Complex demethylation patterns at Sp1 binding sites in F9 embryonal carcinoma cells

John Silke, Kristina I. Rother, Oleg Georgiev, Walter Schaffner*, Koichi Matsuo

Institut für Molekularbiologie II, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received 23 June 1995; revised version received 12 July 1995

Abstract The ubiquitous transcription factor Sp1 has been implicated in the mechanism which maintains CpG islands methylation-free. Plasmids containing GC boxes (Sp1 sites) were in vitro methylated at every CpG dinucleotide. After stable introduction into F9 embryonal carcinoma cells, we analysed the methylation of the sequence around the GC boxes with bisulphite sequencing. In agreement with restriction site analysis by other labs, we found preferential demethylation at GC box DNA versus control DNA. However, the bisulphite sequencing which permits the analysis of every CpG site on a given DNA molecule, revealed a complex pattern of methylated and unmethylated sites. Upon prolonged culture the pattern became simpler, with most sites demethylated but certain sites being consistently methylated.

Key words: DNA methylation; Transcription factor; Bisulphite PCR

1. Introduction

CpG islands are approximately 500 to 2000 bp long and form a distinctive compartment of the mammalian genome which, in contrast to the bulk of genomic DNA, is not depleted in CpG dinucleotides [1,2,3]. They are associated with approximately half of mammalian genes [4], often in promoter and leader regions [2,5], and are usually methylation free throughout development [6]. How CpG islands maintain this unmethylated state is an open question.

The mammalian transcription factor Sp1 has long been suspected to play a role in maintaining CpG islands methylation free [7]. The 10-bp consensus Sp1 binding site is (G/T)(G/ A)GG(C/A)G(G/T)(G/A)(G/A)(C/T) [8], and its core sequence, GGGCGG or GC box, are commonly found in CpG islands [2]. Recent evidence shows the involvement of Sp1 binding sites in demethylation [9,10]. In embryonic stem (ES) cells, subfragments of the adenine phosphoribosyltransferase (aprt) CpG island that had been methylated by HpaII methylase in vitro became demethylated in an Sp1 site dependent manner [9]. Furthermore, in transgenic mice the aprt CpG island remained methylation free provided it contained functional Sp1 binding sites [9,10]. It was shown by using restriction enzymes, that the probability of demethylation at a given CpG dinucleotide in ES cells decreases with distance from an Sp1 site. Other transcription factors, of course, may also be involved as there are exam-

Abbreviations: EC, embryonal carcinoma; ES, embryonic stem cells; aprt, adenine phosphoribosyl transferase; Hprt, hypoxanthine phosphoribosyl transferase.

ples of CpG islands that remain methylation free but contain no Sp1 sites (A. Bird, personal communication).

Using bisulphite sequencing, we show that, despite the requirement of Sp1 sites for protection of a CpG island from methylation [9,10], the presence of an Sp1 site per se is not sufficient to keep surrounding DNA free from methylation. Furthermore, starting with DNA methylated at all CpGs by SssI methylase, we analysed the demethylation of CpG dinucleotides in a 50-bp window containing GC boxes or mutated GC boxes. We show that the demethylation of this heavily methylated substrate is unexpectedly complex and individual CpGs can remain methylated in the close vicinity of an Sp1 site.

2. Materials and methods

2.1. Plasmid constructs

The GC box oligonucleotide used here contains two tandem Sp1 binding site sequences from the ICP4 promoter of *herpes simplex* Virus. The GC box oligo is:

5'-CGAGCCG<u>GCCCCGCCCA</u>TCCCCGGCCCCGCCCATCG-3',
3'-TCGAGCTCGGCCGGGGGGGGGGTAGGGGCC<u>GGGGCGGGT</u>-AGCAGCT-5',

and the GCmut oligo, which is the same as the GC box oligo except for two point mutations (indicated in lower case) within the Sp1 binding site, is: 5'-CGAGCCGGCaCgGCCCATCCCCGGCaCgGCCCATCG-3', 3'-TCGAGCTCGGCCGtGCCGGGTAGGGGCCGtGCCGGGTAGCAGCT-5'.

These oligonucleotides were inserted into the plasmid GOVEC [11,12] to construct the target plasmids GC-OVEC and GCmut-OVEC.

2.2. Bandshifts

2.3. Bisulphite sequencing

Essentially, the previously published protocol [16] was followed, with the light protection suggested in [17]. 2 μ g of genomic DNA or 1 ng plasmid DNA was denatured with 0.3 M NaOH for 15 min at 37°C and the freshly prepared 3.1 M sodium bisulphite, pH 5, was added directly to the denatured DNA, mixed well and placed under mineral oil. The reaction mixture was then incubated for 12 h at 50°C, boiled for 4 min and reincubated for a further 2 h at 50°C. The samples were then placed on ice. The mineral oil was removed and 6 μ l of glass milk was added to the reaction mix. The samples were rotated slowly at 4°C for 30 min. After washing at 4°C, the glass milk (Geneclean) was allowed to dry for 5 min and then incubated at 50°C in 100 μ l 5 mM EDTA, pH 5 for 2 min. The samples were quickly spun and the supernatant was immediately added to 11 μ l 2 M NaOH and incubated at 20°C for 20 min.

^{*}Corresponding author. Fax: (41) (1) 363 8502.

The reaction was neutralised with 47 μ l 10 M NH₄Ac, precipitated with 300 μ l EtOH and resuspended in 100 μ l TE, pH 8.

2.4. PCR

The reaction volume was 50 μ l, with 80 μ M dNTPs, 200 fM primers, 0.4 U Supertaq (P.H. Stehelin) and the recommended reaction buffer except for 20 mM pH 9 AMPSO and 800 μ M MgCl₂. A hot start was performed at 85°C, followed by 95°C for 1 min 20 s, 58°C for 1 min and 72°C for 1 min for 3 cycles and 95°C for 40 s, 52°C for 50 s and 72°C for 1 min for 28 cycles in a cycler (Omnigene).

Primer sequences were:

AOV-Xba10 5'-ACT CTAGATCACTAAACAAATAC AC-3', AOV-Eco16 5'-TAAGAATTCTTTTTTTTTATTAGGAAGATTT-ATTTT-3',

AOV-Eco26 5-CGGAATTCGTATTTATAATTATGGTATAGGT-3', AOV-Bam20 5'-CGGGATCCACAATTATATCAAAAAACAAAT-3'.

Nucleotides in italics represent nucleotides in the original vector OVEC [12]. Underlined and bold nucleotides represent bisulphite conversion and introduced restriction sites respectively. The amplified sequences were cloned into pKSS [18] following *EcoRI/Xba* I or *EcoRI/BamHI* double digestion.

Sequencing was done with SK primer either on an ABI sequencing machine and *Taq* polymerase or manually using Sequenase (UBS).

2.5. Southern blots

High molecular weight genomic DNA was isolated from a 10 cm diameter plate of F9 cells as described [19]. Blots were hybridised with the ³²P-labelled *Hin*dIII–*Xho*I probe (Fig. 1) and a Phospho-imager (Molecular Dynamics) was used to quantify the intensity of bands. To calculate the percentage of demethylation the formula '% demethylation = 0.9-kb *Hpa*II band intensity/(0.9-kb *Hpa*II band intensity + 4.2-kb *Hin*dIII band intensity)' was used.

2.6. DNA transfections

F9 cells were cultured in DMEM medium supplemented with 10% fetal calf serum (GibcoBRL), 100 U/ml penicillin and 50 μ g/ml streptomycin. The F9 cells were always kept in a non-confluent and undifferentiated state [20]. Stable F9 transfections were performed by co-transfecting 3 μ g of OVEC-based plasmid and 0.3 μ g of the puromycin resistance vector pBSpac Δ P [21] with LipotectAMINE (7 μ l/ml) (GibcoBRL) and then selecting for transformants with 0.4 μ g/ml of puromycin (Sigma) for 10 days. Colonies were pooled. Second round transfections were performed with the neomycin resistance vector pMC1NeoPolyA (Stratagene) and stable transformants were selected for with 0.5 mg/ml of G418 (GibcoBRL).

2.7. In vitro methylation

SssI and HpaII methylases were purchased from New England Biolabs. Mock methylation was performed by omitting the enzyme. Completeness of methylation of plasmid DNA, and oligonucleotides, was checked with HpaII digestion and agarose gel electrophoresis. Plasmid DNAs used for transfection were also checked with bisulphite sequencing.

3. Results

3.1. The factors Sp1 and Sp3 are present in F9 cells and bind to a GC box regardless of methylation

In bandshift experiments using nuclear extracts from F9 cells, at least three distinct protein–DNA complexes were detected with the GC-box oligo but only weakly with the mutated GC box (Fig. 1B). The specificity of the three bands was demonstrated with competition by excess GC box oligonucleotide (data not shown). To identify the nature of the bands we performed a supershift experiment using antibodies raised specifically against the transcription factors Sp1 or Sp3 [15], with both F9 and control Hela nuclear extracts. In both extracts, the slowest migrating band contained Sp1 and the two faster migrating complexes contained Sp3. Therefore, the transcription

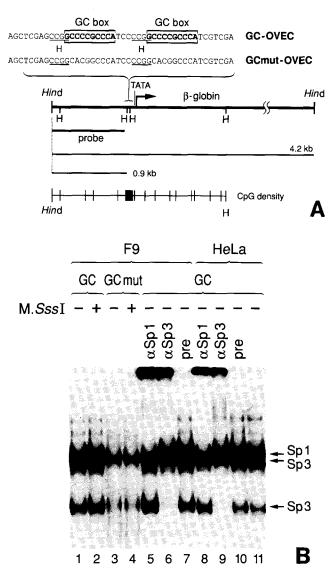


Fig. 1. Schematic representation of the eukaryotic segment of the plasmids, GC-OVEC and GCmut-OVEC. (A) The 0.84-kb HindIII-XhoI probe and bands visualized with Southern blotting are shown below the restriction map. TATA box and the transcription start are indicated. Hind: HindIII, H: HpaII/MspI. In the lower part of the figure a CpG density plot of a 1.85-kb region containing the GC box or GCmut sequences is depicted. (B) Bandshift performed with GCmut oligonucleotide and an oligonucleotide containing a single GC box. Similar results were obtained with the GC box oligo that contained two binding sites. Lanes 1-7, F9 nuclear extract; lanes 1 and 2 contain Mock and SssI methylated end-labelled GC box oligo respectively; lanes 3 and 4 Mock and SssI methylated GCmut oligo, respectively; lanes 5-11 a nonmethylated GC box oligo. Lanes 5-10 contain specific antibody to Sp1(α Sp1), Sp3 (α Sp3), or pre-immune serum (pre) as indicated. Lanes 8-11, HeLa nuclear extract. The lowest fastest migrating band that was shifted by the anti-Sp3 antibody could sometimes be resolved into 2 bands.

factors Sp1 and Sp3 are present in F9 cells and bind specifically to the GC box oligonucleotide whether it is methylated or not. An incidental finding of this study was that the binding of transcription factor Sp3 is not inhibited by methylation of its binding site (Fig. 1B, lane 2), similarly to previous findings with the transcription factor Sp1 [7,22].

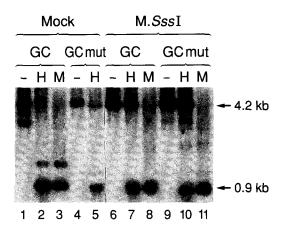


Fig. 2. Southern blot analysis of genomic DNA recovered from F9 cells stably transformed with either mock or SssI methylated GC-OVEC, or mock or SssI methylated GCmut-OVEC. Sizes are indicated in kb. -: HindIII digestion alone, H: HindIII and HpaII double digestion, M: HindIII and MspI double digestion. Lanes 1–3 genomic DNA recovered from F9 cells stably transfected with mock methylated GC-OVEC, lanes 4–5 mock methylated GCmut-OVEC, lanes 6–8 SssI methylated GC-OVEC, lanes 9–11 SssI methylated GCmut-OVEC.

3.2. Demethylation of SssI in vitro methylated Sp1 sites determined by Southern analysis

F9 cells were stably transfected with either SssI methylated GC-OVEC, SssI methylated GCmut-OVEC, mock methylated (i.e. unmethylated) GC-OVEC, or mock methylated GCmut-OVEC, as described in section 2. Genomic DNA was purified and digested with HindIII alone or in combination with either the methylation sensitive HpaII or methylation insensitive MspI.

The percentage of demethylation was quantified using a phospho-imager as described in section 2. The *HpaII* digestion revealed that the SssI methylated GC-OVEC became demethylated by 41%, (Fig. 2, lane 7), whereas only 23% of the GCmut oligo was demethylated, (Fig. 2, lane 10). Therefore, in agreement with previous observations [9], intact GC boxes indeed enhanced demethylation. Further, the Southern blot shows that GC boxes inhibited de novo methylation as the mock methylated GC-OVEC became de novo methylated to 30% but the GCmut-OVEC was de novo methylated to 60% (Fig. 2, lanes 2 and 5). Since both the GC box and GCmut oligonucleotides contain two *HpaII* sites in close proximity to each other (Fig. 1A), digestion at one site cannot be distinguished from digestion at the other. Thus, to observe the 0.9-kb band that shows whether demethylation of SssI methylated DNA has occurred in the GC box oligonucleotides, only one of the two sites need to be demethylated. On the other hand, to observe the HpaII resistant 4.2-kb band that reveals de novo methylation of the originally unmethylated plasmid, both sites must become de novo methylated. The Southern blot therefore gives an indication of GC box function in promoting demethylation and partially preventing de novo methylation.

3.3. Bisulphite sequencing of GC boxes

The samples that had been analysed by Southern blot were examined by bisulphite sequencing [23], which allows the positive identification of methyl cytosines from individual genomic DNA molecules [23]. Consistent with the results that we obtained by Southern blotting, demethylation in the GC box con-

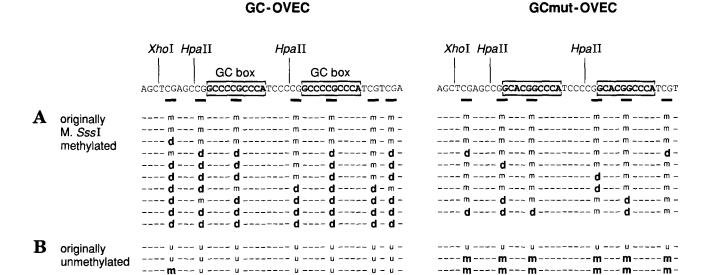
taining construct was high, with approximately 60% of CpGs becoming demethylated, whereas a significantly lower amount was observed in the mutated GC box containing construct (20% of CpGs, Fig. 3A). The individual promoters did not reveal any noteworthy pattern of demethylation, with the promoters showing many possible combinations of remaining methyl CpGs. Consistent with the Southern blot, the mock-methylated, i.e. unmethylated GC-OVEC construct became only sporadically de novo methylated, in contrast to the mock methylated GCmut-OVEC construct which, in different cell clones, either remained unmethylated or became de novo methylated to a high extent, (Fig. 3B).

To examine the effect of prolonged culture on demethylation, we further studied the methylation status of the SssI methylated GC-OVEC constructs after a further 2 weeks of culture. To avoid any possible bias, which might have arisen due to a small number of clonal cells outgrowing other cells, we selected a sub-population of the cells containing SssI methylated GC-OVEC by the transfection of a neomycin resistance gene cassette followed by G418 selection for a period of 2 weeks. Over 300 resistant colonies from three separate transfections were pooled and genomic DNA was bisulphite sequenced. The extent of demethylation was now greater with 75% of CpGs becoming demethylated (Fig. 3C); this was also consistent with a Southern blotting (data not shown). The pattern of the remaining methyl CpGs was, however, surprising. Even though the starting methylation status of the integrated plasmid before this second round of selection was in all probability diverse, the methylation status after this further round of selection was much more homogeneous. In particular, the site 3' to the promoter was uniformly demethylated and 2 CpGs, one actually between the 2 GC boxes, were consistently methylated.

4. Discussion

Although many immortalised cell lines have a proportion of methylated CpG islands, the CpG islands of embryonic carcinoma (EC) and embryonic stem (ES) cell lines obey the general rule that CpG islands are unmethylated [24]. It is consistent with analyses performed in EC and ES cells [25,26], and results obtained in developing mice [6], that the unmethylated state of CpG islands is established during a critical period of high demethylation activity that occurs early in development. The establishment and maintenance of the unmethylated state may require an activity that can recognise methyl CpGs in CpG islands and remove them. EC cells can efficiently convert methyl CpGs present at low density in CpG islands to unmethylated CpG [27], probably via an excision repair activity [28]. Such situations, where a CpG island contains a low level of methylation, would occur if sites in a CpG island became accidentally methylated.

Transcription factors have recently been implicated in the establishment and maintenance of the unmethylated state of CpG islands [9,10]. Sp1 has long been suspected to play such a role [7] because it can bind to a GC box regardless of methylation; and we show incidentally that Sp3 also appears to be insensitive to methylation of its binding site. Transcription factors might establish or maintain CpG islands by, among other possibilities, merely excluding the DNA methyltransferase, or alternatively by creating a secondary molecular flag, such as ordered chromatin [10], which either excludes methyltrans-



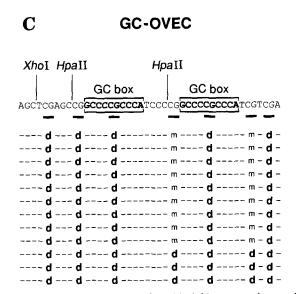


Fig. 3. Schematic representation of the bisulphite sequencing results. (A) Sequences of promoters from SssI in vitro methylated plasmids, GC-OVEC and GCmut-OVEC, recovered from stably transfected F9 cells. m, CpG dinucleotide remained methylated, d, CpG dinucleotide became demethylated. (B) Sequences of promoters from mock methylated plasmids, GC-OVEC and GCmut-OVEC, recovered from stably transfected F9 cells. u, CpG dinucleotide remained unmethylated, m, CpG dinucleotide became de novo methylated. (C) Sequences of GC-OVEC promoters originally in vitro methylated with SssI methylase and recovered from stably transfected F9 cells further selected with G418. m, CpG dinucleotide remained methylated, d, CpG dinucleotide became demethylated.

ferase or attracts a demethylase activity. Whatever the mechanism is, it is still not clear, as in the case of the mouse aprt gene promoter, how it is that three Sp1 sites clustered at the 3' end of the CpG island maintain a methylation free zone of 1-kb upstream [10]. Our results emphasise that the GC box induced demethylation observed in the EC cell system does not encompass every CpG in the neighbourhood of GC boxes, and individual promoters may contain different patterns of methyl groups 10 days after transfection, although after a further two

weeks a quite uniform pattern of methylated and unmethylated sites is observed. In our experiment it was difficult to assess the role of transcription itself, as opposed to mere binding of transcription factor, as the transcription from GC-OVEC was not significantly above the basal level (data not shown).

High density methylated substrates do exist naturally [29], in the dosage compensated X-chromosome and become demethylated, probably during the transition from oogonia to oocyte [30]. It is interesting to note that F9 EC cells can reactivate the inactive X-chromosome of mice thymocytes [31] when fused as somatic cell hybrids. This experiment is not immediately reconcilable with another in F9 cells that shows that an SssI methylated CpG island, with a high density of in vitro introduced methyl CpGs, is very inefficiently demethylated when stably transfected into F9 cells [27]. A possible reconciliation lies in the observation that not all CpGs in a CpG island need necessarily be methylated on the inactive X-chromosome; with the frequency of methylation at each of the 32 CpGs in the hprt CpG island varying from 3-100% on the inactive X-chromosome [29]. Thus, although SssI methylase has the same CpG dinucleotide specificity as the mammalian methyltransferase, a CpG island methylated at every CpG might not provide a real substrate for a demethylation activity in vivo. We were able to observe significant demethylation of SssI methylated DNA, however, and this may be because our constructs were designed so that the GC boxes were surrounded by DNA with a low G + C%, (20%), and low CpG content, (O/E = 0.2), see Fig. 1A.

EC and ES cells paradoxically contain both a high demethylation activity and a high de novo methylation activity [32,33]. We show that two GC boxes can promote demethylation, and can partially protect these short sequences themselves from the high de novo methylation observed in F9 cells. Furthermore we show that mutated GC boxes were less able to promote low level demethylation, and at the same time, were less able to protect themselves from the high de novo methylation activity. It is not obvious whether a GC box can only enhance demethylation or can also directly interfere with de novo methylation. Whatever the mechanism, it is clear that although Sp1 sites

are required to maintain the CpG island of the aprt gene in a methylation free state [9,10], the presence of Sp1 binding sites per se does not cause complete demethylation or complete protection of surrounding DNA from methylation. Further analysis is necessary to determine whether a completely unmethylated or completely methylated substrate, transfected into F9 cells, over a long culture period eventually converge to obtain the same pattern of methylcytosines which may be characteristic for any particular sequence.

Acknowledgements: We thank Prof. Guntram Suske for the kind gift of specific antibodies to Sp3 and Sp1, Dr. Susan Clark, Dr. Tony Raizis, and Martin Kopp for technical advice concerning the bisulphite reaction. We also thank Drs. Donald Macleod and A. Bird for providing us with unpublished information.

References

- [1] Bird, A.P. (1987) Trends Genet. 3, 342-347.
- [2] Gardiner, G.M. and Frommer, M. (1987) J. Mol. Biol. 196, 261–282
- [3] Matsuo, K., Clay, O., Takahashi, T., Silke, J. and Schaffner, W. (1993) Somat. Cell Mol. Genet. 19, 543-555.
- [4] Antequera, F. and Bird, A. (1993) Proc. Natl. Acad. Sci. USA 90, 11995–11999.
- [5] Larsen, F., Gundersen, G., Lopez, R. and Prydz, H. (1992) Genomics 13, 1095–1107.
- [6] Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H. and Razin, A. (1992) Genes Dev. 6, 705-714.
- [7] Höller, M., Westin, G., Jiricny, J. and Schaffner, W. (1988) Genes Dev. 2, 1127–1135.
- [8] Courey, A.J., and Tjian, R., (1992) in: Transcriptional Regulation (McKnight, S.L. and Yamamoto, K.R. eds) Book 2, pp. 743-769, Cold Spring Harbor Laboratory Press.
- [9] Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. and Cedar, H. (1994) Nature 371, 435–438.
- [10] Macleod, D., Charlton, J., Mullins, J. and Bird, A.P. (1994) Genes Dev. 8, 2282–2292.

- [11] Seipel, K., Georgiev, O. and Schaffner, W. (1992) EMBO J. 11, 4961–4968.
- [12] Westin, G., Gerster, T., Muller, M.M., Schaffner, G. and Schaffner, W. (1987) Nucleic Acids Res. 15, 6787–6798.
- [13] Schreiber, E., Matthias, P., Muller, M.M. and Schaffner, W. (1989) Nucleic Acids Res. 17, 6419.
- [14] Westin, G. and Schaffner, W. (1988) EMBO J. 7, 3763-3770.
- [15] Hagen, G., Muller, S., Beato, M. and Suske, G. (1994) EMBO J. 13, 3843–3851.
- [16] Clark, S.J., Harrison, J., Paul, C.L. and Frommer, M. (1994) Nucleic Acids Res. 22, 2990–2997.
- [17] Feil, R., Charlton, J., Bird, A.P., Walter, J. and Reik, W. (1994) Nucleic Acids Res. 22, 695–696.
- [18] Kast, P. (1994) Gene 138, 109-114.
- [19] Hogan, G., Constantini, F., Lacy, E., (1986) in: Manipulating the mouse embryo: a laboratory manual, pp. 174, Cold Spring Harbor Laboratory Press.
- [20] Bestor, T.H., Hellewell, S.B. and Ingram, V.M. (1984) Mol. Cell Biol. 4, 1800–1806.
- [21] de la Luna, S., Soria, I., Pulido, D., Ortin, J. and Jimenez, A. (1988) Gene 62, 121-126.
- [22] Harrington, M.A., Jones, P.A., Imagawa, M. and Karin, M. (1988) Proc. Natl. Acad. Sci. USA 85, 2066–2070.
- [23] Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L. and Paul, C.L. (1992) Proc. Natl. Acad. Sci. USA 89, 1827–1831.
- [24] Antequera, F., Boyes, J. and Bird, A. (1990) Cell 62, 503-514.
- [25] Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A. and Cedar, H. (1991) Nature 351, 239-241.
- [26] Szyf, M., Tanigawa, G. and McCarthy, P.J. (1990) Mol. Cell Biol. 10, 4396–4400.
- [27] Choi, Y.C. and Chae, C.B. (1993) Mol. Cell Biol. 13, 5538-5548.
- [28] Jost, J.P. (1993) Proc. Natl. Acad. Sci. USA 90, 4684-4688.
- [29] Park, J.G. and Chapman, V.M. (1994) Mol. Cell Biol. 14, 7975–7983.
- [30] Grant, S.G. and Chapman, V.M. (1988) Annu. Rev. Genet. 22, 199–233.
- [31] Takagi, N., Yoshida, M.A., Sugawara, O. and Sasaki, M. (1983) Cell 34, 1053–1062.
- [32] Gautsch, J.W. and Wilson, M.C. (1983) Nature 301, 32-37.
- [33] Stewart, C.L., Stuhlmann, H., Jahner, D. and Jaenisch, R. (1982) Proc. Natl. Acad. Sci. USA 79, 4098–4102.