Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 3349–3353, May 1987
Genetics

Comparative anatomy of the human *APRT* gene and enzyme: Nucleotide sequence divergence and conservation of a nonrandom CpG dinucleotide arrangement

(housekeeping gene/DNA sequence/evolution)

THOMAS P. BRODERICK*, DENNIS A. SCHAFF*, AMY M. BERTINO*, MICHAEL K. DUSH*, JAY A. TISCHFIELD†‡, AND PETER J. STAMBROOK*

*Department of Anatomy and Cell Biology, University of Cincinnati College of Medicine, Cincinnati, OH 45267; and †Department of Anatomy, Medical College of Georgia, Augusta, GA 30912

Communicated by Frank H. Ruddle, January 21, 1987 (received for review November 6, 1986)

ABSTRACT **The functional human adenine phosphoribosyltransferase (*APRT*) gene is <2.6 kilobases in length and contains five exons. The amino acid sequences of APRTs have been highly conserved throughout evolution. The human enzyme is 82%, 90%, and 40% identical to the mouse, hamster, and *Escherichia coli* enzymes, respectively. The promoter region of the human *APRT* gene, like that of several other "housekeeping" genes, lacks "TATA" and "CCAAT" boxes but contains five GC boxes that are potential binding sites for the Sp1 transcription factor. The distal three, however, are dispensable for gene expression. Comparison between human and mouse *APRT* gene nucleotide sequences reveals a high degree of homology within protein coding regions but an absence of significant homology in 5′ flanking, 3′ untranslated, and intron sequences, except for similarly positioned GC boxes in the promoter region and a 26-base-pair region in intron 3. This 26-base-pair sequence is 92% identical with a similarly positioned sequence in the mouse gene and is also found in intron 3 of the hamster gene, suggesting that its retention may be a consequence of stringent selection. The positions of all introns have been precisely retained in the human and both rodent genes, as has an unusual AG/GC donor splice site in intron 2. Particularly striking is the distribution of CpG dinucleotides within human and rodent *APRT* genes. Although the nucleotide sequences of intron 1 and the 5′ flanking regions of human and mouse *APRT* genes have no substantial homology, they have a frequency of CpG dinucleotides that is much higher than expected and nonrandom considering the G+C content of the gene. Retention of an elevated CpG dinucleotide content, despite loss of sequence homology, suggests that there may be selection for CpG dinucleotides in these regions and that their maintenance may be important for *APRT* gene function.**

Adenine phosphoribosyltransferase (APRT, EC 2.4.2.7) catalyzes the formation of AMP and inorganic pyrophosphate from adenine and 5-phosphoribosyl-1-pyrophosphate (PRPP). Its importance in metabolism probably relates to the production of adenine as a by-product of the ubiquitously distributed polyamine biosynthesis pathway (1). Deficiency of APRT activity is inherited as an autosomal recessive condition characterized by high urinary levels of adenine and 2,8-dihydroxyadenine (DHA), which may lead to a clinically significant DHA urolithiasis appearing during childhood or at a later age (2). The gene for human APRT has been mapped to chromosome 16 (3) at 16q24 (4). We have described the cloning of this gene and a relatively frequent *Taq* I restriction fragment length polymorphism within its largest intron (5).

The *APRT* gene is constitutively expressed in all adult tissues with only moderate variation between different cell types (6). APRT activity increases about 2-fold during S phase of the cell cycle (7), probably reflecting a doubling of the number of gene copies. As is the case for some other "housekeeping" genes (8), the transcription promoter of the mouse *APRT* gene from liver or cultured fibroblasts lacks TATA or CCAAT-like sequences but contains three CCGCCC repeats (GC boxes) that form the core of a decanucleotide sequence that can interact with Sp1 transcription factor (9–12).

We have determined the nucleotide sequence of the human *APRT* gene for several reasons. First, by distinguishing features that are conserved between our previously sequenced mouse gene (13, 14) and the Chinese hamster gene (15), we may be able to identify anatomic features that are important to their constitutive expression. Second, the sequence of a wild-type human *APRT* gene is necessary for our ongoing studies of mutant genes from human populations and cell cultures. Third, because of the relatively small size of *APRT* genes and the availability of media for selecting cells in culture that either express or fail to express APRT activity (3), *APRT* genes have been extensively utilized for studying mutagenesis in mammalian cells (16–19).

METHODS

Bacteria, Bacteriophage, and Plasmids. The previously described λ clone λHuap15 (5), which contains a 17.5-kilobase (kb) genomic insert including a functional human *APRT* gene, was propagated in *Escherichia coli* LE 392. An 8-kb *Hin*cII–*Eco*RI fragment and a 2.8-kb *Cla* I–*Bgl* II fragment (5) were subcloned into the polylinker of the pIBI 20 vector (International Biotechnologies, New Haven, CT), as were a *Sma* I fragment extending from position 1 to 752 (Fig. 1) and a 2.2-kb *Bam*HI fragment (5), extending from position 602 to 2763 (Fig. 1). All of the above subclones, which were used for nucleotide sequence determination, were maintained in *E. coli* JM109. Single-stranded DNA was obtained by superinfecting transformed cells with helper phage as described by Dente *et al.* (20) but by using M13 K07 (International Biotechnologies) as the helper phage.

DNA Libraries and Screening. Three human cDNA libraries were screened for APRT cDNAs using the purified internal 2.2-kb *Bam*HI fragment as a probe. Fibroblast (21), T-lymphoblast (22), and hepatoma (HepG2) (23) cDNA libraries were kindly provided by H. Okayama (National Institutes of Health), D. A. Wiginton and J. J. Hutton (Children's Hospital, Cincinnati), and R. Moore (Monsanto),

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: APRT, adenine phosphoribosyltransferase; PRPP, 5-phosphoribosyl-1-pyrophosphate.

‡To whom reprint requests should be addressed.

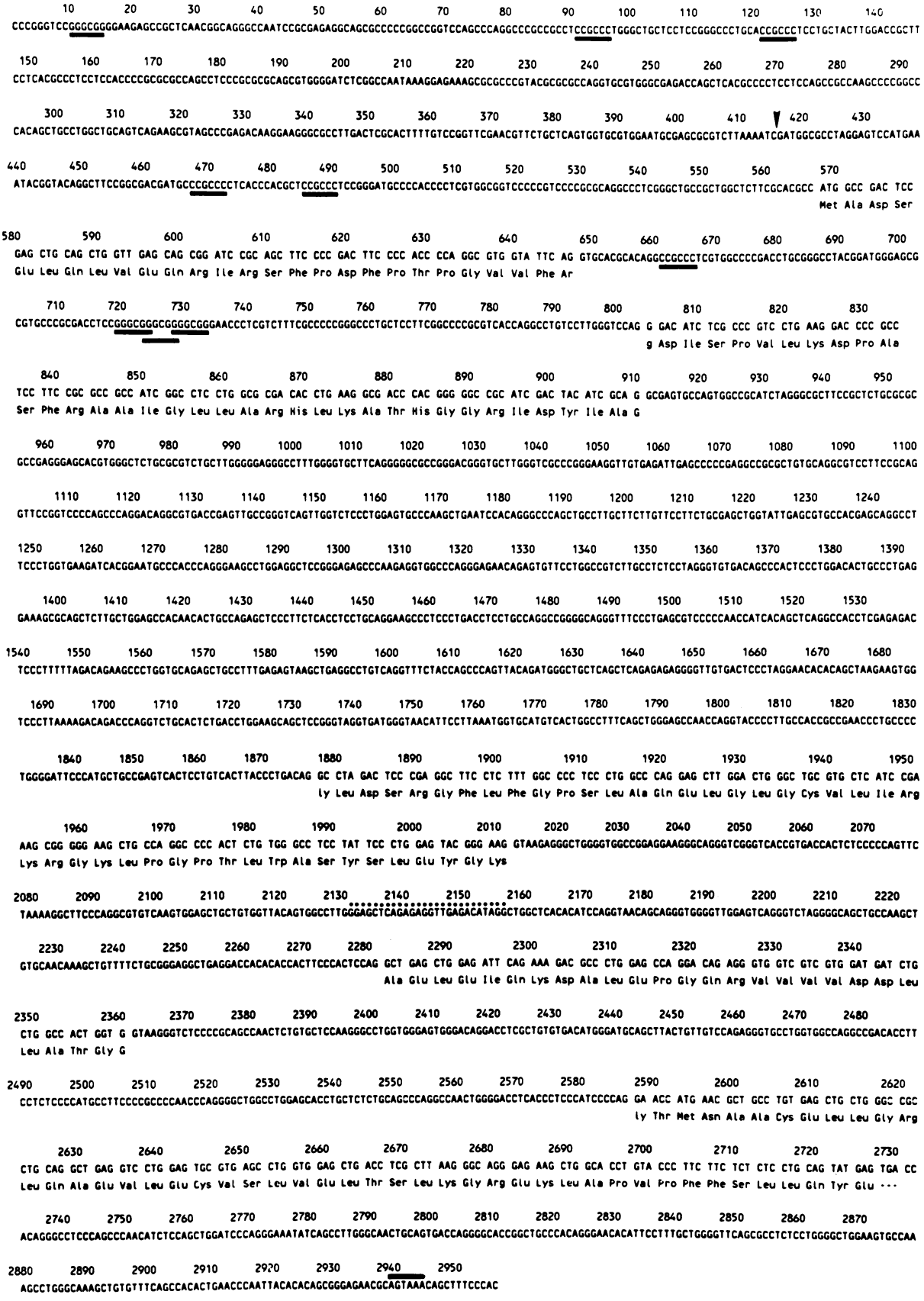

FIG. 1. Nucleotide sequence of the human *APRT* gene and its 5′ flanking region. The deduced amino acid sequence encoded by the five exons is shown beneath. GC boxes within the 5′ flanking sequence and intron 1 are underlined. The arrowhead identifies a unique *Cla* I site (nucleotide 415). The conserved sequence in intron 3 is highlighted by a series of superscripted dots, and the poly(A) signal is identified by a superscript bar. See text for determination of the genomic sequence. The sequences of the protein coding and 3′ untranslated regions were confirmed from cDNAs.

respectively. The fibroblast library was screened by colony hybridization (24), whereas plaques produced by the HepG2 and T-cell libraries were screened as described by Benton and Davis (25).

DNA Sequencing and Homology Analyses. Genomic and cDNA restriction fragments were subcloned in both orientations into pIBI 20 and subjected to nucleotide sequence analysis using a modification (26) of the dideoxynucleotide chain-termination procedure (27). Since the region between nucleotides 805 and 900 (Fig. 1) consistently yielded ambiguous results due to band compression, the sequence of both strands of this segment was confirmed by the Maxam and Gilbert method (28). Nucleotide sequence homology searches were performed using an International Biotechnologies program based on that described by Pustell and Kafatos (29).

RESULTS

The nucleotide sequence of the human *APRT* gene is displayed in Fig. 1. The genomic sequence was determined from a series of overlapping subclones obtained from λHuap15, which we had demonstrated by transfection to contain an entire functional human *APRT* gene (5). The sequence of the coding region was confirmed by sequencing several independently isolated cDNAs retrieved from plasmid or λ phage libraries. The longest cDNA sequenced extended from the 3′ poly(A) site to within 17 nucleotides of the ATG translation start codon at position 568 (Fig. 1). We deduce that translation initiates at this start codon since it is in the same reading frame as the remainder of the cDNA and since this amino terminus precisely coincides with that of mouse *APRT*. Though there are other potential ATG start sites further upstream, only the one at position 568 is preceded by the consensus eukaryotic initiation sequence described by Kozak (30, 31).

Like the mouse *APRT* gene (14) the promoter region lacks TATA or CCAAT-like sequences. However, there are 5 GC boxes 5′ to the coding region that may serve as potential binding sites for the Sp1 transcription factor (9–12). Since removal of DNA upstream of the *Cla* I site at position 413 (arrowhead in Fig. 1) permits efficient *APRT* expression, as assayed by transfection of Aprt$^-$ recipient mouse L cells (data not shown), it appears that the GC boxes distal to the *Cla* I site are dispensable. The two remaining GC boxes, beginning at position 467 and position 485, respectively (Fig. 1), are located 101 and 83 base pairs (bp) from the translation start site. This arrangement is similar to that of the mouse *APRT* promoter region, where the two most proximal GC boxes are 99 and 81 bp upstream from the ATG start codon (14). Curiously, the core hexanucleotide of the Sp1 recognition sequence also appears four times within the first intron. Three of these sequences have an overlapping arrangement; however, only one, GGGGCGGGAA, conforms to the consensus decanucleotide sequence that apparently is required for efficient Sp1 binding (12). The above sequence conforms to the consensus decanucleotide at 9 of the 10 positions, with only the 3′ terminal nucleotide diverging.

Comparison of APRT amino acid sequences reflects strong evolutionary conservation. The human amino acid sequence, deduced from the nucleotide sequence of cloned cDNAs and the functional gene, is 82% and 90% identical to the mouse and hamster sequences, respectively (Fig. 2*a*). The hamster amino acid sequence was deduced from a published nucleotide sequence (15). Though this sequence as published lacks a sufficient open reading frame, insertion of any single nucleotide 37 bp downstream from the translation initiation site corrects the reading frame and produces a protein very similar to its human and mouse counterparts. The extent of APRT amino acid conservation is most vividly illustrated by comparison of prokaryotic and mammalian sequences. The *E. coli* APRT amino acid sequence (32) is about 40% identical to that of the human and rodent enzymes (Fig. 2*a*).

We had previously identified an amino acid sequence conserved between several bacterial and mammalian phosphoribosyltransferases (14). This sequence is underlined in Fig. 2*a* and is invariant over much of its length. Where substitutions have occurred, they are mostly neutral and isosteric. The amino acid sequence of an *E. coli* phosphoribosylpyrophosphate synthetase has been determined (33) and also contains at least part of this conserved sequence (Fig. 2*b*). Since phosphoribosyltransferases and PRPP synthetase (in its reverse reaction) have PRPP as a common substrate, this sequence is a likely candidate for at least part of a PRPP binding site.

a

```
Human:  M A D S E L Q L V E Q R I R S F P D F P T P G V V F R D I S P V L K D P A
Mouse:    S E P     K     A R       V           I       L             L         D
Hamster:    E               A                   I       L             L
E. coli:    T A T A Q     E Y L K N S I K S(EQ) Y   K     I L       V T S L   E     K

Human:  S F R A A I G L L A R H L K A T H G G R I D Y I A G L D S R G F L F G P S
Mouse:          S   R       S       S     S   K
Hamster:        S   R       S       S         K
E. coli:  A Y A L S   D       E R Y   N A G I T K V V G T E A - - -               A P

Human:  L A Q E L G L G C V L I R K R G K L P G P T L W A S Y S L E Y G K A E L E
Mouse:              V                 Q             V S
Hamster:                                          V S       A
E. coli:  V   L G     V   F   P V     P         R E   I S E T   D         T D Q

Human:  I Q K D A L E P G Q R V V V V D D L L A T G G T M N A A G E L L R A L Q A
Mouse:                        I                           F       D     H Q   R
Hamster:                    K                             C             G Q
E. coli:    H V     I K     D K   L                       I E   T V K   I   R   G G

Human:  E V L E C V S L V E L T S L K G R E K L A P V P F F S L L Q Y E
Mouse:    V                                 R   G   I                   D
Hamster:  V                                     G S
E. coli:    A D A A F I I N   F D   G(EQ)(L)   Q G - I T S Y     V P F P G H
```

b

```
Human APRT:                 G Q R V V V V D D L L A T G G T M N A A
E. coli PRPP Synthetase:      R D C   L       M I D         L C K
```

FIG. 2. Amino acid conservation within APRT. (*a*) Amino acid sequence comparison of human APRT with rodent and *E. coli* APRTs. The deduced human sequence is presented in full. Amino acids deduced from mouse (14), hamster (15), and *E. coli* (32) gene sequences are shown only when they differ from the corresponding amino acid of the human enzyme. A blank space indicates amino acid identity. Dashes indicate deletions, and subscripted amino acids bounded by parentheses indicate insertions in the *E. coli* sequence with respect to the mammalian enzymes. The conserved sequence common to mammalian and prokaryotic phosphoribosyltransferases (14) is underlined. (*b*) Amino acid sequence comparison between the putative PRPP binding site of human APRT and *E. coli* PRPP synthetase. The human sequence is presented in full. Spaces in the *E. coli* sequence indicate identity with the corresponding amino acid in the human sequence.

```
Human:   G G A G C T C A G A G A G G T T G A G A C A T A G G
Mouse:     T                                           T
Hamster:   G                             A     T       C A
```

FIG. 3. Conserved nucleotides within intron 3 of human and rodent *APRT* genes. The 26-nucleotide sequence from intron 3 of the human gene is presented in its entirety. Only those nucleotides that differ are shown in the mouse and hamster sequences.

The nucleotide sequence of the human *APRT* gene was compared to that of the mouse (14) to identify conserved sequences in noncoding regions. It is reasonable to speculate that such sequences may be important for *APRT* expression. As expected, the protein coding sequences are highly homologous between species. However, DNA sequences within 5′ flanking and 3′ untranslated regions as well as within introns are extensively diverged. Upstream divergence begins four nucleotides 5′ to the ATG translation start codon and displays little significant homology, except within a region in which both genes contain two closely spaced GC boxes. Despite sequence divergence, the introns all interrupt protein coding sequences at precisely the same positions in the human and mouse genes. Interestingly, intron 2 also contains an unusual AG/GC splice donor site that is common to intron 2 of the human and both rodent species. The only conserved intron nucleotide sequences occur within intron 3, which contains two nucleotide stretches with >75% homology to similarly positioned mouse sequences. The first, which extends from position 2026 to 2055 (Fig. 1), is homologous at 23 of 30 bp; the second (Fig. 3), which extends 26 bp, differs at only 2 bp and its position within the gene is highlighted in Fig. 1. Significantly, the second sequence, differing only at 5 bp, also occurs within intron 3 of the hamster gene (Fig. 3), suggestive of a conserved function. The poly(A) signal in the human gene is AGTAAA and was present in both genomic and cDNA clones. This sequence, which differs from the canonical AATAAA (34, 35) associated with most genes, is not found in the *APRT* genes of either mouse or hamster.

The *APRT* genes display a peculiar distribution of CpG dinucleotides. The dinucleotide CpG is underrepresented in mammalian DNA but appears in clusters within the genome (36). The distribution of CpG dinucleotides within and upstream of the human and mouse *APRT* genes is similar and nonrandom (Fig. 4). In the human gene, there is a cluster of CpG dinucleotides, with frequencies ranging to >10 per 100 bp, that begins at least 500 bp upstream of the ATG translation start site and extends about 200 bp into intron 2. The extent of the CpG-rich region in the mouse gene is more restricted, beginning about 170 bp upstream of the ATG translation initiation codon and extending about 100 bp into intron 2. In both species, the remainder of the *APRT* gene contains fewer CpG dinucleotides than expected based on the G+C content of the gene and a random distribution. That this CpG distribution is not a reflection of G+C content and is apparently not random is demonstrated by the relatively constant GpC distribution over the length of the mouse and human genes.

DISCUSSION

Unlike most mammalian genes that have been characterized, the gene encoding APRT is constitutively expressed and subject to little, if any, regulation. It is possible, therefore, that only minimal sequence information is required for appropriate levels of *APRT* gene transcription. With this assumption in mind, we compared human and rodent *APRT* genes to detect conserved characteristics that may have functional significance. As expected, the protein coding regions of the mouse and human genes are very similar to each other. Furthermore, the *E. coli* enzyme is about 40% identical to the mammalian enzymes, indicating that these purine salvage enzymes have a common origin and that retention of enzymatic function imposes significant mutational constraints upon *APRT* genes.

There are several interesting features associated with the human and rodent *APRT* genes. Their organization is very similar; and though their introns have undergone extensive sequence divergence and vary somewhat in size (Fig. 4), they share identical splice sites. The rare splice donor site AG/GC is present in intron 2 in each of the three genes rather than the almost invariant AG/GT tetranucleotide, which, with the exception of avian α-globin (37, 38) and murine α-crystalline (39), occurs in all of the ≈400 vertebrate genes in the GenBank data base (40). The high degree of conservation of the 26-bp region within intron 3 and its retention within that intron suggest that this nucleotide sequence is subject to stringent selection and possibly serves a functional role. Though it is common to find nucleotide sequence conservation within 5′ and 3′ flanking regions of genes from different species encoding tissue-specific proteins, or genes subject to cell cycle and/or hormonal regulation (e.g., refs. 41–45), no such conservation is apparent within flanking sequences of *APRT* genes from different species. There is no substantial homology between >220 bp of their 3′ untranslated regions or between >500 bp of their 5′ flanking regions, except within a region about 100 bp upstream of the ATG start codon where both genes contain two closely spaced potential Sp1 binding sites that account for the homology.

Although the *APRT* genes lack functional TATA and CCAAT sequences, they are not unique in this respect. Several other genes encoding housekeeping functions have promoters

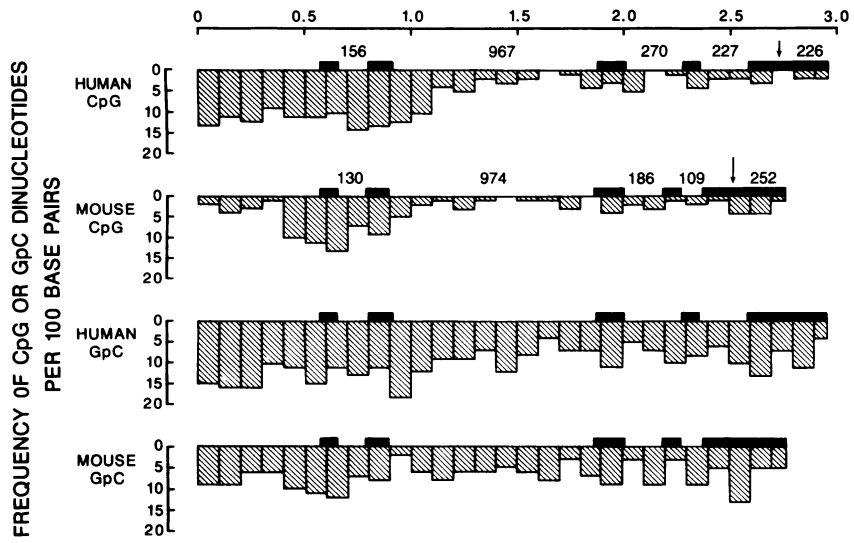

FIG. 4. Distribution of CpG and GpC dinucleotides within the human and mouse *APRT* genes and their 5′ flanking sequences. The positions of exons are defined by solid boxes. The human and mouse genes are aligned at their respective ATG translation start codons. The size of each intron is indicated in base pairs, as is the length of the untranslated 3′ region. The positions of the translation termination signals are indicated by vertical arrows. The width of each bar represents 100 bp and the height denotes the number of CpG or GpC dinucleotides within that 100 bp. The scale is in kilobases.

also lacking these sequences but containing multiple GC boxes (12, 42, 46–49). These GC boxes have the potential for binding the Sp1 transcription factor (9–12) and, in some cases, can promote bidirectional transcription (10, 50), but in the mouse gene transcription appears to be unidirectional (M.K.D., unpublished observations). In some genes described (12, 42, 47–49), the GC boxes are in the opposite orientation of the CCGCCC sequences contained in the 5′ region of human and mouse *APRT* genes. The significance, if any, of the orientation of putative Sp1 binding sites is unclear.

A striking and conserved feature of the *APRT* genes is the distribution of CpG dinucleotides. Although the mammalian genome is about 40% G+C (51), CpG dinucleotides, which may serve as substrates for methylation, are underrepresented, occurring with a frequency of about only 2%. However, Bird (36) has underscored that clusters of CpG dinucleotides primarily occur as discrete islands. Although the mouse and human *APRT* genes have a G+C content >55%, the CpG dinucleotide frequency over the 3′ two-thirds of the genes is less than half than expected were the CpG distribution random. In contrast, the 5′ ends of mouse and human *APRT* genes have a greater than random representation of CpG dinucleotides. Particularly striking is that although the intron 1 sequences of mouse and human *APRT* genes have no apparent homology, both have retained a very high CpG content. This suggests that sequence divergence within intron 1 was not random but subject to selection for a high CpG dinucleotide content. The same observation can be made for the 200 nucleotides extending upstream from the ATG start signal, indicating selection and a possible functional role for the high CpG content in these regions. The possibility that CpG-rich regions may interact with one or more proteins to influence gene transcription has been addressed (36). *In vitro* methylation of the 5′ end but not the 3′ half of the hamster *APRT* gene rendered it nonfunctional (52). Thus, it appears that the CpG-rich domain needs to be unmethylated to exert its effect. It is, therefore, plausible to speculate that in its hypomethylated state, the CpG-rich domains common to the human and murine *APRT* genes may contribute to their constitutive expression.

We thank Ms. Estrella Feliciano for expert technical assistance, Mrs. Susan Eder for help with the manuscript, and Dr. Sandra Degen for help with the Maxam and Gilbert sequencing. This work was supported by National Science Foundation Grant PCM 8118283 and National Institutes of Health Grants CA-36897, DK37762, and DK38185.

1. Kamatani, N., Kubota, M., Willis, E. H., Frinke, L. A. & Carson, D. A. (1984) *Adv. Exp. Med. Biol.* **165,** Part B, 83–88.
2. Simmonds, H. A. & Van Acker, K. J. (1982) in *The Metabolic Basis of Inherited Disease*, eds. Stanbury, J. B., Fredrickson, D. S., Goldstein, J. L. & Brown, M. S. (McGraw-Hill, New York) 5th Ed., pp. 1144–1183.
3. Tischfield, J. A. & Ruddle, F. H. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 45–49.
4. Fratini, A., Simmers, R. N., Callen, D. F., Hyland, V. J., Tischfield, J. A., Stambrook, P. J. & Sutherland, G. R. (1986) *Cytogenet. Cell Genet.* **43,** 10–13.
5. Stambrook, P. J., Dush, M. K., Trill, J. J. & Tischfield, J. A. (1984) *Somatic Cell Mol. Genet.* **10,** 359–367.
6. Epstein, C. (1970) *J. Biol. Chem.* **245,** 3289–3294.
7. Hordern, J. & Henderson, J. F. (1982) *Can. J. Biochem.* **60,** 422–433.
8. Melton, D. W., Konecki, D. S., Brennand, J. & Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2147–2151.
9. Dynan, W. S. & Tjian, R. (1983) *Cell* **35,** 79–87.
10. Dynan, W. S., Saffer, J. D., Lee, W. S. & Tjian, R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4915–4919.
11. Jones, K. A. & Tjian, R. (1985) *Nature (London)* **317,** 179–182.
12. Dynan, W. S., Sazer, S., Tjian, R. & Schimke, R. T. (1986) *Nature (London)* **319,** 246–248.
13. Sikela, J. M., Khan, S. K., Feliciano, E., Trill, J. A., Tischfield, J. A. & Stambrook, P. J. (1983) *Gene* **22,** 219–228.
14. Dush, M. K., Sikela, J. M., Khan, S. A., Tischfield, J. A. & Stambrook, P. J. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2731–2735.
15. Nalbantoglu, J., Phear, G. A. & Meuth, M. (1984) *Nucleic Acids Res.* **14,** 1914.
16. Adair, G. M., Stallings, R. L., Nairn, R. S. & Siciliano, M. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5961–5964.
17. Dickerman, L. H. & Tischfield, J. A. (1978) *Mutat. Res.* **49,** 83–94.
18. Grosovsky, A. J., Drobetsky, E. A., DeJong, P. & Glickman, B. W. (1986) *Genetics* **113,** 405–415.
19. Nalbantoglu, J., Hartley, D., Phear, G., Tear, G. & Meuth, M. (1986) *EMBO J.* **5,** 1199–1204.
20. Dente, L., Cesarini, G. & Cortese, R. (1983) *Nucleic Acids Res.* **11,** 1645–1655.
21. Okayama, H. & Berg, P. (1983) *Mol. Cell. Biol.* **3,** 280–289.
22. Wiginton, D. A., Adrian, G. S., Friedman, R. L., Suttle, P. D. & Hutton, J. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7481–7485.
23. Faust, P. L., Kornfeld, S. & Chirgwin, J. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4910–4914.
24. Hanahan, D. & Meselson, M. (1983) *Methods Enzymol.* **100,** 333–342.
25. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
26. Brunner, A. M., Schimenti, J. C. & Duncan, C. H. (1986) *Biochemistry* **25,** 5028–5035.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
28. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
29. Pustell, J. & Kafatos, F. C. (1984) *Nucleic Acids Res.* **12,** 643–655.
30. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
31. Kozak, M. (1986) *Cell* **44,** 283–292.
32. Hershey, H. V. & Taylor, M. W. (1986) *Gene* **43,** 287–293.
33. Hove-Jensen, B., Harlow, K. W., King, C. J. & Switzer, R. L. (1986) *J. Biol. Chem.* **261,** 6765–6771.
34. Proudfoot, N. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
35. Wickens M. & Stephenson, P. (1984) *Science* **226,** 1045–1051.
36. Bird, A. P. (1986) *Nature (London)* **321,** 209–213.
37. Erbil, C. & Niessing, J. (1983) *EMBO J.* **2,** 1339–1343.
38. Dodgson, J. B. & Engel, J. D. (1983) *J. Biol. Chem.* **258,** 4623–4629.
39. King, C. R. & Piatigorsky, J. (1983) *Cell* **32,** 707–712.
40. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55,** 1119–1150.
41. Gunning, P., Mohun, T., Ng, S. Y., Ponte, P. & Kedes, L. (1984) *J. Mol. Evol.* **20,** 202–214.
42. Chen, M. J., Shimada, T., Moulton, A. D., Cline, A., Humphries, J. & Nienhuis, A. W. (1984) *J. Biol. Chem.* **259,** 3933–3943.
43. Stevens, B. & Luskey, K. L. (1985) *J. Biol. Chem.* **260,** 10271–10277.
44. Jameson, L., Chin, W. W., Hollenberg, A. N., Chang, A. S. & Habener, J. F. (1984) *J. Biol. Chem.* **259,** 15474–15480.
45. Li-yuan, Y. L., Richter-Mann, L., Couch, G. H., Stewart, A. F., Mackinlay, A. G. & Rosen, J. M. (1986) *Nucleic Acids Res.* **14,** 1883–1901.
46. Reynolds, G. A., Basu, S. K., Osborne, T. F., Chin, D. J., Gil, G., Brown, M. S., Goldstein, J. L. & Luskey, K. L. (1984) *Cell* **38,** 275–285.
47. Valerio, D., Duyvesteyn, M. G. C., Dekker, B. M. M., Weeda, G., Berkvens, Th. M., van der Voorn, L., van Ormondt, H. & van der Eb, A. J. (1985) *EMBO J.* **4,** 437–443.
48. Melton, D. W., McEwan, C., McKie, A. B. & Reid, A. M. (1986) *Cell* **44,** 319–328.
49. Singer-sam, J., Keith, D. H., Tani, K., Simmer, R. L., Shively, L., Lindsay, S., Yoshida, A. & Riggs, A. D. (1984) *Gene* **32,** 409–417.
50. Gidoni, D., Kadonaga, J. T., Barrera-Saldana, H., Takahashi, K., Chambon, P. & Tjian, R. (1985) *Science* **230,** 511–517.
51. Maio, J. J. (1976) in *Handbook of Biochemistry and Molecular Biology Nucleic Acids*, ed. Fasman, G. D. (CRC, Boca Raton, FL), Vol. 2, pp. 391–399.
52. Keshet, I., Yisraeli, J. & Cedar, H. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2560–2564.