The EMBO Journal vol.10 no.9 pp.2559-2567, 1991

Analysis of CpG methylation and genomic footprinting at the tyrosine aminotransferase gene: DNA methylation alone is not sufficient to prevent protein binding *in vivo*

Falk Weih, Doris Nitsch, Andreas Reik, Günther Schütz and Peter B.Becker[1]

Institut für Zell- und Tumorbiologie, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-6900 Heidelberg, FRG

[1]Present address: European Molecular Biology Laboratory, Postfach 10.2209, D-6900 Heidelberg, FRG

Communicated by G.Schütz

Specific DNA sequences from several DNase I hypersensitive sites located upstream of the tyrosine aminotransferase (TAT) gene are bound by ubiquitous nuclear factors *in vitro*. Genomic footprinting has shown, however, that proteins are excluded from their potential binding sites in cells where the gene is inactive and that the absence of *in vivo* footprints is correlated with CpG methylation and altered chromatin structures at these sites. *In vitro*, interactions of proteins with sequences of the TAT gene, including binding of the transcription factor CREB to the cAMP-responsive element (CRE), are prevented by a methylated CpG dinucleotide in the respective binding sites, suggesting that methylation of DNA might be sufficient to exclude proteins from their sites *in vivo*. To test directly whether the absence of *in vivo* footprints is the result of DNA methylation, we treated two different cell lines with 5-azacytidine to demethylate CpG dinucleotides. While genomic sequencing confirmed demethylation at two widely separated regions upstream of the TAT promoter, no footprints appeared in these cell lines, even though proteins capable of binding these sites *in vitro* were present in the nuclei. Thus, the simple model whereby protein exclusion *in vivo* is caused solely by DNA methylation is not appropriate in this case. The nucleosomal organization of the potential binding sites suggests that chromatin structure is a dominant determinant in maintaining the inactive state of these sites.

Key words: chromatin structure/CpG methylation/CREB-binding/genomic sequencing/*in vivo* footprinting

Introduction

The interaction of DNA-binding proteins with putative target sequences is often studied *in vitro* by bandshift and footprinting assays (Dynan and Tjian, 1985; McKnight and Tjian, 1986). A tight and specific binding of a protein to a cloned sequence is usually considered as an indication that the protein will also interact with the corresponding site within the cell. Given the complex organization of cellular DNA with nucleosomes and non-histone proteins into chromatin with its hierarchical levels of condensation, *a priori* such a conclusion is not justified. In the nucleus, large stretches of DNA are likely to be inaccessible to nuclear factors, resulting in an effective decrease in complexity of sequences available for protein interaction. It is an attractive hypothesis that changes in chromatin structure could regulate the number of accessible sites as an early step in controlling gene activity (for review see Wolffe, 1990).

Cytosine methylation at CpG dinucleotides, the main modification of vertebrate DNA, has a potential role in the regulation of gene expression. First, hypomethylation of promoters is correlated in general with gene activity and, conversely, *de novo* methylation is associated with the silencing of sequences (for reviews see Doerfler, 1983; Cedar, 1988). Second, changes in methylation patterns at specific sites have been observed to accompany changes in gene activity (Saluz *et al.*, 1986; Shimada *et al.*, 1987; Hansen *et al.*, 1988; for review see Holliday, 1987). Third, methylated sites can be propagated stably through many rounds of replication and thus may constitute epigenetic mutations (Wigler *et al.*, 1981; Holliday, 1987). Mutation-like gene inactivation due to CpG island methylation in many cell lines may account for the loss of cell type-specific functions in culture (Antequera *et al.*, 1990). In addition, effects of DNA methylation on chromatin structure have been demonstrated in several systems (Keshet *et al.*, 1986; Antequera *et al.*, 1989; Michalowsky and Jones, 1989).

If methylation has an effect on gene activity and/or chromatin structure, some protein factors must be able to distinguish between methylated and unmethylated sites to interpret the signals encoded by the modification. These proteins could be *bona fide* transcription factors, accessory proteins or structural components of chromatin. The presence of 5-methylcytosine (5mC) in binding sites for transcription factors inhibits the interaction of some factors *in vitro*, but not of others (Becker *et al.*, 1987; Kovesdi *et al.*, 1987; Harrington *et al.*, 1988; Höller *et al.*, 1988; Watt and Molloy, 1988; Iguchi-Ariga and Schaffner, 1989; Comb and Goodman, 1990). Recently, cellular proteins have been identified that only bind to methylated DNA either at a specific sequence (Khan *et al.*, 1988) or with relaxed specificity (Meehan *et al.*, 1989).

We reported on protein–DNA interactions *in vivo* and *in vitro* within a stretch of DNA centred around 1 kb upstream of the transcriptional start site of the tyrosine aminotransferase (TAT) gene (Becker *et al.*, 1987). This site is hypersensitive to DNase I in the hepatoma cell line FTO-2B in which the TAT gene is faithfully expressed in response to all known stimuli, but strongly reduced in the fibrosarcoma cell line XC where the gene is not active (Nitsch *et al.*, 1990). Genomic footprinting has identified multiple protein-binding sites characteristic for TAT gene expression in FTO-2B cells, none of which were detected in XC cells. Nuclear extracts prepared from FTO-2B cells yielded footprints that correlated remarkably well with the binding sites *in vivo*. Surprisingly, all DNA-binding activities were also present in nuclei of TAT non-expressing XC cells, indicating that simply the presence of factors is not sufficient for their interaction with a binding site *in vivo*. The

absence of *in vivo* footprints in XC cells correlated with complete methylation at a cluster of CpG residues within the −1 kb site. *In vitro* CpG methylation within two footprinted regions completely inhibited the interaction of one factor and impaired the binding of the second (Becker *et al.*, 1987). This result suggested that DNA methylation alone may be responsible for excluding proteins from their binding sites *in vivo* and thus contributes to the establishment of a DNase I insensitive chromatin structure and possibly to the inactive state of the gene (Becker *et al.*, 1987; Cedar, 1988). Therefore, we examined whether CpG methylation alone could be sufficient for the exclusion of proteins from their binding sites *in vivo*.

The mCpG-sensitive footprint within the −1 kb DNase I hypersensitive (HS) site provides us with a suitable test case. Because the identity of the proteins binding to sequences located in the −1 kb HS site is still unclear, we extended our studies on protein−DNA interactions *in vivo* and *in vitro* to the functionally well characterized cAMP-responsive enhancer located in a DNase I HS site 3.6 kb upstream of the TAT cap site (Boshart *et al.*, 1990; Weih *et al.*, 1990). Recently, it has been shown that interaction of the CRE-binding protein (CREB) with the cAMP-responsive element (CRE) of the TAT gene *in vivo* and *in vitro* can be stimulated by cAMP induction or protein kinase A treatment, respectively (Weih *et al.*, 1990; Boshart *et al.*, submitted; Nichols *et al.*, in preparation). Since the CRE contains a CpG dinucleotide in its centre, the influence of cytosine methylation within the TATCRE on binding of purified CREB was analysed.

The genomic sequencing and footprinting techniques were used to assess the methylation status of each CpG dinucleotide *in vivo* and to visualize directly protein−DNA interactions at both HS sites within the intact cell. We have used 5-azacytidine treatment of cells in culture to demethylate completely all CpG dinucleotides within the regions of interest but find that the mCpG-sensitive footprints are not established. Thus, the absence of *in vivo* protein−DNA interactions at these sites cannot be causally linked to CpG methylation alone. Rather, the accessibility of a binding site may be determined by its nucleosomal organization.

Results

5-azacytidine treatment leads to complete demethylation

The pyrimidine analogue 5-azacytidine (5-azaCR) has frequently been used to demethylate 5-methylcytosine residues in vertebrate DNA (Flatau *et al.*, 1984; Hsiao *et al.*, 1984; Konieczny and Emerson, 1984; Michalowsky and Jones, 1989; Avvedimento *et al.*, 1989; for review see Jones, 1985). 5-azaCR causes demethylation presumably by binding to and inhibiting the maintenance methylase (references in Holliday, 1987). We have used multiple 5-azaCR treatments (Flatau *et al.*, 1984) of XC fibrosarcoma cells (Svoboda, 1960) to demethylate CpG dinucleotides. First, the region 1 kb upstream of the TAT transcriptional start site was analysed. Within this region a cluster of 9 CpG dinucleotides is found all of which are unmethylated in the hepatoma cell line FTO-2B (Killary and Fournier, 1984) but fully methylated in XC cells. The methylation-sensitive footprint previously observed *in vitro* lies between 1070 and 1090 bp upstream of the cap site (Becker *et al.*, 1987; Figure 1A).

One day after seeding, XC cells were treated with 5-azaCR for 24 h and subsequently cultured without the drug for an additional 8 days after which the treatment was repeated. Altogether, cells were cultured for ~90 days during which eight 5-azaCR treatments were carried out (see Figure 1D). The cells showed no obvious changes in morphology or generation time as compared with untreated sister cultures. DNA was isolated from aliquots of cells after each 5-azaCR treatment to monitor the process of demethylation. For convenience, the methylation status of two *Hha*I sites (at −1044 and −1077, see Figure 1A) was followed. Genomic DNA was cleaved to completion with *Bam*HI, *Pst*I and *Hha*I and analysed on a Southern blot probed with labelled HP111 sequences (Figure 1B). Methylated *Hha*I sites are resistant to cleavage and thus only the 494 bp *Bam*HI−*Pst*I fragment is detected in methylated DNA. Cleavage at demethylated *Hha*I sites results in the shorter *Hha*I−*Pst*I fragment (Figure 1C). The analysis shows that the XC cell population used as starting material is methylated at ≥95% of the *Hha*I sites whereas FTO-2B cells are completely unmethylated (Figure

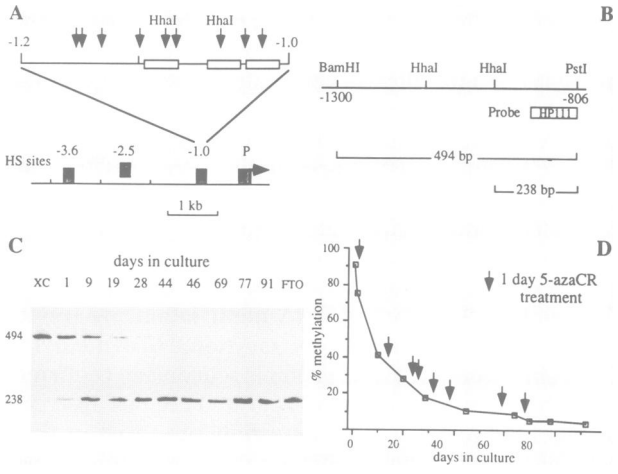

Fig. 1. 5-azaCR treatment leads to complete demethylation of *Hha*I restriction sites in the −1 kb region. (**A**) Hallmarks of active chromatin within 4 kb upstream of the rat TAT gene. Grey boxes indicate the positions of DNase I hypersensitive (HS) sites in FTO-2B cells. Numbers refer to kilobases upstream of the transcriptional start site (horizontal arrow). P: promoter; −2.5 kb: hormone-inducible DNase I hypersensitivity corresponding to the glucocorticoid-responsive enhancer; −3.6 kb: cAMP-responsive tissue-specific enhancer; −1 kb: HS site located 1 kb upstream of the transcriptional start site. Open boxes in the enlargement display sites of protein−DNA interactions *in vivo* and *in vitro*. Vertical arrows point to the positions of CpG dinucleotides, two of which are contained in *Hha*I restriction sites. (**B**) Probing strategy to analyse the methylation status of the *Hha*I restriction sites. Genomic DNA was cleaved with *Bam*HI, *Pst*I, and *Hha*I. The Southern blot was probed with HP111 which reveals a *Bam*HI−*Pst*I fragment when the *Hha*I sites are methylated or a *Pst*I−*Hha*I fragment when the *Hha*I sites are unmethylated and thus cleavable. (**C**) Southern blot documenting the process of demethylation during the successive 5-azaCR treatments of XC cells. The strategy outlined in (B) was employed. The numbers above the lanes refer to the number of days the cells had been cultivated. Cells that are methylated at both *Hha*I sites yield the 494 bp fragment, unmethylated *Hha*I sites result in the 238 bp fragment. (**D**) The autoradiograph in (C) was scanned with a densitometer to obtain a quantitative measure of the fraction of demethylated sites in the cell population. The graph displays the degree of methylation in XC/AzaCR cells calculated as the ratio of the 494 bp band and the sum of the two bands in percent. The 5-azaCR treatments are indicated by vertical arrows.

1C). Successive treatments with 5-azaCR result in a progressive demethylation. After 91 days in culture and eight drug treatments, <5% of XC cells remain methylated at the *Hha*I sites (Figure 1D). At this point the cells (XC/AzaCR) were finally expanded and frozen in aliquots. The demethylated status was stable during the multiple rounds of replication necessary to grow the cells to the numbers required for the following *in vivo* and *in vitro* analyses.

Analysis of the effect of CpG methylation on protein–DNA interactions at the −1 kb HS site: demethylation does not induce protein binding in vivo

The methylation status of all cytosine residues within the −1 kb HS site in XC/AzaCR cells was analysed by genomic sequencing (Church and Gilbert, 1984). Genomic DNA was digested with *Pst*I prior to hydrazine treatment. Hydrazine does not react with 5mC and thus missing bands in the cytosine pattern correspond to modified nucleotides. The blot shown in Figure 2 (right panel) resolves six CpG dinucleotides all of which are methylated in XC cells (absence of corresponding bands) and demethylated on both strands in XC/AzaCR cells (compare with FTO-2B control). Genomic footprinting using dimethylsulphate (DMS) to probe for protected guanosine (G) residues in intact cells (Ephrussi *et al.*, 1985) was employed to reveal whether proteins now bound within the newly unmethylated region in XC/AzaCR cells (Figure 2, left panel). Whereas the characteristic G protections were readily detectable in FTO-2B cells, as previously described (Becker *et al.*, 1987), no protein binding was observed in XC/AzaCR cells nor in XC cells. We conclude that methylation alone cannot be responsible for the exclusion of these proteins from their binding sites *in vivo*.

It is conceivable that in XC cells the TAT gene is packaged into an inert, higher order chromatin structure that would prevent the access of protein factors to TAT control regions in these cells. Therefore, we made use of the hepatoma cell line HTC (Thompson *et al.*, 1966). The TAT gene is expressed in HTC cells but at ~10-fold lower levels than in FTO-2B cells. This lowered expression has been related to reduced activity of the two defined enhancers. Furthermore, several HS sites characteristic of the TAT gene in FTO-2B hepatoma cells are absent in HTC cells including

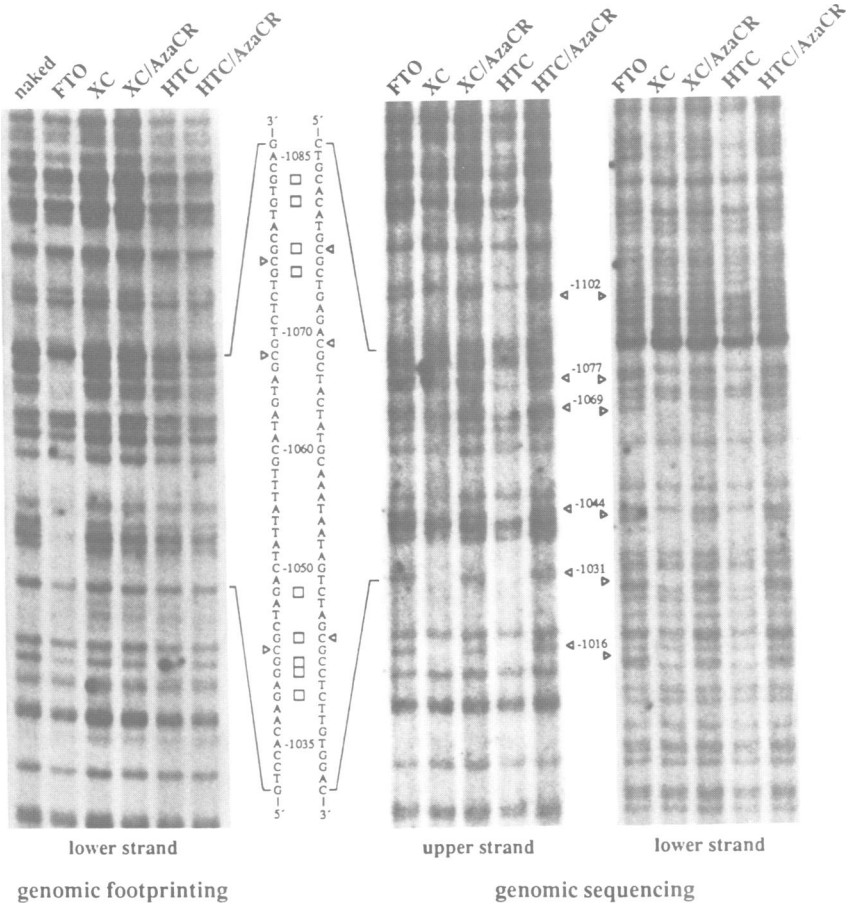

Fig. 2. 5-azaCR treatment leads to complete demethylation of all mCpG dinucleotides within the −1 kb HS site but does not allow protein–DNA interactions *in vivo*. For genomic footprinting (left panel), the cell populations indicated above each lane were treated with dimethylsulphate (DMS). Purified genomic DNA was digested with *Pst*I and cleaved with piperidine at positions of methylated guanosines (Gs). For genomic sequencing (right panel), protein-free DNA was digested with *Pst*I and subjected to a standard cytosine-specific sequencing reaction with hydrazine. After piperidine treatment, 30 μg of genomic DNA per lane were separated on a 6% denaturing polyacrylamide gel, transferred and covalently bound to GeneScreen membrane (New England Nuclear) by UV irradiation. The membrane was hybridized with a single-stranded probe (see Figure 1B) of high specific activity. G residues which show lower DMS reactivity in FTO-2B cells than in XC cells are marked with open squares. Triangles mark the positions of CpG dinucleotides which are demethylated during the 5-azaCR treatments. Numbers indicate the positions in base pairs relative to the start site of transcription.

the HS site at −1 kb (Nitsch *et al.*, 1990). In contrast to the situation in XC cells where no HS site can be found up to −16 kb of the TAT gene, the −1 kb region in HTC cells must be embedded in active chromatin: the promoter HS site (−0.1 kb) and the hormone-inducible enhancer at −2.5 kb, induction of which is accompanied by the rapid formation of a HS site (Jantzen *et al.*, 1987; Nitsch *et al.*, 1990).

The analysis of the methylation status of CpG dinucleotides within the −1 kb region by genomic sequencing (Figure 2, right panel) reveals that also in HTC cells the absence of the HS site correlates with methylation at all CpG dinucleotides. The characteristic *in vivo* footprints are correspondingly absent (Figure 2, left panel). Again, we asked whether methylation was directly responsible for the apparent absence of *in vivo* protein binding and therefore subjected HTC cells to 5-azaCR treatments analogous to those of XC cells. After 90 days in culture and eight 5-azacytidine treatments, the pool of HTC/AzaCR cells had lost all methylations at the *Hha*I sites (data not shown). Genomic sequencing (Figure 2, right panel) showed that all CpGs in the region had been demethylated during the course of drug treatments. Genomic footprinting (Figure 2, left panel) showed again that demethylation did not result in the establishment of footprints within these cells.

Proteins binding to the −1 kb HS site are present in nuclei of AzaCR cells

5-azacytidine has been reported to trigger major differentiation events in 10T1/2 cells (Konieczny and Emerson, 1984; Jones, 1985). It is thus possible that prolonged drug treatment leads to changes in the composition of nuclear DNA-binding proteins. To rule out the possibility that the absence of *in vivo* footprints in 5-azaCR treated cell lines was simply due to the absence of the corresponding proteins as a consequence of drug treatment, we analysed nuclear extracts from all cell lines used in this study for footprinting activities *in vitro*. Figure 3 shows that nuclear extracts from both 5-azaCR treated and untreated cells contain proteins that bind to cloned DNA between −1009 and −1090. Hence proteins that can potentially interact with these sequences are present in all nuclei but are excluded from their binding sites in cells in which the TAT gene is not active even in the absence of CpG methylation.

Role of cytosine methylation at the CRE for CREB binding: demethylation of the cAMP-responsive enhancer of the TAT gene at −3.6 kb does not allow protein binding in vivo

Since it has not been possible to demonstrate a role for the −1 kb HS site in TAT gene transcription, we similarly analysed a functionally characterized enhancer. The TAT gene enhancer at −3.6 kb is active in FTO-2B but not in XC and HTC cells (Nitsch *et al.*, 1990) and contains a CRE (Boshart *et al.*, 1990; Weih *et al.*, 1990) which allows us to evaluate the influence of CpG methylation on binding of a well known transcription factor. In order to examine the methylation status in this region by genomic sequencing, protein-free DNA from the various cell lines was digested with *Sty*I prior to hydrazine treatment and further processed as described (Weih *et al.*, 1990). Figure 4 (middle and right panel) shows three CpG dinucleotides, all of which are methylated in HTC but fully demethylated on either strand in FTO-2B and HTC/AzaCR cells. In XC cells, two CpG dinucleotides (at −3648 and −3676) are methylated and become demethylated after 5-azaCR treatment. When protein binding at the TATCRE was analysed by genomic footprinting (Figure 4, left panel), the characteristic pattern of G enhancements and protections was readily detectable in FTO-2B cells. Furthermore, the *in vivo* footprint could be induced upon forskolin treatment, as described previously (Weih *et al.* 1990). However, no protein binding was detected in HTC nor in HTC/AzaCR cells even after forskolin induction. Again, we conclude that CpG methylation alone is not responsible for the exclusion of the TATCRE-binding protein from its recognition site *in vivo*. The results of the genomic sequencing and footprinting experiments at the −3.6 kb HS site are summarized in the lower part of Figure 4.

Binding of CREB to the CRE of the TAT gene is prevented by cytosine methylation

The fact that the *in vivo* footprint at the CRE of the TAT gene covers a CpG dinucleotide in its centre and the finding of Iguchi-Ariga and Schaffner (1989) that CpG methylation of the CRE consensus sequence results in loss of specific factor binding in crude nuclear extracts prompted us to

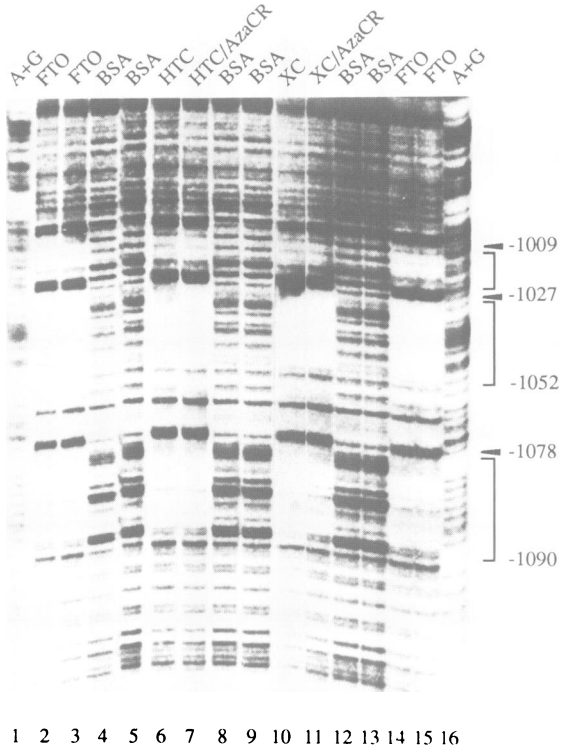

Fig. 3. *In vitro* binding activity of nuclear factors to the −1 kb HS site is not affected by 5-azaCR treatment of cells. DNase I footprints on the lower strand of the −1 kb region of the TAT gene are shown. 50 μg nuclear extract protein from the cells indicated above the lanes were incubated with an end-labelled fragment spanning sequences from −944 to −1225 bp. Incubation on ice was followed by digestion with DNase I at 25°C (6.25 ng/ml in lanes 4, 8 and 12; 12.5 ng/ml in lanes 5, 9, and 13; 25 ng/ml in lanes 2 and 14; 50 ng/ml in lanes 3, 6, 7, 10, 11 and 15). Lanes 1 and 16 show a purine-specific sequencing reaction (A+G). Brackets on the right indicate the characteristic footprints, arrows mark positions of enhanced DNase I cleavage as compared with the BSA control.

analyse the effect of cytosine methylation on binding of purified CREB protein to the TATCRE. Either unmethylated (wt), hemimethylated (um, upper strand methylated; lm, lower strand methylated) or fully methylated (bm, both strands methylated) double-stranded oligodeoxynucleotides containing the TATCRE were labelled and used in bandshift experiments (Figure 5). In comparison to unmethylated templates, binding of purified CREB protein to hemimethylated TATCRE sequences is strongly reduced. When both strands contain a 5mC residue, binding is completely abolished (lanes 1–4). The complex formed with the TATCRE is not competed by unlabelled C/EBP binding sites (Landschulz *et al.*, 1988), but can be competed specifically by either TAT or somatostatin CRE (Montminy *et al.*, 1986) sequences (lanes 5–7).

To control whether 5-azaCR treatment altered the binding activity to the TATCRE, bandshift assays with nuclear extracts prepared from HTC and HTC/AzaCR cells were performed. Both extracts contain TATCRE-binding activity and the complexes formed show identical behaviour with respect to DNA methylation of the oligodeoxynucleotides (lanes 8–11 and 15–18) and in competition experiments (lanes 12–14 and 19–21). From this we conclude that the TATCRE-binding protein (CREB) is sensitive to CpG methylation and that this or a related binding activity is present in HTC and HTC/AzaCR nuclei but excluded from its binding site even after CpG demethylation.

Inaccessibility of DNA is correlated with the presence of a nucleosome

In HTC cells the TAT gene is expressed at low levels and hallmarks of open chromatin—the promoter region and the glucocorticoid response element at −2.5 kb—are adjacent to the closed −1 kb and −3.6 kb regions. Our observations in these cells suggest that the mechanism of protein exclusion must operate locally in chromatin. We therefore compared the nucleosomal organization of these regions in the different cell lines using micrococcal nuclease (MNase). MNase has been used extensively to probe chromatin structure because it cleaves DNA wrapped around a nucleosome core only very poorly as compared with the more accessible linker DNA. Sequences associated with a nucleosome are resistant to extensive digestions and can be recovered as fragments of nucleosome core size (146 bp). Upon partial

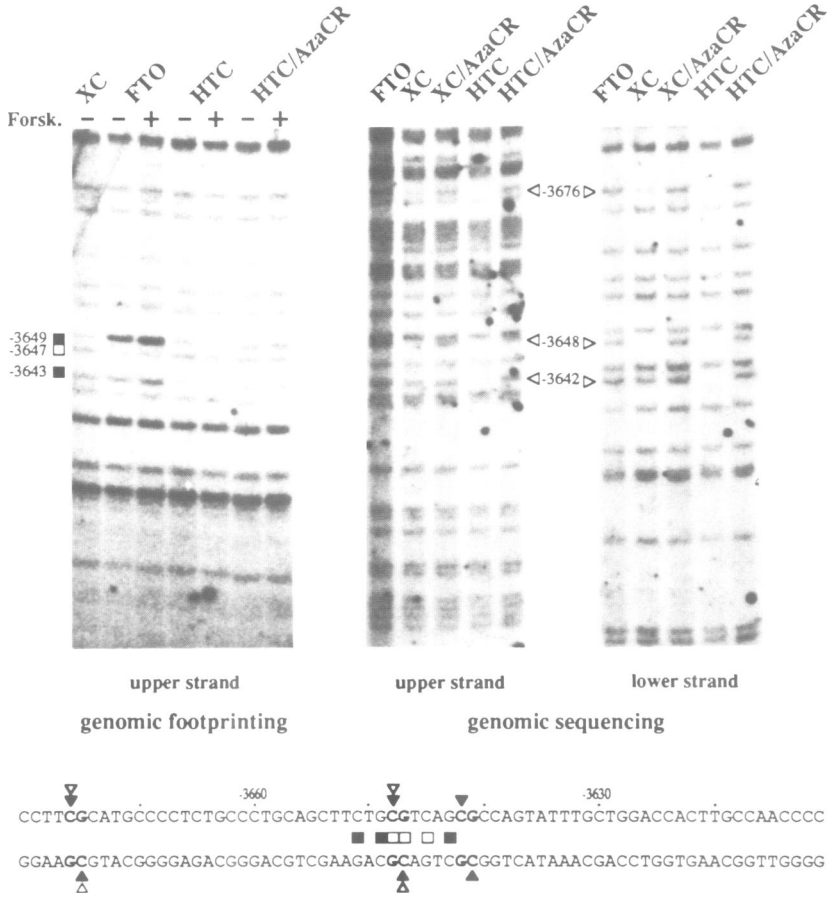

Fig. 4. Complete demethylation at the cAMP-responsive enhancer does not lead to protein–DNA interactions *in vivo*. For genomic footprinting (left panel) cells were either treated with 10 μM forskolin (+) or with 0.1% ethanol as a solvent control (−). After *in vivo* methylation, purified DNA was digested with *Sty*I and cleaved with piperidine at positions of methylated Gs. For genomic sequencing (middle and right panel), protein-free DNA was digested with *Sty*I and cleaved at positions of unmethylated cytosines with hydrazine followed by piperidine treatment. DNA samples were further processed as detailed in Figure 2. Hybridization was performed with a single-stranded DNA probe, recognizing the sequence from −3516 to −3643 (probe HS127, Weih *et al.*, 1990). Altered DMS reactivity of guanosine residues on the upper strand is marked with solid squares for enhancements and with open squares for protections. Numbers indicate the positions in base pairs relative to the start site of transcription. Triangles indicate 5mC residues which become demethylated after 5-azaCR treatment. The characteristic pattern of enhancements and protections at the TATCRE for the upper and lower strand (Weih *et al.*, 1990) is summarized at the bottom (closed triangles: 5mC residues in HTC cells; open triangles: 5mC residues in XC cells).

MNase digestion regularly spaced nucleosomes will give rise to a ladder of fragments representing oligonucleosomes. Isolated nuclei from the various cell lines were digested with MNase to different degrees. The purified DNA was analysed by Southern blotting using a short DNA probe (SH140) located within the −1 kb HS site in FTO-2B cells. The probe labels fragments of nucleosome core size in digests of XC as well as HTC nuclei indicating that the corresponding sequences are associated with a nucleosome in the cells (Figure 6A). Less extensive digestion produces a ladder of fragments corresponding to oligomers of nucleosomes suggesting that the region is part of an array of regularly spaced nucleosomes. This nucleosomal structure is maintained in the corresponding AzaCR populations. In FTO-2B cells, however, no fragments of mononucleosomal size are detected, instead the region is bound by specific DNA-binding proteins.

As a control, the membrane was reprobed with a short DNA fragment (XA163) located 9 kb upstream of the TAT gene transcriptional start site. Fragments up to decanucleosome size are now labelled in the FTO-2B digests (Figure 6B), indicating that this region in FTO-2B cells is organized into nucleosomes. A probe fragment located immediately adjacent to the −1 kb HS site essentially gives the same result (not shown). The hybridization patterns from the other cell lines confirm the extended array of regularly spaced nucleosomes stretching over the region of potential binding sites. Thus, the absence of *in vivo* footprints is correlated with the presence of a nucleosome at the binding sites. Chromatin analysis in the −3.6 kb HS site also demonstrates that the TATCRE region is associated with a nucleosome in XC, HTC and 5-azaCR treated cells. In FTO-2B cells, however, this nucleosomal structure is perturbed, correlating with specific DNA binding of transcription factors (not shown).

Discussion

Our initial observation (Becker *et al.*, 1987) that the inability of nuclear proteins to bind to their recognition sites in chromatin correlated with methylation of these sequences, together with the direct demonstration *in vitro* that a methylated CpG within a binding site prevented the interaction of the corresponding factor, suggested that methylation could be directly responsible for excluding proteins from their sites in some cells and contribute to the repression of gene activity. Other reports of *bona fide* transcription factors for which binding is sensitive to mCpG (Kovesdi *et al.*, 1987; Watt and Molloy, 1988; Iguchi-Ariga and Schaffner, 1989; Comb and Goodman, 1990) as well as the phenomenon of methylation-free islands in the 5′-regions of housekeeping genes (Bird, 1986) provided additional support for this hypothesis. In this report we limit our analysis to a defined question: is methylation at CpG dinucleotides sufficient to exclude ubiquitous proteins from their binding sites *in vivo*? Direct analysis by genomic footprinting allowed us to answer

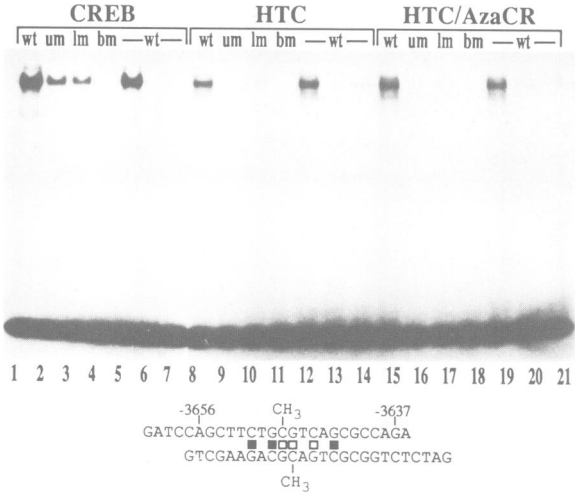

Fig. 5. *In vitro* binding of CREB to the TATCRE is sensitive to cytosine methylation. The oligodeoxynucleotides containing the TATCRE are shown below the figure. Only the version methylated at both strands is depicted ($-CH_3$: 5-methyl group). The *in vivo* DMS footprint is indicated by filled and open squares. Purified rat CREB protein (CREB, lanes 1−7) and nuclear extract protein (10 μg) from HTC cells (HTC, lanes 8−14) or from HTC cells after 5-azaCR treatment (HTC/AzaCR, lanes 15−21) were incubated on ice with 3 fmol of end-labelled double-stranded (ds) oligodeoxynucleotides as detailed in Materials and methods. Lanes 1, 5−7, 8, 12−14, 15 and 19−21: unmethylated ds-oligodeoxynucleotides (wt); lanes 2, 9 and 16: hemimethylated ds-oligodeoxynucleotides (upper strand methylated, um); lanes 3, 10 and 17: hemimethylated ds-oligodeoxynucleotides (lower strand methylated, lm); lanes 4, 11 and 18: both strands methylated (bm). Binding reactions were competed by non-labelled ds-oligodeoxynucleotides, either C/EBP binding site (lanes 5, 12 and 19), TATCRE (lanes 6, 13 and 20), or somatostatin CRE (lanes 7, 14 and 21). Molar ratios of the probe and the competitor were 1:40 (lanes 5−7) and 1:330 (lanes 12−14 and 19−21).

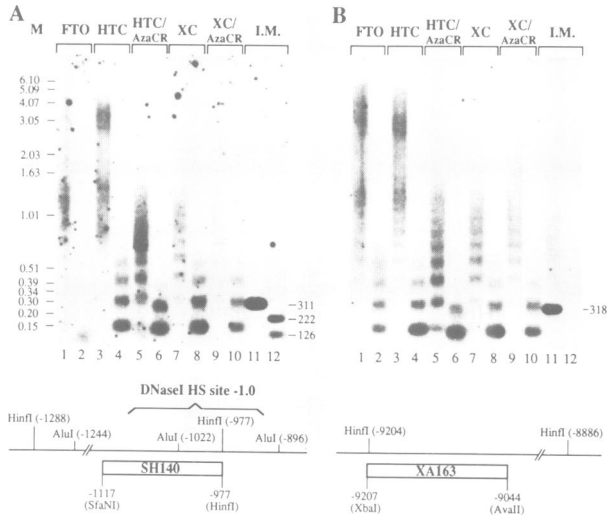

Fig. 6. Nucleosomal structure of the −1 kb region in different cell populations. Comparable amounts of isolated nuclei from the various cell populations as indicated at the top were digested with micrococcal nuclease (lanes 1, 3, 5, 7 and 9: 400 U/ml, 5 min; lanes 2, 4, 6, 8 and 10: 400 U/ml, 50 min). The isolated DNA was separated on a 1.3% agarose gel and transferred to a nylon membrane.
(**A**) Hybridization of the membrane with the probe SH140, specific for the HS site at −1 kb; I.M.: internal markers; lanes 11 and 12: genomic DNA digested with *Hin*fI or *Alu*I, respectively; sizes of resulting fragments are indicated on the right. M: positions of marker fragments (1 kb ladder, BRL). (**B**) Rehybridization of the same membrane with the probe XA163. Lanes are described in (A). *Alu*I sites at positions −9043, −9112 and −9182 evidently produce fragments too small to hybridize efficiently with probe XA163. The probes together with the location of restriction sites used as internal markers and the extent of the HS site at −1 kb are depicted below each panel. Coordinates refer to positions in base pairs upstream of the TAT cap site.

this question. Our results indicate that, at least in the two examples studied, CpG methylation is not the sole mechanism in preventing access of a protein to DNA. Rather, the nucleosomal organization of the potential binding sites implies an important role for chromatin structure in determining the accessibility of DNA. The results also emphasize the necessity of direct *in vivo* analyses to verify conclusions derived from *in vitro* studies.

The available data do not allow a conclusion to be made about what role methylation of DNA sequences may play in gene inactivation. Experiments in which the timing of gene inactivation and accompanying methylation were compared, e.g. during the process of X chromosome inactivation (Lock *et al.*, 1987; Kaslow and Migeon, 1987) or during the developmental inactivation of γ-globin genes (Enver *et al.*, 1988), have shown that DNA modification occurred significantly after the primary inactivation event. In his critical review, Bird (1986) has proposed that methylation could serve as a mechanism to maintain or 'lock in' the repressed state even in the absence of the primary inhibitor signal. Gene inactivation would be initiated by the presence of an inhibitor or the absence of specific activating factors. Vacant protein binding sites may 'invite' *de novo* methylation which then would determine assembly of the gene into inactive chromatin. Our observation that the nucleosomal organization of the potential binding sites in XC and HTC cells at −1 kb and −3.6 kb remains unchanged after 5-azaCR treatment suggests that methylation may participate in the establishment of these closed chromatin configurations but is not required for their maintenance.

A number of reports imply that methylation has a role in the establishment of inert chromatin structures. Using *in vitro* methylated and unmethylated genes in transfection studies, Keshet *et al.* (1986) observed that while unmethylated DNA adopts a DNase I sensitive conformation when stably integrated into the genome, methylated sequences were packaged into nuclease-inaccessible structures. Buschhausen *et al.* (1987) demonstrated that the effect of methylation on transcriptional activity is indirect as chromatin formation is required for methylation-dependent repression to be observed. How could CpG methylation promote an inert chromatin structure? The finding of Ball *et al.* (1983) that mCpG dinucleotides are predominantly located in nucleosomes containing histone H1 and, conversely, the very low amount of H1 in CpG islands compared with bulk chromatin (Tazi and Bird, 1990) provide evidence for a link between methylation and chromatin organization. In addition, a nuclear protein (methyl-CpG binding protein, MeCP) that specifically binds to multiple mCpG dinucleotides may influence nuclease resistance of chromatin and methylation-mediated transcriptional repression *in vitro* and *in vivo* (Antequera *et al.*, 1989; Meehan *et al.*, 1989; Boyes and Bird, 1991). In addition, MeCP could be involved in modifying the nucleosomal structure at positions of methylated DNA.

The regular array of nucleosomes stretching over the −1 kb and −3.6 kb regions of the TAT gene in XC and HTC cells is not perturbed by demethylation and suggests that it is this particular nucleosomal structure which prevents access of DNA-binding proteins. The observation that some transcription factors may bind to their recognition sequences only after removal of positioned nucleosomes has been made previously (Almer *et al.*, 1986; Cordingley *et al.*, 1987; Richard-Foy and Hager, 1987; Fascher *et al.*, 1990; Reik *et al.*, 1991). Clearly other possibilities remain to explain our observation, such as the putative compartmentalization of factors in the nucleus or subtle differences in factor concentration. Further, modification of proteins by post-translational mechanisms or auxiliary factors could alter their DNA-binding activities *in vivo* without affecting *in vitro* protein−DNA interactions. Finally, *in vivo* modifications of histones could prevent nucleosome displacement by DNA-binding proteins. This idea is supported by the demonstration that glucocorticoid receptor-dependent disruption of a specific nucleosome on stably integrated MMTV promoter constructs is prevented by sodium butyrate, a short-chain fatty acid that induces histone hyperacetylation (Bresnick *et al.*, 1990).

The cAMP-responsive and cell type-specific enhancer located 3.6 kb upstream of the TAT transcriptional start site is active and bound by protein in FTO-2B but not in XC or HTC cells (Boshart *et al.*, 1990; Weih *et al.*, 1990). Mutations of the TATCRE in its natural position not only abolish cAMP-inducibility completely but also strongly reduce basal level activity of the −3.6 kb enhancer (Weih *et al.*, 1990). Furthermore, in fibroblasts and in hepatoma × fibroblast hybrids, the CRE is the target for repression of TAT gene transcription by the tissue-specific extinguisher locus *TSE1* (Boshart *et al.*, 1990). High levels of *TSE1* expression are responsible for exclusion of the TATCRE-binding protein (CREB) from its binding site. HTC and HTC/AzaCR cells, however, show identical low levels of *TSE1* expression as compared to FTO-2B cells and therefore this mechanism is not operative in these cell lines (Boshart *et al.*, submitted). When crude nuclear extracts prepared from either FTO-2B or HTC cells were analysed in bandshift experiments, no differences in binding of factors to the TATCRE could be observed (Weih, 1990). The results on protein−DNA interactions shown here clearly indicate that demethylation of CpG dinucleotides, one of which is located in the CRE and interferes with CREB-binding when methylated, does not lead to interaction of the CREB protein or other factors with TAT gene sequences.

In agreement with results obtained by Iguchi-Ariga and Schaffner (1989) with crude nuclear extracts, we demonstrate that CpG methylation of the TATCRE on both strands completely abolished binding of purified CREB protein. Moreover, binding of CREB was strongly reduced when the CRE contains only one methylated cytosine residue on either the upper or the lower strand. Consequently, after replication the hemimethylated sequence would also be a weak binding site for CREB and thus nucleosomes could compete more efficiently. This finally could lead to the establishment of an inert chromatin configuration which then may no longer be dependent on cytosine methylation to be maintained.

Materials and methods

Cell culture and 5-azacytidine treatment of cells

XC cells were grown in Dulbecco modified Eagle medium (DMEM), 10% fetal calf serum, 2 mM glutamine, 100 U/ml penicillin, 100 μg/ml streptomycin, 10 mM HEPES pH 7.4. FTO-2B as well as HTC cells were cultured in 1:1 (v/v) DMEM-HAM F12 instead of DMEM only. For 5-azaCR treatments, cells were grown to confluency, replated at a density of 10^6 (XC) or 2×10^6 (HTC) cells per 25 cm dish and allowed to attach for 24 h. Then the medium was changed and 3 μM 5-azacytidine (Sigma) from a 10 mM stock in 50 mM Tris−HCl pH 8 was added to the culture. Parallel cultures did not receive the drug. After 24 h the medium was changed to medium without the drug and the surviving cells were allowed to grow

to confluency. After replating and growth to confluency the process (in total about 10 days) was repeated.

Genomic footprinting and sequencing

In vivo methylation of cells, purification and hydrazine treatments of genomic DNA as well as preparation of membranes for hybridization were carried out as described (Becker and Schütz, 1988). Synthesis of single-stranded DNA probes of high specific activity by reverse transcription of RNA templates followed the procedure detailed by Weih *et al.* (1988). The probes HP111 and HS127 were cloned into *Sac*I and *Sph*I-digested Bluescribe M13− (Stratagene) after creating blunt ends of vector and insert with T4 DNA polymerase. Identity and orientation were confirmed by sequencing.

Nuclear extracts and DNase I footprinting

Nuclear extracts of XC/AzaCR and HTC/AzaCR cells were prepared in parallel with extracts of untreated control cells as described previously (Becker *et al.*, 1987) with the modification of disrupting the nuclei by sonication. DNase I footprints were performed according to Lee *et al.* (1987) using 50 μg of nuclear extract or BSA and 2.5 μg/ml salmon sperm DNA as unspecific competitor DNA. The binding reaction was performed for 15 min on ice in 10 mM HEPES (pH 7.9), 70 mM KCl, 4 mM $MgCl_2$, 4 mM spermidine, 10% glycerol and 0.1 mM EDTA. Protein binding was assayed on a restriction fragment containing the sequences from 944 to 1225 bp upstream of the TAT transcriptional start as previously described (Becker *et al.*, 1987). Purified DNA was separated on a 6% sequencing gel.

Oligodeoxynucleotides and bandshift assays

Oligodeoxynucleotides were synthesized on an Applied Biosystems synthesizer and purified by HPLC. Complementary oligodeoxynucleotides were denatured and subsequently annealed in 100 mM NaCl, 10 mM Tris−Cl pH 7.5, 1 mM EDTA at a concentration of 1 pmol/μl. For bandshift probes, double-stranded oligodeoxynucleotides were end-labelled with [α-^{32}P]dATP using Klenow enzyme.

wt (unmethylated) TATCRE oligodeoxynucleotides:

G A T C C A G C T T C T G C G T C A G C G C C A G A
G T C G A A G A C G C A G T C G C G G T C T C T A G

um (upper strand methylated) TATCRE oligodeoxynucleotides:

G A T C C A G C T T C T G M G T C A G C G C C A G A
G T C G A A G A C G C A G T C G C G G T C T C T A G

lm (lower strand methylated) TATCRE oligodeoxynucleotides:

G A T C C A G C T T C T G C G T C A G C G C C A G A
G T C G A A G A C G M A G T C G C G G T C T C T A G

bm (both strands methylated) TATCRE oligodeoxynucleotides:

G A T C C A G C T T C T G M G T C A G C G C C A G A
G T C G A A G A C G M A G T C G C G G T C T C T A G

(where M stands for 5-methyl cytosine)

Ten micrograms of nuclear extract protein were incubated with 0.5 μg of sonicated calf thymus DNA for 20 min on ice. To this, 1 μl (3 fmol) of specific end-labelled oligodeoxynucleotide duplex (10 000 c.p.m./μl) was added, and incubation in a total volume of 10 μl was continued for 20 min on ice. Further details of the bandshift reactions are described in Weih *et al.* (1990). Experiments with purified CREB protein were done without unspecific competitor DNA as described elsewhere (Nichols *et al.*, in preparation).

Micrococcal nuclease digestion of nuclei

Nuclei from cultured cells were isolated as described (Becker *et al.*, 1984; Jantzen *et al.*, 1987). Micrococcus nuclease (Boehringer) was added to a final concentration of 400 U/ml and the digestion was performed at room temperature for the times indicated in the figure legend. DNA was purified and further processed as in Jantzen *et al.* (1987) except that a 1.3% agarose gel was used and blotted onto Pall Biodyne B membrane according to the manufacturer's specifications. The probes SH140 and XA163 were cloned into Bluescript M13+ (Stratagene) and used as templates for T3 or T7 polymerase transcription, respectively, by standard procedures (Melton *et al.*, 1984).

Acknowledgements

We are grateful to A.F.Stewart for important and helpful methodical suggestions during the course of the experiment. We thank G.Kelsey as well as A.F.Stewart for critically reading the manuscript. We gratefully acknowledge M.Nichols for purified rat CREB protein, W.Fleischer for synthesis of oligodeoxynucleotides, C.Schneider for secretarial assistance and S.Mähler as well as W.Fleischer for photography. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 229 and the Leibniz Programm and the Fonds der Chemischen Industrie.

References

Almer,A., Rudolph,H., Hinnen,A. and Hörz,W. (1986) *EMBO J.*, **5**, 2689−2696.
Antequera,F., Macleod,D. and Bird,A.P. (1989) *Cell*, **58**, 509−517.
Antequera,F., Boyes,J. and Bird,A. (1990) *Cell*, **62**, 503−514.
Avvedimento,E.V., Obici,S., Sanchez,M., Gallo,A., Musti,A. and Gottesman,M.E. (1989) *Cell*, **58**, 1135−1142.
Ball,D.J., Gross,D.S. and Garrard,W.T. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5490−5494.
Becker,P.B. and Schütz,G. (1988) In Setlow (ed.), *Genetic Engineering*, Plenum Press, New York, Volume 10, pp. 1−19.
Becker,P.B., Renkawitz,R. and Schütz,G. (1984) *EMBO J.*, **3**, 2015−2020.
Becker,P.B., Ruppert,S. and Schütz,G. (1987) *Cell*, **51**, 435−443.
Bird,A.P. (1986) *Nature*, **321**, 209−213.
Boshart,M., Weih,F., Schmidt,A., Fournier,R.E.K. and Schütz,G. (1990) *Cell*, **61**, 906−916.
Boyes,J. and Bird,A. (1991) *Cell*, **64**, 1123−1134.
Bresnick,E.H., John,S., Berard,D.S., LeFebvre,P. and Hager,G.H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3977−3981.
Buschhausen,G., Wittig,B., Graessmann,M. and Graessmann,A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1177−1181.
Cedar,H. (1988) *Cell*, **53**, 3−4.
Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991−1995.
Comb,M. and Goodman,H.M. (1990) *Nucleic Acids Res.*. **18**, 3975−3982.
Cordingley,M.G., Riegel,A.T. and Hager,G.L. (1987) *Cell*, **48**, 261−270.
Doerfler,W. (1983) *Annu. Rev. Biochem.*, **52**, 93−124.
Dynan,W.S. and Tjian,R. (1985) *Nature*, **316**, 774−778.
Enver,T., Zhang,J., Papayannopoulou,T. and Stamatoyannopoulos,G. (1988) *Genes Dev.*, **2**, 698−706.
Ephrussi,A., Church,G.M., Tonegawa,S. and Gilbert,W. (1985) *Science*, **227**, 134−140.
Fascher,K.-D., Schmitz,J. and Hörz,W. (1990) *EMBO J.*, **9**, 2523−2528.
Flatau,E. Gonzales,F.A., Michalowsky,L.A. and Jones,P.A. (1984) *Mol. Cell. Biol.*, **4**, 2098−2102.
Hansen,R.S., Ellis,N.A. and Gartler,S.M. (1988) *Mol. Cell. Biol.*, **8**, 4692−4699.
Harrington,M.A., Jones,P.A., Imagawa,M. and Karin,M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2066−2070.
Höller,M., Westin,G., Jiricny,J. and Schaffner,W. (1988) *Genes Dev.*, **2**, 1127−1135.
Holliday,R. (1987) *Science*, **238**, 163−170.
Hsiao,W.W., Gattoni-Celli,S., Kirschmeier,P. and Weinstein,I.B. (1984) *Mol. Cell. Biol.*, **4**, 634−641.
Iguchi-Ariga,S.M.M. and Schaffner,W. (1989) *Genes Dev.*, **3**, 612−619.
Jantzen,H.M., Strähle,U., Gloss,B., Stewart,A.F., Schmid,W., Boshart,M., Miksicek,R. and Schütz,G. (1987) *Cell*, **49**, 29−38.
Jones,P.A. (1985) *Cell*, **40**, 485−486.
Kaslow,D.C. and Migeon,B.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6210−6214.
Keshet,I., Lieman-Hurwitz,J. and Cedar,H. (1986) *Cell*, **44**, 535−543.
Khan,R., Zhang,X.-Y., Supakar,P.C., Ehrlich,K.C. and Ehrlich,M. (1988) *J. Biol. Chem.*, **236**, 14374−14383.
Killary,A.M. and Fournier,R.E.K. (1984) *Cell*, **38**, 523−534.
Konieczny,S.F. and Emerson,C.P. (1984) *Cell*, **38**, 791−800.
Kovesdi,J., Reichel,R. and Nevins,J.R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2180−2184.
Landschulz,W.H., Johnson,P.F., Adashi,E.Y., Graves,B.J. and McKnight,S.L. (1988) *Genes Dev.*, **2**, 786−800.
Lee,W., Mitchell,P. and Tjian,R. (1987) *Cell*, **49**, 741−752.
Lock,L.F., Takagi,N. and Martin,G.R. (1987) *Cell*, **48**, 39−46.
McKnight,S. and Tjian,R. (1986) *Cell*, **46**, 795−805.
Meehan,R.R., Lewis,J.D., McKay,S., Kleiner,E.L. and Bird,A.P. (1989) *Cell*, **58**, 499−507.

Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T. Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Michalowsky,L.A. and Jones,P.A. (1989) *Mol. Cell. Biol.*, **9**, 885–892.
Montminy,M.R., Sevarino,K.A., Wagner,J.A., Mandel,G. and Goodman,R.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6682–6686.
Nitsch,D., Stewart,A.F., Boshart,M., Mestril,R., Weih,F. and Schütz,G. (1990) *Mol. Cell. Biol.*, **10**, 3334–3342.
Reik,A., Schütz,G. and Stewart,A.F. (1991) *EMBO J.*, **10**, 2569–2576.
Richard-Foy,H. and Hager,G.L. (1987) *EMBO J.*, **6**, 2321–2328.
Saluz,H.P., Jiricny,J. and Jost,J.P. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7167–7171.
Shimada,T., Inokuchi,K. and Nienhuis,A.W. (1987) *Mol. Cell. Biol.*, **7**, 2830–2837.
Svoboda,J. (1960) *Nature*, **186**, 980–981.
Tazi,J. and Bird,A. (1990) *Cell*, **60**, 909–920.
Thompson,E.B., Tomkins,G.M. and Curran,J.F. (1966) *Proc. Natl. Acad. Sci. USA*, **56**, 293–303.
Watt,F. and Molloy,P.L. (1988) *Genes Dev.*, **2**, 1136–1143.
Weih,F. (1990) Ph.D. thesis, Das cAMP-Wirkelement des Tyrosin-Aminotransferase Gens: In vivo und in vitro Analyse von Protein-DNA-Wechselwirkungen, seine funktionelle Charakterisierung und die Rolle der cAMP-abhängigen Proteinkinase. University of Heidelberg, Heidelberg, FRG.
Weih,F., Stewart,A.F. and Schütz,G. (1988) *Nucleic Acids Res.*, **16**, 1628.
Weih,F., Stewart,A.F., Boshart,M., Nitsch,D. and Schütz,G. (1990) *Genes Dev.*, **4**, 1437–1449.
Wigler,M., Levy,D. and Perucho,M. (1981) *Cell*, **24**, 33–40.
Wolffe,A. (1990) *New Biol.*, **2**, 211–218.

Received on September 26, 1990; revised on May 6, 1991