MOLECULAR AND CELLULAR BIOLOGY, Aug. 1993, p. 4784–4792
0270-7306/93/084784-09$02.00/0

Vol. 13, No. 8

Coexpression of Two Closely Linked Avian Genes for Purine Nucleotide Synthesis from a Bidirectional Promoter†

ANTHONY GAVALAS,[1] JACK E. DIXON,[2] KELLY A. BRAYTON,[1] AND HOWARD ZALKIN[1]*

Department of Biochemistry, Purdue University, West Lafayette, Indiana 47907-1153,[1] and Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109-0606[2]

Received 12 March 1993/Returned for modification 21 April 1993/Accepted 7 May 1993

Two avian genes encoding essential steps in the purine nucleotide biosynthetic pathway are transcribed divergently from a bidirectional promoter element. The bidirectional promoter, embedded in a CpG island, directs coexpression of *GPAT* and *AIRC* genes from distinct transcriptional start sites 229 bp apart. The bidirectional promoter can be divided in half, with each half retaining partial activity towards the cognate gene. *GPAT* and *AIRC* genes encode the enzymes that catalyze step 1 and steps 6 plus 7, respectively, in the de novo purine biosynthetic pathway. This is the first report of genes coding for structurally unrelated enzymes of the same pathway that are tightly linked and transcribed divergently from a bidirectional promoter. This arrangement has the potential to provide for regulated coexpression comparable to that in a prokaryotic operon.

De novo synthesis of purine nucleotides occurs by a 14-step branched pathway via IMP. The pathway appears to be invariant in all organisms known to synthesize purine nucleotides. Despite the central importance of the pathway, only for *Escherichia coli* is there an overall understanding of gene organization, gene regulation, and enzyme regulation (51). Recently, most of the cDNAs for vertebrate enzymes have been cloned and characterized (reference 51 and references therein), setting the stage for investigations of gene expression.

Chinese hamster ovary (CHO) cell mutants corresponding to each of the loci required for synthesis of AMP have been isolated, and human chromosome assignments were derived from cytogenetic analysis of CHO-human somatic cell hybrids (summarized in reference 1). With one exception, genes for the 12 steps to AMP synthesis are on individual human chromosomes. However, *GPAT* and *AIRC*, genes encoding enzymes for step 1 and steps 6 plus 7, respectively, were both assigned to the q arm of chromosome 4, and the possibility of physical linkage was considered (1).

Characterization of genes of the pathway is limited to the *Gart* locus. The *Gart* locus in *Drosophila melanogaster* consists of two genes. A purine gene encodes a trifunctional enzyme for de novo purine nucleotide synthesis containing activities for glycinamide ribonucleotide synthetase (EC 6.3.4.13), aminoimidazole ribonucleotide (AIR) synthetase (EC 6.3.3.1), and glycinamide ribonucleotide transformylase (EC 2.1.2.2) (21). A cuticle protein gene is encoded on the opposite strand, entirely within an intron of the purine gene (20). The human *GART* gene has been isolated from chromosome 21 on a yeast artificial chromosome and shown to function in CHO cells (17).

There is at present no information about regulation of genes for de novo purine nucleotide synthesis in vertebrates or invertebrates. To initiate studies of gene regulation, it was our objective to clone the avian *GPAT* gene encoding glutamine phosphoribosylpyrophosphate (PRPP) amidotransferase (EC 2.4.2.14). This enzyme catalyzes step 1 of the presumed rate-limiting reaction in the de novo pathway. *GPAT* is thus a likely target for gene regulation. The cDNA for the avian glutamine PRPP amidotransferase (52) was used in the present work to isolate the *GPAT* gene. Here we report that the *AIRC* gene flanks the 5′ end of *GPAT*. The 44-kb *GPAT* and 8-kb *AIRC* genes are closely linked and are divergently transcribed from a bidirectional promoter. This is the first example of divergent transcription of tightly linked genes from a bidirectional promoter in which the two enzyme products catalyze different steps in the same pathway. This arrangement has the potential to provide for regulated coexpression of functionally related genes comparable to that in a prokaryotic operon. To describe this unit of genetic function, we suggest the term dioskourion, taken from the Greek Dioskouri, the mythological inseparable twin sons of Zeus.

MATERIALS AND METHODS

Library screening. An EMBL4 chicken oviduct genomic library (provided by Ming-Jer Tsai, Baylor College of Medicine) was plated at a density of 25,000 plaques per 100-mm-diameter dish on Luria-Bertani agar supplemented with 50 μg of thymidine per ml and 10 mM $MgSO_4$ with *E. coli* LE392 as the host. A total of 500,000 plaques was screened with each probe. The plaques were lifted on nitrocellulose filters (Schleicher & Schuell) as described previously (47). Prehybridization was carried out at 42°C in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–1× Denhardt's solution (0.2% Ficoll, 0.2% polyvinylpyrrolidone, 0.2% bovine serum albumin)–0.5% sodium dodecyl sulfate (SDS)–0.1 mg of salmon sperm DNA per ml–50% formamide for 4 h. Hybridization was carried out at 42°C for 16 h in prehybridization solution with SDS omitted and containing 2×10^6 cpm of radiolabeled probe per ml (13). Washes were carried out at room temperature in 2× SSC–0.1% SDS and then in 2× SSC–0.1% SDS at 65°C. The filters were autoradiographed at −80°C with an intensifying screen. Positive

* Corresponding author.

† Journal paper 13794 from the Purdue University Agricultural Experiment Station.

plaques were picked for second and third rounds of screening until plaque pure. Plaque-pure phages were propagated in 5-ml cultures. A rapid DNA isolation procedure (44) was used to extract DNA for restriction and Southern hybridization analyses.

Plasmid constructions. Genomic subclones for restriction mapping and sequencing were made in pBluescript SK. The bireporter plasmid pLUC/CAT-1 (see Fig. 4) was assembled from a derivative of pA3LUC (50) and pCAT-CON (Promega). A 1,350-bp *Sma*I genomic subclone (see Fig. 3) was digested with *Bgl*I, blunted, and digested with *Not*I, and the resulting 600-bp fragment carrying the promoter was transferred into *Eco*RV-*Not*I-digested pBluescript SK$^-$. From that construct, a 360-bp *Kpn*I-blunted *Hin*dIII fragment containing the promoter and 60 bp from each 5′ untranslated region (UTR) was cloned in pLUC/CAT-1, giving rise to construct pLC2. Construct pLC4 was made similarly, by subcloning a 380-bp *Kpn*I-blunted *Sal*I fragment into pLUC/CAT-1. Plasmids pCAT-AT and pCAT-C were made by cutting pLC2 and pLC4 plasmids, respectively, with *Ava*I and *Bam*HI, filling in, and religating. Plasmid pCAT-Δ, which was used as a promoterless background control, was prepared similarly by digesting pLUC/CAT-1 with *Xho*I and *Bam*HI, filling in, and religating. Plasmid pSK-PRO1 was constructed by isolating a 600-bp *Pvu*II-*Nar*I restriction fragment from pLC2 and cloning it into the *Sma*I-*Cla*I sites of the pBluescript SK$^-$ polylinker.

RNA preparations and Northern hybridizations. Total RNA preparations from chicken liver and HepG2 cells were done essentially by acid-guanidinium thiocyanate-phenol-chloroform extraction (5). Poly(A)$^+$ RNA was prepared by oligo(dT)-cellulose chromatography (Pharmacia). Northern (RNA) blots were prepared by running the RNA on formaldehyde–1.4% agarose gels and blotting the gel on GeneScreen Plus membranes (Dupont) by capillary transfer according to the manufacturer's instructions. Prehybridization was carried out at 65°C in 1% SDS–1 M NaCl–10% dextran sulfate. For hybridization, salmon sperm DNA at 100 μg/ml and 2 × 10^5 cpm of radiolabeled probe per ml were added and the bag was incubated for at least 12 h at 65°C with constant agitation. The filters then were washed with 2× SSC at room temperature and with 2× SSC–1% SDS at 65°C and autoradiographed for 12 to 48 h at 80°C with an intensifying screen on Kodak X-Omat AR film. RNA molecular weight markers were bought from Promega.

Genomic PCR. A typical 100-μl polymerase chain reaction (PCR) mixture contained 10 μl of PCR buffer (100 mM Tris-HCl [pH 8.3], 500 mM KCl, 15 mM $MgCl_2$, and 0.1% [wt/vol] gelatin), 8 μl of a deoxynucleoside triphosphate stock solution (2.5 mM each), 8 μl of each 30-mer (at 30 ng/μl), 2.5 U of *Taq* DNA polymerase, and 100 ng of chicken genomic DNA. Two cycles were then performed at 94°C for 1 min, 45°C for 2 min, and 72°C for 2 min and were followed by 35 cycles at 92°C for 1 min, 55°C for 2 min, and 72°C for 2 min with a 3-s extension on each cycle. The product was run on an agarose gel, and the unique fragment was isolated with the GeneClean kit (Bio 101). The following primers were used to isolate PCR-AT1: 5′ primer, AAG GGG AAG TGA GCT CAC TGC CAA TAA AAC; 3′ primer, AGT CAG CAG TGT CGA CAT TTT CTA GAG GAG.

Southern hybridizations. For genomic Southern blots, 5 μg of chicken genomic DNA was digested with appropriate restriction enzymes, run on a 0.8% agarose gel, and transferred to a GeneScreen Plus membrane (Dupont). Prehybridization and hybridization were carried out essentially as described for the Northern hybridizations. The Kodak X-Omat film was exposed for 1 to 2 days at −80°C with an intensifying screen. Southern blots of genomic clones and subclones were done on nylon filters (Schleicher & Schuell), and hybridizations were carried out according to the manufacturer's instructions.

DNA sequencing. Genomic subclones were sequenced by the dideoxy sequencing method (40) with ExoIII-deleted clones (19) or with primers derived from the cDNA sequence.

Transient transfections and reporter assays. Transient transfections were done by the lipofection method (14). HepG2 cells were grown to 60% confluency in 35-mm-diameter plates, washed with phosphate-buffered saline (PBS), and incubated with 1 ml of transfection reagent containing 20 μl of Lipofectin reagent (Bethesda Research Laboratories, Inc.), 8 μg of the plasmid to be assayed, 2 μg of RSV-lacZ plasmid (11), and Eagle minimal essential medium for 8 to 12 h at 37°C in 5% CO_2. The transfection reagent was then removed, and cells were fed with 2 ml of Eagle minimal essential medium supplemented with 10% fetal bovine serum and incubated at 37°C in 5% CO_2 for 48 to 60 h. Then the cells were washed with PBS and incubated with 1 ml of TEN buffer (30 mM Tris-HCl [pH 7.5], 1 mM EDTA, 15 mM NaCl) for 10 to 15 min. After removal from the plate, two-thirds of the cells were used to prepare extracts for the chloramphenicol acetyltransferase (CAT) and β-galactosidase assays by the freeze-thaw method and one-third was used for preparation of extract for the luciferase (LUC) assay by Triton X-100 lysis. For the CAT assay, cells were pelleted and lysed in 150 μl of buffer A (35) in three freeze-thaw cycles. The debris were pelleted and 20 to 50 μl of the supernatant was heated to 65°C for 20 min in order to inactivate endogenous acetylases. Subsequently, CAT was assayed by the liquid scintillation counting method (42). CAT extracts that were not heat treated were used for the β-galactosidase assay (37). LUC activity was assayed as described by the supplier of the reagents (Promega). The cell culture lysis reagent provided was supplemented with 0.4 mM phenylmethylsulfonyl fluoride. The light emission was measured with a Monolight 2010 instrument (Analytical Luminescence Laboratory) as relative light units. Protein was quantitated by the Bradford assay, and CAT and LUC specific activities were normalized for β-galactosidase specific activity. Plasmids pGL2-CON and pCAT-CON (Promega) containing a simian virus 40 (SV40) promoter-enhancer to drive LUC and CAT, respectively, were used as positive controls.

RNase protection assays. Plasmid DNA was cut with the appropriate enzymes and, where necessary, 3′ overhangs were trimmed with T4 DNA polymerase. The DNA was then phenol-chloroform extracted twice and ethanol precipitated. For each reaction mixture, 100 to 200 μg of template DNA was mixed with 2.5 μl of 100 mM dithiothreitol, 5 μl of a nucleoside triphosphate stock (5 mM ATP, UTP, and GTP), 1 μl of 300 μM CTP, 5 μl of [α-^{32}P]CTP (400 mCi/mmol), 5 μl of 5× transcription buffer (200 mM Tris [pH 7.5], 30 mM $MgCl_2$, 10 mM spermidine, and 50 mM NaCl), 7 to 12 U of T3 or T7 RNA polymerase, and 40 U of RNasin (Promega) and brought to a 25-μl total volume with H_2O. The reaction mixture was incubated for 1 to 2 h at 30°C and subsequently treated with RNase-free DNase (Promega) for 15 min. Next, 175 μl of solution containing RNase-free 0.1 M NaCl, 10 mM Tris-HCl (pH 8.0), and 1 mM EDTA was added and the mixture was passed through a Sephadex G-25 microspin column to remove the unincorporated label. The hybridization of the probe with the RNA and the RNase treatment were done as described elsewhere (39), and the reaction

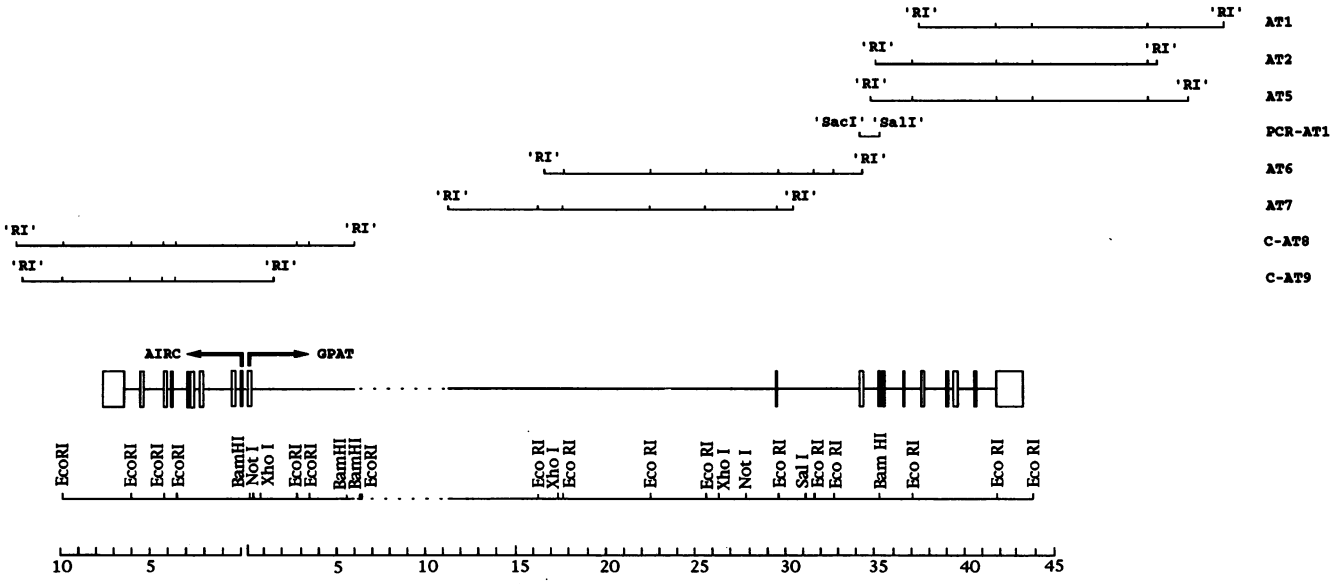

FIG. 1. Organization and restriction map of the chicken *GPAT-AIRC* genomic locus. Representative genomic clones and their *Eco*RI restriction patterns are shown at the top (adapter sites used for cloning are shown in quotation marks). Exons (boxes) and introns (thin lines) are indicated. The dotted line represents a part of the genomic locus that was not cloned. Arrows indicate the direction of transcription.

products were run on a sequencing gel and exposed at −80°C to Kodak X-Omat AR film with an intensifying screen. RNA probes of known lengths were also run alongside the sequencing ladder. Under the electrophoresis conditions used, RNA species ran 5% more slowly than DNA species. Therefore, we applied this correction when the sizes of the RNA protected fragments were calculated (39).

Nucleotide sequence accession number. The nucleotide sequence of the *AIRC-GPAT* intergenic region (shown in Fig. 3) has been deposited in the GenBank data base under accession no. L12533.

RESULTS

Chicken *GPAT* and *AIRC* genes are physically linked. For isolation of *GPAT* clones, a chicken oviduct genomic library was initially screened with the intact, previously isolated, avian cDNA (52). Phage clones AT1, AT2, and AT5, representative of five multiple isolates, are diagrammed in Fig. 1. These clones contain glutamine PRPP amidotransferase sequences downstream of codon 125 in the cDNA. In order to isolate the 5′ end of the *GPAT* gene, it was necessary to rescreen the library with a cDNA probe corresponding to nucleotides (nt) 75 to 521 of the cDNA. This led to the isolation of phages AT6, AT7, C-AT8, and C-AT9 (Fig. 1). Restriction mapping and Southern hybridization established that AT6 and AT7 did not overlap with the 5′ end of AT5 or the 3′ end of C-AT8. The missing DNA between AT5 and AT6, which carries part of *GPAT* exon 3, was isolated by genomic PCR using sequence information from the edges of clones AT2 and AT6. The PCR product was a single 1.1-kb fragment that was cloned and designated PCR-AT1. The 5′ end of PCR-AT1 overlaps with exon 3 in AT6, and the 3′ end overlaps with intron sequences in AT2 and AT5. The gap between clones C-AT8 and AT7 was estimated to be 5.5 kb by genomic Southern blots (data not shown). Subsequent nucleotide sequence analysis indicated that this region was part of *GPAT* intron 1. We were not successful in cloning the missing intronic segment by PCR. The phage clones depicted in Fig. 1 account for all of the *Eco*RI fragments in a genomic Southern blot probed with *GPAT* cDNA and thus contain all of the gene except for the missing portion of *GPAT* intron 1. Genomic Southern blots also suggest that *GPAT* is a single-copy gene (data not shown).

Since the 5′ end of the glutamine PRPP amidotransferase cDNA was not known, it was important to determine whether there was another *GPAT* exon upstream of the exon encoding Met-1. Accordingly, chicken liver poly(A$^+$) RNA blots were probed with various restriction fragments from genomic clones C-AT8 and C-AT9, located 5′ of *GPAT* exon 1. Surprisingly, several of these fragments hybridized with a 2.3-kb mRNA distinct from the 2.8- and 3.1-kb *GPAT* mRNAs (Fig. 2). Nucleotide sequence data from selected restriction fragments that hybridized to the 2.3-kb mRNA were used to scan the EMBL and GenBank data bases. A perfect match with the sequence of chicken AIR carboxylase-5-aminoimidazole-4-*N*-succinocarboxamide ribonucleotide (SAICAR) synthetase cDNA (4) was found, thus establishing the close physical linkage of the *GPAT* and

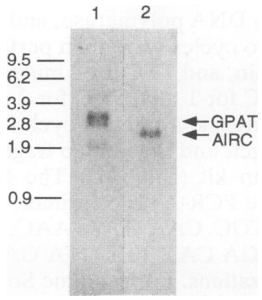

FIG. 2. Northern blots of the 2.3-kb *AIRC* mRNA (lane 2) and the 2.8- and 3.1-kb mRNAs of *GPAT* (lane 1). Both blots came from the same formaldehyde gel but were probed separately, the *AIRC* with a genomic subclone of phage genomic clone C-AT8 and the *GPAT* blot with the *GPAT* cDNA.

TABLE 1. Exon-intron organization[a]

Gene	Exon		5′ splice donor	Intron length (nt)	3′ splice acceptor	Codon phase[b]	Amino acid(s)
	No.	Length (nt)					
Glutamine PRPP amidotransferase	1	297	CAGgtaagg	~29,000	ttgtccacagGG	II	Arg-32
	2	67	AAGgtagaa	4,500	tctcctgaagGG	0	Lys-54/Gly-55
	3	207	AAGgttagt	900	gtcttcaaagCT	0	Lys-125/Leu-126
	4	113	AAGgtgagg	90	ctcattgcagAA	II	Arg-163
	5	146	AAGgtaagc	1,040	tttattacagAA	I	Gly-217
	6	73	CGAgtaagt	910	ctctttcaagGT	II	Glu-241
	7	152	AAGgtttgt	1,230	ctgttcttagGT	I	Gly-292
	8	128	AAGgtttgt	280	ctgtttgtagTG	0	Lys-339/Cys-340
	9	222	GAGgtaagt	920	ttttgttcagGT	0	Glu-413/Val-414
	10	121	TAGgtaaat	1,100	tttcctgtagGG	I	Gly-454
	11	1,480					
AIR carboxylase/SAICAR synthetase	1	77	CAGgtaaac	500	ttttctctagAA	I	Glu-7
	2	198	CAGgtttgt	1,400	ttctatttagGA	I	Gly-73
	3	179	AAGgtaagt	400	tccttaatagGA	0	Lys-32/Asp-33
	4	180	AAGgtatga	90	gtctttatagAT	0	Lys-189/Ile-190
	5	114	CAGgtaaca	840	ccttttttcagTC	0	Gln-229/Ser-230
	6	84	GAGgtattc	165	tcttctgtagTT	0	Gln-258/Leu-259
	7	181	AAGgtacat	1,100	ccccctgcaagGA	I	Gly-319
	8	156	GTGgtaaga	950	ttctttgcagGA	I	Gly-372
	9	1,190					

[a] All intron-exon boundaries conform to the consensus splice junctions (3).
[b] 0, positioning of introns between codons; I, interruption after the first base of a codon; II, interruption after the second base.

AIRC genes. A partial restriction map of the ~52-kb *GPAT-AIRC* region is shown in Fig. 1.

Exon-intron structure. Exons were initially positioned by correlating the restriction map of the *GPAT-AIRC* locus with sites in the exons and by probing Southern blots of genomic clones with cDNA fragments. Exact intron sizes were obtained either by nucleotide sequencing or by sizing the products of PCRs that used primers located on flanking exons. The nucleotide sequences of all *GPAT* and *AIRC* exons and intron-exon boundaries were determined. Splice sites (Table 1) agree with the consensus (3), and exons agree with the cDNA sequences obtained earlier (4, 52), with a few exceptions. The main revision to the cDNA sequence that is required is in the GC-rich 5′ UTR of *GPAT*. nt 1 to 61 in the truncated cDNA sequence (52) were inverted. The corrected 5′ UTR sequence in *GPAT* exon 1 is numbered in Fig. 3 from positions 881 to 1049. Additional corrections to the cDNA sequences are an A-to-T change at nt 496 of the glutamine PRPP amidotransferase cDNA (52), which results in no codon change, and the following changes in the AIR carboxylase-SAICAR synthetase cDNA: from G to T at nt 1, from C to T at nt 37, and from CG to GC at nt 47 and 48 (4). The 5′ UTR of *AIRC* is numbered nt 651 to 592 (Fig. 3).

Nucleotide sequencing and transcript mapping data presented below show that the two genes are transcribed divergently, with their Met-1 codons 458 bp apart and their transcriptional start sites just 229 bp apart (Fig. 3). The 44-kb *GPAT* gene is composed of 11 exons of similar sizes, except the last one, which contains the information for amino acids 454 to 511 of the protein and alternative 3′ UTRs as well (Table 1). *GPAT* mRNAs of 2.8 and 3.1 kb from chicken liver are accounted for by 3′ UTRs of 650 bases and 1 kb, respectively (52). Since exon 11 of *GPAT* contains the information for both 3′ UTRs and includes two polyadenylation sites, the two *GPAT* mRNAs arise by use of different polyadenylation sites rather than by differential splicing. The importance of the different, relatively long 3′ UTRs is not known. Examination of exon-intron boundaries indicates no preference for a particular splicing phase (Table 1). The first two introns of the gene are considerably longer than the rest, with the first intron alone accounting for more than half of the 52-kb locus.

The 8-kb *AIRC* gene is considerably more compact than *GPAT*. It is composed of nine exons, the last of which is considerably longer, coding for amino acids 372 to 426 and the 1,030-nt 3′ UTR (Table 1). There are two polyadenylation signals, 110 bp apart, but evidence for a shorter mRNA has not been obtained. The first six exons encode the SAICAR synthetase domain (amino acids 1 to 258), and exons 7 to 9 encode the carboxylase domain (residues 259 to 426). Domain boundaries were previously inferred by sequence alignments with monofunctional SAICAR synthetase and AIR carboxylase from bacteria and functional analysis of the avian AIR carboxylase-SAICAR synthetase cDNA in *E. coli* (4). The exon structure is thus consistent with a model in which AIR carboxylase-SAICAR synthetase evolved by fusion of ancestral genes encoding monofunctional enzymes. Exon-intron boundaries indicate a strong preference for splicing phases 0 and I, with splicing of exons 4 to 7 occurring in phase 0 and the rest in phase I.

The *GPAT-AIRC* intergenic region is GC rich, lacks TATA boxes, and shows a symmetrical pattern. A 1,350-bp *Sma*I restriction fragment hybridizing with both glutamine PRPP amidotransferase and AIR carboxylase-SAICAR synthetase cDNAs was isolated from clone C-AT8 and was subcloned, mapped, and sequenced on both strands (Fig. 3). The putative bidirectional promoter is 229 bp long, has an 80% GC content, and contains no recognizable TATA homologies. Two pairs of Sp1 and CCAAT boxes are located at the optimal distance (10, 49) from the transcriptional start site of *AIRC*, and one pair of Sp1 and CCAAT boxes is located at the optimal distance from the transcriptional start site of *GPAT* (Fig. 3). Two additional Sp1 boxes flank the intergenic region and are accompanied by a number of other potentially

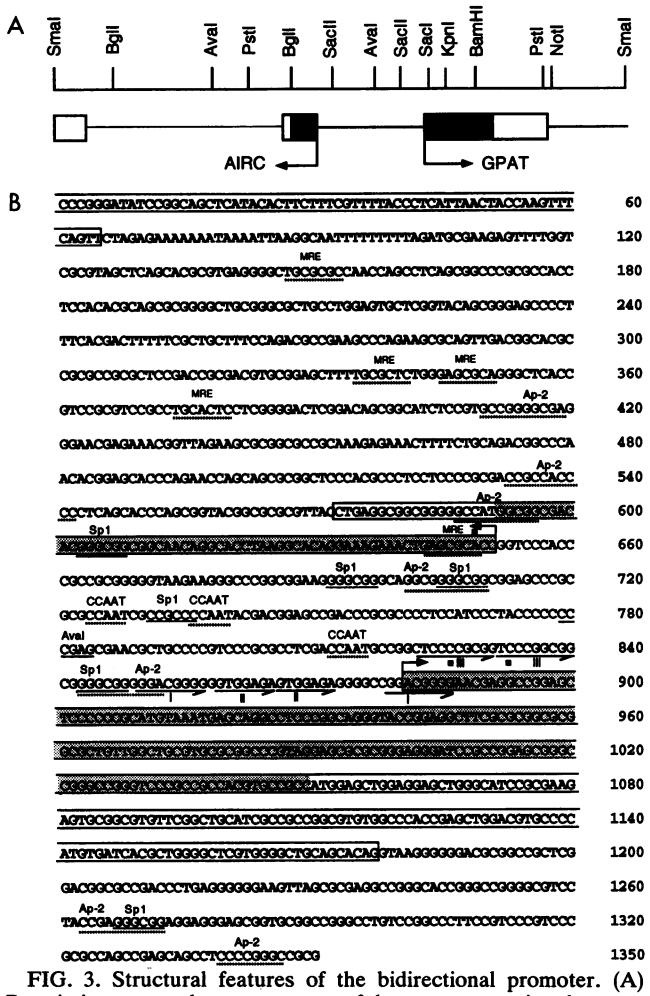

FIG. 3. Structural features of the bidirectional promoter. (A) Restriction map and exon structure of the promoter region drawn to scale. Exons (boxes) and 5′ UTRs (shaded areas) are depicted. Arrows show the direction of transcription. (B) Sequence and consensus sites of the promoter region. Transcription start sites of *GPAT* and *AIRC* as determined by RNase protection assays from a reporter minigene in HepG2 cells (dark arrows) and transcription start sites determined by primer extension for *AIRC* or RNase protection assay of chicken liver RNA for *GPAT* (filled squares) are shown. Putative Sp1 and CCAAT boxes, MRE, and AP-2 consensus sites are underlined. Three sets of direct repeats (I, II, and III) are underlined by half arrows. The AP-2 consensus sequences used are CCSCRGGC (32) and YCSCCMNSSS (24), and the MRE consensus sequence used is TGCRCNC (7).

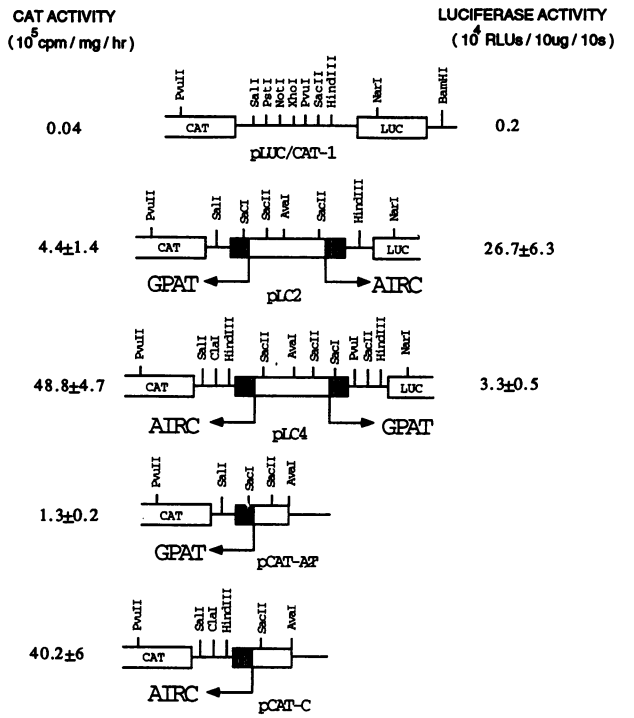

FIG. 4. Constructs used to assay promoter function. Values shown are the averages (± standard deviations) of at least three transfections, each done in duplicate. All results are normalized against β-galactosidase activity of a cotransfected plasmid, RSV-βgal. Background values of the relevant control vectors have been subtracted. RLUs, relative light units.

interesting sites. Five metal response element (MRE) consensus sites (7) flank the 5′ end of the *GPAT* gene, with two of them arranged as a perfect palindrome with a 3-bp spacer. It remains to be determined whether the putative MREs function to coordinate *GPAT* transcription with availability of Fe for the glutamine PRPP amidotransferase Fe-S cluster (reference 52 and references therein). Several AP-2 consensus sites (25, 33) are clustered around the first exon of each of the two genes. A number of perfect or nearly perfect direct repeats are found close to the *GPAT* transcriptional start site. These elements are candidates for *cis*-acting regulatory sites for *GPAT* and *AIRC* and could serve to coordinate expression with other genes for de novo purine nucleotide synthesis.

The *GPAT-AIRC* intergenic region is a bidirectional promoter in vivo. In order to determine whether the DNA region between the first exons of the two genes is sufficient for simultaneous bidirectional transcription, we constructed a bireporter plasmid (pLUC/CAT-1) with the reporter genes, *LUC* and *CAT*, in a head-to-head arrangement separated by a polylinker (Fig. 4). A 343-bp *Bgl*II-*Kpn*I fragment spanning the intergenic region (Fig. 3) was cloned in both orientations in pLUC/CAT-1, yielding plasmids pLC2 and pLC4 (Fig. 4). This fragment contains the 5′ UTR of *AIRC*, the intergenic region, and the first 60 bp of the *GPAT* 5′ UTR. Parts of the 5′ UTRs were included in order to retain potential promoter elements around the transcriptional start sites. With these constructs, bidirectional transcription was assayed from a single dish of transfected cells. With both orientations of the bidirectional promoter in plasmids pLC2 and pLC4, two independent assessments of transcription were obtained, one based on *CAT* expression and the other based on *LUC* expression.

HepG2 cells were transiently transfected with these plasmids. The results depicted in Fig. 3 show that this intergenic fragment can indeed function as a bidirectional promoter. Transcription towards *AIRC* was considerably stronger than towards *GPAT* (8-fold stronger on the basis of the LUC assay and 11-fold stronger on the basis of the CAT assay. The bidirectional promoter exhibited activity comparable to that of an SV40 promoter-enhancer system used to drive the LUC reporter in plasmid pGL2-CON and 2.5-fold lower than that of an SV40 promoter-enhancer used for expression of the CAT reporter in plasmid pCAT-CON. The difference between the expression of the *GPAT* and *AIRC* reporters

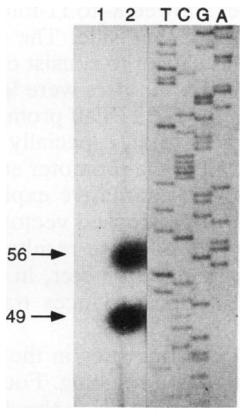

FIG. 5. Results of RNase protection assay to determine the *GPAT* transcription start site. Labeled *GPAT* antisense RNA encompassing the *Kpn*I-*Ava*I fragment of the *Sma*I genomic subclone S2 (Fig. 2) was hybridized to 20 μg of total chicken liver RNA (lane 2) or 20 μg tRNA (lane 1) and run against a sequencing ladder. The estimated sizes for the two protected fragments after a 5% correction to account for the slower mobility of RNA are given (see Materials and Methods).

correlates with the presence of twice as many optimally located Sp1 and CCAAT boxes on the stronger *AIRC* promoter side.

To further characterize the bidirectional promoter, the intergenic region was dissected. The objective was to determine whether half promoters could be obtained and, if so, to what extent they retained function. Starting with plasmids pLC2 and pLC4, the LUC reporter and the proximal half promoter attached to it were deleted by using a conveniently located *Ava*I site at nt 779 (Fig. 3), yielding plasmids pCAT-AT and pCAT-C, respectively (Fig. 4). Each half promoter retained the optimally located consensus CCAAT and Sp1 boxes. A deletion of the LUC reporter from the pLUC/CAT-1 promoterless vector provided plasmid pCAT-Δ, which served as a negative control. Transient expression in HepG2 cells showed that each half promoter retained function (Fig. 4). However, promoter strength decreased in comparison with that of the intact promoter. The *GPAT* side retained 30% and the *AIRC* side showed 80% of intact-promoter function. The *AIRC* half promoter was thus 30-fold more active than the *GPAT* half promoter. This result implies that *cis*-acting sites in the *AIRC* half promoter augment transcription from the *GPAT* half promoter, although an alternative possibility, that decreased expression was due to juxtaposition of vector sequences, was not eliminated.

***AIRC* and *GPAT* transcriptional start sites.** The 5′ end of the *AIRC* mRNA has been determined by primer extension and anchor-PCR cloning (4). A similar approach for *GPAT* mRNA yielded a truncated 5′ end (52) in which cDNA nt 1 to 61 were inverted with respect to the genomic sequences. RNase protection mapping was used to identify the *GPAT* mRNA cap site. In an initial experiment, an antisense RNA probe corresponding to nt 941 to 778 (Fig. 4) and chicken liver total RNA gave two protected fragments of approximately 56 and 49 nt, the longer of which corresponds to a cap site at nt 885 (Fig. 5). Efforts to confirm this result with different RNA probes failed because other RNA probes produced high background levels and/or uninterpretable protected fragments. As an alternative approach to map the 5′ end of *GPAT* and to confirm the *AIRC* primer extension

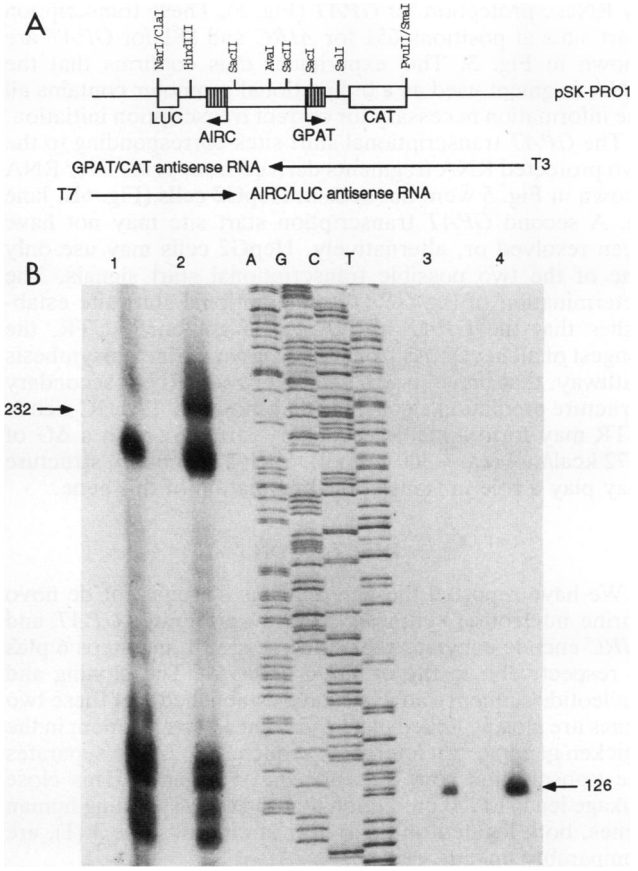

FIG. 6. RNase protection assays of total RNA from HepG2 cells transfected with construct pLC2. (A) Construct pSK-PRO1 was used for the preparation of antisense RNA probes of *LUC* and *CAT* mRNAs. Linearization of the plasmid with *Ava*I and in vitro transcription with T3 RNA polymerase yielded an antisense RNA to the *GPAT-CAT* message. Linearization of the construct with *Sac*II and in vitro transcription with T7 RNA polymerase yielded an antisense RNA to the *AIRC-LUC* message. (B) Results of the RNase protection assays. The *GPAT-CAT* antisense RNA probe was hybridized to 20 μg of tRNA (lane 1) and 20 μg of total RNA from transient transfectants (lane 2). A unique band with an estimated length of 232 nt in lane 2 supports a unique transcription start site at position 881 (Fig. 2). The *AIRC-LUC* antisense RNA probe was hybridized to 20 μg of tRNA (lane 3) and 20 μg of total RNA from transient transfectants (lane 4). A much stronger band (126 nt) relative to the control suggests a unique transcription start site at position 651 (Fig. 2). Estimated sizes of the RNA bands were corrected for an observed 5% slower mobility of RNA under the conditions used.

result, we determined the *GPAT* and *AIRC* transcriptional start sites from transient expression of the pLC2 reporter plasmid in HepG2 cells. The plasmid was transiently transfected into HepG2 cells, and total RNA was prepared from the transfectants. A *Nar*I-*Pvu*II restriction fragment from pLC2 was subcloned into pBluescript SK$^-$, and antisense RNA probes were made for the *GPAT-CAT* and *AIRC-LUC* regions, as shown in Fig. 6A. The RNase protection assays gave a single protected fragment for each reporter; the estimated sizes were 232 nt for the *CAT-GPAT* probe and 126 nt for the *LUC-AIRC* probe (Fig. 6B). The transcription start sites thus derived for the reporter minigenes are very close to those determined by anchor PCR for *AIRC* (4) and

by RNase protection for *GPAT* (Fig. 5). These transcription start sites at positions 651 for *AIRC* and 881 for *GPAT* are shown in Fig. 3. This experiment thus confirms that the DNA fragment used as a bidirectional promoter contains all the information necessary for correct transcription initiation.

The *GPAT* transcriptional start sites corresponding to the two protected RNA fragments derived from avian liver RNA shown in Fig. 5 were not seen in HepG2 cells (Fig. 6B, lane 2). A second *GPAT* transcription start site may not have been resolved or, alternatively, HepG2 cells may use only one of the two possible transcriptional start signals. The determination of the *GPAT* transcriptional start site establishes that the *GPAT* mRNA has a 169-bp 5′ UTR, the longest of all in cDNAs from the de novo purine biosynthesis pathway that have been characterized. RNA secondary structure prediction algorithms (26) show that this GC-rich 5′ UTR may form a stable secondary structure with a ΔG of −72 kcal/mol (ca. −300 kJ/mol). Such a secondary structure may play a role in translational regulation of this gene.

DISCUSSION

We have reported the first analysis of genes for de novo purine nucleotide synthesis from a vertebrate. *GPAT* and *AIRC* encode enzymes that catalyze step 1 and steps 6 plus 7, respectively, in the de novo pathway. The cloning and nucleotide sequence analysis have established that these two genes are closely linked in a head-to-head arrangement in the chicken genome. An intergenic sequence of 229 bp separates the transcription start sites for the two genes. This close linkage leads to the prediction that the corresponding human genes, both located on the q arm of chromosome 4 (1), are comparably linked.

For the purpose of simplifying this discussion, a bidirectional promoter is defined operationally as a short segment of DNA that promotes bidirectional transcription in vivo. This operational definition does not distinguish between mechanisms for transcription initiation that involve assembly of a single basal transcription complex at one site or complexes at two contiguous sites that may or may not interact. In this context, the *AIRC-GPAT* intergenic region is a bona fide bidirectional promoter that drives transcription of the two genes. This promoter exhibits 10-fold-higher activity on the *AIRC* side than on the *GPAT* side. Interestingly, this difference of transcriptional activity, if maintained in the chromosomal context, could contribute to ensuring that glutamine PRPP amidotransferase activity is rate limiting for purine nucleotide synthesis. It is assumed, but has not been shown directly, that reaction 1 is the rate-limiting step in the pathway for de novo purine nucleotide synthesis.

The *GPAT-AIRC* bidirectional promoter contains no recognizable TATA homologies. Despite this and the lack of known initiator elements (32, 45), transcription of *GPAT* and *AIRC* genes appears to start at distinct sites, in contrast to that of most housekeeping promoters, which usually generate heterogeneous mRNA 5′ ends. The basis for unique mRNA 5′ ends from the *GPAT* and *AIRC* genes is not known.

A number of Sp1 and CCAAT boxes are arranged within the promoter in a symmetrical pattern. Sp1 and CCAAT boxes are bidirectional (16, 49). Sp1 protein is present in all cell types, resulting in constitutive activation (27), whereas a multiplicity of proteins is known to bind the CCAAT box (9), which may have regulatory roles apart from activating constitutive gene expression. In agreement with the arrangement of Sp1 and CCAAT boxes, the *AIRC* side of the bidirectional promoter showed 8- to 11-fold-higher transcriptional activity than the *GPAT* side. The *GPAT-AIRC* bidirectional promoter was shown to consist of two half promoters. However, both half promoters were less active than the intact promoter, with the *GPAT* half promoter losing 70% of its activity. This suggests that, especially in the *GPAT* half promoter, the distal half of the promoter enhances transcriptional activity. However, alternative explanations, such as an inhibitory context of juxtaposed vector sequences, were not ruled out. Nevertheless, these results contrast with the murine *DHFR* bidirectional promoter, in which removal of the *DHFR* proximal region enhances transcription of the flanking gene (41).

Of other putative *cis*-acting sites in the bidirectional promoter, the MREs are most intriguing. Four potential MREs in *AIRC* intron I and one proximal to the *AIRC* transcription start site were identified. MREs function to activate transcription of metallothionein genes in order to detoxify heavy metals (34). Since glutamine PRPP amidotransferase contains an essential Fe-S cluster (reference 52 and references therein), it is attractive to think that MREs may function to coordinate Fe availability with de novo purine nucleotide synthesis via expression of *GPAT*. Fe availability is known to regulate the synthesis of other key metabolic enzymes by a translational mechanism (8, 18). It is important, however, that MREs have not been reported to respond to Fe.

A number of AP-2 sites are clustered in and around the bidirectional promoter. AP-2 is implicated in mediating induction of genes responding to protein kinase and cyclic AMP signal transduction pathways (25) and is itself induced by retinoic acid (48). If these sites are relevant in vivo, they will provide some possibilities for the regulation of these genes.

The difficulties in mapping the 5′ UTR by primer extension and RNase protection assays as well as in computer modelling suggest that the 5′ UTR is capable of forming a stable secondary structure. Secondary structure in a 5′ UTR can decrease the efficiency of translation initiation and may also contribute to the regulation of translation (29).

The promoter and flanking regions that include the first exons of both genes, most of *AIRC* intron 1, and the proximal part of *GPAT* intron 1 (Fig. 3, nt 100 to 1350) have a 74% GC content with a nonsuppressed CpG content and thus resemble a CpG island (2). There is a growing body of evidence that CpG islands, regions which contain clustered, nonmethylated CpG dinucleotides resulting from a high GC content and a lack of CpG suppression, are located at the 5′ ends of many genes in vertebrates (2, 15). Given the suppression of CpG content in the genome due to methylation of CpG and the evolutionary mutability of mCpG (2), it is evident that such loci are under positive selective pressure. A likely explanation for this conservation is that these islands are associated with the 5′ ends of predominantly housekeeping genes and host the promoters for these genes.

There are a number of clustered genes in vertebrates that are divergently transcribed from bidirectional promoters. These include housekeeping genes of the surfeit locus (6, 24), the murine and human α1(IV) and α2(IV) collagen genes (28, 46), the histone H2a and H2b genes (22), *DHFR* and an uncharacterized gene homologous to one for bacterial DNA mismatch repair (31, 43), the Wilms' tumor locus (23), a proliferating cell nuclear antigen gene (36), an SV40-like monkey genomic locus (38), and the CpG island HTF9 (30). The *GPAT* and *AIRC* genes can now be added to this list as perhaps the best-characterized case. These genes are at present unique among the group of divergently transcribed

genes in that both gene products are well defined and both function in a single metabolic pathway. In addition, *GPAT* and *AIRC* are not homologous, indicating that the cluster did not evolve by gene duplication as is likely for the α1 and α2 genes of type IV collagen (46). Except for the Wilms' tumor locus, the bidirectional promoter invariably resides in a CpG island and contains no TATA boxes. Sp1 boxes are common, whereas CCAAT boxes and distinct transcriptional start sites may or may not occur.

In addition to the examples of divergent transcription from vertebrates cited above, there is a two-gene cluster for pyrimidine synthesis in *Dictyostelium discoideum* (12). *PYR1-3*, encoding the first three enzymes of the pathway, is separated by 1.5 kb from *PYR4*, which codes for enzyme 4 in the six-step pathway to UMP synthesis. Interestingly, these divergently transcribed genes exhibit the same pattern of developmental regulation.

Divergent transcription of two genes from a bidirectional promoter can provide coexpression and coregulation of expression. Proximal *cis*-acting elements, as well as *trans*-acting factors, have the potential to modulate the expression of both genes simultaneously. The resulting coordinated expression would be not unlike that obtained from a bacterial operon. In view of the emerging number of examples for this arrangement and the unique opportunities for coregulation, this genetic unit which we have termed a dioskourion may represent a functional unit of expression in higher eukaryotes which is comparable in some respects to a bacterial operon.

The basis for the *GPAT-AIRC* dioskourion is not known. The evolution of the *GPAT-AIRC* dioskourion presumably provides a distinct biological advantage. It is not known whether expression of the dioskourion is regulated. Possibly, transcription of genes for step 1 and steps 6 plus 7 is coregulated, whereas other coexpressed genes in the pathway are constitutive. Whatever the basis for the *GPAT-AIRC* dioskourion, the tight linkage of two sites for transcription initiation within a CpG island suggests that this structure may be conserved in mammals. Preliminary work from our laboratories supports this prediction (2a).

ACKNOWLEDGMENTS

We thank Jon LeBowitz and Steve Broyles for helpful comments on the manuscript.

This research was supported by NIH grants GM46466 (to H.Z.) and DDKD18849 (to J.E.D.). Computer facilities were supported by NIH grant AI27713. Oligonucleotides were synthesized by the Purdue Laboratory for Macromolecular Structure, supported by the Diabetes Research and Training Center (NIH grant P60DK20524).

REFERENCES

1. **Barton, J. W., I. M. Hart, and D. Patterson.** 1991. Mapping of a locus correcting lack of phosphoribosylaminoimidazole carboxylase activity in Chinese hamster ovary cell Ade$^-$D mutants to human chromosome 4. Genomics **9**:314–321.
2. **Bird, A. P.** 1986. CpG-rich islands and the function of DNA methylation. Nature (London) **321**:209–213.

2a. **Brayton, K. A., Z. Chen, A. Gavalas, G. Zhou, J. E. Dixon, and H. Zalkin.** Unpublished data.

3. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
4. **Chen, Z., J. E. Dixon, and H. Zalkin.** 1990. Cloning of a chicken liver cDNA encoding 5-aminoimidazole ribonucleotide carboxylase and 5-aminoimidazole-4-*N*-succinocarboxamide ribonucleotide synthetase by functional complementation of E. coli pur mutants. Proc. Natl. Acad. Sci. USA **87**:3097–3101.
5. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid-guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156–159.
6. **Colombo, P., J. You, K. Garson, and M. Fried.** 1992. Conservation of the organization of five tightly clustered genes over 600 million years of divergent evolution. Proc. Natl. Acad. Sci. USA **89**:6358–6362.
7. **Culotta, V. C., and D. H. Hamer.** 1989. Fine mapping of a mouse metallothionein gene metal response element. Mol. Cell. Biol. **9**:1376–1380.
8. **Dandekar, T., R. Stripecke, N. K. Gray, B. Goosen, A. Constable, H. E. Johansson, and M. W. Hentze.** 1991. Identification of a novel iron-responsive element in murine and human erythroid δ-aminolevulinic acid synthase mRNA. EMBO J. **10**:1903–1909.
9. **Darn, A., J. Bollekens, A. Stowb, C. Benoist, and D. Mathis.** 1987. A multiplicity of CCAAT box-binding proteins. Cell **50**:863–872.
10. **Dynan, W. S., and R. Tjian.** 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. Nature (London) **316**:774–778.
11. **Edlund, T., M. D. Walker, P. J. Barr, and W. T. Rutter.** 1985. Cell specific expression of the fat insulin gene: evidence for role of two distinct 5′ flanking elements. Science **230**:912–916.
12. **Faure, M., J. Camonis, and M. Jacquet.** 1989. Molecular characterization of a *Dictyostelium discoideum* gene encoding a multifunctional enzyme of the pyrimidine pathway. Eur. J. Biochem. **179**:345–358.
13. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
14. **Felgner, P. L., T. R. Gadek, M. Holm, R. Roman, H. W. Chan, M. Wenz, J. P. Northrop, G. M. Ringold, and M. Danielsen.** 1987. Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. Proc. Natl. Acad. Sci. USA **84**:7413–7417.
15. **Gardiner-Garden, M., and M. Frommer.** 1987. CpG islands in vertebrate genomes. J. Mol. Biol. **196**:261–282.
16. **Gidoni, D., J. T. Kadonaga, H. B. Saldona, K. Takahashi, P. Chambon, and R. Tjian.** 1985. Bi-directional SV40 transcription mediated by tandem Sp1 binding interactions. Science **230**:511–517.
17. **Gnirke, A., T. S. Barnes, D. Patterson, D. Schild, T. Featherstone, and M. Olsen.** 1991. Cloning and in vivo expression of the human GART gene using yeast artificial chromosomes. EMBO J. **10**:1629–1634.
18. **Haile, D. J., T. A. Rouault, C. K. Tang, J. Chin, J. B. Harford, and R. D. Klausner.** 1992. Reciprocal control of RNA-binding and aconitase activity in the regulation of the iron-responsive element binding protein: role of the iron-sulfur cluster. Proc. Natl. Acad. Sci. USA **89**:7536–7540.
19. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
20. **Henikoff, S., M. A. Keene, K. Fechtel, and J. W. Fristrom.** 1986. Gene within a gene: nested Drosophila genes encode unrelated proteins on opposite DNA strands. Cell **44**:33–42.
21. **Henikoff, S., M. A. Keene, J. S. Sloan, J. Bleskan, R. Hards, and D. Patterson.** 1986. Multiple purine pathway enzyme activities are encoded at a single genetic locus in Drosophila. Proc. Natl. Acad. Sci. USA **83**:720–724.
22. **Hentschel, C. C., and M. L. Birnstiel.** 1981. The organization and expression of histone gene families. Cell **25**:301–313.
23. **Huang, A., C. E. Campbell, L. Bonetta, M. S. Hill, S. McNeill, M. J. Coppes, D. J. Law, A. P. Feinberg, H. Yeger, and B. R. G. Williams.** 1990. Tissue, developmental, and tumor-specific expression of divergent transcripts in Wilms tumor. Science **250**:991–994.
24. **Huxley, C., and M. Fried.** 1990. The mouse surfeit locus contains a cluster of six genes associated with four CpG-rich islands in 32 kilobases of genomic DNA. Mol. Cell. Biol. **10**:605–614.
25. **Imagawa, M., R. Chiu, and M. Karin.** 1987. Transcription factor AP-2 mediates induction by two different signal transduction pathways: protein kinase C and cAMP. Cell **51**:251–260.
26. **Jaeger, J. A., D. H. Turner, and M. Zuker.** 1989. Predicting

optimal and suboptimal secondary structure for RNA. Methods Enzymol. **183:**281–306.
27. **Jones, N. C., P. W. J. Rigby, and E. B. Ziff.** 1988. Trans-acting protein factors and the regulation of eukaryotic transcription. Genes Dev. **2:**267–281.
28. **Killen, P. D., P. Burbelo, Y. Sakurai, and Y. Yamada.** 1988. Structure of the amino-terminal portion of the murine α1(IV) collagen chain and the corresponding region of the gene. J. Biol. Chem. **263:**8706–8709.
29. **Koromilas, A. E., A. Lazaris-Karatzas, and N. Sonenberg.** 1992. mRNAs containing extensive secondary structure in their 5′ non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. EMBO J. **11:**4153–4158.
30. **Lavia, P., D. MacLeod, and A. Bird.** 1987. Coincident start sites of divergent transcripts at a randomly selected CpG-rich island of mouse. EMBO J. **6:**2773–2779.
31. **Linton, J. P., J.-Y. J. Yen, E. Selby, Z. Chen, J. M. Chinsky, K. Liu, R. E. Kellems, and G. F. Crouse.** 1989. Dual bidirectional promoters at the mouse *dhfr* locus: cloning and characterization of two mRNA classes of the divergently transcribed *Rep-1* gene. Mol. Cell. Biol. **9:**3058–3072.
32. **Means, A. L., and P. J. Farnham.** 1990. Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. Mol. Cell. Biol. **10:**653–661.
33. **Mitchell, P. J., C. Wang, and R. Tjian.** 1987. Positive and negative regulation of transcription in vitro: enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell **50:**847–861.
34. **Mueller, P. S., S. Salser, and B. Wold.** 1988. Constitutive and metal-inducible protein: DNA interactions at the mouse metallothionein I promoter examined by in vivo and in vitro footprinting. Genes Dev. **2:**412–427.
35. **Pothier, F., M. Ouelette, J. P. Julien, and S. L. Guérin.** 1992. An improved CAT assay for promoter analysis in either transgenic mice or tissue culture cells. DNA Cell Biol. **11:**83–90.
36. **Rizzo, M. G., L. Offavio, S. Travali, C. Chang, B. Kaminska, and R. Baserga.** 1990. The promoter of the human proliferating cell nuclear antigen (PCNA) is bi-directional. Exp. Cell. Res. **188:**286–293.
37. **Rosenthal, N.** 1987. Identification of regulatory elements of cloned genes with functional assays. Methods Enzymol. **152:** 704–720.
38. **Saffer, J. D., and M. F. Singer.** 1984. Transcription from SV40-like monkey DNA sequences. Nucleic Acids Res. **12:** 4769–4788.
39. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
40. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
41. **Schilling, L. J., and P. J. Farnham.** 1989. Identification of a new promoter upstream of the murine dihydrofolate reductase gene. Mol. Cell. Biol. **9:**4568–4570.
42. **Seed, B., and J.-Y. Sheen.** 1988. A simple phase-extraction assay for chloramphenicol acetyltransferase activity. Gene **67:** 271–277.
43. **Shimada, T., H. Fujii, and H. Lin.** 1989. A 165-base pair sequence between the dihydrofolate reductase gene and the divergently transcribed upstream gene is sufficient for bi-directional activity. J. Biol. Chem. **264:**20171–20174.
44. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions, p. 142–143. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
45. **Smale, S. T., and D. Baltimore.** 1989. The initiator as a transcription control element. Cell **57:**103–113.
46. **Soininen, R., M. Huotari, S. L. Hostikka, D. J. Prockop, and K. Tryggvason.** 1988. The structural genes for α1 and α2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region. J. Biol. Chem. **263:**17217–17220.
47. **Vogeli, G., and P. S. Kaytes.** 1987. Amplifications, storage and replication of libraries. Methods Enzymol. **152:**407–415.
48. **Williams, T., A. Admon, B. Lüscher, and R. Tjian.** 1988. Cloning and expression of AP-2, a cell-type-specific transcription factor that activates inducible enhancer elements. Genes Dev. **2:**1557–1569.
49. **Wingender, E.** 1990. Transcription regulating proteins and their recognition sequences. Crit. Rev. Eukaryotic Gene Exp. **1:**11–48.
50. **Wood, W. M., M. Y. Kao, D. F. Gordon, and E. D. Ridgway.** 1989. Thyroid hormone regulates the mouse thyrotropin β-subunit gene promoter in transfected primary thyrotropes. J. Biol. Chem. **264:**14840–14847.
51. **Zalkin, H., and J. E. Dixon.** 1992. De novo purine nucleotide biosynthesis. Prog. Nucleic Acid Res. Mol. Biol. **42:**259–287.
52. **Zhou, G., J. E. Dixon, and H. Zalkin.** 1990. Cloning and expression of avian glutamine phosphoribosylpyrophosphate amidotransferase. J. Biol. Chem. **265:**21152–21159.