Maintenance of a functional higher order chromatin structure: The role of the nuclear matrix in normal and disease states

Review Article

Amelia K. Linnemann1#, Stephen A. Krawetz1,2,3*
1The Center for Molecular Medicine and Genetics
2Department of Obstetrics and Gynecology
3Institute for Scientific Computing, Wayne State University School of Medicine

*Correspondence: Wayne State University School of Medicine, C.S. Mott Center for Human Growth and Development, 275 E Hancock, Detroit, MI 48201; Phone: (313)-577-6770; Fax: (313)-577-8554; Email: steve@compbio.med.wayne.edu
# Present Address: University of Wisconsin, Cardiovascular Research Center and Department of Pharmacology, 389 Medical Sciences Center, 1300 University Ave., Madison, WI 53706
Key words: S/MAR attachment, Nuclear Matrix, disease, scaffold, model, gene regulation, LIS, NaCl

Received: 14 May 2009; Revised: 29 June 2009; Accepted: 14 May 2009; electronically published: July 2009

Summary

The ordered packaging of DNA within the nucleus of somatic cells reflects a dynamic supportive structure that facilitates stable transcription interrupted by intermittent cycles of extreme condensation. This dynamic mode of packing and unpacking chromatin is intimately linked to the ability of the genome to specifically complex with both histones and non-histone proteins. Understanding the underlying mechanism that governs the formation of higher order chromatin structures is a key to understanding how local architecture modulates transcription. In part, the formation of these structures appears to be regulated through genomic looping that is dynamically mediated by attachment to the nuclear scaffold/matrix at S/MARs, i.e., Scaffold/Matrix Attachment Regions. Although the mechanism guiding the formation and use of these higher-ordered structures remains unknown, S/MARs continue to reveal a multitude of roles in development and the pathogenesis of disease.

I. Introduction

While the sequencing of the human genome has provided an invaluable resource for high throughput genomic studies, sequence alone has not been able to explain how the genome encodes the cellular program. This ability appears to be linked to the push–pull of compaction and potentiation, i.e., the formation of an open chromatin domain from a closed state. These changes in state are intimately linked to the ability of the DNA to form regulatory protein complexes. The ordered packaging of DNA within the nucleus must reflect a dynamic mechanism that supports stable transcription interrupted by intermittent cycles of condensation. On one hand, the four core histones and linker act to physically organize the eukaryotic DNA. On the other hand, they provide a medium to epigenetically reflecting the interphase condensation states that are recapitulated through many cycles of replication. The addition of modifiers at dozens of sites [reviewed in (Bernstein et al. 2007)] provides long-term epigenetic memory. These modifications contribute to cellular differentiation by marking loci for multi-generational activation or silencing (Jenuwein and Allis 2001). Working in concert as part of a combinatorial mechanism they only influence ~25% of gene expression (Wang et al. 2008). The question remains: what other controls participate to determine cell- specificity through transcription? The mechanics of transitioning between different states of condensation contributing to transcriptional control must be carefully orchestrated to maintain order and inhibit the intertwining of DNA fibers. Functional condensation is aided considerably because chromatin can be periodically constrained by a dynamic mesh of non- histone nuclear scaffold or matrix proteins. The nuclear scaffold/matrix was described well over 50 years ago (Zbarskii and Debov 1951), however its
characterization has remained elusive. As a result, the regions of the genome that bind to the nuclear scaffold/matrix, also known as S/MARs, are poorly defined at a genomic level. It has been suggested that S/MARs function to help maintain chromatin states while demarcating points of transition between states (Berezney and Coffey 1974; Mirkovitch et al. 1984). Interactions of the genome within this nuclear framework may coordinately act with histone modifications to generate the stable chromatin states that maintain expression or silencing. In concert they could ensure that this differentiated state is replicated.

II. Genome organization from nucleosome to chromosome territory

In somatic cells, the first order of DNA packaging is achieved by nucleosomes that wrap 146 bp in ~1.7 helical turns around an octamer of H2A
2, H2B
2, H3
2 and H4
2 proteins (Kornberg 1974; Luger et al. 1997). The nucleosomes are separated by 10 to 50 bp stretches of histone H1 bound linker DNA. Nucleosome positioning is influenced by several factors. These include the ability of the DNA to stretch and/or bend sharply around the histone core (Richmond and Davey 2003; Ong et al. 2007) and the direction of the bend of the major groove (Drew and Travers 1985). The energy of packaging is reduced within flexible AT rich regions and these sites of maximum curvature provide a means to precisely position nucleosomes. Local sequence composition can also influence nucleosome positioning despite the absence of a consensus binding sequence. Even with this level of understanding, only ~50% of the nucleosomal interactions can be predicted by sequence characteristics alone (Segal et al. 2006; Valouev et al. 2008).

In solutions of low ionic strength, nucleosomes for a 10 nm fiber that is visualized as ‘beads on a string’. The higher order folding of the 10 nm fiber becomes evident in the presence of the H1 linker (Yaneva et al. 1995) appearing as a 30 nm fiber (Shen et al. 1995). Several models of genome packaging are presented in Figure 1, although the one start and two start helical structures are still debated (Wu et al. 2007). The one-start helix, or solenoid, models are characterized by interactions between neighboring nucleosomes. This determines the dimensions of the coiled fiber as the DNA is bent into a solenoid (Finch and Klug 1976). The two start helix models are based on a nucleosomal arrangement that appears as a zigzag at low ionic strength (Worcel et al. 1981; Woodcock et al. 1984) with the linker maintaining a straight extended conformation (Williams et al. 1986; Smith et al. 1990). Recent evidence supporting the solenoid model suggests that nucleosome interdigitation among neighboring helical gyres creates a high-density nucleosome fiber (Robinson et al. 2006). However, these models fail to consider the effect of additional changes that are imposed upon condensation. For example, the effect of N-terminal acetylation of histone H4 that is necessary for the formation of the 30 nm fiber [reviewed in (Tremethick 2007)]. Irrespective, the 30 nm fiber is further condensed into a series of chromatin loops that are attached at their bases to non-histone nuclear scaffold or matrix proteins.

The first evidence for constrained looping of the 30 nm fiber in mammalian cells was based on sedimentation rates of nucleoids, i.e., nuclei depleted of proteins, through sucrose gradients (Cook and Brazell 1976). A similar study in Drosophila melanogaster suggested that there were approximately 400 nucleosomes or ~80 kb of DNA per loop (Benyajati and Worcel 1976). Looping was visualized by electron microscopy of histone depleted metaphase nuclei (Paulson and Laemmli 1977; McCready et al. 1979).

Figure 1: Folding of the 10 nm Fiber. In the presence of linker histone H1, the 10 nm fiber folds into the 30 nm a fiber. Several models of folding have been proposed. The one start helix model (bottom) requires the coiling of the 10 nm fiber around a central protein core into a solenoid. In comparison, the two-start helix model (top) requires two 10 nm fibers to associate, and form a secondary 30 nm fiber.
Several models have been proposed for the native organization of the chromosomal loops into higher order metaphase structures including the radial loop model of metaphase chromosome organization (Marsden and Laemmli 1979). Electron micrographs clearly showed the attachment of radially arranged loops to a central scaffold. Building upon this model, others proposed minibands for med as the 30 nm solenoid filaments were organized into 60 kb loops grouped 18 per turn (Pienta and Coffey 1984). Each mitotic chromatid was then derived from ~106 minibands arranged along a central axis.

It is now well-accepted that individual interphase chromosomes are organized into discrete territories [reviewed in (Cremer and Cremer 2001)] but the structural organization within the chromosome territories remains to be discerned. Figure 2 summarizes several models including the chromonema model (Belmont and Bruce 1994), the random walk/giant loop model (Sachs et al. 1995) and the multi-loop sub compartment model (Munkel et al. 1999). The chromonema model of organization was derived by following the various states of condensation and intermediates of interphase chromatin through G1. As G1 progressed, folded structures of chromonema fibers of approximately 100-130 nm in diameter decondensed to 60-80 nm. It was suggested that condensed chromonema fibers created looped domains that are stabilized by attaching to either specific non-histone nuclear scaffold proteins or to condensed chromatin within adjacent chromonema fibers. The random walk/giant loop model considered the large-scale aggregation of chromatin into DNA foci. It was developed by assessing the distances between defined genomic sequences visualized by fluorescence in situ hybridization (Sachs et al. 1995). This mathematical model assumed that a series of flexible 3 million base-pair chromatin domains were randomly attached to a non-chromatin nuclear substructure. However, this did not consider the presence of smaller loops addressed in the multi-loop sub compartment model of chromosome territories (Munkel et al. 1999). Among the random walks/giant loops, the multi-120 kb loop sub compartment model considered the intertwining of fibers. This would promote long-range interactions that were constrained by requiring that individual sub compartments provide a high level of specification within each chromosome territory.

Figure 2: Models of Chromosomal Looping. Several models have been proposed for the looping of the 30 nm fibers. The chromonema model proposes the formation of a higher order structure through the progressive compaction of the 30 nm fiber into ~130 nm chromonema fibers. Each chromonema fiber is composed of condensed ~28 kb zig-zags of DNA. The chromonema fibers are then coiled to create looped domains that may be attached to the nuclear scaffold/matrix (dots). The random walk/giant loop model suggests that large condensed chromatin loops of 3-5 Mb are randomly attached to a non-DNA backbone (green dots) within the nucleus. The multi-loop sub compartment model is based on individual ~120 kb chromatin loops anchored to the nuclear scaffold/matrix (green dots). They are organized into rosettes that contain a total of 1-2 Mb of DNA connected to one another by linker DNA.
III. Nucleosome remodeling, modification and transcription

It is well established that post-translational modifications of the amino terminal tails of the core histone proteins that are typified by, but certainly not limited to, acetylation, methylation and phosphorylation, can alter DNA structure. These individual modifications serve to mark regions for active transcription or silencing. It appears that a dynamic balance between multiple histone modifications maintains a chromatin environment that is either permissive or restrictive to transcription (for a review, see Jenuwein and Allis 2001). For example, a histone modification module consisting of a combination of 17 modifications has been defined. This module tends to be overrepresented within the promoters of the highly expressed genes in human CD4+ T cells (Wang et al. 2008). However, these histone modifications do not uniquely determine expression, as other modifications are required to coordinately regulate structure and poised a gene for transcription.

The regions immediately surrounding active genes can be completely devoid of nucleosomes. Several studies have shown that nucleosome presence may not be conducive to transcription (Wasyley et al. 1979; Lorch et al. 1987) requiring their removal for transcriptional activation (Han and Grunstein 1988). For example, chromatin-remodeling complexes such as the multiprotein SWI/SNF complex can alter nucleosome positioning to reduce steric hinderence to promote recognition of specific sequences by transcription factors (Richmond and Davey 2003). This can be achieved through several mechanisms including sliding/moving the histone octamer to a new position to expose the DNA (Hamiche et al. 1999; Langst et al. 1999; Whitehouse et al. 1999), ejecting the octamer (Lorch et al. 1999; Boeger et al. 2003; Reinke and Horz 2003; Boeger et al. 2004), removing the H2A-H2B dimer (Bruno et al. 2003; Yang et al. 2007) and/or replacing dimers with histone variants e.g., exchanging H2A-H2B with H2B-H2A.Z (Mizuguchi et al. 2004).

Transcriptional regulation by structure must also be considered. For example, regions of Z DNA upstream of yeast promoters demarcate the boundaries of neighboring nucleosome rich regions (Wong et al. 2007). These alternative conformations provide a means to absorb torsional stress thereby relieving local super coiling induced by active transcription. This highlights one specific mechanism for altering DNA structure. Additional mechanisms have been identified including the binding of high mobility group proteins (Javaherian et al. 1978) as well as the action of enzymes such as topoisomerase II to stabilize and/or relax the DNA [reviewed in (Giaever et al. 1988)].

IV. Higher order levels of organization and transcription

The regulation of gene expression by higher order structure is not well understood. Perhaps this reflects the poor understanding of chromatin organization beyond the level of the 30 nm fiber. As summarized in Figure 3, it is well established that chromosomes occupy distinct territories within the nucleus (Cremer and Cremer 2001) and genes undergoing active transcription can loop out away from their respective chromosome territories into structurally distinct interchromatin compartments (Albiez et al. 2006). On one hand, looping has been shown to coincide with the activation of the major histocompatibility locus (Volpi et al. 2000) and the HoxB gene cluster (Chambeyron and Bickmore 2004). Perhaps chromosomal looping may be a result of the local density of expressed genes (Mahy et al. 2002) reflecting the aggregation of transcriptionally active genes into transcription factories [reviewed in (Marenduzzo et al. 2007)]. On the other hand, X inactivation activates multiple domains of higher order looping. These may form separate chromatin hubs at the noncoding Xite and Tsix loci as part of the sex-specific X inactivation center induced developmental chromosomal looping (Tsai et al. 2008).

There is a large body of evidence for transcription factories [reviewed in (Jackson 2005)], i.e., complex centers of transcription where active genes from distant locations on the same or different chromosomes are clustered together (Osborne et al. 2007). Joining non-linear regions of the genome within the same transcription factory suggests that chromatin may physically relocate to the factory. One of the most well studied examples of looping is that of the β-globin locus. Activation of this gene cluster relies on the developmentally timed interaction of the locus control region with specific promoters by looping of the intervening DNA (Tolhus et al. 2002). Looping appears to be mediated by the attachment of at least one hypersensitive site within the LCR to the nuclear matrix (Ostermeier et al. 2003). Similar to the concept of a transcription factory, where distant genomic regions are brought together within one nuclear complex, the β-globin locus provided a key developmental example of an active chromatin hub. Active chromatin hubs that are part of transcription factories are assembled through looping to promote the interactions of distant elements to facilitate tissue specific expression [reviewed in (de Laat and Groverseld 2003)].

The positioning of specific genes and or chromosomes relative to the boundaries of the nucleus may also affect transcription. For example, gene dense chromosomes are often located towards the nuclear interior while gene poor chromosomes are located peripherally, just interior to the nuclear lamina (Bridger et al. 2000; Tanabe et al. 2002). Interestingly, genes targeted to the inner nuclear membrane and nuclear laminas are silenced (Reddy et al. 2008). Tethering of genomic regions to the nuclear lamina may serve as a silencing mechanism as lamin B1 associated domains (LADs) appear preferentially bound in gene poor regions (Guelen et al. 2008). Silencing may also be achieved by specific promoter region binding to type A lamins (Lee et al. 2009). While recruitment of genes to the nuclear periphery can alter expression, its effect is not always
The seemingly conflicting data of gene silencing at the nuclear periphery suggests that other factors may play a role. For example, Ikaros regulates lymphocyte cell expression by developing interphase heterochromatinization foci (Brown et al. 1997). As expected, inactive genes were associated with heterochromatic foci whereas active genes were located elsewhere. This is consistent with the view that genes are selectively recruited to repressive foci of heterochromatization.

![Diagram of Gene Expression Coordination with Looping](image)

**Figure 3: Induction of Gene Expression Coordinates with Looping.** Groups of genes that are poised for expression are located at the periphery of the chromosome territory, but not necessarily expressed (‘silent genes’). Upon induction of gene expression, the domain loops out away from the chromosome territory, perhaps into a transcription factory (‘expressed genes’).

V. The Nuclear scaffold, matrix and nucleoskeleton

The proteinaceous nuclear matrix was first observed when interphase nuclei were treated with high salt solutions (Zbarskii and Debov 1951). The protein components of the residual body of the interphase nuclear matrix were subsequently characterized from rat liver nuclei (Berezney and Coffey 1974). This clearly showed that the protein constituents of the high salt extracted nuclear matrix could be differentiated from that of untreated nuclei. Looping constrained by a residual protein body was subsequently verified (Cook and Brazell 1976). This structure was shown to persist in metaphase nuclei after histones were depleted by extraction with 2 M NaCl (Paulson and Laemmli 1977; McCready et al. 1979). Subsequent removal of the DNA component revealed a protein scaffold (Adolphs et al. 1977) suggesting attachment to a static structure. Similarly, isolation of the interphase nuclear matrix was visualized as loops of super coiled DNA that appeared as a halo of DNA constrained by the nuclear matrix (Vogelstein 1980). This respectively defined the nuclear scaffold and nuclear matrix. As expected, their protein constituents differed and their nuclear constituents were classified into three components: the peripheral nuclear lamina, an internal network of proteins and residual nucleoli upon which the nuclear scaffold composed primarily SC1, i.e., Topoisomerase II (Earnshaw et al. 1985; Gasser et al. 1986) and SC2 (Lewis and Laemmli 1982) was overlaid.

It was suggested that the ability to resolve sites of attachment may be compromised by sliding or the rearrangement of the sites of DNA attachment to the nuclear matrix when prepared with high salt. To address this limitation, extraction with LIS, or 25 mM lithium diiodosalicylate, was introduced as a “milder” means to extricate the non-histone nuclear protein fraction. When applied, this procedure also revealed a nonrandom interphase loop structure organized by the residual extracted proteins (Mirkovitch et al. 1984). However, lamins along with their corresponding sites of attachment were absent from LIS scaffold preparations (Mirkovitch et al. 1988). Nevertheless other protein complexes (4 of the 6 sites remained attached to both metaphase and interphase scaffolds) appeared constant throughout the cell cycle (Mirkovitch et al. 1988). The terms nuclear scaffold referring to LIS extraction and nuclear matrix referring to extraction with 2M NaCl have since been adopted.

The methodology of nuclear matrix isolation continues to develop (Fey et al. 1986; Wan et al. 1999) as part of the toolset to study both replication (Vogelstein 1980) and transcription (Jackson et al. 1981). This includes ‘physiological’ methods that encapsulate cells in agarose beads prior to hypotonic lysis followed by the gentle salt extraction of histones that resolves the nucleoskeleton (Jackson et al. 1988). Studies comparing some of these methods have shown method dependent resolution of sites of attachment (Donev 2000) suggesting that the various extraction methods may isolate different subgroups of proteins. However, throughout the literature the various methods are often interchangeably used without addressing the implications. This has fueled the debate of the existence of a matrix/scaffold within the nucleus (Pederson 2000). It has recently been resolved (Heng et al. 2004) with the aid of a comprehensive description of the biological role of the sites of nuclear matrix and scaffold attachment (Linnemann and Krawetz 2009; Linnemann et al. 2009).
VI. Specific constituents of the nuclear scaffold/matrix

Isolation of the metaphase scaffold by different methods of extraction does not appear to affect the types of proteins that are isolated. Initial characterization of the scaffold proteins revealed that the major constituents are the protein scaffold component 1 (SC1), i.e., topoisomerase II (Earnshaw et al. 1985; Gasser et al. 1986) and scaffold component 2 (SC2) (Lewis and Laemmli 1982). Although at a reduced level to metaphase cells, topoisomerase II persists in interphase cells as a part of the nuclear scaffold/matrix (Berrios et al. 1985; Berrios and Fisher 1988; Mirkovitch et al. 1988). Multiple variants of topoisomerase II have been shown to differentially mark potentiation of the proteamine gene domain during mouse spermatogenesis (Martins and Krawetz 2007). This includes sites of matrix attachment that position the locus within a small looped region coincident with expression.

Some, but not all, of the non-histone proteins identified from interphase nuclei differ based on the type of isolation method used. Visualization of the interphase nuclear matrix and nucleoskeleton has revealed proteins that appear to have a structural role in the nucleus. Electron microscopy of thick resin less sections of HeLa cells revealed that residual chromatin loops were associated with a network of proteins that appeared as intermediate filaments (Jackson and Cook 1988; He et al. 1990). Interestingly, cytoplasmic type intermediate filaments also interact with SARs and are isolated as a part of the nuclear scaffold (Tostonomic et al. 2002). The presence of nuclear lamins supporting the nuclear envelope as well as their dispersion throughout the nucleus suggests that these proteins may play a significant structural role as a part of the nuclear scaffold/matrix [reviewed in (Goldman et al. 2002)]. The lamins that dominate and define the nuclear lamina are resistant to high salt solubilization while the more diffusely organized lamins throughout the nucleus are less resistant. For example, lamin B1 is widely distributed throughout the nucleus as a part of the nucleoskeleton. It was recently shown to be necessary for RNA synthesis (Tang et al. 2008). The relative ease that the lamins can be extracted with solutions of high salt suggests that this method can be used to differentiate proteins that appear fixed and static from those serving a more dynamic role. It is likely that these intermittently attached proteins specifically function as needed during dynamic processes like transcription. Thus their differential isolation may provide insight into the direct role of these nuclear scaffold/matrix components of transcription.

As depicted in Figure 4, the inner and peripheral networks within the nucleus appear physically continuous but remain functionally distinct. Their differential isolation by high salt extraction with or without cation/heat stabilization has led to their description as either the Type I or Type II nuclear matrix respectively (Lebkowski and Laemmli 1982a; Lebkowski and Laemmli 1982b). The isolation of the type II nuclear matrix without stabilization yielded peripheral nuclear matrix components including proteins of the lamina and nuclear pore complexes. Similarly, it has been shown that MARs do specifically bind to type II matrix components of the nuclear lamina (Luderus et al. 1992). In comparison the protein constituents of the type I or inner nuclear matrix have been characterized with the use of cation/heat stabilization. The array of proteins characteristic of type I nuclear matrices is similar to those isolated as part of the LIS nuclear scaffold.

Electron microscopy showed Intermediate like filaments that span clusters of globular type proteins (Jackson and Cook 1988) suggesting clustered foci of additional functional non-histone proteins. At least 200 proteins have been detected by 2D gel electrophoresis (Capco et al. 1982) and over 300 by tandem mass spectrometry (Ishii et al. 2008). The wide range of proteins that have been identified as constituents of the nuclear scaffold/matrix attests to its complex nature. However, the distribution of proteins present in relatively small proportions begs the question: does this merely represent a snapshot of dynamic interactions such as the induction of expression of a group of functionally related genes?

One of the most thoroughly investigated nuclear matrix proteins with functional impact is SATB1. This protein is expressed predominantly in the cells of the thymus, where it binds to MARs at the bases of chromatin loops to regulate global gene expression during T cell development (Dickinson et al. 1992; de Belle et al. 1998; Alvarez et al. 2000). SATB1 has recently been shown to direct a significant increase in the number of loops present at the cytokine locus upon T-helper type 2-cell activation and is necessary for the activation of the locus (Cai et al. 2006) Another protein that has gained interest as a potential nuclear matrix interacting protein is CTCF. Interestingly, the binding of CTCF insulates discrete regions of a genome (Antes et al. 2001; Magdinier et al. 2004; Yusufzai and Felsenfeld 2004). However, CTCF may not be a nuclear matrix protein (Goetze et al. 2005) requiring reconciliation. Additional proteins such as scaffold attachment factors A and B [SAF-A (Romig et al. 1992) and SAF-B (Renz and Fackelmayer 1996)] also bind both SARs and MARs in vitro, but are only extracted as a part of the nuclear scaffold. SAF-A is analogous to hn-RNPU and analysis of the two protein isoforms has indicated that they both bind to SAR elements with similar affinity (Fackelmayer and Richter 1994). Several transcription factors, such as the pit-1 and the histone acetyltransferase/ transcriptional coactivator CBP, have also been isolated as part of the cadre of nuclear matrix proteins [reviewed in (Hendzel et al. 2001)].

Characterization of the nuclear substructure has established a link between the local physical packing of DNA and the global reorganization of chromosomes during cellular differentiation (Mayer et al. 2005; Martins and Krawetz 2007). It has been suggested that homeotic proteins (Boulikas 1992) like the homeobox domain containing protein Oct-1, may anchor the genome to the nuclear lamina (Imai et al. 1997). RUNX, or Run-related transcription
factors, are also associated with the nuclear matrix. Their targeting to discrete foci within the nucleus essential to regulate RUNX dependent genes and in turn tissue specific differentiation during embryonic development (Choi et al. 2001). This intranuclear targeting to transcriptionally active foci occurs in partnership with Smad, which attaches to the nuclear matrix as a RUNX- Smad protein complex (Zaidi et al. 2002). Recently it was suggested that the sperm nuclear matrix is necessary for zygotic development (Shaman et al. 2007). Curiously, it is enriched in RNAs (Lalancette et al. 2008).

Figure 4: The Nuclear Matrix. Within the nucleus, the nuclear matrix can be subdivided into the outer/peripheral, or type II nuclear matrix. This can be isolated without the use of heat or cation stabilization. The type II nuclear matrix includes proteins of the nuclear lamina as well as proteins of the nuclear pore complexes. The inner, or type I nuclear matrix can be isolated after nuclei stabilization with heat or the addition of cations such as Cu^{2+}. This portion of the nuclear matrix is complex, comprised of hundreds of different types of proteins. These appear to form clusters on a backbone of filaments of similar diameter as the intermediate filaments.

VII. Attachment of the genome to the nuclear scaffold/matrix

The human genome is organized into loop domains attached at their bases to the nuclear scaffold/matrix by S/MARs that exhibit a range of sizes depending on extraction method (Jackson et al. 1990). Considering an average loop size of approximately 100 kb, it can be estimated that there are 32,000 looped domains throughout the human genome. If each looped domain is demarcated by an attachment site on either side, this would correspond to approximately 64,000 MARs throughout the human genome. In mature spermatozoa, toroid loops are estimated to be approximately 50 kb in length (Brewer et al. 1999), or half that of somatic cells. If every toroid loop were attached to the nuclear matrix in the intervening linker DNA, this would also correspond to ~64,000 sites of attachment in the sperm genome.

Several studies have suggested that MARs are not evenly distributed throughout the mammalian genome. For example, loops as small as 1 kb have been observed at telomeres (Luderus et al. 1996). Similarly, the imprinted region on mouse distal chromosome 7 resolves 52 MARs in a 1 Mb domain (Purbowasito et al. 2004). Although the average MAR spacing was ~20 kb, the majority of attachments clustered into almost equal groups of 58 kb and 169 kb regions flanked by MARs. These observations have now been reconciled following the recent genome-wide analysis (Linnemann and Kräwetz 2009; Linnemann et al. 2009). This has shown that S/MARs segment the genome into a wide range of cell-type dependent differentially sized loops. Taken together, the estimates of the number of
S/MARs are merely predictions. It is important to note that both cellular environment and cell type likely influence the number of attachments observed at any given point in time.

VIII. Sequence specificity and in-silico prediction of nuclear scaffold / matrix attachment

Although many S/MARs that have been identified in vitro and in vivo have been studied in depth by sequence analysis, the identification of an obvious consensus sequence for scaffold/matrix attachment has proven difficult. Perhaps this reflects the many different protein interactions that appear to contribute to the operational definition of an S/MAR. While to date no consensus has been derived, S/MARs have displayed several reiterative characteristics. Initial observations based on a small subset of all genomic sites suggest that S/MARs are approximately 300 to 1000 bp in length, AT enriched, containing both oligo d[A] and d[T] tracks (Gasser and Laemmli 1986) as well as ATATAT boxes (Cockerill and Garrard 1986). As S/MARs are preferentially bound and cleaved by topoisomerase II both in vitro (Adachi et al. 1989; Razin et al. 1991) and in vivo (Kas and Laemmli 1992), many are expected to contain topoisomerase II binding sites.

While S/MARs have often been generally characterized as AT rich regions or sequences containing topoisomerase II both in vitro (Platts et al. 2006), there is evidence that these two characteristics alone are not sufficient for nuclear matrix binding (Das et al. 1993; Bode et al. 2006). Further, this simplicity reflects the lack of examples or a conformational dependence (Yamamura and Nomura 2001). For example, there is direct evidence of an intrinsically curved structure in the tobacco S/M II MAR where the nuclear proteins appear to bind to the MAR because of structure and not sequence (Fukuda 2000). A body of evidence is also accumulating to suggest that mammalian MARs coincide with Base Unpairing Regions (BURs) that alleviate the torsional strain induced by transcription of nearby regions (Bode et al. 2006).

These general characteristics have been used in silico to independently and in combination predict the locations of S/MARs [reviewed in (Platts et al. 2006)]. The application of these algorithms often yields weak and/or seemingly over predictive models of nuclear scaffold/matrix attachment. However, because S/MARs are known to display cell type specificity, what appears as an over prediction may represent potential attachment in other cell types. Additionally, the sequence characteristics identified thus far may work in combination with structural features such as BURs and perhaps additional unidentified sequences or local chromatin modifications to direct attachment to the nuclear scaffold/matrix are required. Recent evidence suggests that in some cases the inclusion of structural elements can lead to stronger prediction algorithms (Girod et al. 2007).

IX. The multifactorial nature of MARs

S/MARs can be classified as constitutive or facultative, perhaps possessing either structural or functional characteristics respectively. In this respect, the human epsilon globin gene has been shown to have both facultative and constitutive MARs (Yan and Qian 1998). Additionally, several MARs within a 1 Mb region of human chromosome 19 have been identified and classified as either structural or regulatory (Chernov et al. 2002). What is perhaps of note is that the S/MARs are merely a snapshot of a dynamic environment. Differences in conditions or the time of measurement could yield varying levels of attachment for an individual S/MAR. Thus their observed roles, be that constitutive or facultative, may not be easily subdivided through the study of a single point in time, environment or cell type.

As a function of both the cell cycle and cell type specificity, only a subset of potential S/MARs are expected to be observed at any given time. In specific cases S/MARs have been shown to regulate conformation in both a cell type specific and cell cycle context specific manner (Britanova et al. 2005; Koina and Piper 2005; Martins and Krawetz 2007). Evidence for cell type specific nuclear scaffold/matrix attachment suggests facultative and constitutive attachment can be differentiated. For example, MARs of the osteocalcin gene are tissue specific and transiently associated with the nuclear matrix (Bidwell et al. 1993) while the tyrosine hydroxylase MARs bind the nuclear matrix independent of tissue type (Lenartowski et al. 2003). Removal of the chicken lysozyme MARs increases ectopic expression (Bonifer et al. 1994) confirming their role in tissue specific gene expression. Recent genome wide surveys have shown that SARs and MARs provide complementary activities that either facilitate or suppress gene expression (Linnemann et al. 2009). In this manner cell type specific attachment mediates identity (Linnemann and Krawetz 2009).

X. The role of MARs in disease pathogenesis

Epigenetic contributions to disease have been long suspected in studies of monozygotic and dizygotic twins (Wong et al. 2005). The linkage of the epigenetic control of gene expression, often via aberrant methylation and acetylation, to disease is well appreciated (Kalebic 2003). While little is known about the direct effect of S/MARs on tissue specific genes in a normal or diseased state, changes in nuclear scaffold/matrix attachment by either the loss of association or through binding to a previously cryptic site have been implicated in the onset of several genetic disorders. Even synonymous mutations that might not affect protein sequence could result in changes in scaffold/matrix binding resulting in disease if the local structure is perturbed. However, in this case they would perhaps be better characterized as non-synonymous. For example, the loss of a MAR may arrest expression of a male factor and present as infertility (Kramer et al. 1997). Similarly, in vitro...
loss of a MAR has been correlated with the onset of FSHD, Fascioscapulohumeral muscular dystrophy (Petrov et al. 2006). Deletions in conserved noncoding sequences of heritable diseases (Kleinjan and van Heyningen 2005; Sjakste and Sjakste 2005) are often underscored by regions containing MARs and insulators. Although the nuclear lamins bind to S/MARs, the key laminopathies resulting from defects in the A/C type lamins, most often muscular dystrophy and cardiomyopathy, are not necessarily directly linked to the function of the nuclear matrix. The primary phenotypic effect of these mutations is an inability of cells with contractile behavior to survive the stress associated with cellular function. The result is that the nuclei are damaged and thus destroyed so that the muscle fibers harbor reduced numbers of nuclei and cannot function. Other laminopathies induced through lamin A mutations including Hutchinson-Gilford progeria syndrome that also presents atherosclerosis (McClintock et al. 2006), and dilated cardiomyopathy (Taylor et al. 2003) may act by directly affecting nuclear scaffold/matrix attachment to destabilize the gene expression program.

The role of MARs in cancer is well established. For example, sites of chromosomal fragmentation preferentially localize to MARs (Schoenlein et al. 1999; Bode et al. 2000a; Bode et al. 2000b). In addition, progression to a malignant phenotype may be promoted by the binding of mutant p53 to the nuclear matrix thereby dysregulating the cell-cycle (Deppert 1996). Similarly, aberrant binding to the nuclear matrix may destabilize the expression of a host of genes that maintain cells in a non- cancerous state (Linnemann and Krawetz 2009). The most direct evidence for cancer by perturbed nuclear scaffold/matrix attachment is that induced by SATB1. As a genomic organizer, SATB1 regulates transcription during T cell development (Dickinson et al. 1992; de Belle et al. 1998; Alvarez et al. 2000; Cai et al. 2006). SATB1 is not dominant in other cell types but its expression in breast cancer cells is indicative of an aggressive phenotype (Han et al. 2008). Interestingly, ectopic expression of SATB1 in non-invasive cancer cells induces an invasive phenotype whereas removal of SATB1 from metastatic cells reverses the invasive phenotype and inhibits tumor growth.

S/MARs continue to reveal a multitude of varied roles in development and disease. Continued analyses should provide a window to their underlying mechanism(s) of action reflective of their complex nature. This is essential as we begin to take these elements from the bench to the bedside for use as effective gene therapeutics.

**Acknowledgements**

This work was supported in part by National Institutes of Health grant HD35612, the Wayne State University Research Enhancement Program in Computational Biology and the Charlotte B. Failing Professorship if Fetal Therapy and Diagnosis to SAK.

**References**


The Krawetz Laboratory: Front row from left to right, Amelia Linnemann, Kathryn Drennan, Claudia Lalancette, Back row from left to right, Robert Goodrich, Adrian Platts, Stephen A. Krawetz, Seema Mahadev, Graham Johnson