CMLS, Cell. Mol. Life Sci. 60 (2003) 1647–1658
1420-682X/03/081646-12
DOI 10.1007/s00018-003-3088-6
© Birkhäuser Verlag, Basel, 2003

CMLS Cellular and Molecular Life Sciences

Review

Structure, function and evolution of CpG island promoters

F. Antequera

Instituto de Microbiología Bioquímica, CSIC/Universidad de Salamanca, Edificio Departamental, Campus Miguel de Unamuno, 37007 Salamanca (Spain), Fax: +34 923 224876, e-mail: CpG@usal.es

Received 27 February 2003; accepted 24 March 2003

Abstract. Mammalian promoters belong to two different categories in terms of base composition and DNA methylation. In humans and mice, approximately 60% of all promoters colocalize with CpG islands, which are regions devoid of methylation that have a higher G+C content than the genome average, while the rest have a methylation pattern and base composition indistinguishable from bulk DNA. Recent comparative studies between both organisms have refined our understanding of how CpG island promoters are organized in terms of protein-DNA interactions and patterns of expression. In addition, the finding that DNA replication initiates at CpG islands in vivo suggests that their distinctive properties could be a consequence of such activity and opens the possibility of a coordinated regulation of transcription and replication. These new data shed light on the origin and evolution of the CpG islands and should contribute to improving methods for promoter prediction in the human and mouse genomes.

Key words. CpG islands; promoter; transcription; replication origins; DNA methylation; evolution.

Introduction

It is difficult to imagine a better contribution to the wealth of knowledge about the biology of mice and humans already available than reading the genetic instructions upon which they are built. The recent sequencing of the human and mouse genomes [1–3] and the possibility of comparative analyses between their genomes and those of other organisms opens unprecedented and fascinating possibilities in all fields of biology. One of the immediate benefits of comparing the mouse and human genomes – given the conservation between their coding sequences – has been the identification or confirmation of many predicted exons and genes [4]. This high degree of homology implies that, perhaps, even more interesting than the identification of their genes will be the analysis of the regulatory networks that specify when and where they are expressed during development. Regulatory regions are known to be more tolerant to changes than coding regions since, in principle, there should be more ways to achieve a particular pattern of expression of a gene than to encode a specific polypeptide. Therefore, such regions have been long thought to be a major source of variability upon which natural selection can act [5–8].

Regulatory regions are made up of a variable number of short modules to which activators and repressors bind in such a way that their integrated contributions result in the correct expression of the gene [9]. Despite the conservation of some of these sequence elements, detailed analysis of the promoter of many genes has revealed a bewildering level of complexity in protein-protein and protein-DNA interactions, suggesting that there might be as many ways to activate transcription as there are genes (see for example, [10–12]). The implication is that each promoter is probably unique and must be understood on its own. This significantly limits bioinformatics attempts to provide general rules for their identification and accounts for the rather limited success of promoter prediction on the basis of sequence analysis [13].

A related approach to the localization of regulatory elements is 'phylogenetic footprinting', which exploits the fact that regulatory modules are sometimes conserved

across species. Two examples of the efficiency of this method are the identification of conserved sequence blocks between mouse and the teleost *Fugu rubripes Hoxb-4* genes [14] and between the mouse *pax6* and *Drosophila eyeless* genes [15]. The main limitation of 'phylogenetic footprinting', however, is that the functional conservation of regulatory pathways does not guarantee conservation of the relevant regulatory sequences involved. This is well illustrated in the case of the enhancer that determines the expression of the *even skipped* gene in stripe number 2 of *Drosophila* embryos. Despite considerable divergence of the enhancer sequence between *D. melanogaster* and *Drosophila pseudoobscura*, the latter can drive the correct expression of *even skipped* in the second stripe in *D. melanogaster* [16]. The key finding of that study was that chimeric enhancers made up of combinations of elements from both species failed to generate the correct spatial pattern of expression. This implied the existence of compensatory mutations capable of maintaining the function of the enhancer in each species. We shall later suggest that a more extreme example of this situation could apply to CpG island promoters in mammals.

The field of transcriptional regulation is continuously being reviewed in the literature and some excellent recent examples include [17–19]. The present review will focus on the organization and evolution of CpG island promoters and how they manage to operate in the repressive context of the highly methylated mammalian genome.

CpG density defines two classes of RNA polymerase II promoters

Despite the sequence diversity among promoters, genes transcribed by RNA polymerase II can be classified in two different and mutually exclusive groups according to the distribution of CpG dinucleotides across their 5′ ends. In one class, the frequency of CpGs is the same as the genome average, which is roughly one every 100 nucleotides. This class invariably includes genes whose expression is restricted to a limited number of cell types (fig. 1 A). In contrast, the 5′ end of the genes belonging to the other group is surrounded by a region ~1 kb long where the frequency of CpGs is approximately 10 times higher than the genome average. These regions were very appropriately called CpG islands [20] and show such a conspicuous clustering of CpG dinucleotides that it can be readily detected by visual inspection of the CpG plot (fig. 1 B). The consistent association of CpG islands with the upstream region of many genes immediately suggested a possible involvement in transcriptional regulation and their potential use as markers to localize genes in genome sequences [21–23]. Even now, 20 years after

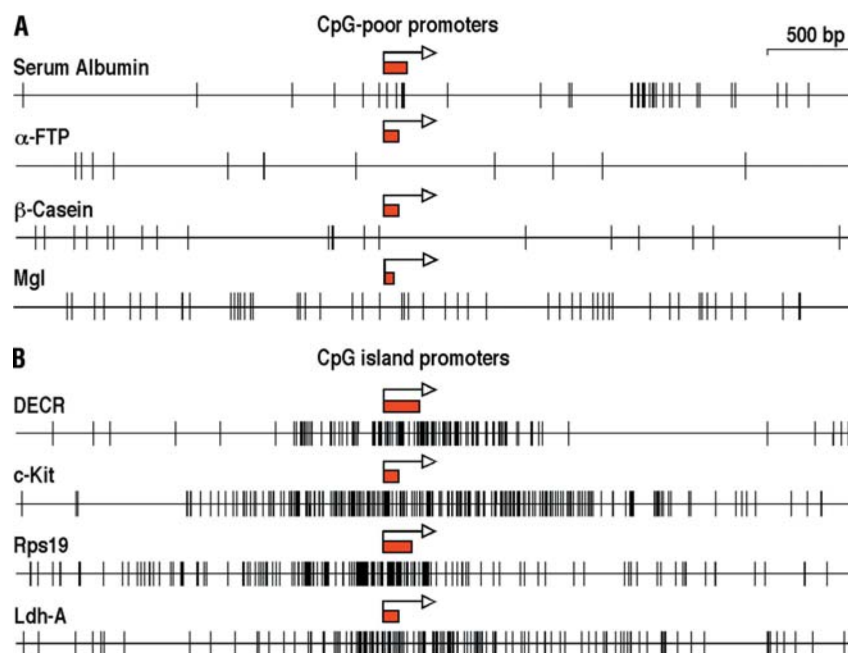

Figure 1. Two classes of promoters in human and mouse. (*A*) Examples of CpG-poor promoters. Red boxes and arrows represent the first exon and the transcription initiation site of the human serum albumin and α-fetoprotein (α-FTP) genes and the mouse β-casein and the macrophage galactose/*N*-acetylgalactosamine-specific C-type lectin (Mgl) genes. (*B*) Examples of CpG island promoters. The first exons of the human 2,4-dienoyl-CoA reductase (DECR) and c-Kit and the mouse ribosomal protein S19 (Rps19) and lactate dehydrogenase A (Ldh-A) genes are shown. Vertical lines indicate the distribution of CpG dinucleotides. Their frequency at CpG-poor promoters is similar to the genome average, while at CpG islands promoters it is approximately 10-fold higher.

their discovery [24, 25], CpG islands still are the most reliable feature for promoter prediction in the mammalian genome [13, 26].

What is so special about CpGs relative to the other 15 possible dinucleotides in DNA? CpGs are the sites where methylation takes place, and ~ 80% of them are methylated at position 5 of the cytosine ring in humans and mice. Somewhat paradoxically, CpGs remain nonmethylated at CpG islands, despite their abundance, whereas the majority of the remaining CpGs scattered across the genome are mostly methylated. In addition to the lack of methylation, human and mouse CpG islands have a G+C content of 67 and 64%, approximately, while the genome averages are 41% and 42%, respectively. The contrast between island and nonisland DNA is so sharp because CpGs occur at the expected frequency at CpG islands on the basis of their G+C content, whereas CpGs in bulk DNA are underrepresented at 20% of their expected frequency. This is due to the spontaneous deamination of methylated cytosines to yield thymine and generate a T:G mismatch that will be fixed as TpG (or CpA in the complementary strand) if not replaced by cytosine before the following round of DNA replication. Deamination of cytosine produces uracyl and generates a U:G mismatch that is repaired far more efficiently than T:G mismatches to restore the original CpG dinucleotide [27]. If methylated CpGs have not been mutated out of existence in the genome, it is because the generation of new CpG sites by point mutation counterbalances their decay in a dynamic equilibrium that maintains a constant level of CpGs at about 20% of the expected frequency [28].

These distinctive features of the CpG islands in terms of a lack of methylation and an elevated G+C content are accompanied by an equally distinctive chromatin organization. Chromatin analysis at global genomic level has revealed that CpG islands show the properties usually ascribed to 'open' or 'active' chromatin. This includes hyperacetylation of histones H3 and H4, a deficiency in histone H1, positioned nucleosomes and nucleosome-free regions that coincide with enhanced sensitivity to nucleases relative to bulk DNA [29–32]. These properties highlight CpG islands as regions that are particularly well suited for direct access to DNA, which is consistent with their colocalization with the promoters of many genes.

Previous studies based on the biochemical isolation of the CpG island fraction estimated approximately 45,000 and 37,000 in the human and mouse genomes, respectively [33], although recent computational predictions have lowered these figures to about 27,000 and 15,500 [1–3]. While the first estimate was limited by the biochemical nature of the assay, the second – although in principle more accurate – is also subject to some uncertainty given that slight variations in any of the parameters used to define the CpG islands mathematically yield a very different final figure [34]. Regardless of the absolute number of CpG islands in the genome, a more relevant issue is what kind of genes are associated with them, since only ~ 60% of all human genes are associated with CpG islands. This includes all the housekeeping genes – those expressed in all cell types – and about half of the tissue-specific genes [23, 33]. Two intriguing features relative to their distribution are that despite their restricted patterns of expression, a significant proportion of the brain or neurally expressed genes is associated with them [35]. The other is that ~ 20% of human promoters associated with CpG islands are CpG deficient in the corresponding mouse orthologues [3, 33]. The presence or absence of CpG islands at orthologous genes of both organisms implies either that some human genes have 'acquired' a CpG island or that the corresponding mouse genes have 'lost' it since both species diverged from a common ancestor about 65 million years ago. The possible mechanisms responsible for the origin and evolution of the CpG islands will be discussed later.

Transcription from CpG-rich and CpG-poor promoters

How does methylation affect transcription from CpG island and non-island promoters? Since CpG islands are nonmethylated in sperm and remain consistently devoid of methylation in somatic tissues, regardless of the expression of the genes associated with them, it is unlikely that DNA methylation would play any role in their regulation. Exceptions to this rule are the CpG islands of imprinted genes, those in the mammalian X inactive chromosome and those associated with the MAGE genes that become methylated during normal mammalian development [36, 37]. Even in this case, methylation is not the primary inactivating signal but takes place at a stage when transcription has been switched off by other means [38, 39]. The MAGE genes were found as antigens in a wide variety of tumours and represent an interesting case. They are specifically expressed in the male germ line, but their function is unknown. They have a CpG-rich region 300–650 bp long at their 5′end that, although shorter than average CpG islands, remains nonmethylated in sperm but is methylated in somatic tissues, where the genes are not expressed. 5-Azacytidine-induced demethylation of these genes causes their expression in various somatic cell types, which is probably driven by ubiquitous transcription factors that are now capable of binding to their CpG-rich promoters [37].

Transcription is strongly repressed upon unscheduled de novo methylation of CpG islands in cell lines and tumour cells, a phenomenon that occurs at high frequency in these situations but never in the organism under normal physiological conditions with the exceptions mentioned above [40–42]. Despite the continuously growing list of examples in many tumours, the mechanisms leading to

aberrant CpG island methylation remain unknown. Transcriptional repression of methylated DNA is mediated by the MeCP and MBD family of proteins that bind specifically to methylated CpGs and are capable of recruiting histone deacetylases and transcriptional corepressors [43–45]. For example, detailed analysis of the binding of the E2F transcriptional activator to a TTTCCCGCG site in the CpG island promoter of the human retinoblastoma *Rb-1* gene has shown that methylation of the CpG dinucleotide abolishes E2F binding and transforms the site into a target for MeCP2 [46]. Additional examples of how the methylation of CpG islands prevents the binding of transcription factors to DNA include in vivo footprinting analysis of the human *PGK-1* and *HPRT* gene promoters in the inactive X chromosome [47, 48] and of the human α-globin CpG island in cells lines in which it has become methylated [49]. In all these cases, the high density of methylated CpGs is likely to cause strong binding of MeCP and MBD proteins and elicits a very effective and stable transcriptional repression [50, 51]. Once methylated, CpG islands never become demethylated in somatic cells unless cells are treated with demethylating agents. Reactivation of aberrantly methylated CpG islands by 5-azacytidine has sometimes been interpreted as evidence that methylation of CpG islands could act as a transcriptional regulator.

In contrast with CpG islands, CpG-poor promoters are methylated in sperm and are always associated with tissue-specific genes. A direct role of DNA methylation in the regulation of this class of promoters predicts a correlation between their methylation profile and their level of expression. Many examples and also several exceptions to this correlation have been described, suggesting that although DNA methylation affects gene expression, it is unlikely to play a general role as a transcriptional regulator. For example, Hpa II sites at the promoter regions of the chicken δ-crystallin genes and at the mouse tissue-specific genes *Acta 1, Mylc* and *Prf1* are nonmethylated in all tissues tested, regardless of the expression of the genes [52, 53]. In the same line, demethylation of the rat tyrosine aminotransferase gene promoter by 5-azacytidine does not lead to its activation in cells where it was previously methylated and inactive, even though proteins capable of binding to the promoter in vitro are present in the nonexpressing cells [54]. Furthermore, no widespread activation of tissue-specific genes has been observed in cancer cells with significantly reduced levels of genomic methylation [55] or in mouse embryonic stem (ES) cells where genomic methylation is reduced to a third of the wild-type level upon disruption of the *Dnmt1* DNA-methyltransferase gene [56].

Detailed analysis of the kinetics of demethylation and gene expression has shown that in some cases where demethylation correlates with expression, the former follows the binding of transcription factors rather than being a prerequisite for it. For example, binding of nuclear factor kappa B (NF-κB) transcription factor to an intronic enhancer of a k-chain gene is required for demethylation in B cells [57]. Also, the binding of Sp1 and of several hybrid activators harbouring various transactivation domains causes promoter demethylation of plasmid constructs in *Xenopus* eggs, even under conditions where ongoing transcription has been blocked by α-amanitin [58]. In this case, replication is required for demethylation, suggesting that it could be achieved passively by preventing the access of the methyltransferase to DNA. Another example of a correlation between demethylation and expression has been found at the human tissue-specific *SERPINB5* gene. The promoter of this gene is associated with a GC-rich region fulfilling the defining criteria for CpG islands but significantly shorter than the average [59]. This region is fully methylated in nonexpressing cell types, and it is reminiscent of the situation of the MAGE genes. Unlike them, however, it is unknown whether this putative CpG island remains nonmethylated in sperm DNA. In addition to passive loss of methylated CpGs, it has been proposed that demethylation might be mediated by an active demethylase activity that could be recruited to some promoters directly by transcription factors or by the associated chromatin remodelling complexes. For a discussion of the current evidence for DNA-demethylase activities, see [60].

That DNA methylation, despite its prominent presence in the mammalian genome, is not a general regulator of gene expression is not surprising since proper gene regulation takes place in invertebrates, many of whose genomes have a low or undetectable level of methylation. Given the conservation of many developmental pathways across the animal kingdom, it would be unlikely that vertebrates would depend on an entirely new regulatory logic. DNA methylation, however, is essential for mammalian development, as shown by the embryonic or perinatal lethal phenotypes caused by disruption of any of the mouse *Dnmt1, Dnmt3a* or *Dnmt3b* DNA-methyltransferase genes [56, 61]. What, then, could be its role? Undesirable consequences of having a large genome, most of which is devoid of genes, is the possible titration of regulators with weak affinity for sequences related to their 6–8 bp cognate binding sites and the risk of abnormal transcription from weak cryptic promoters, whose probability of occurrence increases with genome size [62]. This possibility is supported by recent large-scale transcriptional analyses in mouse embryonic fibroblasts where the *Dnmt1* methyltransferase gene has been deleted. DNA microarray analysis reveals that extensive genomic demethylation in these cells correlates with an at least twofold overexpression of about 10% of the 6000 genes tested [63]. This suggests that methylation could be required for silencing certain tissue-specific genes, although a limitation of genome-wide analyses is that it is

not possible to distinguish between the direct and indirect consequences of demethylation in the activation of specific genes. DNA methylation can also prevent transcription from invading molecular parasites, many of whose promoters mimic CpG islands and can therefore be driven by ubiquitous transcription factors [64], as shown by the widespread expression of intracisternal A particle retroviruses in mouse embryos deficient in the *Dnmt1* gene [65]. In addition to acting as a global repressor of unscheduled transcription, DNA methylation could also contribute to preventing recombination between the thousands of repetitive DNA elements scattered across the genome and to the maintenance of genome stability. Such a function is suggested by the high rate of deletions and chromosomal abnormalities observed in mouse ES cells with reduced levels of methylation [66], although other studies have shown that genomic hypomethylation is associated with a reduction in intestinal polyp formation [67] and in various pathways leading to gene inactivation [68]. In any case, it is possible that the evolution of large genomes might have been facilitated by DNA methylation. Once established, DNA methylation could have incorporated further refinements to its original general repressive role as an adaptation to the specific biology of different organisms [69, 70]. In fact, even between different groups of vertebrates, methylation seems to play different developmental roles [71].

Organization of CpG island promoters

Given that CpG islands are associated with thousands of genes that are active in all cell types of the organism, they are expected to contain many binding sites for ubiquitous transcription factors. This is indeed the case, as shown by in vivo and in vitro analyses of many CpG island promoters [47, 72–76]. An interesting feature of this kind of promoters is the elevated frequency of bidirectional transcription [77–79]. This could be due to the high density of transcription factors bound to them and to the presence of sequence elements capable of activating bidirectional transcription in vitro [72]. Such an organization would allow the coordinated regulation of the two genes, but it could also represent an opportunistic arrangement between two genes to ensure their transcription, given the aparent lack of functional relationship between some of them [79].

Does every CpG island promoter require a specific array of factors, or is there a certain degree of tolerance, such that different arrangements could result in comparable levels of transcription? This possibility is suggested by the slightly different patterns of protein-DNA interactions in vivo found at the mouse HTF9 CpG island promoter in different cell types of the same animal [80]. This flexibility would represent a robust strategy for housekeeping genes in terms of ensuring their expression, regardless of possible variations in the concentration of activators in different cell types. For example, the level of messenger RNA (mRNA) and protein of the Sp1 transcription factor fluctuates in a 100-fold range in different tissues and developmental stages in the mouse [81]. Sp1 is a ubiquitous transcription factor that binds the CCCGCC sequence, and some variations of it, that is present in many CpG island promoters because of their high G+C content [82]. In fact, this could be the only factor driving the transcription of some genes, such as the mouse *Aprt*, where three Sp1 sites seem to be the only requirement for its transcription in vitro and in vivo [31, 83]. Despite this, and as a further proof of the tolerance of CpG island promoters to fluctuations in transcriptional activators, disruption of the mouse Sp1 gene does not affect transcription from several CpG island promoters previously shown to contain Sp1 binding sites, including *Aprt* [84]. This tolerance would be difficult to explain if each promoter had a strict requirement for a particular array of activators, and suggests some degree of promiscuity at CpG island promoters.

One difficulty with this scenario is that approximately half of all the CpG islands in the human genome are associated with genes that are expressed in some cell types only [33]. How are these promoters organized in expressing and nonexpressing tissues, considering that CpG islands remain nonmethylated in both situations? In vivo footprinting analysis of the CpG islands associated with the human *c-JUN* [85] *PCNA* [76] and α-globin genes [49] reveals that they are constitutively bound by many ubiquitous transcription factors and that only minor changes in the pattern of binding can be detected between cell types whose differences in the level of expression are of orders of magnitude. The constitutive binding of ubiquitous transcription factors is probably the reason why a very low level of transcription of tissue-specific CpG island genes such as the mouse *TrkA* [86] and the human α-globin [49, 87] is present in many cell types. Extreme examples of this situation are the expression of the tissue-specific CpG island genes α-globin, myotonin protein kinase (*MPK*), *SRY* and *ZFY* in human preimplantation embryos from the one-cell stage to the eight-cell stage, while parallel analysis fails to detect transcription of the β-globin gene and other tissue-specific genes not associated with CpG islands [88, 89]. Even in cases like the mouse pro-opiomelanocortin gene and the major histocompatibility complex (MHC) class II I-Aβ gene, where the CpG islands are not localized at the 5′ end of the genes, transcription from them has been detected in early embryonic and germ cells [35, 90]. The case of the mouse *MyoD1* gene is unusual given that CpGs across its first long exon are nonmethylated in all tissues tested [91]. The frequency of CpGs, however, significantly decreases upstream from the transcription initiation site (fig. 3, see

later). Is is possible that this low density of CpGs prevents expression of the *MyoD1* gene in early embryos unlike other CpG island-associated genes [90]. This suggests that, in a sense, all CpG island promoters are housekeeping promoters and what is really tissue specific is the high rate of transcription in the relevant cell types. This differential rate of transcription could be achieved by repressing the expression in most cell types or by the presence of cell-type-specific activators in those where the gene should be actively expressed. Tissue-specific regulators could account for the subtle footprint differences at some CpG island promoters between cells with high and low levels of expression [85, 92].

The close association between CpG islands and promoters does not necessarily imply that they overlap precisely and raises the question how they relate to one another, or – in other words – where promoters are located within the CpG island framework. Mouse and human CpG islands are often differentially positioned relative to the transcription initiation site at orthologous genes that are likely to play similar functions in both organisms, especially in the case of housekeeping genes (fig. 2 A). Recent comparative in vivo footprinting analyses across several human and mouse CpG islands have shown that the pattern of protein-DNA interactions is very different between both organisms. These differences contrast with the high degree of conservation of their coding sequences and support the notion mentioned above that regulatory regions are more tolerant to changes than coding regions, at least at CpG island promoters. Despite these interspecific differences, a common theme is that the promoter region is precisely contained between the 5′ boundary of the CpG island and the transcription initiation site [49]. For example, the CpG island of the mouse *Aprt* gene extends approximately 80–100 bp upstream from the transcription initiation site, and all the sequences involved in the expression of this gene – both in vitro [83] and in vivo [31] – are confined within this short region. In contrast, the 5′ boundary of the human orthologue lies 600 bp upstream from the transcription initiation site and the pattern of protein-DNA interactions in vivo extends throughout this length (fig. 2 A). Functional analysis by transient transfection indicates that the entire region participates in the regulation of the human *APRT* gene. These differences in organization also apply to the CpG island promoters of the human and mouse adenosine deaminase and telomerase RNA genes [49] and are consistent with the findings of previous studies that defined the localization of these promoters by functional analysis [93, 94]. Could this precise circumscription of the cis-regulatory region of genes associated with CpG islands be used as a general rule for the identification of promoters directly in DNA sequences? This possibility can be tested by using the inverse approach: namely, asking whether housekeeping promoters that have previously been mapped by in vivo footprinting such as those of the human *CDC2*,

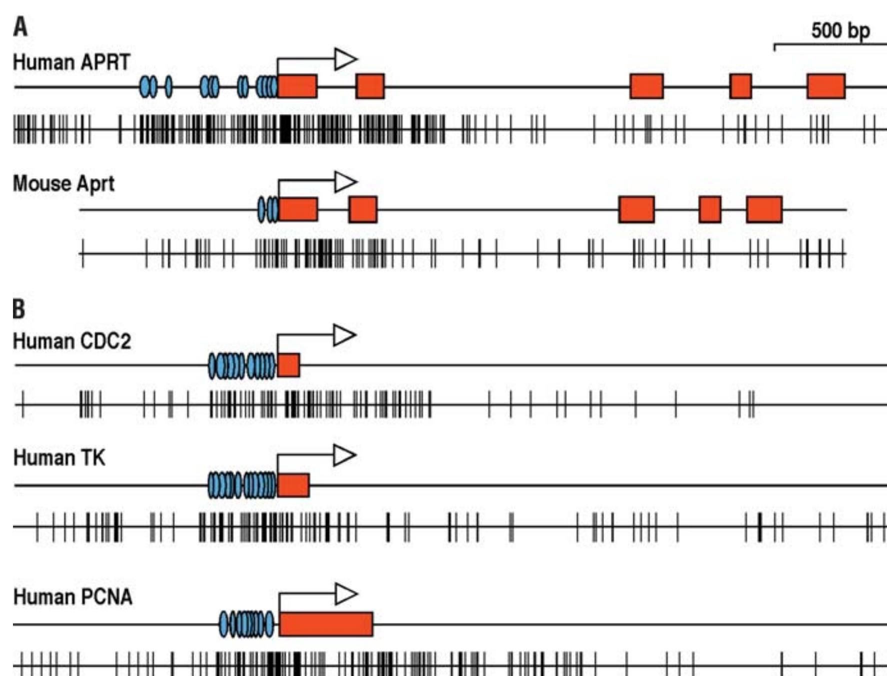

Figure 2. Pattern of protein-DNA interactions at CpG island promoters. (*A*) The five exons of the human and mouse adenine phosphoribosyltransferase (APRT) orthologous genes are shown. Transcription factors are represented by blue ovals upstream from the transcription initiation site. (*B*) The first exons of the human *CDC2*, thymidine kinase (*TK*) and *PCNA* genes are shown. In the five examples, the protein-DNA interactions, as detected by in vivo footprinting, are limited to the region between the 5′ boundary of the CpG-rich region and the transcription initiation site. For further details about the binding sites for the putative transcription factors involved, see [31, 49, 74, 76, 92].

thymidine kinase and *PCNA* genes [74, 76, 92] would also follow this rule. Analysis of the distribution of CpG sites across them shows that in all cases they are contained between the 5′ boundary of their respective CpG island and the transcription initiation site (fig. 2 B). Together, these data suggest that this could be a general rule for direct identification of promoters associated with CpG islands in the mouse and human genomes. The identification of these cis-regulatory regions does not preclude the existence of other regulatory regions in introns or at some distance from the gene.

Initiation of DNA replication at CpG islands

Transcription factors have been shown to stimulate replication in many systems, ranging from viruses [95] to yeast [96]. This stimulation could be accomplished by direct interaction with components of the replication machinery or by facilitating access of the replication complexes to DNA through recruitment of chromatin remodelling complexes. These lines of evidence, along with the fact that some chromosomal replication origins have been mapped close to gene promoters in rat [97] and human cells [98–100], suggested that CpG islands might serve simultaneously as promoters and replication origins. In principle, they should be particularly well suited for replication initiation because of their high density of transcription factors and their open chromatin organization. This possibility was also consistent with the early replication time of housekeeping genes during S phase in synchronous cell cultures [101]. As predicted by this hypothesis, analysis of genomic DNA fractions containing short replication intermediates were found to be enriched in GC-rich sequences showing the high CpG frequency typical of CpG islands [102]. Measurement of the relative abundance of replication intermediates by competitive polymerase chain reaction (PCR) has identified DNA replication origins at the CpG islands associated with the human *TRKA* gene and three hamster genes [102], with several genes in the human X chromosome [103], between the chicken *cLys* and *cGas41* genes [104, 105], and with the human *HPRT* gene [106].
Recently, the role of CpG islands as DNA replication origins has been further supported by chromatin immunoprecipitation analyses with antibodies against the human Orc1 and Orc2 proteins, which are components of the origin recognition complex (ORC). The immunoprecipitated DNA fraction showed the same CpG island properties as the fraction derived from short replication intermediates, strongly suggesting that the ORC complex binds to CpG islands [107]. Detailed characterization of two of the immunoprecipitated fragments revealed that they were derived from the CpG island containing the promoter of the *TOP1* gene [108] and from a CpG island between the bidirectionally transcribed *MCM4* and *PRKDC* genes [107]. To date, the best-characterized human DNA replication origin is localized downstream from the Lamin B2 gene and immediately adjacent to the CpG island promoter of the *PPV1* gene [99]. High-resolution analyses have shown that the transcription and replication initiation sites are only 400 bp apart [109], suggesting a possible coordinated regulation between transcription and replication. The possibility of a physical interaction between the transcription factor Sp1 and Orc2 is supported by the presence of both proteins in the same immunoprecipitated DNA fragments [108]. In addition to the examples mentioned above, the colocalization of replication origins and gene promoters has been reported in eukaryotes as phylogenetically diverse as *Drosophila* [110], *Physarum polycephalum* [111] and *Schizosaccharomyces pombe* [112].

Origin and evolution of the CpG islands

CpG islands are not intrinsically refractory to methylation, as shown by those at imprinted or X-inactivated genes and in tumour cells, and this raises the question of how they manage to remain free of methylation in the heavily methylated genomic context.
The obvious possibility that the binding of transcription factors might render them less accessible to DNA-methyltransferases is supported by several lines of evidence. First, mutation of the Sp1 binding sites required for the expression of the mouse *Aprt* gene results in the de novo methylation of its CpG island [31, 113]. Also, the remethylation kinetics of CpG islands associated with the first and second exons of the *p16* gene in human T24 bladder carcinoma cells after demethylation induced by 5-azacytidine reveals that the 3′ CpG island, which presumably does not act as a promoter in those cells, is remethylated at a higher rate than the 5′ CpG island [114]. In addition, the 3′ end of the human c-Ha-*Ras* and mouse *myoD1* CpG islands is more prone to methylation than their 5′ end during normal development and upon oncogenic transformation [115, 116]. A similar situation has been observed at the 3′ end of the human *p16* and α-globin CpG islands in several tumour cell lines [S. Delgado, M. Cuadrado and F. Antequera, unpublished]. Impaired access to the CpG island region bound by transcription factors might also account for the slow rate of repair relative to the remaining part of the island or to the transcribed region of the *JUN* gene in human fibroblasts after ultraviolet (UV) irradiation [117].
Although the binding of factors could contribute to the maintenance and protection (but not invulnerability) of CpG islands against de novo methylation, it is unlikely to account for their origin in the mammalian genome. This leads us to the key question of why CpG islands exist at

all. This is not merely a rhetorical question, because ~ 40% of all human promoters are CpG poor, indicating that gene regulation can operate without the benefit of the CpG islands. A clear example of this situation is the human α-globin gene, which is associated with a CpG island, while the β-globin gene is not, although both genes are expressed in erythroid cells to produce the hemoglobin tetramer. In contrast, neither the α- nor β-globin mouse orthologous genes is associated with a CpG island (fig. 3). This difference in the presence or absence of CpG islands between humans and mice is not unusual, since ~ 20% of human CpG island promoters are CpG poor in the mouse orthologues [3, 33, 58]. This situation is by far more common than the opposite case, where a human CpG island would be absent or smaller relative to the mouse orthologue. Have some human promoters 'acquired' a CpG island or have some mouse promoters 'lost' it since the time they diverged in evolution? CpG islands can be lost by de novo methylation in the germ line and replacement of CpG by TpGs through deamination of methylated CpGs, as commented above. This could have happened, for example, to the human α-globin pseudogene [118] or to the CpG island at the 3′ end of the mouse ζ-globin gene [33]. While this process could lead to the disappearance or 'erosion' of CpG islands [119], it cannot explain how they appeared in the first place. Given their high G+C content relative to the genome average, it is reasonable to assume that DNA polymerases could have been involved. The finding that CpG islands colocalize not only with promoters but also with DNA replication origins opens the possibility that their two most distinctive properties – the lack of methylation and high G+C content – might be a consequence of such activity. We have previously suggested a model that proposes the existence of a replication initiation stage with different properties from the subsequent elongation phase [120]. The assembly of the replication machinery at this initial stage could involve some specific components other than those at the mature replication forks, and it could be incompatible with, or hinder, the access of DNA-methyltransferases. Also, the rate of errors or the efficiency of repair might also be different and could favour a progressive shift in base composition towards a higher G+C content. Precedents of an initiation stage of DNA replication showing specific properties have been found in the mitochondrial [121] and SV40 replication initiation regions [122]. Moreover, a high rate of mitotic recombination has recently been found at DNA replication origins in the yeast *S. pombe* [123], although it is unknown whether this feature also applies to CpG island replication origins.

One implication of this speculative model is that CpG islands might have emerged without a positive value to be selected for, but as a genomic footprint left in the chromosomes by the replication initiation event. It would be expected that depending on the local conditions that generated the footprint, CpG islands would show differences in their average G+C contents and in their association with genes among the different classes of vertebrates. This is in fact the case. For example, fish have unmethylated regions and reduced suppression of CpGs at some promoters, but their base composition is not significantly higher than their surrounding sequences [124, 125], while CpG islands in *Xenopus* are shorter and have lower G+C contents than in mammals [71]. By contrast,

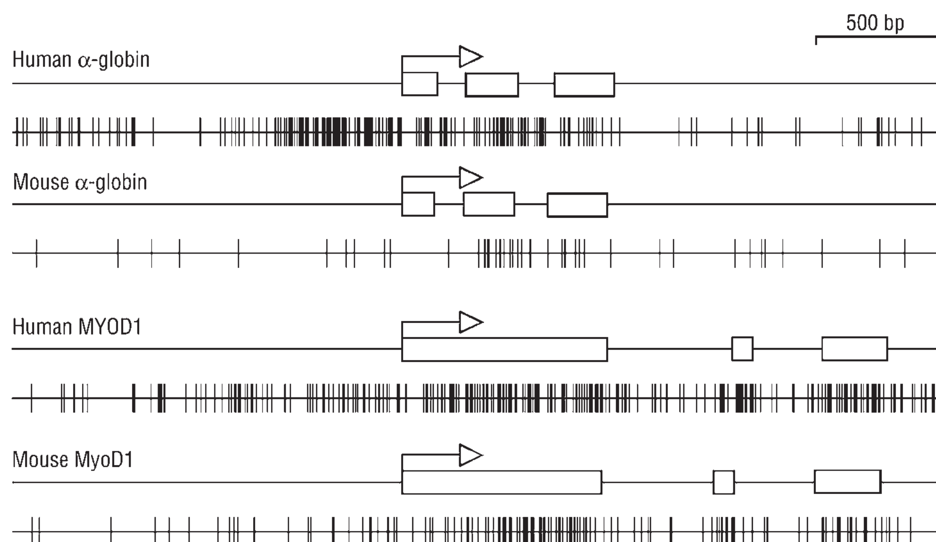

Figure 3. Distribution of CpG sites at the promoters of human and mouse α-globin and MyoD1 orthologous genes. CpG islands can be similar in size, smaller (as in the case of the *APRT* genes shown in fig. 2) or can be absent in the mouse relative to their human orthologues. This is the case of the α-globin and *MyoD1* genes, which are associated with a CpG island in humans but not in the mouse despite similar genomic organization of the genes. The three exons of the four genes are shown. The homology at the nucleotide level between the coding regions of the α-globin and *MyoD1* orthologous genes is 80.9% and 85.4%, respectively.

chicken CpG islands have higher G+C contents than those seen in mammals [126]. Even between mice and humans there are significant differences, as illustrated in figures 2 and 3. According to this scenario, promoters active in the germ line or in early embryonic cells could act as replication origins and as a consequence would have acquired a CpG island. This is consistent with the fact that genes associated with CpG islands – either housekeeping or tissue specific – are expressed in the germ line or very early on in embryonic development [35, 88–90, 127]. Once established, CpG island promoters could be driven by sharing ubiquitous factors that would ensure their constitutive transcription. While this is probably convenient for housekeeping genes, there is no obvious advantage for tissue-specific genes, which would then be at greater risk of being transcribed in the wrong cell types. Whatever the possible benefits granted by CpG islands, they are counterbalanced by the risk of undergoing de novo methylation and irreversible transcriptional shutdown that, depending on the gene affected, may have devastating consequences for the cell or for the organism.

Acknowledgments. I am very grateful to Adrian Bird for insightful comments on the manuscript.

1 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature **409:** 860–921
2 Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G. et al. (2001) The sequence of the human genome. Science **291:** 1304–1351
3 Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. Nature **420:** 520–562
4 Xuan Z., Wang J. and Zhang M. Q. (2002) Computational comparison of two mouse draft genomes and the human golden path. Genome Biol. **4:** R1–R10
5 Tautz D. (2000) Evolution of transcriptional regulation. Curr. Op. Gen. Devel. **10:** 575–579
6 Peterson K. J. and Davidson E. H. (2000) Regulatory evolution and the origin of the bilaterians. PNAS **97:** 4430–4433
7 Rockman M. V. and Wray G. A. (2002) Abundant raw material for cis-regulatory evolution in humans. Mol. Biol. Evol. **19:** 1991–2004
8 Carroll S. B. (2000) Endless forms: the evolution of gene regulation and morphological diversity. Cell **101:** 577–580
9 Arnone M. I. and Davidson E. H. (1997) The hardwiring of development: organization and function of genomic regulatory systems. Development **124:** 1851–1864
10 Yuh C. H., Bolouri H. and Davidson E. H. (1998) Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. Science **279:** 1896–1902
11 Cosma M. P., Tanaka T. and Nasmyth K. (1999) Ordered recruitment of transcription and chromatin remodelling factors to a cell cycle-and developmetally regulated promoter. Cell **97:** 299–311
12 Soutoglou E. and Talianidis I. (2002) Coordination of PIC assembly and chromatin remodeling during differentiation-induced gene activation. Science **295:** 1901–1904
13 Hannenhalli S. and Levy S. (2001) Promoter prediction in the human genome. Bioinformatics **17:** S90–S96
14 Aparicio S., Morrison A., Gould A., Gilthorpe J., Chaudhuri C., Rigby P. et al. (1995) PNAS **92:** 1684–1688
15 Xu P. X., Zhang X. Heaney S., Yoon A., Michelson A. M. and Maas R. L. (1999) Regulation of *Pax6* expression is conserved between mice and flies. Development **126:** 383–395
16 Ludwig M. Z., Bergman C., Patel N. H. and Kreitman M. (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. Nature **403:** 564–567
17 Lemon B. and Tjian R. (2000) Orchestrated response: a symphony of transcription factors for gene control. Genes Dev. **14:** 2551–2569
18 Cosma M. P. (2002) Ordered recruitment: gene-specific mechanism of transcription activation. Mol. Cell **10:** 227–236
19 Narlikar G. J., Fan H. Y. and Kingston R. E. (2002) Cooperation between complexes that regulate chromatin structure and transcription. Cell **108:** 475–487
20 Bird A. (1986) CpG-rich islands and the function of DNA methylation. Nature **321:** 209–213
21 Bird A. (1987) CpG islands as gene markers in the vertebrate nucleus. Trends Genet. **3:** 342–347
22 Gardiner M. and Frommer M. (1987) CpG islands in vertebrate genomes. J. Mol. Evol. **196:** 261–282
23 Larsen F., Gundersen G., Lopez R. and Prydz H. (1992) CpG islands as gene markers in the human genome. Genomics **13:** 1095–1107
24 McClelland M. and Ivarie R. (1982) Asymmetrical distribution of CpG in an 'average' mammalian gene. Nucleic Acids Res. **10:** 7865–7877
25 Cooper D. N., Taggart M. H. and Bird A. (1983) Unmethylated domains in vertebrate DNA. Nucleic Acids Res. **11:** 647–658
26 Ioshikhes I. P. and Zhang M. Q. (2000) Large-scale human promoter mapping using CpG islands. Nature Genet. **26:** 61–63
27 Yang A. S. and Jones P. A. (1996) The mutational burden of 5-methylcytosine. In: Epigenetic Mechanisms of Gene Regulation, pp. 77–94, Russo V. E., Martienssen, R. A. and Riggs A. D. (eds), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y.
28 Sved J. and Bird A. (1990) The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. PNAS **87:** 4692–4696
29 Antequera F., Macleod D. and Bird A. (1989) Specific protection of methylated CpGs in mammalian nuclei. Cell **58:** 509–517
30 Tazi J. and Bird A. (1990) Alternative chromatin structure at CpG islands. Cell **60:** 909–920
31 Macleod D., Charlton J., Mullins J. and Bird A. (1994) Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. **8:** 2282–2292
32 Gilbert S. L. and Sharp P. A. (1999) Promoter-specific hypoacetylation of X-inactivated genes. PNAS **96:** 13825–13830
33 Antequera F. and Bird A. (1993) Number of CpG islands and genes in human and mouse. PNAS **90:** 11995–11999
34 Takai D. and Jones P. A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. PNAS **99:** 3740–3745
35 Gardiner M. and Frommer M. (1994) Transcripts and CpG islands associated with the pro-opiomelanocortin gene and other neurally expressed genes. J. Mol. Endocrinol. **12:** 365–382
36 Reik W. and Walter J. (2001) Genomic imprinting: parental influence on the genome. Nat. Rev. **2:** 21–32
37 De Smet C., Lurquin C., Lethe B., Martelange V. and Boon T. (1999) DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. Mol. Cell. Biol. **19:** 7327–7335

38 Lock L. F., Takagi N. and Martin G. R. (1987) Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. Cell **48:** 39–46
39 Keohane A. M., Lavender J. S., O'Neill L. P. and Turner B. M. (1998) Histone acetylation and X inactivation. Dev. Genet. **22:** 65–73
40 De Bustros A., Nelkin B. D., Silverman A., Ehrlich G., Poiesz B. and Baylin S. B. (1988) The short arm of chromosome 11 is a 'hot spot' for hypermethylation in human neoplasia. PNAS **85:** 5693–5697
41 Antequera F., Boyes J. and Bird, A. (1990) High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell **62:** 503–514
42 Esteller M. (2002) CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene **21:** 5427–5440
43 Nan X., Hg H. H., Johnson C. A., Laherty C. D., Turner B. M., Eisenman R. N. et al. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature **393:** 386–389
44 Hendrich B. and Bird A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol. Cell. Biol. **18:** 6538–6547
45 Wade P. A. (2001) Methyl CpG-binding proteins and transcriptional repression. BioEssays **23:** 1131–1173
46 DiFiore B., Palena A., Felsani A., Palitti F. Caruso M. and Lavia P. (1999) Cytosine methylation transforms an E2F site in the retinoblastoma gene promoter into a into binding site for the general repressor methylcytosine-binding protein 2 (MeCP2). Nucleic Acids Res. **27:** 2852–2859
47 Pfeifer G. P., Tanguay R. L., Steigerwald S. D. and Riggs A. D. (1990) In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. Genes Dev. **4:** 1277–1287
48 Hornstra I. K. and Yang T. P. (1992) Multiple in vivo footprints are specific to the active allele of the X- linked human hypoxanthine phosphoribosyltransferase gene 5′ region: Implications for X chromosome inactivation. Mol. Cell. Biol. **12:** 5345–5354
49 Cuadrado M., Sacristán M. and Antequera F. (2001) Species-specific organization of CpG island promoters at mammalian homologous genes. EMBO Rep. **2:** 586–592
50 Boyes J. and Bird A. (1992) Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. EMBO J. **11:** 327–333
51 Hsieh C. L. (1994) Dependence of transcriptional repression on CpG methylation density. Mol. Cell. Biol. **14:** 5487–5494
52 Sullivan C. H., Norman J. T., Borras T. and Grainger R. M. (1989) Developmental regulation of hypomethylation of δ-crystallin genes in chiken embryo lens cells. Mol. Cell. Biol. **9:** 3132–3135
53 Walsh C. P. and Bestor T. H. (1999) Cytosine methylation and mammalian development. Genes Dev. **13:** 26–34
54 Weih F., Nitsch D., Reik A., Schütz G. and Becker P. B. (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. EMBO J. **10:** 2559–2567
55 Baylin S. B., Herman J. G., Graff J. R., Vertino P. M. and Issa J. P. (1998) Alterations in DNA methylation: a fundamental aspect of neoplasia. Adv. Cancer Res. **72:** 141–196
56 Li E., Bestor T. H. and Jaenisch R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69:** 915–926
57 Kirillov A., Kistler B., Mostoslavsky R., Cedar H., Wirth T. and Bergman Y. (1996) A role for nuclear NF-κB in B-cell-specific demethylation of the Igκ locus. Nat. Genet. **13:** 435–441
58 Matsuo K., Silke J., Georgiev O., Marti P., Giovannini N. and Rungger D. (1998) An embryonic demethylation mechanism involving binding of transcription factors to replicating DNA. EMBO J. **17:** 1446–1453
59 Futscher B. W., Oshiro M. M., Wozniak R. J., Holtan N., Hanigan C. L,. Duan H. et al. (2002) Role for DNA methylation in the control of cell type-specific maspin expression. Nat. Genet. **31:** 175–179
60 Bird A. (2002) DNA methylation and epigenetic memory. Genes Dev. **16:** 6–21
61 Okano M., Bell D. W., Haber D. A. and Li E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell **99:** 247–257
62 Bird A. (1995) Gene number, noise reduction and biological complexity. Trends Genet. **11:** 94–100
63 Jackson-Grusby L., Beard C., Possemato R., Tudor M., Fambrough D., Csankovszki G. et al. (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. **27:** 31–39
64 Yoder J. A., Walsh C. P. and Bestor T. H. (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. **13:** 335–340
65 Walsh C. P., Chailler J. R. and Bestor T. H. (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. **20:** 116–117
66 Chen R. Z., Pettersson U., Beard C., Jackson-Grusby L. and Jaenisch R. (1998) DNA hypomethylation leads to elevated mutation rates. Nature **395:** 89–93
67 Laird P. W., Jackson-Grusby L., Fazeli A., Dickinson S. L., Jung W. E., Li E. et al. (1995) Suppression of intestinal neoplasia by DNA hypomethylation. Cell **81:** 197–205
68 Chan M., van Amerongen R., Nijjar T., Cuppen E., Jones P. A. and Laird P. W. (2001) Reduced rates of gene loss, gene silencing and gene mutation in *Dnmt1*-deficient embryonic stem cells. Mol. Cell. Biol. **21:** 7587–7600
69 Regev A., Lamb M. J. and Jablonka E. (1998) The role of DNA methylation in invertebrates: developmental regulation or genome defense? Mol. Biol. Evol. **15:** 880–891
70 Colot V. and Rossignol J. L. (1999) Eukaryotic DNA methylation as an evolutionary device. Bioessays **21:** 402– 411
71 Stancheva I., El-Maari O., Walter J., Niveleau A. and Meehan R. R. (2002) DNA methylation at promoter regions regulates the timing of gene activation in *Xenopus laevis* embryos. Dev. Biol. **243:** 155–165
72 Somma M.P., Pisano C. and Lavia P. (1991) The housekeping promoter from the mouse CpG island HTF9 contains multiple protein-binding elements that are functionally redundant. Nucleic Acids Res. **19:** 2817–2824
73 Rozek D. and Pfeifer G. P. (1995) In vivo protein-DNA interactions at the c-Jun promoter in quiescent and serum-stimulated fibroblasts. J. Cell. Biochem. **57:** 479–487
74 Tommasi S. and Pfeifer G. P. (1995) In vivo structure of the human *cdc2* promoter: release of a p130-E2F-4 complex from sequences immediately upstream of the transcription initiation site coincides with induction of cdc2 expression. Mol. Cell. Biol. **15:** 6901–6913
75 Litt M. D., Hornstra I. K. and Yang T. P. (1996) In vivo footprinting and high-resolution methylation analysis of the mouse hypoxanthine phosphoribosyltransferase gene 5′ region on the active and inactive X chromosomes. Mol. Cell. Biol. **16:** 6190–6199
76 Tommasi S. and Pfeifer G. P. (1999) In vivo structure of two divergent promoters at the human PCNA locus. J. Biol. Chem. **274:** 27829–27838
77 Lavia P., Macleod D. and Bird A. (1987) Coincident start sites for divergent transcripts at a randomly selected CpG-rich island of mouse. EMBO J. **6:** 2773–2779

78 Williams T., Yon J., Huxley C. and Fried M. (1988) The mouse surfeit locus contains a very tight cluster of four 'housekeeping' genes that is conserved trough evolution. PNAS **85:** 3527–3530
79 Adachi, N. and Lieber, M. R. (2002) Bidirectional gene organization: a common architectural feature of the human genome. Cell **109:** 807–809
80 Stapleton G., Somma M. P. and Lavia P. (1993) Cell type-specific interactions of transcription factors with a housekeeping promoter in vivo. Nucleic Acids Res. **21:** 2465–2471
81 Saffer J. D., Jackson S. P. and Annarella M. B. (1991) Developmental expression of Sp1 in the mouse. Mol. Cell. Biol. **11:** 2189–2199
82 Suske G. (1999) The Sp-family of transcription factors. Gene **238:** 291–300
83 Dush M. K., Briggs M. R., Royce M. E., Schaff D. A., Khan S. A., Tischfield J. A. et al. (1988) Nucleic Acids Res. **16:** 8509–8524
84 Marín M., Karis A., Visser P., Grosveld F. and Philipsen S. (1997) Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. Cell **89:** 619–628
85 Rozek D. and Pfeifer G. P. (1993) In vivo protein-DNA interactions at the c-jun promoter: preformed complexes mediate the UV response. Mol. Cell. Biol. **13:** 5490–5499
86 Lomen-Hoerth C. and Shooter E. M. (1995) Widespread neurotrophin receptor expression in the immune system and other nonneuronal rat tissues. J. Neurochem. **64:** 1780–1789
87 Whitelaw E., Hogben P., Hanscombe O. and Proudfoot N.J. (1989) Transcriptional promiscuity of the human α-globin gene. Mol. Cell. Biol. **9:** 241–251
88 Daniels R. and Monk M. (1997) Gene expression in human preimplantation embryos. In: Molecular Genetics of Early Human Development, pp. 155–170, BIOS Scientific Publishers, Oxford
89 Daniels R., Lowell S., Bolton V. and Monk M. (1997) Transcription of tissues-specific genes in human preimplantation embryos. Hum. Reprod. **12:** 2251–2256
90 Macleod D., Ali R. R. and Bird A. (1998) An alternative promoter in the mouse MHC class II I-Ab gene: implications for the origin of CpG islands. Mol. Cell. Biol. **18:** 4433–4443
91 Jones P. A., Wolkowicz M.J., Rideout W. M., Gonzales F. A., Marziasz C. M., Coetzee G. A. et al. (1990) De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines. PNAS **87:** 6117–6121
92 Tommasi S. and Pfeifer G. P. (1997) Constitutive protection of E2F recognition sequences in the human thymidine kinase promoter during cell cycle progression. J. Biol. Chem. **272:** 30483–30490
93 Ingolia D. E., Al-Ubaidi M. R., Yeung C. Y., Bigo H. A., Wright D. A. and Kellems R. E. (1986) Molecular cloning of the murine adenosine deaminase gene from a genetically enriched source: identification and characterization of the promoter region. Mol. Cell. Biol. **6:** 4458–4466
94 Zhao J. G., Hoare S. F., McFarlane R., Muir S., Parkinson E. K., Black D. M. et al. (1998) Cloning and characterization of human and mouse telomerase RNA gene promoter sequences. Oncogene **16:** 1345–1350
95 DePamphilis M. L. (1993) Eukaryotic DNA replication: anatomy of an origin. Ann. Rev. Biochem. **62:** 29–63
96 Li R., Yu D. S., Tanaka M., Zheng L., Berger S. L. and Stillman B. (1998) Activation of chromosomal DNA replication in *Saccharomyces cerevisiae* by acidic transcriptional activation domains. Mol. Cell. Biol. **18:** 1296–1302
97 Zhao Y., Tsutsumi R., Yamaki M., Nagatsuda Y., Ejiri S. and Tsutsumi K. (1994) Initiation zone of DNA replication at the aldolase B locus encompasses transcription promoter region. Nucleic Acids Res. **22:** 5385–5390
98 Kitsberg D., Selig S., Keshet I. and Cedar H. (1993) Replication structure of the human β-globin gene domain. Nature **366:** 588–590
99 Giacca M., Zentilin L., Norio P., Diviacco S., Dimitrova D., Contreas G. et al. (1994) Fine mapping of a replication origin of human DNA. PNAS **91:** 7119–7123
100 Taira T., Iguchi-Ariga S. M. and Ariga H. (1994) A novel DNA replication origin identified in the human heat shock protein 70 gene promoter. Mol. Cell. Biol. **14:** 6386–6397
101 Holmquist G. P. (1989) Evolution of chromosome bands: Molecular ecology of noncoding DNA. J. Mol. Evol. **28:** 469–486
102 Delgado S., Gómez M., Bird A. and Antequera F. (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J. **17:** 2426–2435
103 Rivella S., Palermo B., Pelizon C., Sala C., Arrigo G. and Toniolo D. (1999) Selection and mapping of replication origins from a 500 kb region of the human X chromosome and their relationship to gene expression. Genomics **62:** 11–20
104 Phi-van L. and Strätling W. H. (1999) An origin of bidirectional DNA replication is located within a CpG island at the 3′ end of the chicken lysozyme gene. Nucleic Acids Res. **27:** 3007–3017
105 Chong S., Riggs A. D. and Bonifer C. (2002) The chicken lysozyme chromatin domain contains a second, widely expressed gene. Nucleic Acids Res. **30:** 463–467
106 Cohen S. M., Brylawski B. P., Cordeiro-Stone M. and Kaufman D. G. (2002) Mapping of an origin of DNA replication near the transcriptional promoter of the human HPRT gene. J. Cell. Biochem. **85:** 346–356
107 Ladenburger E. M., Keller C. and Knippers R. (2002) Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication. Mol. Cell. Biol. **22:** 1036–1048
108 Keller C., Ladenburger E. M., Kremer M. and Knippers R. (2002) The origin recognition complex marks a replication origin in the human TOP1 gene promoter. J. Biol. Chem. **277:** 31430–31440
109 Abdurashidova G., Deganuto M., Klima R., Riva S., Biamonti G., Giacca M. et al. (2000) Start sites of bidirectional DNA synthesis at the human lamin B2 origin. Science **287:** 2023–2026
110 Sasaki T., Sawado T., Yamaguchi M. and Shinomiya T. (1999) Specification of regions of DNA replication initiation during embryogenesis in the 65-kilobase DNA*polα-dE2F* locus of *Drosophila melanogaster*. Mol. Cell. Biol. **19:** 547–555
111 Pierron G., Pallota D. and Benard M. (1999) The one-kilobase DNA fragment upstream of the ardC actin gene of *Physarum polycephalum* is both a replicator and a promoter. Mol. Cell. Biol. **19:** 3506–3514
112 Gómez M. and Antequera F. (1999) Organization of DNA replication origins in the fission yeast genome. EMBO J. **18:** 5683–5690
113 Brandeis M., Frank D., Keshet I., Siegfried Z., Mendelsohn M., Nemes A. et al. (1994) Sp1 elements protect a CpG islands from de novo methylation. Nature **371:** 435–438
114 Bender C. M., Gonzalgo M. L., Gonzales F. A., Nguyen C. T., Robertson K. D. and Jones P. A. (1999) Roles of cell division and gene transcription in the methylation of CpG islands. Mol. Cell. Biol. **19:** 6690–6698
115 Ghazi H., Gonzales F. A. and Jones P. A. (1992) Methylation of CpG-island-containing genes in human sperm, fetal and adult tissues. Gene **114:** 203–210
116 Rideout W. M., Eversole-Cire P., Spruck C. H., Hustad C. M., Coetzee G. A., Gonzales F. A. et al. (1994) Progressive increases in the methylation status and heterochromatinization of the MyoD CpG island during oncogenic transformation. Mol. Cell. Biol. **14:** 6143–6152

117 Tu Y., Tornaletti S. and Pfeifer G. P. (1996) DNA repair domains within a human gene: selective repair of sequences near the transcription initiation site. EMBO J. **15:** 675–683
118 Bird A., Taggart M. H., Nicholls R. D. and Higgs D. R. (1987) Non-methylated CpG-rich islands at the human α-globin locus: implications for the evolution of the α-globin pseudogene. EMBO J. **6:** 999–1004
119 Matsuo K., Clay O., Takahashi T. Silke J. and Schaffner W. (1993) Evidence for erosion of mouse CpG islands during mammalian evolution. Somat. Cell Mol. Genet. **19:** 543–555
120 Antequera F. and Bird A. (1999) CpG islands as genomic footprints of promoters that are associated with replication origins. Curr. Biol. **9:** R661–R667
121 Clayton D. A. (1991) Replication and transcription of vertebrate mitochondrial DNA. Ann. Rev. Cell Biol. **7:** 453–478
122 Decker R. S., Yamaguchi M., Possenti R. and DePamphilis M. L. (1986) Initiation of Simian Virus 40 DNA replication in vitro: Aphidicolin causes accumulation of early-replicating intermediates and allows determination of the initial direction of DNA synthesis. Mol. Cell. Biol. **6:** 3815–3825
123 Segurado M., Gómez M. and Antequera F. (2002) Increased recombination intermediates and homologous integration hot spots at DNA replication origin. Mol. Cell **10:** 907–916
124 Cross S., Kovarik P., Schmidtke J. and Bird A. (1991) Nonmethylated islands in fish genomes are GC-poor. Nucleic Acids Res. **19:** 1469–1474
125 Armes, N., Gilley, J. and Fried, M. (1997) The comparative genomic structure and sequence of the surfeit gene homologs in the puffer fish *Fugu rubripes* and their association with CpG-rich islands. Genome Res. **7:** 1138–1152
126 McQueen H. A., Fantes J., Cross S. H., Clark V. H., Archibald A. L. and Bird A. (1996) CpG islands of chicken are concentrated on minichromosomes. Nat. Genet. **12:** 321–324
127 Ponger L., Duret L. and Mouchiroud D. (2001) Determinants of CpG islands: expression in early embryo and isochore structure. Genome Res. **11:** 1854–1860

To access this journal online:
http://www.birkhauser.ch