Volume 285, number 2, 155–159 FEBS 09972 July 1991

ADONIS 0014579391006855

Minireview

DNA methylation and chromatin structure

Joe Lewis and Adrian Bird

Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, UK

Received 16 May 1991

Studies of the whole genome by molecular and cytogenetic methods have implicated DNA methylation in the formation of 'inactive chromatin'. This has been confirmed by analysis of specific endogenous sequences, and has been mimicked by introducing methylated and non-methylated sequences into cells. As well as affecting chromatin structure, DNA methylation also represses transcription. A protein (MeCP) which binds specifically to methylated DNA has been identified. The properties of MeCP could account for the effects of DNA methylation on both chromatin and transcription.

DNA methylation; 5-Methyl cytosine; Chromatin; Heterochromatin; Methyl-CpG binding protein; CpG islands

1. INTRODUCTION

The organisation of DNA in eukaryotic nuclei is only partially understood. The basic repeating subunit of chromatin, the nucleosome, has been extensively studied and its structure solved by X-ray crystallography. This, however, has shed little light on how higher order structures are formed. Nuclear chromatin is heterogeneous in structure as defined by a number of criteria. For example, its conformation can be probed and classified into categories according to its nuclease sensitivity. Regions of the genome containing active genes are more sensitive to nuclease digestion than non-transcribed regions. Heterogeneity is also detectable by microscopy. Transcriptionally inert regions of the nucleus stain differentially and hence are termed heterochromatin. The underlying molecular mechanisms for these differences are unknown.

At the DNA level, the genome can be divided into two categories containing methylated or non-methylated DNA. In animal genomes, the modified base 5-methylcytosine (m^5C) is restricted to the dinucleotide CpG. In mammals approximately 70% of all CpGs are methylated, the remaining CpGs being non-methylated. DNA containing non-methylated CpGs comprises about 1–2% of the genome. This means that the vast majority of the chromatin is associated with methylated DNA. In this review we discuss the evidence that differential methylation can be a determinant of higher order structure.

Correspondence address: J. Lewis, Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, UK.

2. ORGANISATION OF CpGs IN BULK CHROMATIN

The first indication that methylated DNA was in a specific chromatin structure came from experiments by Razin and Cedar [1]. Mouse L cell nuclei were digested with micrococcal nuclease (MNase) and the m^5C content of the solubilised chromatin analysed. It was observed that m^5C in chromatin was relatively resistant to degradation by MNase. After extensive digestion, 50% of the DNA remained unsolubilised, but this fraction contained over 70% of the total m^5C.

In a subsequent paper Solage and Cedar [2] characterised the kinetics of chromatin digestion and the copy number of associated DNA sequences more thoroughly. Nuclear DNA of L cells was labelled in vivo at thymidine or m^5C residues and the nuclei were then isolated. After varying degrees of digestion with MNase, the amount of solubilised thymidine and m^5C was analysed. Early time points showed that although thymidine was readily released, m^5C was not detected until 15% of the genomic DNA had been solubilised. In addition, m^5C was released with slower kinetics than thymidine. These experiments showed that DNA containing m^5C is organised in chromatin that is relatively resistant to nucleases such as MNase.

Keshet and Cedar [3] have provided a direct demonstration that methylation can cause DNAs to adopt a nuclease insensitive structure. When methylated constructs were transfected into L cells and integrated into the genome, it was shown that they were nuclease resistant. By comparison, unmethylated constructs were found to be relatively hypersensitive to

nucleases. This showed that artificially introduced DNAs were able to adopt a chromatin structure specified by their methylation status.

In an attempt to understand the mechanism underlying the relative nuclease insensitivity of m^5C-containing chromatin, Ball et al. [4] fractionated nucleosomes on agarose gels after MNase treatment. Under the conditions used, nucleosomes could be resolved into histone H1-containing and H1-deficient fractions. Using an antibody directed against m^5C, they assayed for the presence of this modified base in the DNA associated with the different classes of nucleosomes. They observed that nucleosomes which were associated with histone H1 were significantly enriched for m^5C. At least 80% of the m^5C was located in nucleosomes containing H1. Nucleosomes which lacked H1 had levels of m^5C reduced to about 1.6–2.3-fold below the expected level. They proposed on the basis of this evidence that the preferential association of H1 with nucleosomes containing methylated DNA may contribute to alterations in nuclease sensitivity.

In our laboratory, Antequera et al. [5] have shown that the accessibility of sites in nuclei for CpG enzymes depends on their methylation status. This result was unexpected because these enzymes normally cut naked DNA irrespective of methylation. On digestion of bulk chromatin, it was demonstrated that *MspI* and *TthI* are preferentially blocked from cutting methylated sites under conditions of complete digestion. The simplest explanation of these results is that there are factors which bind to m^5CpGs, preventing cleavage of methylated sites. In contrast, bulk chromatin is extensively cleaved by enzymes which do not contain a CpG in their recognition sequence. This suggests the observed differential sensitivity was specific to m^5CpGs.

3. CpG ISLAND CHROMATIN

The chromatin structure of non-methylated CpGs has been studied by utilising some specific properties of CpG islands. These are non-methylated regions of DNA about 0.5–2 kb long which are characteristically GC-rich and which contain CpGs at the frequency expected from the base composition [6]. They are found associated with the 5′ domains of all housekeeping genes and a small proportion of tissue specific genes. Since the CpG islands usually contain the promotors of housekeeping genes which are by definition transcriptionally active, it follows that islands should represent 'active chromatin'. In agreement of this hypothesis, analysis of two X-linked genes, hypoxanthine phosphoribosyltransferase (HPRT) and glucose 6-phosphate dehydrogenase (G6PD), has shown that their CpG islands are hypersensitive to nucleases [7] such as *MspI*, DNase I and S_1 nuclease on the active X chromosome (X_a). Tazi and Bird [8] used the sensitivity of CpG islands to CpG enzymes in order to develop a method for isolating CpG island chromatin. The method was based on the earlier observation of Antequera et al. [5] that complete digestion of nuclei with CpG enzymes gives short oligonucleosomes derived from CpG islands, but leaves bulk chromatin intact. This means that transcriptionally active chromatin can be very highly enriched and its protein composition analysed. The approach circumvented a major drawback of previous studies on active chromatin, since complete rather than partial nuclease digestions were used and cross-contamination was therefore minimal. On analysis of the digestion products, several differences between CpG island chromatin and bulk chromatin were apparent. In addition to the short oligonucleosomes observed on end-labelling of the CpG enzyme released fraction, a fast migrating species of DNA, below mononucleosomal size, was observed. The very small fragments were not observed on digestion of bulk chromatin, implying that they were specific to CpG island chromatin. The most likely explanation for the occurrence of these fragments was the presence of a nucleosome-free region in most, if not all, constitutively transcribed CpG islands. This was confirmed by Southern hybridisation of the fractionated fragments to a *c-fos* probe. The results showed that fragments in this fraction came from a region of the *c-fos* promoter which coincided with a DNAse I hypersensitive site, and therefore most probably lacks a nucleosome.

To investigate the associated proteins, nuclear digestion products of CpG and non-CpG enzymes were fractionated away from uncut chromatin and then enriched for histones. On analysis of the protein components it was striking that the nucleosomes obtained by CpG enzymes were deficient in histone H1 (approximately 10% of the level observed for bulk chromatin). In addition to this, histones H3 and H4 were found to be highly acetylated in comparison to those in bulk chromatin. Acetylation of these two histones has been previously associated with active chromatin structures. For example, Hebbes et al. [9] have shown that actively transcribed regions of chromatin can be significantly enriched by immunoprecipitation using an anti-acetylated H4 antibody. In summary it seems most likely that the presence of highly acetylated H3 and H4, deficiency of H1 and a missing nucleosome in CpG islands may all be a consequence of the presence within these sequences of constitutively active promoters.

4. EFFECTS OF DNA METHYLATION IN HETEROCHROMATIN

Different chromatin structures can be distinguished under the light and electron microscopes. Densely staining regions in the nucleus correspond to heterochromatin, which may be either constitutive or facultative. Constitutive heterochromatin often contains highly repeated DNA sequences, whereas

facultative heterochromatin arises from inactivation of previously euchromatic regions. One of the best known examples of facultative heterochromatin is the inactivation of one X chromosome during development of the mammalian female. There is reason to believe that DNA methylation is relevant to the formation of both facultative and constitutive heterochromatin.

In early embryogenesis, one of the X chromosomes undergoes inactivation by heterochromatisation and the CpG islands become methylated [10]. At this stage the chromosome is visible cytologically in metaphase spreads of female chromosomes as a densely staining structure known as the Barr body. Concurrent with inactivation of the X chromosome is a shift of its replication time to late in the S phase of the cell cycle. There is evidence that methylation may play a role in the shift of the replication time of the chromosome [11]. When cells containing an inactive X chromosome (X_i), are grown in 5-azacytidine, a potent inhibitor of DNA methylation, the inactive X becomes progressively demethylated and its replication time shifts to early S phase. This demonstrates that artificial alteration of chromosomal methylation levels can lead to changes in replication timing.

Further support for the idea that DNA methylation is involved in determining replication timing comes from studies on satellite DNA. These sequences which are present at very high copy number in most vertebrates are normally heavily methylated. For example, the mouse major satellite comprises a 234 bp repeat unit reiterated approximately one million times per haploid genome in *Mus musculus*. Since a repeat unit contains 8 m^5CpGs on average, it follows that approximately 50% of mouse m^5C is located in satellite. In keeping with this, staining with anti-m^5C antibodies localises to the centromeric C-banding regions of constitutive heterochromatin in mouse [12], which is also the location of the major satellite sequences.

Although the major satellite is normally heavily methylated and late replicating, it is undermethylated at early stages in development. In F9 embryonal carcinoma cells, which are derived from an early developmental stage, the replication time of the hypomethylated major satellite is shifted to early S phase [13]. Treatment of the RAG mouse fibroblast cell line, in which the satellite is hypermethylated, with 5-azacytidine causes a dramatic shift of the satellite replication time from very late to the middle of S phase. Thus as in the case of X_i, the evidence suggests that methylation plays a causal role in keeping these sequences late replicating.

Another characteristic in common between the X_i and satellite heterochromatin is reduced sensitivity to nucleases. For example, the X_i is less sensitive to nuclease digestion than the X_a indicating an alternative chromatin structure. Using DNase I directed nick translation on specially prepared metaphase spreads, Kerem et al. [7] were able to differentiate the X_a from the X_i due to extensive DNase I sensitive regions in the former. By comparison, the X_i was largely untouched by this nuclease except for a small region known to be early replicating. It has also been shown by Jablonka et al. [11] that the relative insensitivity of the X_i to nuclease digestion could be reversed by inhibiting methylation with 5-azacytidine. This treatment increased the nuclease sensitivity of X_i to levels comparable with X_a.

The highly methylated major satellite of mouse is also nuclease insensitive. Solage and Cedar [2], in their study of MNase digestion kinetics and m^5C release, analysed the types of sequences that were digested. Utilising C_0t analysis, they showed that the fraction of the chromatin most resistant to nuclease digestion originated from the highly methylated satellite and contained approximately 50% of the m^5C. In contrast to the X_i results, however, these sequences were not nuclease sensitive in F9 cells although they were hypomethylated. Moreover, creation of the undermethylated state with 5-azacytidine did not result in increased nuclease sensitivity. The differences in the nuclease sensitivity of the hypomethylated X and undermethylated satellite may reflect differences in transcriptional activity. Transcription of activated genes on the hypomethylated X would be expected to maintain the chromatin in an active conformation. The satellite sequences, being non-transcribed, may adopt an inactive chromatin structure which is resistant to nucleases regardless of methylation.

5. CpG ACCESSIBILITY AT SPECIFIC SEQUENCES

CpG islands are normally non-methylated irrespective of transcriptional activity. There are however a few exceptions to this general rule. In mammalian females, one of the two X chromosomes is inactivated as part of the dosage compensation mechanism and the island genes are then methylated. In addition, extensive methylation of some CpG islands is observed in tissue culture cell lines, especially those which have been in culture for many years [14]. In keeping with the results described in the previous section, the inaccessibility of methylated DNA in chromatin to nucleases is observed at specific sequences. Wolf and Migeon [15] analysed nuclease sensitivity of DNA sequences surrounding the X-linked HPRT and G6PD genes. They found that sequences on X_a were hypersensitive to digestion at *Msp*I sites. In contrast, sequences on the X_i, in which the CpG islands are methylated and inactive, were not digested at all. Similar blockage of *Msp*I sites in nuclei was found by comparing the accessibility of specific methylated and non-methylated *Msp*I sites in the α-globin locus [5]. The α-globin CpG island in HeLa cells has become extensively methylated during passage in

tissue culture. It was shown that none of the methylated CpGs were cleaved significantly in nuclei, whereas the unmethylated sites at the locus were cut. This and above examples demonstrated that specific methylated sites in nuclei are resistant to *Msp*I. This contrasts with the relatively unhindered access of non-CpG enzymes to regions of methylated DNA. The conclusion is that m^5C-containing DNAs adopt a chromatin structure which protects the m^5CpGs, but not neighbouring sequences, from nuclease digestion.

This conclusion is further supported by studies on nuclease sensitivity of marsupial inactive X chromosomes [16]. Unlike placental mammals, X chromosome inactivation and heterochromatisation in marsupials is not associated with CpG island methylation. Correspondingly the X_a and X_i are equally accessible to *Msp*I, showing that heterochromatinisation alone is not sufficient to block *Msp*I cleavage. Methylation of CCGG sites, as occurs in placental mammals, is also required.

6. TRANSCRIPTIONAL INHIBITION BY m^5CpGs

There is a strong correlation between the presence of DNA methylation at promoter sequences and transcriptional inhibition. For example, DNA from the X_i is not able to transform $HPRT^-$ cells to $HPRT^+$ [17]. However, DNA from the X_a, or from an X_i chromosome which had previously been treated with 5-azacytidine can transform the recipient cells to $HPRT^+$. Inhibition of transcription by methylation has also been confirmed by in vitro methylation of specific genes and transfection of these constructs into cells. Busslinger et al. [18] showed that complete methylation of the human γ-globin gene could inhibit transcription. In addition, these authors showed that methylation in the 5′ promoter region was sufficient to prevent transcription, since methylation outside this region had no inhibitory effect. Murray and Grosveld [19] extended these results to show that the inhibition was not dependent on methylation of specific sites in the promoter. The presence of m^5CpGs in different regions of the promoter was sufficient to inhibit transcription.

There is a variety of evidence which suggests that methylation exerts its effects through alterations in chromatin structure rather than by interfering with the transcription machinery directly. The simplest model invokes a protein or proteins which can bind specifically to methylated DNA thereby excluding transcription factors. Binding of such a protein would also be responsible for the other effects associated with DNA methylation, such as nuclease resistance and late replication.

Evidence for the above model is accumulating from several directions. For example, upon transfection of various methylated and non-methylated constructs into mouse L cells, the methylated constructs assemble into nuclease insensitive chromatin [3].

Furthermore, Buschhausen et al. [20] have shown that inhibition of transcription of *Hpa*II methylated thymidine kinase (TK) constructs was an indirect effect. On assaying the TK RNA levels from the injected methylated and non-methylated constructs, methylated TK was just as transcriptional active as a mock methylated control. After 8 h, however, a sharp drop in expression levels was observed from the methylated construct. This timing coincided with the assembly of transfected DNAs into chromatin. If constructs were assembled in vitro into chromatin prior to injection, transcription from the methylated template was inhibited immediately. Thus, in this case, methylation alone is not sufficient to inhibit transcription. Interaction with nuclear components is first required for transcriptional inhibition to be observed.

7. METHYLATED DNA BINDING PROTEINS

With the above experiments in mind, we have attempted to identify proteins that bind specifically to methylated DNA and may therefore mediate the effects of methylation on chromatin and transcription. Although a methylated DNA binding protein, MDBP, has been previously described [21] it is highly sequence-specific. Due to the sequence complexity of methylated DNA it seems unlikely that a protein like MDBP plays a general role in exerting the effects of methylation. We have identified a nuclear protein from mouse which is characterised by its ability to complex specifically with DNA containing multiple m^5CpG sites independently of the sequence context of those sites [22]. This protein, which we have named 'methylated CpG binding protein' (MeCP), is by virtue of its relaxed sequence specificity and ubiquitous tissue distribution, a good candidate for mediator of the effects of methylation. MeCP is found in nuclear extracts from various mammals that have been tested and has been shown to require a minimum of 15 m^5CpGs per substrate molecule for strong binding. It is found only at very low levels in F9 and PC13 cell lines, both of which are embryonically derived. Interestingly, Antequera et al. [5] found that the level of m^5CpG protection was reduced in PC13 cells, establishing a correlation between levels of MeCP and degree of m^5CpG protection. These studies also established that binding of MeCP to methylated DNA in vitro reduces the accessibility of methylated *Msp*I sites.

There is also good evidence that MeCP is involved in methylation-mediated transcriptional inhibition. Boyes and Bird [23] have provided evidence that MeCP can inhibit transcription from methylated promotors both in vitro and in vivo. The fact that transcription of methylated templates could be restored by addition of only methylated competitor DNAs which are bound

strongly by MeCP, constituted strong evidence that this protein was inhibiting transcription. Further to this, transcriptional inhibition from methylated templates was not observed in an MeCP-deficient cell line. In conclusion, the effects of DNA methylation on chromatin and transcription can be mimicked by the protein MeCP. An intriguing possibility is that all of the primary biological consequences of CpG-methylation are mediated by the interaction of MeCP or related proteins with DNA. Further work will be required to test these points critically.

Acknowledgements: We are grateful for research support from the Research Institute for Molecular Pathology, Vienna, and more recently to The Imperial Cancer Research Fund and The Wellcome Trust. We also thank Richard Meehan for commenting on the manuscript.

REFERENCES

[1] Razin, A. and Cedar, H. (1977) Proc. Natl. Acad. Sci. USA 74, 2725-2728.
[2] Solage, A. and Cedar, H. (1978) Biochemistry 17, 2934-2938.
[3] Keshet, I., Lieman-Hurwitz, J. and Cedar, H. (1986) Cell 44, 535-545.
[4] Ball, D.J., Gross, D.S. and Garrard, W.T. (1983) Proc. Natl. Acad. Sci. USA 80, 5490-5494.
[5] Antequera, F., Macleod, D. and Bird, A.P. (1989) Cell 58, 509-517.
[6] Bird, A.P. (1986) Nature 321, 209-213.
[7] Kerem, B.-S., Goiten, R., Richler, C., Marcus, M. and Cedar, H. (1983) 304, 88-90.
[8] Tazi, J. and Bird, A.P. (1990) Cell 60, 909-920.
[9] Hebbes, T.R., Thorne, A.W. and Crane-Robinson, C. (1988) EMBO J. 7, 1395-1402.
[10] Grant, S.G. and Chapman, V.M. (1988) Annu. Rev. Genet. 22, 199-233.
[11] Jablonka, E., Goitein, R., Marcus, M. and Cedar, H. (1985) Chromosoma 93, 152-156.
[12] Miller, O.J., Schendl, W., Allen, J. and Erlanger, B.F. (1974) Nature 251, 636-637.
[13] Selig, S., Ariel, M., Goitein, R., Marcus, M. and Cedar, H. (1988) EMBO J. 7, 419-426.
[14] Antequera, F., Boyes, J. and Bird, A.P. (1990) Cell 62, 503-514.
[15] Wolf, S.F. and Migeon, B.R. (1985) Nature 314, 467-469.
[16] Kaslow, D.C. and Migeon, B.R. (1987) Proc. Natl. Acad. Sci. USA 84, 6210-6214.
[17] Venolia, L., Gartler, S.M., Wassman, E.R., Yen, P., Mohandas, T. and Shapiro, L.J. (1982) Proc. Natl. Acad. Sci. USA 79, 2352-2354.
[18] Busslinger, M., Hurst, J. and Flavell, R.A. (1983) Cell 34, 197-206.
[19] Murray, E.J. and Grosveld, F. (1987) EMBO J. 6, 2329-2335.
[20] Buschhausen, G., Wittig, B., Graessman, M. and Graessman, A. (1987) Proc. Natl. Acad. Sci. USA 84, 1177-1181.
[21] Khan, R., Zhang, X.-Y., Supakar, P.C., Ehrlich, K.C. and Ehrlich, M. (1988) J. Biol. Chem. 263, 14374-14383.
[22] Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E. and Bird, A.P. (1989) Cell 58, 499-507.
[23] Boyes, J. and Bird, A.P. (1991) Cell 64, 1123-1134.