Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 2731–2735, May 1985
Biochemistry

Nucleotide sequence and organization of the mouse adenine phosphoribosyltransferase gene: Presence of a coding region common to animal and bacterial phosphoribosyltransferases that has a variable intron/exon arrangement

(housekeeping gene/DNA sequence/evolution/promoters)

MICHAEL K. DUSH*, JAMES M. SIKELA*†, SOHAIB A. KHAN*, JAY A. TISCHFIELD‡, AND PETER J. STAMBROOK*§

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

Communicated by Igor B. Dawid, January 17, 1985

ABSTRACT **We have determined the nucleotide sequence of a functional mouse adenine phosphoribosyltransferase (*APRT*) gene and its cDNA. The amino acid sequence of the enzyme is deduced from an open reading frame in the cDNA and predicts a protein with a molecular weight of 19,560. The protein coding region of the gene is ≈2 kilobases, and it is composed of five exons and four introns. While the body of the gene is 53% G+C, the 200 nucleotides upstream from the ATG translation start codon are 66% G+C and contain three copies of the sequence C-C-G-C-C-C. The mouse APRT enzyme shares a homologous 20-amino acid sequence with mouse, hamster, and human hypoxanthine phosphoribosyltransferases (HPRTs) and several bacterial phosphoribosyltransferases. This sequence has previously been shown to be a likely catalytic domain in human HPRT and *Escherichia coli* glutamine phosphoribosyltransferase. Because of the similarities in function of these proteins, both eukaryotic and prokaryotic, it is not unexpected that they should exhibit one or more regions of homology, particularly at the 5-phosphoribosyl-1-pyrophosphate and purine binding sites, especially if they are related via a common evolutionary lineage. This homologous sequence is interrupted by a single intron in the mouse *APRT* gene and by two introns in the mouse *HPRT* gene. Furthermore, the positions of both introns in the *HPRT* sequence are different from that of the single intron in the corresponding sequence of the *APRT* gene.**

The gene encoding adenine phosphoribosyltransferase (APRT) has attracted considerable interest because of its utility as a selectable marker (1, 2). The enzyme, which is a member of a family of phosphoribosyltransferases, constitutes a purine salvage pathway that utilizes adenine and 5-phosphoribosyl-1-pyrophosphate (PRPP) to form AMP. Whether or not the enzyme is expressed provides the basis for sensitive forward and backward selection systems that permit selection of $Aprt^-$ or $Aprt^+$ cells, respectively (1, 2). The APRT enzyme has been purified from several mammalian species (3–5) and is a homodimer (3, 5). However, its amino acid sequence has not been determined, and only the amino acid composition of human APRT has been reported (3). Since the amino acid sequences of several eukaryotic and prokaryotic enzymes that bind PRPP and have phosphoribosyltransferase activity are known, it would be instructive to have available a mammalian APRT sequence for comparative purposes. We have previously described the cloning of mouse (6) and human (7) *APRT* genes, and in this report we present the nucleotide sequence of a functional mouse *APRT* gene and the deduced amino acid sequence of the protein.

We also describe a highly conserved (or possibly convergent) nucleotide sequence and its encoded amino acid sequence that is shared by prokaryotic and eukaryotic phosphoribosyltransferases. These include the products of *Escherichia coli gpt* (xanthine guanine phosphoribosyltransferase) (8, 9), *E. coli pur F* (glutamine phosphoribosylpyrophosphate amidotransferase) (10), *Bacillus subtilis pur F* (11), and *E. coli pyr E* (orotate phosphoribosyltransferase) (12), as well as mouse APRT (this report) and mouse (13, 14), hamster (13, 14), and human (15, 16) hypoxanthine phosphoribosyltransferases (HPRTs). A structural analysis of the human HPRT enzyme, in which the amino acid sequence had been directly determined, previously identified this common sequence as an apparent catalytic domain (16). A peculiar feature of this short homologous sequence is that it is divided by introns in mouse *APRT* and *HPRT* genes and that the positions of the introns in these genes are different.

MATERIALS AND METHODS

Cloning of an *APRT* Gene and Its cDNA. We have previously reported that a 3.1-kilobase (kb) *Eco*RI/*Sph* I fragment derived from a recombinant λ phage contains a functional mouse genomic *APRT* gene (6). This DNA fragment was subsequently cloned between the *Eco*RI and *Sph* I sites of pBR328 and the resultant 6.2-kb recombinant plasmid was designated pSAM 3.1. The single copy insert in pSAM 3.1 was used as a probe to screen two different cDNA libraries of mouse liver (17) and L-cell origin (kindly provided by H. Okayama). Several independent APRT cDNA clones were identified.

Nucleotide Sequence Determination of Genomic APRT DNA and cDNA Sequences. The nucleotide sequence of the entire 3.1-kb fragment contained in pSAM 3.1 and of several cloned APRT cDNAs was determined by using the chemical cleavage method of Maxam and Gilbert (18). The sequencing strategy is presented in Fig. 1.

Source of Phosphoribosyltransferase Amino Acid Sequences. The amino acid sequence of each of the prokaryotic phosphoribosyltransferases has previously been deduced from the nucleotide sequence of their respective cloned genes (8–12). The human HPRT amino acid sequence has been determined directly from purified enzyme (16) and confirmed

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; HPRT, hypoxanthine phosphoribosyltransferase; PRPP, 5-phosphoribosyl-1-pyrophosphate; kb, kilobase(s).

†Present address: Department of Anatomy, University of Colorado Medical Center, Denver, CO 80262.
§To whom reprint requests should be addressed.

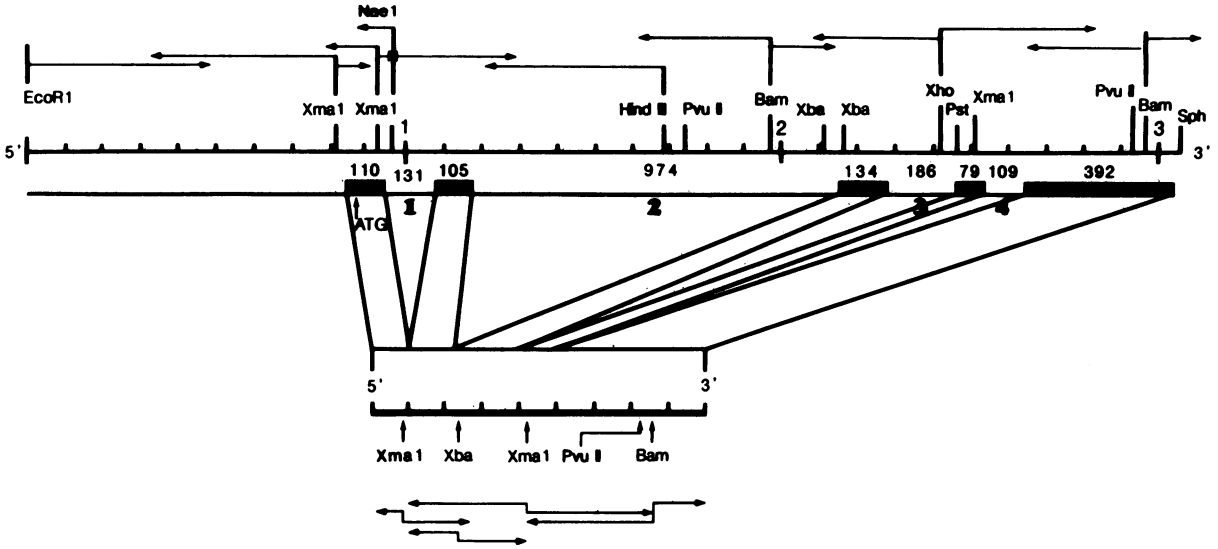

FIG. 1. Organization and restriction map of the mouse *APRT* gene contained within the insert of plasmid pSAM 3.1. Solid blocks represent exons, which are aligned with a mouse APRT cDNA restriction map below. The nonblocked regions within the gene represent introns. Horizontal arrows indicate direction of nucleotide sequencing and the positions from which sequencing was initiated. The position of the ATG translation start codon is indicated. The cDNA nucleotide sequence was determined from independent cloned cDNAs, and it is identical to that of the boxed regions represented in the genomic DNA.

by analysis of cloned cDNAs (15). The mouse and hamster HPRT amino acid sequences have been deduced from cloned cDNAs (13, 14) and in mouse, analysis of the cloned gene identified the position of each of the known introns.

RESULTS

Nucleotide Sequence and Organization of the Mouse *APRT* Gene. The organization of the *APRT* gene was established by comparison of nucleotide sequences derived from the genomic insert of the plasmid pSAM 3.1 and several independent cDNA clones (Fig. 1). Sequence analysis of the cDNAs revealed a single open reading frame of sufficient length to encode APRT. Comparison with the genomic sequence indicates that the protein coding portion of the gene is composed of five exons and four introns and is ≈2 kb. The nucleotide sequence of the 3.1-kb genomic insert, the nucleotide sequence encoding the protein, and the deduced amino acid sequence of mouse APRT are presented in Fig. 2. The molecular weight of mouse APRT monomer, calculated from its deduced amino acid sequence, is 19,560 and agrees with molecular weight estimates of purified APRT protein from other mammals (3–5). Its amino acid composition is similar to that reported for human APRT (3).

The overall organization of the *APRT* gene is similar to that of other mammalian genes. Except for an A-G/G-C instead of an A-G/G-T at the intron 2 donor splice site, the nucleotide sequence at each of the splice junctions conforms to the consensus sequences previously established (19). There is a canonical A-A-T-A-A-A polyadenylylation signal (20) at position 3043, consistent with the 255-nucleotide 3′ untranslated region that we detect in the sequenced cDNAs. Immediately 5′ to the AUG translation start signal of most eukaryotic mRNAS, Kozak (21) has reported the consensus sequence C-C-$\frac{A}{G}$-C-C. Consistent with this observation, the mouse gene contains the sequence C-G-G-C-C at precisely this position (Fig. 2). The body of the gene has a 53% G+C content, while the 200 nucleotides upstream of the ATG translation start codon are 66% G+C. One characteristic of constitutively expressed housekeeping genes appears to be G+C-rich 5′ and 5′ flanking sequences. The mouse *APRT* gene G+C-rich region displays the sequence C-C-G-C-C-C three times within 150 nucleotides of the ATG start codon (position 873) at positions 737, 774, and 792. This same sequence, which occurs in the G+C-rich region upstream of hydroxymethylglutaryl-CoA reductase transcription start sites (22), in the simian virus 40 promoter region (23), and 5′ to the Herpes thymidine kinase (*tk*) gene (24), may be functionally important.

Amino Acid Homology Between Prokaryotic and Eukaryotic Phosphoribosyltransferases. Mouse APRT contains a homologous 20 residue amino acid sequence common to *E. coli gpt* (8, 9), *E. coli pur F* (10), *B. subtilis pur F* (11), and *E. coli pyr E* (12) gene products and to human (15, 16), hamster (13, 14), and mouse (13, 14) HPRTs (Fig. 3). A previous analysis of human HPRT and *E. coli* glutamine phosphoribosylpyrophosphate amidotransferase suggested that this region may be part of a catalytic domain (16). All of the above proteins exhibit a region of homology that encompasses 20 amino acids and contains an invariant aspartic acid and glycine. There are also three positions at which only a single amino acid change occurs (Val → Phe; Val → Ile; Thr → Ala), one site in which a glycine is replaced by an aspartic acid in the *E. coli pur F* product and by a glutamine in the *E. coli pyr E* product, and one site where an aspartic acid is replaced by a glutamic acid in all of the HPRTs. The majority of amino acid substitutions in this region are isosteric and neutral in charge, with minimal potential for affecting function. The major exception is the second amino acid of the sequence that is a glutamic acid in the *E. coli* GPT but a lysine in the HPRTs and the *E. coli* and *B. subtilis pur F* proteins. This substitution produces a net charge change of +2 at this site.

The Position of Introns Within the Homologous Sequence Is Different in Mouse *APRT* and *HPRT* Genes. Although mouse HPRT and APRT perform similar catalytic functions, their genes differ in size and intron number (6, 14). The gene encoding mouse HPRT is 33 kb and contains nine exons and eight introns (14). The functional *APRT* gene is contained on a 3-kb fragment, and it is composed of five exons and four introns within the protein coding region (Figs. 1 and 2). It is significant that the only amino acid sequence that exhibits striking homology between these two enzymes is that sequence described in Fig. 3. Comparison of the intron–exon organization within this short region common to both genes reveals that not only is the homologous sequence interrupted

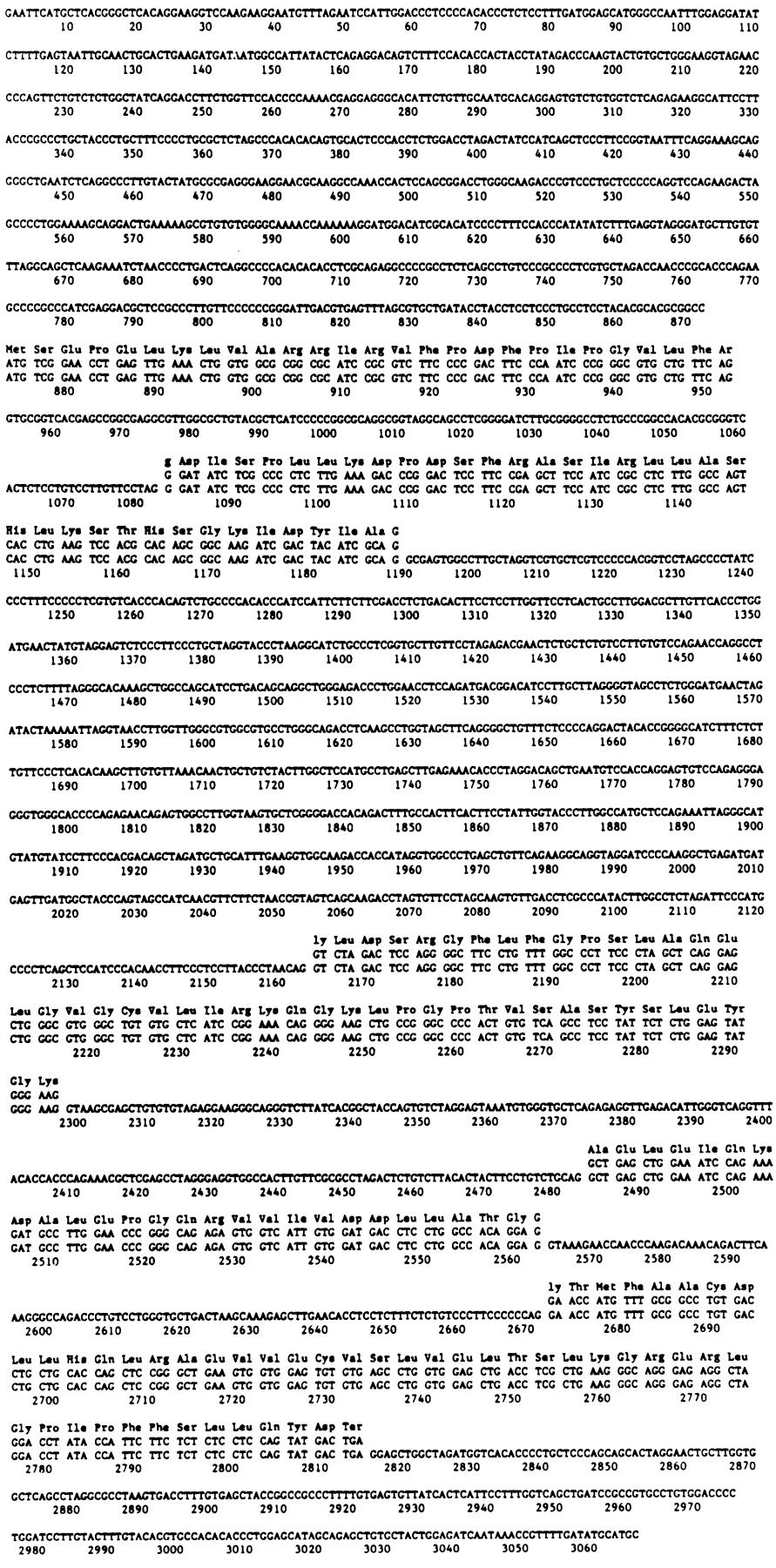

FIG. 2. The complete nucleotide sequence of the 3.1-kb fragment containing the mouse *APRT* gene. The cDNA nucleotide sequence and the corresponding amino acid sequence of the protein are shown for the exon regions of the gene. Numbering begins at the 5′ side of the insert.

in both genes, but the number of introns and their positions in each gene differs (Fig. 4). The *APRT* gene is interrupted once in this region, within the glycine codon representing the 15th amino acid of the sequence. In *HPRT*, the conserved sequence is interrupted twice. The first intron occurs between the 2nd and 3rd amino acids, while the second intron

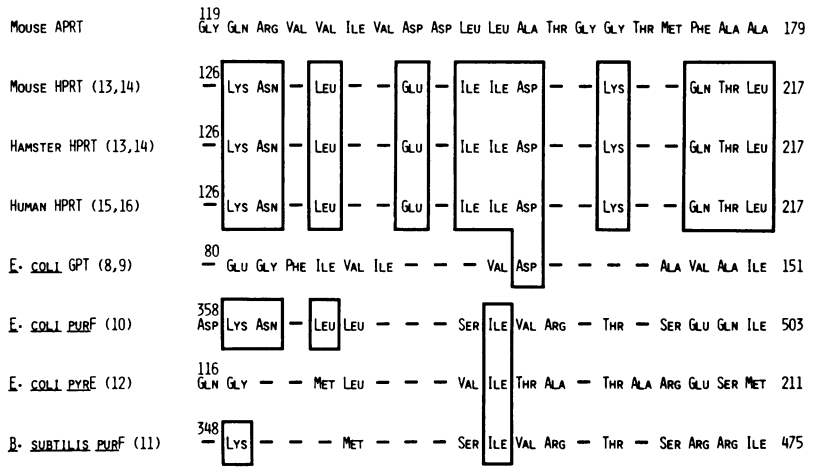

FIG. 3. Homology of amino acid sequence within mammalian and prokaryotic phosphoribosyltransferases. The amino acids within a 20-residue region are aligned to illustrate homology. Amino acids 119–138 of the mouse APRT sequence, derived from the corresponding nucleotide sequence, are presented. Dashes within the sequence of other phosphoribosyltransferases indicate identity with mouse APRT residues. The enclosed amino acids identify residues identical to mouse HPRT at the same position but different from mouse APRT. Numbers at the start of each peptide identify the amino acid position within the complete protein, while the number at the end of each sequence indicates the full length of the protein. Reference from which each sequence was obtained is indicated in parentheses.

separates the 8th amino acid (glutamate) from the invariant aspartate.

DISCUSSION

We have previously shown that the 3.1-kb DNA fragment, whose nucleotide sequence is presented in Fig. 2, has the capacity to transform Aprt$^-$ cells to the Aprt$^+$ phenotype via DNA-mediated transfection (6). The gene is small, which will facilitate its use for *in vitro* mutagenesis, for studying regulation of expression, and for identifying sites within the protein that are important for enzymatic activity. The protein coding regions were established from the nucleotide sequences of several cloned cDNAs, and the positions and sizes of introns were established by their comparison with the genomic nucleotide sequence. It is interesting to note that one of the cDNAs sequenced had retained the entire first intron (data not shown), indicating that it may have been derived from an incompletely processed messenger RNA. If there is an ordered excision of introns, the organization of this cDNA suggests that intron 1 may be the last to be excised.

While we have not yet mapped where transcription begins, we have shown that a cloned processed *APRT* pseudogene diverges from the genomic sequence 31 nucleotides upstream from the ATG translation start codon (data not shown). The position of divergence falls within a sequence at position 842 (an adenine flanked by a string of pyrimidines) that comprises a consensus cap site. Similarly, the 5′ end of our longest cDNA terminates at about the same position (data not shown). This cDNA was retrieved from a cDNA library constructed by Okayama and Berg, which selects for cDNAs approaching full length (25). Together, these observations suggest that transcription may start in this region. Although there is a consensus cap site (26) in this region, there is no "TATA"-like or "CCAAT" sequence (26). This is consistent with the 5′ sequences of other housekeeping genes such as *HPRT* (14), hydroxymethylglutaryl-CoA reductase (22), and dihydrofolate reductase (27), which also appear to lack the sequence signals commonly associated with RNA polymerase II promoters. Furthermore, the 200 nucleotides upstream from the ATG start codon, which encompass the pseudogene divergence site, have a 66% G+C content compared to 53% for the body of the gene. A characteristic of sequences upstream of several housekeeping genes described (14, 22, 27) is that they are G+C-rich. Also, the sequence C-C-G-C-C-C occurs three times within 135 nucleotides upstream from the ATG start codon of the mouse *APRT* gene. This same sequence occurs twice within each of the 21-base-pair repeats that comprise part of the simian virus 40 promoter (23) and three times in the G+C-rich region upstream of the hydroxymethylglutaryl-CoA reductase cap sites (22). In the Herpes thymidine kinase (*tk*) gene, which possesses all of the sequence signals associated with strong polymerase II promoters, the sequence C-C-G-C-C-C appears at position −105 and is apparently critical for maintaining high levels of transcription (24). It is, therefore, attractive to speculate that as with other housekeeping genes analyzed, the promoter of the mouse *APRT* gene may lack characteristic TATA and CCAAT boxes and that transcription may begin in the G+C-rich region described.

The homology between a 20-amino acid sequence in mouse APRT and other phosphoribosyltransferases, coupled with the likely catalytic function of this region of the protein (16), suggests that this sequence may have an ancient evolutionary origin. It has been argued that functional domains of proteins are encoded by individual exons and that one mechanism of evolution is the sorting and shuffling of exons (28). There are now several examples to support the original suggestion that introns separate portions of genes that encode functional and structural domains in their corresponding proteins (29–35). The most detailed analysis has been that describing the globin gene family. There are two introns whose positions are constant within the mammalian α-type and β-type globin genes (32) and in the seal myoglobin gene (31), although introns in the myoglobin gene are much larger than their globin gene counterparts. In the globin genes, introns separate structural units within the protein (32). The soybean leghemoglobin gene has an additional central intron (36–38) whose position was predicted by Go's structural analysis of globin (32). Curiously, insect globin genes lack introns altogether (39). One way to rationalize these observations has been to invoke a putative ancestral gene with at least three introns, all of which have been retained in leghemoglobin genes, two of which have been retained in mammalian globin and myoglobin genes, and none in the insect genes. Intron

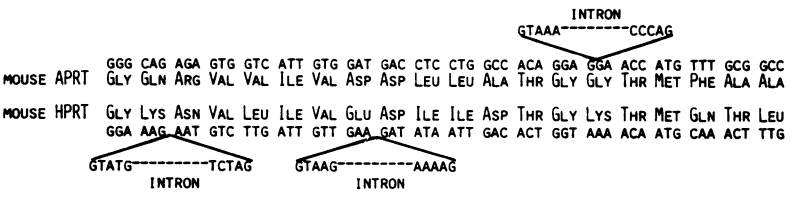

FIG. 4. Position of introns within the nucleotide sequences of mouse *APRT* and *HPRT* (14) gene regions encoding the region of homology. The first five nucleotides within the donor and acceptor sequences of each of the introns are also presented.

deletion has also been postulated for the evolution of intron organization in actin genes from sources as diverse as plants, mammalian skeletal muscle and nonskeletal tissues, and invertebrates. In contrast to globin and actin genes and to members of other gene families, such as genes encoding renin and pepsinogen (35), the structural homology between human α_1-antitrypsin and chicken ovalbumin is not reflected by the positions of introns in their respective genes (40). Although these two proteins are encoded by representatives of a common gene family, the numbers, sizes, and positions of their introns are different (40) and may reflect the very different functions of these proteins.

The homologous amino acid sequence we describe embraces a family of enzymes (i.e., phosphoribosyltransferases) obtained from prokaryotic and mammalian sources. Since these enzymes perform similar catalytic functions, it is not surprising that there should be at least one region of homology between them. That the prokaryotic phosphoribosyltransferases share a common sequence with mammalian *APRT* and *HPRT* argues that these genes may have arisen from a common primordial gene. Although we cannot eliminate the possibility that these enzymes and their genes arose by convergent evolution, we consider this possibility improbable because convergent evolution is more likely to independently produce a similar three-dimensional organization rather than a common amino acid sequence. In support of this argument, *Salmonella typhimurium* ATP phosphoribosyltransferase, which like the other phosphoribosyltransferases utilizes PRPP, lacks the homologous sequence common to the enzymes presented in Fig. 3 (41).

If the mammalian *APRT* and *HPRT* genes and the prokaryotic genes encoding the phosphoribosyltransferases shown in Fig. 3 arose from a common ancestral sequence, the introns were likely introduced into the mammalian genes after development of enzymatic function. The prokaryotic phosphoribosyltransferase genes indicated in Fig. 3 lack introns entirely. However, since there are occasional examples of prokaryotic genes with an intron (42, 43), we cannot exclude an ancestral gene with three introns within the homologous sequence, two of which were lost in the mouse *APRT* gene, one in the mouse *HPRT* gene, and all in the prokaryotic genes. The existence of introns within the genome of a prokaryotic ancestor to both eukaryotes and current common prokaryotes has previously been postulated (44). If three introns were introduced into the common sequence early in eukaryotic evolution, the intron organization in this region of mouse *APRT* and *HPRT* genes could be accounted for by intron deletion. Alternatively, if *APRT* and *HPRT* genes arose independently with no common ancestral origin, so that the homologous region is a consequence of convergent evolution, intron insertion could have occurred independently at any time during the evolution of each of these genes. Regardless of the evolutionary scenario, introns came to reside within a genomic region encoding part of an apparent catalytic domain (16), and in each gene the pattern of insertion is different. The fact that introns occur at all in this region indicates that DNA sequences encoding functional domains may be interrupted by introns as well as being bounded by them.

Note Added In Proof. Primer extension and RNA protection experiments localize transcription initiation to the site proposed in the text within the G+C—rich region. There is also a consensus polymerase II promoter 660 bp further upstream from this point with a possible cap site at position 176, a TATA box at position 153, and C-C-A-A-T-T sequence at position 95 (Fig. 2). The amino acid sequences of *E. coli* APRT (M. W. Taylor and H. V. Hershey, personal communication) and mouse APRT are 42% and 45% identical for the entire protein and putative catalytic domain, respectively, supporting evolutionary conservation.

We thank Ms. Estrella Feliciano for her technical help and Drs. Craig Duncan and Ken Blumenthal for their helpful comments. M.K.D. is supported by a fellowship from the Albert J. Ryan Foundation. This work was supported by National Science Foundation Grant PCM 8118283 and National Institutes of Health Grant CA-36897.

1. Kusano, T., Long, C. & Green, H. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 82–86.
2. Tischfield, J. A. & Ruddle, F. H. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 45–49.
3. Holden, J. A., Meredith, G. S. & Kelley, W. H. (1979) *J. Biol. Chem.* **254**, 6951–6955.
4. Hershey, H. V. & Taylor, M. W. (1978) *Prep. Biochem.* **8**, 453–462.
5. Tischfield, J. A., Jourgenson, D. & Trill, J. J. (1981) *J. Cell Biol.* **91**, 388a (abstr.).
6. Sikela, J. M., Kahn, S. K., Feliciano, E., Trill, J., Tischfield, J. A. & Stambrook, P. J. (1983) *Gene* **22**, 219–228.
7. Stambrook, P. J., Dush, M. K., Trill, J. J. & Tischfield, J. A. (1984) *Somatic Cell Mol. Genet.* **10**, 359–367.
8. Richardson, K. K., Fostel, J. & Skopek, T. (1983) *Nucleic Acids Res.* **11**, 8809–8816.
9. Pratt, D. & Subramani, S. (1983) *Nucleic Acids Res.* **11**, 8817–8823.
10. Tso, J. Y., Zalkin, H., van Cleemput, M., Yanofsky, C. & Smith, J. (1982) *J. Biol. Chem.* **257**, 3525–3531.
11. Makaroff, C. A., Zalkin, H., Switzer, R. L. & Vollmer, S. J. (1983) *J. Biol. Chem.* **258**, 10586–10593.
12. Poulsen, P., Jensen, K. F., Valentin-Hansen, P., Carlsson, P. & Lundberg, L. G. (1983) *Eur. J. Biochem.* **135**, 223–229.
13. Konecki, D. S., Brennand, J., Fuscoe, J. C., Caskey, C. T. & Chinault, A. C. (1982) *Nucleic Acids Res.* **10**, 6763–6775.
14. Melton, D. W., Konecki, D. S., Brennand, J. & Caskey, C. T. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2147–2151.
15. Jolly, D. J., Okayama, H., Berg, P., Esty, A. C., Filpula, D., Bohlen, P., Johnson, G. G., Shively, J. E., Hunkapillar, T. & Friedman, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 477–481.
16. Argos, P., Hanei, M., Wilson, J. M. & Kelley, W. N. (1983) *J. Biol. Chem.* **258**, 6450–6457.
17. Norgard, M. V., Tocci, M. J. & Monahan, J. J. (1980) *J. Biol. Chem.* **255**, 7665–7672.
18. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–574.
19. Mount, S. (1982) *Nucleic Acids Res.* **10**, 459–472.
20. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
21. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
22. 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.
23. Everett, R. D., Batty, D. & Chambon, P. (1984) *Nucleic Acids Res.* **11**, 2447–2464.
24. McKnight, S. L. & Kingsbury, R. (1982) *Science* **217**, 316–324.
25. Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2**, 161–170.
26. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
27. Yang, J. K., Masters, J. N. & Attardi, G. (1984) *J. Mol. Biol.* **176**, 169–187.
28. Gilbert, W. (1978) *Nature (London)* **271**, 501.
29. Sakano, H., Rogers, J. H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. & Tonegawa, S. (1979) *Nature (London)* **277**, 627–633.
30. Stein, J. P., Catterall, J. M., Kristo, P., Means, A. R. & O'Malley, B. W. (1980) *Cell* **21**, 681–687.
31. Blanchetot, A., Wilson, V. & Jeffreys, A. J. (1983) *Nature (London)* **301**, 732–734.
32. Go, M. (1981) *Nature (London)* **291**, 90–92.
33. Campbell, R. D. & Porter, R. R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4464–4468.
34. Verde, P., Stopelli, M. P., Galeffi, P., DiNocera, P. & Blasi, F. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4727–4731.
35. Hobart, P. M., Foglianai, M., O'Connor, B. A., Schaeffer, I. M. & Chirgwin, J. M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5026–5030.
36. Jensen, E. O., Paludan, K., Hyldig-Nielsen, J. J., Jorgensen, P. & Marker, K. A. (1981) *Nature (London)* **291**, 677–679.
37. Brisson, N. & Verma, D. P. S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4055–4059.
38. Wiborg, O., Hyldig-Nielsen, J. J., Jensen, E. O., Paludan, K. & Marker, K. A. (1982) *Nucleic Acids Res.* **10**, 3487–3494.
39. Antoine, M. & Niessing, J. (1984) *Nature (London)* **310**, 795–798.
40. Leicht, M., Long, G. L., Chandra, T., Kurachi, K., Kidd, V. J., Mace, M., Davie, E. W. & Woo, S. L. C. (1982) *Nature (London)* **297**, 655–659.
41. Piszkiewicz, D., Tilley, B. E., Rand-Meir, T. & Parsons, S. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1589–1592.
42. Kaine, B. P., Gupta, R. & Woese, C. R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3309–3312.
43. Chu, F. R., Maley, G. F., Maley, F. & Belfort, M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3049–3053.
44. Doolittle, W. F. (1978) *Nature (London)* **272**, 581–582.