CJ (2001) Re-defining the chromatin loop domain. *Cytogenet Cell Genet* **93**: 155–161.

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.

Hoertnagel K, Mautner J, Strobl LJ *et al.* (1995) The role of immunoglobulin kappa elements in c-myc activation. *Oncogene* **10**: 1393–1401.

Hozak P, Sasseville AMJ, Raymond Y, Cook PR (1995) Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells. *J Cell Sci* **108**: 635–644.

Kipp M, Goehring F, Ostendorp T *et al.* (2000) SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Mol Cell Biol* **20**: 7480–7489.

Kiyama R, Trifonov EN (2002) What positions nucleosomes? *FEBS Lett* **523**: 7–11.

Knopp A (2001) *Studien zur Struktur von S/MARs am Beispiel einer genomischen Domaene aus dem humanen Interferon-Gencluster*. Dissertation, Universitaet Braunschweig.

Kohwi-Shigematsu T, Maass K, Bode J (1997) A thymocyte factor SATB1 suppresses transcription of stably integrated matrix-attachment region-linked reporter genes. *Biochemistry* **36**: 12005–12010.

Kramer JA, Krawetz SA (1996) Nuclear matrix interactions within the sperm genome. *J Biol Chem* **271**: 11619–11622.

Kramer, JA, Adams, MD, Singh, GB, Doggett NA, Krawetz SA (1998a) A matrix associated region localizes the human SOCS-1 gene to chromosome 16p13.13. *Somatic Cell Mol Genet* **24**:131–133.

Kramer JA, McCarrey JR, Djakiew D, Krawetz SA. (1998b) Differentiation: the selective potentiation of chromatin domains. *Development* **125**: 4749–4755.

Laengst G, Becker PB (2001) Nucleosome mobilization and positioning by ISWI-containing chromatin remodeling factors. *J Cell Sci* **114**: 2561–2568.

Leblanc PB, Benham JC, Clark JD (2000) An initiation element in the yeast CUP1 promoter is recognized by RNA polymerase II in the absence of TBP if the DNA is negatively supercoiled. *Proc Natl Acad Sci USA* **97**: 10745–10750.

Lipps HJ, Jenke A, Nehlsen K, Scinteie M, Stehle IM, Bode J (2002) Chromosome-based vectors for gene therapy. *Gene* **304**: 23–33.

Ludérus MEE, De Graaf A, Mattia E *et al.* (1992) Binding of matrix attachment regions to lamin B1. *Cell* **70**: 949–959.

Ludérus MEE, Den Blaauwen JL, De Smit OJB, Compton DA, Van Driel R (1994) Binding of matrix attachment regions to lamin polymers involves single-stranded regions and the minor groove. *Mol Cell Biol* **14**: 6297–6305.

Martens JHA, Verlaan M, Kalkhoven E, Dorsman JC, Zantema A (2002) Scaffold/matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound to acetylated nucleosomes. *Mol Cell Biol* **22**: 2598–2606.

Mielke C, Christensen MO, Westergaard O, Bode J, Benham CJ, Breindl M (2002) Multiple collagen 1 gene regulatory elements have sites of stress-induced DNA duplex destabilization and nuclear scaffold/matrix association potential. *J Cell Biochem* **84**: 484–496.

Neri LM, Raymond Y, Giordano A, Capitani S, Martelli AM (1999) Lamin A is part of the internal nucleoskeleton of human erythroleukemia cells. *J Cell Physiol* **178**: 284–295.

Norton VG, Imai BS, Yau P, Bradbury EM (1989) Histone acetylation reduces nucleosome core particle linking number change. *Cell* **57**: 449–457.

Paul A-L, Ferl RJ (1993) Osmium tetroxide footprinting of a scaffold attachment region in the maize Adh1 promoter. *Plant Mol Biol* **22**: 1145–1151.

Pemov A, Bavykin S, Hamlin JL (1998) Attachment to the nuclear matrix mediates specific alterations in chromatin structure. *Proc Natl Acad Sci USA* **95**: 14757–14762.

Scheuermann RH, Chen U (1989) A developmental-specific factor binds to suppressor sites flanking the immunoglobulin heavy-chain enhancer. *Genes Dev* **3**: 1255–1266.

Schmid C, Heng HHQ, Rubin C, Ye CJ, Krawetz SA (2001) Sperm nuclear matrix association of the PRM1-PRM2-TNP2 domain is independent of Alu methylation. *Hum Mol Reprod* **7**: 903–911.

Schuebeler D, Mielke C, Maass K, Bode J (1996) Scaffold/matrix-attached regions act upon transcription in a context-dependent manner. *Biochemistry* **35**: 11160–11169.

Singh GB, Kramer JA, Krawetz SA (1997) Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucleic Acids Res* **25**: 1419–1425.

Struhl K (1998) DNA-methylation and histone-acetylation: Multistep model for generating an accessible chromatin domain. *Genes Dev* **12**: 599–606.

Tikhonov AP, Lavie L, Tatout C, Bennetzen JL, Avramova Z, Deragon JM (2001) Target sites for SINE integration in Brassica genomes display nuclear matrix binding activity. *Chromosome Res* **9**: 325–337.

Trifonov ENJ (1995) Segmented structure of protein sequences and early evolution of genome by combinatorial fusion of DNA elements. *J Mol Evol* **40**: 337–342.

Trifonov NE, Kirzhner A, Kirzhner MV, Berezovsky IN (2001) Distinct stages of protein evolution as suggested by protein sequence analysis. *J Mol Evol* **53**: 394–401.

Tsutsui K (1998) Synthetic concatemers as artificial MAR: importance of a particular configuration of short AT-tracts for protein recognition. *Gene Ther Mol Biol* **1**: 581–590.

Wan KM, Nickerson JA, Krockmalnic G, Penman S (1999) The nuclear matrix prepared by amine modification. *Proc Nat Acad Sci USA* **96**: 933–938.

West AG, Gaszner M, Felsenfeld G (2002) Insulators: many functions, many mechanisms. *Genes Dev* **16**: 271–288.

Yarovaya OV, Razin SV (1998) New approaches to studying the structural and functional organization of the eukaryotic genome. *Mol Biol Engl Tr* **32**: 36–43.

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.

Zhao K, Kaes 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.