

Transcription-promoting genomic sites in mammalia: their elucidation and architectural principles

Jürgen Bode¹, Jörg Bartsch², Teni Boulikas³, Michaela Iber¹, Christian Mielke⁴, Dirk Schübeler¹, Jost Seibler¹, and Craig Benham⁵

¹ GBF, National Research Center for Biotechnology, Genregulation und Differenzierung, D-38124 Braunschweig, Mascheroder Weg 1, Germany.

² Entwicklungsbiologie, Universität Bielefeld, D-33501 Bielefeld, Germany.

³ Institute of Molecular Medical Sciences, 460 Page Mill Road, Palo Alto, California 94306 USA.

⁴ Department of Molecular Biology, University of Aarhus, DK-8000 Aarhus, Denmark.

⁵ The Mount Sinai Med. Center, New York/Biomathematical Sciences New York, 10029 USA.

Corresponding author: Jürgen Bode Tel./Fax: +49 531 6181 251/262, E-mail: jbo@gbf-braunschweig.de

Summary

Scaffold/matrix attached regions (S/MARs) represent a relatively novel addition to the class of cis-acting DNA sequences in the eukaryotic genome. These elements are thought to operate via functional contacts to the protein backbone of the nucleus. S/MARs of several kilobases are found at the putative borders of several chromatin domains, and shorter elements with basically the same physicochemical properties occur in close association with certain enhancers or in introns. Accordingly, S/MARs can be situated either in nontranscribed regions or within transcription units. Biological roles that have been assigned to them include insulating and chromatin domain opening functions. These activities apparently are not separable, and both are compatible with the same kind of structure.

In this contribution we present a series of recent results suggesting that S/MARs act as topological gauges with the potential to adapt their functions to environmental stresses. We also suggest that previously noted uncertainties regarding their activities may have arisen from inadequacies in the methods that were used for their characterization. We discuss the application of new, highly controlled site specific recombination methodologies that integrate single copies into controlled genomic positions to the study of transgene and S/MAR functions in cell cultures and in transgenic organisms.

I. Organization of the eukaryotic genome

The eukaryotic genome is organized on at least four levels. At the lowest level the double-stranded DNA molecule combines with octamers of core histones, wrapping around each in two left-handed superhelical turns. This produces a string of nucleosomes whose spacing is largely determined by the presence of a linker histone.

Since these basic features emerged, evidence has accumulated that the orderly arrangement of nucleosomes can be affected by transacting factors and structural features of the DNA. In particular, DNA sequences with an intrinsic curvature or bendability prefer to be accommodated within a nucleosome, causing phased arrays even in the absence of additional proteins (reviewed by Wolffe, 1994b).

The nucleosome string, also called the 10 nm filament, shows a propensity to fold into a fiber with a diameter of 30 nm. This fiber in turn is organized into looped domains. Early evidence suggesting this domain structure included the observation that neither micrococcal nuclease nor restriction enzymes are able to release from nuclei soluble chromatin with a DNA chain length in excess of 75000 base pairs (Igó-Kemenes and Zachau, 1977). Around the same time the existence of topologically independent domains was established by microscopic studies of histone-depleted metaphase chromosomes (Paulson and Laemmli, 1977) and nuclei (Cook and Brazell, 1978). These studies revealed the presence of a supporting structure, the nuclear scaffold or matrix, to which DNA was periodically attached to form superhelical loops. Experiments with intercalating agents, which at low concentrations cause an expansion and at higher concentrations a contraction of the halo, were explained by dye-induced relaxation of negative superhelical loops and to their subsequent overwinding. These loops evidently were held in a way that constrained their topologies by preventing changes in their linking numbers. Attached regions of DNA were subsequently characterized by several extraction procedures (Mirkovitch et al., 1984; Cockerill and Garrard, 1986; review: Boulikas, 1995) and accordingly they were either termed scaffold- or matrix-attached regions (S/MARs). Since the same elements are recovered by various protocols, the original distinction of SAR- and MAR-elements seems no longer justified (Kay and Bode, 1994).

Besides the common extraction approach there are other, supposedly milder methods aimed at detecting attachment sequences. However, these mostly fail to establish the existence of functional S/MARs (Jackson et al., 1990; Eggert and Jack, 1991; Hempel and Strätling, 1996). A critical evaluation of these experiments shows that the S/MAR elements either were not probed in their genuine transcriptional context and/or that the topological state of their domains had been perturbed by restriction (Bode et al., 1996). A careful study by Ferraro et al. (1995, 1996) used cis-diamminedichloroplatinum to form reversible crosslinks between matrix proteins and DNA in intact cells. The use of authentic S/MAR probes for Southwestern blotting strongly suggested that the separated matrix proteins in fact are the interacting partners of the S/MARs.

A. Chromatin domains and boundary elements

Genes which are committed to transcription are generally accessible to the action of DNaseI (Weisbrod et al., 1982). An elevated sensitivity has been demonstrated to extend several kbp from the transcribed region until an area of lower accessibility is reached. It is tempting to

speculate that the boundaries between these regions could be formed by S/MARs, which would prevent the topological changes within an active domain from propagating into quiescent ones (Bode et al., 1992). Examples for which this situation has been documented include the domains of the chicken lysozyme gene (Phi-Van and Strätling, 1988), the human apolipoprotein gene (Levy-Wilson and Fortier, 1989), the human β -globin cluster (Dillon and Grosveld, 1993), and the human interferon- β gene (Bode et al., 1995). These findings led to the idea that the DNA loops defined above represent functional units within the genome, so-called chromatin domains.

The group of DNA "boundary elements" that have been implicated in the functional compartmentalization of the eukaryotic genome share certain common attributes. One defining property is insulation: a boundary element placed between two cis-acting elements inhibits their interactions. When a promoter is separated from an enhancer in this way, for example, the enhancer is no longer able to interact with the transcription initiation complex at the promoter (review: Corces et al., 1995). Early work found certain sequences that exhibited insulation, although they did not appear to be S/MARs. Examples are the *scs* and *scs'* sequences flanking the *Drosophila* heat shock locus (Kellum and Schedl, 1992, Vazquez et al., 1994). Although *scs* and *scs'* did not behave as S/MARs, at least in the initial assay, they have a number of properties in common with them. Each contains a large, nuclease-resistant core spanning a DNA segment that is very AT rich and flanked by DNaseI hypersensitive sites. After heat shock, both elements are primary targets for the action of topoisomerase II, which is an abundant S/MAR-associated protein (Laemmli et al., 1992). Other examples are a sequence within the gypsy transposon of *Drosophila* (Wolffe, 1994a), and a flanking element in the β -globin locus of chicken (Chung et al., 1993). The latter element coincides with a constitutive hypersensitive site (HS4), and blocks the action of enhancers in a way resembling *scs* and *scs'*. Although this GC-rich sequence is similar in many aspects to the HS5-associated sequence in the human β -globin locus, which is a S/MAR (Li and Stamatoyannopoulos, 1994), it has no S/MAR activity *in vitro*. If it were matrix-attached *in vivo*, it would have to be by a different, but possibly related, mechanism (see chapter V).

B. Structural factors affecting transgene expression levels

Several factors conspire to make the expression levels of transgenes highly unpredictable when transfections are performed using conventional techniques. The two most

important of these are positional effects and copy number effects.

Position effect variegation (PEV) can be defined as a position-dependent inactivation of gene expression in a fraction of cells that generate a particular tissue. The first and best documented instance of PEV is a chromosomal rearrangement in *Drosophila* in which an allele of the white (w^+) gene is transferred to a site close to the centromere. After this translocation its previously uniform expression becomes "variegated," producing patches of pigmented and unpigmented cells in the eye. It is thought, but still unproven, that a pericentromeric location renders the gene susceptible to the spreading of heterochromatic condensation. PEV has also been demonstrated in yeast and in mammals for gene sequences within centromeres or close to telomeres (Dobie et al., 1997).

Variegated expression can occur in transgenes as well as in endogenous genes. As the expression level of a transgene is highly dependent on its integration site, which cannot be predetermined with conventional transfection techniques, the forces leading to PEV can cause large and unpredictable variations in expression levels. In the case of mice, it has been reported that transgene integration into pericentromeric regions is the most frequent inactivating process (Festenstein et al., 1996).

However, in some remarkable cases transgenes are expressed in an integration-site independent manner (Chamberlain et al., 1991; Greaves et al., 1989; Aronow et al., 1992; Palmiter and al., 1993; Schedl et al., 1993; Thorey et al., 1993; Neznanov et al., 1996). Attempts to identify the sequences responsible for this insulating effect have not yet led to unambiguous candidate elements. One simple explanation might be that these constructs are delimited by boundary structures (Eissenberg and Elgin, 1991). Alternatively or additionally, they may contain elements that prevent mislocalization into heterochromatic nuclear compartments. The latter class of elements has originally been termed "dominant control regions" (DCR; Collis et al., 1990) and now, more commonly, "locus control regions" (LCR; Epner et al., 1992). They are thought to form extraordinarily stable complexes with their coordinated promoters in a way that overcomes external influences.

In vertebrate transfection experiments the transgenes frequently insert in large tandem arrays. In these arrays expression levels are not strictly correlated with copy number, the extreme case being where expression is completely absent. This phenomenon is termed "repeat-induced gene silencing" (RIGS) in animals (Dobie et al., 1997), and "cosuppression" in plant systems (Matzke and Matzke, 1995). It is presumed that the close repetition of sequences leads to the formation of unproductive multiprotein complexes between transcription factors

and/or to the sequestering of these complexes in a heterochromatic nuclear compartment.

It is obvious that studies of PEV promise insights into the basis for heterochromatin formation and the role of higher order chromatin and chromosome structure in gene regulation. On the other hand, positional and copy number effects on transgene expression levels are serious obstacles to the straightforward application of reverse genetics. The various ways in which transgene expression patterns are affected by multiple copy integration events and inadvertent occupation of certain integration sites will be discussed further below (chapter III). Chapter IV presents a proposed transfection strategy that does not have these problems.

C. Locus control elements

A prototype LCR has been defined upstream from the human β -globin genes. It contains five DNaseI hypersensitive sites, each of which is a small region of 200-300 bp containing a high density of transcription factor binding sites. This LCR is absolutely required for expression, and it confers an altered chromatin structure on a region of more than 150 kbp. (Dillon and Grosfeld, 1993). The existence of an LCR has also been demonstrated at the human CD2 locus (Festenstein et al., 1996) and in the chicken lysozyme gene domain. In the latter case a group of proximal regulatory sites (between -1 and -3 kb) and two distal sites (at -6.1 and -7.9 kb) are all required for high-level, position-independent expression. It follows that the collection of these separated functions together constitutes the LCR (Sippel et al. 1993, Bonifer et al., 1994, 1996).

Many transgenic studies involving either a complete LCR or its core sequences have relied on the analysis of cells possessing more than one copy. More refined studies are possible using retroviral vectors, which enable a single copy of the transcription unit per cell to be integrated in a precise way and in the absence of selection. (This approach will be described in detail below). In retroviral transfection experiments performed to date, the LCR and its components act more like a classical enhancer than as an element dominating chromatin structure. These unexpected observations raise the question of whether an LCR can truly confer position-independent expression when present in one copy per cell (Novak et al. 1990).

II. The elusive roles of scaffold/matrix attached regions

S/MARs have been observed to have both boundary functions and to act upon transcriptional rates. Due to their locations at the putative ends of chromatin domains, they have been considered as domain borders (Phi-Van and Strätling, 1988; Bode and Maass, 1988; Levy-Wilson and

Fortier, 1989; Dillon and Grosveld, 1993). In some systems they were found to dampen positional effects, as expected of insulating boundary elements (Stief et al., 1989; Phi-Van et al., 1990, 1996; Kalos and Fournier, 1995). Although S/MARs are clearly distinguishable from enhancers, they augment transcriptional levels by a distinct, enhancer-independent mechanism (Klehr et al., 1991, 1992; Dietz et al., 1994; Poljak et al., 1994). Conversely, in some systems enhancers are fully active only when associated with S/MAR elements (reviews: Bode et al., 1995, 1996, and section V).

A. An operational definition

Originally, S/MARs were defined using only the protocols that led to their detection. These procedures involve the isolation of interphase nuclei, followed by extraction of non-matrix proteins to yield a nuclear halo. It has been shown that the total number of loops per cell can depend on details of the procedure used. For every attachment existing *in vivo*, several new attachments may be created *in vitro* as nuclei are prepared, stabilized and lysed (Jackson et al., 1990). Moreover, the matrix-S/MAR contacts of degraded, topologically unrestrained halos are accessible to competing, soluble S/MAR elements (Mielke et al., 1990; Kay and Bode, 1994, 1995; Bode et al., 1995 and references therein). These observations suggest that not all S/MARs are constitutively associated with the nuclear matrix *in vivo* (Dillon and Grosveld, 1993). Techniques able to assess the occupation of S/MARs *in vivo* are currently emerging (Ferraro et al. 1995, 1996).

Although the central biological effects of S/MAR elements are clearly compatible with their affinity for the nuclear matrix, others have been harder to explain. Below we describe the more prominent properties of these elements. These diverse properties are reconciled by recent experiments that have been performed in our laboratory (section IVA). In Section V we will discuss novel transfection strategies using S/MAR elements. These have the capacity to avoid the uncertainties arising from positional and copy number effects, providing a comprehensive and unambiguous analysis of the associated functional aspects of transfected elements, including S/MARs.

B. Transcriptional augmentation

S/MARs are a relatively new addition to the list of cis-acting elements known to elevate transcriptional rates. In our experience the simultaneous presence of an enhancer is not required for this S/MAR activity (Klehr et al., 1991). This effect, which we call 'transcriptional augmentation', is clearly separable from prototypical enhancement since enhancers, but not S/MARs, are active in transient assays.

But like enhancers, S/MARs act independent of orientation and independent of distance, provided this is at least several kilobases (Klehr et al., 1991 and section IVA). This activity is found for minimal, viral and cellular promoters, of both inducible and housekeeping genes. If an enhancer is also present, enhancement and augmentation factors act roughly multiplicatively (Bode et al., 1995).

Although enhancers are generally believed to interact with the basal transcription machinery by looping, the experimental data supporting such a mechanism are not conclusive. To explain the correct choice of promoter, it has been suggested that the initial contacts must be checked by a tracking mechanism (Weintraub, 1993). Both looping and tracking may be modulated by the occurrence of S/MARs, although there are divergent views of how this is accomplished. On one hand, S/MARs may associate nearby enhancers to the nuclear matrix, and thereby assure their proximity to transcribed units (Boulikas, 1995). Such a mechanism could enable the formation of alternative functional units, activating their respective promoters at the matrix by a directional transfer of transcription factors. On the other hand, by acting as domain boundaries (IA), S/MARs could limit the effect of an enhancer to the domain in which it occurs. This view is compatible with the results from enhancer-blocking experiments (Bode et al., 1995; Li and Stamatoyannopoulos, 1994). Finally, S/MARs might serve as domain openers, as first proposed by Zhao et al. (1993).

Although a direct correlation has been demonstrated between the binding and augmenting activities of S/MARs (Mielke et al., 1990; Kay and Bode, 1995; Allen et al. 1996), it is not clear whether nuclear matrix binding *per se* is sufficient to augment transgene expression in stably transfected cells (Phi-Van and Strätling 1996). If this were the case one could use a simple *in vitro* binding test to predict *in vivo* properties. Very recent studies show that the general augmenting effect of S/MARs can be disrupted by the overexpression of certain S/MAR binding proteins (Kohwi-Shigematsu et al, 1997). A simple explanation for this phenomenon would be the interference of such a (possibly soluble) factor with genuine S/MAR-matrix contacts, but more sophisticated explanations also are possible (see the model by Scheuermann and Chen, 1989; Zong and Scheuermann, 1995).

The biological effects of S/MAR elements commonly are studied by transfecting a S/MAR-reporter construct and relating its expression level to that of a S/MAR-free control that has been transfected in parallel. In animal cells, such an approach consistently leads to a significant elevation of the reporter signal. Although this usually is intuitively ascribed to cis-actions of the S/MAR, such an effect could in principle arise in several ways:

(i) as an immediate effect of the S/MAR on transcriptional initiation. This is the proposed explanation for almost every transcriptional S/MAR effect published to date.

(ii) as a copy number effect. Many (but not all) S/MAR elements tend to increase the number of integrated copies relative to S/MAR free control (Schlake, 1994). This has been ascribed to the recombinogenic nature of these sequences (Sperry et al., 1989; Bode et al., 1995, 1996). It should be noted that, due to PEV-related phenomena (Section IB), the activity of a single copy can normally not be derived by just dividing the overall expression by the number of integrated copies.

(iii) as a targeting effect. Owing to the affinity of S/MARs for the nuclear matrix, which is thought to be the site of active transcription, a preferential integration of S/MAR constructs into actively transcribed regions or active nuclear compartments could occur. S/MAR free controls would not exhibit this preference.

C. Insulator functions

The presence of S/MARs has been shown to be required to prevent the ectopic expression of transgenes (Sippel et al., 1993). In one case a S/MAR is needed to enable correct, hormonal gene regulation (McKnight et al., 1992, 1996). These and other results from a range of insulation experiments strongly implicate S/MARs in defining the boundaries of autonomously regulated chromatin domains.

S/MARs have been studied in both animal and plant systems. As a rule, transcriptional augmentation by S/MARs is most easily seen in mammalian cells, while in plants their insulator functions are more frequently observed (Bode et al., 1995). We ascribe this difference to the fact that most plant studies are based on T-DNA vectors which, when integrated into genomic loci, yield expression levels high enough to cope with applied selection pressure (Dietz et al., 1994). In this case the maximum attainable level of gene expression does not depend on the presence of S/MARs, whose most easily observed function then becomes to stabilize an elevated (but not extreme) expression level shielded from negative and positive influences of the surroundings (Dietz et al., 1996, Bode et al., 1996). This would agree with the insulating or bordering function suggested by Mlynarova et al. (1995).

After the enhancer-blocking assay (IA), described previously, copy number dependence is by far the most common principle used to detect and assay insulator functions. However, there are major difficulties inherent in this method (demonstrated by Poljak et al., 1994; Schlake, 1994). These are due, at least in part, to the fact that

multiple copies often do not integrate at separate genomic locations, but rather as a single tandem array at a site which is unique for each cell. This induces many unexpected interactions, rearrangements and cellular 'defense mechanisms' (Mehtali et al., 1990; Krickler et al., 1992; Dorer and Henikoff, 1994; Kalos and Fournier, 1995; Dorer, 1997). For this reason there frequently is not a linear correlation between copy number and gene transcription, even in the case of low copy numbers (see IB and IIIC).

D. Intronic S/MARs

S/MARs cannot simply be considered to represent static delimiters of functional domains. This is demonstrated by the detection of S/MARs within introns. Examples of single genes that are apparently divided over two domains include the genomic sequences encoding hamster DHFR (Käs and Chasin, 1987), human topo I (Romig et al., 1992), human interleukin-2 (Artelt and Bode, unpublished) the mouse immunoglobulin γ - and μ -chains (Cockerill and Garrard, 1986, Cockerill et al., 1987) and a light-inducible plant gene (Stockhaus et al., 1987, Mielke et al., 1990). Studies on MPC-11 plasmacytoma cells have shown that 9% of poly(A) mRNA arises from the κ -locus, whose transcription must be completed on average once every 3.2 s (Cockerill and Garrard, 1986).

Since by definition intronic S/MARs are transcribed, and since they do not impede passage of RNA polymerase II, their occupation must be regulated. Functional analyses have recently revealed complex roles for S/MARs in gene expression, which can only be appreciated after structural analyses of these composite elements (section VB).

E. Current uses of S/MAR elements

Retroviral delivery systems enable the efficient transfer and expression of transgenes in primary cells. Sometimes, a pitfall of these methods has been a continuous downregulation of the transgene (but see IVA). Because S/MAR elements augment transcription and stabilize its level over extended periods of time (Bode et al., 1995), they could be used to construct a new generation of expression vectors. The mechanism of retroviral replication enables one to establish a minidomain by simply introducing a S/MAR element into the 3'-LTR (Schübeler et al., 1996 and **Figure 3**). In this minidomain the expression unit is flanked by two S/MARs, which can stabilize transgene expression. A number of current pharmaceutical developments are exploiting this approach.

Although short and apparently functional S/MAR elements can be constructed by oligomerizing certain sub-S/MAR motifs (core unwinding elements, see Mielke et

al., 1990, Bode et al., 1992 and VB), the practical use of such synthetic elements for retroviral transfection is limited by the intrinsic genomic instability within and between the multimers. One can avoid this problem by using minidomain bordering elements from different sources (Bode et al., 1995, Dietz et al., 1994).

A whey acidic protein (WAP) transgene was found to be active in just 1 out of 17 lines of transgenic mice, demonstrating copy-number independent (position dependent) expression. After ligating S/MARs to the transgene, 7 of 9 lines exhibited transgene activity, and correct hormonal regulation occurred in the majority of cases (McKnight et al., 1992, 1996). These experiments show that S/MARs are able to prevent ectopic expression (Sippel et al., 1993).

S/MARs have found widespread use in the manipulation of plant cells, callus cultures and complete plants. Although increased and position-independent transcription has been described by various authors, these phenomena are only linked in some cases (Schöffl et al., 1993, Mlynarova et al., 1994, van der Geest et al., 1994). Examples are known of plant cells transformed by *Agrobacterium* in which the use of S/MARs improves copy-number dependent (Breyne et al., 1992) or position-independent (Dietz et al., 1996) expression without affecting transcriptional levels (Breyne et al., 1992). In other cases where microprojectile bombardment was used, transcription was dramatically raised independent of copy number (Allen et al., 1993, 1996). These differences may, in part, be due to the gene transfer technique used. The first method tends to target transcriptionally competent sites, while random positions are hit in the second one (Dietz et al., 1994).

III. Problems associated with reverse genetics

The level of *in vivo* expression of a transfected gene may be quite different from that which occurs when the same gene is in its natural context. Although transient assays in cultured cells have been used in the past to characterize tissue-specific enhancers, their potential is limited because the actions of these and other cis-acting elements depend sensitively on chromosomal context. Proximity of enhancers to S/MAR elements provides a paradigm example of context-specific behavior (reviews: Bode et al., 1995, 1996). These observations underscore the difficulty of establishing a dependable approach to the study of gene function. The analysis of gene expression must be performed in a native chromatin context. However, this approach is complicated by a number of problems arising from the following facts (section IIIA-C).

A. Need for a selection marker

Since only a small percentage of cells incorporate foreign DNA after transfection, selection for a drug resistance marker is required to isolate those cells that harbor the construct. This marker has to be expressed at a threshold level, which creates a selection bias against low producers (Blasquez et al., 1992). Positive selection markers are constitutively expressed, which can seriously interfere with the transcription even of remote genes. An extreme example has been reported of a gene that became completely deregulated by a selection marker which was placed 50 kbp away (Fiering et al., 1993). Further, some prokaryotic vector sequences have been found to inhibit (Palmiter and Brinster, 1986) or to promote gene expression (Seibler and Bode, 1997).

B. Integration occurs at random

Since there is no evidence for site-directed insertion, the generation of random chromosomal breaks is thought to be the rate-limiting step of the prevailing integration mechanism. This would explain the predominance of unique integration sites in mammalian cells and the observation that certain (but not all; see Mielke et al., 1990) cell types prefer linearized over circular templates for integration (Palmiter and Brinster, 1986). Random integration can subject transgene expression to unwanted local influences (position effects) resulting from their proximity to regulatory elements or heterochromatic regions (IB).

C. Multiple copy integration

If DNA is introduced by standard transfection techniques or by microinjection, multiple gene copies are usually integrated at a single site. This is probably due to homologous recombination events occurring among the transfected molecules (Phi-Van and Strätling, 1996). There are several ways in which this tandem arrangement can alter the effect of an element from what it would be as a single copy. One usually cannot determine how many copies of a gene are functional templates, or whether different copies perform different functions (Oancea et al., 1997). Tandem repetition of promoter elements could trigger the formation of multiprotein complexes between transcription factors, with largely unpredictable consequences (IB). It also may lead to a more than proportional effect of a cis-acting regulatory sequence, such as a S/MAR (Stief et al., 1989). More typical are shutoff processes occurring with time. These have been associated with methylation followed by mutagenesis (Mehtali et al., 1990; Dorer, 1997) and/or a genomic instability resulting from the continuous loss of members of the array (Palmiter et al., 1982; Weidle et al., 1988). As a

consequence, expression levels cannot be expected to be proportional to copy number (**Figure 1**).

**Figure 1.
Inadequacy of
conventional gene
transfer techniques.**

Conventional transfection techniques lead to multi-copy integration events, usually at a single site and in tandem, head-to-tail orientation. Whether all members of a multigene complex are transcribed at the same rate is an unsolved question

While a tandem head-to-tail integration at a given site is considered typical, more complicated integration forms were recently demonstrated for single S/MAR constructs (Phi-Van and Strätling, 1996). Most of the copies were colocalized as usual, but they clearly differed regarding their relative orientation: whereas head-to-head (hh) and head-to-tail (ht) were preferred, hh *plus* tt and ht *plus* tt tandems were strongly disfavored and exclusive tt integrations were lacking altogether. It may be speculated in this case that the unequal ends, which contained a S/MAR at their 5' end and a unique vector sequence at their 3' terminus, were responsible for this effect. This nonrandom distribution could result from preintegration ligations of two molecules. Alternatively, the close juxtaposition of two identical copies could induce inversions. The recombinogenic nature of S/MAR elements (Sperry et al.

1989) might add to the complications associated with conventional gene transfer techniques.

In conclusion, with current transformation methods, insertion of DNA into the genomes occurs at random, and in the case of plant systems in many instances at multiple sites. The associated position effects, copy number differences and multigene interactions can make conventional gene expression experiments difficult to interpret.

1. An immediate solution: gene transfer by electroporation can be optimized

Classical gene transfer experiments are based on the cotransfer of a reporter and a selector construct which are coprecipitated by Ca⁺⁺-phosphate and taken up by endocytosis (transfection). As an alternative, DNA can be

transferred from solution by transiently permeabilizing the cells in an electric field (electroporation). This technique is most efficient if the reporter and the selector gene are physically linked in a single construct. For some cell lines the electrical parameters and cell survival rates can be optimized in a way that yields predominantly single copy integration events (Mielke et al., 1990, 1996; J. Bartsch, unpublished). This technique eliminates the unpredictable effects of multicopy integration.

This method has been used to compare expression levels of S/MAR-constructs to those of S/MAR-free controls. Representative data are found in **Figure 2**. The average level of expression from a S/MAR-free luciferase control (Lu: 85 700 light units) was increased eightfold if an upstream S/MAR element was present (E-Lu: 698 000 light units), and 26-fold if it was transfected as a minidomain with upstream and downstream S/MAR elements attached (E-Lu-W: 2 253 000 light units). These results are similar to those previously reported (Klehr et al., 1991, 1992; Bode et al., 1995, 1996; Phi-Van et al., 1990; Poljak et al., 1994).

This analysis has been extended to individual clones to derive information about the insulator function of S/MARs that is difficult to obtain by either the enhancer-blocking assay (IA) or the copy number dependence of expression (section IIC). Since expression levels vary by three orders of magnitude, they have been plotted in **Figure 2** on a logarithmic scale. The largest inter-clone variation of

expression was 340 fold, found for the S/MAR-free control. A single S/MAR element reduces the variation to 4.6 fold. This E-Lu construct has a single insulator at its 5' end and could, in principle, experience position effects acting through its "open" 3' terminus. If the domain is closed by a second S/MAR element (E-Lu-W) the variation between clones is reduced to a factor of two. These data are in general agreement with a recent study by Kalos and Fournier (1995), who demonstrated that in the presence of both apolipoprotein B and S/MARs the expression of a test construct was increased from the detection limit to at least 200 fold higher levels. At the same time the variation between the clones decreased from >10-fold to <3-fold. For technical reasons these studies were restricted to a small number of clones, and apparently they have not been generally accepted (see Wang et al., 1996).

We believe that the combined results of our studies strongly support the notion that S/MARs can enable position independent expression of transfected genes. Whether this is due to insulator function, as currently believed, is not clear. Ours is the first study to demonstrate the dramatic influence even of a single S/MAR, which would also be compatible with a dominant (domain opener) effect in case of the E-Lu construct. This information cannot be derived from any assay based on the presence of several copies (**Figure 1**), since in the common case of tandem integration each individual member but one will be an effective double-S/MAR construct (...E-Lu-E-Lu...).

Figure 2. Insulator function of S/MARs demonstrated at the single copy level

A luciferase-neomycin resistance construct was introduced into CHO cells either as a S/MAR-free control (Lu), flanked by an upstream S/MAR element (E-Lu) or by a minidomain flanked by an upstream and a downstream S/MAR (E-Lu-W). Electroporation of these constructs resulted in 10 clones (Lu), 20 clones (E-Lu) and 90 clones (E-Lu-W), resp.. 9-10 clones with a single integrated copy of each construct (solid bars) and 2-3 clones with 2-3 copies (hatched bars) were selected for determinations of the luciferase expression level. Average luciferase levels (light units) were $86 \text{ E}3 \pm 124\%$

(Lu), 590E3 +/-39% (E-Lu), 2273E3+/-14% (E-Lu-W).

Figure 3. Use of retroviral vectors

Because those retroviral genes that are necessary for retrovirus replication have been replaced by a reporter gene (secretory alkaline phosphatase, SEAP) and a selection marker (fusion gene from Hygromycin-B-phosphotransferase and HSV-Thymidine kinase, HygTK), these functions have to be provided *in trans* by a packaging cell line. These helper cells produce infectious retroviral particles which can be used to transfer the construct into the recipient cell. After reverse transcription of viral RNA, the transgene is stably linked with the genome (provirus state). Promoter and terminator functions reside in the long terminal repeats (LTRs). Moderate manipulation of the LTRs is compatible with their function and the introduction of S/MAR elements (Mielke et al., 1996) and Flp-recognition target sites (FRTs) has been described (Schübeler and Bode, 1997; see also section IVA).

Interestingly, the properties documented in our study appear to be independent of the source of the S/MAR. Elements of human origin (E) behave in our system like their counterparts from plants (W), provided that they both show a corresponding affinity for the scaffold in the *in vitro* test system (cf. also Dietz et al., 1994). To date it has not been possible to separate the insulation and augmentation properties of S/MARs (Phi-Van and Strätling, 1996) which both require a threshold length of DNA in excess of about 300 bp. Together, these results suggest that the basic functions of insulation and transcriptional augmentation are widely conserved, and are based on structural features which are recognized by ubiquitous proteins rather than by specialized factors.

We note that five clones in our experimental series contained more than a single copy (hatched bars in **Figure 2**). Since their expression is in the same range found for the single copy integrates, a standard copy-number dependence test (IIC) would have failed to reveal any evidence for an insulator function in this system. This re-emphasizes the fact that any test of sophisticated S/MAR functions must use single copy integrants.

IV. Advanced approaches: A progress report

While the electroporation method has effectively eliminated the detrimental effects of multicopy

integrations, the other critical problems (numbered (i) to (iii) in section IIB) still need to be addressed. The following studies describe recent developments in this direction. They allow the integration of single, intact copies, in some cases without the cotransfer of vector sequences. They also are aimed at preventing any integration bias which may arise from the biochemical properties of S/MAR elements (criterion 3 in IIB).

A. What retroviruses can tell us

We used a retroviral vector infection procedure as it facilitates the introduction of single copies with strictly defined ends, the long terminal repeats (LTRs). For these studies we have constructed an expression cassette consisting of a reporter gene (SEAP, secretory alkaline phosphatase) and a selector gene (PAC, puromycin N-acetyltransferase). Both members of this bi-cistronic cassette are synchronously expressed owing to the fact that an internal ribosome entry site (IRES) enables the translation of the second cistron. To obtain an infectious retrovirus particle, accessory functions have to be provided *in trans*. This is done by first transfecting the construct into a helper cell line (Psi2). The helper cells contribute the gag, env and pol functions, which package the viral mRNA and secrete the virus into the medium. These viruses can be harvested and used for an infection of the ultimate target cells. There copies of the viral mRNA are reverse-transcribed into DNA. During this process the

sequence of the 3'-LTR is copied to form an identical 5'-LTR (**Figure 3**). Elements cloned into the 3'-LTR (marked by a half arrow) will then appear also in the 5'-terminus. This strategy has been used both to introduce a minidomain (S/MARs at both ends, see section IIE), and to generate two Flp recombinase target sites which are recognized and recombined by the Flp enzyme (Schübeler et al, 1997 and section IVB).

Before making extensive use of these possibilities, we decided to thoroughly characterize the sites of provirus integration. As an initial step, the identical construct was introduced into target cell by standard transfection, electroporation and infection techniques. Expression levels of the SEAP reporter were compared and related to the number of integrated copies (**Figure 4**). While the overall expression level turned out to be mostly independent of the procedure used for transferring the vector, the per-copy level for transfection is clearly inferior to electroporation and electroporation in turn is inferior to infection. Only infection guarantees the incorporation of single copies which remain stable over extended time intervals (Schübeler et al, 1997 and unpublished). This indicates that there is no gradual inactivation due to methylation. If such an event were to occur, selection pressure could be applied to eliminate it.

1. Anatomy of retroviral integration sites

It is the prevailing view that transcriptionally active genomic regions and regions associated with DNase I hypersensitive sites are preferred by the integration machinery of a retrovirus (Rohdewohld et al., 1987). We wanted to study the architecture of these integration sites in order to obtain information about the factors which mediate a consistently high and stable level of expression.

Until recently, most methods for isolating proviral flanking sequences involved plasmid rescue. This approach led to the identification of a small number of highly preferred integration targets (Shih et al., 1988). Some of these results could not be confirmed by subsequent PCR-based techniques that were developed to study integration without prior selection by molecular cloning. This has led to a critical re-evaluation of the prevailing views (Withers-Ward et al., 1994). We have applied retroviral vectors in conjunction with inverse PCR techniques to reconstruct a number of these sites for further characterization. As in many previous studies, the recovered integration sites conformed to no obvious consensus sequence. This suggests that the site of integration into the host genome may be determined by other factors, such as DNA secondary structure and/or host proteins. While the retroviral integrase performs the central cutting and joining steps as part of a 160 S nucleoprotein complex, the final resolution steps require host functions which may lead to a further selection among a number of initial target sites.

Figure 4. Transcriptional properties of transgenes introduced by three different gene transfer techniques.

A single construct (pM5sepa) was introduced into NIH3T3 murine embryo fibroblasts by three gene transfer techniques. pM5sepa contains a reporter (SEAP, secretory alkaline phosphatase) and a selector gene (PAC, puromycin N-acetyltransferase) linked by an internal ribosomal binding site (IRES) to enable coupled expression of both cistrons.

Solid bars represent total expression from an entire pool of PAC-resistant clones. Hatched bars refer to expression levels divided by copy numbers.

.... S/MAR-type DNA:

- recombinogenic
- supporting transcription

.... bent DNA

- no positioned nucleosomes (Boulikas, unpublished)
- backbone strain retrievable for unpairing (Ramstein & Lavery, 1988)

Figure 5. Architecture of highly expressing genomic sites which are the preferred targets for retroviral integration. Integration occurs into S/MAR type DNA which is flanked by bent segments. Bending prevents the association of nucleosomes as to keep these sites accessible. It is therefore different from the type of curvature which accommodates nucleosomes (Mielke et al., 1996).

Remarkably, all investigated examples conformed to a unique pattern (Mielke et al., 1996), summarized in **Figure 5**. All integration events occurred into scaffold/matrix attached regions (S/MAR-elements) which were flanked by DNA that was bent in a way that discourages nucleosome assembly (Boulikas et al., unpublished). These S/MARs belong to a novel class that does not conform to the AT-rich prototype. On the other hand they exhibit most of the *in vitro* and *in vivo* properties of S/MARs. Their binding is competed by ssDNA, and they induce significant transcriptional augmentation (as opposed to enhancement) if they are investigated as parts of mammalian expression vectors. We have suggested that the entire insertion process might be guided by the nuclear matrix, which also could provides the enzymatic functions for the final steps of the integration process (Mielke et al., 1996). Integration into S/MARs may also be ascribed to their recombinogenic properties while the selection of non AT-rich subtypes seems to be directed by the mechanism of integration. In summary, retrovirus integration occurs at sites which are available due to their lack of nucleosomes, and which show structural and functional features reminiscent of S/MARs (see analyses under VB/C).

2. S/MAR effects on transcriptional initiation and elongation

At this stage of our investigation it was unclear if retroviral integration sites were appropriate for the study of

S/MAR constructs, as S/MAR functions also resided in their immediate vicinity. But these tools permit integration into a subclass of genomic sites whose selection is dictated by the retroviral integration machinery, not by the presence or absence of the S/MARs within the vectors (criterion 3 under IIB).

We inserted an 800 bp S/MAR element from the upstream border of the human interferon- β domain into the vector shown in **Figure 4**. This element was cloned into various positions, both within and outside a transcribed region of 4.3 kb. Insertion into the 3'-LTR yielded a minidomain (two flanking S/MAR elements) according to the scheme in **Figure 3**. This study revealed a range of unexpected S/MAR effects that were obscured when the same constructs were introduced by transfection (Schübeler et al., 1996). The most striking observation was that at a distance of about 4 kb, the S/MAR supported transcriptional initiation whereas at distances below 2.5 kb transcription was essentially shut off. Controls proved the functionality of all constructs in the transient expression phase, and ruled out any influence of S/MAR position on transcript stability. Moreover, no pausing or premature termination was observed within these elements.

One interpretation of these results is presented in **Figure 6**. This assumes that S/MARs are kept in a single-stranded state by association with ssDNA binding proteins. This "unwound" structure is able to facilitate the progression of an approaching polymerase if the buildup of positive supercoils is sufficient for breaking the contacts

between the single-strands and the binding proteins (transition D2 → D2.1). If positive superhelicity is minor,

Figure 6. Scaffold/Matrix-attached regions act upon transcription in a context-dependent manner.

S/MARs can be trapped *in vivo* as a single-stranded structure by chloroacetaldehyde (CAA). It is hypothesized that single strands are kept apart by ssDNA binding proteins (triangles). Unconstrained positive supercoils arising from a S/MAR-scaffold association are ultimately removed by the action of topoisomerases (B, C). If the S/MAR is located immediately downstream from a transcription initiation site, the polymerase cannot pass the attachment point as the buildup of positive superhelicity is insufficient to rupture S/MAR-scaffold contacts (D1). If it is situated further upstream the contacts can be broken by the approaching polymerase and positive superhelicity can be relaxed as it is compensated by the now unconstrained unwound DNA structure of the S/MAR segment (complementarity of plectonemic and paranemic structures, see Yagil, 1991). Other polymerases can initiate in the wake of the first one.

binding of the ssDNA is strong enough to prevent a rotation of the DNA helix about its axis, which inhibits the progress of the polymerase (state D1). This model explains the bordering and insulator functions of a S/MAR, which will prevail if the element is long enough. It also explains the observation that, in principle, a polymerase can progress through a short S/MAR, and even benefit from the fact that S/MARs are repositories of unwound DNA. This may explain the essential fact that transcriptional augmentation is only found after anchoring the construct in the genome of the host cell. It is unnecessary to postulate the existence of different S/MAR types to perform bordering and augmenting functions. In our experiments both functions are provided by an

intermediate-size S/MAR isolated from the center of an extended putative domain border.

B. A novel concept: genomic reference integration sites

The study of *cis*-acting regulatory elements is often confounded by the variability of gene expression among independent transformants. This variability is ascribed to chromosomal position effects (PEV, see IB) at the sites of transgene integration. Inserting single copy test constructs into the same genomic target would control these effects and facilitate valid comparisons of expression levels. Targeting a given site via homologous recombination, though successful in fungal and some animal systems, is not always practical because it occurs at very low

frequencies compared with the high background of illegitimate recombination events.

Figure 7. Reactions catalyzed by site-specific recombinases: Excision/integration and inversion

Excision is the consequence of removing a stretch of DNA which is flanked by two equally-oriented FRT sites; the reaction is mediated by a crossover between these sites. In principle, this reaction is reversible in the sense that a circular vector can be accommodated at a genomic site carrying an FRT tag. The native role of Flp recombinase in yeast is the inversion of the origin of replication region on the 2 plasmid which is localized between two inversely-oriented FRTs (right-hand part).

A full Flp-recognition target site consists of three 13 bp repeats and an 8 bp spacer within 48 bp of DNA (bottom). Each of the 13 bp repeats represents an Flp binding element (FBE) and the spacer is the region in which single strand cuts (vertical arrows) are introduced in preparation of the crossover and resolution steps. There is no physical contact of the recombinase to the spacer which determines the polarity and identity of the site. Hence, spacer mutants will be recombined with an identical FRT site but not with a wild type FRT (Schlake and Bode, 1994). For convenience, the FRT-site is symbolized by an half arrow.

Flp recombinase from the 2 plasmid of *Saccharomyces cerevisiae* can be introduced into mammalian cells to perform site-specific integration reactions (O’Gorman et al., 1991). This enzyme excises any piece of DNA that is flanked by two Flp-recognition target (FRT-) sites of identical orientation (Walters et al., 1996). Even more important in the present context, it also performs the reverse reaction of integrating an FRT-labelled circular vector into an FRT-tag placed in the genome (**Figure 7**; cf. Schlake and Bode, 1994 and

references therein). This enables site-specific integration while avoiding the problems of earlier methods. Related approaches are being developed on the basis of the Cre/loxP1 system of bacteriophage P1 which is mostly used for excision-type reactions and with an increasing number of other members of the Int recombinase family (reviews: Kilby et al., 1993; Sauer, 1994).

Suitable sites for integration can be prepared by introducing an FRT-tagged construct into an endogenous locus with known properties. This is only possible in

systems permissive of homologous recombination, and for *loci* that are redundant and constitutively expressed such as the histone *locus* in mice (Wigley et al., 1994). Alternatively, the construct can be transferred by electroporation (section IIIC1). Among the multiple clones that result, those integrants will be chosen which mediate a high and consistent expression in the absence of continued selection pressure. This will guarantee that the respective site does not become inactivated by heterochromatization, by DNA methylation, or by being situated in a locus that is genetically unstable. Both procedures are initially laborious, but screening has only to be performed once. As an additional advantage of the approach, if an FRT-tagged construct tends to form head-to-tail multimers, these concatemers will be reduced by a continuous action of the recombinase which (according to **Figure 7**) will excise pieces of DNA that are flanked by equally-oriented sites (Lakso et al., 1996). If the individual vector contains just one of these sites, ideally the excision will continue until a single, intact copy is left.

For a convenient characterization of expression parameters, it is advantageous first to introduce a reporter gene which can be monitored easily. Since integration can be reversed by a second pulse of Flp activity, the remaining FRT site is then open for other rounds of integration during which any gene of interest can be directed to the pre-defined locus. The general validity of this concept was demonstrated using the Cre/loxP system (Fukushige and Sauer, 1992). However, the straightforward application of the scheme in **Figure 7** faces a number of problems which will be discussed after describing some molecular features of site specific recombinases, exemplified by Flp/FRT.

1. Mechanism of recombination: design and function of FRT sites

The amino acid sequence of Flp bears no resemblance to any known DNA binding motif. The full 48 bp FRT site consists of three individual 13 bp Flp binding elements (FBE a-c), two of which form an inverted repeat around an 8 bp spacer. The third 13 bp element (c) represents a direct repeat and is separated from b by a single base pair. While most reactions are also possible with a minimal 34 bp site consisting of the inverted repeat and the spacer, the integration reaction appears to benefit from the presence of the extra FBE (Lyznik et al., 1993).

Association of Flp with its site causes bending of two types. While each Flp monomer introduces an individual bend of 40°; a larger bend (>140°) is caused by the interaction of the two Flp monomers across the spacer. This is a consequence of strong interprotomer contacts which occur despite the fact that the protein binding sites lie on opposite faces of the DNA. The Flp monomer

bound by the third element does not participate in these strong cooperative interactions, which require flexibility of the spacer. This spacer becomes single stranded upon Flp binding (Kimball et al., 1995).

Strand cutting can be initiated when one individual FRT site is occupied by two Flp monomers. This “trans-horizontal” cleavage mode does not require the presence of a synaptic complex. Each monomer of the recombinase has only a partial active site and contributes to the formation of a full active center by donating the catalytic tyrosine to the Arg-His-Arg cleft of the partner that is bound across from it on the other side of the spacer. This mechanistic property of Flp guarantees that during the two-step strand transfer of a complete recombination reaction, the activation of one pair of active sites is coupled to the disassembly of the other (Kwon et al., 1997).

Synapsis is mediated by protein-protein interactions between the bound recombinase molecules. In this way the paired strand cleavage steps become coordinated (**Figure 8**). This underlines the importance of the synaptic complex, which channels the chemistry of strand breakages so that recombination, not self-healing, is the final result. After the initial strand breaks, strand transfer, and ligation, a Holliday intermediate is formed. Homology permits the Holliday junction to undergo branch migration and isomerization, during which the crossover strands and the helical strands switch functions. This isomerization probably results from a preference for one pair of stacking isomers over another (Li and al., 1997).

2. The homology checkpoints

A central requirement for recombinations catalyzed by members of the INT family is an absolute homology between the partner substrates in a strand-exchange reaction. This implies a DNA-DNA interaction at some point in the reaction. Recent evidence suggests that homology is not checked before strand cleavage, so the first strand transfer can occur in spite of one or more mismatches. Subsequently there are two homology checkpoints. First, the strand-joining step requires complementary base pairing to orient the 5-OH group for its attack on the phosphotyrosyl bond at the cleavage point. Second, the branch-migration event and associated isomerization of the recombination complex require homology (see insert to **Figure 8**).

3. Construction and use of reference integration sites

An early application of the Flp/FRT technology investigated chromatin domains *in situ* (**Figure 9A**). It was based on the assumption that, in a single-S/MAR construct, the reporter remains sensitive to influences from

the surrounding chromatin. We transfected the S/MAR-Luciferase-FRT cassette and subsequently isolated a series

Figure 8. Formation and resolution of the Holliday structure during the action of Flp recombinase.

Four molecules of the recombinase (ellipses) participate in site specific recombination which are bound next to the crossover region (8 bp spacer). After single strand cuts are introduced into the recombining partner, a first strand transfer step occurs, followed by ligation. During a branch migration process, which depends on absolute homology of the interacting strands, the end of the spacer is reached. During or following this process an isomerization is thought to produce a structure in which the crossover and noncrossover strands are switched. The structure is resolved by two more single strand cuts and ligation. Insert: Relative movement of two (more extended) homologous helices during branch migration.

of clones with widely different expression characteristics. After closing the domain by Flp-mediated targeting of a second S/MAR element to the endogenous site, clones were isolated and the levels of luciferase expression were compared to the initial ones. Among the recovered clones there was only a minority showing an augmented luciferase activity. When a second pulse was applied to these to re-excise the second S/MAR, the old expression characteristics were re-established. These clones underwent the modification-demodification cycle depicted in **Figure 9A** (Bode et al., 1996). The majority of them did not permanently acquire the second S/MAR element, but rather gained hygromycin resistance due to a faulty integration of the circular FRT-S/MAR-HygTk vector. In other cases integration of the circular Flp expression vector occurred, leading to a permanent base level of recombinase activity (Iber, 1997). These results underline a major problem with

the Flp and Cre recombinase systems: because the reactions are reversible and excision is highly favored (being an intramolecular process), they do not allow one to control the direction of recombination.

This approach did therefore not permit the investigation of a large number of events to establish an insulation function of S/MARs, especially at sites with an initially mediocre expression. Nevertheless, it provided the clear demonstration of a transcriptional augmentation at a singular integration reference site. This experiment is therefore considered additional evidence that augmentation is - at least in part - due to a cis-effect of S/MARs on transcriptional initiation.

At present the fastest progress in the field is expected from a simple reversal of the above strategy (**Figure 9B**): A complete minidomain is constructed, and the flanking

S/MAR elements are removed by the successive action of two different site specific recombinases. For this approach

Figure 9. Engineering the genome: Addition, excision and exchange reactions catalyzed by Flp recombinase

A. Circular vectors can be integrated at a genomic locus that has been tagged with an identical site. The reaction has to be driven by an excess of the vector and will easily reverse if Flp activity persists after its dilution or degradation. The depicted experiment has been used to complete an artificial domain by the addition of a second S/MAR element (Bode et al., 1996)

B. Excision of boundary elements by site-specific recombinases (Flp, Cre) can be used to decompose an artificial domain in a stepwise fashion. The upstream S/MAR is surrounded by FRT sites and is removed by the action of Flp recombinase. In a related reaction, the downstream S/MAR can be excised by Cre recombinase acting upon the lox sites.

C. An expression cassette (gene 1) that is surrounded by a wild type FRT site and an FRT linker-mutant can be exchanged for a cassette (gene 2) with an analogous set of sites. This double-reciprocal crossover reaction will delete the plasmid sequences contained in the circular gene 2 vector and will hence result in a so called “clean exchange”.

we initially used the same 800 bp S/MAR element in both the 5' and 3' positions so the results could be related to the position of the element without the need to consider different S/MAR structures. The upstream S/MAR was surrounded by two FRT sites, hence could be excised by a pulse of Flp recombinase. Similarly, the downstream S/MAR was surrounded by two loxP sites, and could be excised by Cre recombinase. Selection of cells which had received the Flp construct was facilitated by a bicistronic vector encoding Flp (first cistron) and GFP (green-

fluorescent protein, second cistron). By sorting for green fluorescence a population of cells was obtained in which excision had occurred with over 90% efficiency. This underscores the fact that excision is a spontaneous event. In a separate step, the downstream S/MAR was removed by Cre recombinase. Here, Cre was expressed stably and expressors were selected via hygromycin resistance, generated by a cotransfected selection marker. The results of a pilot experiment on a cell population, generated by electroporation as for **Figure 2** (thereby mostly

consisting of single copy constituents), are shown in **Figure 10**. These data show a moderate effect (25% decrease of expression) for the removal of the 5' S/MAR and a strong one (85%) for the removal of the corresponding 3' element. It is noted that these findings agree with the model of S/MAR action depicted in **Figure 6** (transition D2 -> D2.1): the unwound structure stored in a S/MAR element downstream from the site of transcriptional initiation can be utilized to release the positive superhelical strain that accumulates in front of a transcribing polymerase. Recent experiments by Wang and Dröge (1996) have demonstrated the persistence of supercoiling even in the presence of topoisomerase

activities which resolve topological problems on a longer time scale.

The experiments outlined in **Figure 8B** will permit an extended series of tests for the properties of individual clones, even at the single cell level. This possibility arises from the use of the LacZneo gene (Walters et al. 1995) the product of which confers neomycin resistance and at the same time allows fluorescence-activated cell sorting (FACS analysis) of clones according to their lacZ expression level. In another system it was shown that integration occurs next to the centromere in up to 50% of clones, which causes position-effect variegation (PEV) unless the chromatin is either kept open by an enhancer-

**Figure 10. Probing
the activity of
S/MARs in situ:
excision of the 5'-
S/MAR and the 3'-
S/MAR from an
artificial minidomain.**

The experiment followed the outline given in **Figure 9** and is based on two identical 800 bp S/MAR elements flanking a LacZ/neo fusion gene. Excision of the 3'-S/MAR by Cre results in a larger effect than excision of the 5' element.

like function (Festenstein et al., 1996) or flanked by insulator-type elements. Both the domain-opener (Zhao et al. 1993) and the insulator functions (Stief et al., 1989; Phi-Van et al., 1996) that have been ascribed to S/MARs should prevent PEV. We hope to trace these functions by successively eliminating the 5'- and the 3'- elements. The effect of each of these deletions on the decay of expression will be traced over an extended period of time (cf. Walters et al., 1996).

4. The equilibrium problem and its solutions

The simple integration/excision system of **Figure 9A** has one major drawback, caused by the reversibility of the recombination reactions. Since intramolecular excision is kinetically favored over bimolecular integration, insertion products are inherently unstable in the presence of recombinase (Seibler and Bode, 1997). As an example, we have reported the facile excision of retroviral sequences between FRT sites which were strategically placed into the long terminal repeats (LTRs) of a provirus (Schübeler et al, 1997; cf. also **Figure 3**). Reversal of this step, i.e. use of the remaining site for re-integration, proved to be unfeasible. We have recently shown that this goal can be achieved by applying a very stringent selection system (integration of a promoter- and ATG-free cassette next to a preexisting promoter and translation-initiation site) and by maintaining an open chromatin structure around the target site (Seibler et al, submitted).

Several measures have been taken to limit the activity of the recombinase to a time interval where a high concentration of the circular exchange vector drives the integration. Conventionally, this is done by generating a pulse of Flp activity from an appropriate concentration of a transiently expressed construct. Since in many cases the transient expression phase is followed by the integration/stable expression of the vector, recombinase-mRNA or -protein has been used instead. In another approach the recombinase gene has been placed under the control of an inducible vector (Logie and Stewart, 1995). Alternatively, one could abolish recombinase activity following integration by directing the integration event so that it separates the promoter from its coding sequence (Kilby et al., 1993). Another strategy is to use mutant target sites. For both loxP and FRT exact 13 bp inverted repeats are the recombinase binding sites, which implies a stringent sequence requirement. If a point mutation is introduced in one of the repeats, a recombination between the site with a mutation in the left element (LE) and another site with a mutation in the right element (RE) would yield two recombination product sites, i.e. a wild-type one and one with mutations in LE and RE. Since the LE plus RE site has a dramatically reduced affinity for the recombinase, a subsequent excisional recombination

between this and the wild-type site becomes less probable favoring the forward (integration) reaction (Senecoff et al., 1988; Araki et al., 1997). Unfortunately, when compared with wild type FRTs, point mutations also lead to a reduced recombination between the LE and RE sites, resp., and there are even cases where the enhanced stability conferred upon the integrated molecule was outweighed by the far fewer integration events which resulted from an inefficient forward reaction.

There are no identified protein-DNA interactions in the 8 bp spacer sequence of an FRT site. At least six (and possibly all) of these bases can be changed without destroying Flp binding activity. Such changes produce a mutant site that will recombine with a second mutant site of the same composition, but not with one having a different spacer sequence (Schlake and Bode, 1994). We have indicated above (section IVB2) that this is an immediate consequence of the homology check points occurring during the branch migration step of the recombination cycle.

Based on these results we have studied the feasibility of a cassette exchange reaction mediated by Flp (RMCE concept, see **Figure 8C** and Seibler and Bode, 1997). We demonstrated that the double-reciprocal crossover events occurring between FRT couples of identical composition enable the efficient substitution of a recombination target flanked by a wild-type FRT site and an FRT- mutant for another cassette designed in the same way. Since RMCE is a true equilibrium reaction (both the forward and the reverse reaction are bimolecular processes) it proceeds to near completion if the exchange plasmid can be provided at a sufficiently high excess (Seibler and Bode, 1997). So far, the RMCE concept not only provides the most efficient solution for the equilibrium problem but it also enables a "clean" replacement of one expression cassette for another in the sense that prokaryotic vector parts can be deleted during the exchange step by an appropriate placing of the wild type FRT site (F) and the FRT mutant (F_n)

5. Outlook

The ability to perform a clean exchange of one expression cassette for another provides an entirely new way to gently manipulate the genome. In its most stringent form this approach requires that expression patterns remain unperturbed, as assessed by the simultaneous transcription of a selection marker which has to be removed in a second step. This goal can be achieved by a novel two-step strategy called "tag-and-exchange" (Askew et al., 1993) or double replacement (Stacey et al., 1994). In our modification of the concept, an expression cassette carrying the HygTk positive/negative selection marker is introduced in step 1. In the appropriate cell types, this can either be achieved by an 3-type homologous

recombination which substitutes an endogenous gene for the HygTk- "tag" or by a random integration followed by the selection of suitable integrants. The presence of the tag can be assessed because the HygTk gene product mediates resistance to hygromycin. Following the outline in **Figure 9C**, the HygTk tag is removed in step 2 by an exchange reaction utilizing the RMCE principle. Successful events can be screened for the absence of the HygTk cassette (negative selection). This is done in the presence of ganciclovir which is converted to a toxic compound by the thymidine kinase activity of the HygTk fusion gene product. As a result, the initial HygTk tag serves as a selection marker in both steps of the procedure, obviating the need for a marker on the final DNA. This procedure can be performed in embryonic stem (ES-) cells with high efficiency since this cell type does not spontaneously integrate circular DNA. In this way a circular Flp expression construct can be provided at high enough concentrations to generate sufficient recombinase activity. Although it can disappear by dilution or degradation, it will not be incorporated into the genome by an unspecific integration event. An excess of the circular exchange plasmid also can be used whereby the specific exchange mediated by the two sets of Flp sites becomes the favored pathway (Seibler et al, submitted).

The combination of targeted gene modification and production of animals derived from ES cells has established a powerful method for studying gene function in the developing animal. Genes can be disrupted, inserted, or modified in the ES cell genome, and the altered cells can be used to generate chimaeric animals. If ES cells contribute to the germ line, chimeras can be outcrossed to produce progeny that are heterozygous or homozygous for the genomic modification. Therefore, using the tag-and-exchange concept in combination with the RMCE technology the way is open to exchange an endogenous gene for an analogue carrying gentle mutation(s). Since the method avoids any further modification at this particular locus, the effect of the mutation will become immediately obvious.

V. Can S/MAR functions be derived from sequence information?

S/MARs are polymorphic and appear to be distributed throughout the eukaryotic genome. They are specific for eukaryotes, 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 and Bode, 1994, 1995). Although prototype S/MARs are AT-rich, they do not share sufficient sequence similarity to allow cross-hybridization (Gasser and Laemmli, 1987; Phi-Van and Strätling, 1990). Biologists have physically identified S/MARs and tried to correlate

their presence with the occurrence of sequence and structural motifs which have subsequently been used to develop algorithms for the prediction of S/MARs from sequence data.

A. Six prominent rules

The predictive scheme introduced by Krawetz and colleagues (Kramer et al., 1996; Singh et al., 1997) and Boulikas (1993a,b) makes use of six features which in various combinations confer an affinity for the nuclear matrix or scaffold: **(i)** DNA replication occurs in association with the matrix, so sites of matrix attachment share certain AT-rich tracts with homeotic protein recognition sites (including several ATTA and ATTTA tracts) and origins of replication. **(ii)** A number of genes contain TG-rich sequences in their 3'-UTRs which can be S/MARs (Boulikas et al, 1996). **(iii)** Intrinsically curved DNA occurs within or near several S/MARs, although curvature or bending is no prerequisite for S/MAR activities *in vitro* (von Kries et al., 1990). **(iv)** Certain dinucleotides, TG, CA or TA, that produce a kink when separated by 2-4 or 9-12 nucleotides, are prominent features of some S/MARs (Boulikas, 1993a). **(v)** Topoisomerase II consensus sequences and cleavage sites are concentrated at sites of nuclear attachment. These have been used to excise complete chromatin domains (Targa et al., 1994). Although this enzyme responds to topology rather than to a strict consensus, the presence of *Drosophila* and vertebrate consensus sequences have served as important criteria for the prediction of S/MARs. **(vi)** Many S/MARs contain significant stretches of AT-rich sequences and both the occurrence of A_n runs (Käs et al., 1993) and $(AT)_n$ tracts (Bode et al., 1992) has been implicated in S/MAR functions. These six patterns have been used to define a set of decision rules with which DNA sequences can be searched to find regions having S/MAR potential. Several examples were published which show a reasonable correspondence between these predictions and wet-lab results (Kramer et al., 1996; Singh et al., 1997; Krawetz and Bode, unpublished).

B. Stress-induced duplex destabilization (SIDD)

The above criteria indicate that S/MAR activity may be related to structural or topological features which are not strictly linked to primary sequence. Chemical probing and 2D gel analyses of S/MAR constructs under superhelical tension revealed that these elements readily relieve strain by becoming stably base-unpaired. In all cases, unpairing could be shown to initiate at a nucleation site, the core unwinding element (CUE), then extend to a wider region (**Figure 11**). These observations have led to the suggestion that S/MARs contain efficient base-unpairing

Figure 11. Stress-induced structures in particular double stranded DNA sequences.

Increasing negative superhelical densities lead to base unstacking which initiates at core unwinding elements (CUEs). This process is followed by a more extensive strand separation which finally involves an entire base-unpairing region (BUR). The energy stored in the open structures may be retrieved by nearby cruciform-, Z-DNA- or triplex-forming sequences. CUEs and BURs can be trapped by chloroacetaldehyde (CAA) which forms etheno-derivatives with cytosine and adenosine bases that are located in single-stranded regions. The figure shows a derivatized adenine which is no longer able to base pair.

regions (BURs). Since single-stranded character could also be found at S/MARs in living cells, it is possible that duplex destabilization mediates at least some of their functions (Bode et al, 1992, 1995, 1996). This hypothesis is supported by observations that base-unpairing properties correlate with the strength of binding of S/MARs to nuclear scaffold/matrix preparations *in vitro*, and to the potential of these elements to augment transcriptional initiation rates *in vivo* (Mielke et al., 1990, Bode et al., 1992, see also Allen et al., 1996).

We have recently put these hypotheses to a critical test by calculating the stress-induced duplex destabilization (SIDD) profiles for prototype S/MARs for which chemical reactivity data were already available. Sample results are shown in **Figure 12** for several S/MARs integrated into plasmids. These are an 800 bp fragment from the 7 kb S/MAR upstream from the human interferon- β (huIFN- β)

gene (Bode and Maass, 1988; element IV in Mielke et al., 1990), an inactive mutant of that sequence, and the immunoglobulin μ -chain enhancer-associated S/MAR.

For these analyses the S/MAR sequences were placed in the pTZ-18R plasmid. This is the same plasmid as was used to experimentally determine the reactivity of the huIFN- β S/MAR and its mutant with the single-strand specific reagent chloroacetaldehyde (CAA) (Bode et al., 1992 and **Figure 11**). In all cases a superhelix density of -0.05 was used, simulating the conditions existing in a bacterial plasmid (Benham et al., 1997). In the resulting destabilization plot a value near zero indicates an essentially completely destabilized base pair, which is predicted to denature with almost no input of additional free energy. But partial destabilization, indicated by intermediate energy values, may also be important, as it may enable protein binding or other events to occur.

Figure 12. Stress-induced duplex destabilization (SIDD-)profiles for prototype S/MARs and a mutant.

A. 800 bp fragment from the huIFN- β upstream S/MAR shown in a vector backbone. The S/MAR insert is indicated by the horizontal bar and the CUE is marked by an asterisk. The peaks at map positions 2.2, 3.2 and 3.7 kb are destabilized sites at the amp terminator, amp promoter and f1ori of the vector backbone

B. same as A but after mutagenesis of the CUE (light asterix)

C. SIDD profile for the murine IgH-enhancer-S/MAR sequence, superimposed on CAA- modification data according to Kohwi-Shigematsu and Kohwi (1990, 1997). Sites for some S/MAR-binding proteins have been added (from Dickinson et al., 1992).

The calculated destabilization profiles show some distinct features which recurred in all other analyses performed to date:

(i) Those parts of the sequence derived from the original plasmid are generally stabilized by 8-10 kcal/mole. However, there are three sharp and well separated minima between 2.0 and 3.7/0 kbp which correspond to the terminator and promoter of the β -lactamase gene, resp. and to the f1 origin. All these elements have been the subject of earlier analyses (Figures 2 and 3 in Benham, 1993). For our purposes they serve the role of well defined internal standards;

(ii) in striking contrast to the prokaryotic part, the S/MAR sequence (present between 0.2 and 1.0 map units) is chaotically destabilized exhibiting a characteristic succession of minima with a spacing of 200-400 bp over its entire length. Such a modular design is thought to be related to function and may thereby be of diagnostic value (see the analyses by Okada et al., 1996);

(iii) a core unwinding element (AATATATTT in this case), mapped by chemical labeling techniques to position 0.72 occurs at one of the most destabilized sites on the molecule (Bode et al., 1992);

The β -lactamase-associated sites mentioned above were among the first for which local denaturation had been demonstrated by nuclease digestion on superhelical pBR322 DNA (Kowalski et al., 1988). The free-energy parameters governing these transitions have been calculated from these experimental results (Benham, 1992). The plasmid-derived peak centered at position 3.7/0 coincides with the f1 ori which, in the context of this multipurpose plasmid, enables the generation of single stranded DNA after superinfection with a helper phage. Although the f1ori is too short to constitute a S/MAR *per se*, we have found that it contributes synergistically to scaffold binding in the presence of other S/MAR-sequences (Figure 1 in Mielke et al., 1990).

pCL (**Figure 12A**) and pCL^{mut} (**Figure 12B**) contain the unwinding core of the huIFN- β S/MAR in its wild type and mutagenized form, respectively. A critical comparison shows that the core unwinding element which is stabilized by less than 1 kcal/mol in pCL reaches 8 kcal/mole after mutagenesis, again in perfect agreement with the chemical reactivity data (Bode et al, 1992).

We also have calculated destabilization profiles for an extended series of S/MAR elements whose relative binding strengths have been established by S. Michalowski and S. Spiker (in preparation). A prediction of the relative binding strength of these elements was obtained by relating the area covered by these S/MARs in the SIDD profile to the area covered by the amp^r-related peaks. The results agreed well with the experimental data (correlation coefficient 0.89). This was better than predictions based on the occurrence of other criteria, such as A-boxes (0.58), AT richness (0.77) and even the occurrence of a motif common to this particular set of sequences (0.81).

Evidence from several laboratories shows that some S/MARs cohabit with enhancers (Gasser and Laemmli, 1986). This association is particularly intriguing, as S/MARs have the capacity to augment transcription via a non-enhancer mechanism (IIB). The most thoroughly studied examples are the immunoglobulin γ - and μ -chain intronic enhancers, which are associated with one and with two distinct S/MAR elements, respectively. (Cockerill and Garrard, 1986a,b; Cockerill et al., 1987). They function in domain opening (Zhao et al., 1993; Bode et al., 1996), which operates during embryonic development (Jenuwein et al., 1993, 1994, 1997; Forrester et al., 1994; Oancea et al., 1997). A regional demethylation occurs in a process that relies on several cis-acting modules, including the S/MAR (Lichtenstein et al., 1994; Kirillov et al., 1996; Jenuwein et al., 1997). While any S/MAR sequence appears to be able to function in this reaction, tissue specificity is contributed by the intronic enhancer (Kirillov et al., 1996).

For the μ -chain intronic enhancer (**Figure 12c**) Kohwi-Shigematsu & Kohwi (1990, 1997) have demonstrated an overlap between specific protein binding sites and locations that become stably and uniformly unpaired when this region is subjected to torsional stress. Prominent destabilized sites coincide with both the 3'- and the 5'-S/MAR which have been characterized by Cockerill et al., 1987 and Mielke et al., 1990 (cf. elements XVI₁ and XVI₂).

We also have evaluated the destabilization properties of the 992 bp XbaI fragment (**Figure 12c**) by computation. A striking tripartite destabilization profile occurs in the insert region, in which the (stable) enhancer is bounded by two strongly destabilized flanks. The latter regions coincide with the S/MARs, and also with the regions accessible to

CAA. The precision of this analysis becomes evident from the fact that the unwinding feature which initiates at a core unwinding element, AATATATTT, then spreads in the 5' direction and is stalled at the enhancer border (Kohwi-Shigematsu and Kohwi, 1990), is precisely predicted. This directional preference is not readily explained by the mere A+T-contents of neighboring sequences which are both 70%. In contrast to this, the unwinding region 5' of the enhancer does not influence neighboring regions, i.e. it shows an all or none reactivity towards CAA. This property is reflected by the two steep flanks bordering the CAA-reactive region in the destabilization profile.

C. Occurrence of secondary structures

These results raise the question of whether the presence of long stretches of base-unpaired or destabilized duplex DNA is sufficient to account for S/MAR activities, or if these are modulated by alternative stress-induced structures. Schroth and Ho (1995) have demonstrated that strong cruciform forming sequences (inverted repeats, IR) occur at relatively high frequency in yeast (1/19700 bp) and humans (1/41800 bp) whereas triple-helix promoting sequences (mirror repeats, MR) are abundant only in humans (1/49400 bp). While eukaryotic IRs are very A+T-rich, prokaryotic ones have a relatively high G+C-content and occur almost exclusively in transcription termination sites. Base composition is important because cruciforms form more easily in AT-rich sequences. Since strong cruciform and triplex DNA forming sequences are not abundant in the *E. coli* genome, these results suggest they may have specific roles in eukaryotes, where they are concentrated in S/MARs and in ORIs (Boulikas & Kong, 1993; Boulikas, 1995; Mielke et al., 1996 and below). Still other observations hint at a regulatory role of direct repeats in the replication of eukaryotic genomes. Direct repeats are also common in S/MARs (Opstelten et al., 1989, Mielke et al., 1996).

Among the possible supercoiled-induced alternate DNA structures, triplexes are now felt to be the best candidates for serving a role in gene expression, as their requirements for specific environmental conditions and negative supercoiling are the least stringent (Palacek, 1976, 1991). After strand separation, the experimentally best characterized structural change in a negatively supercoiled DNA is the C-type cruciform extrusion. This transition initiates with a coordinated opening of many base pairs to form a large bubble that can be trapped by a single-strand specific reagent like chloroacetaldehyde (**Figure 11**). As a prelude to opening, the base pairs must be unstacked. This partial relaxation already mediates reactivity to osmium tetroxide, which allows the researcher to visualize subsequent intermediates on the extrusion pathway (Furlong et al., 1989).

For C-type extrusions, AT-rich BURs may be positioned at the center of a cruciform as in **Figure 11**. These also could be responsible for a coordinate destabilization of a large domain in the supercoiled DNA, and thereby increase the probability that more distant sites "harvest" the energy stored in these structures. In this case a BUR could be separated from a stress-induced non-B structure, and might be recognized and thereby stabilized by certain scaffold proteins. Clearly, this point needs clarification and it can only be approached when the appropriate experiments have been performed to determine the precise energetics for various alternatives under the same environmental conditions. For the time being, analyses have to be restricted to the sequence features which would be compatible with secondary structure formation in and around the destabilized regions.

1. Potential cruciforms in S/MAR-type sequences

Several independent lines of evidence support the idea that cruciform structures might be enriched in S/MARs (Boulikas, 1993, 1995): First, hnRNA, a component of the nuclear matrix, is anchored by regions that correspond to DNA inverted repeats. That these features direct origins of replications and S/MARs to the nuclear matrix is suggested by the results of a recent random cloning experiment which found a number of elements that were significantly enriched in IRs. In 77 kb of the human β -globin locus 22 potential cruciforms have been found, some of which coincide with S/MAR sequences. The recognition of these sequences could be due to several established protein components of the matrix (review: Bode et al., 1996), among these HMG1 and a number of transcription factors.

Using a computer program for a search of regions with the sequence requirements for cruciform formation, we noted an extended inverted repeat in the human interferon- β upstream S/MAR. The corresponding stem-loop structure (**Figure 13**) would expose the core unwinding element (ATATTT) on the tip of a loop. We have recently determined, that retroviruses prefer S/MAR-type sequences for their integration into the host's genome (**Figure 5**). Within these sequences we noted a high proportion of direct and inverted repeats (Mielke et al., 1996). When we calculated the corresponding SIDD profiles there were again several coincidences between stabilization minima and potential stem-loop structures (see Int-19 and Int-26 in **Figure 13**).

A much investigated retroviral integration event which leads to an extensive deregulation of the collagen(I) expression is localized in the gene's first intron (Breindl et al., 1984). As in the other reported cases, a 300 bp sequence around this site behaves as a prototype S/MAR

element of the same extension (H. Rühl, unpublished). An SIDD analysis of the intron reveals a succession of peaks, one of which represents the actual site of integration. While a more extensive *in vitro* analysis of the S/MAR potential along the entire sequence is in progress (C. Mielke, unpublished), the question arises whether this particular minimum was preferred by incidence or whether there are signals superimposed to it which might have attracted the retroviral integration machinery. It is demonstrated in **Figure 13** that integration had again occurred at the tip of a potential stem loop structure i.e. between the two guanidine residues shown in bold print. While none of these analyses can yet be considered proof for the existence or relevance of these structures (VC), it is certainly tempting to consider their contribution to explain the range of properties associated with S/MAR-type sequences (IIB-E).

VI. Conclusions and perspectives

S/MARs are typically found at the borders of eukaryotic gene domains. A recent compilation (Boulikas, 1995) covers 50 well characterized S/MAR-elements for mammalian, rodent, chicken, *Drosophila* and plant genes defining domains of 5-400 kb (average size 60 kb). Since there is an inverse relation between domain size and potential gene activity and a frequent association of enhancers with S/MARs, the study of S/MAR organization yields valuable initial information about the nature and expression of those genes that are associated with them.

The human genome project is aimed at the localization of all 50-100000 human genes. Progress depends, to a large extent, on the availability of markers that are polymorphic, very common ($1E5$ to $1E6$ copies per genome) and evenly dispersed. Recovery of all S/MARs from human chromosome 19 by Nikolaev et al. (1996) has shown that these are indeed individual and do not belong to any family of repeated sequences (IIB-D, VB). These attributes could make S/MARs valuable genomic markers in sequencing projects.

Many S/MAR-related functions seem to depend on particular DNA structures (Boulikas 1995, Singh et al., 1997) which are recognized by distinct sets of single- or double-strand specific binding proteins (Bode et al., 1996). Their pronounced base unpairing character (Bode et al., 1992) together with a possible propensity subsequently to form non-B DNA structures under superhelical tension (Boulikas 1993) may explain the observation that some S/MARs coincide with recombination hot spots (Sperry et al., 1989, Kohwi and Panchenko, 1993). Recently investigated examples include sites of translocation in the human type I interferon gene cluster (Pomykala et al., 1994; Diaz 1995) and at the MLL breakpoint cluster region

Figure 13. Is base-unpaired DNA stabilized by secondary structure formation?

The figure shows sequences at the sites of duplex destabilization which have the potential to yield cruciforms. The CUE (ATATTT) of the human interferon upstream S/MAR is localized at the top of a potential stem-loop structure. Related sites are also the targets for retrovirus integration which is exemplified by a Moloney murine leukemia integration site (Breindl et al., 1984) and by two targets for a retroviral vector (Mielke et al., 1996).

(Broecker et al., 1996a, b), all of which occur within S/MARs. Currently several projects address the question of whether S/MAR elements, besides structuring the human genome, might also give rise to the extreme instability of these *loci*.

Until recently, the identification of genomic segments associated with the nuclear matrix has essentially relied on biochemical strategies. Sequence searches for S/MARs have met with serious difficulties since, although several characteristic motifs are known, no true consensus is apparent (Boulikas, 1993, Kramer & Krawetz, 1995, Kramer et al., 1996). The novel, structure-related approach that is discussed above (VB-C) suggests that it may be possible to recognize S/MARs on the basis of subtle underlying properties of their sequences. This would allow rapid progress in the localization of functional genes and some of their associated regulatory features (Benham et al., 1997). The availability of an increasing number of S/MAR elements and the characterization of both their common and individual properties will provide valuable insights regarding genomic organization and regulation.

In the emerging fields of improving agricultural crops, and human gene therapy the inclusion S/MARs that regulate chromatin structure in transgene constructs appears of immediate use to obtain consistent and authentic expression patterns. Any of these protocols relies on the efficiency of DNA delivery as well as on expression properties. These are profoundly influenced by the nature of the insertion site and the presence of DNA elements with the potential to overcome chromosomal position effects. Site specific recombination systems are being developed which will ultimately allow successive rounds of transformation with different genes inserted into the same locus. This locus could either be an endogenous site which can be targeted without interrupting central genomic functions or a site which has been constructed *in situ* by the inclusion of elements which define an autonomously regulated gene domain (IVB-C).

References

- Allen G. C., Hall Jr. G. E., Childs L. C., Weissinger A. K., Spiker S., and Thompson W. F. (1993). Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. **Plant Cell** 5, 603-613.
- Allen G. C., Hall Jr. G., Michalowski S., Newman W., Spiker S., Weissinger A. K., and 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.
- Araki K., Araki M., and Yamamura K. I. (1997). Targeted integration of DNA using mutant lox sites in embryonic stem cells. **Nucleic Acids Res** 25 (4), 868-872 (4).
- Aronow B. J., Silbiger R. N., Dusing M. R., Stock J. L., Yager K. L., Potter S. S., Hutton J. J., and Wigington D. A. (1992). Functional analysis of the human adenosine deaminase gene thymic regulatory region and its ability to generate position- independent transgene expression. **Mol Cell Biol** 12, 4170- 4185.
- Askew G. R., Doetschman T., and Lingrel J. B. (1993). Site-directed point mutations in embryonic stem cells: A gene targeting tag-and-exchange strategy. **Mol Cell Biol** 13, 4115- 4124.
- Benham C. (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. (1992). Energetics of the strand separation transition in superhelical DNA. **J Mol Biol** 225, 835-847.
- Benham C., Kohwi-Shigematsu T., and Bode J (1997) Stress-induced duplex DNA destabilization in scaffold/matrix attachment regions. **J Mol Biol** 272, 181-196.
- Blasquez V. C., Hale M. A., Trevorow K. W., and Garrard W. T. (1992). Immunoglobulin kappa gene enhancers synergistically activate gene expression but independently determine chromatin structure. **J Biol Chem** 267, 23888-23893.
- Bode J., and Maass K. (1988). Chromatin domain surrounding the human interferon-beta gene as defined by scaffold-attached regions. **Biochemistry** 27, 4706-4711.
- Bode J., Kohwi Y., Dickinson L., Joh R. T., Klehr D., Mielke C., and 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., and Klehr-Wirth D. (1995). Scaffold/Matrix-Attached Regions (S/MARs): Structural Properties Creating Transcriptionally Active Loci. In Structural and Functional Organization of the Nuclear Matrix - International Review of Cytology, Jeon K. W., and Berezney R., eds., vol. 162A, chapt. 8 (Orlando Academic Press), pp. 389-453.
- Bode J., Stengert-Iber M., Schlake T., Kay V., and Dietz-Pfeilstetter A. (1996). Scaffold/matrix-attached regions: Topological switches with multiple regulatory functions. **Crit Rev Eukaryot Gene Expr** 6, 115-138.
- Bonifer C., Yannoutsos N., Krueger G., Grosveld F., and Sippel A. E. (1994). Dissection of the locus control function located on the chicken lysozyme gene domain in transgenic mice. **Nucleic Acids Res** 22, 4202-4210.
- Bonifer C., Huber M. C., Jaegle U., Faust N., and Sippel A. E. (1996). Prerequisites for tissue specific and position independent expression of a gene locus in transgenic mice. **J Mol Med** 74, 663-671.
- Boulikas, T. (1993a) Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. **J. Cell. Biochem.** 52, 14-22.

- Boulikas, T (1993b) Homeodomain protein binding sites, inverted repeats, and nuclear matrix attachment regions along the human β -globin gene complex. **J. Cell. Biochem.** 52, 23-36.
- Boulikas T. (1995). Chromatin and prediction of MAR sequences. In International Review of Cytology, Jeon K. W., and Berezney R., eds., vol. 162A (Orlando Academic Press), pp. 279-388.
- Boulikas T., and 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.
- Boulikas T, Kong CF, Brooks D, and Hsie L (1996) The 3' untranslated region of the human poly(ADP-ribose) polymerase gene is anchored to the nuclear matrix. **Int J Oncol** 9, 1287-1294.
- Breindl M., Harbers K., and Jaenisch R. (1984). Retrovirus-induced lethal mutation in collagen I gene of mice is associated with an altered chromatin structure. **Cell** 38, 9-16.
- Breyne P., Van Montagu M., Depicker A., and Gheysen G. (1992). Characterization of a plant scaffold attachment region in a DNA fragment that normalizes transgene expression in tobacco. **Plant Cell** 4, 463-471.
- Broeker P. L., Harden A., Rowley J. D., and Zeleznik-Le N. (1996). The mixed lineage leukemia (MLL) protein involved in 11q23 translocations contains a domain that binds cruciform DNA and scaffold attachment region (SAR) DNA. **Curr Top Microbiol Immunol** 211, 259-268.
- Broeker P. L. S., Super H. G., Thirman M. J., Pomykala H., Yonebayashi Y., Tanabe S., Zeleznik-Le N., and Rowley J. D. (1996). Distribution of 11q23 breakpoints within the MLL breakpoint cluster region in de novo acute leukemia and in treatment-related acute myeloid leukemia: Correlation with scaffold attachment regions and topoisomerase II consensus binding sites. **Blood** 87, 1912-1922.
- Chamberlain J. W., Vasavada H. A., Ganguly S., and Weissmann S. M. (1991). Identification of cis sequences controlling efficient position-independent tissue-specific expression of human major histocompatibility complex class I genes in transgenic mice. **Mol Cell Biol** 11, 3564-3572.
- Chung J. H., Whiteley M., and Felsenfeld G. (1993). A 5' element of the chicken Beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. **Cell** 74, 505-514.
- Cockerill P. N., and Garrard W. T. (1986). 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., and 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.
- Collis P., Antoniou M., and Grosveld F. (1990). Definition of the minimal requirements within the human β -globin gene and the dominant control region for high level expression. **EMBO J** 9, 233-240.
- Cook P. R., and Brazell I. A. (1977). The superhelical density of nuclear DNA from human cells. **Eur J Biochem** 74, 527-531.
- Corces V. G. (1995). Chromatin insulators: Keeping enhancers under control. **Nature** 376, 462-463.
- Diaz M. O. (1995). The human type I interferon gene cluster. **Seminars in Virology** 6, 143-149.
- Dickinson L., Joh T., Kohwi Y., and Kohwi-Shigematsu T. (1992). A tissue-specific MAR/SAR binding protein with unusual binding site recognition. **Cell** 70, 631-645.
- Dietz A., Kay V., Schlake T., Landsmann J., and Bode J. (1994). A plant scaffold attached region detected close to a T-DNA integration site is active in mammalian cells. **Nucleic Acids Res** 22, 2744-2751.
- Dietz-Pfeilstetter A., Kay V., Landsmann J., and Bode J. (1996). Characterization of a plant scaffold attached region (SAR) from a T-DNA integration site /Biosafety of transgenic organism, horizontal gene transfer and expression of transgenes., Schmidt E. R., ed., vol. /25 (Heidelberg Springer), pp. 261-268.
- Dillon N., and Grosveld F. (1993). Transcriptional regulation of multigene loci: Multilevel control. **Trends Genet** 9, 134-137.
- Dobie K., Mehtali M., McClenaghan M., and Lathe R. (1997). Variegated gene expression in mice. **Trends Genet** 13, 127- 130.
- Dorer D. R. (1997). Do transgene arrays form heterochromatin in vertebrates?. **Transgenic Res** 6, 3-10.
- Dorer D. R., and Henikoff S. (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. **Cell** 77, 993-1002.
- Eggert H., and Jack R. S. (1991). An ectopic copy of the Drosophila ftz associated SAR neither reorganizes local chromatin structure nor hinders elution of a chromatin fragment from isolated nuclei. **EMBO J** 10, 1237-1243.
- Eissenberg J. C., and Elgin S. C. R. (1991). Boundary functions in the control of gene expression. **Trends Genet** 7, 335-340.
- Epner E., Kim C. G., and Groudine M. (1992). What does the locus control region control?. **Curr Biol** 2, 262-264.
- Ferraro A., Eufemi M., Cervoni L., and Turano C. (1995). The binding of DNA to the nuclear matrix studied by crosslinking reactions. **J Cell Biochem Suppl.** 21B, 134 (J7-205).
- Ferraro A., Cervoni L., Eufemi M., Altieri F., and Turano C. (1996). Comparison of DNA-protein interactions in intact nuclei from avian liver and erythrocytes: A cross-linking study. **J Cell Biochem** 62, 495-505.

- Festenstein R., Tolaini M., Corbella P., Mamalaki C., Parrington J., Fox M., Miliou A., Jones M., and Kioussis D. (1996). Locus control region function and heterochromatin-induced position effect variegation. **Science** 271, 1123-1125.
- Fiering S., Kim C. G., Epner E. M., and Groudine M. (1993). An "in-out" strategy using gene targeting and F1p recombinase for the functional dissection of complex DNA regulatory elements: Analysis of the β -globin locus control region. **Proc Natl Acad Sci USA** 90, 8469-8473.
- Forrester W. C., Van Genderen C., Jenuwein T., and Grosschedl R. (1994). Dependence of enhancer-mediated transcription of the immunoglobulin Mu gene on nuclear matrix attachment regions. **Science** 265, 1221-1225.
- Fukushige S., and Sauer B. (1992). Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. **Proc Natl Acad Sci USA** 89, 7905-7909.
- Furlog J. C., Sullivan K. M., Murchie A. I. H., Gough G. W., and 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., and 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., and Laemmli U. K. (1987). A glimpse at chromosomal order. **Trends in Genetics** 3, 16-22.
- Greaves D. R., Wilson F. D., Lang G., and Kioussis D. (1989). Human CD2 3'flanking sequences confer high-level T-cell specific, position-independent gene expression in transgenic mice. **Cell** 56, 979-986.
- Hempel K., and Strätling W. H. (1996). The chicken lysozyme gene 5'MAR and the *Drosophila* histone SAR are electroelutable from encapsulated and digested nuclei. **J Cell Sci.** 109, 1459-1469.
- Iber, M. (1997). Sequenzspezifische Rekombinationssysteme zur Charakterisierung von SAR-Elementen innerhalb genomischer Referenz-Integrationsstellen. **PhD Thesis**, Technische Universität Braunschweig
- Igó-Kemenes T., and Zachau H. G. (1977). Domains in chromatin structure. **Cold Spring Harb Symp Quant Biol** 42, 109-118.
- Jackson D. A., Dickinson P., and Cook P. R. (1990). The size of chromatin loops in HeLa cells. **EMBO J** 9, 567-571.
- Jenuwein T., Forrester W., and Grosschedl R. (1993). Role of enhancer sequences in regulating accessibility of DNA in nuclear chromatin. **Cold Spring Harbor Symp Quant Biol** 58, 97- 103.
- Jenuwein T., Forrester W. C., Qiu R.-G., and Grosschedl R. (1993). The immunoglobulin Mu enhancer coreestablishes 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., and Grosschedl R. (1997). Extension of chromatin accessibility by nuclear matrix attachment regions. **Nature** 385, 269-272.
- Kalos M., and Fournier R. E. K. (1995). Position-independent transgene expression mediated by boundary elements from the apolipoprotein B chromatin domain. **Mol Cell Biol** 15, 198-207.
- Käs E., and Chasin L. A. (1987). Anchorage of the chinese hamster dihydrofolate reductase gene to the nuclear scaffold occurs in an intragenic region. **J Mol Biol** 198, 677-692.
- Käs E., Poljak L., Adachi Y., and 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., and Bode J. (1994). Binding specificity of a nuclear scaffold: Supercoiled, single-stranded, and scaffold-attached-region DNA. **Biochemistry** 33, 367-374.
- Kay V., and 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., and King S. L., eds., vol. 5 (Inc. Wiley-Liss), pp. 186-194.
- Kellum R., and Schedl P. (1992). A group of scs elements function as domain boundaries in an enhancer-blocking assay. **Mol Cell Biol** 12, 2424-2431.
- Kimball A. S., Kimball M. L., Jayaram M., and Tullius T. D. (1995). Chemical probe and missing nucleoside analysis of F1p recombinase bound to the recombination target sequence. **Nucleic Acids Res** 23, 3009-3017.
- Kilby N. J., Snaith M. R., and Murray J. A. H. (1993). Site-specific recombinases: Tools for genome engineering. **Trends Genet** 9, 413-421.
- Kirillov A., Kistler B., Mostoslavsky R., Cedar H., Wirth T., and Bergman Y. (1996). A role for nuclear NF- B in B-cell specific demethylation of the Ig kappa locus. **Nature Genet** 13, 435-441.
- Klehr D., Maass K., and Bode J. (1991). SAR elements from the human IFN- β domain can be used to enhance the stable expression of genes under the control of various promoters. **Biochemistry** 30, 1264-1270.
- Klehr D., Schlake T., Maass K., and Bode J. (1992). Scaffold-attached regions (SAR elements) mediate transcriptional effects due to butyrate. **Biochemistry** 31, 3222-3229.
- Kohwi Y., and Panchenko Y. (1993). Transcription-dependent recombination induced by triple-helix formation. **Genes Dev** 7, 1766-1778.
- Kohwi-Shigematsu T., and Kohwi Y. (1990). Torsional stress stabilizes extended base-unpairing in DNA flanking the

- immunoglobulin heavy chain enhancer. **Biochemistry** 29, 9551- 9560.
- Kohwi-Shigematsu T., and Kohwi Y. (1997). High unwinding capability of matrix attachment regions and ATC-sequence context-specific MAR-binding proteins. in **Nuclear Structure and Gene Expression**, (Berezney R and Stein G, Editors), pp111-114, Academic Press.
- Kohwi-Shigematsu, T., Maass K, and Bode J (1997). A thymocyte factor, SATB1, suppresses transcription of stably integrated MAR-linked reporter genes. **Biochemistry**, 36, 12005-12010.
- Kowalski D., Natale D. A., and 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., Singh G. B., and Krawetz S. A. (1996). Computer- assisted search for sites of nuclear matrix attachment. **Genomics** 33, 305-308.
- Kricker M. C., Drake J. W., and Radman M. (1992). Duplication- targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. **Proc Natl Acad Sci USA** 89, 1075-1079.
- Kwon H. J., Tirumalai R., Landy A., and Ellenberger T. (1997). Flexibility in DNA recombination: Structure of the lambda integrase catalytic core. **Science** 276, 126-131.
- Laemml U. K., Kaes E., Poljak L., and Adachi Y. (1992). Scaffold-associated regions: cis-Acting determinants of chromatin structural loops and functional domains. **Curr Opin Genet Dev** 2, 275-285.
- Lakso M., Pichel J. G., Gorman J. R., Sauer B., Okamoto Y., Lee E., Alt F. W., and Westphal H. (1996). Efficient in vivo manipulation of mouse genomic sequences at the zygote stage. **Proc Natl Acad Sci USA** 93, 5860-5865.
- Levy-Wilson B., and Fortier C. (1989). The limits of the DNase I-sensitive domain of the human apolipoprotein B gene coincide with the location of chromosomal anchorage loops and define the 5' and 3'-boundaries of the gene. **J Biol Chem** 264, 21196- 21204.
- Li Q., and Stamatoyannopoulos G. (1994). Hypersensitive site 5 of the human Beta locus control region functions as a chromatin insulator. **Blood** 84, 1399-1401.
- Li X. J., Wang H., and Seeman N. C. (1997). Direct evidence for Holliday junction crossover isomerization. **Biochemistry** 36, 4240-4247.
- Lichtenstein M., Keini G., Cedar H., and Bergman Y. (1994). B cell-specific demethylation: A novel role for the intronic kappa chain enhancer sequence. **Cell** 76, 913-923.
- Logie C., and Stewart A. F. (1995). Ligand-regulated site-specific recombination. **Proc Natl Acad Sci USA** 92, 5940- 5944.
- Lyznik L. A., Mitchell J. C., Hirayama L., and Hodges T. K. (1993). Activity of yeast Flp recombinase in maize and rice protoplasts. **Nucleic Acids Res** 21, 969-975.
- Matzke M. A., and Matzke A. J. M. (1995). Homology-dependent gene silencing in transgenic plants: What does it really tell us?. **Trends Genet** 11, 1-3.
- McKnight R. A., Shamay A., Sankaran L., Wall R. J., and 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., and 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.
- Mehtali M., LeMeur M., and Lathe R. (1990). The methylation- free status of a housekeeping transgene is lost at high copy number. **Gene** 91, 179-184.
- Mielke C., Kohwi Y., Kohwi-Shigematsu T., and 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., Tuemmler M., and Bode J. (1996). Anatomy of highly expressing chromosomal sites targeted by retroviral vectors. **Biochemistry** 35, 2239-2252.
- Mirkovitch J., Mirault M. E., and Laemml U. K. (1984). Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. **Cell** 39, 223-232.
- Mlynarova L., Loonen A., Heldens J., Jansen R. C., Keizer P., Stiekema W. J., and Nap J. P. (1994). Reduced position effect in mature transgenic plants conferred by the chicken lysozyme matrix-associated region. **Plant Cell** 6, 417-426.
- Mlynarova L., Jansen R. C., Conner A. J., Stiekema W. J., and Nap J.-P. (1995). The MAR-mediated reduction in position effect can be uncoupled from copy number-dependent expression in transgenic plants. **Plant Cell** 7, 599-609.
- Neznanov N., Kohwi-Shigematsu T., and Oshima R. G. (1996). Contrasting effects of the SATB1 core nuclear matrix attachment region and flanking sequences of the keratin 18 gene in transgenic mice. **Mol Biol Cell** 7, 541-552.
- Nikolaev L. G., Tsevegiyn T., Akopov S. B., Ashworth L. K., and Sverdlov E. D. (1996). Construction of a chromosome specific library of human MARs and mapping of matrix attachment regions on human chromosome 19th. **Nucleic Acids Res** 24, 1330-1336.
- Novak U., Harris E. A. S., Forrester W., Groudine M., and Gelinas R. (1990). High-level β -globin expression after retroviral transfer of locus activation region-containing

- human β -globin derivatives into murine erythroleukemia cells. **Proc Natl Acad Sci USA** 87, 3386-3390.
- O'Gorman S., Fox D. T., and Wahl G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. **Science** 251, 1351-1355.
- Oancea A. E., Berru M., and Shulman M. J. (1997). Expression of the (recombinant) endogenous immunoglobulin heavy-chain locus requires the intronic matrix attachment regions. **Mol Cell Biol** 17 (5), 2658-2668 (5).
- Okada S., Tsutsui K., Tsutsui K., Seki S., and Shohmori T. (1996). Subdomain structure of the matrix attachment region located within the mouse immunoglobulin kappa gene intron. **Biochem Biophys Res Comm** 222, 472-477.
- Opstelten R. J. G., Clement J. M. E., and Wanka F. (1989). Direct repeats at nuclear matrix-associated DNA regions and their putative control function in the replicating eukaryotic genome. **Chromosoma** 98, 422-427.
- Palacek E. (1976). Premelting changes in DNA conformation. **Progr Nucl Acid Res Mol Biol** 18, 151-183.
- Palacek E. (1991). Local supercoil-stabilized DNA structures. **Crit Rev Biochem Mol Biol** 26, 151-159.
- Palmiter R. D., and Brinster R.L. (1986). Germ-line transformation of mice. **Ann Rev Genet** 20, 465-499.
- Palmiter R. D., Chen H. Y., and Brinster R. L. (1982). Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. **Cell** 29, 701-710.
- Palmiter R. D., Sandgren E. P., Koeller D. M., and Brinster R. (1993). Distal regulatory elements from the mouse metallothionein locus stimulate gene expression in transgenic mice. **Mol Cell Biol** 13, 5266-5275.
- Paulson J. R., and Laemmli U. K. (1977). The structure of histone-depleted metaphase chromosomes. **Cell** 12, 817-828.
- Phi-Van L., and Strätling W. H. (1988). The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. **EMBO J** 7, 655-664.
- Phi-Van L., and Strätling W. H. (1996). Dissection of the ability of the chicken lysozyme gene 5' matrix attachment region to stimulate transgene expression and to dampen position effects. **Biochemistry** 35, 10735-10742.
- Phi-Van L., von Kries J. P., Ostertag W., and Strätling W. H. (1990). The chicken lysozyme 5' matrix attachment region increases transcription from a heterologous promoter in heterologous cells and dampens position effects on the expression of transfected genes. **Mol Cell Biol** 10, 2302-2307.
- Poljak L., Seum C., Mattioni T., and Laemmli U. K. (1994). SARs stimulate but do not confer position independent gene expression. **Nucleic Acids Res** 22, 4386-4394.
- Pomykala H. M., Bohlander S. K., Broeker P. L., Olopade O. I., and Diaz M. O. (1994). Breakpoint junctions of chromosome 9p deletions in two human glioma cell lines. **Mol Cell Biol** 14, 7604-7610.
- Rohdewohld H., Weiher H., Reik W., Jaenisch R., and Breindl M. (1987). Retrovirus integration and chromatin structure: Moloney Murine Leukemia proviral integration sites map near DNase I hypersensitive sites. **J Virol** 61, 336-343.
- Ramstein J., and Lavery R. (1988). Energetic coupling between DNA bending and base pair opening. **Proc Natl Acad Sci USA** 85, 7231-7235.
- Romig H., Ruff J., Fackelmayer F. O., Patil M. S., and Richter A. (1994). Characterisation of two intronic nuclear-matrix-attachment regions in the human DNA topoisomerase I gene. **Eur J Biochem** 221, 411-419.
- Sauer B. (1994). Site-specific recombination: Developments and applications. **Curr Opin Biotechnol** 5, 521-527.
- Schedl A., Montoliu L., Kelsey G., and Schuetz G. (1993). A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic mice. **Nature** 362, 258-261.
- Scheuermann R. H., and Chen U. (1989). A developmental-specific factor binds to suppressor sites flanking the immunoglobulin heavy-chain enhancer. **Genes Dev** 3, 1255-1266.
- Schlake T. (1994). Entwicklung einer auf SAR-Elementen und sequenzspezifischer Rekombination basierenden Strategie zur stabilen Hochexpression in Mammaliazellen. **PhD thesis**, Technische Universität Braunschweig.
- Schöffl F., Schröder G., Kliem M., and Rieping M. (1993). An SAR sequence containing 395 bp DNA fragment mediates enhanced, gene-dosage-correlated expression of a chimaeric heat shock gene in transgenic tobacco plants. **Transgenic Res** 2, 93-100.
- Schroth G. P., and Ho P. S. (1995). Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. **Nucleic Acids Res** 23, 1977-1983.
- Schübeler D., Mielke C., Maass K. and Bode J. (1996). Scaffold/Matrix-attached regions act upon transcription in a context-dependent manner. **Biochemistry** 35, 11160-11169.
- Schübeler D., Mielke C. and Bode J. (1997) Excision of an integrated provirus by the action of FLP-recombinase. **In Vitro Cell Dev Biol Anim** 33, 825-830.
- Seibler J., and Bode J. (1997). Double-reciprocal crossover mediated by FLP-recombinase: A concept and an assay. **Biochemistry** 36, 1740-1747.
- Senecoff J. F., Rossmeissl P. J., and Cox M. M. (1988). DNA recognition by the Flp recombinase of the yeast 2 plasmid: A mutational analysis of the Flp binding site. **J Mol Biol** 201, 405-421.
- Shih C.-C., Stoye J. P., and Coffin J. M. (1988). Highly preferred targets for retrovirus integration. **Cell** 53, 531-537.

- Singh G. B., Kramer J. A., and Krawetz S. A. (1997). Mathematical model to predict regions of chromatin attachment to the nuclear matrix. **Nucleic Acids Res** 25, 1419-1425 .
- Sippel A. E., Schaefer G., Faust N., Saueressig H., Hecht A., and Bonifer C. (1993). Chromatin domains constitute regulatory units for the control of eukaryotic genes. **Cold Spring Harb Symp Quant Biol** 58, 37-44.
- Sperry A. O., Blasquez V. C., and Garrard W. T. (1989). Dysfunction of chromosomal loop attachment sites: Illegitimate recombination linked to matrix association regions and topoisomerase II. **Proc Natl Acad Sci USA** 86, 5497- 5501.
- Stacey A., Schnieke A., McWhir J., Cooper J., Colman A., and Melton D. W. (1994). Use of double-replacement targeting to replace the murine β -lactalbumin gene with its human counterpart in embryonic stem cells and mice. **Mol Cell Biol** 14, 1009-1016.
- Stief A., Winter D. M., Strätling W. H., and Sippel A. E. (1989). A nuclear DNA attachment element mediates elevated and position-independent gene activity. **Nature** 341, 343- 345.
- Stockhaus J., Eckes P., Blau A., Schell J., and Willmitzer L. (1987). Organ-specific and dosage-dependent expression of a leaf/stem specific gene from potato after tagging and transfer into potato and tobacco plants. **Nucleic Acids Res** 15, 3479- 3491.
- Targa F. R., Razin S. V., Moura Gallo C. V. De, and Scherrer K. (1994). Excision close to matrix attachment regions of the entire chicken Alpha-globin gene domain by nuclease S1 and characterization of the framing structures. **Proc Natl Acad Sci USA** 91, 4422-4426.
- Thorey I. S., Cecena G., Reynolds W., and Oshima R. G. (1993). Alu Sequence involvement in transcriptional insulation of the keratin 18 gene in transgenic mice. **Mol Cell Biol** 13, 6742-6751.
- Van der Geest A. H. M., Hall Jr. G. E., Spiker S., and Hall T. C. (1994). The Beta-phaseolin gene is flanked by matrix attachment regions. **Plant J** 6, 413-423.
- Vazquez J., and Schedl P. (1994). Sequences required for enhancer blocking activity of scs are located within two nuclease-hypersensitive regions. **EMBO J** 13, 5984-5993.
- von Kries J. P., Phi-Van L., Diekmann S., and Strätling W. H. (1990). A non-curved chicken lysozyme 5'matrix attachment site is 3'followed by a strongly curved DNA sequence. **Nucleic Acids Res** 18, 3881-3885.
- Walters M. C., Fiering S., Eidemiller J., Magis W., Groudine M., and Martin D. I. K. (1995). Enhancers increase the probability but not the level of gene expression. **Proc Natl Acad Sci USA** 92, 7125-7129.
- Walters M. C., Magis W., Fiering S., Eidemiller T., Scalzo D., Groudine M., and Martin D. I. K. (1996). Transcriptional enhancers act in cis to suppress position-effect variegation. **Genes Dev** 10, 185-195.
- Wang D. M., Taylor S., and Levy-Wilson B. (1996). Evaluation of the function of the human apolipoprotein B gene nuclear matrix association regions in transgenic mice. **J Lipid Res** 37, 2117-2124.
- Wang Z. Y., and Dröge P. (1996). Differential control of transcription-induced and overall DNA supercoiling by eukaryotic topoisomerases in vitro. **EMBO J** 15, 581-589.
- Weidle U. H., Buckel P., and Wienberg J. (1988). Amplified expression constructs for human tissue-type plasminogen activator in Chinese hamster ovary cells: instability in the absence of selective pressure. **Gene** 66, 193-203.
- Weintraub H. (1993). Summary: Genetic tinkering--Local problems, local solutions. **Cold Spring Harb Symp Quant Biol** 58, 819-836.
- Weisbrod S. (1982). Active chromatin. **Nature** 297, 289-295.
- Wigley P., Becker C., Beltrame J., Blake T., Crocker L., Harrison S., Lyons I., McKenzie Z., Tearle R., Crawford R., and Robins A. (1994). Site-specific transgene insertion: An approach. **Reprod Fertil Dev** 6, 585-588.
- Withers-Ward E. S., Kitamura Y., Barnes J. P., and Coffin J. M. (1994). Distribution of targets for avian retrovirus DNA integration in vivo. **Genes Dev** 8, 1473-1487.
- Wolffe A. P. (1994a). Gene regulation: Insulating chromatin. **Curr Biol** 4, 85-87.
- Wolffe A. P. (1994b). The transcription of chromatin templates. **Curr Opin Genet Dev** 4, 245-254.
- Yagil G. (1991). Paranemic structures of DNA and their role in DNA unwinding. **CRC Crit Rev Biochem Mol Biol** 26, 475-559.
- Zhao K., Kaes E., Gonzalez E., and Laemmli U. K. (1993). SAR- dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG- I/Y is enriched in H1-depleted chromatin. **EMBO J** 12, 3237- 3247.
- Zong R. T., and Scheuermann R. H. (1995). Mutually exclusive interaction of a novel matrix attachment region binding protein and the NF-MuNR enhancer repressor - Implications for regulation of immunoglobulin heavy chain expression. **J Biol Chem** 270, 24010-24018.