Non-methylated islands in fish genomes are GC-poor

Sally Cross*+, Pavel Kovarik§, Jorg Schmidtke[1] and Adrian Bird+
Research Institute for Molecular Pathology, Dr. Bohr Gasse 7, 1030 Vienna, Austria and
[1]Department of Human Genetics, Medizinsche Hochschule, Konstanty-Gutschow Strasse 8, D-3000 Hannover 61, FRG

Received January 11, 1991; Revised and Accepted March 1, 1991

ABSTRACT

In the vertebrate genomes studied to date the 5′ end of many genes are associated with distinctive sequences known as CpG islands. CpG islands have three properties: they are non-methylated; the dinucleotide CpG occurs at the frequency predicted by base composition; and they are GC-rich. Unexpectedly we have found that CpG islands in certain fish only have the first two properties; that is, their GC-content is not elevated compared to bulk genomic DNA. Based on this finding, we speculate that the GC-richness of CpG islands in vertebrates other than fish is a passive consequence of a higher mutation rate in regions of open chromatin under conditions where the nucleotide precursor pools are biased.

INTRODUCTION

The major known modification of DNA in animals is methylation of cytosine residues in the dinucleotide CpG. Methylation of DNA is found in both invertebrates and vertebrates but its extent and distribution differs between the two groups. In invertebrate genomes, long tracts of DNA are stably methylated at most CpGs while the remainder of the genome is stably non-methylated. Genes are found in the non-methylated tracts which make up the majority of the genome. Some invertebrate organisms, for example *Drosophila*, have undetectable levels of methylation. Unlike the invertebrates, vertebrate genomes are highly methylated, with 60–90% of CpG's containing a methylated cytosine (reviewed in 1). The dinucleotide CpG is under-represented within bulk DNA, occurring at about 20% of the expected frequency. CpG suppression is thought to be due to mutation following deamination of methylated cytosine to thymine (2).

Although most vertebrate genomes are globally heavily methylated and relatively AT-rich (typically 40% GC), they are found to contain short stretches (0.5–2kb) of non-methylated GC-rich DNA known as CpG (or HTF for HpaII Tiny Fragment) islands (1). These are associated with the 5′ ends of housekeeping genes and of some tissue-specific genes. CpG islands have an open chromatin structure and may mark sites of interaction between transcription factors and promoters (3). They are normally constitutively non-methylated, but become methylated on the inactivated X chromosome (4) and at the promoters of non-essential genes in established cultured cell-lines (5). One consequence of the non-methylated state of islands is that the frequency of the dinucleotide CpG is not suppressed as it is in most of the genome, but is maintained at the level predicted by base composition. This, together with their relative GC-richness (60–80% GC), means that CpG islands contain many sites for methylation-sensitive restriction enzymes which are rare and blocked by methylation in the rest of the genome.

In this study we have used a simple end-labelling assay to test a wide range of vertebrates, notably some fish species, for the presence of CpG islands. Using this assay we had previously concluded that all vertebrates have CpG islands in their genomes (6). In this more extensive study, however, several fish species did not appear to have a typical 'HTF fraction' in their genome. This prompted us to investigate in fish species the 5′ ends of two housekeeping genes which have been shown to have islands in other vertebrate species. The analysis showed that the fish species do in fact have CpG islands which are unmethylated, but they differ from those in mammals and birds in that they are not GC-rich.

MATERIALS AND METHODS

End-labelling and DNA isolation

The end-labelling method used was essentially as described in ref 6. Genomic DNA used for end-labelling, PCR and Southern blots was isolated using standard procedures.

Amplification of 5′ end of trout metallothionein gene

Polymerase chain reactions were carried out as described in ref 7 with some modifications. Reactions contained 40 mM KCl, 10 mM Tris (pH 8.3), 1.5 mM $MgCl_2$, 200 μM of each dNTP, 1 μM of each primer and 2.4 units of TaqI polymerase (Cetus Corporation) in a 50 μl reaction mix. A fragment of 630 bp from the 5′ end of the trout metallothionein B gene was amplified from

* To whom correspondence should be addressed
Present addresses: +Institute of Cell and Molecular Biology, University of Edinburgh, Darwin Building, Kings Buildings, Edinburgh EH9 3JR, UK and §Institute of Biochemistry, University of Vienna, Vienna, Austria

100ng of genomic DNA using the oligonucleotides metal-3 (5'-GCGAATTCCTGATTAAGTTTTGTATAGTTAA-3') and metal-4 (5' GCCTGCAGACGTGCCACCAAGTCACCA-TTGA -3') as primers. These primers were chosen from published sequence data (8). Thirty-five cycles of amplification of 1 min at 94°C followed by 2 min at 60°C and 3 min at 72°C were performed. When the 630 bp fragment was gel-purified and used as a probe it was found not to be single-copy. Therefore another PCR reaction was used to amplify the 5' 520 bp sequence from 0.1 ng of the 630 bp fragment using oligonucleotides metal-3 and metal-8 (5'-ACTTTTCTTACAACTGGTGC-3') as primers. Fifteen cycles of amplification of 1 min at 94°C followed by 2 min at 35°C and 3 min at 72°C were performed. The 520 bp fragment was gel-purified and used as a single-copy probe for Southern blots.

Amplification and sequencing of 5' end of carp β-actin gene

The 5' end of the carp β-actin gene was amplified using PCR as described above with the following changes. Oligonucleotides carp-1 (5'-GGCGAATTCGTCGACAAACCCCCCCAAACC-TAAG-3') and carp-2 (5'-GCCGGATCCCCAGTGCGGC-AATTTCATCATCCAT-3') were used as primers. GC clamps and cloning sites were included at the 5' ends of both oligonucleotides. These primers were chosen from the available sequence of the common carp β-actin gene (EMBL Accession No. M24113). Genomic DNA (100ng) was used as a substrate for 35 cycles of amplification of 1 min at 94°C followed by 2 min at 55°C and 5 min at 72°C. The expected fragment of 1.3 kb was generated and after restriction with BamHI and EcoRI it was sub-cloned into pUC19 using standard procedures. The sub-cloned fragment was used as a probe for Southern blots. In order to confirm that the PCR products were derived from the β-actin gene, four separate PCR products were sequenced from both ends using T7 DNA polymerase (Pharmacia) following the instructions of the suppliers and compared to the published common carp β-actin sequence. Each clone was identical at one end. This sequence (193 bp) has 70% identity with the published common carp β-actin sequence. Three of the PCR clones were identical at the other end and one was different. As there were multiple differences between the two sequences they are unlikely to be PCR artifacts. There was 90% similarity and 90% identity between these two sequences so it is probable that the two different PCR fragments were derived from two different alleles. When compared to the common carp β-actin sequence one sequence (202 bp) had 94% identity and the other (183 bp) had 85.4% identity. The sequence of the PCR clones also has a similar identity with the published grass carp β-actin sequence (9), data not shown.

Southern analysis

Ten μg of digested genomic DNA was used per lane of 1.0–2.0% agarose gels which were transferred to Genescreen after electrophoresis using standard procedures. For denaturing polyacrylamide gels 30 μg of digested genomic DNA was precipitated and redissolved in formamide loading buffer prior to loading. After electrophoresis the DNA fragments were electroblotted onto Genescreen using a LKB Multiphor II Electrophoresis System. Filters were hybridized at 68°C in 0.5 M Na_2HPO_4 (pH 7.2), 7% SDS, 0.5% dried milk powder. All probes were labelled by the random priming method (10). Filters were washed at 68°C with 0.1×SSC, 0.1% SDS, except when using the metallothionein 520 bp probe when filters were washed at the same temperature but in 2×SSC, 0.1% SDS.

Computer analysis

All sequence analysis was carried out using the University of Wisconsin sequence analysis software package (11).

RESULTS

Some vertebrate species have a reduced or undetectable HTF fraction

Genomic DNA of vertebrates is heavily methylated and hence is poorly cleaved by methylation-sensitive restriction enzymes such as HpaII. A small fraction of the genome that is derived from CpG islands can, however, be detected by end-labelling HpaII digests of genomic DNA (6). HpaII cuts frequently in the GC-rich, methylation-free islands and rarely in the rest of the genome. Figure 1A shows the result of end-labelling HpaII digests of genomic DNA from several vertebrate species. The patterns obtained fall into three groups. The pattern of the first group which comprises crocodile (lane 1) and frog (lane 2) is typical of the patterns obtained previously with bird and mammalian DNAs. Unexpectedly the second group (lanes 3–6) which

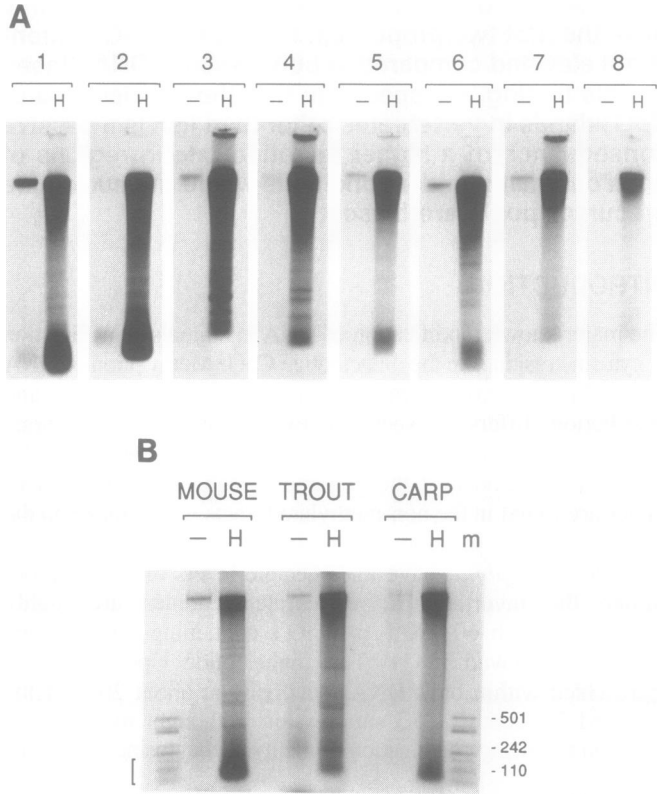

Figure 1. HTF fractions in different vertebrate species. Genomic DNA, undigested (–) or digested with HpaII (H) was end-labelled, resolved by electrophoresis on 1.2% agarose gels and autoradiographed to compare released HTF fractions. Undigested samples served to control for the extent of nicks and broken ends in the DNA which would also be labelled. (A) 1, *Caiman crocodilus* (crocodile); 2, *Odontophynus am.* (frog); 3, *Osmerus eperlanus* (smelt); 4, *Tinca tinca* (tench); 5, *Coregonus fera* (whitefish species); 6, *Salmo gairdneri irideus* (American steelhead); 7, *Leuciscus cephalus* (chub); 8, *Rutilus rutilus* (roach). The variation of DNA separation between tracks is because the data shown are from several different end-labelling experiments. (B) Comparison of mouse, trout and carp. Bracket indicates HTF fraction in the mouse. m = marker DNA and sizes are in bp.

includes the fish *Tinca tinca* (lane 4) has a much reduced HTF fraction, and the third group (lanes 7 and 8) which includes the fish *Rutilus rutilus* (lane 8) does not have a detectable HTF fraction at all. In a separate experiment we extended the analysis to two more fish species. Figure 1B shows the result of end-labelling HpaII digests of genomic DNA from trout, carp, and for comparison, mouse. Trout DNA did not show a prominent HTF fraction, while carp DNA gave an HTF fraction that was reduced compared to that in mouse.

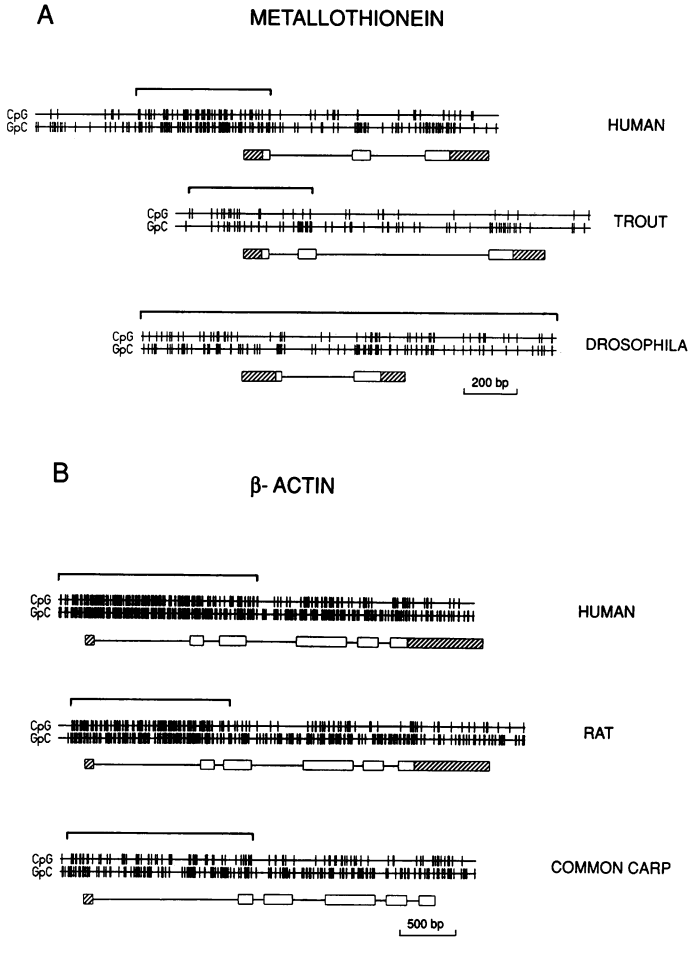

Figure 2. CpG frequency in metallothionein and β-actin genes. The diagram shows the positions of CpG and GpC dinucleotides and the brackets indicate regions where CpG occurs at the frequency predicted by base composition. The 5′ ends of the transcripts are aligned. Boxes represent exons, open and cross-hatched portions representing translated and untranslated regions respectively. (A) Comparison of metallothionein genes from human (metallothionein-II; 12), trout (metallothionein B; 8) and *Drosophila* (25). (B) Comparison of β-actin genes from human (13, 14), rat (15) and common carp (EMBL Accession No. M24113).

Island regions of the β-actin and metallothionein genes

The end-labelling data could be explained in two ways. Either some vertebrate species have few CpG islands, or they have CpG islands which have a different structure from the typical mammalian-type island and this renders them undetectable using the end-labelling approach. Since all known housekeeping genes in mammalian species have CpG islands, we attempted to distinguish these possibilities by comparing the structure of the 5′ regions of specific housekeeping genes in fish species with reduced HTF fractions and in mammals. We chose to analyze two genes for which sequence data was available in both mammalian and fish species; the metallothionein and β-actin genes. Both of these genes have prominent CpG islands in several mammalian species based on published sequence data (12, 13, 14, 15). Figure 2 shows comparisons of the frequency of the dinucleotides CpG and GpC in the metallothionein genes of human, trout and *Drosophila*, and in the β-actin genes of human, rat and common carp. Table 1 shows the base composition (% GC) of the island and non-island portions for each gene, and the observed and expected frequencies of the dinucleotides CpG and GpC in the two portions.

The human metallothionein II gene has a CpG island which covers the promoter region and the first exon. The island can easily be seen because there is a dense clustering of CpGs at the 5′ end of the gene which coincides with a region where the number of CpGs equals the number of GpCs. The data in Table 1 confirm that CpG occurs at the expected frequency in this region. In the equivalent part of the trout metallothionein B gene there is no striking clustering of CpGs, but in fact the numbers of CpGs and GpCs in this region is the same (Table 1). Table 1 confirms that CpG occurs here at the expected frequency. The reason for the difference between the number of CpGs in the human and trout islands becomes apparent if the GC-contents of the two islands are compared (Table 1): the human island has a high GC-content (68%) and the trout island has a low GC-content (40%). The discrepancy in number of CpGs present in the two islands merely reflects this difference in their GC-contents. Outside the islands the human and trout gene are similar,

Table 1. Base composition, CpG and GpC frequency of island and non-island regions of the metallothionein and β-actin genes. The per cent GC content (% G+C); the number of observed (o) and expected (e) occurences of the dinucleotides CpG and GpC; and the observed/expected frequency (o/e) of both dinucleotides in the island and non-island parts of the genes are given. Regions with CpG island-like character are marked by brackets in Figure 2.

	ISLAND						NON-ISLAND					
	%G+C	CpG number o	CpG number e	CpG o/e	GpC number o	GpC number e	GpC o/e	%G+C	CpG number o	CpG number e	CpG o/e	GpC number o	GpC number e	GpC o/e
human metallothionein II	68.42	54	59	0.9	70	59	1.2	50.33	29	76	0.4	77	76	1.0
trout metallothionein B	39.91	20	18	1.1	25	18	1.4	39.11	12	41	0.3	39	41	0.9
Drosophila metallothionein	42.84	60	71	0.9	91	71	1.3	–	–	–	–	–	–	–
human β-actin	70.64	201	218	0.9	242	218	1.1	52.83	57	133	0.4	122	133	0.9
rat β-actin	65.83	140	151	0.9	163	151	1.1	48.74	54	161	0.3	150	161	0.9
carp β-actin	44.46	67	83	0.8	103	83	1.2	46.07	46	123	0.4	112	123	0.9

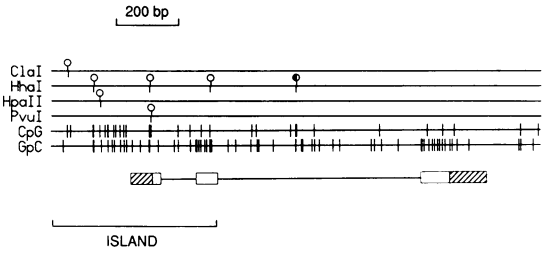

Figure 3. Methylation status of the trout metallothionein gene. Boxes represent exons, open and cross-hatched portions representing translated and untranslated portions respectively. Open and half solid circles denote unmethylated and partially methylated sites respectively. The bracket indicates the putative CpG island. The probe used coincides with the island. HaeII was used to determine the methylation status of the HhaI site which is shown as partially methylated.

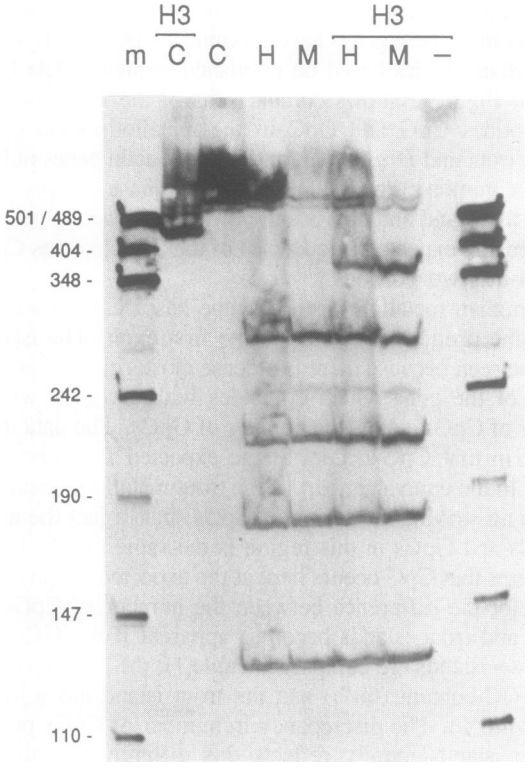

Figure 4. Methylation status of the carp β-actin gene. Carp genomic DNA was digested as indicated; C = CfoI, H3 = HindIII, H = HpaII, M = MspI. The samples were electrophoresised on a 6% 8M urea polyacrylamide gel and transferred to Genescreen. The filter was probed with the sub-cloned carp 1.3 kb PCR product. The same pattern of fragments is detected if the mixed PCR product of 1.3 kb is used as a probe. m = marker and sizes are in bp.

both having a low GC-content and a suppressed frequency of the dinucleotide CpG (compare GpC and CpG frequencies in Table 1). For comparison the *Drosophila* gene is also shown in Figure 2. Table 1 shows that the *Drosophila* metallothionein gene has a GC-content of 40% and that CpG occurs throughout at the frequency predicted by base composition, as would be expected for an organism in which there is no methylation. The 5′ end of the trout gene resembles the 5′ end of the *Drosophila* gene in GC composition and numbers of CpGs, but elsewhere it resembles the human gene.

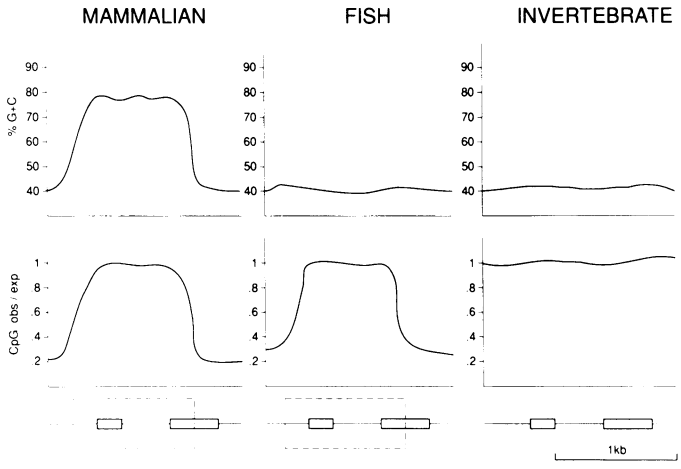

Figure 5. GC-content and CpG frequency of a housekeeping gene. A schematic representation of GC-content (% G+C) and frequency of the dinucleotide CpG (CpG obs/exp) of an imaginary house-keeping gene in three animal groupings, mammalian, fish and invertebrate are shown. Open boxes represent exons and dashed boxes represent islands. Imaginary genes are used to facilitate comparison, but the profiles are based upon available data.

Turning to the β-actin genes, an equivalent difference between the mammalian and fish genes can be seen. The human and rat β-actin genes have obvious CpG islands which are GC-rich (65–70% GC) and in which CpG occurs at the expected frequency. The islands cover the promoter region and part of the transcribed region. In the corresponding part of the common carp β-actin gene, CpG also occurs at the frequency predicted by base composition, while elsewhere in the gene and its flanking sequences CpG is suppressed. The GC-content of the putative carp β-actin island, however, is low (44% GC), and consequently the density of CpGs is not as dramatically enriched in this region as it is in the rat and human genes.

The island-like regions are unmethylated

One prediction arising from the fact that CpG occurs at the expected frequency in the 'GC-poor' islands of these fish genes is that CpGs in these regions will be non-methylated. In order to test the methylation status of the island-like region of the trout metallothionein B gene a 520 bp fragment from the 5′ end of the gene was used to probe various methylation-sensitive restriction digests of trout genomic DNA (data not shown). Figure 3 is a map showing the position of unmethylated and partially methylated sites in trout genomic DNA derived from this analysis. Although only a small proportion of the CpG's could be tested by restriction enzymes, six CpGs in the putative island-like region were found to be unmethylated. One CpG mapping to the 3′ side of the island-like region was found to be partially methylated. These data suggest that the island-like region is non-methylated.

A similar strategy was used to test the methylation status of the island-like region at the 5′ end of the β-actin gene in carp genomic DNA. A fragment of the expected size of 1.3 kb was amplified by PCR from the 5′ end of the gene and sub-cloned into pUC19. The sub-cloned PCR fragment was shown to be single-copy (data not shown) and it was then used to probe a Southern blot of various digests of carp genomic DNA (Figure 4). The patterns generated with HpaII and its methylation-insensitive isoschizomer MspI are the same, indicating that the genomic region was unmethylated. The sizes of the HpaII and

CfoI fragments detected were not the same as those expected from the published common carp β-actin sequence, but sequencing of the fragments generated by PCR showed that they are similar, but not identical to, the available common carp β-actin sequence. Therefore it is likely that the carp DNA which was used for the PCR and mapping is from a different variety of carp from that which was used for the published cloning and sequencing. Another possibility is that since carp is polyploid (16) this gene represents a diverged allele of the previously cloned gene. The differences between the sequence of the fragment generated by PCR and the published sequence are all in the first intron of the gene where there are presumably fewer constraints on sequence changes. Whatever the reason for the differing sequences the analysis nevertheless indicates that the 5′ end of the β-actin gene in the carp variety used here has a similar GC-content (40%) and lack of CpG suppression to that found in the published sequences of the common and grass carp β-actin genes. Therefore the β-actin gene of the carp variety used here probably also has a methylation-free, but GC-poor, island in common with the trout metallothionein B gene.

DISCUSSION

The low GC-content of non-methylated islands in fish can account for the reduced or absent HTF fractions

We initially observed that DNA from some fish species, unlike other tested vertebrates, have a much reduced HTF fraction. Either these fish species did not have CpG islands, or they had CpG islands which were not detectable using the end-labelling approach. By studying two specific housekeeping genes we found that the 5′ regions of the homologous genes in fish resemble those in mammals and birds in that they are unmethylated and have CpG at the frequency predicted by base composition. In contrast to CpG islands in mammals and birds fish CpG islands are not GC-rich compared to the surrounding DNA. Recent data concerning the trout metallothionein A gene (17) makes it likely that this gene too has a GC-poor CpG island, although the methylation status of the region is not known. Based on the end-labelling experiments and sequence analysis, it is reasonable to expect that most promoters in those fish species with reduced HTF fractions are not GC-rich. It is also interesting to note that several *Xenopus laevis* genes, including heat-shock protein hsp70 and some histone genes, have CpG islands with a comparitively low GC content (18). It is therefore possible that vertebrates other than fish have GC-poor CpG islands. Thus, whereas lack of methylation appears to be a universal feature of functional housekeeping promoters in vertebrate genomes, GC-richness is not universal.

The low GC-content of CpG islands in certain fish can explain why the HTF fractions of these species are reduced. The frequency of HpaII sites is extremely sensitive to GC-content. For example, CpG islands of 1kb in length with a GC-content of 70% are expected to contain 15 HpaII sites, whereas islands of 40% GC-content will contain fewer than 2 HpaII sites. It follows that HpaII fragments generated from GC-poor CpG islands will be relatively large and will usually extend into heavily methylated flanking sequences. Consequently the HpaII Tiny Fragment fraction characteristic of other vertebrate genomes will not be seen.

Figure 5 shows a schematic generalisation of our findings. The GC composition and frequency of the dinucleotide CpG is shown across an imaginary house-keeping gene in three animal groups: mammalian (vertebrate with an HTF fraction), fish (vertebrate with no HTF fraction) and invertebrate. The two vertebrate genes have islands where the dinucleotide CpG occurs at the expected frequency, but the islands differ in GC-content. The mammalian island has a high GC-content when compared to the flanking DNA, whereas the fish island has the same GC-content as the flanking DNA. In the invertebrate gene and its flanking sequences the GC-content is constant and CpG occurs at the expected frequency. The islands of the vertebrate genes and the whole of the invertebrate gene are non-methylated. It is apparent that the fish gene represents an intermediate case between the mammalian and the invertebrate genes.

Among organisms whose genomes had prominent HTF fractions were several cold blooded vertebrates (for example crocodile and frog; see Figure 1, lanes 1 and 2). Therefore there is no simple correlation between low body temperature and absence or reduction of the HTF fraction. Neither does the absence of a prominent HTF fraction correlate with a particularly large or small genome size. The genomes of species with a reduced HTF fraction ranged between 1.39 pg and 6.86 pg, which is not substantially different from the genome sizes found in mammals and birds.

Evolution of GC-rich CpG islands and GC-rich isochores

The difference in GC-content between fish and other vertebrate promoters is paralleled by another distinctive feature of many fish genomes: namely absence of GC-rich isochores. An isochore is a long region of homogeneous base composition which can extend over many hundreds of kilobases (reviewed in 19). Genomes of mammals and birds are mosaics of isochores with different base compositions, whereas fish genomes tend to be more homogeneous, lacking in particular isochores that are GC-rich (20, 21). (Note that CpG islands are different from isochores in that the former are very small, and can be found within both GC-rich and AT-rich isochores.) Is it possible that the absence of GC-rich isochores and the absence of GC-rich CpG islands in fish have a common cause? Several explanations have been put forward to account for the origin of GC-rich isochores, but recent evidence has suggested that isochores arose as a passive consequence of the time of replication during S phase (22). It was speculated that changing bias in the balance of free deoxynucleotide triphosphate (dNTP) pools (23), has meant that mutations in sequences replicating early in S phase are more likely to result in incorporation of GC base pairs, whereas mutations in late replicating sequences preferentially incorporate AT base pairs. According to this hypothesis, sequences that have consistently replicated at a particular time-window of S phase will, in the course of evolution, acquire a base composition that reflects the dNTP pool composition at that time.

It is difficult to extend this model unchanged to CpG islands, as they are are almost certainly too short to behave as independent replicons. Instead of replication, however, we suggest that DNA repair within islands could be sensitive to dNTP pool imbalances in the same way. It is possible to imagine that CpG islands are subject to higher levels of damage and repair than bulk DNA, because, unlike bulk DNA, CpG islands adopt an open and accessible chromatin configuration (3). Indeed it has been found that a region including the CpG island of the DHFR gene is preferentially repaired in CHO cells subjected to UV irradiation (24). Our hypothesis is that regions of open chromatin, such as those flanking active promoters, are subject to increased levels of repair, and that errors in the repair process lead to the

substitution of bases that reflect the balance of dNTPs in the available pool. To explain the existence of GC-poor islands and isochores in fish, we must presume that the dNTP pools are evenly balanced at all times in fish. In other vertebrates, however, we expect that the pools in early S, and at other times in the cell-cycle when repair may occur, are biased in favour of G and C. Thus the same pool bias that may have lead to the appearance of GC-rich isochores in mammals and birds, could also have given rise to the GC-richness of CpG islands.

ACKNOWLEDGEMENTS

We are grateful to Hannes Tkadletz for photography, Elke Kleiner for technical assistance, and Gotthold Schnaffner and colleagues for sequencing and oligonucleotide synthesis. We would also like to thank the other members of our laboratory in the IMP for helpful comments and suggestions.

REFERENCES

1. Bird, A.P. (1987) Trends Genet., **3**, 342–347.
2. Coulondre, C., Miller, J.H., Farabough, P.J. and Gilbert, W. (1978) Nature, **274**, 775–778.
3. Tazi, J. and Bird, A. (1990) Cell, **60**, 909–920.
4. Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedman, T. and Migeon, B.R. (1984) Proc. Natl. Acad. Sci USA, **81**, 2806–2810.
5. Antequera, F., Boyes, J. and Bird, A. (1990) Cell, **62**, 503–514.
6. Cooper, D.N., Taggart, M.H. and Bird, A.P. (1983) Nucl. Acids Res., **11**, 647–658.
7. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988) Science, **239**, 487–491.
8. Zafarullah, M., Bonham, K. and Gedamu, L., (1988) Mol. Cell Biol., **8**, 4469–4476.
9. Liu, Z., Zhu, Z., Roberg, K., Faras, A.J., Guise, K.S., Kapuscinski, A.R. and Hackett, P.B. (1989) Nucl. Acids Res., **17**, 5850
10. Feinberg, A.P. and Vogelstein, B. (1984) Anal. Biochem., **137**, 266–267.
11. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucl. Acids Res., **12**, 387–395.
12. Karin, M. and Richards, R.I. (1982) Nature, **299**, 797–802.
13. Ponte, P., Ng, S.-Y., Engel, J., Gunning, P. and Kedes, L. (1984) Nucl. Acids Res., **12**, 1687–1696.
14. Ng, S.-Y., Gunning, P., Eddy, R., Ponte, P., Leavitt, J.,Shows, T. and Kedes, L. (1985) Mol. Cell Biol., **5**, 2720–2732.
15. Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. and Yaffe, D. (1983) Nucl. Acids Res., **11**, 1759–1771.
16. Leipoldt, M. and Schmidtke, J. (1982) In Dover, G.A. and Flavell, R.B. (eds.), Genome Evolution. Academic Press, London, pp. 219–236.
17. Murphy, M.F., Collier, J., Koutz, P. and Howard, B. (1990) Nucl. Acids Res., **18**, 4622.
18. Gardiner-Garden, M. and Frommer, M. (1987) J. Mol. Biol., **196**, 261–282.
19. Bernardi, G. (1989) Annu. Rev. Genet., **23**, (637–661).
20. Bernardi, G. and Bernardi, G. (1990) J. Mol. Evol., **31**, 265–281.
21. Bernardi, G. and Bernardi, G. (1990) J. Mol. Evol., **31**, 282–293.
22. Wolfe, K.H., Sharp, P.M. and Li, W-H (1989) Nature, **337**, 283–285.
23. Leeds, J.M., Slaubaugh, M.B. and Matthews, C.K. (1985) Mol. Cell Biol., **5**, 3443–3450.
24. Bohr, V.A., Okumoto, D.S., Ho, L. and Hanawalt, P.C. (1986) J. Biol. Chem., **261**, 16666–166672.
25. Maroni, G., Otto, E. and Lastowski-Perry, D. (1986) Genetics, **112**, 493–504.