The EMBO Journal vol.6 no.4 pp.999–1004, 1987

Non-methylated CpG-rich islands at the human α-globin locus: implications for evolution of the α-globin pseudogene

Adrian P.Bird, Mary H.Taggart, Robert D.Nicholls[1,2], Douglas R.Higgs[1]

MRC Mammalian Genome Unit, King's Buildings, West Mains Road, Edinburgh EH9 3JT, and [1]MRC Molecular Haematology Unit, John Radcliffe Hospital, Headington, Oxford, UK

[2]Present address: Department of Genetics, The Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA

Communicated by A.P.Bird

We have analysed CpG frequency and CpG methylation across part of the human α-globin locus. Clusters of CpG at the α1 and α2 genes resemble the '*Hpa*II tiny fragment (HTF) islands' that are characteristic of mammalian 'housekeeping' genes: CpG frequency is not suppressed; testable CpGs are not methylated in DNA from erythroid or non-erythroid tissues, although flanking CpGs are methylated; CpG clusters are ~1.5 kb long and extend both upstream and downstream of the α-globin transcription start site. These features are not found at genes of the β-globin locus. The α-globin pseudogene (ψα1) is highly homologous to the α2 and α1 genes, but it lacks an HTF island. Sequence comparison shows that a high proportion of CpGs in the α2 gene are substituted by TpG or CpA in the pseudogene. This strongly suggests that an ancestral HTF island at the pseudogene became methylated in the germline, and was lost due to the mutability of 5-methylcytosine.

Key words: globin genes/HTF islands/methylation/pseudogenes

Introduction

The vertebrate genome is A+T-rich and within it the doublet CpG is found to be both heavily methylated and rare. Typically between 60 and 90% of CpGs are methylated (Sinsheimer, 1955; Gruenbaum *et al.*, 1981) and CpG occurs at about one-fifth of the frequency predicted from base composition (Swartz *et al.*, 1962; Russell *et al.*, 1976). This average picture, however, conceals sequences that differ significantly from the mean. In particular the HTF (*Hpa*II tiny fragment) fraction (Cooper *et al.*, 1983; Bird *et al.*, 1985) is rich in G+C and contains non-methylated CpG at the expected frequency. The HTF fraction amounts to ~1% of the nuclear DNA in a range of vertebrates, and it is dispersed throughout the genome in islands of ~1000 bp. There are of the order of 30 000 islands in the haploid genome, and the average spacing between them is ~100 kb, although variation about this mean is large (Brown and Bird, 1986).

Speculation about the function of HTF islands centres on their frequent association with genes. Vertebrate housekeeping genes, for example, have nucleotide sequences at their 5′ ends with the characteristics of islands, and in a few cases it has been shown that these sequences are non-methylated in a variety of tissues (reviewed by Bird, 1986). Tissue-specific genes, on the other hand, are normally heavily methylated in non-expressing tissues, and are deficient in CpG. There are, however, several genes whose expression appears to be confined to a single tissue, but which are associated with CpG-rich sequences. In this study we have determined whether the discrete clusters of CpG near one family of these genes, the human α-globin family, are non-methylated in germ cells and in erythroid and non-erythroid somatic tissues. We have found that the CpG clusters are non-methylated in all tested cell types, while the inter-island regions are methylated. By contrast the α-globin pseudogene (ψα1) (Proudfoot and Maniatis, 1980), which lacks HTF-like clusters of CpG, is heavily methylated. Sequence comparisons strongly suggest that there was once an island at this gene, but that it became methylated and was subsequently lost by mutation of methyl-C to T.

Results

Members of the α-globin gene family are clustered at 16p13.1 on the short arm of chromosome 16 (Deisseroth *et al.*, 1977; Nicholls *et al.*, 1986). The locus comprises genes for embryonic and adult α-globins, their pseudogenes (Lauer *et al.*, 1980) and an α-like gene of unknown function (Marks *et al.*, 1986; see Figure 5B). Since much of the region has been sequenced, we analysed C+G content and also the distribution of the dinucleotide sequences CpG and GpC across the ψα1, α2 and α1 genes and their flanking sequences (Figure 1). G+C content is on average 60%, which is well above the average for the human genome as a whole (40%). This non-average base composition is presumably the reason for the relatively high buoyant density of α-globin genes in caesium chloride gradients (Bernardi *et al.*, 1985). The distribution of GpC sequences (Figure 1, broken line)

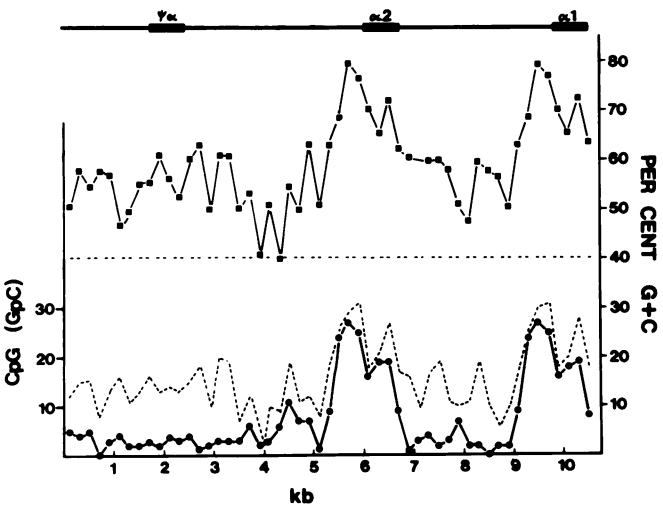
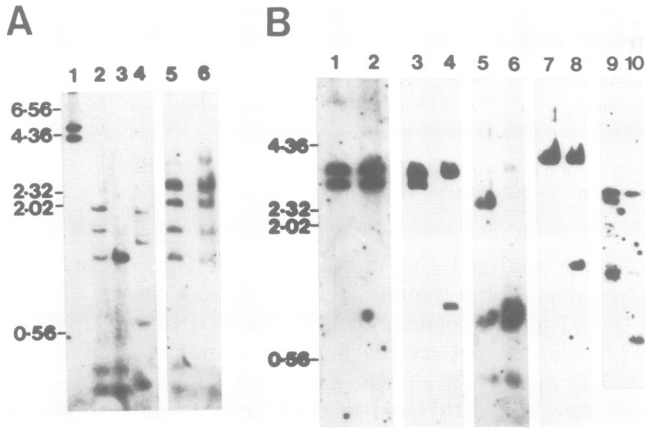


Fig. 1. Base composition and dinucleotide frequencies across part of the α-globin locus. The ordinate scale to the left refers to the number of CpGs (solid circles) or GpCs (broken line) in 200-bp steps across the region. The right hand ordinate denotes the percentage of G+C in 200-bp steps. The horizontal broken line shows the average G+C content of human chromosomal DNA.

roughly parallels G+C content, increasing in density by almost 2-fold at the α2 and α1 genes and their 5′ flanking regions. CpG density, however, increases ~10-fold at the α2 and α1 genes. In the CpG clusters the densities of GpC and CpG are approximately equal, while outside the clusters the CpG density falls to about one-fifth that of GpC, in keeping with the CpG deficiency of the bulk genome.

Gene sequences that show tissue-specific methylation patterns are generally fully methylated in germline. In order to test the state of methylation around the α-globin genes in sperm DNA we used two enzyme pairs that distinguish methylated CpG (*Hpa*II/*Msp*I and *Sma*I/*Xma*I) and a variety of enzymes that are sensitive to methylation of CpG (*Nar*I, *Ava*I, *Nae*I). The probe for the α2 and α1 gene was a 1.5-kb *Pst*I fragment (Figure 2C). This probe hybridizes to both genes due to their extensive sequence identity. We were generally able to distinguish the two regions by digestion with *Hpa*I, which cleaves downstream of the α2 gene, but not the α1 gene (Figure 2C). Figure 2A and B show examples of the Southern blots that were used to compile the methylation map for sperm DNA. *Hpa*II digestion of sperm DNA gave a group of small fragments of ~0.32 and 0.23 kb which were present at the same intensity after *Msp*I digestion (Figure 2A, lanes 2 and 3). These fragments originate from the coding region of the α2 and α1 genes (Figure 2C). Fragments <0.23 kb were not detectable in this experiment. A series of larger fragments was also generated by *Hpa*II. These must originate downstream of the α1 gene as they were not altered by *Hpa*I digestion (lanes 5 and 6). Also *Eco*RI removed bands in the series >2.1 kb, giving a new 2.0 kb band (see Figure 2C). Since *Msp*I (a methyl-insensitive isoschizomer of *Hpa*II) only gave the lower 1.0 kb band in the series, we concluded that the larger *Hpa*II fragments are partially methylated as shown in



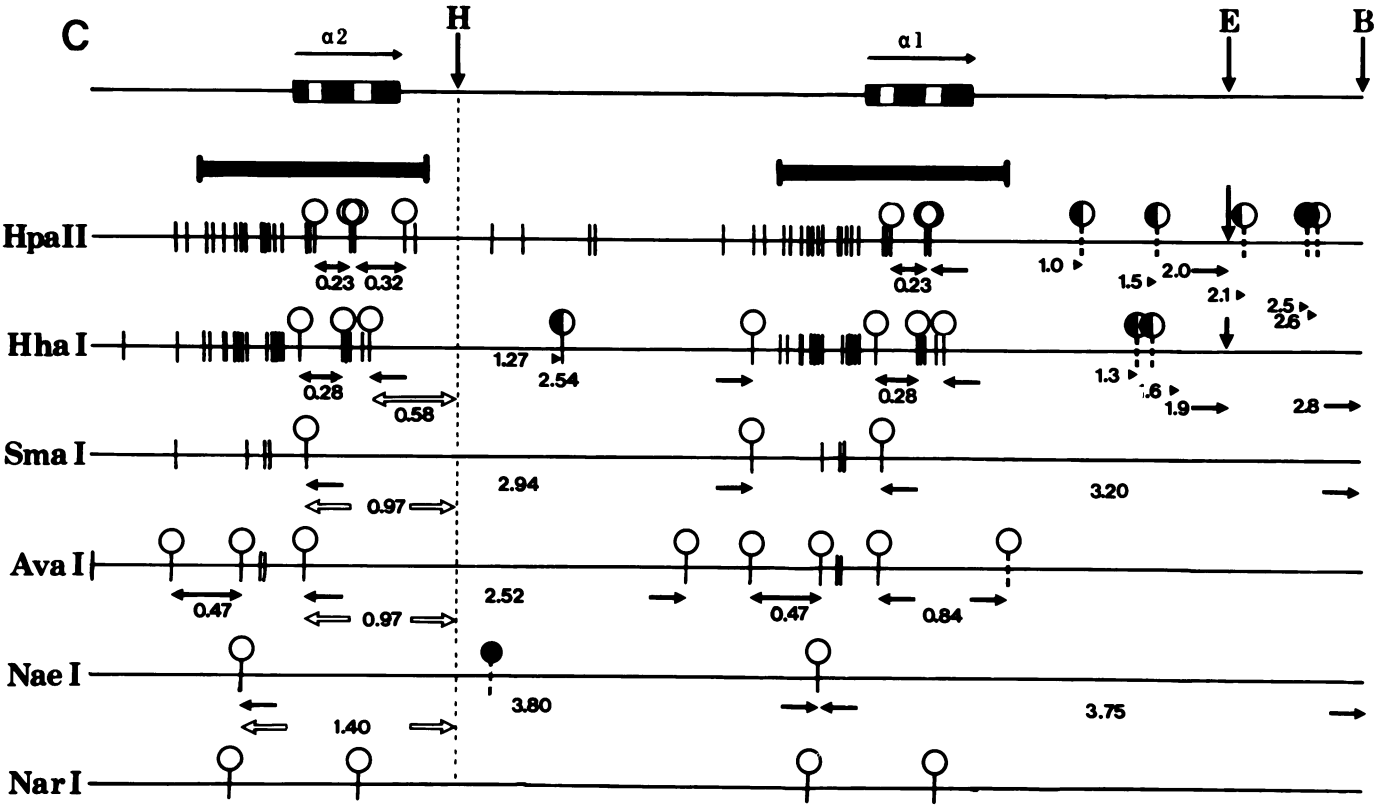


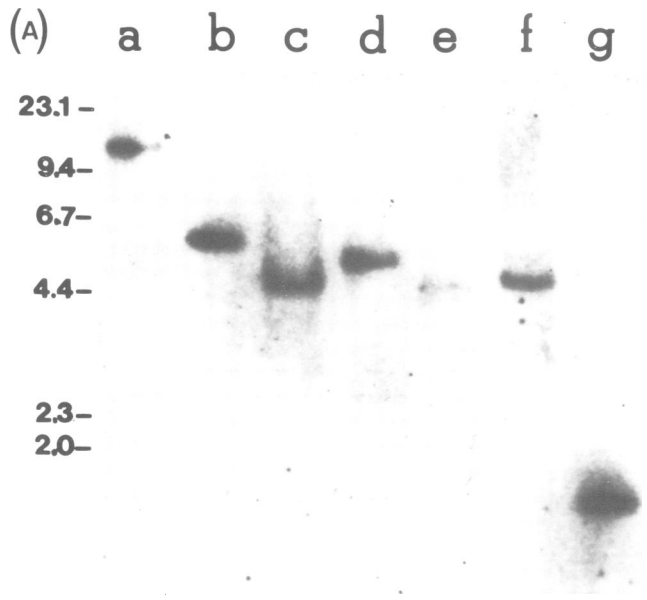
Fig. 2. Methylation of CpG enzyme sites at the α-globin genes in sperm DNA. **(A)** Southern blots of human sperm DNA probed with the α1 *Pst*I fragment (see diagram) after double digestion with *Hpa*I + *Eco*RI **(lane 1)**; or triple digestion with *Hpa*I + *Eco*RI and either *Hpa*II **(lane 2)**, *Msp*I **(lane 3), or** *Hha*I **(lane 4). Lanes 5** and **6** show digests with *Hpa*II and *Hpa*II plus *Hpa*I respectively. **(B)** Southern blots as above but double digested with *Bam*HI plus *Sma*I **(lanes 1** and **3)**, *Xma*I **(lane 2)**, *Ava*I **(lane 5)**, *Nae*I **(lane 7)** or *Hha*I **(lane 9). Lanes 4, 6, 8** and **10** were triple digested with *Bam*HI plus *Hpa*I plus either *Sma*I **(lane 4)**, *Ava*I **(lane 6)**, *Nae*I **(lane 8)** or *Hha*I **(lane 10). (C)** Maps derived from the data shown in (A) and (B) and other data. Thick bars show the 1.5-kb *Pst*I−*Pst*I probe which is homologous to both α2 and α1 genes. Fragments that are visible on the appropriate blots are indicated by double arrows and their size in kilobases is shown. Vertical bars show restriction sites from sequence data. Broken vertical bars to the right of the map are sites outside the sequenced region that were deduced from this work. The broken bar at the *Nae*I site located between the α2 and α1 genes signifies that no cleavage was seen at this site in chromosomal DNA. The site is either fully methylated or absent (see text). Open circles denote non-methylated sites, solid circles, heavily methylated sites, and half-solid circles, partially methylated sites. H (top map) shows the *Hpa*I site used to distinguish α2 from α1 genes on the blots. The new fragments generated by *Hpa*I digestion are denoted by open arrows. E − *Eco*RI; B − *Bam*HI. *Nar*I blots are not shown.

Figure 2C. *Hha*I digests gave a 0.28-kb fragment from within the α2 and α1 genes, and a pair of partially methylated sites downstream of the α1 gene (Figure 2A, lane 4). In addition, a 2.5-kb fragment and a 1.3-kb fragment are seen after *Hha*I digestion, and both are reduced to ~0.6 kb after additional digestion with *Hpa*I (Figure 2B, lanes 9 and 10 and data not shown). The 2.5-kb fragment therefore spans a partially methylated *Hha*I site between the α2 and α1 genes. Digests with *Sma*I, *Ava*I, *Nae*I (Figure 2B and C) and *Nar*I (data not shown) gave fragments that were identified by additional digestion with *Hpa*I (Figure 2B and C) or other enzymes (data not shown). The presence of a 3.8-kb band in *Nae*I digests which is cut by *Hpa*I (Figure 2B, lane 7) indicates that the *Nae*I site between α2 and α1 genes is fully methylated. Formally, however, we cannot exclude the possibility that this site, though present in cloned sequences covering this region, is absent from the genomic DNA that was tested.

Methylation of CpGs upstream of the α2 gene was tested using a *Pst*I-*Sst*I probe derived from genomic DNA. The probe includes the ψα1 sequence. Sperm DNA was digested with *Bam*HI (Figure 3A, lane a) and then cleaved with *Sma*I, *Xma*I, *Ava*I, *Hha*I, *Hpa*II and *Msp*I (Figure 3A, lanes b–g). Methylated sites were clearly demonstrated by comparing *Xma*I (a methyl-insensitive isoschizomer of *Sma*I) with *Sma*I digests, and *Msp*I with *Hpa*II digests. The blots are interpreted in Figure 3B. Non-methylated sites on the left of the map were identified in *Hpa*II and *Hha*I digests of *Hpa*I plus *Bam*HI-cleaved DNA (not shown). A summary map of methylation in the region for sperm DNA is shown in Figure 4.

We tested for methylation at the α2 and α1 genes in a wide range of cell types by comparing *Hpa*II and *Msp*I fragment patterns (Figure 5A). Genomic DNA was digested with *Pst*I to give identical 1.5-kb fragments from each α gene, and then further digested with either *Hpa*II or *Msp*I. No difference between *Hpa*II and *Msp*I patterns could be detected in any of the tested tissues, indicating lack of methylation at *Hpa*II sites of both genes. A broader picture of DNA methylation was obtained by comparing *Hpa*II and *Msp*I fragment patterns for probes dispersed across the entire α-globin gene locus (summarized in Figure 5B, data not shown). As expected, identical *Hpa*II and *Msp*I patterns were found at the α1 and α2 regions in many tissues. Sequences further upstream, however, consistently showed partial cleavage or no cleavage by *Hpa*II, indicating partial or complete methylation of sites in these regions.

A striking feature of the CpG distribution in Figure 1 is the lack of a CpG-rich region at the pseudo-α sequence. Since this pseudogene is 75% homologous to α2 and α1 in nucleotide sequence, we had expected that an identifiable vestige of the CpG cluster would remain. Our finding that the pseudo-alpha region is methylated at testable CpGs offers an explanation for the absence of clustered CpGs: the pseudogene may have become



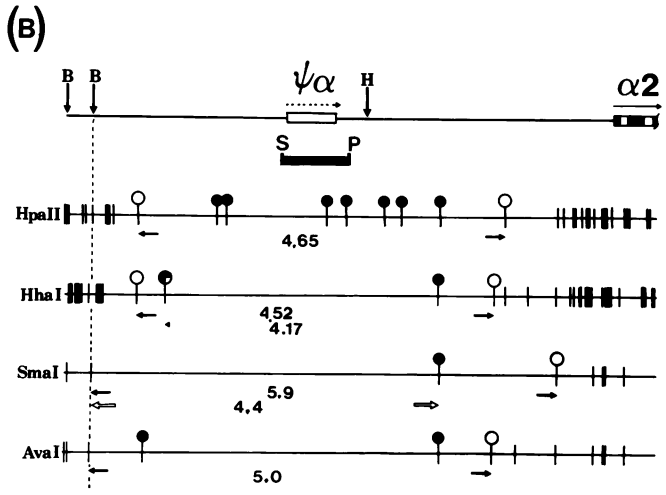


Fig. 3. Methylation of CpG-enzyme sites at the ψα1 gene in sperm DNA. **(A)** Southern blots of human sperm DNA probed with a genomic *Sst*I–*Pst*I fragment including ψα1 (see diagram). Digestion was with *Bam*HI alone (**lane a**) or with *Bam*HI plus *Sma*I, (**b**); *Xma*I, (**c**); *Ava*I, (**d**); *Hha*I, (**e**); *Hpa*II, (**f**); and *Msp*I, (**g**). The low intensity band in **lane e** is due to underloading of this lane. **(B)** Maps derived from data shown in (A) and other data. The solid bar indicates the *Sst*I–*Pst*I probe. Fragment sizes correspond to bands visible on the autoradiograph. S – *Sst*I; P – *Pst*I; B – *Bam*HI; H – *Hpa*I. Non-methylated sites on the left of the maps were defined by digestion with CpG enzymes of *Bam*HI plus *Hpa*I-digested DNA (not shown).

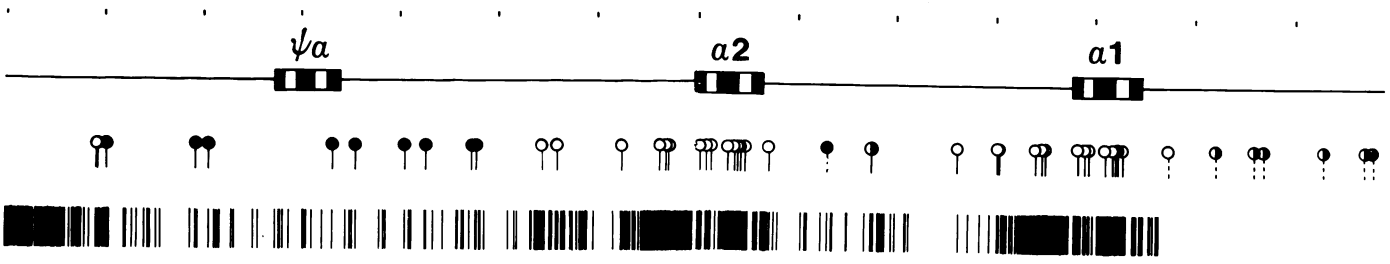


Fig. 4. Methylated and non-methylated sites at the α-globin locus in sperm DNA. Vertical bars represent CpGs. Methylated, non-methylated and partially methylated sites are represented by solid, half-solid and open circles respectively. The scale is in kilobase pairs.

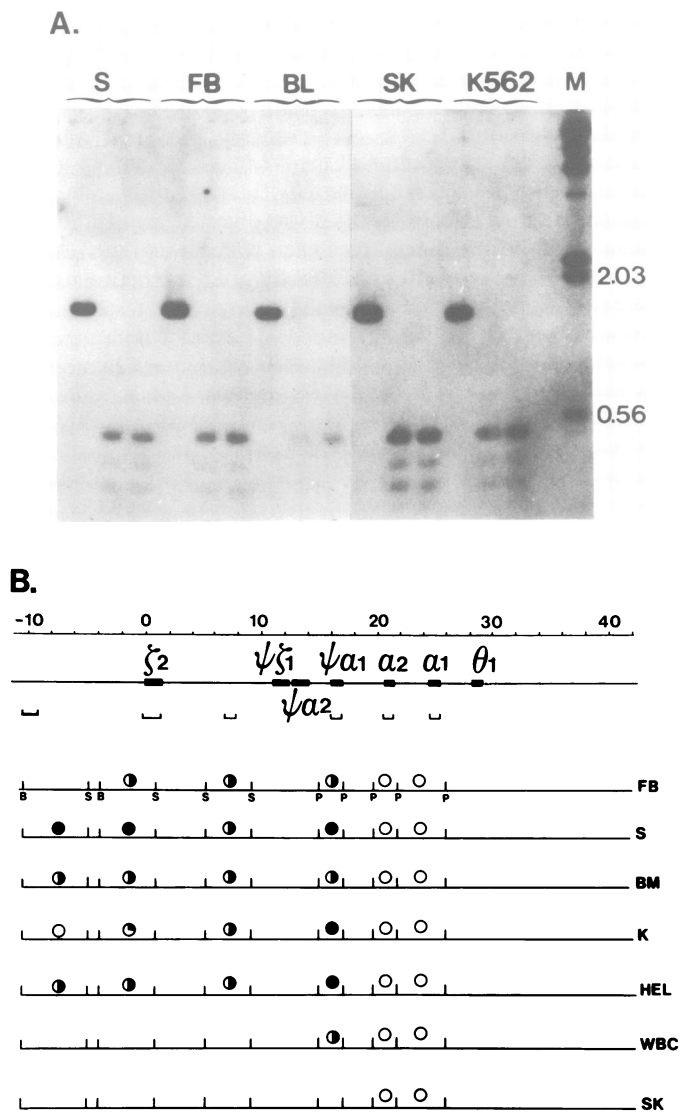


Fig. 5. Lack of methylation at the α-globin genes in a wide variety of tissues. (**A**) Comparison of *Hpa*II and *Msp*I digestion patterns of the identical *Pst*I fragments of α2- and α1-globin genes. For each tissue the order of digests (left to right) is *Pst*I, *Pst*I plus *Hpa*II, *Pst*I + *Msp*I. The marker (M) is bacteriophage λ DNA digested with *Hin*dIII. (**B**) A summary of the results of *Hpa*II and *Msp*I comparisons at various parts of the α-globin locus. Brackets below the map of the locus denote probes. Circles on the map for each tissue show levels of methylation ranging from little or none as judged by identity of *Hpa*II and *Msp*I patterns (open circles) to heavy as judged by failure of *Hpa*II to cleave at potential sites (solid circles). Partially solid circles indicate partial methylation. Each circle only refers to methylation of *Hpa*II sites between the restriction sites that flank it on the map. The tissues are FB, foetal brain; S, semen; BM, bone marrow; K, K562 human erythroleukaemic cells; HEL, human erythroleukaemic cells; WBC, white blood cells; SK, skin; BL, blood, P = *Pvu*II, S = *Sac*I and B = *Bam*HI.

methylated in an ancestral human germ line leading to mutation of CpGs due to the instability of 5-methylcytosine (Coulondre *et al.*, 1978). This scenario makes the prediction that CpGs in the presumed 'parent' gene (α2 or α1) will have been preferentially converted into TpG and CpA, as these are the direct products of mCpG deamination (Bird, 1980). Figure 6A shows the α2 and pseudo-α sequences compared. Out of 70 CpGs in the α2 sequence, only four occur at equivalent positions in the pseudogene (Figure 6B). Of those which change, 46 show single base changes and 32 of these are to TpG (20) and CpA (12).

(A)

```
6896  .....TGCCCCCGCGCCCCAAGCATAAACCCTGGCGCGCTCGCGGCCCGG   6945
2640  GCCAA   A   T  AT    G       GAG  CCTA T   C CCA  TT  2689

6946  CACTCTTCTGGTCCCCACAGACTCAGAGA.GAACCCACCATGGTGCTGTC   6995
2690  TCACT  TGA CTTA            A TT  G TG  G             2735

6996  TCCTGCCGACAAGACCAACGTCAAGGCCGCCTGGGGTAAGGTCGGCGCGC   7045
2740       AG      G T   AC      G T     AG  A T    AC    2789

7046  ACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAGGTGAG...GCTCCCTC   7095
2790    A   CT GC    CCA                   CA  AACC   TCT  2839

7096  CCCTGCTCCGACCCGGGCTCC....TCGCCCGCCCGGACCCACAGGCCAC   7145
2840          AC   TT  G   AACGC CA TC AG  CT    T         2889

7146  CCTCAACCGTCCTGGCCCCGGACCCAAACCCCA...CCCCTCACTCTGCT   7195
2890    CT   TA T  TA   T       GC     GCC  T A T  T       2939

7196  TCTCCCCGCAGGATGTTCCTGTCCTTCCCCACCACCAAGACCTACTTCCC   7245
2940    C   T A  C          A     ....................T    2989

7246  GCACTTCGACCTGAGCCACGGCTCTGCCCAGGTTAAGGGCCACGGCAAGA   7295
2990  T     G C       TT T  G  A      A C   AAA  AT     T  3039

7296  AGGTGGCCGACGCCCTGACCAACGCCGTGGCGCACGTGGACGACATGCCC   7345
3040            A G         T       TC   T A  T           3089

7346  AACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCACAAGCTTCGGGTGGA   7395
3090    T ATG    T AGG    GA G    T TC   G    GT          3139

7396  CCCGGTCAACTTCAAGGTGAGCGGCGGGCCGGGAGCGATCTGGGTCGAGG   7445
3140     A G    A   G  A    TTT    T     GA    A  GT T    3189

7446  GGCGAGATGGCGCCTTCCTCTCAGGGCAGAGGATCACGCGGGTTGCGGGA   7495
3190    GC  C  ..          AT   A    CCC  C A  C   CA      3239

7496  GGTGTAGCGCAGGCGGCGGCTGCGGGCCTGGGCCGCACTGACCCTCTTCT   7545
3240       G   A  ...  T   A TA TATCT CATC C     T  ...    3289

7546  CTGCACAGCTCCTAAGCCACTGCCTGCTGGTGACCCTGGCCGCCCACCTC   7595
3290    C     T    GG TA A                  TA    TG   . A  3339

7596  CCCGCCGAGTTCACCCCTGCGGTGCACGCCTCCCTGGACAAGTTCCTGGC   7645
3340    TT  CG   T    A  T   ....       A        A  AT T   3389

7646  TTCTGTGAGCACCGTGCTGACCTCCAAATACCGTTAAGCTGGAGCCTCGG   7695
3390    T       TG T   T       AA   C   A          A    A T  3439

7696  TAGCCGTTCCTCCTGCCCGATGGGCCTCCCAACGGGCCCTCCTCCCCTCC   7745
3440   G T A G  C       T C             C A      G          3489

7746  TTGCACCGGCCCTTCCTGGTCTTTGAATAAAGTCTGAGTGGGCGGCA...   7795
3490  C   C  A  A         A          G      C    A   A   ATG  3539
```

(B)

	G	A	T	C
C	④	12	6	4
T	20	2	3	4
A	0	0	2	3
G	4	2	0	1

Fig. 6. Loss of CpGs in the ψα1 gene compared to the α2 gene. (**A**) Comparison of pseudogene and gene sequences (after Proudfoot and Maniatis, 1980). The upper sequence is the α gene (5′ to 3′) in which the ATA box, ATG and polyadenylation signals have been underlined. The lower sequence indicates base changes, deletions or insertions in the α1 gene. Identical bases are not shown. Brackets above the gene sequence denote CpGs. A solid triangle above the bracket indicates that this CpG is represented by a TpG or CpA in ψα1. (**B**) Matrix showing the frequency of all possible dinucleotides at positions in the ψα1 gene corresponding to CpGs in the gene sequence. The CpG figure (circled) is the number of conserved CpGs. Three CpGs in the gene that correspond to deletions in the pseudogene have been omitted.

(Both nucleotides have changed in 17 cases and three CpGs are altered due to small deletions.) The great preponderance of transitions from CpG/CpG in the gene to TpG/CpA in the pseudogene

strongly supports the view that methylation has led to loss of the HTF islands which characterize the functional α-globin genes.

Discussion

Typical HTF islands at the α globin genes

The HTF islands at the α2 and α1 globin genes are typical of islands found near other genes: they are ~1.5 kb in length and contain DNA that is non-methylated in all tested tissues, is G+C rich, and lacks CpG suppression. HTF islands usually extend both upstream and downstream of the site of transcriptional initiation, including one or more exons, and in this respect also the α-globin islands are typical. What is atypical is that the expression of α-globin genes as globin protein is highly tissue specific compared to most other island-associated genes, being confined to erythroid cells. Other potential HTF-associated genes that show a high degree of tissue specificity are the genes of the type II major histocompatibility complex (Tykocinski and Max, 1984), and the gene for retinol binding protein (d'Onofrio *et al.*, 1985) though DNA methylation has not been investigated in these cases. Non-methylated CpG-rich islands have been reported previously at genes that are active in several, but not all, tissues of the organism: the chicken α2(I) collagen gene (McKeon *et al.*, 1982) and the murine Thy-1 gene (Kolsto *et al.*, 1986).

It is worth noting that other sequences in the α-globin locus contain clusters of CpGs that are potential HTF islands. Although we have not systematically studied these putative islands, our data indicate that a CpG cluster upstream of the pseudo-α gene contains CpG-enzyme sites that are non-methylated in sperm DNA (see left of Figure 4). Other clusters located close to the pseudo-zeta gene (see Figure 5B) and at the 3′ ends of both ζ and ψζ genes have potential HTF islands based on DNA sequence analysis (Nicholls, 1986).

As shown previously by Bernardi *et al.* (1985) the α-globin genes are part of a long G+C-rich DNA sequence. The HTF islands, however, are only a small part of this tract and are easily distinguished by both lack of methylation and CpG frequency. In general, HTF islands can be found in either A+T-rich or G+C-rich regions of the genome. We suggest, therefore, that these two features of the mammalian genome, HTF islands and long-range fluctuations in G+C content, are unrelated phenomena.

Differences between α- and β-globin genes

The present study highlights differences between α and β globin genes in mammals. Unlike the α genes, the human β genes are A+T-rich, deficient in CpG, and heavily methylated in sperm and several other non-expressing tissues (van der Ploeg and Flavell, 1980). The two loci also differ in expression. Developmental activity of α-globin occurs very early (5−6 weeks) whereas the β-globin gene is only fully activated in the final stages of development (Peschle *et al.*, 1985). Transfection of the α and β genes into cultured cells also gives divergent results. The β-like genes require a transcriptional enhancer in *cis* for efficient transcription, whereas the transfected α genes are expressed at a high rate in several cell types without the presence of an enhancer (Mellon *et al.*, 1981). It is possible that the HTF islands which α-globin genes have in common with housekeeping genes are relevant to the differences in expression.

Loss of an HTF island at the ψα1 gene

Since HTF-islands are thought to be CpG-rich by virtue of their lack of methylation, it follows that if methylation does occur in the germ line it would lead to the relatively rapid loss of CpGs by mutation (Bird, 1986). Our analysis of the α-globin pseudogene supports this proposal, as only 7% of CpGs in the functional gene are present in the pseudogene, despite an overall nucleotide homology between the sequences of 75%. The preponderance of CpA and TpG at changed sites in the pseudogene implicates failure to correct deamination of 5-methylcytosine as the primary source of CpG loss (see Coulondre *et al.*, 1978; Bird, 1980). This mechanism may be one important aspect of the function of DNA methylation in animals. If CpG-rich sequences somehow signal a class of promoter regions, then methylation may ensure that other regions of the genome remain CpG-poor.

The loss of an HTF island at the pseudogene supports the view that CpG is depleted in any DNA sequence that becomes heavily methylated in the germline. Adams and Eason (984) have observed that G+C-rich sequences in vertebrate DNA often show no CpG suppression, and they propose that this is because deamination of 5-methylcytosine occurs at very low rates in DNA with this base composition. Thus they suggest that the G+C-rich sequences that they have analysed are methylated at CpG but tend not to lose CpG by mutation. The α-pseudogene data are not compatible with this theory. The α-globin islands and the pseudogene are both G+C-rich (see Figure 1), yet the evidence strongly suggests that methylation of the island has led to loss of CpG. It seems possible that many of the G+C-rich sequences analysed by these authors are in fact part of non-methylated HTF islands.

The association between HTF islands and globin genes is variable. Apart from the difference between human α-like and β-like genes in this respect (see above), the α-like globin genes themselves are not associated with HTF-like sequences in all species. In the chicken, for example, both α and β gene families are known to be heavily methylated in sperm and in non-expressing somatic tissues, while developmental loss of methylation occurs at α- and β-like genes in erythroid cells (McGhee *et al.*, 1981; Weintraub *et al.*, 1981; reviewed by Shen, 1984). At present the evolutionary origin of HTF islands is uncertain, though it is possible that they represent regions of the genome that have consistently been protected from methylation by bound factors (Bird, 1986). According to this view, all globin-like genes possessed HTF islands in an ancestral vertebrate, and the present day patterns are the result of evolutionary loss of islands at the β-like genes in man, or α- and β-like genes in chicken. An alternative possibility is that HTF islands can arise in non-island DNA during evolution. Thus the human α-like genes may have acquired islands, while the chicken globin genes failed to do so. It will be of interest to study the globin genes further in order to distinguish these extreme possibilities, and to determine the relevance of islands to gene expression.

Materials and methods

DNA was isolated using standard procedures from bone marrow, semen, whole blood and white blood cells of adult, from brain and skin of foetus, and from K562 and HEL human erythroleukaemic cell lines.

Restriction endonucleases were from Boehringer, Biolabs and Anglian Biotechnology and were used according to the supplier's instructions. We used a 20- to 40-fold excess of all CpG enzymes. Gels were blotted (Southern, 1975) onto Sleicher and Schuell nitrocellulose membranes, or Amersham Hy-Bond-N-membranes, and hybridized in 3 × SSC at 65°C in the presence of competitor DNA, dextran sulphate and Denhardt's solution.

Probes for α1 and α2 genes and pseudo-α genes were obtained from plasmids that have been described previously (Higgs *et al.*, 1986). Labelling was by nick translation or more frequently by the random primer method using fragments recovered from agarose gels (Feinberg and Vogelstein, 1984).

Acknowledgements

We are grateful to Donald Macleod for assistance with computer analysis and to Anne Deane for typing the manuscript. We also thank Susan Lindsay for a critical reading of the manuscript. The work was supported by the Medical Research Council.

References

Adams,R.L.P. and Eason,R.E. (1984) *Nucleic Acids Res.*, **12**, 5869–5877.
Bernardi,G., Olofsson,B., Filipski,J., Zerial,M., Salinas,J., Curry,G., Meunier-Rotival,M. and Rodier,F. (1985) *Science*, **228**, 953–958.
Bird,A.P. (1986) *Nature*, **321**, 209–213.
Bird,A.P. (1980) *Nucleic Acids Res.*, **8**, 1499–1504.
Bird,A., Taggart,M., Frommer,M., Miller,O.J. and Macleod,D. (1985) *Cell*, **40**, 91–99.
Brown,W.R.A. and Bird,A.P. (1986) *Nature*, **322**, 477–481.
Cooper,D.N., Taggart,M.H. and Bird,A.P. (1983) *Nucleic Acids Res.*, **11**, 647–658.
Coulondre,C., Miller,J.H., Farabaugh,P.J. and Gilbert,W. (1978) *Nature*, **274**, 775–780.
Deisseroth,A., Nienhuis,A., Turner,P., Velez,R., Anderson,W., Ruddle,F., Lawrence,J., Creagan,R. and Kucherlapati,R. (1977) *Cell*, **12**, 205–218.
d'Onofrio,C., Colantuoni,V. and Cortese,R. (1985) *EMBO J.*, **4**, 1981–1989.
Feinberg,A. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266–267.
Higgs,D.R., Wainscoat,J.M., Flint,J.V.S., Hill,A., Thein,S., Nicholls,R.D., Teal,H., Ayyub,H., Peto,T.E., Falusi,A.C., Jarman,A.P., Clegg,J.B. and Weatherall,D.J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5165–5169.
Kolsto,A.-B., Kollias,G., Giguere,V., Isobe,K.-I., Prydz,H. and Grosveld,F. (1986) *Nucleic Acids Res.*, **14**, 9667–9678.
Lauer,J., Shen,C.-K. and Maniatis,T. (1980) *Cell*, **20**, 119–130.
Marks,J., Shaw,J. and Shen,C.-K. (1986) *Nature*, **321**, 789–793.
McGhee,J.D., Wood,W.I., Dolan,M., Engel,J.D. and Felsenfeld,G. (1981) *Cell*, **27**, 45–55.
McKeon,C., Ohkubo,H., Paston,I. and de Crombrugghe,B. (1982) *Cell*, **29**, 203–210.
Mellon,P., Parker,V., Gluzman,Y. and Maniatis,T. (1981) *Cell*, **27**, 279–288.
Nicholls,R.D. (1986) D.Phil. Thesis, University of Oxford.
Nicholls,R.D., Jonasson,J., McGee,O., Patil,S., Ionasescu,V., Weatherall,D.J. and Higgs,D.R. (1986) *J. Med. Genet.*, in press.
Peschle,C., Mavilio,F., Care,A., Miliaccio,G., Miliaccio,A., Salvo,G., Sarnoggia,P., Petti,S., Guerriero,R., Marinucci,M., Lazzaro,D., Russo,G. and Mastroberardino,G. (1985) *Nature*, **313**, 235–238.
Proudfoot,N.J. and Maniatis,T. (1980) *Cell*, **21**, 537–544.
Russell,G.J., Walker,P.M.B., Elton,R.A. and Subak-Sharpe,J.H. (1976) *J. Mol. Biol.*, **108**, 1–23.
Shen,C.-K. (1984) In Razin,A., Cedar,H. and Riggs,A.D. (eds), *DNA Methylation: Biochemistry and Biological Significance*. Springer Verlag, pp. 249–268.
Sinsheimer,R.L. (1955) *J. Biol. Chem.*, **215**, 579–583.
Swartz,M.N., Trautner,T.A. and Kornberg,A. (1962) *J. Biol. Chem.*, **237**, 1961–1967.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Tykocinski,M.L. and Max,E.C. (1984) *Nucleic Acids Res.*, **12**, 4385–4396.
van der Ploeg,L.H.T. and Flavell,R.A. (1980) *Cell*, **7**, 771–779.
Weintraub,H., Larsen,A. and Groudine,M. (1981) *Cell*, **24**, 333–344.

Received on December 16, 1986; revised on February 6, 1987