MOLECULAR AND CELLULAR BIOLOGY, Feb. 1986, p. 393–403
0270-7306/86/020393-11$02.00/0

Vol. 6, No. 2

Fine Structure of the Human Hypoxanthine Phosphoribosyltransferase Gene

PRAGNA I. PATEL,[1,2] PAUL E. FRAMSON,[1] C. THOMAS CASKEY,[1,2,3,4] AND A. CRAIG CHINAULT[1,2,3]*

Howard Hughes Medical Institute[1] and Departments of Medicine,[2] Biochemistry,[3] and Cell Biology,[4] Baylor College of Medicine, Houston, Texas 77030

Received 5 June 1985/Accepted 23 October 1985

The human hypoxanthine phosphoribosyltransferase (*HPRT*) gene has been characterized by molecular cloning, mapping, and DNA sequencing techniques. The entire gene, which is about 44 kilobases in length, is composed of nine exon elements. The positions of the introns within the coding sequence are identical to those of the previously-characterized mouse *HPRT* gene, although there are significant differences between intron sizes for the two genes. *HPRT* minigenes have been used in a transient expression assay involving microinjection into HPRT$^-$ cells to demonstrate functional promoter activity within a 234-base-pair region upstream from the ATG codon. The promoter of this gene resembles those of other recently characterized "housekeeping" genes in that it lacks CAAT- and TATA-like sequences, but contains several copies of the sequence GGGCGG. Both RNase protection and primer extension analysis indicate that human *HPRT* mRNA is heterogeneous at the 5′ terminus, with transcription initiation occurring at sites located ≈104 to ≈169 base pairs upstream from the ATG codon. Comparison of the mouse and human *HPRT* 5′-flanking sequences indicates that there are only limited stretches of conserved sequence, although there are other shared features, such as an extremely high density of potential methylation sites, that may have functional significance.

Hypoxanthine-guanine phosphoribosyltransferase (HPRT; IMP:pyrophosphate phosphoribosyltransferase, EC 2.4.2.8) catalyzes the conversion of hypoxanthine and guanine to their respective 5′-mononucleotides and plays a crucial role in the metabolic salvage of purines in mammalian cells. Partial deficiency of the enzyme in human males causes gouty arthritis, whereas total deficiency leads to a severe neurological disorder, the Lesch-Nyhan syndrome (28). The functional *HPRT* gene, which is constitutively expressed, has been mapped to the X chromosome in both rodents (20) and humans (3). The *HPRT* locus has been a primary focus of studies on mutational mechanisms because of the ability to select for and against gene expression in cultured animal cells (32) and because of the predicted high level of new mutations at this locus in humans (24).

Earlier studies of spontaneous and induced *HPRT* mutants were restricted to genetic analyses and biochemical characterization of the protein (reviewed in references 9 and 12). Sequencing of altered peptides from mutant HPRT proteins of several patients with gout or Lesch-Nyhan syndrome has led to the identification of single amino acid changes leading to altered enzymatic activity (53). However, this approach is limited to the study of mutants that produce significant amounts of enzymatically active or immunologically cross-reactive proteins. Examination of mutational alteration at the nucleic acid level is not subject to these limitations; the use of cloned DNA sequences complementary to rodent and human *HPRT* mRNAs as probes in Southern hybridization experiments has allowed preliminary analysis of a diverse collection of mutations at this locus, including deletions and a partial internal duplication (57). Interpretation of these experiments relied upon the ability to discriminate fragments comprising the functional gene from cross-hybridizing autosomal sequences. Determination of the overall genomic organization of the human *HPRT* gene by hybridization of cDNA probes representing three different regions of the mRNA to X-derived fragments in Southern blot experiments (43) was also helpful in delineating the approximate positions of major alterations.

In this report, we describe the isolation and detailed characterization of cloned sequences representing the human *HPRT* gene. By restriction site mapping and direct sequencing of exon-intron boundaries, this gene was shown to be composed of nine exon elements dispersed over approximately 44 kilobases (kb). Structural comparisons between the human *HPRT* gene and the previously characterized mouse *HPRT* gene are presented. Among the interesting features of the *HPRT* genes is the lack of conventional promoter elements in the 5′-flanking sequences. We have localized functional promoter activity to sequences within 234 base pairs (bp) of the initiator methionine codon for the human gene by fusing 5′-flanking genomic sequences with a human *HPRT* cDNA to generate active minigenes. The data presented here will enable fine structure comparative analyses of mutant *HPRT* genes and will also be valuable in the study of the control of *HPRT* expression associated with X-chromosome inactivation.

MATERIALS AND METHODS

Enzymes and radioisotopes. Restriction endonucleases were obtained from either New England Biolabs or Bethesda Research Laboratories and used under conditions specified by the supplier. T4 polynucleotide kinase, bacterial alkaline phosphatase, and T4 DNA ligase were obtained from Bethesda Research Laboratories, Inc., DNA polymerase I was from New England Nuclear Corp., SP6 polymerase and RNasin were from Promega-Biotec, and reverse transcriptase was from Life Sciences, Inc. Isotopes, including [α-^{32}P]dCTP, [γ-^{32}P]ATP, [α-^{32}P]GTP, and [α-^{32}P]CTP, were obtained from Amersham Corp.

* Corresponding author.

Screening of genomic libraries. Genomic Charon 4A libraries of human DNA partially digested with *Hae*III and *Alu*I (30) or *Eco*RI (21) were kindly provided by Tom Maniatis (Harvard University). A third library constructed in Charon 4A, with DNA from a 49,XXXXY cell line (GM1202) partially digested with *Eco*RI, was generously supplied by Brian Crawford (Los Alamos National Laboratories). Phage were plated on bacterial lawns of $DP50_{supF}$, and those with homology to *HPRT* cDNA were identified by the plaque hybridization procedure of Benton and Davis (4). Labeled probes were prepared by nick translation (33) with [α-^{32}P]dCTP to a specific activity of 1×10^8 to 3×10^8 cpm/μg. cDNA probes used for screening genomic libraries were separated from vector DNA by digestion with appropriate restriction enzymes and electrophoresis in low-melting-point agarose gels followed by extraction of the DNA (33). Filters containing immobilized phage DNA were prehybridized overnight at 65°C in 6× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0) and 2× Denhardt solution (1× Denhardt solution is 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), hybridized overnight at 65°C with 5×10^5 cpm of ^{32}P-labeled probe per filter, and washed in 2× SSC–0.1% sodium dodecyl sulfate (SDS) at 65°C with shaking for 2 h with three changes. If necessary, filters were washed at 50°C in 0.5× SSC–0.1% SDS for 30 min to reduce background before exposure to X-ray film at −70°C with Dupont Cronex intensifying screens. Phage showing positive hybridization signals were subjected to at least three rounds of plaque purification. DNA was prepared from purified clones by the plate lysis method (33).

Gel transfer and hybridization. DNA fragments were transferred to nitrocellulose filters either by the method of Southern (49) or by the bidirectional transfer method of Smith and Summers (48). Filters were baked at 68°C in vacuo for 2 h. Hybridizations were performed at 65°C in 6× SSC or at 42°C in 6× SSC–30% formamide or 6× SSC–50% formamide depending upon the stringency desired. Dextran sulfate (10%), 100 μg of denatured, sonicated herring sperm DNA per ml, and ^{32}P-labeled probe (1×10^6 to 2×10^6 cpm of DNA per lane) were included in the hybridization solution. After hybridization, filters were rinsed free of formamide at room temperature in 2× SSC–0.1% SDS and then washed at 65°C in a similar solution for 1 to 2 h with three changes. Filters were reused after removal of the probe by incubation in 50% formamide–0.5× SSC–0.1% SDS at 65°C for ≥1 h.

Oligonucleotide hybridization. Approximately 100 ng of synthetic oligonucleotide (kindly provided by J. Habener, Massachusetts General Hospital) was labeled at the 5′ end with [γ-^{32}P]ATP (>3,000 Ci/mmol) to a final specific activity of 6×10^8 to 8×10^8 cpm/μg. Filters were prehybridized in 5× SSC–1× Denhardt solution–0.02% SDS–100 μg of herring sperm DNA per ml at room temperature for 4 h, after which 10^6 cpm of the 5′ end-labeled oligomer per ml was added. After hybridization for 48 h at room temperature, filters were subjected to two 10-min washes with 2× SSC–0.1% SDS at 37°C.

Subcloning of genomic fragments into plasmids. Restriction fragments from phage inserts were subcloned into plasmids by ligating a mixture of restriction fragments obtained by digestion of phage DNA with the appropriate restriction enzyme(s) to vector DNA linearized with the same enzyme(s). Alternatively, DNA fragments were purified from low-melting-point agarose gels and ligated to vector DNA that had been appropriately linearized. The 5′ termini of linear vector DNA (pUC8) were dephosphorylated with bacterial alkaline phosphatase before ligation. Ligations were performed by incubating at 12 to 14°C in 66 mM Tris (pH 7.6)–6 mM $MgCl_2$–10 mM dithiothreitol–0.4 mM ATP with 1 U of T4 DNA ligase per 20-μl reaction (33). Competent *Escherichia coli* JM83 cells were prepared and transformed by the method of Dagert and Ehrlich (16), except that L-ampicillin plates (100 by 15 mm) were spread evenly with 40 μl of 2% 5-bromo-4-chloro-3-indolyl-β-D-galactoside in dimethylformamide before plating of cells. Recombinant pUC8 plasmids were retransformed into RR1 cells to maximize the yield of plasmid DNA. Small-scale plasmid DNA preparations were performed as described by Birnboim and Doly (5), and in large-scale preparations lysozyme and Triton X-100 were used to lyse cells as described by Shepard and Polisky (46).

DNA sequencing. Plasmid DNA was purified of contaminating RNA by chromatography on Sepharose 4B columns. DNA fragments generated by cleavage with restriction enzymes were labeled at their 5′ ends with [α-^{32}P]ATP by using T4 polynucleotide kinase after treatment with bacterial alkaline phosphatase (33). Labeled fragments were isolated from low-melting-point agarose gels and sequenced as described by Maxam and Gilbert (35) on 8 and 25% acrylamide–urea gels.

RNase protection analysis. Total cellular RNA was isolated from HeLa or RJK859 (57) cells by guanidinium isothiocyanate extraction and centrifugation through a cesium chloride cushion (13). Polyadenylated RNA was prepared by oligo(dT) cellulose chromatography (2). To synthesize a riboprobe for RNase protection analysis, a 1.8-kb *Eco*RI-*Bam*HI fragment containing the first exon was cloned into the vector pSP64 (37). The DNA was digested with *Ava*II and transcribed with SP6 polymerase as described by Melton et al. (39), except that [α-^{32}P]GTP and [α-^{32}P]CTP were each used at 12 μM (114 Ci/mmol). The transcripts were monitored for size after glyoxalation by electrophoresis on formaldehyde-agarose gels since they migrated aberrantly on 5% acrylamide–urea gels (possibly due to secondary structure). A 40-μg sample of total RNA was mixed with 5×10^5 cpm of riboprobe in a volume of 30 μl of 80% formamide–40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)] (pH 6.7)–0.4 M NaCl–1 mM EDTA, denatured by boiling for 5 min, and incubated for 8 to 12 h at 65°C. After hybridization, 300 μl of 0.3 M NaCl–10 mM Tris (pH 7.5)–5 mM EDTA–RNase A (50 μg/ml)–RNase T_1 (6 μg/ml) was added to the mix and incubated at 30°C for 1 h. A 10-μl portion of 20% SDS and 50 μg of proteinase K was added, and the incubation was continued at 37°C for 15 min. After phenol extraction, the ^{32}P-RNA was precipitated with carrier tRNA and ethanol. The products were analyzed by electrophoresis on 5% acrylamide–7 M urea gels.

Primer extension analysis. A synthetic oligonucleotide complementary to nucleotide positions −4 to −53 in the genomic sequence shown in Fig. 2 (kindly provided by J. Habener, Massachusetts General Hospital) was labeled at the 5′ end with [γ-^{32}P]ATP (≈3,000 Ci/mmol) to a specific activity of 3×10^4 cpm/pmol. The oligomer was purified of unincorporated nucleotide on a DEAE-cellulose column, and 1.5 pmol of the primer was coprecipitated with 20 μg of polyadenylated RNA from HeLa cells. The mixture was suspended in 10 μl of a buffer containing 40 mM PIPES (pH 6.7)–0.4 M NaCl–1 mM EDTA, sealed in a glass capillary, denatured by incubation in boiling water for 5 min, and allowed to hybridize for 12 h at 65°C. The RNA was precipitated with ethanol and extended in 50 mM Tris (pH 8.3), 40 mM KCl, 10 mM $MgCl_2$, 28.4 mM β-mercaptoeth-

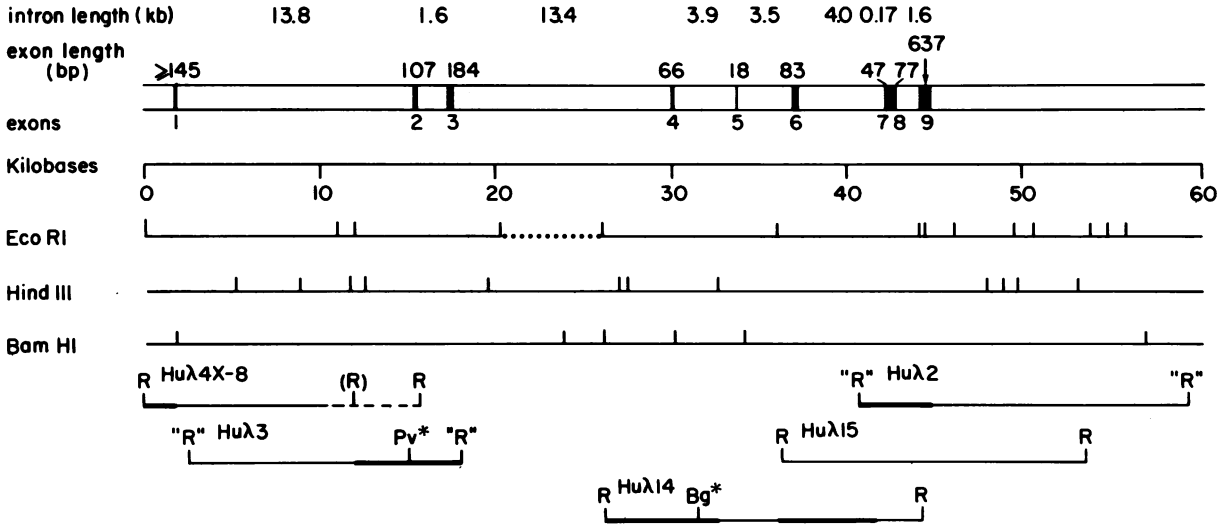

FIG. 1. Summary of human *HPRT* gene organization. The overall structure of the gene is represented by the box at the top. Exons 1 through 9 are indicated by the solid boxes. In the center, composite *Eco*RI, *Hin*dIII, and *Bam*HI maps of the locus, based on restriction maps of the representative lambda clones shown below, are diagrammed. The distance between Huλ3 and Huλ14 was estimated on the basis of the hybridization of probes from these phage to a genomic *Pvu*II-*Bgl*I fragment (Pv*, *Pvu*II; Bg*, *Bgl*I) as discussed in the text. The *Bam*HI and *Hin*dIII sites in the region between Huλ3 and Huλ14 were also similarly mapped; the dotted line in the *Eco*RI map represents an unmapped region for this enzyme. Natural and artificial *Eco*RI sites bordering phage inserts are represented as R and "R," respectively. The dashed line in Huλ4X-8 indicates sequences that are apparently rearranged during cloning; (R) indicates an artifactual *Eco*RI site in Huλ4X-8. The dark lines within phage inserts represent regions that were subcloned for localization of intron-exon boundaries, detailed maps of which are shown in Fig. 6.

anol, 1 mM each of dATP, dGTP, dCTP, and TTP, 25 U of RNasin, and 100 U of reverse transcriptase at 46°C for 1 h. After phenol extraction and ethanol precipitation, the products were analyzed by electrophoresis in 5% acrylamide–7 M urea gels.

Construction of *HPRT* minigenes. A previously constructed expression plasmid, pHPT44, which had 5′-flanking sequences of the mouse *HPRT* gene ligated to human *HPRT* cDNA followed by a polyadenylation signal from the hamster gene, was modified and used in the construction of *HPRT* minigenes (38). Plasmid DNA from pHPT44 was digested with *Eco*RI and *Xma*III to allow separation and removal of a 1-kb fragment, containing 850 bp of 5′-flanking sequence and 113 bp of the first exon of the mouse gene as well as 75 bp of the first exon from the human cDNA, from the remainder of the plasmid. The 1.8-kb *Eco*RI-*Bam*HI fragment from the 5′ genomic subclone pλ4X8-RB1.8 was partially digested with *Xma*III, and a 1.7-kb fragment that contained part of the first exon and ≈1.6 kb of 5′-flanking sequences was isolated. This was ligated to the 3.9-kb *Eco*RI-*Xma*III fragment of pHPT44 to generate the expression plasmid designated pHPT36.

To obtain derivatives of pHPT36 that had various lengths of 5′-flanking sequence, pHPT36 DNA was linearized at the unique *Sma*I site located approximately 930 bp upstream from the ATG codon. The DNA was then digested for times ranging from 2 to 10 min with the *Bal* 31 exonuclease, treated with the large fragment of DNA polymerase I to generate blunt ends, and ligated to phosphorylated *Eco*RI linkers (33). After digestion with *Eco*RI, excess linkers and fragments from the region between the *Sma*I site and the *Eco*RI site in pHPT36 (≤800 bp) were removed by electrophoresis in a low-melting-point agarose gel. The deleted derivatives (≤4.75 kb) were isolated, recircularized by ligation, and used to transform *E. coli* RR1 cells.

Microinjection assay for *HPRT* expression. Approximately 250 copies of the minigene pHPT36 and a deleted derivative, pHPT36-D2, were microinjected into the nuclei of 100 HPRT$^-$ hamster cells (RJK88) (22) that had been grown on cover slips for 2 days. The cells were then incubated in Dulbecco modified Eagle medium with 10% fetal calf serum at 37°C for 24 h. The medium was removed by aspiration, and the incubation continued for 16 h in fresh Dulbecco modified Eagle medium containing 10% fetal calf serum, 10^{-6} M aminopterin, 10^{-5} M thymidine, and 1 μCi of [^{3}H]hypoxanthine (2.93 Ci/mmol). The cells were fixed with 2% glutaraldehyde for 10 min and autoradiographed for 3 days at 4°C.

RESULTS

Isolation of *HPRT* recombinants from genomic libraries. Approximately 10^6 phage from each of two recombinant Charon 4A libraries of human genomic DNA were screened with a 1,200-bp *Msp*I fragment from a mouse *HPRT* cDNA clone, pHPT2 (7), or a 900-bp *Pst*I-*Hha*I fragment of a human *HPRT* cDNA clone, pHPT30 (8). A total of 15 λ phage clones were identified that gave positive hybridization signals. An additional eight recombinants were isolated by using the *Pst*I-*Hha*I fragment from pHPT30 to screen about 10^6 phage from a Charon 4A phage library constructed with DNA from a human cell line containing four X chromosomes.

These recombinant clones were initially categorized by restriction site mapping and by Southern analysis with cDNA probes. Correlation between *Eco*RI, *Hin*dIII, or *Bam*HI fragments that hybridized to the cDNA and genomic fragments identified previously as X linked or autosomal by comparative Southern blots of DNA from cells containing one or four X chromosomes (43) was used to identify recombinants containing sequences from the X-linked gene. Alternatively, in certain cases a repeat-free probe was iso-

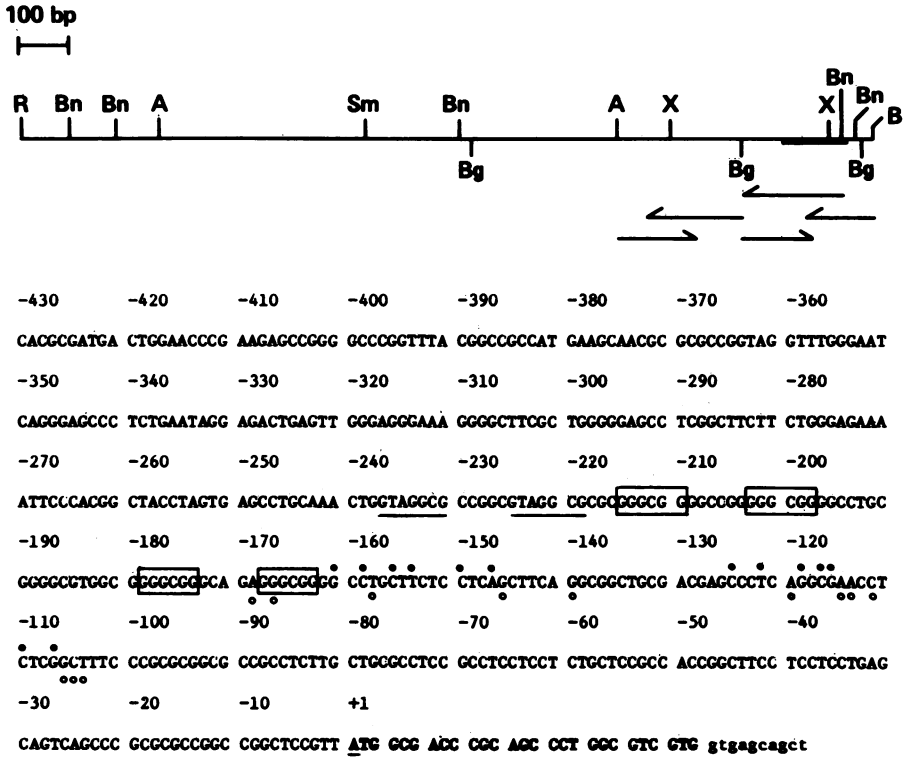

FIG. 2. Map of the 5′ end of the human *HPRT* gene. A restriction map of pλ4X8-RB1.8, containing exon I and 5′-flanking sequences of the human *HPRT* gene, is shown at the top. The strategy used to sequence this region is summarized below the map; all fragments were labeled at their 5′ ends, and the direction and approximate extent of the sequence determined for each fragment are indicated by the arrows. Restriction sites: A, *Ava*II; B, *Bam*HI; Bg, *Bgl*II; Bn, *Bst*NI; R, *Eco*RI; S, *Sma*I; X, *Xma*III. The DNA sequence of the 5′ end of the gene is shown in the lower part of the figure. Nucleotide residue +1 denotes the A of the ATG used to encode the initiator methionine codon of HPRT protein, and residues preceding it are indicated by negative numbers. The multiple transcription initiation sites predicted by RNase mapping and primer extension are indicated above (●) and below (○) the nucleotide, respectively. The sequence GGGCGG is enclosed by a box, and a 7-bp direct repeat sequence is underlined. The protein-coding sequence is in boldface type, and the first intron sequence is represented in lowercase letters.

lated from the cloned insert and examined for X linkage by hybridization to genomic DNA from cells containing one or four X chromosomes or to DNA from a hamster-human hybrid cell line carrying a single human X chromosome. These studies indicated that 2 of 8, 2 of 7, and 8 of 8 phage clones isolated from the *Hae*III-*Alu*I, *Eco*RI (normal), and *Eco*RI (four X chromosomes) libraries, respectively, represented parts of the functional X-linked gene. Restriction mapping data compiled from representative overlapping clones encompassing the functional *HPRT* gene are summarized in Fig. 1. Regions of the genomic phage clones that hybridized to the cDNA were subcloned into plasmids to enable fine structure mapping of these regions and identification of the individual exon elements on the basis of colinearity between regions of genomic DNA sequence and the published cDNA sequence (27).

Structural analysis of the 5′ end of the human *HPRT* gene. The 5′-most phage clone shown in Fig. 1, Huλ4X-8, was one of seven similar phage clones isolated from the library constructed from DNA of a 49,XXXXY cell line (GM1202). The size of the *Eco*RI fragment showing hybridization to a 5′ cDNA probe in these clones was variable, ranging from 5.0 to 11.8 kb, apparently as a result of deletions or rearrangement or both. Only sequences downstream from the *Bam*HI site appeared to be involved in these events, because the hybridizing *Eco*RI-*Bam*HI fragment (1.8 kb) was identical in each of the clones. A 600-bp repeat-free *Msp*I fragment located ≈80 bp 3′ to the *Bam*HI site hybridized to 11- and 11.8-kb *Eco*RI fragments from GM1202 genomic DNA and Huλ4X-8 DNA, respectively (data not shown). Other studies using repeat-free probes further 3′ of the *Bam*HI site allowed a rough delineation of the genomic sequences within Huλ4X-8 that showed unambiguous overlap with the apparently authentic genomic insert in Huλ3, the 5′ boundary of which fell just short of the first exon.

The 1.8-kb *Eco*RI-*Bam*HI fragment from Huλ4X-8 was subcloned into pUC8 (pλ4X8-RB1.8), and restriction endonuclease sites were mapped to develop a sequencing strategy (Fig. 2). The 3′ boundary of the first exon was found by sequencing to lie ≈40 bp upstream from the *Bam*HI site. The sequence of 430 bp upstream from the AUG codon is shown in Fig. 2. A numbering scheme is used in which +1 denotes the A of the AUG codon for the protein and upstream nucleotides in the genomic sequence are numbered with consecutive negative numbers. The 5′ end of the *HPRT* cDNA isolated by Jolly et al. (27) corresponds to position −85 in this sequence. Sequences normally associated with eucaryotic promoters such as TATA and CAAT boxes could not be identified in the immediate 5′ region. Instead, this region is highly G+C rich (91% in bases −170 to −235 in Fig. 2). A 7-bp direct repeat that is underlined in Fig. 2 was also noted.

To define the 5′ end of the mRNA, we used RNase protection and primer extension analysis. For the RNase

protection analysis, the 1.8-kb *Eco*RI-*Bam*HI fragment from Huλ4X-8 was cloned in the vector pSP64. In vitro-generated transcripts of 450 bp, which contained sequences complementary to the entire first exon of the *HPRT* mRNA, were hybridized to total RNA from HeLa cells and from lymphoblasts of a Lesch-Nyhan patient with a total *HPRT* gene deletion (RJK859) (57). A number of protected fragments ranging in size from 134 to 186 bp, which corresponded to transcripts that initiated 107 to 161 bp from the AUG codon, were seen with RNA from HeLa cells (Fig. 3). The absence of any protected fragments with RNA from RJK859 ruled out the possibility of artifactual protection of the riboprobe. In addition, since digestion of the hybrids with twofold lower or higher levels of RNase A and RNase T_1 than was used in the experiment shown in Fig. 3 did not have a significant effect on the overall pattern observed (data not shown), the multiplicity of bands does not appear to be an artifactual result of the digestion conditions.

To confirm these results, the 5′ end of the gene was mapped by primer extension analysis. A ^{32}P-labeled oligonucleotide primer that was complementary to nucleotides −4 to −53 (Fig. 2) was hybridized to 20 μg of polyadenylated RNA from the HeLa cells and extended with reverse transcriptase. The products of this reaction had lengths of 101 to 166 bp, corresponding to transcripts that initiated 104 to 169 bp from the AUG codon (Fig. 4). No extended products were detectable when 40 μg of total RNA from HeLa cells was used due to the low abundance of *HPRT* mRNA. Thus, both the RNase protection and primer extension analyses gave similar results, indicating that there are multiple initiation sites for human *HPRT* mRNA which appear to be clustered in several regions. Although exact correlation between the start sites predicted from the two analyses could not be demonstrated, this may be due, at least in part, to inherent differences in the generation and electrophoretic behavior of the DNA or RNA products of the two methods. In addition, contributory effects of secondary structure in the 5′ end of the mRNA, which could potentially affect the results from both types of analysis, cannot be completely ruled out.

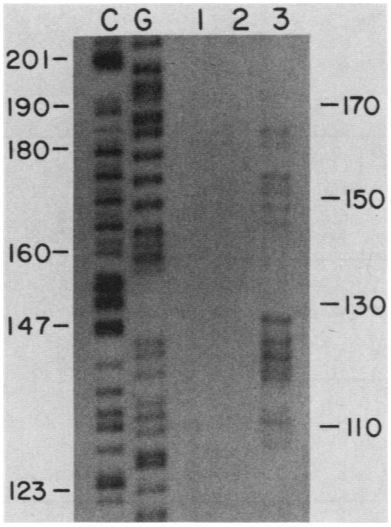

FIG. 3. Mapping of the 5′ end of the human *HPRT* gene by RNase protection analysis. Samples (40 μg each) of yeast tRNA (lane 1), total RNA from the HPRT$^-$ lymphoblast line RJK859 (lane 2), and total RNA from HeLa cells (lane 3) were hybridized at 65°C for 10 h with 5×10^5 cpm of uniformly labeled riboprobe to the first exon and flanking sequence synthesized as described in Materials and Methods. The RNA-RNA hybrids were digested with 40 μg of RNase A per ml and 6 μg of RNase T_1 per ml at 30°C, followed by digestion with proteinase K, phenol extraction, and electrophoresis of one-fifth of the reaction products on 5% acrylamide–7 M urea gels. The gel was dried and exposed to X-ray film for 14 h at −70°C with an enhancing screen. Lanes C and G represent a known DNA sequencing ladder employed as size standards. The numbers on the left represent migration of *Hpa*II fragments of pBR322 used as molecular weight markers. To allow comparison of these data with primer extension analysis, the numbers on the right indicate the calculated distances from the translation start site and thus correspond to the negative numbers used in the sequence shown in Fig. 2.

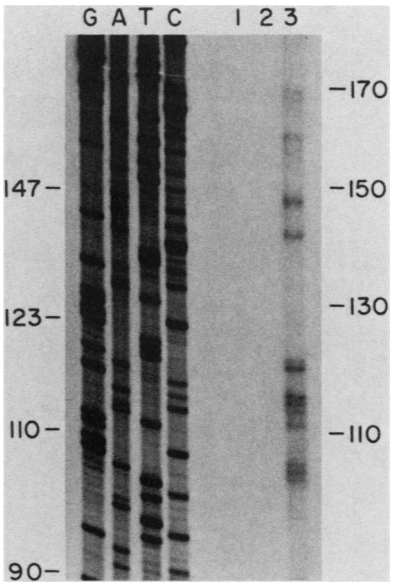

FIG. 4. Mapping of the 5′ end of the human *HPRT* gene by primer extension analysis. A synthetic oligonucleotide complementary to nucleotide positions −4 to −53 in the genomic sequence coding for the 5′ untranslated region of human *HPRT* mRNA was 5′ end labeled and hybridized to the RNA listed below as described in Materials and Methods. Primer-extended products were synthesized and characterized by electrophoresis in 5% acrylamide–7 M urea gels. Lanes: 1, 20 μg of yeast tRNA; 2, 40 μg of total RNA from HeLa cells; 3, 20 μg of polyadenylated RNA from HeLa cells; G, A, T, and C, known sequencing ladder used as size standards. The numbers on the left represent migration of *Hpa*II fragments of pBR322 used as molecular weight markers. The numbers on the right represent calculated distances from the translation start site and thus correspond to the negative numbers used in the sequence shown in Fig. 2.

Functional analysis of 5′-flanking sequences. For analysis of functional promoter activity in 5′-flanking sequences, minigenes were constructed (Fig. 5) and tested for expression. A previously constructed expression plasmid, pHPT44, that had the 5′-flanking sequence of the mouse gene ligated to the human cDNA and a polyadenylation signal from hamster *HPRT* cDNA was modified and used in this construction (38). The 5′-flanking sequence and approximately half of the first exon in the human cDNA was removed from pHPT44, and 1.6 kb of human 5′-flanking sequence was inserted in its place via a common *Xma*III site in the first exon of the human gene. This minigene, designated pHPT36, was introduced by $Ca(PO_4)_2$-mediated gene transfer (52) into RJK88, a hamster cell line with a

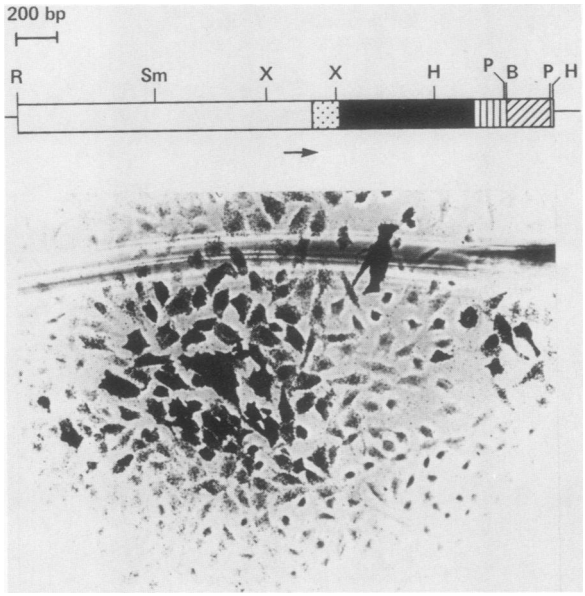

FIG. 5. Transient expression assay for functional HPRT activity. A restriction map of pHPT36, a human *HPRT* minigene constructed with the vector pUC9 for demonstration of functional promoter sequences, is shown at the top. Symbols: □, 5′ untranscribed sequence; ⊞, 5′ untranslated sequence; ■, coding sequence; ▥, 156 bp of 3′ untranslated sequence; ▨, fragment isolated from Chinese hamster HPRT cDNA (29) containing the polyadenylation signal AATAAA. Restriction sites: B, *Bam*HI; H, *Hin*dIII; P, *Pst*I; R, *Eco*RI; S, *Sma*I; X, *Xma*III. Shown in the lower part of the figure is an autoradiograph of representative HPRT$^-$ hamster cells (RJK88) after microinjection with pHPT36 and incubation in hypoxanthine-aminopterin-thymidine medium containing [^{3}H] hypoxanthine.

deletion of the entire *HPRT* gene (22) that has been used successfully in previous gene transfer experiments (8). HPRT$^+$ cells were produced at a frequency of $\simeq 1.7 \times 10^{-6}$ per μg of DNA, which, although relatively low, is at least 10-fold higher than the frequency obtained with cDNA constructs lacking the 5′-flanking region. Furthermore, upon microinjection into HPRT$^-$ hamster cells, this minigene effectively promoted the incorporation of [^{3}H]hypoxanthine into nucleic acids (Fig. 5); constructs lacking a functional promoter gave little or no incorporation by this autoradiographic analysis (data not shown). Thus, this minigene appeared to have sufficient sequence to allow production of functional HPRT protein. To further localize the promoter activity, deleted derivatives of pHPT36 containing different lengths of the 5′-flanking sequence were constructed and tested for expression in HPRT$^-$ hamster cells by the microinjection assay. A deletion mutant pHPT36-D2, containing 234 bp upstream from the ATG codon, functioned efficiently in this assay, whereas plasmid pHPT36-D6, containing 49 bp upstream from the ATG codon, was inactive (data not shown). Thus the G+C-rich region between the two endpoints contains the functional *HPRT* promoter, although further analysis will be required to pinpoint the element more precisely.

Determination of intron-exon boundaries. A 5.8-kb *Eco*RI fragment that hybridized to *HPRT* cDNA was identified in Huλ3, which overlapped the phage clone containing exon 1. Because the 3′ *Eco*RI site of this fragment was an artificial one introduced in the cloning procedure, direct correlation with a genomic *Eco*RI fragment was not possible. However, a 600-bp *Pst*I-*Eco*RI fragment from the 3′ end of this clone hybridized to a 8.3-kb fragment of *Eco*RI-digested genomic DNA; thus, the 5.8-kb *Eco*RI fragment in Huλ3 was derived from this larger genomic fragment (data not shown). This 5.8-kb fragment was subcloned into pUC8, and a partial restriction map of the recombinant plasmid, pλ3-R5.8, is shown in Fig. 6. A *Bgl*II site in this subclone divided the fragment into 3- and 2.8-kb fragments, both of which hybridized to the cDNA. The 3-kb *Eco*RI-*Bgl*II fragment had two internal *Msp*I sites, one of which generated a 280-bp *Msp*I-*Bgl*II fragment that hybridized to the cDNA. Determination of the nucleotide sequence of this small fragment by sequencing inward from the *Bgl*II and *Msp*I sites enabled the identification of a 107-bp exon (II) with the 3′ boundary located 13 bp upstream from the *Bgl*II site (Fig. 6). The sequence data also allowed orientation of the *Eco*RI fragment cloned in pλ3-R5.8.

A 1.3-kb *Pst*I fragment downstream from the *Bgl*II site in Huλ3, which also hybridized to the cDNA, was further subcloned (pλ3-P1.3) (data not shown). This 1.3-kb *Pst*I fragment had a single *Xho*I site that allowed separation of a hybridizing fragment from a nonhybridizing fragment. Since the cDNA also contains a unique *Xho*I site, pλ3-P1.3 was sequenced outward in both directions from this site; a 184-bp exon (III) was identified.

A region of genomic sequence between the third and

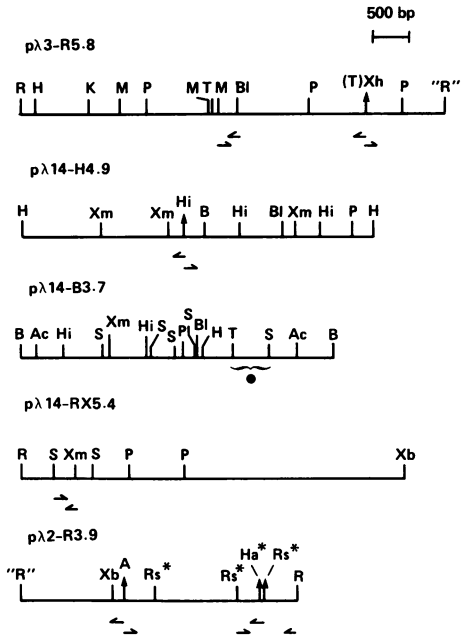

FIG. 6. Restriction maps and sequencing strategies for subclones containing exons 2 through 9. The arrows below the maps represent sequencing strategies used to locate exons. pλ3-R5.8 contains exons 2 and 3, pλ14-H4.9 contains exon 4, pλ14-B3.7 contains exon 5; pλ14-RX5.4 contains exon 6, and pλ2-R3.9 contains exons 7, 8, and 9. The closed circle below pλ14-B3.7 indicates the fragment containing exon 5, whose boundaries were not sequenced. Restriction sites: A, *Ava*II; Ac, *Acc*I; B, *Bam*HI; Bl, *Bgl*II; H, *Hin*dIII; Ha, *Hae*III; Hi, *Hin*cII; K, *Kpn*I; M, *Msp*I; P, *Pst*I; R, *Eco*RI, "R," artificial *Eco*RI site; S, *Sau*96I; T, *Taq*I; Xb, *Xba*I; Xh, *Xho*I; Xm, *Xmn*I. A bold arrow indicates a restriction site found within the cDNA sequence. An asterisk indicates that not all sites for a particular enzyme were mapped within the subclone.

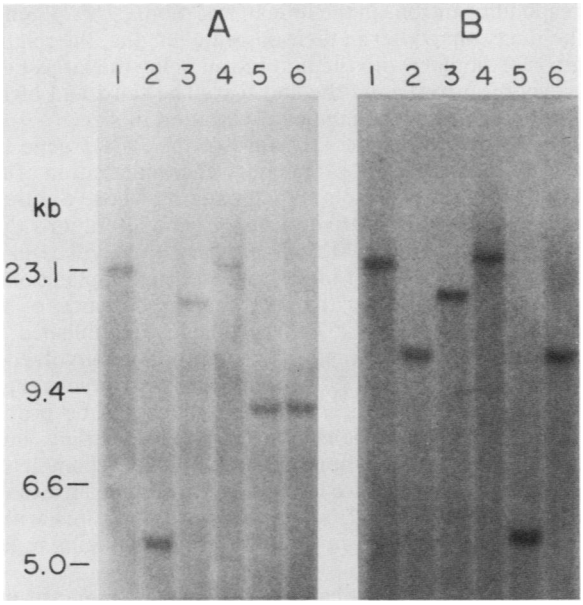

FIG. 7. Southern analysis of genomic DNA with probes from Huλ3 and Huλ14 to determine the size of the third intron. Genomic DNA was digested with *Pvu*II (lanes 1), *Pvu*II-*Eco*RI (lanes 2), *Bgl*I-*Pvu*II (lanes 3), *Bgl*I (lanes 4), *Bgl*I-*Eco*RI (lanes 5), and *Eco*RI (lanes 6), and Southern blots of these were probed with a 600-bp *Pst*I-*Eco*RI fragment from the 3′ end of the insert in Huλ3 (A). After autoradiography, the hybridized probe was removed from the blot as described in Materials and Methods, and the blot was probed with a 310-bp *Eco*RI-*Bam*HI fragment from the 5′ end of the insert in Huλ14 (B).

fourth exons was not present in any of the lambda recombinants isolated. The failure to isolate these sequences from three different genomic libraries is puzzling, although such anomalies have been reported in attempts to isolate other genes (20, 21, 34, 48, 51, 55) and have been attributed to the instability of certain sequences in the bacterial host used or underrepresentation in amplified libraries. The length of the third intron was estimated from Southern analysis of genomic DNA with probes from Huλ3 and Huλ14 (Fig. 7) and from the data shown in Fig. 1. A 600-bp *Pst*I-*Eco*RI fragment from the 3′ end of Huλ3 and a 320-bp *Eco*RI-*Bam*HI fragment from the 5′ end of Huλ14 both hybridized to a 17-kb genomic *Pvu*II-*Bgl*I fragment (Fig. 7). From the size of the genomic *Pvu*II-*Eco*RI fragment (5.3 kb) that hybridized to the probe from Huλ3 and the genomic *Eco*RI-*Bgl*I fragment (5.4 kb) that hybridized to the probe from Huλ14, it can be deduced that sequences spanning 8.3 kb between the third and fourth exons are not represented in any of the isolated recombinants.

A 4.9-kb *Hin*dIII fragment in Huλ14 corresponded to a fragment in genomic DNA that hybridized to a probe from the middle of the cDNA. This fragment was subcloned (pλ14-H4.9), and its restriction map is shown in Fig. 6. An internal *Hin*cII site that split a 500-bp *Xmn*I-*Bam*HI fragment into a hybridizing and a nonhybridizing fragment was examined for correspondence to the single *Hin*cII site in the cDNA. Sequencing outward from this *Hin*cII site in both directions led to the identification of exon IV with a size of 66 bp.

The fifth exon was localized in the phage clone Huλ14 by hybridization of an 18-nucleotide oligomer of human cDNA sequence corresponding to the fifth exon of the mouse gene.

TABLE 1. Exon-intron organization of the human HPRT gene

Intron no.	Sequence of exon-intron junctions[a]		Intron size (kb)	
	5′ boundary	3′ boundary	Human	Mouse
1	GGC GTC GTG gtgagcagct .....tcttttcag 27 Val (9)	ATT AGT GAT 28 Ile (10)	13.8	10.8
2	ATG GAC AG gtaagtaaga .....atttctgtag 134 Arg (45)	G ACT GAA 135 Arg (45)	1.6	2.9
3	AGC TAT TGT gtgagtatat .....ttttaactag 318 Cys (106)	AAT GAC CAG 319 Asn (107)	13.4	6.5
4[b]	ACT GGA AAG gtatatatct ..... 384 Lys (128)	AAT GTC TTG 385 Asn (129)	≈3.9	3.6
5[b]	ATT GTG GGA .....ttttgaaag 402 Glu (134)	GAT ATA ATT 403 Asp (135)	≈3.5	3.9
6	GTC GCA AG gtatgtatga .....taattaacag 485 Ser (162)	C TTG CTG 486 Ser (162)	4.0	3.9
7	CCA GAC T gtaagtgaat .....ttctttttag 532 Phe (178)	TT GTT GGA 533 Phe (178)	0.17	0.2
8	GAT TTG AAT gtaagtaatt .....tttttatag 609 Asn (203)	CAT GTT TGT 610 His (204)	1.6	0.6

[a] Nucleotide sequences of exon-intron boundaries were determined from the genomic subclones using the sequencing strategies shown in Fig. 2 and 6. Exon sequences are in uppercase letters, whereas intron sequences are in lowercase letters. The number shown immediately below the nucleotide sequence refers to the nucleotide position at which the intron interrupts *HPRT* mRNA, with position +1 in the numbering scheme being the A of the ATG codon used to initiate the protein. The identity and position of the amino acid in the protein sequence at which intron interruption occurs is indicated below the corresponding nucleotide positions.

[b] Intron where one of the boundaries was not determined.

Hybridization was detected to a 3.7-kb *Bam*HI fragment within Huλ14. This fragment was subcloned (pλ14-B3.7), and its restriction map is shown in Fig. 6. The fifth exon was further localized to a 500-bp *Taq*I-*Sau*96I fragment within this subclone. Although the boundaries of this exon have not been sequenced, it is reasonable to assume that this exceptionally short sequence is contained in a single exon that corresponds to the analogous exon of the mouse gene.

A 8.0-kb *Eco*RI fragment adjacent to the 10.5-kb *Eco*RI fragment in Huλ14 that contained exon 4 had a single *Xba*I site that split it into two fragments, both of which hybridized to the cDNA. The larger 5.4-kb *Eco*RI-*Xba*I fragment from Huλ14 was subcloned (pλ14-RX5.4), and restriction mapping and Southern analysis indicated that the region with homology to cDNA was between the *Eco*RI and *Xmn*I sites (Fig. 6). An *Sau*96I site within this *Eco*RI-*Xmn*I fragment divided it asymmetrically into a 450-bp *Eco*RI-*Sau*96I fragment and a 270-bp *Sau*96I-*Xmn*I fragment. When the latter fragment, which hybridized to the cDNA, was sequenced inward from both the *Sau*96I and *Xmn*I sites, the 3′ boundary of exon VI (83 bp), was found 20 bp from the *Xmn*I site.

A 3.9-kb *Eco*RI fragment at the 5′ end of the insert in Huλ2, a phage clone that overlapped the 3′ end of Huλ14 and contained the 2.6-kb *Xba*I-*Eco*RI that hybridized to the cDNA was subcloned (pλ2-R3.9) and mapped (Fig. 6). A single *Ava*II site in the insert in pλ2-R3.9 was examined for correspondence to the single *Ava*II site in the cDNA by sequencing outward from it in both directions; from these data exon VII with a size of 47 bp was identified. Furthermore, about 160 bp downstream from this exon, the eighth exon of the gene consisting of 77 bp was found.

The 3′ boundary of the last exon, including the polyadenylation signal of the gene, was localized approximately 20 bp from the 3′ *Eco*RI site in pλ2-R3.9 by sequencing upstream from this *Eco*RI site. Three hybridizing *Rsa*I fragments of 550, 460, and 310 bp were identified and examined by sequencing for the presence of the 5′ boundary of the last exon. The 310-bp *Rsa*I fragment was 5′ end labeled and separated into a 260-bp fragment and a 50-bp fragment by digesting with *Hae*III. Sequencing the 260-bp *Rsa*I-*Hae*III fragment led to the identification of the 5′ boundary of exon 9.

Positions and sizes of the exon elements for the human *HPRT* gene are given in Fig. 1. Sequences of the exon-intron boundaries are summarized in Table 1. The sizes of the introns in the human *HPRT* gene as well as intron sizes for the mouse *HPRT* gene are also listed for comparison.

DISCUSSION

The human *HPRT* gene is estimated to have a length of approximately 44 kb based on Southern hybridization analysis of genomic DNA and characterization of genomic λ clones. Both the human and mouse *HPRT* genes have nine exons, with the introns interrupting the protein-coding sequence at identical positions; the untranslated regions of *HPRT* mRNA are uninterrupted in the genome. The mouse and hamster genomes have an additional *HPRT*-like sequence, apparently a pseudogene, located on an autosome (22); there are four such nonfunctional *HPRT*-like sequences in the human genome; two have been mapped to chromosome 11, one has been mapped to chromosome 3, and one has been mapped to the region between p13 and q11 on chromosome 5 (43).

The splice junctions at the boundaries of each exon sequenced conform closely to the consensus sequences compiled from other eucaryotic genes (41). Although the corresponding introns in the human and mouse *HPRT* genes, unlike the exons, differ in their absolute lengths, the relative lengths are similar, with the first intron being the largest and the seventh intron being the shortest. The third and eighth introns show the most significant variation in size.

The information on the structure of the *HPRT* gene has been used to confirm the preliminary characterization of the molecular defects in Lesch-Nyhan patients whose Southern blot patterns were indicative of major gene alterations (57). One of the patients (GM1662) appears to have a major rearrangement in the first intron and a partial duplication of the coding and possibly the intervening sequence of the second and third exons (57; D. Konecki, unpublished results). Although it is intriguing that the regions involved in this DNA rearrangement fall within areas where recovery of authentic cloned genomic sequences proved to be problematic, any possible relationship between these findings must remain speculative until more precise mapping is completed. The data on gene structure have also allowed localization of two previously observed polymorphic *Bam*HI sites within the gene (42) to the first intron and to the 3′-flanking region of the gene.

Comparison of the 5′-flanking region of the mouse and human *HPRT* genes by a dot matrix analysis is shown in Fig. 8. Although the mRNAs generated from the two genes are highly homologous (>90%), there is limited sequence conservation in the 5′ regions. Further analysis of the promoters will indicate whether sequences that are conserved have functional significance. Both the human and mouse 5′-noncoding sequences are particularly rich in G and C bases. Approximately 80% of the 250 bp of sequence 5′ to the initiation codon is constituted of G+C, in contrast to ≈46% in the main body and ≈32% in the 3′-untranslated region of the genes.

The immediate 5′-flanking sequences of most eucaryotic

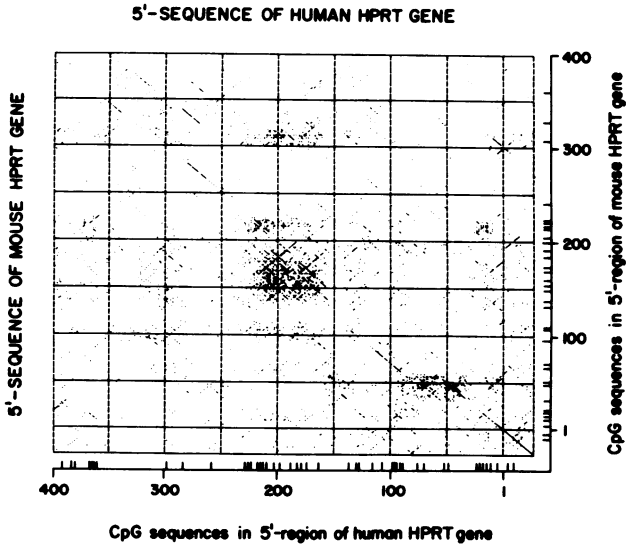

FIG. 8. Dot matrix comparison of the 5′ sequences of human and mouse *HPRT* genes. The 5′ sequences, including the entire first exon and 400 bp of sequence 5′ to the AUG codon of the human and mouse *HPRT* genes, are represented along the horizontal and vertical axes, respectively. Position 1 marks the A of the AUG codon used to initiate the protein. A dot on the matrix represents a 9-for-10 nucleotide match. Potential methylation sites (CpG sequences) within these sequences are also indicated along the axes.

mRNA-coding genes contain two consensus sequences that play a regulatory role in transcription, namely, a TATA box located 20 to 30 bp upstream from the mRNA cap site (6) and a CAAT sequence frequently located 70 to 80 bp upstream. These sequences are not found within a 600-bp 5'-flanking region of the human and mouse *HPRT* genes, even though functional promoter activity has been demonstrated (38). Other genes transcribed by RNA polymerase II that lack TATA-like sequences in their promoters include simian virus 40 late genes (19), the hepatitis virus surface antigen gene (10), the adenovirus IVa2 gene (31), the human epidermal growth factor receptor gene (26), and four other recently characterized eucaryotic housekeeping genes, the 3-hydroxy-3-methylglutaryl coenzyme A reductase (44, 45), adenosine deaminase (50), dihydrofolate reductase (11, 15, 36, 56), and phosphoglycerate kinase (47) genes.

The 5' termini of human *HPRT* mRNAs appear to be quite heterogeneous, with little or no obvious preference for a particular initiation point; similar results have been published recently for mRNA encoding the human epidermal growth factor receptor protein (26). The other housekeeping genes mentioned above also appear to produce mRNAs with 5' heterogeneity, although in general one or a few preferred sites of initiation were detected. The promoter regions of these genes, like the *HPRT* gene, contain several copies of the sequence 5'-CCGCCC-3' or its inverted complement 5'-GGGCGG-3'. These sequences are also found in several viral promoters, including that of the simian virus 40 early and late genes (25) and the herpesvirus thymidine kinase gene (37). In the latter case, these sequences serve as flexible distal transcriptional signals in which the distance from the cap site, orientation, and number of copies can vary, although changes in each of these parameters do influence the magnitude of the distal signal effect. More detailed functional characterization of the promoters is needed to determine how these sequences influence gene expression. It is possible that they represent binding sites for promoter-specific transcription factors, since factors that allow discrimination between simian virus 40 promoters by RNA polymerase II have been described (17, 18).

Methylation of cytosine residues has frequently been implicated in the control of gene expression and has been suggested to play a crucial role in the maintenance of X-chromosome inactivation. The strongest evidence for this is the ability of the potent methylation inhibitor 5-azacytidine to reactivate *HPRT* and other genes on an inactive X chromosome in somatic cell hybrids (40). With cloned fragments from the 5' region of the human *HPRT* gene and methylation-sensitive restriction enzymes, the methylation patterns of active, inactive, and reactivated human *HPRT* genes have been examined (54, 58). These studies have shown that there is hypomethylation of 5' clustered sites in the active *HPRT* gene. In contrast, methylation of inactive *HPRT* alleles is nonuniform, and there is less extensive hypomethylation of the 5' cluster. It has therefore been suggested that the overall pattern of methylation plays a role in determining transcription. As shown in Fig. 8, there are 46 and 35 CpG sequences within 400 bases upstream of the translation start sites of the human and mouse *HPRT* genes, respectively. The abundance and clustering of these CpG sequences, each of which is a potential substrate for methylating enzymes, illustrates the problems inherent in the feasibility and analysis of the studies mentioned above with cloned *HPRT* probes. Complete studies on the correlation of critical methylation at these sites with X inactivation will require direct examination of methylated cytosine residues from active and inactive X chromosomes with a technique such as genomic sequencing (14).

Three other X-linked genes have been recently characterized, namely, the genes for factor VIII (23), factor IX (1), and phosphoglycerate kinase (47). These genes and the *HPRT* gene appear to lack any shared distinguishing features, especially in the promoter region, which might explain their coordinate control as a result of X-chromosome inactivation. The promoters of the genes for factor VIII and IX, which appear to be tissue specific, differ from the phosphoglycerate kinase and HPRT promoters in having lower G+C content (30 to 40%) and very few potential methylation sites. Detailed examination of the methylation patterns of the factor VIII and IX genes from active and inactive X chromosomes will be required to determine the relevance of methylation to the inactivation of these genes.

The information presented here will further enhance the *HPRT* system as a model for the study of spontaneous and induced mutational mechanisms at the molecular level and for the evaluation of the mutagenic potential of various agents. The generation of genomic probes that are relatively free of repetitive sequences from the introns of the gene should be useful in identifying restriction fragment-length polymorphisms that could serve as additional haplotype markers for the locus and allow more detailed studies on methylation patterns and nuclease sensitivity in different regions of *HPRT* genes from active and inactive X chromosomes.

ACKNOWLEDGMENTS

We thank Twee Tsao for performing the microinjection experiment and Barbara Goldmar for typing the manuscript.

P.I.P. was supported by fellowships from the Robert A. Welch Foundation and the Arthritis Foundation. This work was supported by the Howard Hughes Medical Institute, Public Health Service grant AM31428 to C.T.C. from the National Institutes of Health, and American Cancer Society grant CD-171 to A.C.C.

LITERATURE CITED

1. **Anson, D. S., K. H. Choo, D. J. G. Rees, F. Giannelli, K. Gould, J. A. Huddleston, and G. G. Brownlee.** 1984. The gene structure of human anti-haemophilic factor IX. EMBO J. **3**:1053–1060.
2. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
3. **Becker, M. A., R. C. K. Yen, P. Itkin, S. J. Goss, J. E. Seegmiller, and B. Bakay.** 1979. Regional localization of the gene for human phosphoribosylphosphate synthetase on the X-chromosome. Science **203**:1016–1019.
4. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
5. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
6. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
7. **Brennand, J., A. C. Chinault, D. S. Konecki, D. W. Melton, and C. T. Caskey.** 1982. Cloned cDNA sequences of the hypoxanthine/guanine phosphoribosyltransferase gene from a mouse neuroblastoma cell line found to have amplified genomic sequences. Proc. Natl. Acad. Sci. USA **79**:1950–1954.
8. **Brennand, J., D. S. Konecki, and C. T. Caskey.** 1983. Expression of human and Chinese hamster hypoxanthine-guanine phosphoribosyltransferase cDNA recombinants in cultured

Lesch-Nyhan and Chinese hamster fibroblasts. J. Biol. Chem. **258**:9593–9596.
9. **Caskey, C. T., and G. Kruh.** 1979. The HPRT locus. Cell **16**:1–9.
10. **Cattaneo, R., H. Will, N. Hernandez, and H. Schaller.** 1983. Signals regulating hepatitis B surface antigen transcription. Nature (London) **305**:336–338.
11. **Chen, M.-J., T. Shimada, A. D. Moulton, A. Cline, R. K. Humphries, J. Maizel, and A. W. Nienhuis.** 1984. The functional human dihydrofolate reductase gene. J. Biol. Chem. **259**:3933–3943.
12. **Chinault, A. C., and C. T. Caskey.** 1984. The hypoxanthine-guanine phosphoribosyltransferase gene: A model for the study of mutation in mammalian cells. Prog. Nucleic Acid Res. Mol. Biol. **31**:295–313.
13. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
14. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
15. **Crouse, G., E. J. Leys, R. N. McEwan, E. G. Frayne, and R. E. Kellems.** 1985. Analysis of the mouse *dhfr* promoter region: existence of a divergently transcribed gene. Mol. Cell. Biol. **5**:1847–1858.
16. **Dagert, M., and S. D. Ehrlich.** 1979. Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. Gene **6**:23–28.
17. **Dynan, W. S., and R. Tjian.** 1983. The promoter-specific transcription factor SP1 binds to upstream sequences in the SV40 early promoter. Cell **35**:79–87.
18. **Dynan, W. S., and R. Tjian.** 1983. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell **32**:669–680.
19. **Elder, J. T., R. A. Spritz, and A. M. Weissman.** 1981. Simian virus 40 as a eukaryotic cloning vehicle. Annu. Rev. Genet. **15**:295–340.
20. **Francke, U., and R. T. Taggart.** 1980. Comparative gene mapping: Order of loci on the X-chromosome is different in mice and humans. Proc. Natl. Acad. Sci. USA **77**:3595–3599.
21. **Fritsch, E. F., R. M. Lawn, and T. Maniatis.** 1980. Molecular cloning and characterization of the human β-like globin gene cluster. Cell **19**:959–972.
22. **Fuscoe, J. C., R. G. Fenwick, Jr., D. H. Ledbetter, and C. T. Caskey.** 1983. Deletion and amplification of the HGPRT locus in Chinese hamster cells. Mol. Cell. Biol. **3**:1086–1096.
23. **Gitschier, J., W. I. Wood, T. M. Goralka, K. L. Wion, E. Y. Chen, D. H. Eaton, G. A. Vehar, D. J. Capon, and R. M. Lawn.** 1984. Characterization of the human factor VIII gene. Nature (London) **312**:326–330.
24. **Haldane, J. B. S.** 1935. The rate of spontaneous mutation of a human gene. J. Genet. **31**:317–326.
25. **Hansen, U., and P. A. Sharp.** 1983. Sequences controlling *in vitro* transcription of SV40 promoters. EMBO J. **2**:2293–2303.
26. **Ishii, S., X. Young-Hua, R. H. Stratton, B. A. Roe, G. T. Merling, and I. Pastan.** 1985. Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. Proc. Natl. Acad. Sci. USA **80**:477–481.
27. **Jolly, D. J., H. Okayama, P. Berg, A. C. Esty, D. Filpula, P. Bohlen, G. G. Johnson, J. E. Shively, T. Hunkapillar, and T. Friedmann.** 1983. Isolation and characterization of a full-length expressible cDNA for human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. USA **80**:477–481.
28. **Kelley, W. N., and J. B. Wyngaarden.** 1983. Clinical syndromes associated with hypoxanthine-guanine phosphoribosyltransferase deficiency, p. 1115–1143. *In* Stanbury, J. B., J. B. Wyngaarden, D. S. Fredrickson, J. L. Goldstein, and M. S. Brown (ed.), The metabolic basis of inherited disease, 5th ed. McGraw-Hill Book Co., New York.
29. **Konecki, D. S., J. Brennand, J. C. Fuscoe, C. T. Caskey, and A. C. Chinault.** 1982. Hypoxanthine-guanine phosphoribosyltransferase genes of mouse and Chinese hamster: construction and sequence analysis of cDNA recombinants. Nucleic Acids Res. **10**:6763–6775.
30. **Lawn, R. M., E. F. Fritsch, R. C. Parker, G. Blake, and T. Maniatis.** 1978. The isolation and characterization of linked α- and β-globin genes from a cloned library of human DNA. Cell **15**:1157–1174.
31. **Lee, D. C., and R. G. Roeder.** 1981. Transcription of adenovirus type 2 genes in a cell-free system: apparent heterogeneity of initiation of some promoters. Mol. Cell. Biol. **1**:635–651.
32. **Littlefield, J.** 1964. Selection of hybrids from matings of fibroblasts *in vitro* and their presumed recombinants. Science **145**:709–710.
33. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **May, F. E. B., R. Weber, and B. R. Westley.** 1982. Isolation and characterization of *Xenopus laevis* albumin genes: loss of 74K albumin gene sequences by library amplification. Nucleic Acids Res. **10**:2791–2807.
35. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
36. **McGrogan, M., C. C. Simonsen, D. T. Smouse, P. J. Farnham, and R. T. Schimke.** 1985. Heterogeneity at the 5′-termini of mouse dihydrofolate reductase mRNAs. Evidence for multiple promoter regions. J. Biol. Chem. **260**:2307–2314.
37. **McKnight, S. L., R. C. Kingsbury, A. Spence, and M. Smith.** 1984. The distal transcription signals of the herpes virus *tk* gene share a common hexanucleotide control sequence. Cell **37**:253–262.
38. **Melton, D. W., D. S. Konecki, J. Brennand, and C. T. Caskey.** 1984. Structure, expression and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA **81**:2147–2151.
39. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
40. **Mohandas, T., and L. J. Shapiro.** 1983. Factors involved in X-chromosome inactivation, p. 271–297. *In* A. A. Sandberg (ed.), Cytogenetics of the mammalian X-chromosome, part A: basic mechanisms of X-chromosome behavior. Alan R. Liss, Inc., New York.
41. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
42. **Nussbaum, R. L., W. E. Crowder, W. L. Nyhan, and C. T. Caskey.** 1983. A three-allele restriction-fragment-length polymorphism at the hypoxanthine phosphoribosyltransferase locus in man. Proc. Natl. Acad. Sci. USA **80**:4035–4039.
43. **Patel, P. I., R. L. Nussbaum, P. E. Framson, D. Ledbetter, C. T. Caskey, and A. C. Chinault.** 1984. Organization of the *HPRT* gene and related sequences in the human genome. Somat. Cell Mol. Genet. **10**:483–493.
44. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMGCoA reductase: a negatively regulated gene with unusual promoter and 5′ untranslated regions. Cell **38**:275–285.
45. **Reynolds, G. A., J. L. Goldstein, and M. S. Brown.** 1985. Multiple mRNAs for 3-hydroxy-3-methylglutaryl coenzyme A reductase determined by multiple transcription initiation sites and intron splicing sites in the 5′-untranslated region. J. Biol. Chem. **260**:10369–10377.
46. **Shepard, H. M., and B. Polisky.** 1979. Measurement of plasmid copy number. Methods Enzymol. **68**:503–573.
47. **Singer-Sam, J., D. H. Keith, K. Tani, R. L. Simmer, L. Shively, S. Lindsay, A. Yoshida, and A. D. Riggs.** 1984. Sequence of the promoter region of the gene for human X-linked 3-phosphoglycerate kinase. Gene **32**:409–417.
48. **Smith, G. E., and M. D. Summers.** 1980. The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzyloxymethyl-paper. Anal. Biochem. **109**:123–129.
49. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

50. **Valerio, D., M. G. C. Duyvesteyn, B. M. M. Dekker, G. Weeda, Th. M. Berkvens, L. van der Voorn, H. van Ormondt, and A. J. van der Eb.** 1985. Adenosine deaminase: characterization and expression of a gene with a remarkable promoter. EMBO J. **4**:437–443.
51. **Vizard, D. L., and J. Yarsa.** 1985. Comparison of genomic fragment and clone sequences within a long interspersed repeated sequence of the mouse genome. Nucleic Acids Res. **13**:473–484.
52. **Wigler, M., A. Pellicer, S. Silverstein, and R. Axel.** 1978. Biochemical transfer of single-copy eukaryotic genes using total cellular DNA as donor. Cell **14**:725–731.
53. **Wilson, J. M., A. B. Young, and W. N. Kelley.** 1983. Hypoxanthine-guanine phosphoribosyltransferase deficiency. The molecular basis of the clinical syndrome. N. Engl. J. Med. **309**:900–910.
54. **Wolf, S. F., D. J. Jolly, K. D. Lunnen, T. Friedmann, and B. R. Migeon.** 1984. Methylation of the hypoxanthine phosphoribosyltransferase locus on the X-chromosome: implications for X-chromosome inactivation. Proc. Natl. Acad. Sci. USA **81**:2806–2810.
55. **Wyman, A. R., L. B. Wolfe, and D. Botstein.** 1985. Propagation of some human DNA sequences in bacteriophage λ vectors requires mutant *Escherichia coli* strains. Proc. Natl. Acad. Sci. USA **82**:2880–2884.
56. **Yang, J. K., J. N. Masters, and G. Attardi.** 1984. Human dihydrofolate reductase gene organization. Extensive conservation of the G+C-rich 5′ non-coding sequence and strong intron size divergence from homologous mammalian genes. J. Mol. Biol. **176**:169–187.
57. **Yang, T. P., P. I. Patel, A. C. Chinault, J. T. Stout, L. G. Jackson, B. M. Hildebrand, and C. T. Caskey.** 1984. Molecular evidence for new mutation at the *HPRT* locus in Lesch-Nyhan patients. Nature (London) **310**:412–414.
58. **Yen, P. H., P. I. Patel, A. C. Chinault, T. Mohandas, and L. J. Shapiro.** 1984. Altered methylation of *HPRT* genes on active and inactive human X-chromosomes. Proc. Natl. Acad. Sci. USA **81**:1759–1763.