Computational and in Vitro Analysis of Destabilized DNA Regions in the Interferon Gene Cluster: Potential of Predicting Functional Gene Domains †

ARTICLE *in* BIOCHEMISTRY · FEBRUARY 2003

Impact Factor: 3.02 · DOI: 10.1021/bi026496+ · Source: PubMed

4 AUTHORS, INCLUDING:

Sandra Goetze
ETH Zurich
53 PUBLICATIONS **920** CITATIONS
SEE PROFILE

Juergen Bode
Hannover Medical School
199 PUBLICATIONS **5,069** CITATIONS
SEE PROFILE

Computational and in Vitro Analysis of Destabilized DNA Regions in the Interferon Gene Cluster: Potential of Predicting Functional Gene Domains†

S. Goetze,‡ A. Gluch,‡ C. Benham,§ and J. Bode*,‡

German Research Center for Biotechnology/Epigenetic Regulation, Mascheroder Weg 1, D-38124 Braunschweig, Germany, and University of California Davis Genome Center, Davis, California 95616-8536

Received July 23, 2002; Revised Manuscript Received September 10, 2002

ABSTRACT: Recent evidence adds support to a traditional concept according to which the eukaryotic nucleus is organized into functional domains by scaffold or matrix attachment regions (S/MARs). These regions have previously been predicted to have a high potential for stress-induced duplex destabilization (SIDD). Here we report the parallel results of binding (reassociation) and computational SIDD analyses for regions within the human interferon gene cluster on the short arm of chromosome 9 (9p22). To verify and further refine the biomathematical methods, we focus on a 10 kb region in the cluster with the pseudogene IFNWP18 and the interferon α genes IFNA10 and IFNA7. In a series of S/MAR binding assays, we investigate the promoter and termination regions and additional attachment sequences that were detected in the SIDD profile. The promoters of the IFNA10 and the IFNA7 genes have a moderate ∼20% binding affinity to the nuclear matrix; the termination sequences show stronger association (70−80%) under our standardized conditions. No comparable destabilized elements were detected flanking the IFNWP18 pseudogene, suggesting that selective pressure acts on the physicochemical properties detected here. In extended, noncoding regions a striking periodicity is found of rather restricted SIDD minima with scaffold binding potential. By various criteria, the underlying sequences represent a new class of S/MARs, thought to be involved in a higher level organization of the genome. Together, these data emphasize the relevance of SIDD calculations as a valid approach for the localization of structural, regulatory, and coding regions in the eukaryotic genome.

While there is increasing awareness that the eukaryotic nucleus is a highly structured organelle, its functional architecture has remained a largely unresolved enigma of molecular biology. According to recent publications the nucleus is organized into three major compartments: an open euchromatic compartment containing active genes, a heterochromatic compartment containing inactive genes, and an interchromatin compartment mostly consisting of proteins (*1*), which is otherwise referred to as the in vivo nuclear matrix.

Following its discovery in 1974 (*2*), the nuclear matrix has been shown to accommodate the replication and transcription machineries and, accordingly, the genes that are being actively transcribed. The DNA sequences thought to be responsible for mediating such effects, by serving as an anchor to the nuclear matrix, are the scaffold/matrix attachment regions (S/MARs),[1] which are recognized according to topological features that become reinforced by topological stress as it arises during replication and transcription (*3*, *4*). According to a popular model a group of extended, tightly matrix-attached constitutive S/MARs serves as a coordinate system, which enables the formation of independently regulated chromatin loops ranging in size between 5 and 200 kb. A tendency for active genes to be organized into small loops has been noted (*5*).

S/MARs were discovered almost two decades ago and have been defined as the DNA elements that either stay at the nuclear skeleton after the extraction of the histones and soluble factors during a halo-mapping procedure (*6*) or that reassociate with a scaffold or matrix preparation with high affinity in vitro (*7*−*9*). For obvious reasons, only the latter property (reassociation strength rather than the actual status in vivo) lends itself to computerization. While S/MARs do not conform to any obvious consensus sequence, their most consistent feature appears to be the propensity to expose single strands under negative superhelical tension (*10*) in addition to their intrinsic potential to form secondary structures for which strand separation is a prerequisite (*4*). Following this kind of reasoning, the prediction of S/MARs has required entirely new biomathematical concepts. The development of dedicated algorithms is considered important since S/MARs are commonly found at the boundaries of transcription units, typically in association with DNase I hypersensitive sites (*11*), where they may function as genomic insulators (*12*), in the vicinity of enhancers (*13*), or origins of replication (*14*). Thereby this class of elements emerges as a valuable marker enabling the localization of independently regulated genomic units, the so-called chro-

† This work was supported by grants from Deutsche Forschungsgemeinschaft (Bo 419/6-1/-2) and BMBF (01 KW 0003).

* Corresponding author: e-mail jbo@gbf.de; telephone +49 531 6181 251; fax +49 531 6181 262.

‡ German Research Center for Biotechnology/Epigenetic Regulation.

§ UC Davis Genome Center; e-mail cjbenham@ucdavis.edu.

[1] Abbreviations: BUR, base-unpairing region; CUE, core-unpairing element; HS, DNase I hypersensitive site; SIDD, stress-induced duplex destabilization; S/MAR, scaffold/matrix attachment region; UE, unpairing element.

10.1021/bi026496+ CCC: $25.00
Published on Web 12/13/2002

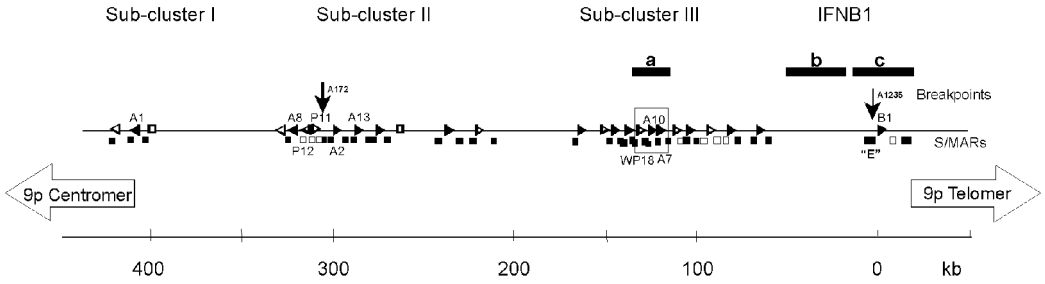

FIGURE 1: Human type 1 interferon gene cluster on the short arm of chromosome 9: a, b, and c indicate the regions that have been subjected to a detailed computational and in vitro analysis. The arrows marked A172 and A1235 are fragile sites adjacent to S/MAR elements (*11*).

matin domains, and as an indicator for sites with a regulatory potential.

Among the biomathematical approaches, the stress-induced duplex destabilization (SIDD) analysis has been proven to be of exceptional value, probably because it goes beyond the features that are directly linked to sequence motifs (*10*; manuscript in preparation). SIDD profiles are calculated by a statistical mechanical procedure in which a superhelical deformation is partitioned between strand separation, twisting within denatured regions, and residual superhelicity, resulting in the identification of unpairing elements (UEs), among these the core unpairing element (CUE), which represents the nucleation point of strand separation somewhere within a base-unpairing region (BUR; *10*, *15*). In all cases examined to date, S/MARs coincide with SIDD regions of eukaryotic DNA, which are predicted to be extensively destabilized. These BURs may consist either of one dominant minimum exceeding a threshold extension or of multiple moderately destabilized UEs, which are spaced according to distinct criteria in order to permit a multisite attachment of the responsible protein complexes (*16*, *17*). If spacing exceeds an upper limit of about 500 bp, individual UEs loose their capacity to interact with each other, concomitant with a loss of binding activity (*18*).

The structure of the rather extended (14 kb) domain surrounding the interferon-β (IFNB1) gene at the telomeric end of the gene cluster (*11*) yielded access to 5 and 7 kb long, highly active S/MAR elements that have found increasing use for the rational design of vectors with novel properties (*14*, *19*, *20*). Overall, the type I interferon gene cluster (depicted in Figure 1) comprises 26 genes spread over 400 kb on the short arm of chromosome 9 at band 9p22 (*21*). While IFNB1 forms a unique domain with many sites of regulatory potential, which together may constitute the LCR (locus-control region), the other interferon genes occur in three subclusters (I−III); within subcluster II, there is an inversion point where transcription direction switches from centromeric to telomeric. Despite this apparent space limitation within the subclusters, the functional members appear to be separated by efficient S/MARs (*11*) and are a potential source for short though highly active elements of this type. In the present context, therefore, our primary choice is a 10 kb sequence for which initial SIDD data have been published before (Figure 7 in ref *10*). This stretch comprises a pseudogene (WP18) beside the IFNα genes A10 and A7. Interestingly, in this and some other cases the pseudogenes (open symbols in Figure 1) have lost flanking S/MAR sequences, indicating that there is selective pressure acting on biological functions that are associated with the physicochemical properties investigated here.

In a previous study (*22*) we have demonstrated the way in which in vitro experiments can profit from the information contained in SIDD profiles in the sense that genuine S/MAR sequences can be pinpointed and discriminated from major portions of non-S/MAR DNA. Such an approach appears to be a prerequisite for valid data bank entries, a need that has been underlined by two recent studies (*23*, *24*), which have been initiated to extract common sequence features from SMARt-DB (http://transfac.gbf.de/SMARtDB/), the most comprehensive database on S/MAR sequences available to date (*25*). These evaluations clearly demonstrated the category of problems associated with conventional S/MAR detection procedures as they call for a more detailed delimitation of active parts within a putative S/MAR sequence and for a standardized way to classify S/MAR activities. Accordingly, the present series of experiments is based on the predictions of the SIDD profile for the rational design of reassociation experiments. By such a combined in silico/in vitro approach, we are in a position to derive the rules that govern reassociation strength. Binding strength seems to be an important indicator for the biological activity of an S/MAR element since it is paralleled by transcriptional potential (*9*, *26*). Since other sites with a particular regulatory potential, among these the so-called nuclease-hypersensitive sites, also show concomitants in a SIDD profile (*11*), it appears entirely conceivable that these features will enable a SIDD-based prediction of functional genes and their associated regulatory and—as we will demonstrate below—structural elements.

MATERIALS AND METHODS

Cell Culture and Isolation of Nuclei. Mouse LTK cells, line 2_4, were cultivated at 37 °C in a humidified atmosphere of 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal calf serum, 100 units of penicillin/mL, and 100 μg of streptomycin/mL.

Cloning and Labeling of Tested DNA Fragments. The DNA fragments tested in S/MAR binding assays were isolated from BAC-DNA (accession numbers AL353732, AL512606, and AL390882). The DNA fragments were amplified by PCR with *Eco*RI/*Mfe*I-flanked primers and either cloned into the *Eco*RI site of the pTZ18R cloning vector (Pharmacia) or used directly in reassociation experiments. PCR was carried out in a total volume of 50 μL with 200 μM dNTP, 1 μM primer mix, and 1.25 units of Taq polymerase (Perkin-Elmer). Amplified DNA fragments were precipitated with ethanol/0.6 M LiCl, digested with *Eco*RI/

	PCR Primers
F1fw	GTAATTCCGAATTCCCTCATTCCATTTAAGGACTCAGTAT
F1rev	GAAATAGCGAATTCCCCTTAAAAGTATCAATCCAAT
P-A10fw	GTAATTCCGAATTCTTCAATACATCAGTACTTGTGTCAA
P-A10rev	GAAATAGCGAATTCCGTATACCTCTACTTCACAGATAAA
T-A10-1fw	GTAATTCCGAATTCAAAGACTCACTTCTATAACCACGAC
T-A10-1rev	GAAATAGCGAATTCCCTTGGCATTTTATCTTAAAGAATA
T-A10-2fw	GTAATTCCGAATTCTATTCTTTAAGATAAAATGCCAAGG
T-A10-3rev	ATCTACATTCTGCTAAGAGGAATTCATTGTTTTTTTACT
F4fw	CTCTGCATATGGCTATCC
F4rev	GTAATAAGTTTGGACCCCCC
F4/5fw	GGGCATTTGGGTTGGTTTCC
F4/5rev	GCCCCATCGAAAAGTGGG
F6fw	GTAATTCCGAATTCTTTAAGGACTCAGTATCTATAGGGC
F6rev	GAAATAGCGAATTCCTACGAGAACTAATTTTTGTACCC
P-A7fw	GTAATTCCGAATTCAATATCTGGATGAATACTGCAGCTA
P-A7rev	GAAATAGCGAATTCAATATCTGGATGAATACTGCAGCTA
T-A7fw	TTACAGCAGTTAGTATAACC
T-A7rev	CCTGTGCAGGCACTAGTCC
P-A2fw	GCCGAATTCGGCACAAAAATATAGTTAGA
P-A2rev	GAAGAATTCCTCACCACATTTTTTTCATT
T-A2(1)fw	GCCGAATTCTTTCTGCTATGACCATGACA
T-A2(1)rev	GAAGAATTCGTGGCTGTGGATGTTTCCTT
T-A2(2)fw	GCCGAATTCTTTCTGCTATGACCATGACA
T-A2(2)rev	GAAGAATTCTGAAATGCAGGACACAGGTCC
P-B1fw	GTGCAATTGCAATTCCCTCTTATAAAACGTT
P-B1rev	CAGCAATTGTATTCAGAGGAATTTCCCACTT
T-B1fw	GCCGAATTCCTTTTACTTCATTAACAGACT
T-B1rev	GAAGAATTCCCCTCCAGCTTTCCTTTTCT
F20fw	GCCGAATTCTTCTAGTTACTTGAGGTGAAAA
F20rev	GAAGAATTCGGGATAATTTCATCAAGAGAAT
F75fw	GTGCAATTGTAAAAGGTTGGGGTAGAGACAT
F75rev	CAGCAATTGTTACCTCTGTATGTAGCTTTGC
F100fw	GTGCAATTGCATCTCATACTCAAGATCTGTC
F100rev	CAGCAATTGACTAATGTCCTAGCTCACACTC
F180fw	GCCGAATTCAATCACAATGAAGCTTCTTATT
F180rev	GAAGAATTCATTGGTTTGAAAAGCCACTTTT
F230fw	GCCGAATTCATTCAGGAAGTTTTAAATGTTC
F230rev	GAAGAATTCTAAGTGAGCAGGGGACATTTAA
F275fw	GCCGAATTCAACAGTTCATTTAAATCATTACCAC
F275rev	GAAGAATTCCTCCTAATCCCTATGTTGTTCA

*Mfe*I for 1 h, and isolated by gel extraction (Qiagen). Ligation was carried out at 16 °C overnight. The DNA sequence of all cloned fragments was confirmed by sequence analysis. For the labeling reaction the *Eco*RI restriction fragments were end-labeled by Klenow polymerase and [α-^{35}S]dATP. Each assay required ∼20−60 ng of ^{35}S-end-labeled fragments.

Nuclear Halos and Nuclear Scaffolds. The detailed procedure has been published in ref *9*. In the following, details are only given in cases where modifications have been introduced since our first publication in 1995. For each experiment two plates of 150 cm^2 (∼5 × 10^7 cells) were used. Cells were washed twice with 15 mL of isolation buffer; they were then scraped off in 15 mL of the same buffer containing 0.1% digitonin, and nuclei were set free by homogenization in a tight-fitting Dounce homogenizer. Isolated nuclei were resuspended in 200 μL of freshly prepared nuclear buffer and incubated at 37 °C for 30 min. Lithium 3,5-diiodosalicylate (10× LIS) buffer was added, and after several strokes with a loosely fitting Dounce homogenizer the suspension was centrifuged at 2400*g* for 5 min. After centrifugation, the resulting pellet consisted of nuclear halos. The halos were washed three times in 20 mL of restriction buffer and could afterward be stored in 500 μL of restriction buffer plus 5% glycerol at −70 °C.

Quantification of S/MAR Activities in Vitro (Halo Reassociation). Nuclear halos were thawed and predigested with 400 units of *Eco*RI at 37 °C for 5 h. They were then aliquoted into seven tubes, and 50 μg of sonicated *Escherichia coli* competitor DNA was added to each assay mixture. After 15 min of preincubation the nuclear remnants were supplied with ∼200 000 cpm of labeled DNA fragments. Samples were then incubated at 37 °C until thermodynamic equilibrium was reached (*9*), preferably overnight. The nuclear remnants containing the S/MARs were separated from the unbound fraction by centrifugation. "Equivalent" fractions for P and S were determined for our pTZ-E standard as detailed before (*9*) and maintained throughout the series of samples. P and S were then treated with proteinase K (1 mg/mL) at 56 °C for several hours. DNA was precipitated with ethanol/0.6 M LiCl, resuspended in TE, and analyzed by gel electrophoresis and subsequent blotting onto a nylon membrane followed by autoradiography. The intensity of the blotted fragments was measured via Phosphoimager and evaluated by ImageQuant. All binding data were referenced to the 95% binding activity of fragment E; 95% (±8%) is the average binding contribution in numerous experiments performed according to this protocol. Relative to this standard, the respective subfragment SAR_{800} has a binding activity of 70% ± 8% (see Figure 2 for an example). In our operational definition an S/MAR has a binding activity exceeding 50%.

SIDD Calculations. SIDD profiles were calculated according to an algorithm developed by Benham (*10, 15, 26*

and references therein). The incremental free energy $G(x)$ needed to separate the base pair at each position x is computed in the context of a defined stretch of DNA. A value of $G(x)$ near or below zero indicates an essentially completely destabilized base pair with a high probability to denature at equilibrium. SIDD profiles, plots of $G(x)$ vs x, visualize regions of the sequence where superhelical stress will destabilize the duplex. The calculations used in this publication correspond to a superhelix density of $\sigma = -0.05$ to -0.06, which reflects the average physiological value. SIDD profiles have been calculated for three continuous sections within the gene cluster, i.e., part of locus AL390882 (complementary, 117 934 bp), AL353732 (complementary, 193 116 bp), and HSIFD1 (direct, 9937 bp).

RESULTS

Human Type 1 Interferon Gene Cluster on Chromosome 9p22. The general architecture of the type 1 interferon gene cluster has been known for some time (*11*, *21*, *27*): it consists of three subclusters (I−III) of α and ω genes in addition to the IFNβ gene (IFNB1) which is located separately (Figure 1). A 14 kb domain size has been derived for IFNB1 from humans and 22 kb for mice (*28*) but considerably smaller domain sizes for the remaining members of the cluster. The present study concentrates on two coding regions (marked a and c in Figure 1) and on an extended noncoding region (b) between IFNB1 and subcluster III. The emerging principles are supported by analyses of additional sequences.

Structure of a Prototype S/MAR with a Known Spectrum of Biological Activities and Physicochemical Properties. The IFNB1 upstream S/MAR element E was among the first domain-boundary elements that has been localized by DNase I accessibility, nucleosome distribution (*28*), and occurrence of constitutive DNase I hypersensitive (HS) sites (*11*, *28*) in addition to conventional halo-mapping and reassociation analyses (*9*). The spectrum of biological activities associated with this and other prototype S/MAR elements has been studied in our laboratory and others (*4*, *11*−*14*, *28*−*31*). Biochemical studies have first indicated that element E has the propensity to separate strands under superhelical tension (*32*, *33*). On the basis of this finding, computational studies (*10*) have shown that this potential is also reflected in the corresponding SIDD profile.

Figure 2 reproduces the SIDD and binding characteristics of a 2.2 kb fragment E and its 858 bp subfragment SAR_{800}, which have become two widely used S/MAR standards. We have marked the nucleation point of base unpairing, which is seen to coincide with one of the most pronounced SIDD minima. Meanwhile, it has been possible to accurately relate the relative binding potential of unknown fragments to a distinct set of SIDD parameters (*30*; Benham et al., manuscript in preparation).

Structure of the IFNB1 Domain. The structure of the largest type 1 interferon domain has been determined, initially according to the availability of restriction sites (*34*). It is remarkable that the coding region (usually investigated as part of *Eco*RI fragment F) has a minor binding potential in the trough of a smooth curve. Using SIDD-based predictions, we have now designed more restricted PCR fragments (details in Figure 3) and find that the transition between nonbinding and binding is much sharper than previously

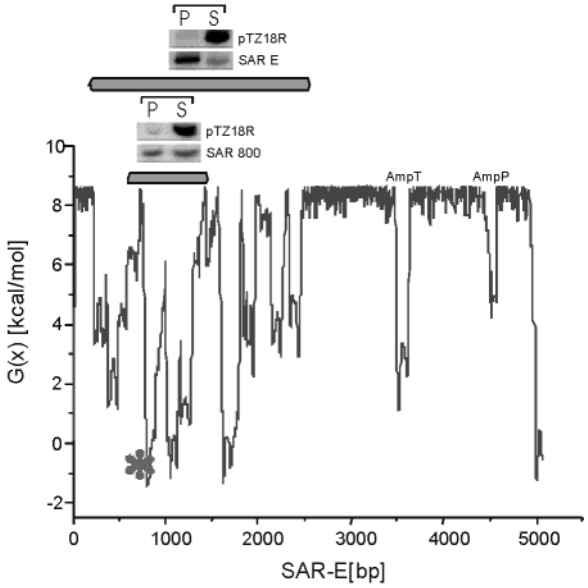

FIGURE 2: SIDD structure of a prototype S/MAR element (S/MAR-standard E). The IFNB1-upstream S/MAR element (*Eco*RI fragment E) was cloned into a pTZ18R backbone as described before (*10*). Binding strength is a function of the fragment's extension, the number of unpairing elements (UEs), the amount of their destabilization (calculated at a superhelical density $\sigma = -0.05$) and their spacing (*10*; manuscript in preparation). The amount of destabilization is determined relative to the classical amp^r-associated UEs (*10*). Two S/MAR standards have been marked: E as a whole and its 858 bp subfragment SAR_{800}. The nucleation center of base-unpairing (core-unpairing element, CUE) has been indicated by an asterisk. E and SAR_{800} have been described as fragments I and IV in Mielke et al. (*32*). Inset: prototype binding assay. Binding of S/MAR E was 86% (95%), and that of SAR_{800} was 56% (62%); values in parentheses are the result of normalizing to our 95% binding standard (Materials and Methods).

thought: while the coding region binds to less than 5%, the promoter region has a normalized binding activity of 22%. Together with the adjacent UE, which coincides with a prominent DNase I hypersensitive site (*11*, *28*), this activity reaches 49% (Figure 3b). On the downstream end, the 3′-terminal flank (further referred to as terminator because of brevity) is embedded in a highly destabilized site with a considerable 79% binding activity.

The properties of the 2.2 kb upstream *Eco*RI element (S/MAR standard E) are inserted in Figure 2. Its 95% binding activity decreases to 70% ± 8% if it is cut down to an 858 bp segment from the most destabilized region and if the binding of this SAR_{800} fragment is normalized to the parent fragment. These results demonstrate two opposing forces: a fragment from the center of a S/MAR will lose binding potential since its extension and the number of participating UEs is reduced. On the other hand, if a stretch of DNA with binding potential comprises non-S/MAR components, its affinity will increase in case non-S/MAR constituents are removed. These findings indicate the category of problems associated with various database entries: S/MARs may qualify as such since they are either long enough or because an active core has been met by incidence. Accordingly, false negatives may arise because the extension of the fragment is subthreshold or because it contains too much of non-S/MAR DNA.

Characterization of Destabilized Regions in the WP18A10A7 Region. For the WP18A10A7 DNA locus (*10*; Figure 4) the

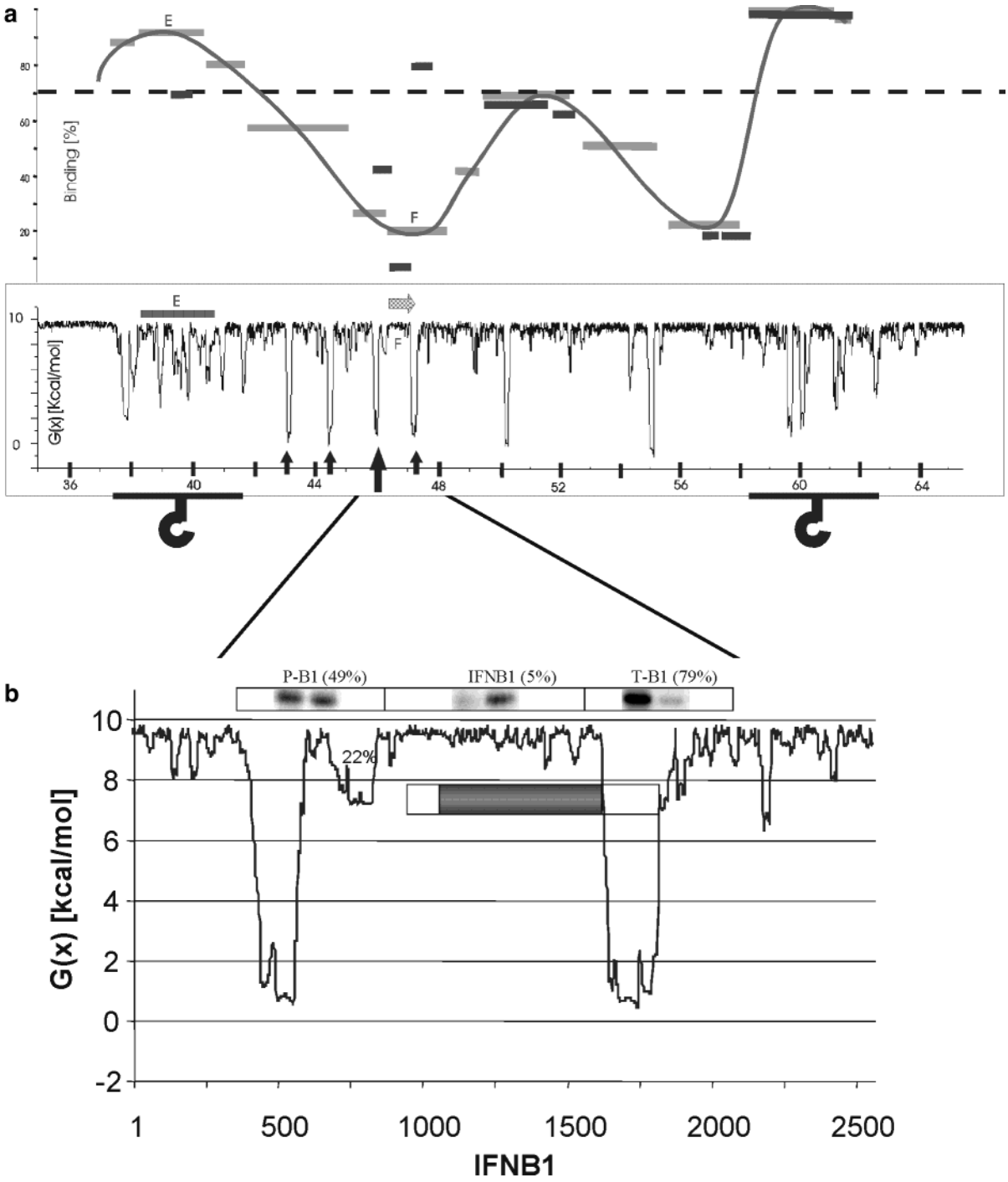

FIGURE 3: SIDD structure of the complete IFNB1 chromatin domain (section c in Figure 1). (a, top panel) Correlation of binding activities and SIDD properties. The two strong flanking S/MAR elements (hooks) are seen to consist of a repetitive series of moderately destabilized unpairing elements (UEs). Isolated, strong UEs correlate with established DNase I hypersensitive sites (arrows; *11*). The fat arrow marks fragile site A1235 from Figure 1. (b, bottom panel) SIDD profile of the IFNB1 gene region. Whereas the promoter in the narrow sense is only moderately destabilized (22% binding), it is accompanied by a stronger upstream unpairing element (total binding activity 49%) representing a DNase I hypersensitive site. A widespread feature is the occurrence of a more strongly destabilized region (79% binding) at the downstream region, while the coding region itself shows both negligible binding and destabilization.

SIDD profile shows a number of base-unpairing regions (BURs) with a predicted potential to mediate attachment to the nuclear matrix. Interestingly, in this particular example, the prediction is mirrored by the results of the MARfinder algorithm, which is based on an entirely different concept (see Figure 7 below and discussion).

To verify the suspected correlation between prediction and in vitro binding, we tested seven individual sequences with calculated base-unpairing potential and, in addition, several predicted negative controls. We cloned plasmid constructs containing the individual PCR-amplified DNA fragments and quantitated the scaffold-reassociation potential of the inserts. As a positive control, the strongly (95%) binding S/MAR-standard E (called I in ref *32*) was used, and the designation S/MAR was given to all elements binding >50% under standardized conditions. To rule out nonspecific binding contributions to the nuclear matrix, the vector backbone served as a convenient negative control. The results of a

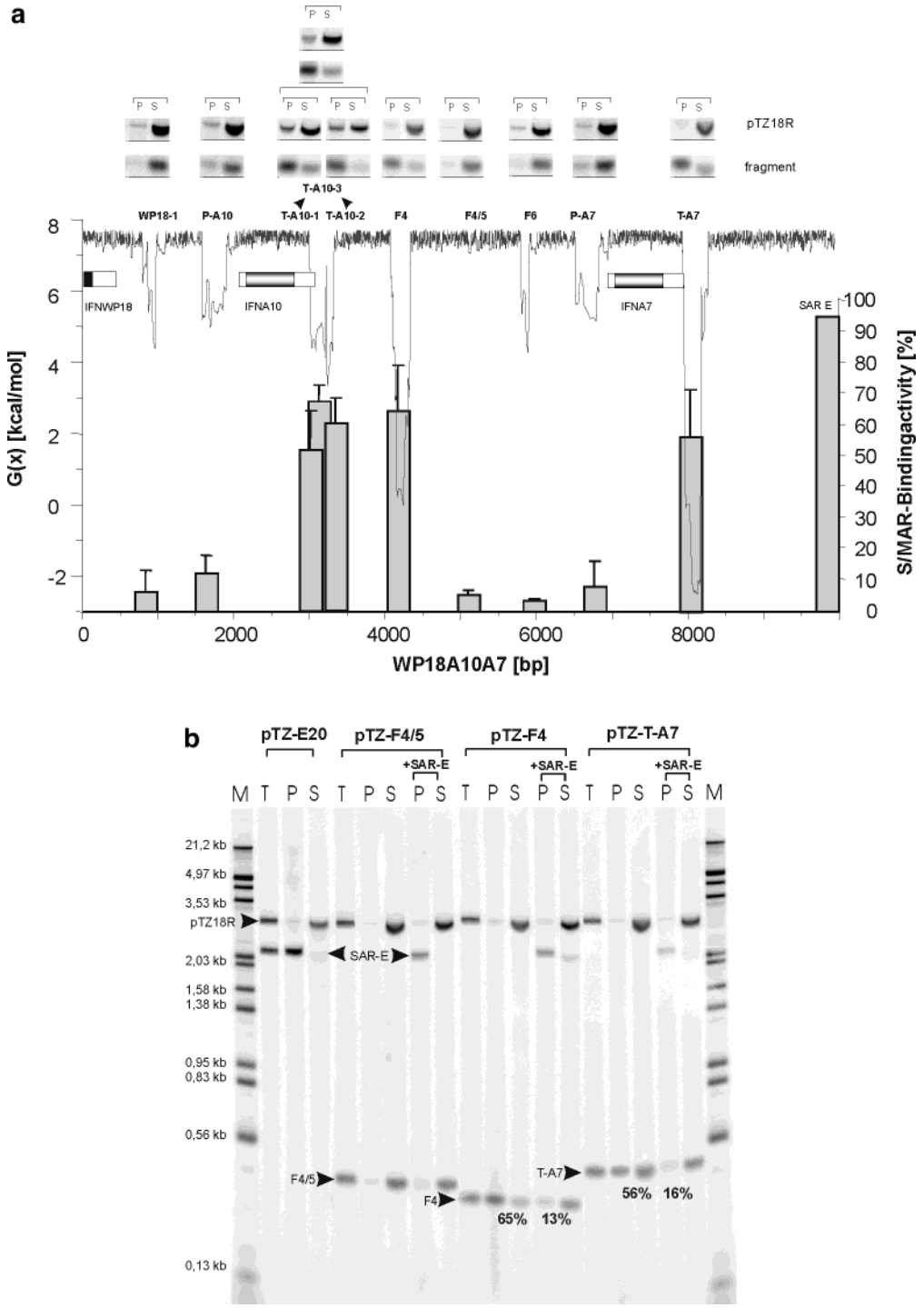

FIGURE 4: Correlations between prediction and in vitro binding potential for fragments from a 10 kb region of the type 1 interferon gene cluster (region a in Figure 1). (a, top panel) The coding regions are regularly surrounded by duplex-destabilized DNA, except for the WP18 pseudogene. Destabilized regions have consistent binding potential whereas stable sections do not. See the top panel for reassociation analyses and Table 1 for their numerical evaluation. (b, bottom panel) Competition analyses for selected fragments. Fragment E serves a dual role: it is a stringent control of scaffold quality. Lanes marked pTZ-E20 with T, total input; P, pellet (S/MAR) fraction; and S, supernatant (non-S/MAR) fraction. Fragment E is also a competitor in the reassociation experiments marked +SAR-E, which is added to discriminate the "prototype" from an "alternative" binding mode. Associating fragments F4 and terminator fragment T-A7 are seen both to be competed off by the S/MAR standard.

binding assay that is performed at a 40 000-fold molar excess of bacterial (*E. coli* genomic) DNA according to ref *9* are depicted in Figure 4a. Already a superficial inspection shows that, in fact, all predicted destabilized sequences show an affinity to the nuclear matrix, i.e., they partition into the pellet fraction while the plasmid part of the vector is found in the supernatant at a consistent abundance of about 98%.

The binding strength of individual fragments is listed in Table 1. The region downstream from pseudogene WP18 is barely destabilized. For this gene interferon-like motifs have been described, although the sequence is interrupted by multiple insertions and deletions such that termination codons are present in all frames (*35*). Remarkably, we could not detect any matrix binding potential for DNA adjacent to this

Table 1: Scaffold Reassociation Strength for Prototype and Novel S/MAR Elements[a]

element	AT content (%)	fragment size (bp)	binding act. (%)	S/MAR
		Standards		
Ig-κ (*7*)	70	370	50	+
Igμ-3′ (*8*)	70	431	86	+
S/MAR E (*32*)	69.2	2200	95	+
SAR_{800} (*32*)	69	858	70	+
		WP18-A7 Range		
WP18-1	73.7	333	7	−
P-A10	75.5	443	12	−
T-A10-1	72.3	328	52	+
T-A10-2	73.8	290	60	+
T-A10-3	72.8	583	69	+
F4	72.7	372	65	+
F4/5	62.9	463	5	−
F6	70.8	308	3	−
P-A7	77.1	429	8	−
T-A7	75.6	508	56	+
average	72.7	406	33.5	
		Promoters/Terminators		
T-A1[b]	65.2	2066	67	
P-A2[b]	74.1	557	50	
T-A2(1)	68.5	896	24	
T-A2(2)	67.7	2275	82	
P-B1[c]	70.9	516	49 (22)	
T-B1[c]	66.4	500	79	
		Periodic Attachment Elements[d]		
F20	73.4	567	95	
F75	73.2	1620	96	
F100	70.5	526	61	
F180	72.7	773	57	
F230	73	504	83	
F275	75.7	515	73	

[a] Quantification of binding activities follows the procedure described by Kay and Bode (*9*) with SAR E as the 95% binding standard. In the present context, we have designated >50% binding fragments as S/MARs (+). See Figure 4a for standard deviations. Abbreviations: P, promoter; T, 3′-terminal flank (terminator); F, fragment. [b] Figure 5b. [c] Figure 4. [d] Figure 6.

pseudogene, in either the SIDD or MARfinder profiles (cf. Figure 7) or in the in vitro test.

For the two peaks flanking the coding region of IFNA10 an activity of about 12% (promoter) and 67% ± 5% (polyadenylation site) was determined in close agreement with the IFNB1 data in Figure 3b. The corresponding observations were made for the IFNA7 gene. In this case the promoter region bound to the matrix preparations with an average activity of 8%, while 56% of the terminator was bound. Another destabilized peak located at map position 4 kb in the intergenic spacer (fragment F4 in Figure 4) has an affinity of about 65%.

Since binding strength generally increases in parallel to the extension of a base-unpairing region, the 583 bp A10 terminator fragment T-A10-3 was subdivided into two overlapping 328 and 290 bp sections in order to see if the binding potential is evenly distributed (as suggested by the shape of the UE) or concentrated somewhere within this region. Although these fragments are close to the minimum effective length of a naturally occurring S/MAR, their binding was only slightly reduced (to 50% for T-A10-1 and to 60% for T-A10-2; see Figure 4a). Still, there is a clear additivity of BURs as commonly found for S/MAR elements (see the standards in Figure 2).

To exclude the possibility that sequences devoid of unpairing potential also attach to the nuclear matrix, we tested a 500 bp fragment located in a stable region of the investigated domain around map position 4500 (F4/5). In accord with the predictions, this region with a duplex stabilization energy of 9−10 kcal/mol shows no affinity for the nuclear matrix. Other controls are the coding regions themselves, which are never destabilized and never show an affinity for isolated matrixes (see Figure 3b for an example).

Specificity of Binding: Do Binding Sites Overlap with Those of the S/MAR Standard? Introduced in Figure 2, our S/MAR standard E (2.2 kb) has so far served the role of a control of scaffold binding strength and capacity. In an extension of the binding experiments E was added together with the fragment in question in order to test the relation of the respective binding sites. Among the fragments investigated here is a strongly binding terminator (T-A7) and a pronounced intergenic signal at map position 4 kb (F4). Here, the most clear-cut results arise from fragment F4, which appears to be in mutual competition with the E standard: presence of E reduces the binding contribution from 65% to 13%, while E itself is reduced from 95% to 85%. Similar results are found for terminator T-A7, which is reduced from 56% to 16%, for terminator T-B1 (Figure 5b), and for T-A10 (not shown).

Table 1 summarizes these results and lists them together with AT content and fragment sizes. These data clearly show that, although within the present series most S/MAR elements are AT-rich, there is no direct correlation between AT content and binding potential (two fragments with AT contents of 73.7% and 73.8% bind at 7% and 60%, respectively; here binding is stronger for the shorter of the two fragments, although it approaches the minimum effective length of natural S/MARs). Another claim is that beyond a certain threshold of AT richness any element will behave as an S/MAR (*23*). If this were the case, fragment P-A10 clearly shows that this threshold must be above 75.5%. Note that our 50% binding strength criterion is met for other prototype S/MARs such as the Ig-κ element (*7*) and the Ig-μ-3′ element (*8*), which comprises the CUE of the region (*10*, *32*).

Association of Promoter and Terminator Elements with Destabilized Regions: A General Phenomenon? So far our results indicate that some duplex destabilization is found for several of the promoter and terminator regions, in agreement with prior work on yeast (*26*). However, the binding potential of these elements does not compare to that of strong S/MARs, which becomes particularly evident in the above competition binding assays. Equally interesting is the fact that the coding region of a gene, which is flanked by such destabilized elements, contains no destabilized regions itself (see example in Figure 3a). It has been proposed (*26*) that this lack of base-unpairing potential in an open reading frame (ORF) may be required as stress-induced alternate secondary structures in this region would impede the progress of the polymerase.

If we inspect the collection of destabilization profiles of IFN genes shown in Figure 5, we have to conclude that presence of destabilized regions upstream and downstream from the coding sequence is a general feature. The surroundings of IFNB1, IFNA10, and IFNA7 are all marked by a moderately destabilized immediate promoter and by a

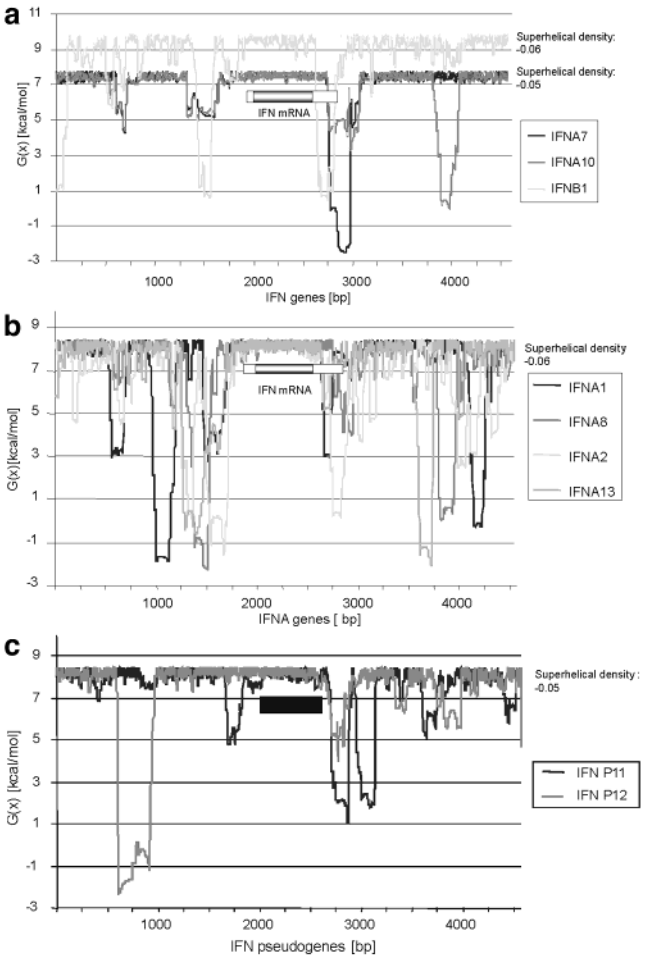

FIGURE 5: Common and distinct features of other type 1 interferons: (a, top panel) genes with a moderately destabilized promoter; (b, middle panel) genes for which the promoter overlaps with a strongly destabilized region; (c, bottom panel) pseudogenes.

terminator that shows a marked destabilization and a considerable matrix-reassociation potential. On the other hand the promoters of IFNA1, IFNA2, IFNA8, and IFNA13 overlap with a strongly destabilized region, which results in genes that are flanked by two regions with scaffold attachment potential (Figure 5b). The terminator sequence of IFNA1 was tested in greater detail and supports our hypothesis (67% affinity of the entire destabilized region; see Table 1). Similar tendencies were also observed for IFNA2, the surrounding region of which is strongly destabilized. Especially the extended promotor area shows a significant binding potential of 50%. It appears that moderate destabilization of a promoter is sufficient for and more extensive destabilization compatible with its function. These characteristics of genes indicate the existence of a structural code for ORFs, their immediate flanks, and the remote domain boundaries. In summary, destabilization features and base unpairing potential seem to be a prerequisite for gene function as they are associated with all genes investigated in this context.

BURs may become abandoned after gene inactivation as indicated by several pseudogenes (Figures 5c and 7). Besides IFNWP18 (Figures 4a and 7), IFNP11 and IFNP12 (Figure 1) are other members of the human interferon gene cluster that became nonfunctional by the introduction of several stop codons. Following this event, IFNP12 seems to have lost any strand separation potential in the promoter area while the terminator sequence still contains a moderate BUR. IFNP11, on the other hand, has maintained its destabilization potential both for its promoter and for its terminator (Figure 5c).

Repetitive Structural Elements in an Extended, Noncoding Region. Binding of S/MARs to whole scaffolds occurs with positive cooperativity (*32*) in accord with the mass-binding mode described for an abundant scaffold component, scaffold attachment factor A (SAF-A/hnRNP-U; *16*). Long S/MARs appear to associate with extended clusters of such a protein, whereby the number of binding sites becomes limited (3000 for fragment E). This mode of binding is the reason for the consistent observation that a shorter fragment associates at a higher "moles of S/MAR per nuclear equivalent" ratio (see the 15 000−20 000 binding sites for SAR_{800} reported in ref *32*).

If a scaffold is provided with 3000 nuclear equivalents of fragment E, this treatment efficiently prevents binding of prototype terminator elements, which is exemplified by T-B1 in Figure 6b and the corresponding binding curve in Figure 6c. In other words, whereas T-B1 exhibits a 79% binding activity in the absence of E (Figure 3b), its apparent activity is reduced to about 20% in the presence of saturating amounts of E (Table 1), confirming that the respective binding sites are mostly overlapping and thereby in competition with each other.

Between subclusters III and IFNB1 we find one of the more extended, gene-free regions. For a 20 kb section and some easily accessible subfragments thereof, we could not trace scaffold attachment potential in our pilot study (*34*), in agreement with later investigations by Strissel and coworkers (*11*, *27*; summarized in Figure 1). On the other hand, the detailed study of this region by SIDD profiling reveals a striking series of 250−500 bp wide minima that are regularly spaced at an approximate 3000 or 2 × 3000 bp periodicity. These observations triggered a more detailed comparative study between fragment T-B1 and F20, which can be directly compared as they are comparable both regarding their extension and the amount of destabilization. For this statement to be valid it is important to know that for both fragments SIDD properties have been calculated in the context of the same 117 kb DNA section where all UEs are in direct competition [Figure 6c (inset) and Materials and Methods]. In the absence of E, both elements can be classified as strong binders (79% for T-B1 and 95% for F20). However, if saturating amounts of E are added, the binding behavior of these two fragments becomes radically different: while binding of T-B1 is efficiently prevented as demonstrated above, F20 is able to bind with characteristics that remain unaffected by the presence of the S/MAR standard E (Figure 5c).

Figure 6c shows the common cooperative binding behavior also for F20. Saturation occurs at 15 900 (±2000) cpm, corresponding to 10 000−15 000 sites per nuclear equivalent. A simulation of this sigmoid function according to Adair (*32*) suggests that affinities of the fragments associating first and last are at least 2 orders of magnitude different, possibly much more. These results suggest the existence of a second set of interacting binding sites that are distinct from standard S/MAR sites. The function of these sites may be exclusively structural. It will be interesting to reveal the nuclear binding

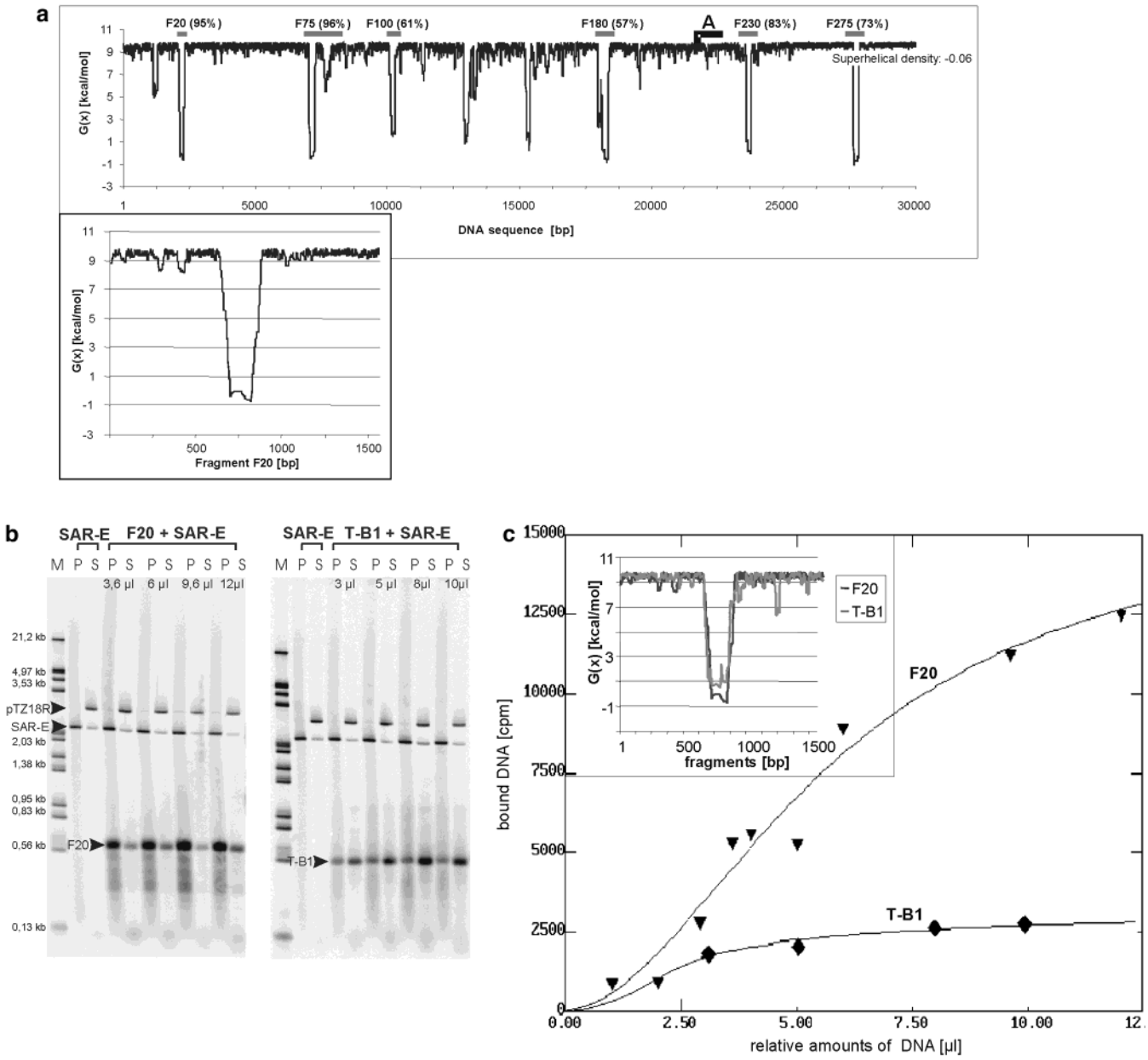

FIGURE 6: Structural signals (UEs) in an extended noncoding region and their relation to S/MARs. (a, top panel) Position of fragment 20 (F20) in an extended noncoding region upstream from the IFNB1 gene (section b in Figure 1). Note the apparent 2800 ± 500 and 5000 ± 700 periodicities of sharp destabilized sites in this region. Considering their limited extension (typically 250−500 bp), binding activities are unusually strong. (b, bottom left panel) Binding sites for F20 and a terminator sequence (T-B1, see Figure 3) are distinct. Saturating amounts of S/MAR-element E prevent binding of an otherwise associating terminator (cf. Figure 3b); they do not affect binding of the external fragment F20. (c, bottom right panel) Characteristics of F20 and T-B1 saturation binding curves in the presence of E (evaluation of panel b). Note the sigmoid shape in the presence of saturating amounts (3000 molecules/nuclear scaffold equivalent) of fragment E. Association of E blocks the majority of T-B1 sites but permits the additional association of approximately 10 000−15 000 F20 molecules.

partners of these novel sites by electrophoretic mobility shift assay (EMSA), cross-linking (*36*), or footprinting technologies and to determine whether this set of attachment sequences shares the spectrum of biological activities with that of the standard S/MARs, which have been studied for many years.

DISCUSSION

Traditionally, S/MARs have been characterized by a variety of classical procedures all of which involve some extraction of nuclei for the removal of histones and other soluble factors. Accordingly, there have been extensive claims that the entities recovered this way and, at the same time, the associating DNAs might represent artifacts (*36*, *37*). However, there are very recent data to suggest that the nucleoskeleton in fact represents a permanent structure within nuclei regardless of their transcriptional activity (*38*). Novel in vivo cross-linking procedures (*39*) as well as mild in situ extraction steps coupled with FISH (fluorescence in situ hybridization) visualization (halo-FISH) have yielded confirmatory evidence that the classical S/MARs are in fact the elements that, by matrix attachment, provide the eukaryotic genome with a higher order organization (*29*, *40*). Finally, functional tests consistently reveal a spectrum of S/MAR-

associated biological activities, which are in full agreement with the postulated role of a boundary element, an insulator, and an enhancer-supporting structure (*3*, *4*, *32*, *40*, *41*).

Using the type 1 interferon gene cluster as a paradigm, we have studied a group of genes and pseudogenes with a promoter/terminator destabilization pattern related to that of classically defined S/MAR elements. Comparison of inducible genes and pseudogenes suggests that a prerequisite for transcription is the organization into functional units with destabilized promoter and terminator sequences. Still, on our scale the association potential of such elements is weak to moderate even for functional genes (Table 1). While in some examples the terminator region gains an activity approaching that of prototype S/MARs, in others such a function is contributed or improved by a separate element in the intergenic space (Figure 4). The elements that embed the coding regions may represent nucleosome-free (*11*) destabilized opening sites important for the entering and release of transcription factor complexes (promoter) and a tool to immediately relax positive superhelical tension as it arises from the action of a tracking protein (terminator) before final relaxation is provided by a topoisomerase (*42*, *43*).

S/MAR Binding Modes: Two Prominent Protein Binding Partners. The spatial distribution of S/MAR binding centers in the nuclear scaffold has been studied by Ludérus et al. (*44*, *45*): S/MAR sites appear to be distributed equally over the peripheral nuclear lamina and the internal fibrogranular network. In case of the lamina, lamins B (and later A) have been determined as the major binding partners. The specificity of a S/MAR−scaffold interaction could largely be reproduced with paracrystallike lamin polymers revealing two different types of interaction that appear to be related to different features of S/MARs. One type involves the regions with single-strand potential, and the other, the minor groove of the DNA double strand. Both modes of association are interdependent since S/MAR binding is largely inhibited by the presence of single-stranded competitors.

The other major component, scaffold attachment factor A (SAF-A), otherwise known as SP120 (*17*) or hnRNP-U, is a major member of the heterogeneous nuclear ribonucleoproteins (hnRNPs) (*16*). SAF-A shows a pronounced propensity to self-polymerize and in this state it associates with multiple S/MAR elements in a process that is marked by positive cooperativity (*16*) and that is optimal if AT tracts (short stretches of consecutive As and Ts), are distributed according to the rules specified by Tsutsui (*17*). UV cross-linking experiments show that SAF-A is likewise associated with DNA in vivo. The protein's primary structure reflects its dual role as there are two independent nucleic acid binding domains: (i) the previously characterized C-terminal RNA/ssDNA binding domain (RGG box) and (ii) the S/MAR-specific 45 amino acid N-terminal domain, which is cleaved during apoptosis (*46*). Such a "SAF box" recurs in several eukaryotic proteins and represents the first characterized protein domain specifically recognizing S/MARs.

For a cloned SAF box, the characteristic association with S/MARs can be demonstrated by pull-down experiments in case a critical protein density is reached on the surface of Sepharose beads. According to the "mass-binding" mode, each individual domain interacts just weakly with a DNA element, which may be a UE according to Figure 2. Only the simultaneous binding of multiple sites will therefore confer a strong, specific interaction, which explains the fact that there are hardly any naturally occurring S/MARs below a critical length of 250 bp. A failure of ssDNA to compete for the interaction of S/MARs with SAF boxes shows that many but not all criteria of scaffold−S/MAR interactions can be explained by SAF box proteins and, therefore, the 50% competition limit found for the complete scaffold (*47*) probably reflects the contribution of the lamins in addition to SAF-A.

Approaches Used for the Development of S/MAR Prediction Tools. Although standard S/MAR elements have been described as being AT-rich, Table 1 reconfirms that AT richness per se is not sufficient to yield a S/MAR, at least in the range below 80% A+T. In fact, some of the most common S/MAR standards have only a 70% abundance of A+T. Recent data have rather shown the relevance of localized AT-rich areas (the AT tracts mentioned in ref *17*), which are related to 90% AT boxes (blocks of 20 nucleotides of which at least 18 are A or T) of Michalowski et al. (*48*). Apparently, for these AT tracts to interact, the intervening DNA must not exceed 500 bp (*18*). In another study (*49*), S/MARs with a balanced (46−61%) A+T content have been recovered from random chromosome 19 fragments by affinity selection; several of these examples contained >20 bp runs of >75% A+T. Compared to this class of structure−affinity correlations, SIDD profiling can be considered a more refined tool to estimate S/MAR activities since it considers a variety of parameters together, such as the presence, number, arrangement, and normalized amount of stress-induced destabilization, whereby the correlation coefficient could be improved as discussed in ref *48* (Benham et al., manuscript in preparation).

The first bioinformatic approach ever devoted to S/MAR prediction, MARfinder (*50*), is based on the statistical occurrence of motifs that are related to DNA secondary structure. Among these are spacing criteria for AT tracts similar the ones mentioned above and, in addition, sequences that are abundant in 3′-UTRs, curvature/kink-inducing motifs, topoisomerase II recognition sites, and motifs found at origins of replication. In Figure 7 we have compared the "matrix association potential" generated by MARfinder with the SIDD calculation. Here we find a remarkable agreement between the two profiles, which are nearly mirror images of one another in the sense that maxima in matrix-associating potential coincide with minima in duplex stability.

As a very recent addition, experimentally determined S/MAR sequences have been used as a training set to generate a library of S/MAR-associated patterns described in the form of weight matrixes (*51*). Thereby a new tool, SMARtest, has been developed that identifies potential S/MARs via a density analysis (http://genomatix.gsf.de/cgi-bin/smartest_pd/smartest.pl). In contrast to the first mentioned algorithms, SMARtest does not depend on the wider sequence context. On balance, it does not yet quantify S/MAR activities. We have applied all three algorithms in parallel and found a reassuring amount of coincidence for a thoroughly studied example, the scaffold attachment profile covering a T-DNA integration site (*52*). Altogether these data emphasize the importance of secondary structure rather than an immediate sequence consensus as a determinant for the strength of an S/MAR−scaffold interaction.

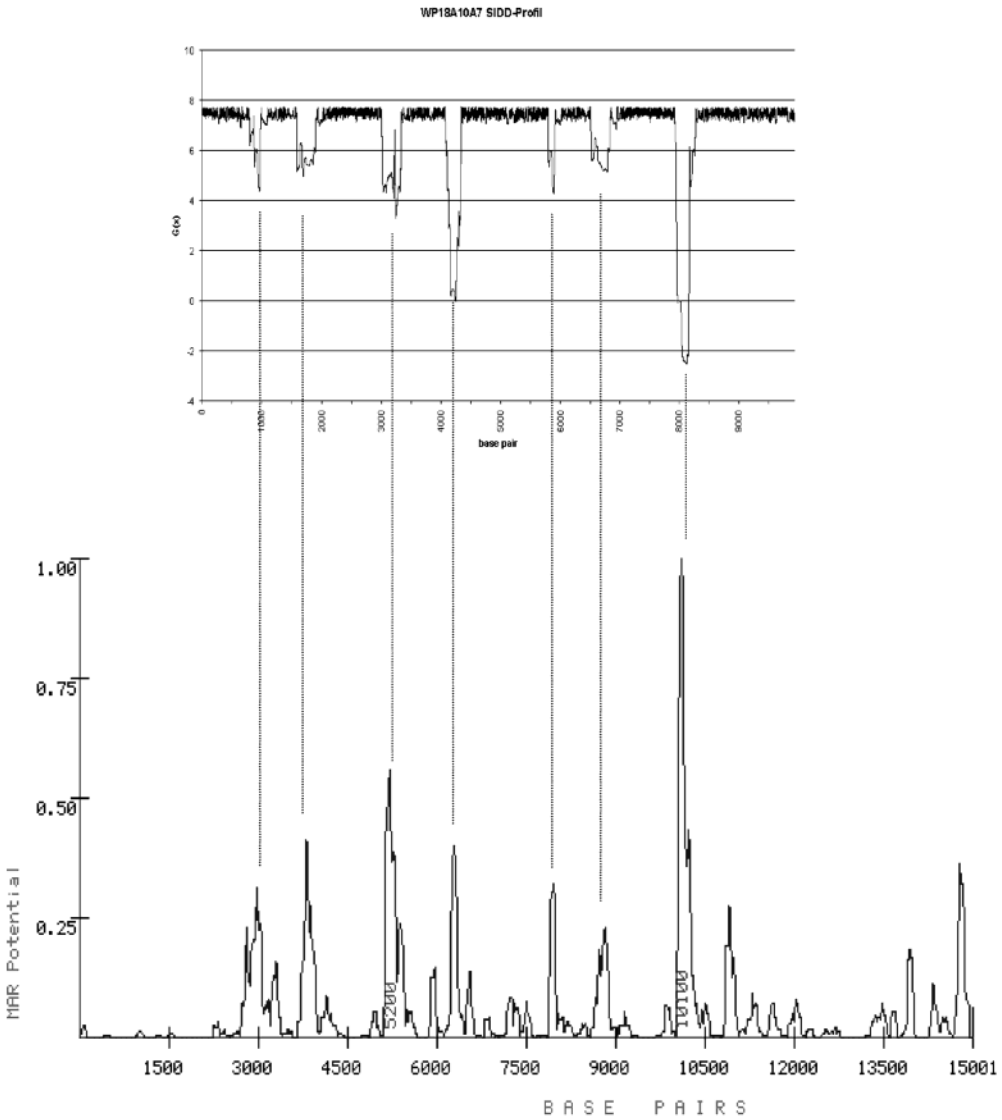

FIGURE 7: Comparative performance of S/MAR prediction algorithms exemplified by region WP18A10A7 from subcluster III (Figure 1). MARfinder analyses were performed with default settings but with window width (100) and window slide distance (10) both narrowed down by a factor of 10.

Imprints of Other Regulatory and Structural Elements in an SIDD Profile. Traditionally, our laboratory has put major emphasis on the SIDD profiling approach. Using SIDD, we have detected a coincidence of sharp unpairing elements with DNase I hypersensitive (HS) sites (*11*), which at the same time represent prominent factor binding regions (unpublished results). Present work also demonstrates that SIDD detects a—possibly novel—class of destabilized sites outside coding regions (Figure 6a). Since these signals do not appear to be HSs in the classical sense, it is likely that they represent structural transitions with a different, possibly constitutive function. Significantly, they appear to be distributed according to a periodicity of $n \times 3000$ bp (which would correspond to 15−20 nucleosomes). The distinct character of these sites is also reflected by the observation that their binding mode is distinct from a standard fragment with very similar duplex-destabilization characteristics (Figure 5c). The SIDD minima themselves may represent nucleosome-positioning signals as demonstrated by Bode et al. (Figure 3 in ref *11*).

There are precedents for periodicities in genomic sequences such as bends at a four-nucleosome interval (680 bp), which is compatible with the genome segmentation rules (350−450 bp units) (*53*, *54*). In the framework of these models, the elements found here may indicate a still higher level of organization for eukaryotic genomes that becomes dominant in regions which do not have to perform regulatory functions in the classical sense. It will be interesting to see if their affinity for SAF-A and lamins is equal to or if they differentiate between these or other binding partners in a way that differs from the classical S/MARs.

Definition of an S/MAR Requires Further Refinement. Recently, we have established a comprehensive database that to date covers 313 entries, collected from 180 references, SMARt-DB (*25*; http://transfac.gbf.de/SMARtDB/). Two groups have tried to subject this material to a systematic analyses of recurring motifs. One of the major conclusions from these studies was, in fact, that any sequence of the respective AT content would contain the same kind and number of conventional factor binding sites. In light of our present work, which disproves any direct link between AT richness and binding activity, at least up to an AT level of 80%, we found that the major obstacles to such an approach may be the following:

(i) Data have been assembled from the results of many laboratories that do not determine activity parameters in any comparable way. While the nature of the scaffold or matrix does not appear to be the major source of discrepancies (*55*), the affinity standards and the kind and amount of competitor DNA are variable as are the rules according to which equivalent fractions have been determined for the pellet and supernatant fractions (*9*). (ii) In the past, the extension of determined fragments was largely dictated by the availability of restriction sites. This appears to be a considerable source of erroneous conclusions as the size of an element may both increase or decrease the level of binding depending on the content of non-S/MAR DNA (see above).

From the work described here, it is our conclusion that an elegant solution to this dilemma is a combined in vitro/in silico approach in the sense that the available prediction schemes are first used for the identification of sequences with S/MAR potential. The critical regions are then PCR-amplified, ideally to a uniform length of 250−500 bp. Longer candidate sections are subdivided into a number of overlapping fragments in this size range and then used in a probe-scanning approach. One of the central lessons from our present study is the proof of this principle.

Perspective. The preliminary sequence of the human genome has been published with an estimation of about 30 000−40 000 genes, about half of which are known. Several attempts have been made to cope with the vast amount of sequence information and to predict genes and regulatory principles. Especially in higher eukaryotes, genes are highly complex, with many alternative splicing sites and large intergenic sequences, calling for entirely novel approaches to delineate the various structural and regulatory levels. Current techniques to search DNA sequences for regulatory elements are mainly based on finding local sequence strings and their combinations, which usually produce lists of possible sites, only a small fraction of which prove to be functional. Our present series of experiments suggests that stress-induced duplex destabilization is tightly associated with both structural and regulatory principles as it reveals the existence on DNA of dominant structural signals that resist a uniform association of nucleosomes. Among these sites are promoters and terminators that may have acquired the additional function of an S/MAR wherever this was compatible with their function. Where this is not the case or where long noncoding regions are to be organized, other elements have been maintained or entered that are S/MARs in the sense of their original definition. Nevertheless, these elements may be specialized as they occupy a separate set of binding sites (Figure 6).

Considering the wealth of biological activities—both constitutive and facultative—that have been associated with S/MARs and insulators, studies of this kind may provide a better insight into the fundamental architecture of eukaryotic genomes and open entirely new avenues for the rational construction of transgenes and vectors (reviewed in ref *40*).

ACKNOWLEDGMENT

We are grateful to our colleagues Volker Kay and Christian Mielke who, in the past, have developed the standardized quantitative methods applied here. We recognize the pivotal contributions by Terumi Kohwi-Shigematsu, who provided early experimental proof for the existence of base-unpairing regions in eukaryotic DNA as part of circular vectors and in the living cell. Fruitful cooperations with the developers of the S/MAR database, MARfinder, and SMARtest have been a most essential platform for the work reported here.

REFERENCES

1. Cramer, T., Kreth, G., Koester, H., Fink, A., R. H., Heintzmann, R., Cremer, M., Solovei, I., Zink, D., and Cremer, C. (2000) *Crit. Rev. Eukaryot. Gene Expr. 12*, 179−212.
2. Berezney, R., and Coffey, D. S. (1974) *Biochem. Biophys. Res. Commun. 60*, 1410−1417.
3. Bode, J., Stengert-Iber, M., Schlake, T., Kay, V., and Dietz-Pfeilstetter, A. (1996) *Crit. Rev. Eukaryot. Gene Expr. 6*, 115−138.
4. Bode, J., Bartsch, J., Boulikas, T., Iber, M., Mielke, C., Schübeler, D., Seibler, J., and Benham, C. (1998) *Gene Ther. Mol. Biol. 1*, 551−880.
5. Gasser, S. M., and Laemmli, U. K. (1987) *Trends Genet. 3*, 16−22.
6. Mirkovitch, J., Mirault, M. E., and Laemmli, U. K. (1984) *Cell 39*, 223−232.
7. Cockerill, P. N., and Garrard, W. T. (1986) *Cell 44*, 273−282.
8. Cockerill, P. N., Yuen, M. H., and Garrard, W. T. (1987) *J. Biol. Chem. 262*, 5394−5397.
9. Kay, V., and Bode, J. (1995) in *Methods in Molecular and Cellular Biology: Methods for studying DNA−protein interactions, an overview* (Papavassiliou, A. G., and King, S. L., Eds.) Wiley-Liss, Inc., New York.
10. Benham, C., Kohwi-Shigematsu, T., and Bode, J. (1997) *J. Mol. Biol. 274*, 181−196.
11. Bode, J., Benham, C., Ernst, E., Knopp, A., Marschalek, R., Strick, R., and Strissel, P. (2000) *J. Cell. Biochem. 35*, 3−22.
12. Antes, J., T., Namciu, J., S., Fournier, K. R. E., and Levy-Wilson, B. (2001) *Biochemistry 23*, 6731−6742.
13. Fernandez, L. A., Winkler, M., and Grosschedl, R. (2001) *Mol. Cell. Biol. 21*, 196−208.
14. Bode, J., Fetzer, C. P., Nehlsen, K., Scinteie, M., Hinrich, Bok-Hee, A. B., Piechazcek, C., Benham, C., and Lipps, H. J. (2001) *Int. J. Gene Ther. Mol. Biol 6*, 33−46.
15. Benham, C. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2999−3003.
16. Kipp, M., Gohring, F., Ostendorp, T., vanDrunen, C. M., vanDriel, R., Przybylski, M., and Fackelmayer, F. O. (2000) *Mol. Cell. Biol. 20*, 7480−7489.
17. Tsutsui, K. (1998) *Gene Ther. Mol. Biol. 1*, 581−590.
18. Okada, S., Tsutsui, K., Tsutsui, K., Seki, S., and Shohmori, T. (1996) *Biochem. Biophys. Res. Commun. 222*, 472−477.
19. Auten, J., Agarwal, M., Chen, J. Y., Sutton, R., and Plavec, I. (1999) *Hum. Gene Ther. 10*, 1389−1399.
20. Dang, Q., Auten, J., and Plavec, I. (2000) *J. Virol. 74*, 2671−2678.
21. Diaz, M. O., Pomykala, H. M., Bohlander, S. K., Maltepe, E., Malik, K., Brownstein, B., and Olopade, O. I. (1994) *Genomics 22*, 540−552.
22. Mielke, C., Christensen, M. O., Westergaard, O., Bode, J., Benham, C. J., and Breindl, M. (2002) *J. Cell. Biochem. 84*, 484−496.
23. Liebich, I., Bode, J., Reuter, I., and Wingender, E. (2002) *Nucleic Acids Res. 15*, 3433−3442.
24. Ganapathy, S., and Singh, G. (2001) *Computational Biology and Genome Information System and Technology*, Duke University Press: Durham, NC, pp 235−239.
25. Liebich, I., Bode, J., Frisch, M., and Wingender, E. (2002) *Nucleic Acids Res. 30*, 372−374.
26. Benham, C. J. (1996) *J. Mol. Biol. 255*, 425−434.
27. Strissel, P. L., Dann, H. A., Pomykala, H. M., Diaz, M. O., Rowley, J. D., and Olopade, O. I. (1998) *Genomics 47*, 217−229.
28. Bode, J., Schlake, T., Ríos-Ramírez, M., Mielke, C., Stengert, M., Kay, V., and Klehr-Wirth, D. (1997) in *Nuclear Matrix* (Berezney, R., Ed.) Academic Press, New York, pp 389−453.
29. Fernandez L. A., Winkler, N. Forrester, W., Jenuwein, T., and Grosschedl, R. (1998) *Cold Spring Harbor Symp. 63*, 515−524.
30. Schlake, T. Klehr-Wirth, D. Yoshida, M., Beppu, T., and Bode, J. (1994) *Biochemistry 33*, 4197−4206.
31. Kirillov, A., Kistler, B., Mostoslavsky, R., Cedar, H., Wirth, T., and Bergman, Y. (1996) *Nat. Genet. 13*, 435−441.

32. Mielke, C., Kohwi, Y., Kohwi-Shigematsu, T., and Bode, J. (1990), *Biochemistry 29*, 7475−7485.
33. Bode, J., Kohwi, Y., Dickinson, L., Joh, R. T., Klehr, D., Mielke, C., and Kohwi-Shigematsu, T. (1992) *Science 255*, 195−197.
34. Bode, J., and Maass, K. (1988) *Biochemistry 27*, 4706−4711.
35. Ullrich, A., Gray, A., Goeddel, D. V., and Tull, T. J. (1982) *J. Mol. Biol. 156*, 467−486.
36. Pederson, T. (2000) *Mol. Biol. Cell 11*, 799−805.
37. Hancock, R. (2000) *Chromosoma 109*, 219−225.
38. Philimonenko, V. V., Flechon, J. E., and Hozak, P. (2001) *Exp. Cell. Res. 264*, 201−210.
39. Ferraro, A., Cervoni, C., Eufemi, M., Altieri, F., and Turano, C. (1996) *J. Cell. Biochem. 62*, 495−505.
40. Bode, J., Götze, S., Ernst, E., Hüsemann, Y., Baer, A., Seibler, J., and Mielke, C. (2002) in *New Comprehensive Biochemistry: Gene Transfer and Expression in Mammalian Cells* (Makrides, S., Ed.) Elsevier, Amsterdam (in press).
41. Bode, J., Benham, C., Knopp, A., and Mielke, C. (2000) *Crit. Rev. Eukaryot. Gene Expr. 10*, 73−90.
42. Schübeler, D. Mielke, C., Maass, K., and Bode, J. (1996) *Biochemistry 35*, 11160−11169.
43. Wang, Z. Y., and Dröge, P. (1996) *EMBO J. 15*, 581−589.
44. Ludérus, M. E. E., De Graaf, A., Mattia, E., Den Blaauwen, J. L., Grande, M. A., De Jong, L., and Van Driel, R. (1992) *Cell 70*, 949−959.
45. Ludérus, M. E. E., Den Blaauwen, J. L., De Smit, O. J. B., Compton, D. A., and Van Driel, R. (1994) *Mol. Cell. Biol. 14*, 6297−6305.
46. Gohring, F., Schwab B. L., Nicotera, P., Leist, M., and Fackelmayer, F. O.(1997) *EMBO J. 16*, 7361−7371.
47. Kay, V., and Bode, J. (1994) *Biochemistry 33*, 367−374.
48. Michalowski, S. M., Allen, G. C., Hall, G. E., Thompson, W. F., and Spiker, S. (1999) *Biochemistry 38*, 12795−12804.
49. Chernov, I. P., Akopov, S. B., Nikolaev, L. G., Sverdlov, E. D., and (2002) *J. Cell. Biochem. 84*, 590−600.
50. Singh, G. B., Kramer, J. A., and Krawetz, S. A.(1997) *Nucleic Acids Res. 25*, 1419−1425.
51. Frisch, M., Frech, K., Klingenhoff, A., Cartharius, K., Liebich, I., and Werner, T. (2002) *Genome Res. 2*, 349−354.
52. Dietz-Pfeilstetter, A. Arndt, N. Kay, V. and Bode, J. (2002) *Transgenic Res.* (in press).
53. Wada-Kiyama, Y., and Kiyama, R. (1996) *Mol. Cell. Biol. 16*, 5664−5673.
54. Trifonov, E. N. (1995) *J. Mol. Evol. 40*, 337−342.
55. Donev, R. M. (2000) *Mol. Cell. Biochem. 214*, 103−110.

BI026496+