doi:10.1016/j.jmb.2005.11.073 J. Mol. Biol. (2006) 358, 597–613

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

# Correlations between Scaffold/Matrix Attachment Region (S/MAR) Binding Activity and DNA Duplex Destabilization Energy

**Jürgen Bode[1]*, Silke Winkelmann[1], Sandra Götze[2], Steven Spiker[3] Ken Tsutsui[4], Chengpeng Bi[5], Prashanth A.K.[5] and Craig Benham[5]***

[1]*German Research Center for Biotechnology, RDIF/Epigenetic Regulation, D-38124 Braunschweig, Mascheroder Weg 1, Germany*

[2]*Swammerdam Institute for Life Sciences, University of Amsterdam (UvA), 1098 EL Amsterdam, The Netherlands*

[3]*North Carolina State University, Genetics Dept. Box 7614, Raleigh, NC 27095-7614 USA*

[4]*Department of Molecular Biology, Okayama University Medical School, 2-5-1 Shikatacho, Okayama 700-8558 Japan*

[5]*UC Davis Genome Center University of California, One Shields Ave., Davis, CA 95616 USA*

**Corresponding authors*

Scaffold or matrix-attachment regions (S/MARs) are thought to be involved in the organization of eukaryotic chromosomes and in the regulation of several DNA functions. Their characteristics are conserved between plants and humans, and a variety of biological activities have been associated with them. The identification of S/MARs within genomic sequences has proved to be unexpectedly difficult, as they do not appear to have consensus sequences or sequence motifs associated with them. We have shown that S/MARs do share a characteristic structural property, they have a markedly high predicted propensity to undergo strand separation when placed under negative superhelical tension. This result agrees with experimental observations, that S/MARs contain base-unpairing regions (BURs). Here, we perform a quantitative evaluation of the association between the ease of stress-induced DNA duplex destabilization (SIDD) and S/MAR binding activity. We first use synthetic oligomers to investigate how the arrangement of localized unpairing elements within a base-unpairing region affects S/MAR binding. The organizational properties found in this way are applied to the investigation of correlations between specific measures of stress-induced duplex destabilization and the binding properties of naturally occurring S/MARs. For this purpose, we analyze S/MAR and non-S/MAR elements that have been derived from the human genome or from the tobacco genome. We find that S/MARs exhibit long regions of extensive destabilization. Moreover, quantitative measures of the SIDD attributes of these fragments calculated under uniform conditions are found to correlate very highly ($r^2 > 0.8$) with their experimentally measured S/MAR-binding strengths. These results suggest that duplex destabilization may be involved in the mechanisms by which S/MARs function. They suggest also that SIDD properties may be incorporated into an improved computational strategy to search genomic DNA sequences for sites having the necessary attributes to function as S/MARs, and even to estimate their relative binding strengths.



*Keywords:* nuclear architecture; nuclear matrix; scaffold-matrix attachment regions (S/MARs); DNA structure; stress-induced duplex destabilization (SIDD)

## Introduction

The organization of the eukaryotic nucleus into chromosomal domains is thought to be mediated by a proteinaceous intranuclear framework, called either the nuclear matrix,[1] or the nuclear scaffold.[2] Branched core filaments provide a supporting structure for the formation of DNA loops, and participate in diverse matrix-supported processes

Abbreviations used: S/MAR, scaffold/matrix attachment region; CUE, core-unpairing element; BUR, base-unpairing region; SIDD, stress-induced duplex destabilization.

E-mail addresses of the corresponding authors: jbo@gbf.de; cjbenham@ucdavis.edu

involved in DNA replication and transcription, RNA processing and transport,[3,4,5] signal transduction and apoptosis.[6,7] The DNA elements that mediate attachment of chromatin loops to this nuclear scaffold are called scaffold/matrix attachment regions (S/MARs).[8] S/MARs are operationally defined according to the protocols that led to their detection (see below). The elements that are recovered by these procedures have been implicated in a variety of biological activities that are compatible with an affinity for the nuclear matrix.[9] These include the insulation of transgenes from negative effects of the genomic surroundings (insulator function),[10,11] augmentation of transcription rates,[12] long-term maintenance of high transcription levels by counteracting DNA methylation, and the support of histone acetylation,[13,14] enhancer,[15–18] and origin-of-replication functions.[19,20] During their replication, episomes are segregated by S/MARs functioning as maintenance elements.[21] All these activities may be consequences either of the structures of S/MARs or of their interactions with proteins and/or other molecules.

In line with the above activities, S/MARs have been reported to mediate domain opening.[22,23] During embryonic development, this process is accompanied by a regional demethylation activity.[16,17] These data, together with the results from various insulation experiments,[9–11,24,25] strongly implicate S/MARs in defining the boundaries of autonomously regulated chromatin domains. The occupancy of at least some of these boundaries can be regulated dynamically *in vivo*.[26,27] Since these dynamic effects have not been fully characterized, we will examine measures of matrix affinity and associated biological activities that have been observed consistently for all S/MARs.

S/MARs have clearly been shown to have a variety of effects on transcription. They can act *in cis* to increase transcriptional initiation rates,[28,29] even in the absence of an enhancer. This so-called augmentation activity[12] is clearly distinct from prototypical enhancement, since enhancers, but not S/MARs, are active in transient assays.[30,31] The presence of S/MARs has been shown to prevent the ectopic expression of transgenes.[32] In one case, the presence of an S/MAR proved to be indispensable for correct hormonal gene regulation.[33,34] It has been shown that short subsections of a domain border can substitute for the function of an enhancer-associated element.[35]

Several investigations have documented a direct correlation between the matrix binding and transcriptional augmentation activities of S/MARs.[36–38] This suggests that a standard assay of *in vitro* S/MAR binding strength might predict this regulatory S/MAR activity, and perhaps others as well. However, the general augmentation effect of an S/MAR can be modulated or even disrupted by the over-expression of distinct S/MAR-binding proteins.[39] This finding suggests that an intricate interplay may occur among various S/MAR-binding proteins, which could alter constitutive S/MAR-matrix contacts.[40–42]

## The relationship between *in vitro* and *in vivo* S/MAR activities

The first studies of the structure/function relationships of S/MARs performed in this laboratory examined the organization of a 14 kb region containing the human interferon β (IFNB1) gene domain that is located at position 9p22 on the short arm of chromosome 9. The transcription unit of IFNB1 is bounded by a strong 7 kb S/MAR upstream, and a strong 5 kb S/MAR downstream (Figure 1(a)). A second, weakly-associating element has been found between the gene and its strong downstream S/MAR.[43] The occupancy of this secondary S/MAR is regulated *in vivo* according to transcriptional activity.

To investigate the structure and transcriptional activity of the native IFNB1 domain, we made a series of transgene constructs and studied their expression in a mouse L host cell line.[31] Next, we created a series of artificial single-S/MAR and double-S/MAR constructs using various reporter systems. These studies again demonstrated that transcriptional augmentation increased as longer S/MAR-like elements were added, and was maximal for constructs having a minidomain structure in which the reporter gene was bracketed on both sides by S/MARs.[31,44]

The properties that confer S/MAR-binding activity have been studied *in vitro* using artificial elements of the appropriate sizes and minimal sequence complexity, whose binding and transcriptional augmentation properties could be determined in parallel. The first study of this kind constructed a series of oligomers of a 166 bp region from the upstream IFNB1 S/MAR.[36] Neither binding nor augmentation was observed for the monomer, but a simultaneous increase in both activities was found with increasing degree of oligomerization (Figure 4). This and other studies indicate that the strength of *in vitro* binding can be used as a qualitative predictor of transcriptional efficiency.

## Specific DNA sequences are not associated with S/MAR-binding activity

The prediction of S/MARs from primary DNA sequence data has proven to be unexpectedly difficult. Despite considerable efforts, no single consensus sequence, pattern or motif has been found to be associated with S/MAR-binding activity. A set of six rules has been proposed which, together or alone, have been suggested to contribute to S/MAR function.[45,46] The observations that suggested these rules, which were based on a relatively small sample set of S/MARs, are: (i) sites of matrix attachment share certain AT-rich tracts with homeotic protein recognition sites and origins of replication; (ii) a number of genes

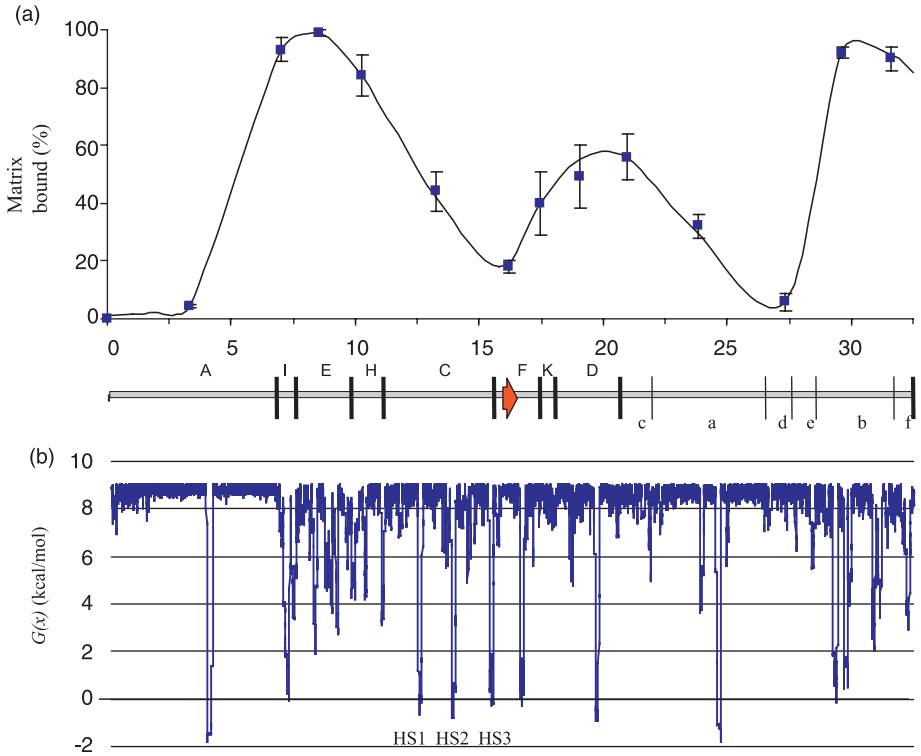


**Figure 1.** The human interferon-ß (huIFN-ß) gene and its upstream bordering (S/MAR-) region analyzed by stress-induced duplex destabilization (SIDD). (a) Binding strength for EcoRI/PstI-fragments (A-F, H, I, K; G is an EcoRI-BglII fragment) and BglII/HindIII fragments (a, b, d, e, f; c equals G) along the IFNB1 chromatin domain. The position of the coding region is marked by an arrow in fragment F. (b) The SIDD profile of this region. Here $G(x)$ is the free energy needed for full separation of base-pair $x$ when this region is at superhelix density $\sigma = -0.055$.

contain TG-rich sequences in their 3′-untranslated regions (UTRs), which can be S/MARs; (iii) intrinsically curved DNA occurs within or near several S/MARs; (iv) the dinucleotides TG, CA or TA, spaced in a way that leads to kinking, are present in some S/MARs; (v) topoisomerase II cleavage sites occur at nuclear matrix attachment sites; and (vi) many S/MARs contain significant stretches of AT-rich sequence, with prominent occurrences of both $A_n$ runs[47] and $(AT)_n$ tracts.[48] (A refinement of this criterion, the 90% AT box, has also been proposed.[49]) These rules were amalgamated into the MARfinder computer program,[45,46] now called MAR-Wiz†. Although sometimes in excellent agreement with the structure-based method described below,[50] this approach alone is not always successful at reliably finding S/MAR locations.

The compilation of an S/MAR database called SMARtDB has enabled a more thorough investigation of the sequence attributes of S/MARs.[51] An analysis of all experimentally characterized S/MARs in SMARtDB in late 2002 showed that most, but not all, were significantly A+T-enriched relative to their genomic averages.[52,53] The statistical significances of occurrences of specific sequence elements in S/MARs were assessed by comparing their occurrences in each actual sequence with their expected frequency distributions in randomized sequences having the same base composition as the actual sequence. This was done for each of the 245 sequences in the database. This analysis found that the apparent enrichments for all the types of elements listed as (ii)–(vi) above were not statistically significant. Their frequencies did not significantly exceed those found for random sequences having the same average base composition as these S/MARs. Moreover, a careful examination was made of the frequencies of occurrence of every possible $n$-mer, from dimers through hexamers, both for the actual S/MAR sequences and for four randomized sequence sets. Of the 2080 non-redundant hexamers, none was present in all S/MARs. Only 34 occurred at apparently elevated frequencies, all of which are highly A+T-rich. It is probable that most or all of these were simply random statistical outliers, an inevitable consequence of asking 2080 different questions of a dataset. Indeed, none was present in sufficiently large numbers of different S/MARs to suggest they formed part of a sequence motif. This analysis suggests strongly that S/MARs do not have identifiable sequence patterns or motifs. So, their binding properties must be determined by other factors. The expectation is that these are structural in nature.

† http://www.futuresoft.org/MAR-Wiz/

### Stress-induced duplex stabilization and base-unpairing in S/MARs

Our own efforts to define a unique property common to many or all S/MARs were guided by observations that these elements have an unusual propensity for DNA strand separation. *In vitro* chemical reactivity studies have shown that S/MARs, when placed in plasmids under negative superhelical tension, contain local regions of strand separation. Depending on the salt conditions, strand separation initiates at the most easily destabilized site, the core-unpairing element (CUE), and from there spreads throughout a base-unpairing region (BUR) having well-defined borders.[48] Base unpairing has been found to occur in the transcriptional regulatory loci controlling several genes, including those encoding chicken $\beta^A$-globin,[54] the human CMV major immediate early protein,[55] interferon-β,[43] and c-myc.[56,57] Several of these BURs have been shown to coincide with S/MARs.[44,58]

The local propensity for the DNA duplex to become destabilized under stress can be calculated using methods from equilibrium statistical mechanics.[59–62] This stress-induced duplex destabilization (SIDD) analysis technique calculates the extent to which the imposed level of superhelical stress decreases the free energy needed to open the duplex at each position along a DNA sequence. The results are displayed as an SIDD profile, in which sites of strong destabilization appear as deep minima. (The SIDD profile of the human IFNβ gene region is shown in Figure 1(b).)

Using this approach, we have shown that all S/MARs investigated to date share a common architecture comprised of one or more very easily destabilized unpairing element(s) within a long region of considerable destabilization.[50,59] These S/MAR-associated SIDD sites commonly extend for at least several hundred base-pairs, and consist of either an extended continuous unpairing element, as in the case of the IFNWP18-IFNA10-IFNA7 segment in the interferon cluster,[50] or a succession of approximately evenly spaced local unpairing elements (Figure 1). The sites of strong destabilization within these regions, which are the deep minima on the SIDD profile, coincide with the experimentally observed BURs. And the deepest minimum commonly is found at the CUE, which acts as a nucleation center for strand separation. The extent of destabilization seen for the S/MAR regions of these plasmids is sufficient to render them partially or entirely single-stranded at a moderate negative superhelicity (viz. at superhelix density $\sigma = -0.05$).

Some activities of S/MARs may be directly related to their propensity to become single-stranded under negative superhelical tension.[63,64] Any mechanism that involves either single strand-specific DNA binding or partial helical unpairing would then be expected to occur preferentially at S/MARs.

Here, we examine the quantitative correlations between the SIDD properties of S/MARs and their *in vitro* binding activities. A variety of S/MARs whose binding strengths have been determined experimentally are placed in a standard plasmid context, and the SIDD profiles of these plasmids are calculated at a standard superhelix density. This is done first for three classes of artificial sequences that were constructed to evaluate how scaffold binding varies with systematic changes in specific attributes. Then, we use the same approach to determine the SIDD properties that correlate most closely with S/MAR-binding activity for naturally occurring S/MARs, whose SIDD profiles are much more complex. The close correlation between SIDD properties and both binding strengths and biological activities has a number of obvious implications regarding regulatory mechanisms.

## Results

### The IFNB1 domain: an overview

Figure 1(a) shows the measured binding strengths for individual EcoRI (upper case letters) or BglII /HindIII fragments (lower case letters) across the human interferon-β (IFNB1) domain. Figure 1(b) shows the calculated SIDD profile of this region, together with previously determined mapping data. The coding sequence of IFNB1 (short arrow) is contained within the most stable interval in the entire region, and is flanked by destabilized sites. A pronounced SIDD site in its 3′ flank, which, as in other cases,[44] coincides with the transcriptional termination site, is among the most highly destabilized in the region. In contrast, the promoter is destabilized less strongly. Three major upstream SIDD sites coincide with DNase I-hypersensitive sites (HS1, HS2 and HS3),[63,64] and represent factor-binding sites with regulatory potential. Among these, HS3 also is a genomic fragile region, and thereby a hotspot for deletion.[65] Sites H3 and H2 bracket a register of six positioned nucleosomes.[63]

The strongest S/MAR activity found within the region depicted in Figure 1(b) occurs in fragments I, E, and H that are taken from the left-most 5kb of the sequence. This region is characterized by a long succession of moderate SIDD minima. If S/MAR character depends on SIDD properties, as has been suggested,[59] the activity must be a composite function of the extent of a destabilized BUR, the overall amount of destabilization throughout the BUR, and possibly of the spacing between individual SIDD peaks. The importance of each of these parameters is assessed in model constructs analyzed below.

### Artificial S/MARs

We analyze three classes of artificially constructed S/MAR elements that were designed to illuminate how binding properties depend on

specific attributes of a S/MAR. First, we consider how binding increases with the length of a simple destabilized site. Next, we assess the synergistic interactions between two sites. Then, we determine how binding varies with the number of identical sites placed in a row, with fixed separation distances between them.

### *Extending a core unpairing element: the $(25)_n$ series*

The S/MAR located 3′to the immunoglobulin heavy chain gene (IgH) enhancer contains the 25 bp A+T-rich sequence CTCTTAATTTCTAATATATT TAGAA. When cloned into a plasmid and subjected to negative superhelicity, this site has been shown to become base-unpaired, while the balance of the plasmid remains double-stranded.[48] This 25 bp tract has been oligomerized to form the $(25)_n$ series. This was done in order to determine experimentally how S/MAR activity varies with sequence length, a question that arose from the observation that active S/MARs are relatively long, commonly containing at least 300 bp, and that there appears to be a minimum size consistent with activity.

The *in vitro* binding strengths of these oligomers were evaluated experimentally by using the modified equal fractions scaffold reassociation procedure, as described.[37] Three measurements were made for each oligomer and the results are shown in Figure 2(b). Significant binding commences with the trimer, and the average measured activity of each oligomer increases monotonically with $n$ thereafter. Binding is half-maximal at the pentamer level [$(25)_5$], and 100% binding is achieved by the octamer. The heptamer $(25)_7$ binds with strength equal to that of our 800 bp standard, and both elements cause substantial transcriptional augmentation.[36,48] (This heptamer was used by Kohwi-Shigematsu and co-workers to isolate S/MAR-binding proteins, including SATB1, nucleolin, and the p114/PARP and Ku/PARP complexes.[66,67])

The destabilization properties of each oligomer in the $(25)_n$ series were computed in the standard pTZ-18R vector context by the method described below. The SIDD profiles of the plasmids containing the monomer through the octomer are shown in Figure 2(a). The oligomer insert sites all show a simple pattern of destabilization: $G(x)$ drops steeply to a uniform minimum whose width at the bottom corresponds to the size of the oligomer. For $n>1$ the minima at the oligomer site all have essentially the same depth. Even at the dimer level, destabilization at the insert site outcompetes that at the terminator of the β-lactamase gene, an internal standard that has been shown to be highly susceptible to stress-induced denaturation.[59,68] This behavior results from the globally interactive character of stress-driven transitions. Denaturation at any site relaxes superhelicity, which is felt by all other sites experiencing the stress. As the amount of relaxation afforded by denaturing the insert region increases with insert length, the competitiveness of denaturing the terminator region diminishes correspondingly.

Extents of destabilization may be assigned to the insert portions of the $(25)_n$ series of plasmids in a variety of ways, based either on lengths or areas, and using any integer threshold $G_d$ between 1 and 9. (In this, as in other analyses, the absolute minimum value of $G(x)$ is not informative, here having essentially the same value for every insert larger than the monomer; for this reason, it is not considered further.) In this analysis we include only insert sizes $n=2,\ldots8$. The monomer was excluded because no activity was measured for it in any experiment. The 9-mer and 10-mer were excluded because complete binding had occurred in the shorter 8-mer. For each threshold chosen, both the area and the length SIDD measures increased linearly with $n$ with highly significant correlation ($r^2>0.99$). This result did not change when 8-mers were also excluded (data not shown.). Next, the average binding strength was determined from the three measurements that were made for each insert length. Linear, exponential, logarithmic and sigmoidal functions of the SIDD area and length parameters were fitted to these averages.

For every threshold $G_d>1$, the best linear fits of either length or area measures gave essentially identical results, in all cases yielding $r^2=0.96$. This is not surprising, given that the linear regression between the average binding strength and insert length $n$ also has $r^2=0.96$, and the fit between the SIDD parameters and $n$ is virtually perfect, as described above. A sigmoidal functional form yielded essentially indistinguishable results. The fit of the exponential form was degraded only slightly, while the logarithmic form yielded somewhat poorer fits. These conclusions are apparent from Figure 2(b) when one notes that the abscissa $n$ is linearly related to both area and length SIDD measures. One sees that the experimentally measured activity values shown in that Figure fit very well to either a sigmoid or a straight line. They fit an exponential or a logarithmic curve slightly less well.

This series of extremely simple artificial constructs does not serve to distinguish whether length or area measures are more useful, or what values of the threshold $G_d$ are most informative. However, their results can be used to determine the minimum value of each SIDD parameter at which onset of binding activity occurs. This is the $x$-intercept of each fit of each SIDD measure at each threshold $G_d$.

There is no reason to expect that simple artificial constructs should behave identically with natural S/MARs. Thus, as has been noted earlier, the threshold for 50% activity found for this $(25)_n$ series is shorter than the threshold length seen in natural S/MARs, whose destabilization patterns are much less uniform. Any biological differences between plant and vertebrate S/MARs also might be reflected in distinct patterns of SIDD association (see Discussion). For these reasons, we usually analyze each series *de novo*, independent of previous results on other series.

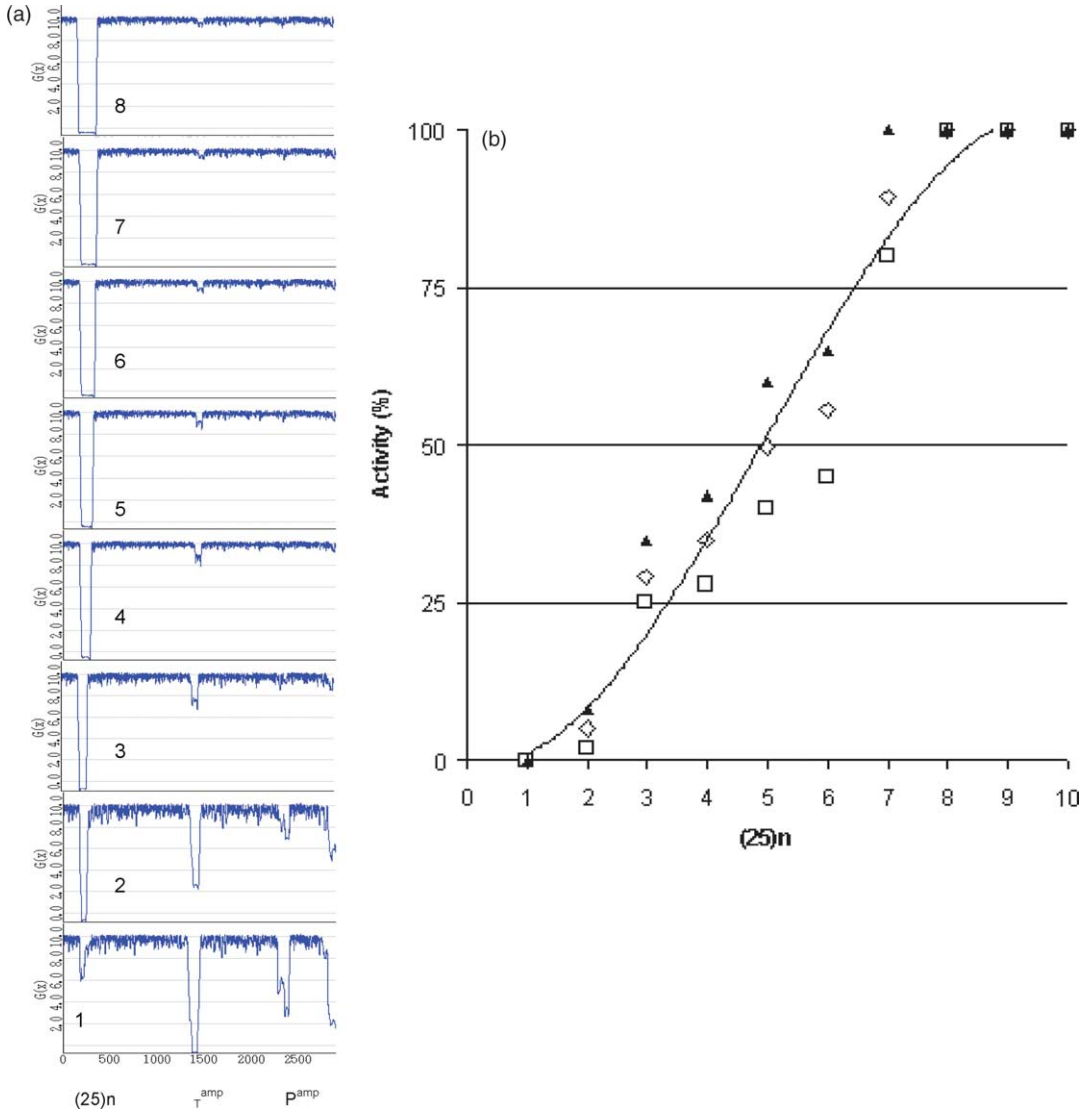


**Figure 2.** Extending an unpairing element creates an active S/MAR element. (a) The 25 bp core-unpairing element defined by Bode *et al*.[46] was oligomerized creating a series $(25)_1$ to $(25)_{10}$. The Figure shows the SIDD profile for the inserts ($n = 1$–8) in a vector background as described.[46] As the 25 bp segments are not separated by a spacer, oligomerization results in a continuous extension of the destabilized region. As the extent of oligomerization increases, the resulting element becomes increasingly capable of competing with the *amp*-associated unwinding elements until, at higher degrees of oligomerization, these are the only regions at which the superhelical strain is released. This behavior illustrates the globally competitive nature of stress-driven transitions, where denaturation at one site causes a corresponding relaxation of the superhelical stresses throughout the domain involved. Because longer oligomers relax more stress on denaturing, their transition more effectively suppresses competing transitions elsewhere. (b) The binding activities of the members of this series were determined by a standardized reassociation assay in the presence of LIS.[37] The binding strength is referenced to an 80% binding standard as described.[59] For the present purposes, the binding strengths were called activities to suggest that they can serve as predictors of the transcriptional augmentation potential of these elements.[46]

### *Synergy and communication between unpairing elements: the Okada dimer series*[69]

The SIDD profiles of all S/MARs that have been analyzed to date show them to be highly susceptible to stress-induced destabilization. Their profiles commonly consist either of a series of closely spaced destabilized sites or of a unique, strongly destabilized site, often with secondary minima nearby (e.g. see Figure 1(b)). The deepest minima on the SIDD plots are seen experimentally as unpairing elements. As the pattern of destabilization within an S/MAR often is complex, it is appropriate to consider how the separation distances between its internal SIDD sites might affect its overall binding.

To address this question, Okada and colleagues have constructed a series of S/MARs derived from the immunoglobulin κ enhancer-associated S/MAR.[47,69] This S/MAR is composed of two unpairing elements, which by themselves have sub-S/MAR extension. These unpairing elements

have been separated and dimerized individually. Six homodimeric S/MAR inserts were constructed, in which the copies were separated by distances of 5 bp, 100 bp, 300 bp, 500 bp, 800 bp and 2900 bp. (The spacer DNA in every case shows no destabilization under stress.) The S/MAR activities of these sequences were determined using the standard procedure, as described below. A strong synergistic effect was seen, in which S/MAR-binding activity dropped sigmoidally as the distance between the individual elements increased. The segment having least separation showed the strongest binding, 37% in this assay. Binding activity dropped sigmoidally with separation distance until at the widest separation (2.9 kb), where the sites presumably bound independently, 7% activity was measured.

We model this situation as follows. Consider a domain containing two SIDD sites, which we call $S_1$ and $S_2$. (In the experiments described by Okada *et al.*,[47,69] these sites are identical copies of a sub-S/MAR, but here we describe a more general case in which the sites need not have identical sequences or SIDD properties). Suppose the separation distance between these sites is $d$. When $d$ is sufficiently small, the sites effectively abut, so they may be regarded as forming a single longer site, which we call $S_{12}$. We denote the S/MAR-binding probabilities of the three individual sites $S_1$, $S_2$ and $S_{12}$ by $p_1$, $p_2$ and $p_{12}$, respectively. (This probability is the binding fraction. Since binding activity is expressed as the percentage binding, these two values are directly equivalent.)

When $d$ is sufficiently large, sites $S_1$ and $S_2$ are effectively independent, so the probability of both binding is the product $p_1p_2$ of their individual binding probabilities. The domain will bind when either or both sites bind; so, by the addition rule, the overall binding probability $p_i$ is:

$$p_i = p_1 + p_2 - p_1\, p_2 \qquad (1)$$

The binding probability of the widely separated, independent dimer was measured to be $p_i = 0.07$. Putting this value into equation (1), setting $p_1 = p_2$ as appropriate for identical repeats, and solving the resulting quadratic equation yields $p_1 = p_2 = 0.036$.

According to the experiment, as the separation distance $d$ increases, the observed binding probability $p_{obs}$ decreases sigmoidally from $p_{12} = 0.37$ to $p_i = 0.07$:

$$p_{obs} = f(d)p_{12} + [1 - f(d)]p_i \qquad (2)$$

We call $f(d)$ the synergy function. We have determined the synergy function that best fits the experimental data reported by Okada *et al.*[69] using the following procedure. We give $f(d)$ the functional form of a standard sigmoidal curve, which is the cumulative distribution function of a normal distribution with mean $\mu$ and standard deviation $\Phi$. We fit this curve to the experimental activities by evaluating this expression for a range of values of $\mu$ and $\Phi$, and determining the pair of values that give the best rms fit to the data using equations (1) and (2). We found this to occur when $\mu = 407$ and $\sigma = 112.5$. The curve that best fits the data, which we call the synergy curve, is shown in Figure 3.

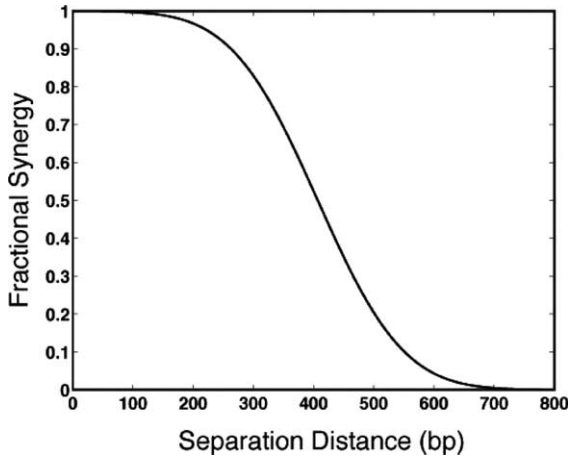


**Figure 3.** Communication between individual unpairing elements is required to create activity. Dimerization of a sub-S/MAR size fragment (130 bp) from the immunoglobulin κ S/MAR creates an element associating with the nuclear scaffold. If neutral spacers with a balanced AT/GC ratio are inserted between the monomers, S/MAR-activity is gradually lost, indicating the spatial binding requirements of the major constituent proteins.[69] This Figure shows the synergy function derived in the text to describe this effect.

In fitting SIDD properties to binding data we first evaluate an SIDD measure $m$, then find a functional form $a = F(m)$ that best fits the binding activity $a$ as a function of this measure. We analyze a variety of measures and functional forms until we find those that provide the best fit. Because synergistic interactions are clearly very important in determining overall activity, this effect must be included in our procedure. For this purpose, we use the synergy function found here. However, this can be done in either of two ways. Suppose the SIDD profile of an S/MAR contains two minima separated by a distance $d$. (In practice, $d$ will usually be determined as the separation distance between their associated minima in the SIDD profile, not by the separation of well-defined inserts as for the above experiment.) One may assume that the synergistic effect occurs either at the level of the measure, where:

$$m_{eff} = f(d)m_{12} + [1 - f(d)]m_i, \;\; a_{eff} = F(m_{eff}) \qquad (3a)$$

or that it occurs at the level of activity, where:

$$a_{eff} = f(d)F(m_{12}) + [1 - f(d)]F(m_i). \qquad (3b)$$

Unless the fitting function is linear, these approaches will not be equivalent. In what follows, we will use both strategies. Thereby, we will determine the best way to incorporate synergistic interactions based on the Okada synergy function $f(d)$ determined above.

The experimental observation reported by Okada *et al.*,[69] that the S/MAR-binding activity of a region

containing two sub-S/MARs increases when they are moved closer together, suggests that some sort of distance-dependent communication occurs between them. There are several possible mechanisms for this communication. Proximity between the sites could increase the chances that both bind to the same complex, which, through allosteric effects, could induce a stronger attachment than is achieved by widely separated sites that must bind to independent complexes. Alternatively, binding could involve a progressive buildup of a protein complex from components that must contact each other in order to mediate binding of sufficient strength.[8] The precise mechanism underlying this synergy remains to be elucidated.[47]

### *Repetition of unpairing elements: their communication in the (VIII)$_n$ series*

The (VIII)$_n$ series is constructed from $n$-mers ($n$=1, 2, 3, or 4) of a 166 bp segment from the huIFN upstream S/MAR.[36] This segment comprises a core-unpairing element similar to that used to construct the (25)$_n$ series. The pCL sequence, an 800 bp segment taken from the same S/MAR, serves as our reference standard for normalizing binding strengths.

We have inserted each of the elements in the (VIII)$_n$ series and the pCL sequence separately into the pTZ-18R vector, and calculated the SIDD profiles of the resulting plasmids. The results for the (VIII)$_n$ series are shown in Figure 4. Strong destabilization is predicted to occur within a subregion of each 166 bp segment, so the strongly destabilized sites in the multimeric inserts in that series are separated from each other by fixed distances. Here, this series is used to evaluate the dependence of binding strength on the number of identical, uniformly separated participating sites, other factors remaining fixed.

The binding activities of these S/MAR constructs have been measured experimentally, as described

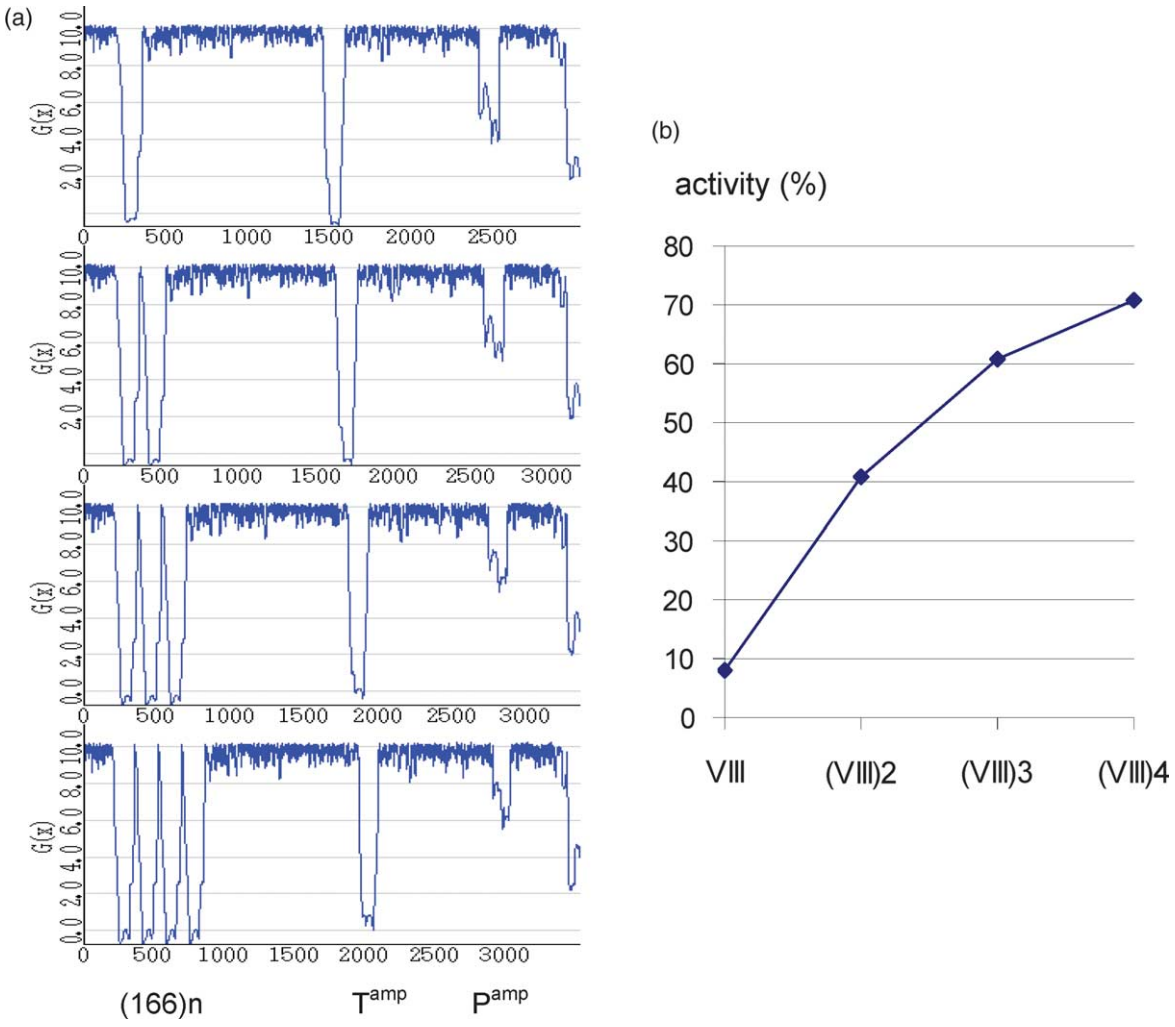


**Figure 4.** Active S/MAR elements can be created by a succession of individual unpairing elements. The 166 bp sequence surrounding a core unpairing element in fragment VIII[36] was oligomerized to form the (VIII)$_n$ series, for $n$=1, 2, 3, 4. The resulting S/MAR-binding activities of these constructs were found to increase with the degree of oligomerization. We have previously reported a correlation of binding activity with the transcriptional augmentation potential.[84]

below. The 166 bp monomer behaves as a sub-S/MAR, exhibiting very slight binding affinity and no transcriptional augmentation potential. However, both of these activities increase monotonically with $n$, the number of monomer units in the insert.[37] At the tetramer level (664 bp) the activity of the 800 bp pCL reference standard S/MAR is attained, both in the binding experiment (Figure 4) and in the biological test of transcriptional augmentation.[12,37] This concordance suggests that the functions of the $(VIII)_4$ artificial S/MAR closely match those of the natural element.

The binding activities of the multimeric $(VIII)_n$ series appears to increase with copy number $n$ in a slightly sub-linear manner. This might suggest that the synergy between nearby sites is not strictly a linear combination of the values when separated widely and when coincident, as it has been modeled in equation (2). However, it is more plausible to incorporate this non-linearity into the functional dependence $a=F(m)$ of activity on destabilization measure. Some form of sub-linearity is expected, as the maximum binding strength cannot exceed 100%, no matter how extensive the destabilization (in this series, how many repeat copies are present). This sub-linearity suggests a logarithmic dependence of activity on destabilization, as the following analyses also find. However, the deviation from linearity is small in this case, within the uncertainty observed in repeated experimental measurements of binding activity under other circumstances. (See the analysis below of the β-interferon S/MARs.)

## Naturally occurring S/MARs

Here, we analyze two collections of naturally occurring S/MARs, the first from the human interferon-β gene region and the second from plants. In these analyses, we use the information gleaned from the simple paradigm cases described above.

The human and plant S/MARs are analyzed separately. We do this for two reasons. First, the binding activities of these two series were measured in different laboratories, which may introduce systematic differences in their measurements. In particular, the human series included the 800 bp reference standard, but the plant series did not. So, measurements of binding activities in these studies are not directly comparable. Second, and more fundamentally, the structural and/or functional properties of S/MARs may not be universal (see below).[70]

### *S/MARs from the human interferon-β chromatin domain*

The human type 1 interferon gene cluster consists of 26 IFN genes and pseudogenes, distributed over 400 kb at Giemsa band 9p22. This cluster has gained particular attention because deletions that initiate here and include the adjacent tumor suppressor gene(s) are related to some of the most common genetic abnormalities that occur in numerous forms of cancer.[65,71] S/MARs have been implicated in the mechanism of these deletions.[72]

The interferon-β gene (IFNB1) is the most telomere-proximal member of this cluster. It also is in the longest domain of the cluster, extending over 14 kb and flanked on both sides by strong S/MARs.[43,65] Immediately downstream from IFNB1 there is an S/MAR of intermediate strength, whose occupancy halo-mapping experiments suggest is regulated upon gene induction.[23,43,63]

Figure 1(b) confirms that the SIDD profile of the region encompassing the interferon β gene coding sequence is complex.[65] While the coding region itself is stable, it is flanked on the 5′ side by a slightly destabilized promoter region, and on the 3′ side by a highly destabilized locus that contains its poly-adenylation site. This architecture has been noted for numerous other eukaryotic genes.[59,62] The other functional members of this gene cluster have similar arrangements of SIDD sites, but some of the pseudogenes do not. Thus, the presence of S/MARs closely parallels the SIDD properties of these intergenic regions.

*Binding under uninduced versus induced conditions.* The binding activities of 11 fragments were examined under conditions where the IFNB1 gene was either induced or uninduced. These were ten EcoRI/PstI fragments (upper case letters in Figures 1(a) and 5(a)), and a BglII/HindIII fragment ("b")

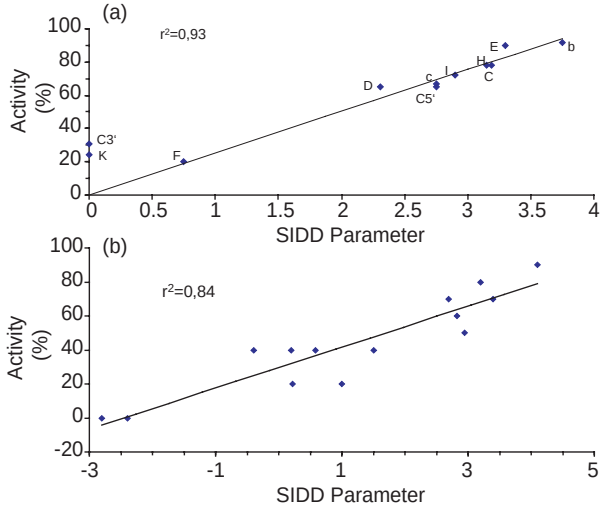


**Figure 5.** Best correlates of SIDD properties with binding strengths for naturally occurring S/MAR restriction fragments. (a) This series of 11 fragments is from the IFNB1 region, as shown in Figure 1. Besides the EcoRI fragments (A–K), the series includes $C^{3'}$ and $C^{5'}$, the PstI-subfragments of EcoRI fragment C. (b) The best fit to the plant series of S/MAR elements described in the text.[47] Fragments and binding strengths are from Michalowski *et al.*,[49] who explored the predictive potential of a number of sequence-oriented parameters. The SIDD-derived parameters used in the present work were superior to these previous approaches, as discussed.[49] Fragments in order of decreasing SIDD parameter (i.e. from right to left along the SIDD axis) are: 116-1 > 211-1 > RB7-6 > 205-2 > 206-1 > 220-1 > 115 > ps1 > 202-2 > ps4 > 217-1 > 202-2 > ARS1 > ps8 > 218 (see Table 1 of Michalowski *et al.*[49]).

spanning the 3′ S/MAR. Two measurements were made under the uninduced condition, and one under the induced condition. The measured binding strengths were correlated very strongly between each pair of experiments ($r > 0.95$ in all cases), indicating that reassociation affinities (in contrast to the *in situ* halo-mapping approach) remain unaffected by prior gene activity.

Parameter sets and procedures were sought that gave good fits to all three experiments. When placed in the standard plasmid context and subjected to SIDD analysis, in all cases the profile of the S/MAR region contains sites with strong destabilization. Often, these consist of multiple strongly destabilized sites in close proximity, extending over several hundred, or more, base-pairs. Less frequently, a pattern of limited destabilization or of destabilization at a single site is observed.[50]

Our evaluations showed that either area-based measures or linear measures could give fits that were as good as the pairwise correlations between these experiments, achieving coefficients of determination $r^2 > 0.9$ for all three experiments. In all cases, the best fitting function was logarithmic. Multiple minima were merged according to equations (1) and (2), using a closest-pair-merges-first strategy. The minimum length $N_{min}$ of the destabilized region needed to be considered an S/MAR ranged between 150 bp and 235 bp using either area or length measures. The best fitting area measures did not consider the relative amount of destabilization in the vector *versus* the insert, but the best fitting length measures did. Excellent fits (i.e. $r^2 > 0.9$) were found either way. The two apparent outliers ($C^{3'}$ and K, with binding affinities between 20% and 30%) represent the flanks of the transcription unit, and may therefore bind to specialized proteins that are not ubiquitous matrix components.

*An extended IFNB1 series.* A set of 16 fragments from the IFNB1 gene region (I, E, H, C, $C^{3'}$, $C^{5'}$, F, K, D, G, c, a, d, e, b, f as shown in Figure 1(a)) were investigated for their S/MAR binding activities. One complete fitting run investigated 426,523 different parameter sets, which is a typical value. Again, excellent fits were achieved using either area (best $r^2 = 0.897$) or length (best $r^2 = 0.865$) measures. In all cases, the best-fitting function was logarithmic. As before, a range of parameter values gave high correlations. Here, the minimum lengths were smaller, $N_{min}$ ranging between 60 bp and 190 bp. In some cases, the SIDD measures were weighted by the fraction of destabilization occurring in the insert, while in other cases they were unweighted.

### S/MARs from plants

Spiker and co-workers have created a library of randomly obtained S/MARs from tobacco by cloning DNA fragments that co-isolate with the nuclear scaffold obtained by the LIS halo-mapping procedure.[49] Subsequently, 34 cloned fragments were tested for binding using a technique that corresponds closely to, but is not identical with, the reassociation protocol used for the IFNB1 S/MAR studies. Of the 34 fragments tested, 13 were chosen for further study. These include three strong scaffold binders, six that bind with medium strength, two that bind weakly, and two non-binders. In addition to these 13 elements, a previously characterized strong tobacco ARS completed the series (see Table 1 of Michalowski *et al.*[49])

Although somewhat different assays for strength of binding were used for the plant S/MARs and the IFNB1 S/MARs, the results found here are similar to those for the IFNB1 S/MARs. High correlations are achieved with either linear or area measures for a fairly broad range of parameter choices. All strong correlations required a logarithmic functional form. The best fitting length and area measures both weighted the SIDD measure by the fraction of destabilization in the insert. Here, however, the coefficients of determination were not as high as for the earlier case. The best fit for area measures was $r^2 = 0.844$, and for length measures the best fit was $r^2 = 0.826$. This is probably as good as can be expected, as the experimental binding activities were measured to only one significant digit of accuracy (i.e. to the nearest 10%).

The work reported here demonstrates that only some of the possible ways of constructing SIDD measures are relevant to S/MAR activity. In particular, the best fits to all natural S/MARs have logarithmic functional forms. Area measures give slightly better results than length measures, while extrema measures are not informative. Regarding the competitiveness of the vector *versus* the insert regions, ratio measures do not give good fits. Fractional measures give the best fits, but in some cases they are not greatly improved over measures that do not consider this factor at all. Beyond this, strong associations (viz. $r^2 > 0.8$) are achieved with a variety of thresholds and with both area and length SIDD measures.

## Discussion

The prediction of S/MAR locations from primary sequence data has proven difficult, because S/MARs do not appear to share a consensus sequence or motif (see Introduction). With hindsight, this is reasonable, as the major scaffold proteins recognize structural attributes rather than specific base sequences. Our efforts to define the DNA properties that are shared by many or all S/MAR were guided by the observation that these elements have an unusual tendency for SIDD, which becomes manifest when an S/MAR is placed under negative superhelical tension in plasmids.[60] We have shown by computation that natural S/MARs share a common architecture consisting of unpairing elements, which together form an extensive base-unpairing region.[59]

Here, we have documented a close correlation between the magnitude and extent of destabilization seen in the SIDD profiles of S/MARs and

their *in vitro* binding strengths. Using suitable models in the form of artificial S/MAR elements, the underlying parameters, such as the size and degree of destabilization of a BUR could be examined (Figure 2). Spacing criteria have been taken into account in cases where BURs were composed of individual unpairing elements (Figures 3 and 4).

As the SIDD properties used here to correlate with S/MAR activity were calculated in a standard context and under uniform conditions, it is not straightforward to apply the conclusions to regions whose SIDD properties were computed in other contexts. In particular, the results presented here do not permit S/MAR prediction directly from the genomic SIDD profile. Instead, one can use the genomic SIDD profile (or any other information) to find possible S/MARs, then apply the procedures presented here to predict their activities. This requires that each site be placed in the standard plasmid context, so all SIDD profiles are computed under uniform conditions. The resulting profile is examined for SIDD sites, whose properties are analyzed using the parameters that we have shown give the strongest correlations. These are the minimum length $N_{\min}$, and the threshold SIDD measure $m_{\mathrm{thr}}$ for whichever attribute (area or length) is of greatest interest (possibly both). Then one must amalgamate close SIDD sites to get a total SIDD measure $M$ for the insert region, plus area measures for the insert ($A_{\mathrm{ins}}$) and the vector ($A_{\mathrm{vec}}$) components. This information may be used to compute a predicted strength $a_{\mathrm{p}}$ of S/MAR binding using the functions that have been found here to give the best fit:

$$a_{\mathrm{p}} = s(g \log_{10}(M)) + b$$

Here, $s$ and $b$ are the slope and $y$-intercept of the best fitting regression line, as determined in the previous section. The fractional destabilization of the insert site is:

$$g = A_{\mathrm{ins}}/[A_{\mathrm{vec}} + A_{\mathrm{ins}}]$$

If this predicted activity is greater than the threshold for strong scaffold binding ($a > 50\%$), this predicts S/MAR binding. This could be done for either area or length measures, or for both. But the parameters appropriate to each case must be used.

It is important to note that this procedure predicts the level of binding that would occur in the standard assay procedure used here (see Methods), not the strength derived from a halo-mapping approach or even the *in vivo* binding situation. Although this assay correlates with *in vivo* activity, our SIDD measures as currently developed do not directly predict that activity. Nevertheless, our results have three important implications. First, the fact that the SIDD properties of S/MARs correlate closely with their activities supports the suggestion that destabilization may be involved in the mechanism of scaffold attachment. Second, calculations of SIDD properties could form the basis of a strategy to computationally identify within genomic DNA sequences the sites that have the necessary attributes to function as S/MARs. Third, the results presented here show that the SIDD properties of naturally occurring S/MARs can be used to guide the construction of biologically active synthetic S/MAR elements with a restricted complexity.[12,36,37]

## Protein–S/MAR interactions and base–unpairing

The backbone of the eukaryotic nucleus can be isolated and characterized according to protocols that have been optimized for the removal of soluble components.[72,73] As expected for an operationally defined entity, the various procedures used to isolate the nuclear matrix/scaffold produce preparations that differ somewhat in ultrastructure,[74] although most of their protein constituents appear to be the same.[75–78]

After LIS-extraction the resulting scaffold contains components from three nuclear compartments, the lamins A–C (lamina), nucleolin (nucleolus) and factors from the fibrogranular internal network. The components of this network have a strong resemblance to intermediate filaments.[79–81] In addition to their established role at the nuclear periphery, internal lamins[81] are now thought to serve regulatory functions in both transcription and replication, and NuMA (nuclear mitotic apparatus protein) also participates in the nuclear skeleton, where it co-localizes with splicing factors that link RNA processing to the nuclear substructure.[82]

If the nuclear matrix is isolated in the absence of ribonuclease (as is usually done), it contains 70% of the nuclear RNA together with a number of known RNA-binding proteins. Among these, the heterogeneous nuclear ribonucleoproteins (hnRNP) persist even after RNA degradation, suggesting that they are involved in the binding of hnRNA to the nuclear matrix. Experiments have shown that actin is also a constituent of interphase nuclei and nuclear matrices, where it is closely associated with small nuclear ribonucleoproteins (snRNPs) and pre-mRNAs.[83]

It has been suggested that either lamin B or SAF-A shows the S/MAR-binding behavior of a complete scaffold. While both of these proteins display comparable recognition profiles for native and artificial S/MAR elements, they clearly differ in how effectively S/MAR binding competes with ssDNA binding.[12,50] The 50% competition limit found for the complete scaffold[84] is thought to reflect the combined contributions of these two major constituents.

*S/MAR elements and the DNA-binding properties of the nuclear scaffold*. Vertebrates, plants and yeast all have S/MARs, and these are functionally interchangeable *in vivo*.[44] Once optimal SIDD predictors are known for each group, one can assess how well they predict the activities of other groups. Any

differences found might suggest systematic differences in the SIDD properties of plant and human S/MARs.

Lamins are essential components of the nuclear scaffold in vertebrates, but are entirely absent from yeast.[70] Plants do have lamins, but they have striking differences from their vertebrate conterparts in both sequence and nuclear localization.[70] The fact that lamins bind ssDNA suggests that the SIDD attributes of plants and vertebrate S/MARs might differ. This is why S/MARs from these sources were analyzed separately here. However, the fact that both families were best fit by the same functional forms with similar parameters suggests that their destabilization properties may not be significantly different.

Mattern *et al*. have identified the 21 most abundant proteins that are present only in the internal network, most of which are hnRNPs.[76,77] This observation is consistent with earlier reports,[85] and supports a model in which the major protein constituents of the internal nuclear network participate in RNA metabolism, packaging and transport. The most prominent scaffold attachment factor in LIS-extracted scaffolds is hnRNP-U, which is also called SP120[86] and SAF-A.[84] Like the lamins, SAF-A also associates with multiple S/MARs. Although hnRNP-U is known to bind RNA,[87] UV-crosslinking experiments show that it also associates with DNA *in vivo*.[84] It has been proposed that S/MAR binding to SAF-A is optimized when "AT patches" (short stretches of consecutive A and T bases), are distributed according to certain rules.[8,47] However, an extensive analysis of the SMARtDB does not find evidence for any specific consensus sequence or motif for overall S/MAR binding.[52]

*Structural features recognized by S/MAR-binding proteins*. The fact that the SIDD properties of S/MARs correlate highly with both their binding and biological activities suggests that destabilization may be involved in their mechanisms of scaffold attachment.[88] A concern may be raised that this correlation is made between SIDD profiles calculated for DNA under global superhelical stress, and the *in vitro* binding activities of DNA fragments under no such stress. However, the fact that S/MAR activity requires a long destabilized site suggests strongly that multiple binding events may be involved. If the proteins that perform this binding associate to form a matrix, they could exert substantial stresses on the DNA between the binding sites in a manner analogous to superhelicity. If the DNA must be strongly bent, for example, the increased flexibility of denatured regions could enable this to occur in an energetically favorable manner. So ease of stress-induced denaturation could be a requirement for strong binding, even in short, linear molecules. A specific model has been suggested,[89] according to which proteins with propensities to bind single strands will bind the fraction of fragments that expose single strands at equilibrium. An analogous process will occur *in vivo*. But in this situation positive superhelicity will arise transiently adjacent to the binding site. After the relaxation by topoisomerases a repository of underwound DNA is formed, which may support the progression of tracking proteins in a context-dependent manner. This model has been used to explain the fact that an S/MAR immediately downstream from a promoter forms an impediment to transcription, while at a more remote location it is able to support progression of RNA polymerase in a process that changes the binding state at the S/MAR, relieving positive supercoils ahead of the approaching enzyme.[28,89]

Besides SAF-A and lamins, BURs have been shown to bind a number of other S/MAR-binding proteins, some of which were first isolated by their preferential binding to an oligomerized $(25\,\mathrm{bp})_7$ CUE (Figure 2(a)) over a mutagenized $(24\,\mathrm{bp})_8$ sequence that lacks the unpairing property. The contribution of these proteins,[66–67,90–97] and their role in the functional network of the nuclear matrix, has been reviewed.[8,12,23] In summary, many proteins associating with S/MARs are attracted by specific features of BURs, such as particular dsDNA structures (SATB1, HMG I(Y), SAF-A, (the recognition of single strands individually (lamins, nucleolin), or conjointly ($p53^{mut}$), or secondary structures depending on prior unpairing (HMG-1/2). Together, these factors have the potential either to activate,[96,97] or to inactivate,[39,41,42] a chromatin domain. These effects explain many of the known effects of S/MARs, including nuclear localization, their influences on the expression characteristics of associated genes,[9] and their role as elements that support the function of episomes and provide for their segregation.[98]

## Methods

### Experimental assay of scaffold binding

Two conventional methods have been used to detect S/MARs. The halo-mapping approach reveals the potentiated status of genes, i.e. either their active transcription or their propensity to be transcribed. In contrast, scaffold/matrix re-association techniques can be used to derive quantitative data on the interaction of S/MARs with the scaffold/matrix. Only the latter method yields results that are suitable for the present purposes.[37] Here, we use the previously described "modified equal fractions approach" to derive reassociation binding data.[37,50] Briefly, a nuclear scaffold preparation is obtained by LIS-extraction of nuclei, typically from cultured mouse L or other rodent cell-lines. After equilibration in digestion buffer, the resulting halo is degraded with a robust restriction enzyme selected for its compatibility with the ends of the fragments that will be subjected to reassociation. (Where possible, EcoRI is the preferred choice.) The resulting scaffolds are supplied with a large amount (8000–60,000-fold molar excess) of bacterial (*Escherichia coli* genomic) DNA. Digested endogenous fragments are left in place to serve as an additional competitor, a step that improves the reproducibility of the procedure.[49] Scaffolds are then divided into aliquots, and each aliquot

is provided with a mixture of end-labeled restriction fragments. As an important control, an S/MAR-containing vector was divided by restriction into its S/MAR part (strongly binding) and its plasmid part (non-binding). This control serves to establish the "equivalent fractions" volumes, i.e. the volumes of supernatant (S) and pellet (P) that contain the same amount of radioactivity (cpm). These volumes are maintained for the remaining samples that contain fragments having presently unknown binding characteristics. Only in cases where a subsequent electrophoretic separation yielded an autoradiograph in which the control contained vector DNA in the S-fraction and S/MAR DNA in the P-fraction (in the absence of any cross-contamination) are the parallel samples subjected to densitometric evaluation. Minor variations between the experimental series are normalized using an EcoRI/HindIII-cut pCL vector, whose 800 bp S/MAR insert serves as a 70% binding standard. In addition, mixture experiments were performed in which this standard sequence was re-associated, together with other unknown samples, and their properties were compared under conditions of enhanced competition.

### Calculation of SIDD profiles

The computational analysis of the destabilization properties associated with S/MARs was performed as follows. First, the base sequence was determined for each S/MAR fragment of interest. Then, this sequence was inserted computationally into a uniform standard context. For this purpose, we use the multiple cloning site of the PTZ-18R plasmid (Pharmacia), a pUC derivative. This is the same context in which the experimental strengths of scaffold binding of many of the S/MAR fragments were assessed.[59]

The SIDD properties of the resulting plasmid were calculated at the standard superhelix density $\sigma = -0.055$ using the method described by Benham.[59–62] This method calculates the SIDD profile of the plasmid, which is the graph of the destabilization energy $G(x)$ associated with the base-pair at each position $x$ along the plasmid. Here, $G(x) = G_x - G$, where $G$ is the ensemble average free energy of the system and $G_x$ is the average free energy of the subset of states in which base-pair $x$ is separated.[61] Thus, it provides a measure of the stability of that base-pair within the superhelical plasmid. The destabilization energy $G(x)$ is inversely related to the probability of denaturation: positions where $G(x)$ is high require large amounts of free energy to open, and hence are stable, while positions where $G(x)$ is low can be opened easily. States that are destabilized sufficiently strongly to experience strand separation at the assumed level of superhelicity will have values of $G(x)$ near or below zero, while sites that are destabilized marginally have intermediate values of $G(x)$. Such locations are not destabilized sufficiently to have a significant probability of opening without further assistance. But they still can be biologically important if other processes contribute free energy to their opening that is not enough to drive separation in the absence of this partial destabilization. Whereas a probability profile describes only complete strand separation, the SIDD profile contains detailed information regarding the extent of destabilization, both within the S/MAR insert and throughout the entire plasmid. For this reason, we use SIDD profiles in this study.[59,62]

Each SIDD profile is partitioned into two components, corresponding to the S/MAR-containing insert and the vector parts of the plasmid. As described below, a variety of measures of the destabilization experienced by each component are evaluated, and those that most closely correlate with the experimentally measured S/MAR activity are determined.

### Correlating SIDD properties with binding activity

Previous calculations have shown that sequences exhibiting S/MAR-binding activity have certain apparently characteristic SIDD attributes.[59] First, the S/MAR-binding sequences analyzed in that study all showed high propensities for destabilization extending over long regions of DNA. This suggests that the strength of *in vitro* scaffold binding may depend in some as yet undetermined way upon both the size of the destabilized region and the extent of its destabilization. Second, within this region there is an internal, strongly destabilized core that corresponds to the CUE. Both of these observations are in accord with experimental results.

Here, we use these insights to design various destabilization measures whose correlations with S/MAR binding strength are evaluated. This process involved three steps:

(1) To develop measures $m$ of the overall extent of destabilization of the SIDD sites within an S/MAR
(2) To specify a functional form expressing the experimentally measured binding activity $a$ as a function of destabilization, $a = F(m)$
(3) To assess the goodness of fit between this function and the experimentally measured binding activity.

This will be done systematically for a wide variety of destabilization measures, $m$, and functions $F$, and those giving the best fit will be determined. Considerations involved at each step include the following.

#### *Step 1. Destabilization measures*

To quantify the extent of destabilization, one must first specify a threshold value $G_d$ that must be surpassed (i.e. $G(x) < G_d$) for base-pair $x$ to be regarded as destabilized. One then finds all regions in each component (i.e. insert *versus* vector) of the plasmid that satisfy this criterion. Different threshold values can be used for different purposes. A low threshold may be used to define a strongly destabilized core, while a more moderate value may be used to delineate the boundaries of the wider destabilized region.

Next, there are several ways to measure the extent of destabilization found in each region. One may use extrema, lengths or areas for this purpose. The extremum measure is the absolute minimum value $G_{min}$ that $G(x)$ attains within a region. A length measure counts the number of base-pairs within the region that are destabilized according to the threshold criterion being applied. Area measures determine the area between the curve $G(x)$ and a threshold value $G_t$. A measure can be applied to the entire sequence in each component of the plasmid, giving one aggregate value for the insert and another for the vector. Alternatively, measures can be associated to individual contiguous destabilized regions within either component.

Previous work suggests that multiple measures may be needed, some to characterize putative CUEs and others to treat their encompassing destabilized regions. The extensive destabilization seen at S/MARs suggests that a length or area measure may need to surpass a threshold before scaffold binding activity occurs. For example, the criterion for a region to be regarded either as destabilized

or as a putative core might be that it contains a site where $G(x)$ falls below a specified value for at least $N_{min}$ contiguous base-pairs. Then the extent of destabilization of a (possibly larger) region containing a site of this type could be assessed by determining the area between the $G(x)$ curve and a (possibly different) threshold value. Clearly, both the locations that are regarded as destabilized and the values of the destabilization measures associated with them will depend on the threshold value(s) used, so calculations are made for each choice within a reasonable range. We also ascribe a minimum value to each type of measure that must be exceeded before active binding is regarded as possible.

#### *Step 2. Functional forms*

There are several alternative functional forms that could express the relationship between the extent of destabilization, as calculated using the above measures, and the experimental *in vitro* binding activity *a* of S/MARs. In the analyses reported below we considered four different possibilities; linear, exponential, sigmoidal and logarithmic.

Previous work has indicated that the presence of nearby destabilized sites synergistically increases the S/MAR-binding activity of a region.[65,69] This effect depends on the number and relative positions of sites involved, and decreases with increasing separation distance. So, we specifically investigate the functional form and distance-dependence of this synergistic enhancement.

The relative strengths of destabilization of different S/MAR elements can be compared by determining how each competes with destabilization in the vector component of the plasmid, which thus serves as an internal standard. For this purpose, the same definition of a destabilized region and the same measure of destabilization must be used throughout. Possible measures of competitiveness include the fraction of the total plasmid destabilization that occurs in the insert, and the ratio of insert destabilization to vector destabilization.

#### *Step 3. Goodness of fit*

As these considerations show that there are many possible ways to assess the relationships between the destabilization characteristics of S/MARs and their binding properties. In this study, we systematically investigate which properties correlate most closely with scaffold binding strength in each of five series of S/MARs. We apply the same SIDD measures and functional relationships to each element within a series, and assess the goodness of fit of the results to the experimental binding activities by determining the correlation coefficient $r$ and the coefficient of determination $r^2$. The latter parameter gives the fraction of the variability that can be ascribed to a linear relationship between the SIDD measure and the experimental activity, so the closer $r^2$ is to 1, the better the fit. (It is sufficient to fit only with linear regression because different functional forms are considered in step 2 above.)

## Acknowledgements

This work was supported, in part, by grants to C.J.B. from the National Science Foundation (DBI-0416764), and from the National Institutes of Health (R01-HG01973). J.B.O. received support from Deutsche Forschungsgemeinschaft (BO 419-6), BMBF (DHGP I/II), INTAS (011-0279), and from the Epi-Vector STREP Program in the EU-FW6.

## References

1. Berezney, R. & Coffey, D. S. (1974). Identification of a nuclear protein matrix. *Biochem. Biophys. Res. Commun.* **60**, 1410–1417.
2. Mirkovitch, J., Mirault, M. E. & Laemmli, U. K. (1984). Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell*, **39**, 223–232.
3. Berezney, R., Mortillaro, M. J., Ma, H., Wei, X. & Samarabandu, J. (1995). The nuclear matrix: a structural milieu for genomic function. In *Structural and Functional Organization of the Nuclear Matrix–International Review of Cytology* (Jeon, K. W. & Berezney, R., eds), vol. 162A, pp. 1–65, Academic Press, Orlando.
4. Jackson, D. A. & Cook, P. R. (1995). The structural basis of nuclear function. In *Structural and Functional Organization of the Nuclear Matrix–International Review of Cytology* (Jeon, K. W. & Berezney, R., eds), vol. 162A, pp. 125–149, Academic Press, Orlando.
5. van Driel, R., Wansink, D. G., van Steensel, B., Grande, M. A., Schul, W. & de Jong, L. (1995). Nuclear domains and the nuclear matrix. In *Structural and Functional Organization of the Nuclear Matrix–International Review of Cytology* (Jeon, K. W. & Berezney, R., eds), vol. 162A, pp. 151–189, Academic Press, Orlando.
6. Gohring, F., Schwab, B. L., Nicotera, P., Leist, M. & Fackelmayer, F. O. (1997). The novel SAR-binding domain of scaffold attachment factor A (SAF-A) is a target in apoptotic nuclear breakdown. *EMBO J.* **16**, 7361–7371.
7. Martelli, A. M., Bortul, R., Fackelmayer, F. O., Tazzari, P. L., Bareggi, R., Narducci, P. & Zweyer, M. (1999). Biochemical and morphological characterization of the nuclear matrix from apoptotic HL-60 cells. *J. Cell. Biochem.* **72**, 35–46.
8. Bode, J., Goetze, S., Heng, H., Krawetz, S. & Benham, C. (2003). From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. *Chromosome Res.* **11**, 435–445.
9. Bode, J., Goetze, S., Ernst, E., Huesemann, Y., Baer, A., Seibler, J. & Mielke, C. (2003). Architecture and utilization of highly-expressed genomic sites. In *New Comprehensive Biochemistry* (Bernardi, G., ed.), vol. 38, pp. 551–572, Elsevier, Amsterdam.
10. Antes, T. J., Namicu, S. J., Fournier, R. E. K. & Levy-Wilson, B. (2000). The 5 boundary of the human apolipoprotein B chromatin domain in intestinal cells. *Biochemistry*, **23**, 6731–6742.
11. Goetze, S., Baer, A., Winkelmann, S., Nehlsen, K., Seibler, J., Maass, K. & Bode, J. (2005). Genomic bordering elements: their performance at pre-defined genomic loci. *Mol. Cell. Biol.* **25**, 2260–2272.
12. Bode, J., Benham, C., Knopp, A. & Mielke, C. (2000). Transcriptional augmentation: modulation of gene expression by scaffold/matrix attached regions (S/MAR elements). *Crit. Rev. Eukaryot. Gene Expr.* **10**, 73–90.
13. Dang, Q., Auten, J. & Plavec, I. (2000). Human beta interferon scaffold attachment region inhibits de novo

methylation and confers long-term, copy number-dependent expression to a retroviral vector. *J. Virol.* **74**, 2671–2678.
14. Kurre, P., Morris, J., Thomasson, B., Kohn, D. B. & Kiem, H. P. (2003). Scaffold attachment region-containing retrovirus vectors improve long-term proviral expression after transplantation of GFP-modified CD34+ baboon repopulating cells. *Blood*, **102**, 3117–3119.
15. Gasser, S. M. & Laemmli, U. K. (1986). Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster*. *Cell*, **46**, 521–530.
16. Jenuwein, T., Forrester, W. C., Fernandez-Herrero, L. A., Laible, G., Dull, M. & Grosschedl, R. (1997). Extension of chromatin accessibility by nuclear matrix attachment regions. *Nature*, **385**, 269–272.
17. Forrester, W. C., Van Genderen, C., Jenuwein, T. & Grosschedl, R. (1994). Dependence of enhancer-mediated transcription of the immunoglobulin Mu gene on nuclear matrix attachment regions. *Science*, **265**, 1221–1225.
18. Fernandez, L. A., Winkler, M. & Grosschedl, R. (2001). Matrix attachment region-dependent function of the immunoglobulin mu. *Mol. Cell. Biol.* **21**, 196–208.
19. Jenke, B. H., Fetzer, C. P., Jönsson, F., Fackelmayer, F. O., Conradt, H. C., Bode, J. & Lipps, H. J. (2001). An episomally replicating vector binds to the nuclear matrix protein SAF-A *in vivo*. *EMBO Rep.* **3**, 349–354.
20. Lipps, H. J. & Bode, J. (2001). Exploiting chromosomal and viral strategies: the design of safe and efficient non-viral gene transfer systems. *Curr. Opin. Mol. Therap.* **3**, 133–141.
21. Bode, J., Fetzer, C. P., Nehlsen, K., Scinteie, M., Hinrich, B. H., Baiker, A. *et al*. (2001). The Hitchhiking Principle: optimizing episomal vectors for the use in gene therapy and biotechnology. *Int. J. Gene Ther. Mol. Biol.* **6**, 33–46.
22. Käs, E., Poljak, L., Adachi, Y. & Laemmli, U. K. (1993). A model for chromatin opening: stimulation of topoisomerase II and restriction enzyme cleavage of chromatin by distamycin. *EMBO J.* **12**, 115–126.
23. Bode, J., Stengert-Iber, M., Schlake, T., Kay, V. & Dietz-Pfeilstetter, A. (1996). Scaffold/matrix-attached regions: Topological switches with multiple regulatory functions. *Crit. Rev. Eukaryot. Gene Expr.* **6**, 115–138.
24. Kalos, M. & Fournier, R. E. K. (1995). Position-independent transgene expression mediated by boundary elements from the apolipoprotein B chromatin domain. *Mol. Cell. Biol.* **15**, 198–207.
25. Li, Q., Zhang, M., Han, H., Rohde, A. & Stamatoyannopoulos, G. (2002). Evidence that DNase I hypersensitive site 5 of the human β-globin locus control region functions as a chromosomal insulator in transgenic mice. *Nucl. Acids Res.* **30**, 2484–2491.
26. Vassetzky, Y., Hair, A. & Mechali, M. (2000). Rearrangement of chromatin domains during development in Xenopus. *Genes Dev.* **14**, 1541–1552.
27. Cai, S., Han, H. J. & Kohwi-Shigematsu, T. (2003). Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nature Genet.* **34**, 42–51.
28. Schübeler, D., Mielke, C., Maass, K. & Bode, J. (1996). Scaffold/matrix-attached regions act upon transcription in a context-dependent manner. *Biochemistry*, **35**, 11160–11169.
29. Klehr, D., Maass, K. & Bode, J. (1991). SAR elements from the human IFN-ß domain can be used to enhance the stable expression of genes under the control of various promoters. *Biochemistry*, **30**, 1264–1270.
30. Stief, A., Winter, D. M., Strätling, W. H. & Sippel, A. E. (1989). A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature*, **341**, 343–345.
31. Schlake, T., Klehr-Wirth, D., Yoshida, M., Beppu, T. & Bode, J. (1994). Gene expression within a chromatin domain: the role of core histone hyperacetylation. *Biochemistry*, **33**, 4187–4196.
32. Sippel, A. E., Schaefer, G., Faust, N., Saueressig, H., Hecht, A. & Bonifer, C. (1993). Chromatin domains constitute regulatory units for the control of eukaryotic genes. *Cold Spring Harbor Symp. Quant. Biol.* **58**, 37–44.
33. McKnight, R. A., Shamay, A., Sankaran, L., Wall, R. J. & Hennighausen, L. (1992). Matrix-attachment regions can impart position-independent regulation of a tissue-specific gene in transgenic mice. *Proc. Natl Acad. Sci. USA*, **89**, 6943–6947.
34. McKnight, R. A., Spencer, M., Wall, R. J. & Hennighausen, L. (1996). Severe position effects imposed on a 1 kb mouse whey acidic protein gene promoter are overcome by heterologous matrix attachment regions. *Mol. Reprod. Dev.* **44**, 179–184.
35. Kirillov, A., Kistler, B., Mostoslavsky, R., Cedar, H., Wirth, T. & Bergman, Y. (1996). A role for nuclear NF-6B in B- cell specific demethylation of the Igkappa locus. *Nature Genet.* **13**, 435–441.
36. Mielke, C., Kohwi, Y., Kohwi-Shigematsu, T. & Bode, J. (1990). Hierarchical binding of DNA fragments derived from scaffold-attached regions: Correlations of properties *in vitro* and function *in vivo*. *Biochemistry*, **29**, 7475–7485.
37. Kay, V. & Bode, J. (1995). Detection of scaffold-attached regions (SARs) by *in vitro* techniques; activities of these elements *in vivo*. In *Methods in Molecular and Cellular Biology: Methods for Studying DNA-Protein Interactions — An Overview* (Papavassiliou, A. G. & King, S. L., eds), vol. 5, pp. 186–194, John Wiley & Sons, Inc. Hoboken, NJ.
38. Allen, G. C., Hall, G., Jr, Michalowski, S., Newman, W., Spiker, S., Weissinger, A. K. & Thompson, W. F. (1996). High- level transgene expression in plant cells: Effects of a strong scaffold attachment region from tobacco. *Plant Cell*, **8**, 899–913.
39. 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.
40. Heng, H. H. Q., Goetze, S., Ye, C. J., Lu, W., Liu, G., Bremer, S., Hughes, M., Bode, J. & Krawetz, S. A. (2004). Dynamic features of scaffold/matrix attached regions (S/MARs) in anchoring chromatin loops. *J. Cell Sci.* **17**, 999–1008.
41. Scheuermann, R. H. & Chen, U. (1989). A developmental-specific factor binds to suppressor sites flanking the immunoglobulin heavy-chain enhancer. *Genes Dev.* **3**, 1255–1266.
42. Zong, R. T. & Scheuermann, R. H. (1995). Mutually exclusive interaction of a novel matrix attachment region binding protein and the NF-MuNR enhancer repressor–implications for regulation of immunoglobulin heavy chain expression. *J. Biol. Chem.* **270**, 24010–24018.
43. Bode, J. & Maass, K. (1988). Chromatin domain surrounding the human interferon-beta gene as defined by scaffold-attached regions. *Biochemistry*, **27**, 4706–4711.

44. Bode, J., Schlake, T., Ríos-Ramírez, M., Mielke, C., Stengert, M., Kay, V. & Klehr-Wirth, D. (1995). Scaffold/matrix-attached regions (S/MARs): structural properties creating transcriptionally active loci. In *Structural and Functional Organization of the Nuclear Matrix–International Review of Cytology* (Jeon, K. W. & Berezney, R., eds), vol. 162A, pp. 389–453, Academic Press, Orlando.
45. Kramer, J. A., Singh, G. B. & Krawetz, S. A. (1996). Computer-assisted search for sites of nuclear matrix attachment. *Genomics*, **33**, 305–308.
46. Singh, G. B., Kramer, J. A. & Krawetz, S. A. (1997). Mathematical model to predict regions of chromatin attachment to the nuclear matrix. *Nucl. Acids Res*, **25**, 1419–1425.
47. 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.
48. Bode, J., Kohwi, Y., Dickinson, L., Joh, R. T., Klehr, D., Mielke, C. & Kohwi-Shigematsu, T. (1992). Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science*, **255**, 195–197.
49. Michalowski, S. M., Allen, G. C., Hall, G. E., Thompson, W. F. & Spiker, S. (1999). Characterization of randomly-obtained matrix attachment regions (MARs) from higher plants. *Biochemistry,* **38**, 12795–12804.
50. 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,* **42**, 154–166.
51. Liebich, I., Bode, J., Frisch, M. & Wingender, E. (2002). S/MARt DB-A tdatabase on scaffold/matrix attached regions. *Nucl. Acids Res.* **30**, 372–374.
52. Liebich, I., Bode, J., Reuter, I. & Wingender, E. (2002). Evaluation of sequence motifs found in scaffold/matrix attached regions (S/MAR). *Nucl. Acids Res.* **30**, 3433–3442.
53. Ganapathy, S. & Singh, G. (2001). Statistical mining of S/MARt Database. In *Proceedings of the Atlantic Symposium on Computational Biology and Genome Information System Technology (CBGIST), March 15–17 2001*, pp. 235–239, Duke University, Durham, NC.
54. Kohwi-Shigematsu, T., Gelinas, R. & Weintraub, H. (1983). Detection of an altered DNA conformation at specific sites in chromatin and supercoiled DNA. *Proc. Natl Acad. Sci. USA*, **80**, 4389–4393.
55. Kohwi-Shigematsu, T. & Nelson, J. A. (1988). The chemical carcinogene, chloroacetaldehyde, modifies a specific site within the regulatory sequence of human cytomegalovirus major immediate early gene *in vivo*. *Mol. Carcinog.* **1**, 20–25.
56. Michelotti, G. A., Michelotti, E. F., Pullner, A., Duncan, R. C., Eick, D. & Levens, D. (1996). Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene *in vivo*. *Mol. Cell. Biol.* **16**, 2656–2669.
57. Kouzine, E., Liu, J., Sanford, S., Chung, H. J. & Levens, D. (2004). The dynamic response of upstream DNA to transcription-generated torsional stress. *Nature Struct. Mol. Biol.* **11**, 1092–1100.
58. Paul, A.-L. & Ferl, R. J. (1993). Osmium tetroxide footprinting of a scaffold attachment region in the maize Adh1 promoter. *Plant Mol. Biol.* **22**, 1145–1151.
59. Benham, C., Kohwi-Shigematsu, T. & Bode, J. (1997). Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions. *J. Mol. Biol.* **274**, 181–196.
60. Benham, C. J. (1992). Energetics of the strand separation transition in superhelical DNA. *J. Mol. Biol.* **225**, 835–847.
61. Benham, C. (1993). Sites of predicted stress-induced DNA duplex destabilization occur preferentially at regulatory loci. *Proc. Natl Acad. Sci. USA*, **90**, 2999–3003.
62. Benham, C. J. (1996). Duplex destabilization in superhelical DNA is predicted to occur at specific transcriptional regulatory regions. *J. Mol. Biol.* **255**, 425–434.
63. Klar, M., Stellamanns, E., Ak, P., Gluch, A. & Bode, J. (2005). Dominant genomic structures: detection and potential signal functions in the interferon-beta chromatin domain. *Gene*. **364**, 79–89.
64. Klar, M. & Bode, J. (2005). Enhanceosome formation over the interferon-beta promoter underlies a remote-control mechanism mediated by YY1 and YY2. *Mol. Cell. Biol.* **25**, 10159–10170.
65. Bode, J., Benham, C., Ernst, E., Knopp, A., Marschalek, R., Strick, R. & Strissel, P. (2000). Fatal connections: when DNA ends meet on the nuclear matrix. *J. Cell Biochem.* **35**, 3–22.
66. Dickinson, L., Joh, T., Kohwi, Y. & Kohwi-Shigematsu, T. (1992). A tissue-specific MAR/SAR binding protein with unusual binding site recognition. *Cell*, **70**, 631–645.
67. Galande, S. & Kohwi-Shigematsu, T. (1999). Poly(-ADP-ribose) polymerase and Ku autoantigen form a complex and synergistically bind to matrix attachment sequences. *J. Biol. Chem.* **274**, 20521–20528.
68. Kowalski, D., Natale, D. A., Eddy, M. J. & Stable, D. N. A. (1988). unwinding, not "breathing" accounts for single-strand-specific nuclease hypersensitivity of specific A+T-rich sequences. *Proc. Natl Acad. Sci. USA*, **85**, 9464–9468.
69. Okada, S., Tsutsui, K., Tsutsui, K., Seki, S. & Shohmori, T. (1996). Subdomain structure of the matrix attachment region located within the mouse immunoglobulin kappa gene intron. *Biochem. Biophys. Res. Commun.* **222**, 472–477.
70. Meier, I. (2001). The plant nuclear envelope. *Cell Mol. Life Sci.* **58**, 1774–1780.
71. Kamb, A. (1995). Cell-cycle regulators and cancer. *Trends Genet*. **11**, 136–140.
72. Kaufmann, H. S. & Shaper, J. H. (1984). A subset of non-histone nuclear proteins reversibly stabilized by the sulfhydryl cross-linking reagent tetrathionate. Polypeptides of the internal nuclear matrix. *Expt. Cell Res.* **155**, 477–495.
73. Fey, E. G., Krochkalnic, G. & Penman, S. (1986). The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. *J. Cell. Biol.* **102**, 1654–1665.
74. Belgrader, P., Siegel, A. J. & Berezney, R. (1991). A comprehensive study on the isolation and characterization of the HeLa S3 nuclear matrix. *J. Cell Sci.* **98**, 281–291.
75. Ivanchenko, M. & Avramova, Z. (1992). Interaction of MAR- sequences with nuclear matrix proteins. *J. Cell. Biochem.* **50**, 190–200.
76. Mattern, K. A., Humbel, B. M., Muijsers, A. O., deJong, L. & van Driel, R. (1996). hnRNP proteins and B23 are the major proteins of the internal nuclear matrix of HeLa S3 cells. *J. Cell. Biochem.* **62**, 275–289.

77. Mattern, K. A., van Driel, R. & de Jong, L. (1997). Composition and structure of the internal nuclear matrix. In *Nuclear Structure and Gene Expression*, pp. 87–110, Academic Press, New York.
78. Gerner, C., Holzmann, K., Grimm, R. & Sauermann, G. (1998). Similarity between nuclear matrix proteins of various cells revealed by an improved isolation method. *J. Cell. Biochem.* **71**, 363–374.
79. He, D., Nickerson, J. A. & Penman, S. (1990). Core filaments of the nuclear matrix. *J. Cell. Biol.* **110**, 569–580.
80. Wang, X. & Traub, P. (1991). Resinless section immunogold electron microscopy of karyo-cytoskeletal frameworks of eukaryotic cells cultured *in vitro*. Absence of a salt-stable nuclear matrix from mouse plasmacytoma MPC-11 cells. *J. Cell. Sci.* **98**, 107–122.
81. Hozak, P., Sasseville, A. M.-J., Raymond, Y. & Cook, P. R. (1995). Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells.. *J. Cell Sci.* **108**, 635–644.
82. Gueth-Hallonet, C., Wang, J., Harborth, J., Weber, K. & Osborn, N. (1998). Induction of a regular nuclear lattice by overexpression of NuMA. *Expt. Cell Res.* **243**, 434–452.
83. Nakayasu, H. & Ueda, K. (1986). Preferential association of acidic actin with nuclei and nuclear matrix from mouse leukemia L5178 Y cells. *Expt. Cell Res.* **163**, 327–336.
84. Kipp, M., Gohring, F., Ostendorp, T., van Drunen, C. M., van Driel, R., Przybylski, M. & Fackelmayer, F. O. (2000). SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA. *Mol. Cell. Biol.* **20**, 7480–7489.
85. Nakayasu, H. & Berezney, R. (1991). Nuclear matrins: Identification of the major nuclear matrix proteins. *Proc. Natl Acad. Sci. USA*, **88**, 10312–10316.
86. Tsutsui, K., Tsutsui, K., Okada, S., Watarai, S., Seki, S., Yasuda, T. & Shohmori, T. (1993). Identification and characterization of a nuclear scaffold protein that binds the matrix attachment region DNA. *J. Biol. Chem.* **268**, 12886–12894.
87. Kiledjian, M. & Dreyfuss, G. (1992). Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.* **11**, 2655–2664.
88. Kay, V. & Bode, J. (1994). Binding specificity of a nuclear scaffold: supercoiled, single-stranded, and scaffold-attached- region DNA. *Biochemistry*, **33**, 367–374.
89. Bode, J., Bartsch, J., Boulikas, T., Iber, M., Mielke, C., Schübeler, D. *et al*. (1998). Transcription- promoting genomic sites in mammalia: their elucidation and architectural principles. *Gene Ther. Mol. Biol.* **1**, 551–580.
90. Herrscher, R. F., Kaplan, M. H., Lelsz, D. L., Das, C., Scheuermann, R. & Tucker, P. W. (1995). The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: A B cell-specific transactivator that describes a new DNA-binding protein family. *Genes Dev.* **9**, 3067–3082.
91. Will, K., Warnecke, G., Wiesmuller, L. & Deppert, W. (1998). Specific interaction of mutant p53 with regions of matrix attachment region DNA elements (MARs) with a high potential for base-unpairing. *Proc. Natl Acad. Sci. USA*, **95**, 13681–13686.
92. Gohler, T., Jager, S., Warnecke, G., Yasuda, H., Kim, E. & Deppert, W. (2005). Mutant p53 proteins bind DNA in a DNA structure-selective mode. *Nucl. Acids Res.* **33**, 1087–1100.
93. Dickinson, L. A., Dickinson, C. D. & Kohwi-Shigematsu, T. (1997). An atypical homeodomain in SATB1 promotes specific recognition of the key structural element in a matrix attachment region. *J. Biol. Chem.* **272**, 11463–11470.
94. Deppert, W. (1996). Binding of MAR-DNA elements by mutant p53: possible implications for its oncogenic functions. *J. Cell. Biochem.* **62**, 172–180.
95. Dickinson, L. A. & Kohwi-Shigematsu, T. (1995). Nucleolin is a matrix attachment region DNA-binding protein that specifically recognizes a region with high base-unpairing potential. *Mol. Cell. Biol.* **15**, 456–465.
96. Zao, K., Kaes, E., Gonzalez, E. & Laemmli, U. K. (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.
97. Yanagisawa, J., Ando, J., Nakayama, J., Kohwi, Y. & Kohwi-Shigematsu, T. (1996). A matrix attachment region (MAR)- binding activity due to a p114 kilodalton protein is found only in human breast carcinomas and not in normal and benign breast disease tissues. *Cancer Res.* **56**, 457–462.
98. Jenke, A. W., Stehle, I. M., Eisenberger, T., Baiker, A., Bode, J., Fackelmayer, F. O. & Lipps, H. J. (2004). Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome. *Proc. Natl Acad. Sci. USA*, **101**, 11322–11327.

***Edited by G. von Heijne***

*(Received 28 July 2005; received in revised form 2 November 2005; accepted 4 November 2005)*
Available online 9 December 2005