J. Mol. Biol. (1997) **274**, 181–196

JMB

Stress-induced Duplex DNA Destabilization in Scaffold/Matrix Attachment Regions

Craig Benham[1]*, Terumi Kohwi-Shigematsu[2] and Jürgen Bode[3]*

[1]*Department of Biomathematical Sciences Box 1023, Mount Sinai School of Medicine, 1 Gustave Levy Place, New York, NY 10029 USA*

[2]*Life Science Division Lawrence Berkeley National Laboratory, Building 70A Room 4405, 1 Cyclotron Road University of California Berkeley, CA 94720, USA*

[3]*GBF, Gesellschaft für Biotechnologische Forschung mbH, Genregulation und Differenzierung/Genetik von Eukaryonten, Mascheroder Weg 1, D-38124 Braunschweig Germany*

**Corresponding authors*

S/MARs are DNA elements 300 to several thousand base-pairs long, which are operationally defined by their affinity for the nuclear scaffold or matrix. S/MARs occur exclusively in eukaryotic genomes, where they mediate several functions. Because S/MARs do not have a clearcut consensus sequence, the characteristics that define their activity are thought to be structural. Ubiquitous S/MAR binding proteins have been identified, but to date no unique binding sequence or structural motif has been found. Here we show by computational analysis that S/MARs conform to a specific design whose essential attribute is the presence of stress-induced base-unpairing regions (BURs). Stress-induced destabilization (SIDD) profiles are calculated using a previously developed statistical mechanical procedure in which the superhelical deformation is partitioned between strand separation, twisting within denatured regions, and residual superhelicity. The results of these calculations show that BURs exhibit a succession of evenly spaced destabilized sites that would render part or all of the S/MAR sequence single stranded at sufficient superhelicity. These analyses are performed for a range of sequenced S/MAR elements from the borders of eukaryotic gene domains, from centromeres, and from positions where S/MARs are known to support the action of an enhancer. The results reported here are in excellent agreement with earlier *in vitro* chemical reactivity studies. This approach demonstrates the potential for computational analysis to predict the points of division of the eukaryotic genome into functional units (domains), and also to locate certain *cis*-regulatory sequences.

Keywords: scaffold-attached regions; nuclear scaffold; chromatin domain; DNA duplex stability; DNA superhelicity.

Introduction

Vertebrate chromatin is organized into loops by periodic attachment to the nuclear scaffold or matrix at positions whose average separations are approximately 60 kb in somatic cells (Gasser & Laemmli, 1987; Mirkovitch *et al.*, 1987; Vogelstein *et al.*, 1980), and 27 kb in the male haploid genome (Barone *et al.*, 1994). The DNA elements mediating this attachment have been termed scaffold-attached or matrix associated regions (S/MARs). S/MARs are unique features of eukaryotic genomes, as demonstrated by the observation that S/MAR–scaffold interactions cannot be disrupted by an up to 60,000-fold excess of double stranded bacterial DNA (Kay & Bode, 1995). Recently it has been demonstrated that centromeric regions have clusters of attachment sites concentrated along the chromosomal axis (Strissel *et al.*, 1996).

S/MAR elements are associated with a variety of biological functions, apparently in consequence either of their structure and/or of their interactions with proteins. Experiments suggest they play important roles during gene expression. S/MARs are thought to support cell type-specific expression of genes (Bonifer *et al.*, 1994; McKnight *et al.*, 1992,

Abbreviations used: α, linking difference; σ, superhelical density; ARS, autonomously replicating sequence; bp, base pair; BRIGHT, B cell regulator of IgH transcription; BUR, base-unpairing region; CAA, chloroacetaldehyde; CUE, core-unwinding element; huIFN-β, human interferon-β; IFNB, gene encoding IFN-β; IR, inverted repeat; LINE, long interspersed reiterated element; ORF, open reading frame; ORI, origin of replication; SAR, scaffold-attached region; MAR, matrix-associated region; S/MAR, consensus term covering SARs and MARs; NF-μNR, nuclear factor-μ negative regulator; SIDD, stress-induced duplex destabilization.

0022–2836/97/470181–16 $25.00/0/mb971385

1996) and have been implicated in gene switching during development (Boulikas, 1995). Evidence from several laboratories shows that some S/MARs coexist with enhancers (Gasser & Laemmli, 1986). This association is particularly intriguing, as most or all S/MARs also have the capacity to augment transcription *via* a non-enhancer mechanism (for a review, see Bode *et al*., 1995). The most thoroughly studied example involves the immunoglobulin κ and μ-chain intronic enhancers, which are associated with one and with two distinct S/MAR elements, respectively (Cockerill & Garrard, 1986a,b; Cockerill *et al*., 1987). They function in domain opening (Bode *et al*., 1996; Käs *et al*., 1993), which occurs in forming accessible chromatin domains during embryonic development (Forrester *et al*., 1994; Jenuwein *et al*., 1993). The regional demethylation that must occur as a prerequisite for opening utilizes distinct *cis*-acting modules, including the intronic enhancer element and the S/MAR (Jenuwein *et al*., 1997; Lichtenstein *et al*., 1994). While any S/MAR sequence appears to be able to function in this reaction, tissue specificity is mediated by sequences within the intronic enhancer (Kirillov *et al*., 1996). Some, and possibly all, of these enhancer-related activities are regulated by specific S/MAR binding factors. Among these, the nuclear factor-μ negative regulator (NF-μNR) and the B cell regulator of IgH transcription (BRIGHT) show a reciprocal lymphoid expression pattern. NF-μNR is expressed in non-B cells, where it attenuates the enhancer (Scheuermann & Chen, 1989; Zong & Scheuermann, 1995), while BRIGHT upregulates IgH expression in mature B cells (Herrscher *et al*., 1995). The sites specific for BRIGHT also are recognized by SATB1, which was the first protein for which a S/MAR-related association was noted (Dickinson *et al*., 1992; Herrscher *et al*., 1995).

Possible roles for S/MARs in replication are suggested by several observations. ORI sites mapped in various eukaryotic genomes appear to coincide with S/MARs. In yeast, chromosomal replication initiates at a subset of the ARS elements present in the genome for which S/MAR functions have been demonstrated (Amati *et al*., 1990). In *Drosophila* there also is a close correlation between ARS and S/MAR activity (Amati and Gasser, 1990). Although in most assays in higher eukaryotes replication initiation seems to occur in broad regions, an overlap between ORI and S/MAR functions at the DHFR replication origin has been demonstrated (Dijkwel and Hamlin, 1995). Active chromosomal ORIs are permanently associated with the matrix (Carri *et al*., 1986). DNA is reeled through the replication machinery, which is a part of the matrix (Hozak & Cook, 1994).

To date, the identification of genomic regions associating with the nuclear matrix has relied primarily on biochemical studies. DNA segments with an affinity for the nuclear matrix have been recovered by a variety of assays (for reviews, see Boulikas, 1995; Kay & Bode, 1995). The associations of active genomic regions with the scaffold are mediated both by S/MARs and by a distinct preference of the nuclear matrix for supercoiled DNA as it occurs during transcription (Kay & Bode, 1994; Tsutsui *et al*., 1988). Assays designed to monitor attachment differ in their ability to recover constitutive and transcription-dependent forms of DNA association.

Sequence searches have been inadequate at revealing putative scaffold attachment sites because, although several characteristic motifs are known, including A+T-richness, no unique consensus S/MAR sequence has been found (Boulikas, 1993; Kramer & Krawetz, 1995; Kramer *et al*., 1996). A map of the S/MARs on human chromosome 19 supports the view that these sites do not share any common repeat sequences (Nikolaev *et al*., 1996).

S/MAR functional activity may be related instead to topological or structural features that are not strictly linked to primary sequence. Structural properties which commonly occur in A+T-rich regions include natural curvature, a narrow minor groove in oligo(dA) tracts, and a susceptibility to denature (Bode *et al*., 1995; Boulikas, 1993, 1995). Chemical probes and two dimensional gel analyses of S/MAR regions placed in plasmids under superhelical tension show that these elements readily relieve imposed torsional strain by stable base-unpairing. In all the cases that have been experimentally analyzed to date, this unpairing initiated at a nucleation site, here called the core unwinding element or CUE (Bode *et al*., 1992; 1995; 1996). Although all S/MAR sequences appear to share a propensity to form non-B DNA structures under superhelical tension (Boulikas, 1993), in some cases their sequence suggests that this could be either a triple helical H-form or a denatured region. In both structures at least one unpaired strand is present.

S/MAR activities are thought to be mediated by distinct sets of DNA binding proteins that recognize specific structural features such as single strands (Bode *et al*., 1996). Support for this hypothesis comes from observations that decreases in the thermodynamic stability of S/MAR regions correlate both with increases in their strength of binding to nuclear scaffold/matrix preparations *in vitro*, and with their potential to augment transcriptional initiation rates *in vivo* (Allen *et al*., 1996; Bode *et al*., 1992; Mielke *et al*., 1990; Schübeler *et al*., 1996).

Here, we perform a statistical mechanical analysis of S/MARs within genomic sequences to predict their stress-induced duplex-destabilization (SIDD) properties. This procedure has proven to be highly accurate in finding sites of denaturation and predicting the extent of opening at each site, as detected by the Kowalski nuclease assay (Benham, 1992, 1993, 1996a; Kowalski *et al*., 1988). Several specific sequenced S/MARs are analyzed, for which *in vitro* data on base-unpairing and matrix binding are available (Bode *et al*., 1992; Kohwi-Shigematsu & Kohwi, 1990; Mielke *et al*., 1990;). This analysis finds that those locations having

S/MAR activity also exhibit strong destabilization patterns with characteristic lengths and structures. The accuracy of these results is assessed by comparison with experimental data on base-unpairing within, and functional activity of, these S/MAR regions.

Calculations

At thermodynamic equilibrium a population of identical superhelical DNA molecules will be distributed among its available conformational states according to Boltzmann's law. If the states are indexed by i, and if the free energy of state i is G_i, then the governing partition function is given by:

$$Z = \sum_i e^{(-G_i/RT)} \tag{1}$$

where R is the gas constant and T is the absolute temperature. The equilibrium probability p_i of state i, which is its fractional occupancy in a population at equilibrium, equals the relative contribution of its Boltzmann factor to the entire partition function:

$$p_i = \frac{e^{(-G_i/RT)}}{Z} \tag{2}$$

This equation shows that at equilibrium the occupancy of individual states decreases exponentially with increasing free energy.

If a parameter ζ has value ζ_i in state i, then its population average value $\bar{\zeta}$ at equilibrium is:

$$\bar{\zeta} = \sum_i \zeta_i p_i. \tag{3}$$

This expression may be used to evaluate any equilibrium property of interest, once the information required to construct the partition function is known. In particular, one can use this approach to calculate the ensemble average probability $p(x)$ of denaturation of the base-pair at each position x along the DNA sequence. The graph of $p(x)$ *versus* x, called the transition profile, displays the regions of the sequence that have significant probabilities of denaturation. A more sensitive measure of destabilization is given by the incremental energy $G(x)$ needed to separate the base-pair at position x (Benham, 1993; 1996a). This quantity is calculated as:

$$G(x) = \bar{G}(x) - \bar{G} \tag{4}$$

where $\bar{G}$ is the ensemble average free energy of the system and $\bar{G}(x)$ is the average free energy of all states in which the base-pair at position x is denatured. $G(x)$ is negative for base-pairs that are favored to denature, and positive for base-pairs where incremental free energy is needed to assure denaturation. Stress-induced duplex destabilization (SIDD) profiles are plots of $G(x)$ *versus* x. These are generally more informative than transition probability profiles because a region may be substantially destabilized without experiencing significant denaturation. This may be important in cases where another process or molecule can contribute a limited amount of energy towards denaturation. In these circumstances transition might only be possible when the site involved already is substantially destabilized by imposed stresses. Thus, a process that requires single stranded DNA may occur at a site that would not by itself entirely denature, unless other molecules contribute energy to assist the separation process.

In principle this formalism may be used to calculate equilibrium distributions once the states and their energies have been determined. Consider a particular state of a superhelical molecule, in which the secondary structure of each base pair is specified. The superhelical constraint is the linking difference α, which is the number of turns by which the helicity of the domain differs from its relaxed value when its relaxed tertiary structure is maintained. In this analysis the imposed superhelicity is partitioned among three types of deformations: local denaturation, helical interwinding of the unpaired strands within these denatured regions, and residual superhelicity (more complete descriptions of this analysis are presented in Benham, 1990, 1992). The free energy associated with each of these deformations has been determined from experiments. Summing their contributions yields the free energy of the state. Here copolymeric denaturation energetics are used, so the free energy needed to separate each base-pair is regarded as depending only on whether it is A:T or G:C. Also, the torsional deformations of denatured regions are assumed to equilibrate with residual superhelicity.

Because there are 2^N states of denaturation available to a superhelical DNA containing N base-pairs, it is not possible to perform these calculations by exhaustive enumeration for the kilobase length DNA molecules of biological interest. Instead, an approximate statistical mechanical technique is used in which the low energy states are included explicitly, and the cumulative influence of the high energy states is estimated. Once free energies have been associated to the states of the DNA molecule, the state having minimum free energy $G_{\min}$ is found. An energy threshold θ is specified, and all states i are found whose free energies exceed $G_{\min}$ by no more than this threshold amount. An approximate partition function Z_{cal} is computed from this collection of low energy states to be:

$$Z_{\text{cal}} = \sum_{i|G_i - G_{\min} < \theta} e^{-G_i/RT} \tag{5}$$

Approximate ensemble average (i.e. equilibrium) values are computed for all parameters of interest, including the transition and stress-induced duplex destabilization (SIDD) profiles. Although individ-

ual high-energy states are exponentially less populated than low energy states at equilibrium, they are so numerous that their cumulative contribution to the equilibrium may be significant. So the next step in this calculation is to estimate the aggregate influence of the states that do not satisfy the threshold condition. This is done by a density of states procedure described elsewhere (Benham, 1990, 1992). Its results can be used to assess the accuracy of the profiles calculated with a given energy threshold θ. This allows the threshold θ to be selected that gives any required degree of accuracy. In practice transition profiles having accuracies exceeding 99% are feasible to calculate, even for highly supercoiled molecules having lengths of 10 kb or more.

In the analyses of S/MAR-containing DNA sequences reported below, values of the energy parameters are used which are appropriate for the conditions used in Kowalski's nuclease digestion experiments measuring the extent and locations of denatured regions in pBR322 DNA *in vitro* (Kowalski *et al*., 1988). The predictions achieved by calculations that use these energetics have been shown to be in precise quantitative agreement with the results of nuclease digestion experiments performed on other molecules (Benham, 1992). The extensive variations in the locations of separated regions that result from minor sequence alterations are predicted precisely. The relative amounts of transition at each site also are in complete agreement with experimental measurements. The superhelicity required to drive a specific amount of separation agrees to within one turn with the observed value, which reflects the limit of accuracy with which extents of transition can be measured in these experiments.

The high degree of quantitative precision achieved by this approximate statistical mechanical method enables its use to predict the duplex destabilization behavior of other DNA sequences on which experiments have not been performed. Although the resulting predictions are only known to be quantitatively accurate under the experimental conditions of Kowalski's *in vitro* assay, their qualitative conclusions regarding locations of destabilized sites have been found to be quite robust, holding under a wide variety of conditions. These qualitative predictions of destabilized sites illuminate a fundamental physical chemical attribute of the sequence of the domain involved. They do not depend on the manner in which untwisting torsional stress is imposed on the DNA or how strand separation is detected. However, the precise quantitative agreement found with nuclease digestion results may not be achieved when denaturation is detected by other experimental methods or under other environmental conditions than those of the standard Kowalski assay.

Numerous calculations clearly show that sites of predicted stress-induced duplex destabilization are closely associated with several types of DNA regulatory regions (Benham, 1993, 1996a). Here, we document a particularly strong association with matrix attachment regions. Most calculations reported below assume a superhelix density of $\sigma = -0.055$, the value found in extracted bacterial plasmids.

Results

The immunoglobulin heavy chain enhancer-associated S/MARs: a paradigm?

Kohwi-Shigematsu & Kohwi (1990) have demonstrated an overlap between binding sites for the regulator NF-μNR and locations that become stably and uniformly unpaired when the IgH gene region is subjected to torsional stress. Prominent destabilized sites coincide with the strong 3′-S/MAR and the weaker 5′-S/MAR of the IgH enhancer (Cockerill *et al*., 1987; cf. elements XVI_l and XVI_r in Mielke *et al*., 1990).

We have evaluated the destabilization properties of this region when it is inserted in the pBluescript vector. The inserted 992 bp *Xba*I fragment containing the intronic enhancer of the murine IgH gene is located between map positions 2265 and 3256 in this constructed plasmid. This region is roughly subdivided into three sections by the *Hin*f I sites at positions 2609 and 2830: the 5′-S/MAR, the core enhancer, and the 3′-S/MAR. The *Pvu*II (2648) and *Dde*I (2782) sites are somewhat more precise markers for the position of the core enhancer (Dickinson *et al*., 1992). This plasmid also contains the β-lactamase gene and the f1 replication origin.

Those parts of the sequence that derive from the cloning vector show several characteristic destabilization features. Most regions remain stable under superhelical stress, with values of $G(x)$ in the range from 8 to 10 kcal/mol. Three sharply delineated and well separated destabilized sites are found. Two of these coincide with the terminator and promoter of the β-lactamase gene, which were among the first sites where local denaturation was demonstrated by nuclease digestion of superhelical pBR322 DNA (Kowalski *et al*., 1988). The third plasmid-derived peak, centered at position 4000, coincides with the f1 ORI site. Although this site is too short to constitute a S/MAR itself, it contributes synergistically to scaffold binding if other S/MAR-sequences are present (Mielke *et al*., 1990). These three plasmid-derived features recur in all analyses which use this or a related plasmid as a cloning vector. All of these elements have been the subject of earlier computational analyses (Benham, 1993). For present purposes they provide well defined internal standards for assessing destabilization profiles.

The 992 bp insert exhibits a striking tripartite pattern, in which the core enhancer is not destabilized but is flanked by two strongly destabilized regions that coincide with the S/MARs. These are precisely the same regions that have been shown to be accessible to chloracetaldehyde (CAA), a single strand-specific reagent (Figure 1B, bottom).

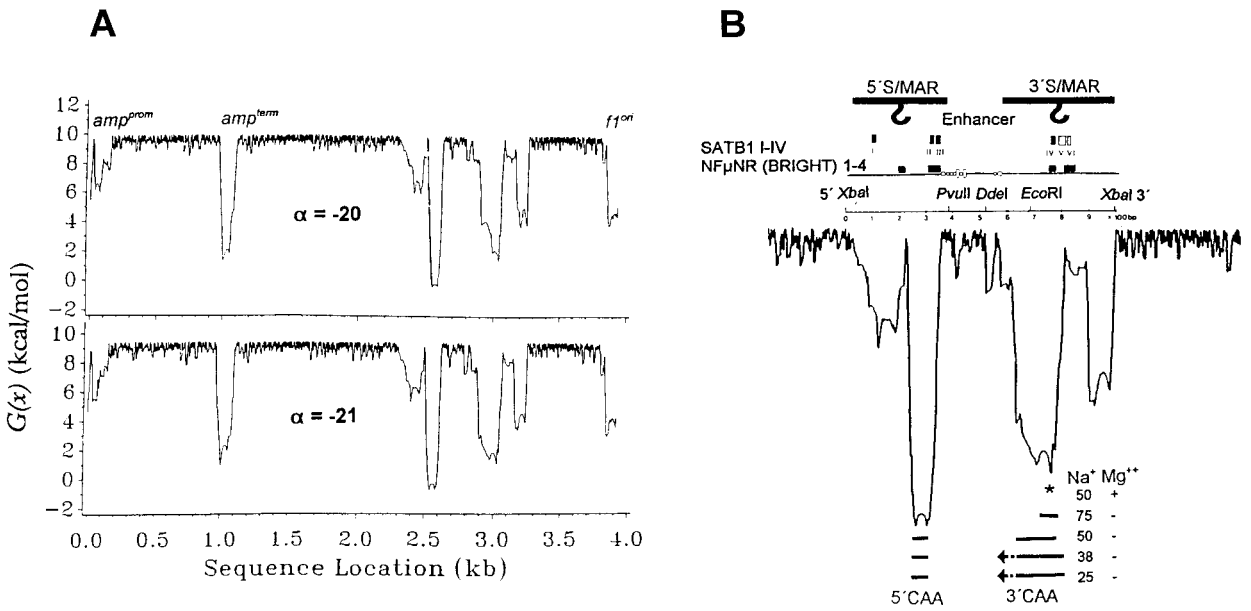

Figure 1. Destabilization profiles for the IgH-enhancer-S/MAR sequence cloned into the circular pBluescript plasmid. A, The results of analyses at linking differences, −20 ($\sigma = -0.058$) and −21 ($\sigma = -0.065$) are shown. B, Magnification of the lower trace from A are superimposed on the data of Kohwi-Shigematsu and Kohwi (1990) and Dickinson *et al.* (1992) . Boxes marked I to VI designate SATB1 binding sites.

Kohwi-Shigematsu & Kohwi (1990) have demonstrated that reactivity at the 3′ S/MAR initiates at the TAATATATTT core unwinding element (CUE, marked by an asterisk), and then spreads unidirectionally in the 5′ direction until it reaches the enhancer border. This directional preference is not readily explained by the relative A+T-contents of the neighboring sequences, which are both 70% A+T. This behavior is precisely reproduced by our calculations (Figure 1), which find that destabilization spreads in the 5′ direction from the CUE as the superhelical deformation is increased. In contrast to the 3′ S/MAR, the unwinding region 5′ to the enhancer revealed an all-or-none reactivity towards CAA that is reflected by the two steep flanks bordering this destabilized region in the calculated profile.

Although the present analysis precisely locates the base-upairing regions, it does not reproduce certain details of the competition between the two CAA reactive segments that have been observed *in vitro*. When both regions were present simultaneously on the same plasmid molecule, the S/MAR region 3′ to the enhancer was found to be significantly more reactive to CAA than was the 5′ S/MAR. Only if the 3′ S/MAR was deleted would the 5′ element become as reactive as the 3′ element had been in the complete construct (Kohwi-Shigematsu & Kohwi, 1990). This contrasts with the results of the calculations, which find the 5′ S/MAR to be slightly more destabilized than the 3′ S/MAR. It is not presently known whether this discrepancy is due to differences between the conditions of the CAA reaction used here and the nuclease digestion procedure on which the calculation energies are based (Benham, 1993). It also might arise from properties specific to the 3′ element.

A series of plasmids has been constructed by inserting into the pBluescript vector *n*-mers of the 25-base-pair sequence TCTTTAATTTC**TAATATATTT**AGAA containing the 3′ S/MAR CUE. These plasmids are denoted as $(25\text{ bp})_n$. A second set of plasmids (denoted $(24\text{ bp}^{\text{mut}})_n$) was constructed by insertion of *n*-mers of the 24 bp sequence in which the CUE sequence TAATATATTT was mutated into the experimentally non-unwinding sequence TACTGC-TTT. These plasmids were originally used for affinity-based isolation of proteins that recognize base-unpairing regions (Dickinson *et al.*, 1992).

Figure 2 shows the results of a computational analysis of superhelical destabilization of the $(25\text{ bp})_n$ oligomer plasmid series at a linking difference of −16 turns, together with the results of an *in vitro* scaffold association assay. The core unwinding element and its oligomers are seen to be significantly destabilized in this context. The size of the destabilized region and the extent of destabilization it experiences increase with the degree of oligomerization (*n*). The probability of strand separation in the oligomeric insert, displayed in the transition profiles in the right-hand panel of Figure 2, depends sensitively on oligomer length. Our calculations predict that a sufficiently long region in this plasmid that has the AT-richness of the insert will contain an average of 68 denatured base-pairs at this linking difference. Only the trimer and longer oligomers have sufficient length to enable 68 base-pairs to open there in a single run. Shorter inserts can accomodate little (monomer), or most but not all (dimer), of the

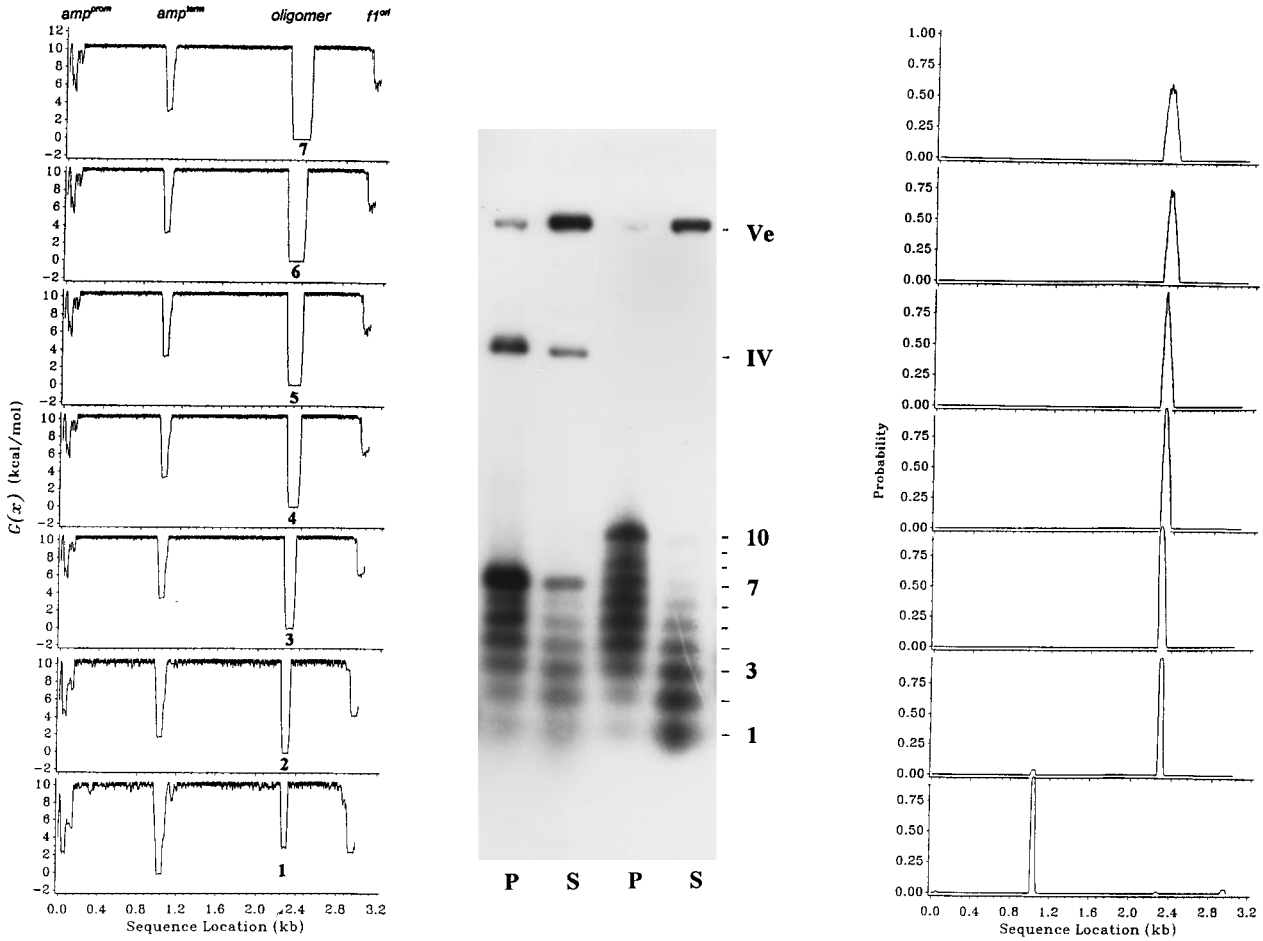

Figure 2. Analysis of oligomerized core unwinding elements *in vitro* and by computation. Center: gels showing the results of two scaffold reassociation analyses performed according to the procedure of Kay & Bode (1995) on oligomers of lengths up to $(25\,bp)_7$ (left) and $(25\,bp)_{10}$ (right), respectively. The oligomers were excised from the pTZ vector (Ve), reassociated with the scaffold and partially degraded *in situ* by *Eco*RI. After electrophoresis on a 2% agarose gel, fragments were transferred to a nylon screen and visualized by hybridization with a nick-translated probe. Oligomers in the P (pellet) lane are associated with the scaffold, whereas those in the S (supernatant) lane are not. The fragment marked Ve is the vector. For calibration, an 800 bp S/MAR-standard (fragment IV of Mielke *et al.*, 1990) has been included. Left panel: destabilization profiles of plasmids containing the 25 bp monomer to heptamer (marked by numerals 1 to 7 from bottom to top), cloned into the *Eco*RV site of pBluescript. Right panel: the site of predicted base unpairing in the same set of constructs is shown in probability profiles. These results mirror the occurrence of the corresponding oligomers in the scaffold attached form (lane P) of the experiment.

superhelical strain. So a change occurs from opening at an alternative site to exclusive opening at the insert site as oligomer length n increases. The value of n at which this happens increases as the imposed superhelicity becomes more extreme. At a linking difference of -18 turns the expected length of the open region is 84 denatured base-pairs. Only the tetramer and longer oligomers are long enough to accomodate 84 unpaired bases (data not shown). Finally, the transition profiles in Figure 2 show that, while the entire tetramer insert is open all the time, longer inserts have a triangular distribution of opening probabilities. This occurs when the length of the insert is substantially greater than the expected number of open base-pairs. Under these circumstances the essentially uniform composition of the insert dictates that the opening region is approximately equally likely to be positioned anywhere within it, leading to a triangular distribution.

To visualize S/MAR activity, a nuclear scaffold was prepared from mouse L-cells by the LIS-extraction procedure detailed in Kay & Bode (1995). Vectors containing either the heptamer, $(25\,bp)_7$, or the decamer, $(25\,bp)_{10}$ were separated into insert- and vector parts by restriction, and the respective fragment mixtures were associated with an aliquot of the nuclear scaffold. Subsequently, lower members of the oligomer series were generated *in situ* by partial *Eco*RI digestion of the DNA-scaffold complex using an appropriate amount of the restriction enzyme. The autoradiograph (Figure 2, center) shows that by this procedure the 25 bp monomer is mostly released from the scaffold (occurrence in the supernatant, or S fraction). Significant retention on the scaffold (occurrence in

the pellet, or P fraction) starts with the dimer and by the tetramer the majority of the oligomer is retained. The affinity of the $(25\,bp)_4$ species for the scaffold is comparable in this assay to that of a natural 800 bp S/MAR fragment (Mielke *et al.*, 1990) The vector control (Ve) is >98% unbound and thus found in the supernatant. These findings show that the strength of the S/MAR–scaffold interaction is reflected by the strand separation properties of the region involved. This strongly supports previous claims that S/MAR activity depends upon having an unwinding propensity extending over a sufficient length of DNA, and perhaps also on an adequate spacing of unwinding elements favoring multisite attachment (Bode *et al.*, 1996; Mielke *et al.*, 1990).

Next, the unwinding $(25\,bp)_7$ oligomer was compared with the non-unwinding mutant $(24\,bp^{mut})_8$ containing the mutagenized ACTGC-TTT core. Both oligomers have similar, very high A+T content. The mutagenized sequence has 18 ATs within its 24 bp repeat unit, while the wild-type has 22 ATs within its 25 bp repeat. In spite of this small difference in A+T content, these plasmids exhibit substantially different destabilization profiles (Figure 3A and B). The insert region is the only place in the $(25\,bp)_7$ plasmid where the duplex is destabilized at high superhelical stresses (linking difference $= -20$ turns). At a linking difference of -16 turns other sites are destabilized, but the insert remains the only one that is sufficiently destabilized to separate with high probability. The behavior of the $(24\,bp^{mut})_8$ plasmid is significantly different. Here destabilization occurs predominantly at the *amp*-terminator around position 1030, with sites near positions 50 and 3100 also being substantially destabilized. The insert is less destabilized than any of these alternate sites, although it becomes more competitive at more extreme linking differences. This is another example of the complexity of stress-induced transition behavior, in which the relative probability of denaturation at the *amp*-terminator does not increase monotonically with the level of destabilizing stress (Benham, 1996a).

Substantial functional differences between these constructs have been found by various assays (Bode *et al.*, 1992). For this S/MAR it has been shown that only the unwinding element confers significant affinity for the nuclear scaffold *in vitro*, and transcriptional activity *in vivo*.

The human interferon-β upstream S/MAR: a prototype domain border

A 2.2 kb *Eco*RI fragment within the upstream region of the human interferon-β (huIFN-β) gene has been shown to possess a strong affinity for the nuclear scaffold. This sequence, which has been called ''I'' in Mielke *et al.* (1990), has served as a prototype S/MAR element in several experimental studies (viz. Dickinson *et al.*, 1992; Kay & Bode, 1994; Kirillov *et al.*, 1996). This region is highly accessible to single-strand specific reagents in the living cell. It appears to function as a topological sink supporting the inducible transcription of the interferon gene (Bode *et al.*, 1995).

The plasmid pTZ_E20 was constructed by cloning this 2.2 kb S/MAR fragment into the *Eco*RI site of the pTZ18R polylinker. Figure 4A shows the destabilization profile of pTZ_E20, calculated at a linking difference of -27 turns. The inserted S/MAR sequence, which lies between map positions 213 and 2413, is predicted to be chaotically destabilized, in striking contrast to the regions of this vector derived from prokaryotic sources. A succession of minima with spacings of 200 to 400 bp extends over the entire length of the insert.

Plasmid pCL was constructed by inserting the 823 bp segment containing the most destabilized region of this S/MAR into the pTZ18R polylinker. (Note that this insertion and the one constructing pCL^{mut}, described below, were made in the opposite orientation from that used in pTZ_E20.) The core unwinding element (CUE) in the pCL plasmid has been mapped previously by chemical labeling techniques to position 719 (Bode *et al.*, 1992). This corresponds to the minimum marked by an asterisk in the SIDD profiles of pTZ_E20 (Figure 4A) and of pCL (Figure 4B). The CUE sequence is AAATATATTT, which is closely related to that of the CUE in the IgH downstream S/MAR. The plasmid pCL^{mut} differs from pCL only in that this core was mutagenized to GAGTGTCTGT, which resists unwinding *in vitro* (Bode *et al.*, 1992).

The destabilization profiles of these two plasmids are shown in Figures 4B (pCL) and 4C (pCL^{mut}). Both profiles were calculated at a linking difference of -20 turns, which corresponds to the same superhelix density ($\sigma = -0.055$) as was used in the pTZ_E20 calculation. The sequence corresponding to the above CUE occurs at positions 719 to 727 in both molecules. This site is stabilized by less than 0.45 kcal/mol in the wild-type pCL sequence, but by more than 8 kcal/mol in the mutated form. These predictions agree precisely with the strong stabilization of the CUE by this mutation observed in chemical reactivity experiments (Bode *et al*, 1992). In Figure 4D we have superimposed the profiles of the (identically oriented) critical regions in these three vectors, with their CUE sequences at positions 119 to 127. Although the profiles for pTZ_E20 and pCL coincide in this range, adjacent regions have somewhat different profiles due their modulation by more remote sequences in these two molecules.

Although the mutation of the CUE sequence stabilizes its immediate region, the 823 bp inserts in both pCL and pCL^{mut} still exhibit strong destabilization over their lengths. Their *in vitro* affinities for the scaffold and their transciptional effects in marker gene constructs are virtually indistinguishable, suggesting that both molecules retain some S/MAR-function. This situation changes if the insert is shortened until the CUE marked in Figure 4B remains the only prominent minimum in

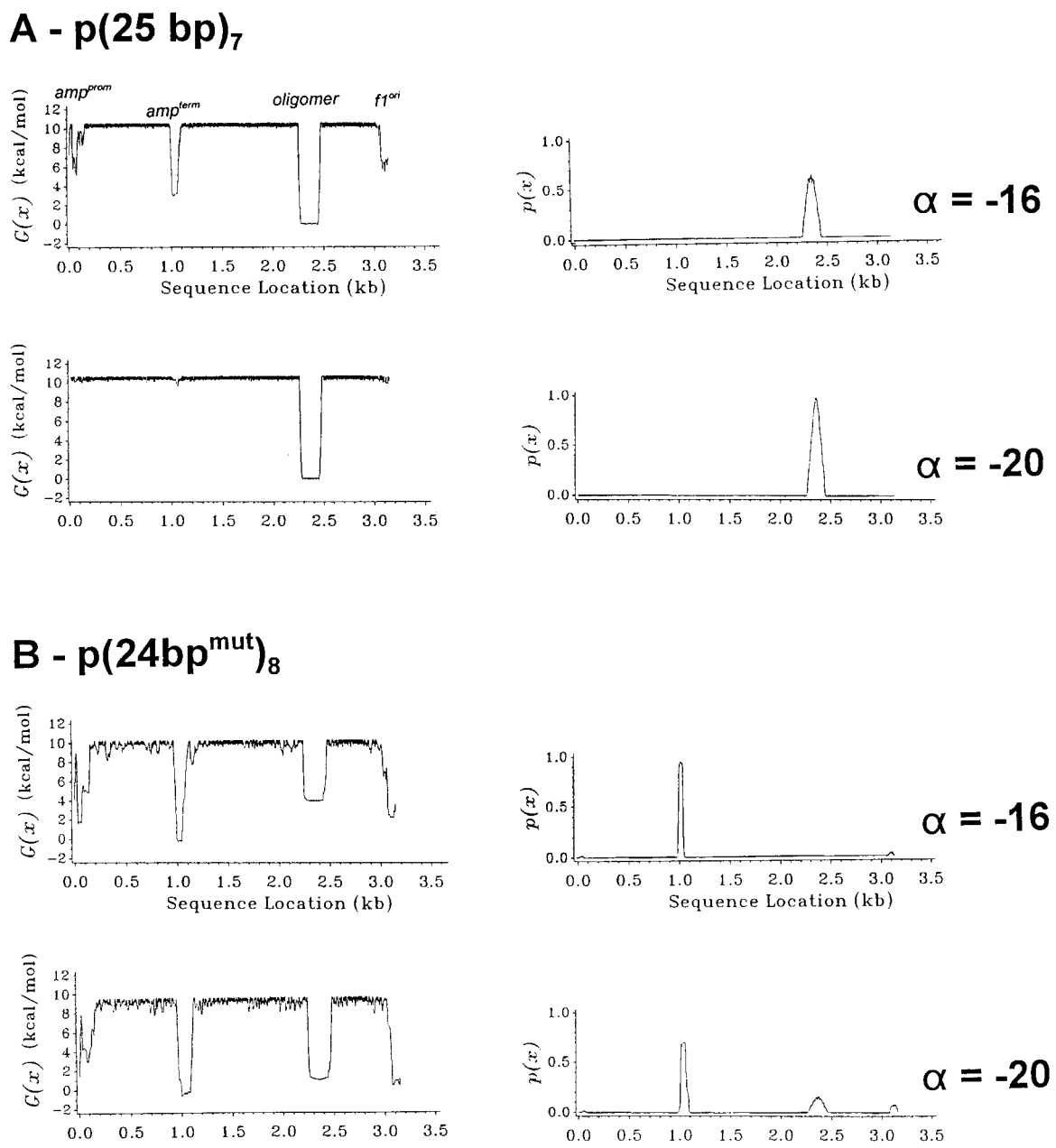

Figure 3. Destabilization properties of an oligomerized core-unwinding element and its mutant simulating an experiment reported in Figure 2 of (Bode *et al.*, 1992). A, A 25 bp sequence around the core unwinding element (AATATATTT) of the IgH enhancer-associated 3′ S/MAR element was oligomerized, and the heptamer $(25\ bp)_7$ was cloned into the *Eco*RV-site of a pBluescript vector. The destabilization energies *G(x)* are plotted at linking differences of $\alpha = -16$ turns and $\alpha = -20$ turns, which correspond to superhelix densities of $\sigma = -0.052$ and $\sigma = -0.065$, respectively. The corresponding opening probabilities *p(x)* are shown in the right hand graphs as functions of sequence location. B, A 24 bp sequence around the mutagenized core unwinding element ACTGC-TTT was oligomerized, and the octamer $(24\ bp^{mut})_8$ was cloned as for A. A small amount of strand opening at the mutated oligomer is predicted to occur only at the more extreme superhelicity.

the SIDD profile (see the 450 bp fragments V_1 and V_1'' in Mielke *et al.*, 1990). In this case the CUE mutation clearly decreases the affinity of this region for the scaffold, as determined in a reassociation assay. These results suggest that destabilization at the precise position of the original CUE is not essential for S/MAR activity, provided a sufficiently long region containing nearby destabilized sites is present. Still shorter fragment sizes exhibit a parallel decline of S/MAR properties both *in vitro* and *in vivo*. However, S/MAR activity returns in the gradual manner analogous to that shown in Figure 2 when such a 150 bp ''sub-S/MAR element'' is oligomerized (Kay & Bode, 1995). Findings of this type demonstrate a modular design of authentic S/MARs, and suggest that the under-

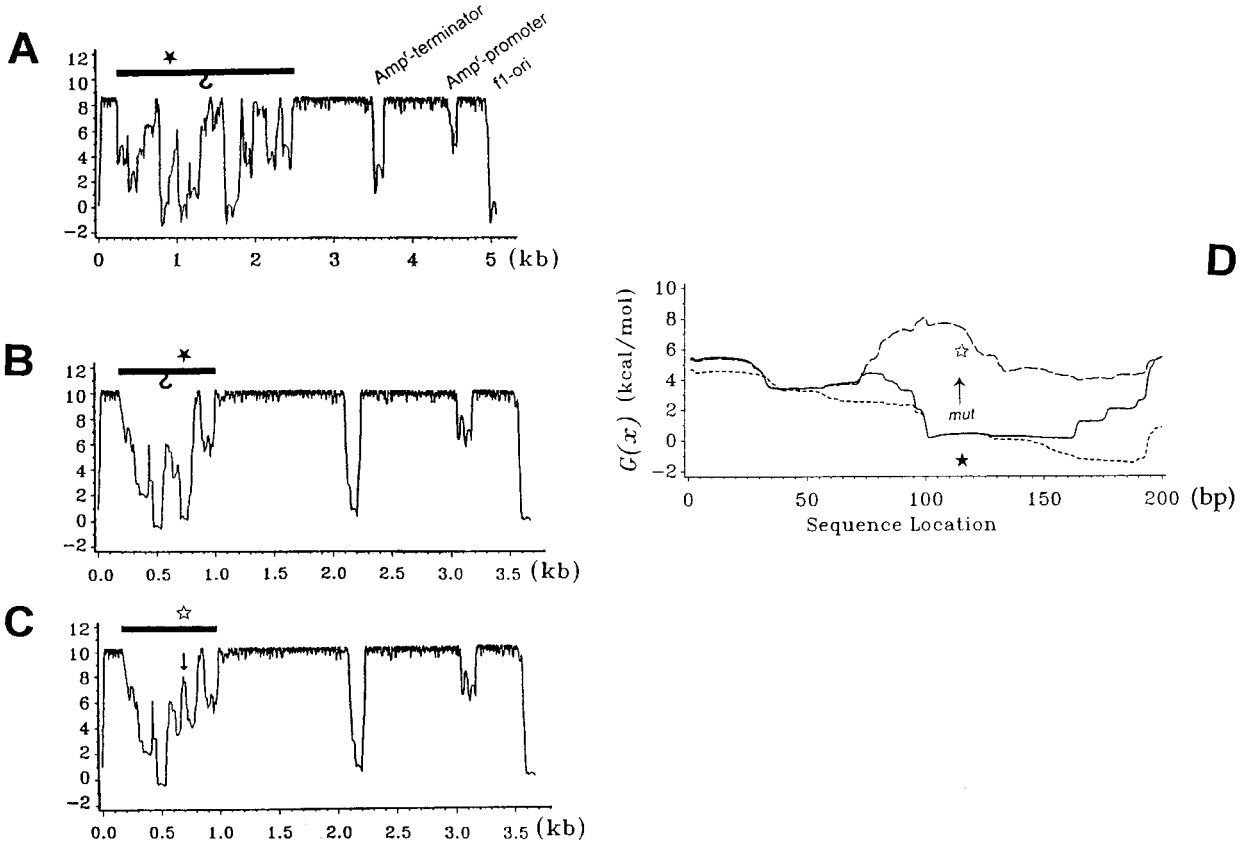

Figure 4. Destabilization profiles for a prototype S/MAR and its mutant. A, The 2.2 kb S/MAR element (called "I" in Mielke *et al.*, 1990), derived from the huIFN-β upstream S/MAR, is cloned into a pTZ18R vector and analyzed at superhelix density −0.055. The S/MAR insert is marked by a horizontal bar on top of the profile. The core unwinding element that has been determined by chemical modification analysis is indicated by a filled asterisk. B, Simulation of an experiment of (Bode *et al.*, 1992, Figure 1). The most destabilized 823 bp subfragment (fragment IV in Mielke *et al.*, 1990) of the 2.2 kb S/MAR was cloned into the pTZ18R vector and analyzed at the same superhelix density as in A. The position of the CUE is denoted by a filled asterisk. C, Similar construction as was used in B, but here the core unwinding element has been mutagenized to GACTGTCTGT (open asterisk). D, The core unwinding element transition profiles are shown at high resolution for the three cases above. The appropriate regions from the profiles in A, B and C are superimposed in identical orientations, so the CUE lies at positions 119 to 127. The effect of mutagenizing the 823 bp insert is indicated by an arrow.

lying sequence elements interact with the abundant scaffold proteins by a multisite-attachment mechanism that requires destabilization to occur over a region of sufficient length.

To this point we have only addressed the unwinding properties of genomic regions with established S/MAR activity. To assess the significance of the destabilization profiles of S/MARs found in the above analyses, we compared their calculated behaviors to that of the coding region of the IFN-β gene located downstream from the S/MAR analyzed in Figure 4A. The transition profile for this region is shown in Figure 5. It fully conforms to a three component pattern which has previously been noted to occur around yeast genes (Benham, 1996a): (1) the strongest destabilization occurs in the 3′ terminal flank of the gene, and coincides with the oligo(A) attachment site; (2) the sequence encoding the primary transcript contains no destabilized regions; (3) moderate destabilization occurs in the promoter region. The lack of potential base-unpairing regions (BURs) in the ORF would be required if the presence of stress-induced alternate secondary structures were to impede the progress of the polymerase. Both reassociation assays and halo mapping procedures show that the 1.9 kb *Eco*RI-F fragment (designated in Figure 5 by the horizontal bar on top of the profile) does not attach to the nuclear matrix (Bode & Maass, 1988). A comparison of the profiles in Figures 4A and 5 suggests that this could be either because: (1) the destabilized regions are too short to function as a S/MAR; (2) the spacing between neighbor destabilized sites is too large for efficient cooperation in multisite binding, and/or; (3) the sequence between these destabilized regions is inappropriate for transmitting single-strand character to adjacent regions, possibly due to its relatively high G+C content.

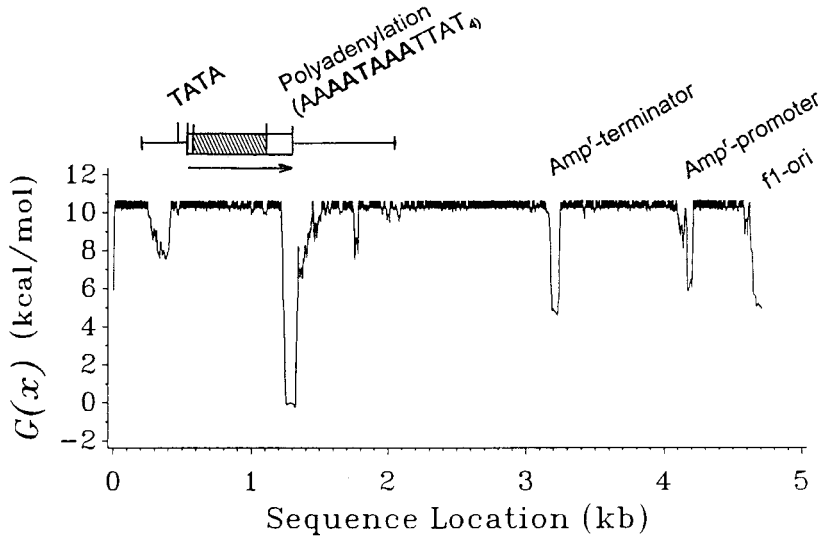

Figure 5. Structural signals in a transcription cassette containing the IFNB gene. The 1.8 kb *Eco*RI-F fragment (Gross *et al.*, 1981) was placed within the pTZ18R plasmid and analyzed at linking difference of −19 turns, which corresponds to superhelix density −0.04.

Destabilization at centromeres

The centromeric regions of chromosomes have been shown to exhibit S/MAR activity (Amati & Gasser, 1988). This may reflect their binding to kinetochore proteins, which are an integral part of the protein scaffold of the chromosome (Earnshaw *et al.*, 1984). S/MARs are present in each 1.9 kb repeat unit in the centromeric region of human chromosome 11, and in each 1.7 kb repeat unit in that of chromosome 16 (Strissel *et al.*, 1996). This unusual density of S/MARs suggests that the DNA within the centromeric regions of several, and possibly all, human metaphase chromosomes may be held in smaller loops than those present in chromosome arms.

The centromeres in the chromosomes of *Saccharomyces cerevisiae* consist of a short site with a characteristic three part structure (Fitzgerald-Hayes *et al.*, 1982). Elements CDEI and CDEIII have specific consensus sequences and fixed lengths. They bracket the element CDEII, which has variable length and no sequence motif other than extreme A+T-richness. The centromere CEN4 in chromosome IV has 8 bp in CDEI, 78 bp in CDEII and 25 bp in CDEIII. This centromere has been shown to experience strand separation of 80 (±10) base pairs under superhelical stresses (Tal *et al.*, 1994).

A 2 kb segment centered on the CEN4 centromeric region has been analyzed theoretically to determine the sites of predicted stress-induced duplex destabilization. The calculated transition profile is shown in Figure 6A. The only site of strong destabilization coincides with CEN4, which is denoted by a bar.

Schizosaccharomyces pombe is a species of yeast which displays many characteristics associated with higher eukaryotes. In particular, its centromeres have much more complex structures, consisting of multiple copies of long repetitive sequences. Figure 6B shows the destabilization profile of the 3964 bp dgIIa repetitive sequence from the centromere of chromosome II of *Sc. pombe*. Sequence comparison with the centromeric region from chromosome IV of *S. cerevisiae* shows a short region of very high homology that coincides with CEN4 in the latter molecule. This region of homology, which is denoted by a bar in each profile in Figure 6, is the most destabilized site in both cases. These centromeres both exhibit a characteristic pattern of destabilization at a conserved site, although they come from distinct species which vary greatly in the complexity of their centromeric organization.

Discussion

Base-unpairing *in vivo* and *in vitro*

Previous work has shown that base-unpairing regions (BURs) are closely associated with several types of DNA functional elements. BURs characterized by their reaction properties with single strand-specific enzymes or reagents have been reported to occur *in vivo* within the transcriptional regulatory regions of several genes, including the chicken β^A-globin gene (Kohwi-Shigematsu *et al.*, 1983), the human CMV major immediate early gene (Kohwi-Shigematsu & Nelson, 1988), the interferon-β gene (Bode *et al.*, 1986), and the c-myc oncogene (Michelotti *et al.*, 1996a,b). A region in the terminal 3′ flank of the yeast FBP1 gene that was previously predicted to be destabilized by stresses has also been observed to be unpaired *in vivo* (Aranda *et al.*, 1997). In replication, a BUR located at a specific position within the *Escherichia coli* origin *oriC* has been shown to be required for *in vivo* activity (Kowalski & Eddy, 1989).

Sites within DNA that are accessible to single-strand specific probes *in vivo* often also are reactive when inserted into supercoiled plasmids (Evans *et al.*, 1984; Schon *et al.*, 1983). This allows the *in vivo* potential for stress-induced duplex destabilization to be assessed by more convenient *in vitro*

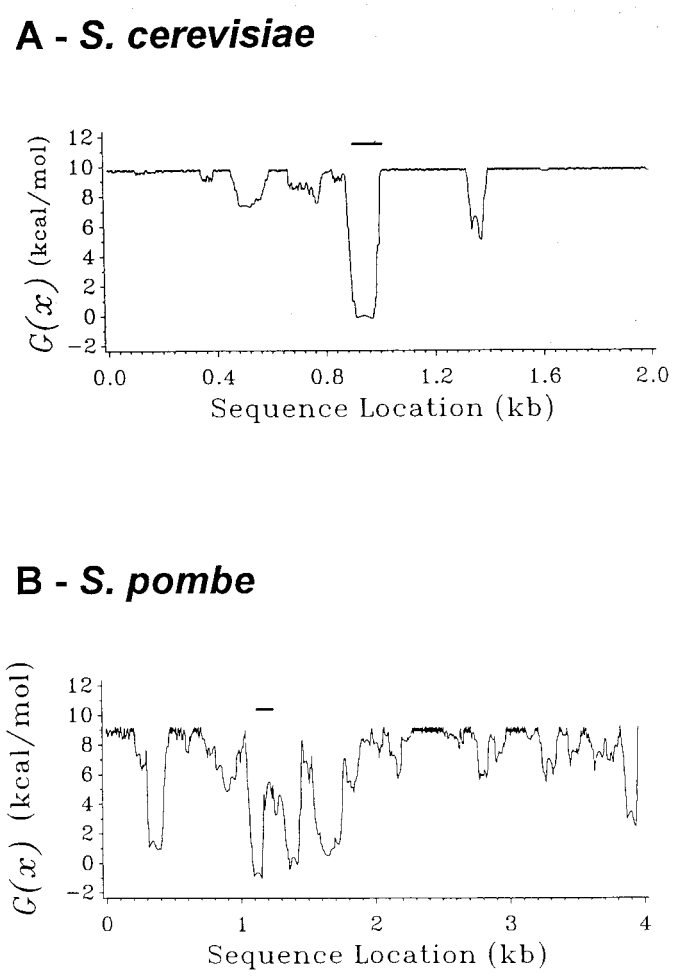

Figure 6. SIDD-profiles for centromeric regions in *S. cerevisiae* and *Sc. pombe*. A, The SIDD profile is shown of the 2kb fragment containing the centromeric region of *S. cerevisiae* chromosome IV. The site of strong destabilization coincides with the centromere sequence. B, The SIDD profile is shown of the 3.9 kb fragment comprising the dgIIa repetitive sequence of *Sc. pombe* chromosome II. The region denoted by a bar in both graphs is the only site of high sequence homology in both centromeres. This homologous region coincides with the most destabilized site in both cases.

assays. In this way BURs have been found at the S/MAR elements surrounding the IgH enhancer. These become reactive when subjected to the torsional stress of negative superhelicity. This is not due to a high rate of DNA breathing, but rather to stable base-unpairing (Kohwi-Shigematsu & Kohwi, 1990).

Predictions of base-unpairing regions at S/MARs

In this paper we have applied the computational technique of Benham (1992) to predict the destabilization characteristics of DNA sequences containing sites with S/MAR activity, including transcriptional regulatory regions and centromeres. This work documents a close association between S/MAR activity and stress-induced destabilization. The S/MARs and the yeast centromeres studied here share a propensity to denature when stressed. These results accord with the single-strand character that has been seen at these S/MAR elements *in vivo* (Bode *et al.*, 1995 and references therein), and with the observation that single-stranded DNA competes with S/MAR–scaffold interactions (Kay & Bode, 1994, Ludérus *et al.*, 1994).

Our studies center around the core-unwinding element (CUE) AATATATTT, which is present in both the IgH enhancer-associated 3′ S/MAR (Figure 1) and the interferon upstream S/MAR element (Figure 4). This is the site where unwinding initiates, as assayed by reactivity with the single-strand specific reagent chloroacetaldehyde. In the IgH S/MAR this element induces progressive unpairing in the direction towards the enhancer to yield a BUR which can include up to 200 base-pairs, depending on ionic composition and superhelix density (Figure 1B and Kohwi-Shigematsu and Kohwi, 1990). If the 25 bp sequence TCTTTAATTTCT**AATATATTT**AGAA surrounding this CUE is oligomerized, an element is created which functions as a S/MAR by all *in vitro* and *in vivo* criteria that have been developed to date (Bode *et al.*, 1992). Figure 3A shows that a 7-mer of this element is strongly destabilized over its entire length. This construct is subject to more extensive unwinding than even the strongest naturally occurring unwinding elements known.

These results suggest that the DNA attachment events mediated by these sites may involve unpaired structures. Holding each strand of the DNA by a separate single-strand binding event would provide a more stringent way to constrain

the topology of the DNA than would duplex binding, as release of this constraint would require two simultaneous detachment events. DNA unwinding might be facilitated *in vivo* by an enrichment of the nuclear scaffold with proteins that are either single-strand specific or have DNA unwinding activities. S/MAR-binding proteins of this type include members of the hnRNP class (Mattern *et al*., 1996) which function in hnRNA packaging and as architectural transcription factors (Michelotti *et al*., 1996a).

Relation of S/MAR unwinding potential to chromosomal functions

Strong associations have been found between SIDD and several other types of DNA regulatory regions. Stress-destabilized sites have been shown to occur within the 3′ flanks of yeast genes (Aranda *et al*., 1997; Benham, 1996a;). This site may act as a sink for absorbing the positive supercoils generated in front of a progressing polymerase transcription complex, increasing the efficiency of transcript elongation by removing a geometric impediment to progress (Benham, 1996a). A similar function has been suggested for intergenic S/MARs (Bode *et al*., 1992, 1996).

Localized supercoiling in eukaryotes is not immediately relaxed by topoisomerases (Dunaway & Ostrander, 1993; Krebs & Dunaway, 1996; Wang & Dröge 1996). Because regions having S/MAR activity have a strong propensity for stress-induced base-unpairing, it is appropriate to consider whether destabilized 3′-terminal gene regions could function as S/MARs. Examples are known that suggest this may occur. A restriction fragment that overlaps the polyadenylation site of the human interleukin-2 gene has been shown to have S/MAR activity *in vitro* (Straube & Bode, unpublished). Another S/MAR, postulated to be involved in regulating polyadenylation, has been localized to a 172 bp fragment containing the first of two polyadenylation signals of the *c-myc* protooncogene (Chou *et al*., 1990; 1991). A 400 bp region containing this site has been predicted to be strongly destabilized by stress (Benham, 1996b). In contrast, the polyadenylation site of the IFNB gene is embedded in a relatively short destabilized region (Figure 5). This gene is associated with two S/MARs, each 5 to 7 kb long, which start at positions −1.7 and +12.3 kb relative to it. (Bode & Maass, 1988). These extensive unwinding regions may be needed to protect this most telomeric member of an extended gene array from position effect variation.

Is base-unpairing the only structural transition associated with S/MARs?

The absence of a clear sequence motif associated with S/MAR activity suggests that the essential attributes of these regions may be structural. The present work documents a clear association between S/MAR activity and stress-induced duplex destabilization. Under different *in vitro* conditions superhelical stresses also can induce the formation of cruciforms, Z-DNA, or triple helical structures in susceptible sequences (Kowalski *et al*., 1988; Sheflin & Kowalski, 1985). This raises the question of whether other types of stress-induced conformational transitions might also be involved in S/MAR function.

Indirect evidence suggests that sequences which can form either cruciforms or triple helices in principle could serve biological functions in eukaryotes (Furlong *et al*., 1989; Palacek, 1991). Schroth & Ho (1995) have demonstrated that strong cruciform-forming sequences (inverted repeats, IR) occur at relatively high frequencies in yeast (1/19,700 bp) and in humans (1/41,800 bp), whereas triple-helix promoting sequences are abundant only in humans (1/49,400 bp). While eukaryotic IRs are very A+T-rich, prokaryotic IRs commonly have a relatively high G+C-content and occur almost exclusively at transcription termination sites. Since triplex-forming sequences are not abundant in the *E. coli* genome, these results suggest the possibility that triplexes and cruciforms may have specific physiological roles in eukaryotes that they do not have in prokaryotes. This suggestion is reinforced by the observation that genomic sequences susceptible to these alternate structures are concentrated at S/MARs, at centromeres and at ORI sequences (Boulikas, 1995; Boulikas & Kong, 1993; Mielke *et al*., 1996).

However, the presence of susceptible sequences alone is not sufficient reason to infer a biological role for the alternative structures which they can form. Inverted repeats must be present in DNA to encode RNA molecules that contain hairpins, which are biologically important structures. Again, any bi-directional process requiring protein recognition of specific DNA sequences must arrange identical binding sites in an inverted repeat pattern. The SV40 replication origin occurs within the loop region of an inverted repeat sequence for this reason. The cruciform which this repeat in principle can form is not thought to play a role in the initiation of replication.

Future directions

We have documented a particularly strong association of BURs with scaffold/matrix-attached regions (S/MARs), and with centromeres. We show that stress-induced duplex destabilization properties have a high predictive potential for the S/MAR elements that are thought to subdivide the genome into functional units. These results strongly suggest that DNA regulatory regions do not serve only as passive repositories of the nucleotide sequence information required for protein binding specificity. Instead, structural polymorphisms may be involved in the mechanisms by which their activities are regulated. Susceptibility to struc-

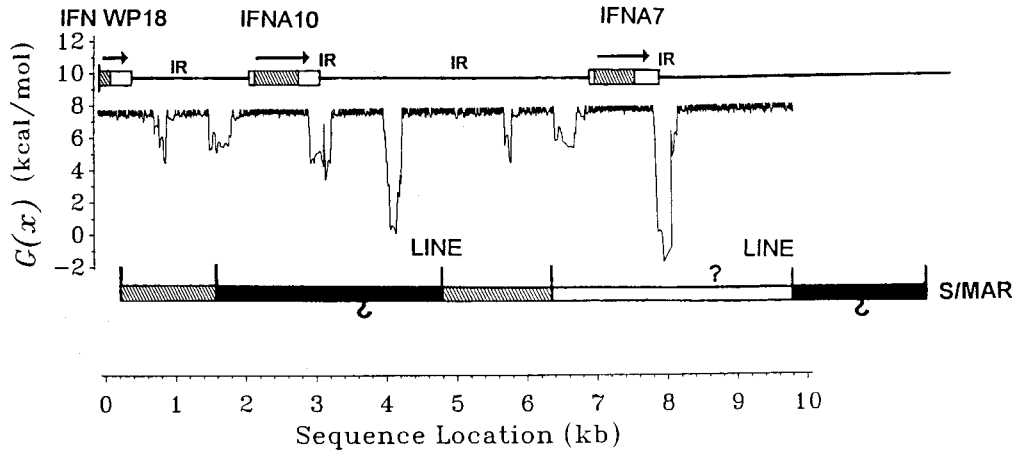

Figure 7. Structural signals in a gene cluster. The 9.9 kb genomic sequence (Ullrich, 1982) covering a pseudogene (IFNWP18) and two IFN-genes (IFNA10 and IFNA7) on the short arm of chromosome 9 are analyzed at linking difference −48 turns, corresponding to superhelix density −0.05. Results of preliminary S/MAR analyses are indicated by the bars below the profile (filled, >90% bound; hatched, >70% bound; hatched, >50% bound; open, no evidence due to the fragment's large size). IR, location of the most prominent inverted repeat sequences.

tural transitions of the proper type then becomes a necessary condition for activity, which, once documented, can be used to search genomic sequences for sites with regulatory potential.

In future work, we will extend the present analysis of the association between S/MARs and stress-destabilized sites. It is our intention to use the strong destabilization found here to occur at S/MARs to predict sites in other sequences that have S/MAR potential, and to test these predictions by subsequent experimentation. The first example of such a prediction is given in Figure 7, which shows the destabilization profile of a sequenced 9 kb portion of the IFN gene cluster (Ullrich *et al.*, 1982). The tripartite pattern documented in Figure 5 for the IFNB gene is seen also to occur in the IFNA10 and IFNA7 regions. However, here the destabilized sites in the 3′-non-coding flanks are more extensive, covering 300 to 400 bp rather than the 100 bp seen with IFNB. This suggests the possibility that this region may have S/MAR activity. Recent experiments in the lab of one of us (J. B.) show that the *Hin*dIII fragment covering this linker sequence indeed has strong S/MAR character (see Figure 7 and Holger Rühl, G. B. F., unpublished results). The linker between the pseudogene IFNWP18 and IFNA10 is unusual in that it lacks a strongly destabilized site at its immediate 3′ end. This may be due a lack of selective pressure to enforce such a feature in pseudogenes. Instead, some destabilization occurs before the IFNA10 control region which could explain the weak S/MAR character found for a 1.4 kb *Hin*dIII fragment containing this site.

Additional experiments will be performed, both *in vitro* and *in vivo*, to test various hypotheses regarding the role of destabilization in the activity of S/MARs and in their various influences on other regulatory events. To this end, a method has been proposed for altering the destabilization characteristics of a region without changing its local base sequence (Benham, 1996a). This can be accomplished by inserting (or removing) sequences at remote positions which are susceptible to transitions that compete with destabilization at the site of interest. This and other strategies will be used in our future research on this topic.

Acknowledgments

This work was supported in part by grants to C. J. B. from the National Institutes of Health and from the National Science Foundation, and to J. B. O. from the Deutsche Forschungsgemeinschaft Bo419/6-1.

References

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.

Amati, B. & Gasser, S. M. (1988). Chromosomal ARS and CEN elements bind specifically to the yeast nuclear scaffold. *Cell,* **54**, 967–978.

Amati, B. & Gasser, S. M. (1990). Drosophila scaffold-attached regions bind nuclear scaffolds and can function as ARS elements in both budding and fission yeasts. *Mol. Cell. Biol.* **10**, 5442–5454.

Amati, B., Pick, L., Laroche, T. & Gasser, S. M. (1990). Nuclear scaffold attachment stimulates, but is not essential for ARS activity in *Saccharomyces cerevisiae*: analysis of the Drosophila ftz SAR. *EMBO J.* **9**, 4007–4016.

Aranda, A., Perez-Ortin, J., Benham, C. J. & del Olmo, M. (1997). Analysis of the *in vivo* structure of a

naturally occurring d(AT)n sequence in yeast chromatin. *Yeast,* **13**, 313–326.

Barone, J. G., De Lara, J., Cummings, K. B. & Ward, W. S. (1994). DNA organization in human spermatozoa. *J. Androl.* **15**, 139–144.

Benham, C. J. (1990). Theoretical analysis of heteropolymeric transitions in superhelical DNA molecules of specified sequence. *J. Chem. Phys.* **92**, 6294–6305.

Benham, C. J. (1992). Energetics of the strand separation transition in superhelical DNA. *J. Mol. Biol.* **225**, 835–847.

Benham, C. J. (1993). Sites of predicted stress-induced DNA duplex destabilization occur preferentially at regulatory loci. *Proc. Natl Acad. Sci. USA,* **90**, 2999–3003.

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

Benham, C. J. (1996b). Computation of DNA structural variability: a new predictor of DNA regulatory regions. *Comp. Appl. Biol. Sci.* **12**, 375–381.

Bode, J. & Maass, K. (1988). Chromatin domain surrounding the human interferon-β gene as defined by scaffold-attached regions. *Biochemistry,* **27**, 4706–4711.

Bode, J., Pucher, H. J. & Maass, K. (1986). Chromatin structure and induction-dependent conformational changes of human interferon-ß genes in a mouse host cell. *Eur. J. Biochem.* **158**, 393–401.

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.

Bode, J., Schlake, T., Ríos-Ramírez, M., Mielke, C., Stengert, M., Kay, V. & Klehr-Wirth, D. (1995). Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. In *Structural and Functional Organization of the Nuclear Matrix: International Review of Cytology* (Brezney, R. & Jeon, K. W., eds), vol. 162A, pp. 389–453, Academic Press, Orlando.

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 Express.* **6**, 115–138.

Bonifer, C., Yannoutsos, N., Krueger, G., Grosveld, F. & Sippel, A. E. (1994). Dissection of the locus control function located on the chicken lysozyme gene domain in transgenic mice. *Nucl. Acids Res.* **22**, 4202–4210.

Boulikas, T. (1993). Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. *J. Cell Biochem.* **52**, 14–22.

Boulikas, T. (1995). Chromatin domains and prediction of MAR sequences. In *Structural and Functional Organization of the Nuclear Matrix: International Review of Cytology* (Brezney, R. & Jeon, K. W., eds), vol. 162A, pp. 279–388, Academic Press, Orlando.

Boulikas, T. & Kong, C. F. (1993). Multitude of inverted repeats characterizes a class of anchorage sites of chromatin loops to the nuclear matrix. *J. Cell Biochem.* **53**, 1–12.

Carri, M. T., Micheli, G., Graziano, E., Pace, T. & Buongiorno-Nardelli, M. (1986). The relationship between chromosomal origins of replication and the nuclear matrix during cell cycle. *Expt. Cell Res.* **164**, 426–436.

Chou, R. H., Churchill, J. R., Flubacher, M. M., Mapstone, D. E. & Jones, J. (1990). Identification of a nuclear matrix-associated region of the c-myc protooncogene and its recognition by a nuclear protein in the human leukemia HL-60 cell line. *Cancer Res.* **50**, 3199–3206.

Chou, R. H., Churchill, J. R., Mapstone, D. E. & Flubacher, M. M. (1991). Sequence-specific binding of a c-myc nuclear-matrix-associated region shows increased nuclear matrix retention after leukemic cell (HL-60) differentiation. *Am. J. Anat.* **191**, 312–320.

Cockerill, P. N. & Garrard, W. T. (1986a). Chromosomal loop anchorage of immunoglobulin genes occurs adjacent to enhancers. *Hoppe Seyler's Z. Physiol. Chem.* **367**, 178a.

Cockerill, P. N. & Garrard, W. T. (1986b). Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell,* **44**, 273–282.

Cockerill, P. N., Yuen, M. H. & Garrard, W. T. (1987). The enhancer of the immunoglobulin heavy chain locus is flanked by presumptive chromosomal loop anchorage elements. *J. Biol. Chem.* **262**, 5394–5397.

Diaz, M. O. (1995). The human type I interferon gene cluster. *Seminars in Virology,* **6**, 143–149.

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.

Dijkwel, P. A. & Hamlin, J. L. (1995). Origins of replication and the nuclear matrix: the DHFR domain as a paradigm. In *Structural and Functional Organization of the Nuclear Matrix* (Berezney, R. & Jeon, K. W., eds), vol. 162A, pp. 455–484, Academic Press, San Diego.

Dunaway, M. & Ostrander, E. A. (1993). Local domains of supercoiling activate a eukaryotic promoter *in vivo*. *Nature,* **361**, 746–748.

Earnshaw, W. C., Halligan, N., Cooke, C. & Rothfield, N. (1984). The kinetochore is part of the metaphase chromosome scaffold. *J. Cell. Biol.* **98**, 352–357.

Evans, T., Schon, E., Gora-Maslak, G., Patterson, J. & Efstratiadis, A. (1984). S1-hypersensitive sites in eukaryotic promoter regions. *Nucl. Acids Res.* **12**, 8043–8058.

Fitzgerald-Hayes, M., Buhler, J. M. & Carbon, J. (1982). Nucleotide sequence comparison and functional analysis of yeast centromere DNAs. *Cell,* **29**, 235–244.

Forrester, W. C., Van Genderen, C., Jenuwein, T. & Grosschedl, R. (1994). Dependence of enhancer-mediated transcription of the immunoglobulin μ-gene on nuclear matrix attachment regions. *Science,* **265**, 1221–1225.

Furlong, J. C., Sullivan, K. M., Murchie, A. I. H., Gough, G. W. & Lilley, D. M. J. (1989). Localized chemical hyperreactivity in supercoiled DNA: evidence for base unpairing in sequences that induce low-salt cruciform extrusion. *Biochemistry,* **28**, 2009–2017.

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.

Gasser, S. M. & Laemmli, U. K. (1987). A glimpse at chromosomal order. *Trends Genet.* **3**, 16–22.

Gross, G., Mayr, U., Bruns, W., Grosveld, F., Dahl, H. H. & Collins, J. (1981). The structure of a thirty-six kilobase region of the human chromosome

including the fibroblast interferon gene IFN-beta. *Nucl. Acids Res.* **9**, 2495–2506.

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 trans-activator that describes a new DNA-binding protein family. *Genes Dev.* **9**, 3067–3082.

Hozak, P. & Cook, P. R. (1994). Replication factories. *Trends Cell Biol.* **4**, 48–52.

Jenuwein, T., Forrester, W. C., Qiu, R.-G. & Grosschedl, R. (1993). The immunoglobulin Mu enhancer core establishes local factor access in nuclear chromatin independent of transcriptional stimulation. *Genes Dev.* **7**, 2016–2032.

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.

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.

Kay, V. & Bode, J. (1994). Binding specificity of a nuclear scaffold: Supercoiled, single-stranded, and scaffold-attached-region DNA. *Biochemistry,* **33**, 367–374.

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, Wiley-Liss, Inc., New York.

Kirillov, A., Kistler, B., Mostoslavsky, R., Cedar, H., Wirth, T. & Bergman, Y. (1996). A role for nuclear NF-KB in B-cell specific demethylation of the Igkappa locus. *Nature Genet.* **13**, 435–441.

Kohwi-Shigematsu, T. & Kohwi, Y. (1990). Torsional stress stabilizes extended base-unpairing in DNA flanking the immunoglobulin heavy chain enhancer. *Biochemistry,* **29**, 9551–9560.

Kohwi-Shigematsu, T. & Nelson, J. A. (1988). The chemical carcinogen, chloroacetaldehyde, modifies a specific site within the regulatory sequence of human cytomegalovirus major immediate early gene *in vivo*. *Mol. Carcinogen,* **1**, 20–25.

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.

Kowalski, D. & Eddy, M. J. (1989). The DNA unwinding element: a novel, *cis*-acting component that facilitates opening of the *E. coli* replication origin. *EMBO J.* **8**, 4335– 4344.

Kowalski, D., Natale, D. A. & Eddy, M. J. (1988). Stable DNA unwinding, not ''breathing'' accounts for single-strand-specific nuclease hypersensitivity of specific A+T-rich sequences. *Proc. Natl Acad. Sci. USA,* **85**, 9464–9468.

Kramer, J. A. & Krawetz, S. A. (1995). Matrix-associated regions in haploid expressed domains. *Mammalia Genome,* **6**, 677–679.

Kramer, J. A., Singh, G. B. & Krawetz, S. A. (1996). Computer-assisted search for sites of nuclear matrix attachment. *Genomics,* **33**, 305–308.

Krebs, J. E. & Dunaway, M. (1996). DNA length is a critical parameter for eukaryotic transcription *in vivo*. *Mol. Cell. Biol.* **16**, 5821–5829.

Lichtenstein, M., Keini, G., Cedar, H. & Bergman, Y. (1994). B cell-specific demethylation: a novel role for the intronic kappa chain enhancer sequence. *Cell,* **76**, 913–923.

Ludérus, M. E. E., Den Blaauwen, J. L., De Smit, O. J. B., Compton, D. A. & Van Driel, R. (1994). Binding of matrix attachment regions to lamin polymers involves single-stranded regions and the minor groove. *Mol. Cell. Biol.* **14**, 6297–6305.

Mattern, K. A., Humbel, B. M., Muijsers, A. O., De Jong, 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.

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.

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.

Michelotti, E. F., Michelotti, G. A., Aronsohn, A. I. & Levens, D. (1996a). Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol. Cell. Biol.* **16**, 2350–2360.

Michelotti, G. A., Michelotti, E. F., Pullner, A., Duncan, R. C., Eick, D. & Levens, D. (1996b). Multiple single-stranded *cis* elements are associated with activated chromatin of the human c-myc gene *in vivo*. *Mol. Cell. Biol.* **16**, 2656–2669.

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.

Mielke, C., Maass, K., Tümmler, M. & Bode, J. (1996). Anatomy of highly expressing chromosomal sites targeted by retroviral vectors. *Biochemistry,* **35**, 2239–2252.

Mirkovitch, J., Gasser, S. M. & Laemmli, U. K. (1987). Relation of chromosome structure and gene expression. *Philos. Trans. Roy. Soc. Lond. ser. B,* **317**, 563–574.

Nikolaev, L. G., Tsevegiyn, T., Akopov, S. B., Ashworth, L. K. & Sverdlov, E. D. (1996). Construction of a chromosome specific library of human MARs and mapping of matrix attachment regions on human chromosome 19. *Nucl. Acids Res.* **24**, 1330–1336.

Palacek, E. (1991). Local supercoil-stabilized DNA structures. *Crit. Rev. Biochem. Mol. Biol.* **26**, 151–183.

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.

Schon, E., Evans, T., Welsh, J. & Efstratiadis, A. (1983). Conformation of promoter DNA: fine mapping of S1-hypersensitive sites. *Cell,* **35**, 837–848.

Schroth, G. P. & Ho, P. S. (1995). Occurence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucl. Acids Res.* **23**, 1977–1983.

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.

Sheflin, L. G. & Kowalski, D. (1985). Altered DNA conformations detected by mung bean nuclease occur in promoter and terminator regions of

supercoiled pBR322 DNA. *Nucl. Acids Res.* **13**, 6137–6154.

Strissel, P. L., Espinosa, R., III, Rowley, J. D. & Swift, H. (1996). Scaffold attachment regions in centromere-associated DNA. *Chromosoma,* **105**, 122–133.

Tal, M., Shimron, F. & Yagil, G. (1994). Unwound regions in yeast centromere IV DNA. *J. Mol. Biol.* **243**, 179–189.

Tsutsui, K., Tsutsui, K. & Muller, M. T. (1988). The nuclear scaffold exhibits DNA-binding sites selective for supercoiled DNA. *J. Biol. Chem.* **263**, 7235–7241.

Ullrich, A., Gray, A., Goeddel, D. V. & Tull, T. J. (1982). Nucleotide sequence of a portion of human chromosome 9 containing a leukocyte interferon gene cluster. *J. Mol. Biol.* **156**, 467–486.

Vogelstein, B., Pardoll, D. M. & Coffey, D. S. (1980). Supercoiled loops and eukaryotic DNA replication. *Cell,* **22**, 79–85.

Wang, Z. Y. & Dröge, P. (1996). Differential control of transcription-induced and overall DNA supercoiling by eukaryotic topoisomerases *in vitro*. *EMBO J.* **15**, 581–589.

Zong, R. T. & Scheuermann, R. H. (1995). Mutually exclusive interaction of a novel matrix attachment region binding protein and the NF-μNR enhancer repressor: implications for regulation of immunoglobulin heavy chain expression. *J. Biol. Chem.* **270**, 24010–24018.

Edited by I. Tinoco

(Received 13 March 1997; received in revised form 3 September 1997; accepted 3 September 1997)