PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

CpG Islands in Genes Showing Tissue-Specific Expression

Yvonne H. Edwards

Phil. Trans. R. Soc. Lond. B 1990 **326**, 207-215
doi: 10.1098/rstb.1990.0005

Phil. Trans. R. Soc. Lond. B **326**, 207–215 (1990)
Printed in Great Britain

CpG islands in genes showing tissue-specific expression

By Yvonne H. Edwards

MRC Human Biochemical Genetics Unit, The Galton Laboratory, University College London, Wolfson House, 4 Stephenson Way, London, NW1 2HE, U.K.

Patterns of DNA methylation at CpG dinucleotides and their relations with gene expression are complex. Methylation-free CpG clusters, so-called HTF islands, are most often associated with the promoter regions of housekeeping genes, whereas genes expressed in a single-cell type are usually deficient in these sequences. However, in the human carbonic anhydrase (CA) gene family, both the ubiquitously expressed CAII and the muscle specific CAIII appear to have such CpG islands although erythrocyte-specific CAI does not. The CAII island is quantitatively more CpG rich than that of CAIII, with a CpG:GpC ratio of 0.94 compared with 0.82 for CAIII. Estimation of CpG:GpC ratios in the proximal-promoter regions of 44 vertebrate genes suggest that 40% of genes with tissue-specific or limited tissue distribution may show methylation-free CpG clusters in their promoter regions. In many cases the CpG:GpC ratio is less than that found in housekeeping genes and this may reflect variation in the interaction of CpG clusters with regulatory factors that define different patterns of tissue expression.

1. Introduction

Vertebrate DNA contains 5-methylcytosine, which is found almost exclusively in CpG dinucleotides (Gruenbaum *et al.* 1981). Many studies indicate that CpG methylation is variable and that methylation changes are associated with changes in gene expression (Razin & Riggs 1980; Cedar 1988, and references cited therein). Four classes of methylatable CpG can be distinguished.

1. CpGs that are clustered at the 5′ ends of genes and are permanently unmethylated in all cell types, including germ cells, and regardless of gene expression – so-called 'HTF islands' (Bird 1986, 1987). These CpGs are only methylated on the inactive X-chromosome (Toniolo *et al.* 1988).

2. CpGs that are unmethylated in active genes and bind methylation-sensitive transcription factors (Watt & Molloy 1988; Murray & Grosveld 1987). In these cases, undermethylation is a prerequisite of gene expression and, in contrast to HTF islands, these sites are methylated in non-expressing tissues. This category may be identical with that class of sites that, when demethylated by azacytidine, gives rise to gene expression in cells which are normally non-expressing (Jones 1985).

3. CpGs that are scattered along the length of the gene and show an inverse correlation between methylation and gene activity (for review see Crisly & Scangos (1986); Cedar (1988)). In these cases methylation–demethylation seems to be secondary to gene repression–expression, but might be involved in setting the functional state by consequential alteration to DNA conformation and chromatin structure (Buschhausen *et al.* 1987; Keshet *et al.* 1986).

4. CpGs that are methylated or not in a manner inconsequential to gene expression; for example, in the albumin and α-fetoprotein genes (Kunnath & Locker 1983) and at the 3′ end of the adenine phosphoribosyl transferase gene (Keshet *et al.* 1985).

2. CpG islands

This paper focuses on the first of these categories, the clustered CpGs. In HTF islands, the G+C content is relatively high, greater than 60%, and CpG dinucleotides occur at a frequency expected on the basis of random dinucleotide occurrence in DNA. This contrasts with bulk genomic DNA where G+C content is about 40% and where CpGs are five times under-represented. Their depletion seems to be the consequence of cytosine methylation and subsequent loss of methylcytosine by its conversion to thymine. The relatively high density of CpGs in HTF islands has been achieved by continuous protection from methylation, and indeed, lack of methylation in all cell types is a diagnostic feature of HTF islands (Bird 1986, 1987). CpG clusters are found most often at the 5′ ends of 'housekeeping' genes and are thought to play a role in their constitutive expression. However, this picture is confused by the finding that a small number of tissue-specific genes such as the α-globin genes (Bird *et al.* 1987), the type II major histocompatibility complex gene (Tykocinski & Max 1985) and the retinol binding-protein gene (d'Onofrio *et al.* 1985) are associated with candidate HTF islands.

Some of the structural features of CpG clusters and their distribution amongst different genes can be illustrated by considering a single gene family, the carbonic anhydrases (CA).

3. Methylation patterns in carbonic anhydrase genes

The human CA gene family provides a comparative system for characterizing DNA methylation patterns as its member genes show diverse tissue and organelle-specific expression (see table 1 for summary and Tashian & Hewett-Emmett (1984) for review). Furthermore, four of the CA genes have been cloned (Barlow *et al.* 1987; Lloyd *et al.* 1987; Shapiro *et al.* 1987; Fernley 1988) and there is evidence for HTF islands at the 5′ ends of at least two genes; the ubiquitously expressed CAII (Shapiro *et al.* 1987) and the muscle specific, CAIII (Edwards *et al.* 1988).

Table 1. Mammalian carbonic anhydrase gene family catalytic action and tissue distribution

Carbonic anhydrase
$CO_2 + H_2O \rightleftharpoons HCO_3^- + H^+$

CAI erythroid cells (? weak others)	CAIV membrane kidney, lung
CAII erythroid cells others	CAV mitochondrial liver
CAIII skeletal muscle (rodent liver)	CAVI secreted salivary glands

In the 5′ flanking regions proximal to the CAII and CAIII genes, the G+C content is 78% and 65%, respectively. In both cases, the CpG:GpC ratio is much higher, 0.94 for CAII and 0.65 for CAIII (table 2), than is found in bulk genomic DNA where the average ratio is about 0.2. In general, the relatively high incidence of CpG within HTF islands is associated with a high frequency of restriction enzymes sites containing this dinucleotide, for both 4-base cutters, such as *Hpa*II (CCGG), and for rarer 6-base cutters such as *Sac*II (CCGCGG). This feature is illustrated in figure 1, that shows *Hpa*II digests of 5′ DNA fragments, which encompass the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

CpG islands of human CAII and CAIII. The high density of *Hpa*II sites (mapped in figure 2) is evidenced by the small sizes of the digested fragments, which vary from 26 base pairs (b.p.) to 330 b.p. The incidence of *Hpa*II sites in these CA genes is, for example, four to eightfold greater than in the human myoglobin gene (Weller *et al.* 1984), which does not contain a CpG island (figure 2).

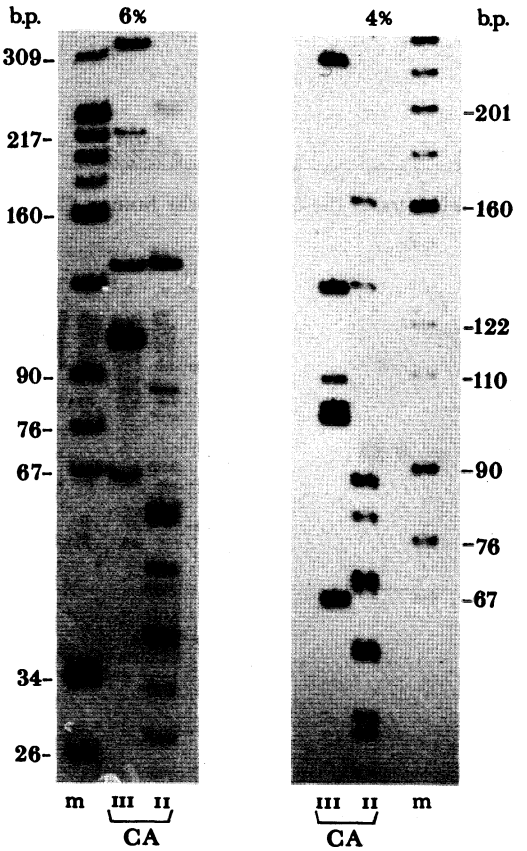

FIGURE 1. Restriction fragment patterns seen after *Hpa*II digestion of a 3.4 kb *Hin* dIII fragment from the CAIII gene (Lloyd *et al.* 1987) and a 3.2 kb *Eco*RI fragment from CAII (Shapiro *et al.* 1987). Both fragments include the promoter region and exon 1. The *Hpa*II fragments are visualised after end-labelling with the Klenow fragment of DNA polymerase I and ^{32}P dCTP and electrophoresis on 6% and 4% non-denaturing polyacrylamide gels. Lane m (markers) is a ^{32}P end-labelled *Hpa*II digest of pBR322.

The CpG cluster is maintained by protection from methylation and it is expected that each individual CpG site will be methylation free in all cells at all stages of differentiation and this includes germ cells. However, despite the fact that candidate HTF islands have been identified in most of the cloned 'housekeeping' genes, methylation status has not been established in the great majority of cases. This paucity of data is due to the difficulties associated with detecting the very small DNA fragments generated using methylation-sensitive enzymes (viz; *Hpa*II and its methylation insensitive isoschizomer *Msp* I). For example, in the analysis of the human CAIII CpG cluster, the smallest fragment detected with reasonable reproducibility by Southern blotting from agarose gels was 228 b.p. (figure 3*a*). The range of detection can be extended by using a polyacrylamide gel–electroblot procedure (Church & Gilbert 1984; Edwards *et al.* 1988) that allows the analysis of fragments ranging in size from 400 b.p. down to 50 b.p. (figure 3*b*). All the expected fragments are seen and intermediate-sized fragments,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

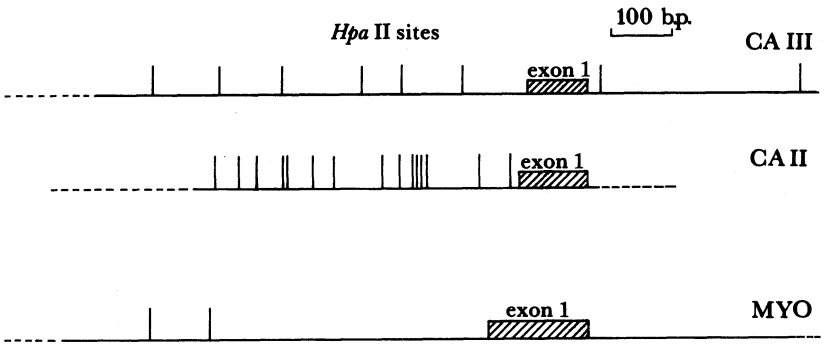

FIGURE 2. The distribution of *Hpa*II sites at the 5′ ends of the human CAII, CAIII and myoglobin genes.

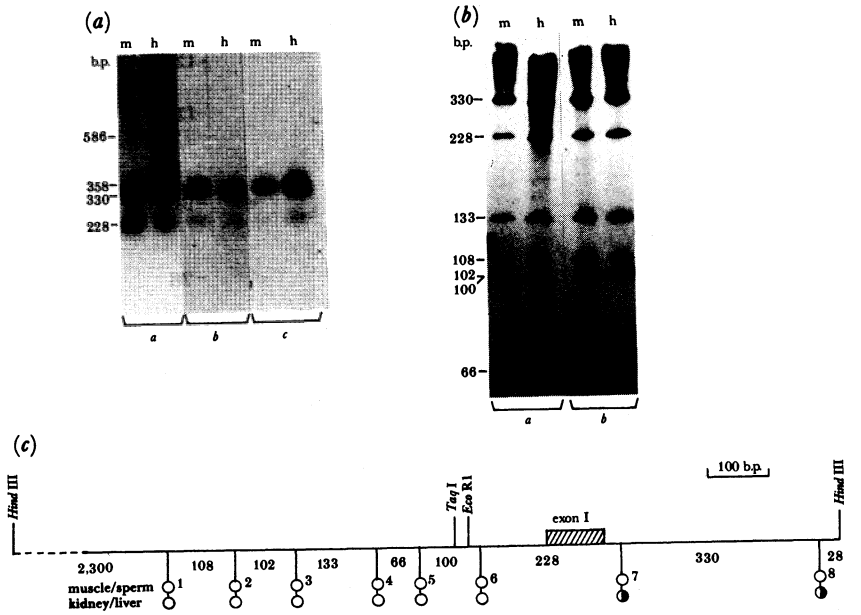

FIGURE 3. Restriction fragment patterns of the CAIII CpG cluster seen after digestion of DNA from human tissues. The DNA was digested with *Hin* dIII before digestion with *Msp*I (lane m) or *Hpa*II (lane h). (*a*) After Southern blotting from agarose gels: *a*, liver; *b*, muscle; *c*, sperm; (*b*) after electroblotting from 8% acrylamide gels: *a*, muscle; *b*, sperm; (*c*) Map showing the position and sizes of *Hpa*II fragments at the 5′ end of the CAIII gene. Methylation status at each site in DNA from muscle or sperm and kidney or liver is indicated. (○) unmethylated, (◑) partially methylated.

which would indicate methylation at one or more sites, are not. Using such a procedure it was possible, in the case of CAIII, to assess methylation at all of the *Hpa*II sequences (Edwards *et al.* 1988); it is worth bearing in mind that not all CpG sites can be tested as only about 30% occur in restriction-enzyme sites.

Our studies have shown that, like many other tissue-specific genes, the CAIII gene shows a tissue-specific pattern of methylation outside the CpG cluster; CpGs are methylated in CAIII non-expressing tissues and unmethylated in the expressing tissue and skeletal muscle (Lloyd *et al.* 1987). However, immediately upstream of exon 1 in the CpG cluster, there is no methylation in any tissue DNA regardless of expression. Methylation of this region is also absent from sperm DNA (Edwards *et al.* 1988). Thus human CAIII is in the unusual category of a tissue-specific gene with a methylation-free CpG island.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

4. CpG islands in tissue-specific genes

It is not clear what role HTF islands play in tissue-specific gene expression. The frequent association of HTF islands with genes that exhibit constitutive expression has led to the suggestion that the 5′ sequences enriched in CpGs may be involved in binding ubiquitous transcription factors and that the clustered arrangement of CpGs allows for efficient cooperative binding of multiple factors (Bird 1986). It is tempting to propose that HTF islands in tissue-specific genes are evolutionary remnants and mark an intermediary stage in the conversion from gene control mediated by 'housekeeping' factors to one involving tissue-specific factors (or vice versa). A comparison of the 5′ flanking regions of three of the carbonic anhydrase genes provides some support for this notion (table 2). CAII, which is expressed in a wide range of tissues, shows the highest G+C content CpG:GpC ratio of 0.94, whereas CAI, which shows a much more limited tissue distribution does not exhibit a CpG cluster. CAIII which is expressed solely in skeletal muscle in man but is also found in rodent liver, shows a CpG island with intermediate characteristics. In this context it is also interesting to compare the 5′ promoter sequences of the human aldolase A gene (Tsutsumi *et al.* 1985; Maire *et al.* 1987). Aldolase A contains three different promoter regions determining mRNAs of different length: a distal promoter (N), which governs the synthesis of mRNAs found in a limited range of tissues; a median promoter (M), which is only active in skeletal muscle and a proximal promoter (H), which shows ubiquitous expression. The sequence that contains the H promoter seems to contain an HTF island. It has the highest G+C content (71%) and a CpG:GpC ratio close to 1. The other two promoter regions are intermediate in these characteristics with moderately high CpG:GpC ratios of 0.44 and 0.58 for the N and M promoters, respectively (table 3). The immediate 5′ flanking region of aldolase B, a separate gene locus, expressed most abundantly in liver and kidney, but also at low levels in several other tissues, shows a low G+C content of 46% and a relatively low CpG:GpC ratio of 0.38 (table 3, data from rat).

Table 2. Incidence of CpG dinucleotides and recognition sequences for *Hpa*II and the Sp1 transcription factor in the promoter regions of the carbonic anhydrase genes

−300 to +100	CAII	CAIII	CAI
G+C%	78	65	49
CpG/GpC	0.94	0.65	0.17
*Hpa*II sites	16	8	0
Sp1 sites	9	2	1
tissue specificity	erythroid others	muscle (rodent liver)	erythroid (? weak others)

Table 3. Incidence of CpG dinucleotides and reconigition sequences for *Hpa*II and the Sp1 transcription factor in the promoter regions of aldolase A and B genes[a]

	Hu ALD.A H	Hu ALD.A N	Hu ALD.A M	Ra ALD.B
G+C%	71	62	60	46
CpG/GpC	0.94	0.44	0.58	0.38
*Hpa*II sites	1	2	2	0
Sp1 sites	6	0	1	0
expression	ubiquitous	limited	muscle	limited

[a] Sequence information from Tsutsumi *et al.* (1985) and Maire *et al.* (1987).

The actin gene family also shows 5′ promoter regions with variable CpG richness. Exceptionally, among the cloned skeletal muscle actins, the chicken-gene 5′ CpG cluster has a high CpG:GpC ratio of 0.96 (for sequence references, see Appendix 1). Some HTF islands occur sporadically within gene families. A well-characterized example is the human α-globin (Bird *et al.* 1987), which contains an unmethylated CpG cluster with a high Cpg:GpC ratio of 0.8. In contrast, mouse α-globin, human β- and γ-globins and human myoglobin all appear to have lost this CpG island.

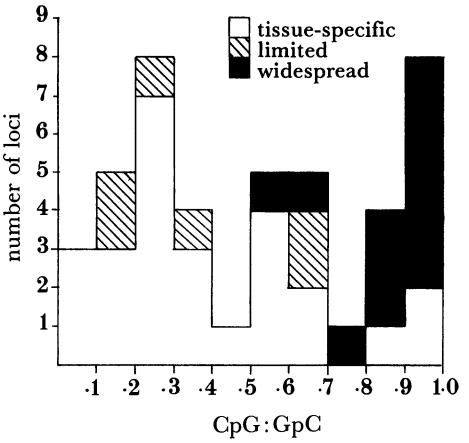

FIGURE 4. Lack of suppression of CpGs in vertebrate genes. 44 vertebrate genes plotted according to the CpG:GpC ratio in their 5′-flanking sequences and including exon 1. Between 200 and 450 b.p. were analysed for each gene. A list of the genes and references to the sequences is given in Appendix 1. The type of tissue expression is indicated.

It is clear from the data that the extent of CpG clustering at the 5′ ends of genes is quite variable and that genes cannot be simply divided into those with CpG suppression (no HTF islands) and those without CpG suppression (with HTF islands). Figure 4 shows the distribution of 44 vertebrate genes according to the CpG:GpC ratio in their immediate 5′ flanking sequences and first exons. The genes, chosen more or less at random, include 12 that are expressed ubiquitously, nine show limited tissue distribution and 23 are tissue-specific. About half of the genes are grouped around a CpG:GpC ratio of 0.2, the average for genomic DNA, and most of these are tissue-specific except for α-1-antitrypsin, phosphoglycerate mutase M, pyruvate kinase L and aldolase B which are limited in their range of expression. The second peak of 12 genes around a CpG:GpC ratio of 0.9 shows a preponderance of genes with ubiquitous expression, or 'housekeeping' genes, but also includes three that are strictly tissue-specific (chicken skeletal-muscle actin, retinol-binding protein and human α-globin).

Thus overall, the general distribution confirms the earlier conclusion that HTF islands are more commonly associated with constitutively expressed genes and that tissue-specific genes are characterized by an absence of such sequences (Bird 1986). However, amongst this collection of 44 genes, there are 12 that show intermediate levels of CpG retention, with CpG:GpC ratios ranging between 0.47 and 0.75 in their 5′ flanking sequences. Most of these are tissue-specific or show limited tissue distribution. It is difficult to be certain whether all of these genes carry bona fide methylation-free HTF islands but in three cases where detailed studies have been made, Thy1 (Kolsto *et al.* 1986), CAIII (Edwards *et al.* 1988) and DHFR (Yen *et al.* 1984), the unmethylated status of the CpG cluster has been confirmed. Gardiner-Garden & Frommer

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(1987) in their survey of vertebrate genes also reported ten 'limited expression' genes, other than those described here, with intermediate CpG:GpC ratios (Obs/Exp CpG, 0.63–0.75).

This survey suggests that a substantial proportion, perhaps as many as 40%, of genes with tissue-specific or limited tissue expression may have methylation-free CpG clusters in their promoter regions. This observation raises various questions; for example, are these promoters continuously available to transcription factors and how is expression confined to one or a few cell types? If the CpGs in the cluster are part of recognition sequences for binding factors does the density of CpGs within the cluster relate to the efficiency of binding? In other words, do the clusters of relatively low CpG:GpC ratio bind ubiquitous transcription (or methylation protection) factors less efficiently?

It has been proposed that tissue-specific control may be achieved through a mechanism in which negative control plays a major role. This implies the existence of repressor molecules in tissue where a certain gene is not expressed. The human retinol-binding protein is a possible example (Colantuoni *et al.* 1987; Bird 1987). The range of nuclear factors and the mode of their binding in CpG islands has yet to be fully established. It has been suggested that the Sp1 transcription factor whose binding site contains CpG (Kadonaga *et al.* 1986), has a special role in CpG islands, both in transcription regulation and their protection from methylation (Holler *et al.* 1988). However, these sites do not occur in multiple copies in all CpG clusters, for example, the human retinol-binding protein gene and chicken skeletal-muscle actin gene promoters each have only one Sp1 binding site. Furthermore, the Sp1 factor binds equally well to both methylated and unmethylated DNA (Holler *et al.* 1988).

Note added in proof (16 *October* 1989)

A second promoter on the CAI gene, active in the colon, with a G+C content of 40% and a CpG:GpC of 0.14 has recently been described (Fraser, P., Cummings, P. & Curtis, P. 1989 *Molec. Cell Biol.* **9**, 3308.)

I am grateful to Dr Peter Butterworth, University College London and Dr Sabura Sakoda, Osaka University Hospital for permission to quote unpublished data and would like to thank Mrs Doreen Trinder for the preparation of the typescript.

References

Barlow, J., Lowe, N., Edwards, Y. & Butterworth, P. 1987 Human carbonic anhydrase I, cDNA. *Nucl. Acids Res.* **15**, 2386–2387.

Bird, A. 1986 CpG-rich islands and the function of DNA methylation. *Nature, Lond.* **321**, 209–213.

Bird, A. 1987 CpG islands as gene markers in the vertebrate nucleus. *Trans. Genet.* **3**, 342–347.

Bird, A., Taggart, M., Nicholls, R. & Higgs, D. 1987 Non-methylated CpG-rich islands at the human α-globin locus: implications for evolution at the α-globin pseudogene. *EMBO J.* **6**, 999–1004.

Buschhausen, G., Wittig, B., Graessmann, M. & Graessmann, A. 1987 Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1177–1181.

Cedar, H. 1988 DNA methylation and gene activity. *Cell* **53**, 3–4.

Church, G. & Gilbert, W. 1984 Genomic sequencing. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995.

Colantuoni, V., Pirozzi, A., Blance, C. & Cortese, R. 1987 Negative control of liver-specific gene expression: cloned retinol binding-protein gene is repressed in Hela cells. *EMBO J.* **6**, 631–636.

Crisly, B. A. & Scangos, G. A. 1986 Control of gene expression by DNA methylation. In *Molecular genetics of mammalian cells* (ed. G. M. Malacinski), pp. 3–29. New York: Macmillan Publishing.

Edwards, Y. H., Charlton, J. & Brownson, C. 1988 A non-methylated CpG-rich island associated with the human muscle-specific carbonic anhydrase III gene. *Gene* **17**, 473–481.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Gardiner-Garden, M. & Frommer, M. 1987 CpG islands in vertebrate genomes. *J. molec. Biol.* **196**, 261–282.
Gruenbaum, Y., Stein, R., Cedar, H. & Razin, A. 1981 Methylation of CpG sequences in eukaryotic DNA. *FEBS Lett.* **124**, 67–71.
Höller, M., Westin, G., Jiricny, J. & Schaffer, W. 1981 Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Devel.* **2**, 1127–1135.
Jones, P. 1985 Altering gene expression with 5-azacytidine. *Cell* **40**, 485–486.
Kadonaga, J. T., Jones, K. A. & Tjian, R. 1986 Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends biochem. Sci.* **11**, 20–23.
Keshet, I., Lieman-Hurwitz, J. & Cedar, H. 1986 DNA methylation affects the formation of active chromatin. *Cell* **44**, 535–543.
Keshet, I., Ysraeli, J. & Cedar, H. 1985 Effect of regional DNA methylation on gene expression. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2560–2564.
Kunnath, L. & Locker, J. 1983 Developmental changes in the methylation of the rat albumin and α-fetoprotein genes. *EMBO J.* **2**, 317–324.
Kolsto, A., Kollias, G., Giguere, V., Isobe, K., Prydz, H. & Grosveld, F. 1986 The maintenance of methylation-free islands in transgenic mice. *Nucl. Acids Res.* **14**, 9667–9678.
Lloyd, J., Brownson, C., Tweedie, S., Charlton, J. & Edwards, Y. 1987 Human muscle carbonic anhydrase: gene structure and DNA methylation patterns in fetal and adult tissues. *Genes Devel.* **1**, 594–602.
Mairie, P., Gautron, S., Hakim, V., Gregori, C., Mennecier, F. & Kahn, A. 1987 Characterizations of three optimal promoters in the 5′ region of the human aldolase A gene. *J. molec. Biol.* **197**, 425–438.
Murray, E. & Grosveld, F. 1987 Site-specific demethylation in the promoter of human γ-globin gene does not alleviate methylation mediated suppression. *EMBO J.* **6**, 2329–2335.
Nudel, U., Greenberg, D., Ordahl, C., Saxel, O., Neuman, S. & Yaffe, D. 1985 Developmentally regulated expression of a chicken muscle specific gene in stably transfected at myogenic cells. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3106–3109.
d'Onofrio, C., Colantuoni, V. & Cortese, R. 1985 Structure and cell-specific expression of a cloned human retinol binding protein gene: The 5′-flanking region contains hepatoma-specific transcriptional signals. *EMBO J.* **4**, 1981–1989.
Razin, A. & Riggs, A. 1980 DNA methylation and gene function. *Science, Wash.* **219**, 604–610.
Shapiro, L., Venta, P. & Tashian, R. 1987 Molecular analysis of G+C-rich upstream sequences regulating transcription of the human carbonic anhydrase II gene. *Molec. Cell Biol.* **7**, 4589–4593.
Tashian, R. & Hewett-Emmett, D. (eds) 1984 Biology and chemistry of the carbonic anhydrases. *Ann. N.Y. Acad. Sci.*, vol. 249.
Toniolo, D., Martini, G., Migeon, B. & Dono, R. 1988 Expression of the G-6-PD locus on the human X chromosome is associated with demethylation of three CpG islands within 100 kb of DNA. *EMBO J.* **7**, 401–406.
Tsutsumi, K., Mukai, T., Tsutsumi, R., Hidaka, S., Arai, Y., Hori, K. & Ishikawa, K. 1985 Structure and genomic organization of the rat aldolase B gene. *J. molec. Biol.* **181**, 153–160.
Tykocinski, M. & Max, E. 1984 CpG clusters in MHC genes and 5′ demethylated genes. *Nucl. Acids Res.* **12**, 4385–4396.
Watt, F. & Molloy, P. 1988 Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Devel.* **2**, 1136–1143.
Weller, P., Jeffreys, A., Wilson, V. & Blanchelot, A. 1984 Organization of the human myoglobin gene. *EMBO J.* **2**, 439–446.
Yen, P. H., Patel, P., Chinault, A. C., Mohandas, T. & Shapiro, L. J. 1984 Differential methylation of hypoxanthine phosphoribosyl-transferase genes on active and inactive human X chromosomes. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1759–1763.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Appendix 1. List of genes and references to their sequences

locus[a]	expression[b]	references
Hu adenosine deaminase	W	Wiginton *et al.* (1986) *Biochemistry* **25**, 8234.
Hu mt aldehyde dehydrogenase	W	Hsu *et al.* (1988) *Genomics* **2**, 57.
rat skeletal muscle actin	S	Nudel *et al.* (1985) *Proc. natn. Acad. Sci. U.S.A.* **82**, 3106.
Hu cardiac actin	S	Minty *et al.* (1986) *Molec. cell Biol.* **6**, 2125.
Chk skeletal muscle actin	L	Nudel *et al.* (1985) *Proc. natn. Acad. Sci. U.S.A.* **82**, 3106.
rat aldolase B	L	Tsutsumi *et al.* (1985) *J. molec. Biol.* **181**, 153.
Hu aldolase A	W/L/S	Mairie *et al.* (1987) *J. molec. Biol.* **197**, 425.
Hu α-1 antitrypsin	L	Long *et al.* (1984) *Biochemistry* **23**, 4828.
Hu carbonic anhydrase I	S	Butterworth *et al.* (personal communication).
Hu carbonic anhydrase II	W	Shapiro *et al.* (1987) *Molec. Cell Biol.* **7**, 4589.
Hu carbonic anhydrase III	S	Lloyd *et al.* (1987) *Genes Devel.* **1**, 594.
Mo creatine kinase M	L	Jaynes *et al.* (1986) *Molec. Cell Biol.* **6**, 2855.
Hu creatine kinase B	L	Maruman *et al.* (1987) *Genomics* **1**, 126.
Mo α-A-crystallin	S	King *et al.* (1983) *Cell* **32**, 707.
Mo dihydrofolate reductase	W	Crouse *et al.* (1982) *J. biol. Chem.* **257**, 7887.
Hu glucose-6-phosphate dehydrogenase	W	Toniolo *et al.* (1984) *EMBO J.* **3**, 1987.
Hu β-globin	S	Efstratiadis (1980) *Cell* **21**, 653.
Hu γ-globin	S	Efstratiadis (1980) *Cell* **21**, 653.
rabbit β-globin	S	Efstratiadis (1980) *Cell* **21**, 653.
Hu γ-globin	S	Baralle *et al.* (1980) *Cell* **21**, 621.
Hu α-globin	S	Proudfoot *et al.* (1980) *Cell* **21**, 537.
Mo α-globin	S	Nishioka *et al.* (1979) *Cell* **18**, 875.
Bo growth hormone	S	Woychik *et al.* (1982) *Nucl. Acids Res.* **10**, 7197.
Hu growth hormone	S	Woychik *et al.* (1982) *Nucl. Acids Res.* **10**, 7197.
rat growth hormone	S	Woychik *et al.* (1982) *Nucl. Acids Res.* **10**, 7197.
Hu glutathione-S-transferase	W	Cowell *et al.* (1988) *Biochem. J.* **255**, 79.
Hu haptoglobin	S	Benoi *et al.* (1985) *EMBO J.* **4**, 119.
Mo hypoxanthine guanine ribosyltransferase	W	Melton *et al.* (1986) *Cell* **44**, 319.
Hu myoglobin	S	Weller *et al.* (1984) *EMBO J.* **3**, 439.
Chk myosin heavy chain f	S	Kropp *et al.* (1986) *J. biol. Chem.* **261**, 6613.
rat myosin light chain 1f	S	Robert *et al.* (1986) UCLA Symp. *Molec. Cell Biol.* **29**, 487.
rat myosin light chain 3f	S	Robert *et al.* (1986) UCLA Symp. *Molec. Cell Biol.* **29**, 487.
Mo nerve growth factor α	S	Evans *et al.* (1985) *EMBO J.* **4**, 133.
Mo nerve growth factor γ	S	Evans *et al.* (1985) *EMBO J.* **4**, 133.
Hu tissue plasminogen activator	W	Degen *et al.* (1986) *J. biol. Chem.* **261**, 6972.
rat phosphoenol pyruvate carboxy kinase	S	Short *et al.* (1986) *J. biol. Chem.* **261**, 9721.
Hu phosphoglycolate mutase M	L	Sakoda & Schon (personal communication).
Hu phosphoglycerate kinase X	W	Singer-Sam *et al.* (1984) *Gene* **32**, 409.
Chk pyruvate kinase M_1	L	Lonberg *et al.* (1985) *Cell* **40**, 81.
rat pyruvate kinase L	L	Cognet *et al.* (1987) *J. molec. Biol.* **196**, 11.
Hu retinol binding protein	S	d'Onofrio *et al.* (1985) *EMBO J.* **4**, 1981.
Mo Thy 1 antigen	S	Giguere *et al.* (1985) *EMBO J.* **4**, 2017.
rat tyrosine aminotransferase	S	Shinomiya *et al.* (1984) *Proc. natn. Acad. Sci. U.S.A.* **81**, 1346.
Hu superoxide dismutase A	W	Levanon *et al.* (1985) *EMBO J.* **4**, 77.

[a] Hu, human; Chk, chicken; Mo, mouse; Bo, bovine.
[b] W, widespread; L, limited; S, tissue specific.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

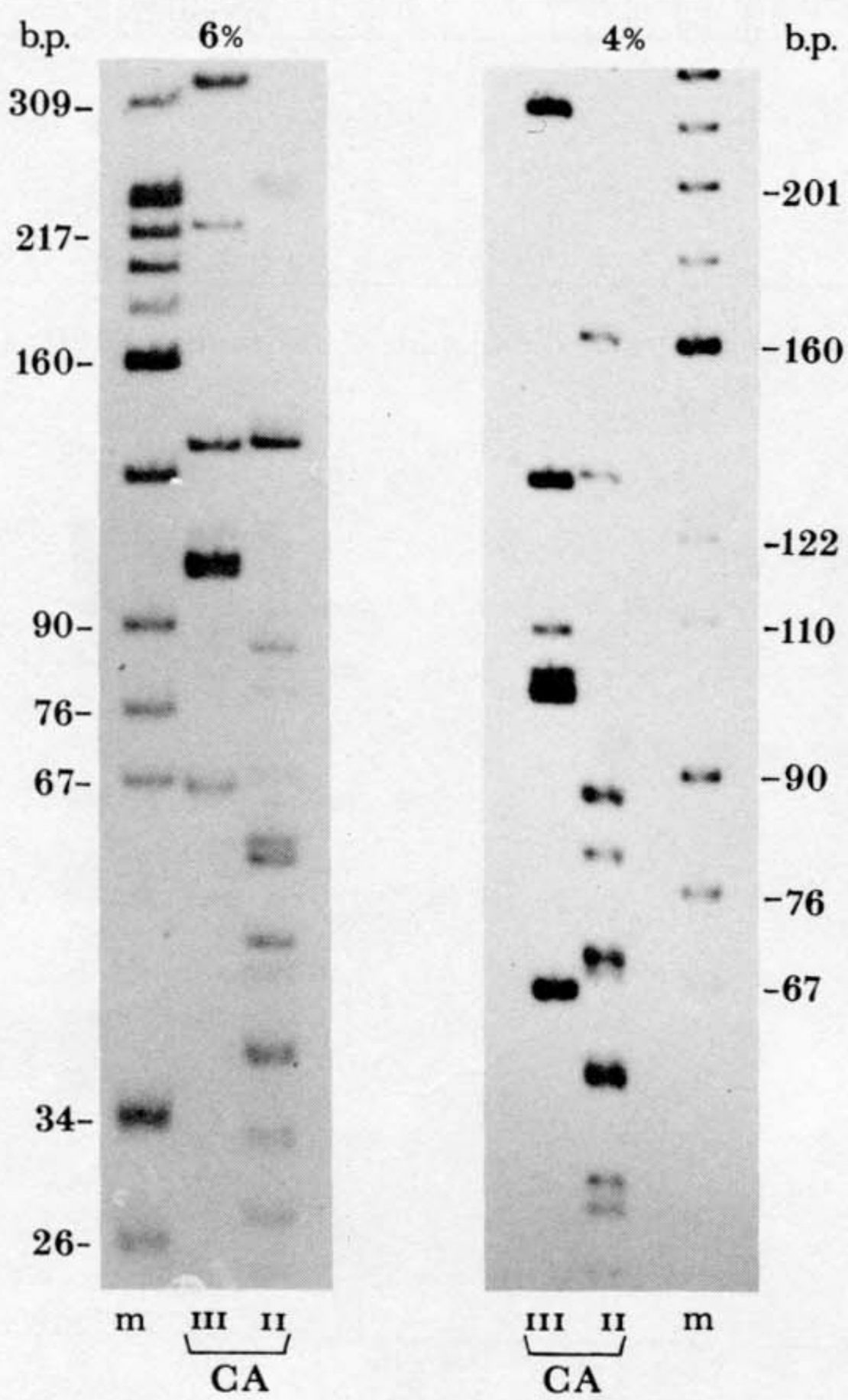

FIGURE 1. Restriction fragment patterns seen after *Hpa*II digestion of a 3.4 kb *Hin* dIII fragment from the CAIII gene (Lloyd *et al.* 1987) and a 3.2 kb *Eco*RI fragment from CAII (Shapiro *et al.* 1987). Both fragments include the promoter region and exon 1. The *Hpa*II fragments are visualised after end-labelling with the Klenow fragment of DNA polymerase I and ^{32}P dCTP and electrophoresis on 6% and 4% non-denaturing polyacrylamide gels. Lane m (markers) is a ^{32}P end-labelled *Hpa*II digest of pBR322.

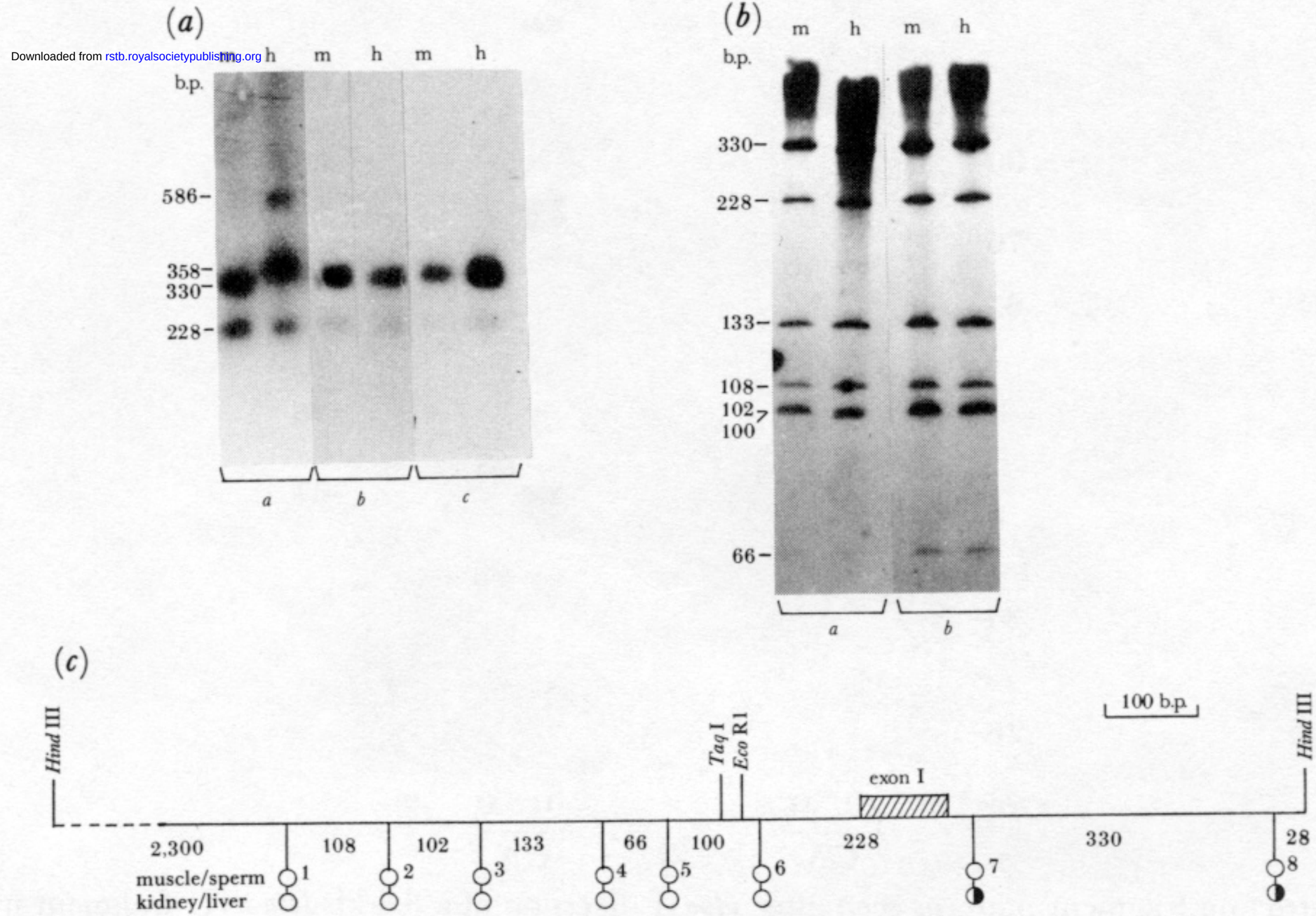

FIGURE 3. Restriction fragment patterns of the CAIII CpG cluster seen after digestion of DNA from human tissues. The DNA was digested with *Hin* dIII before digestion with *Msp*I (lane m) or *Hpa*II (lane h). (*a*) After Southern blotting from agarose gels: *a*, liver; *b*, muscle; *c*, sperm; (*b*) after electroblotting from 8% acrylamide gels: *a*, muscle; *b*, sperm; (*c*) Map showing the position and sizes of *Hpa*II fragments at the 5′ end of the CAIII gene. Methylation status at each site in DNA from muscle or sperm and kidney or liver is indicated. (○) unmethylated, (◑) partially methylated.