Biochemical and Biophysical Research Communications **276,** 117–124 (2000)
doi:10.1006/bbrc.2000.3439, available online at http://www.idealibrary.com on IDEAL®

Molecular Cloning and Genomic Organization of the Mouse AE2 Anion Exchanger Gene

Jon Lecanda, Raquel Urtasun, and Juan F. Medina[1]

Laboratory of Molecular Genetics, Division of Hepatology and Gene Therapy, University Clinic and Medical School, University of Navarra, E-31008 Pamplona, Spain

Received August 8, 2000

The molecular organization of the AE2 (SLC4A2) gene, a member of the multigene family encoding sodium-independent chloride/bicarbonate anion exchangers, has previously been described in both humans and rats. In these two species, AE2 shows alternate promoter usages and tissue-specific expression of isoforms in a similar, but not identical, fashion. Here we report the molecular cloning and organization of the entire mouse AE2 gene. The gene consists of 23 exons and 22 introns and spans about 17 kb. Moreover, it drives transcription of N-terminal truncated isoforms from alternate promoter sequences in a way analogous to that described for rat and/or human orthologs. Thus, sequences within intron 2 function as overlapping alternate promoters for truncated isoforms AE2b$_1$ and AE2b$_2$, and sequences of intron 5 drive transcription of isoforms AE2c$_1$ and AE2c$_2$. Each of these variants has a specific alternative first exon, while remaining exons are common to the complete form of the message AE2a, the diversity at 5′ leading to different N-termini in corresponding encoded proteins. As expected, mouse AE2 promoter sequences and the patterns of tissue expression of AE2 isoforms resemble rat counterparts more closely than human ones.

Key Words: **anion exchanger; gene organization; alternate promoters; RACE; tissue expression.**

Sodium-independent anion exchangers (AE) are a family of membrane proteins that mediate the electroneutral exchange of Cl^- for HCO_3^- ions across cell membranes (1, 2). They are involved in intracellular pH and cell volume regulation. In polarized cells they may also be involved in transepithelial acid/base transport. Thus far, three members of the family, AE1, AE2 and AE3, are well characterized (reviewed in 1, 2). AE1 (also termed Band 3) is expressed in red blood cells and kidney. Mouse, human, and rat AE1 cDNAs were sequentially cloned in the late eighties (3–6). Through cross hybridizations with AE1 cDNA probes and/or through RT-PCR, AE2 and AE3 transcripts (and corresponding cDNAs) were also identified in these species (7–14). AE2 mRNA is expressed in a wide variety of mammalian tissues, including choroid plexus, stomach, intestine, kidney, testis, cochlea, salivary glands, and liver, among others (9–11, 15–22), while AE3 mRNA is mainly encountered in excitable tissues like the nervous system and cardiac muscle (11–14, 23). The chromosomal localization of AE genes has been worked out in humans and in rodent species: AE1, AE2, and AE3 genes locate in human chromosomes (Chr) 17, 7, and 2, in rat Chr 10, 4, and 9, and in mouse Chr 11, 5, and 1, respectively (see Ref. 24 and references therein). Moreover, all these genes except the murine AE2 gene have been cloned in their entirety (13, 14, 23, 25–31).

A common feature of AE genes appears to be the use of alternate promoters, from which transcription of N-terminal truncated isoforms may be driven in a tissue-specific manner (cf. Ref. 22 and references therein). For instance, cardiac AE3, a truncated isoform of brain AE3, is transcribed from intron 6 of the AE3 gene. Also kidney AE1 mRNA is a truncated isoform of erythroid AE1 transcribed from intron 3 of the AE1 gene. The function of this type of truncated isoform may differ, as suggested by the pathological findings in a model of targeted disruption of the AE1 gene in mice (32–34). In that model, the lack of only the erythroid isoform of AE1 was associated with spherocytosis and severe hemolytic anemia but not with renal alterations. Concerning the AE2 gene, several N-truncated variants of the complete form AE2a (i.e., AE2b$_1$, AE2b$_2$, AE2c$_1$, and AE2c$_2$ isoforms) have also been described to be driven from alternate promoters in humans and/or in rat, with some differences between

Abbreviations used: AE, sodium-independent anion exchanger; nt, nucleotides; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-PCR; SMART, switch mechanism at the 5′ of RNA templates.

[1] To whom correspondence should be addressed. Fax: +34 948 425649. E-mail: jfmedina@unav.es.

0006-291X/00 $35.00

TABLE 1

Oligonucleotides Used for 5′ RACE and RT-PCR of Mouse AE2 mRNA Isoforms

Name	Sequence (5′ → 3′)	Location
MAE02	(7397)-TCATGAGGTCAAGGTCGGCT, ←	Exon 6
MAE04	(7266)-TGTCCAGGCTGGGCTTTGG, ←	Exon 6
MAE07	(1364)-CGGGACACGAAATCTAGAGC, →	Exon 1
MAE12	(3543)-CATGGCCAAATCTTAGCCCTTCTCGCT, ←	Exon 2
MAE15	(4194)-TTCACCCCTGCCGCCATGGACTT, →	Intron 2 (exon $1b_2$)
MAE17	(4751)-GTGCTGTCAGCTCCTGCACT, →	Intron 2 (exon $1b_1$)
MAE18	(4784)-AGTCATCCTGCGGGAGTGCAGGAGCTGACA, ←	Intron 2 (exon $1b_1$)
MAE20	(5669)-CGGTGGT/ATTCAAAGTCTTCC, ←	Exons 3/4
MAE22	(5429)-CTTCCTCTTGCTCGGGGAAC, ←	Exon 3
MAE59	(7098)-TCCAGGAGTGGAAGTCAGGT, →	Intron 5 (exon $1c_2$)
MAE61	(6824)-TAGTGTCTCTGAGGGGCAAAGCA, →	Intron 5 (exon $1c_1$)
MAE68	(7289)-GCTCCTGAAGGTTGTAACTTCGATGTCC, ←	Exon 6
MAE70	(7181)-CCTCCACCAGAGTT/CTGGAAATAAGC, ←	Exon $1c_1$/exon 6
MAE72	(5678)-GAGGACTGGCGGTGGT/ATTCAAAGTC, ←	Exons 3/4

Note. Numbers of the first nt in primer sequences (in parentheses) refer to the gene sequence in GenBank (Accession No. AF255774). Details of adaptor primers AP1 & AP2 and UPM & NUP (Marathon and SMART kits from Clontech, respectively) are available from the manufacturer. Symbols: →, forward; ←, reverse.

species in their pattern of tissue expression (22, 30). Although the mouse AE2 gene has not yet been described, partial data obtained in this species indicate similar features (35). Knowledge about AE genes in all three species (humans, rat, and mouse) is likely to be very helpful in future biological studies. Accordingly, we report here the molecular cloning of the entire mouse AE2 gene and the pattern of gene expression from alternative promoters in several murine tissues.

MATERIALS AND METHODS

Isolation of genomic clones and sequence analysis. A mouse genomic library in vector P1 (Genome Systems, St. Louis, MO) was screened by PCR using oligonucleotides MAE01 and MAE02 (nt 760–779 and nt 990–971 in the mouse AE2 cDNA; Ref. 9), as forward and reverse primers respectively. Four positive genomic clones (P1-9317, P1-9318, P1-9319, and P1-9320) were obtained and their plasmid DNA was isolated with Nucleobond AX2000 columns (Macherey-Nagel, Düren, Germany). Sequence analysis was carried out by direct primer walking on plasmidic DNA and/or by PCR amplifications of overlapping DNA stretches followed by sequencing of resultant amplicons, as previously described for the human AE2 gene (31). Sequencing reactions were performed with an ABI Prism Cycle Sequencing kit (PE Biosystems), being further submitted to capillary electrophoresis in an automated ABI Prism 310 Genetic Analyzer. Nucleotide numbering is essentially as described for the human gene (22); introns are numbered with positive integers proceeding from 5′ donor sites to 3′ acceptor sites; for the upstream promoter and untranslated regions in exons 1 and 2, the base preceding the ATG start codon of AE2a within exon 2 is designated −1, and negative integers are used proceeding 5′ (the 2088-bp long intron 1 is not taken into account for this numbering); for alternate promoters and corresponding alternative exons (cf. Fig. 1), each base preceding the ATG translation initiation codons within intron 2 (those of $AE2b_1$ and $AE2b_2$ transcripts), in exon 6 (of $AE2c_1$; alternative intron $1c_1$ is not taken into account when numbering exon $1c_1$ and upstream sequence), and within intron 5 (of $AE2c_2$), are designated $-1b_1$, $-1b_2$, $-1c_1$, and $-1c_2$, respectively, negative integers being used proceeding 5′. The whole genomic DNA sequence of mouse AE2 gene determined in this work (including sequences not shown in this report) has been placed in the GenBank/EMBL database with the Accession No. AF255774. Alignments of some of those mouse sequences with AE2 genomic sequences from other species were carried out with the ABI Prism Navigator software (PE Biosystems).

RACE and RT-PCR. Total RNA was isolated from five mouse tissues (see Fig. 2), according to the guanidinium thiocyanate method (36), employing the TRI Reagent (Sigma). Liver, stomach, and kidney RNA were enriched for poly(A)$^+$ RNA with an mRNA Isolation kit (Roche Molecular Biochemicals). Estimation of the 5′ regions of AE2 mRNA isoforms was carried out through a RACE-based procedure (37), using a Marathon cDNA Amplification kit or a SMART PCR cDNA synthesis kit (both from Clontech). Following the reverse transcription of either total RNA or poly(A)$^+$ RNA isolated from different mouse tissues, and the utilization of kit-specific adaptors to resultant cDNAs (according to manufacturer's recommendations), several rounds of nested PCR were carried out; adaptor oligonucleotides were used as upstream primers and several oligonucleotides from the mouse AE2 gene sequence were employed as downstream primers (see Table 1). For AE2a, a first amplification with the oligonucleotide pair AP1/MAE20 was followed by nested PCR with the pair AP2/MAE12; for $AE2b_2$, the same first amplification carried out for AE2a was followed by nested PCR with oligonucleotides AP2/MAE22; for $AE2b_1$, oligonucleotide pairs UPM/MAE72 and NUP/MAE18 were used in the first- and second-round PCR amplifications, respectively; for $AE2c_1$, the respective pairs were UPM/MAE68 and NUP/MAE70, while for $AE2c_2$ they were AP1/MAE02 and AP2/MAE04. After resin extraction with High PCR Purification kit (Roche Molecular Biochemicals), the 5′-cDNA amplicons from corresponding nested reactions were ligated into pGEM-T Easy vector (Promega) and introduced into *Escherichia coli* strain XL1-Blue, resulting in a set of five libraries with the upstream regions of each isoform. Plating of 5′ amplicon libraries allowed for their screening by transferring colonies to nitrocellulose filters and hybridizing them with ^{32}P-labeled isoform-specific oligonucleotides, followed by the isolation and sequence analysis of the 5′ RACE inserts of all five AE2 mRNA isoforms (see Fig. 1).

The tissue expression of mRNA isoforms was evaluated by semi-quantitative RT-PCR (see Fig. 2). For each tissue, a cDNA synthesis was performed in a 40-μl reaction mixture containing 3 μg of corresponding total RNA, 200 U of M-MLV reverse transcriptase (Gibco), 4 mM of dNTPs, 300 ng of random hexamers (Roche Molecular

Biochemicals), and 40 U of RNase Out Recombinant Ribonuclease Inhibitor (Gibco). PCR fragments for AE2a, AE2b$_1$, AE2b$_2$, AE2c$_1$, and AE2c$_2$ isoforms were produced with 4 μl of each cDNA pool, in a final volume of 50 μl of the reaction mixture, including corresponding forward and reverse primers, 200 nM each (Table 1). As forward primers, the following isoform specific oligonucleotides were used: MAE07 for AE2a, MAE17 for AE2b$_1$, MAE15 for AE2b$_2$, MAE61 for AE2c$_1$, and MAE59 for AE2c$_2$. The reverse primers were MAE20 for "a" and "b" isoforms, and MAE02 for "c" isoforms. As normalizing internal control for each sample, amplification of a fragment of mouse GAPDH cDNA was performed with oligonucleotides 5′CCAAGGTCATCCATGACAAC and 5′TGTCATACCAGGAAATGAGC (upstream and downstream primers respectively). Aliquots were taken from amplification reactions at 20, 25, and 30 cycles for GAPDH, and at 25, 30, 35, and 40 cycles for AE2 isoforms, and electrophoresed in 1% agarose gel stained with ethidium bromide. Bands were visualized with an ultraviolet lamp and analyzed with the software Molecular Analyst/PC (Bio-Rad). In Fig. 2, only amplified bands from aliquots that are in the exponential phase (i.e., with no plateau effect at the specified number of cycles) are shown. Routine PCR amplifications were performed with a highly efficient *Taq* DNA polymerase (Eurobiotaq, from Eurobio, Les Ulis, France) in the reaction mixture, supplemented with 5% dimethyl sulfoxide (Sigma). Particular PCR reactions for the 5′ ends were carried out using "touchdown" conditions as previously described for the human ortholog (cf. Ref. 22).

RESULTS AND DISCUSSION

Cloning and Exon–Intron Organization of the Mouse AE2 Gene

We report here the molecular cloning and organization of the mouse AE2 gene. Four separate genomic clones were obtained from a P1 mouse genomic library (cf. Materials and Methods), two of which were fully sequenced. While clone P1-9318 lacked the 3′ region downstream of exon 20, another clone (P1-9317) was found to contain the entire sequence of the murine AE2 gene. Similarly to the rat and human orthologs (30, 31), the mouse gene spans about 17 kb. The exon–intron boundaries were determined by comparing the genomic sequence with that of the AE2 cDNA originally cloned from mouse kidney and lymphoid cells (9), i.e., with the AE2a message transcribed from the upstream promoter (cf. Fig. 1A). According to this comparison, the mouse AE2 gene consists of 23 exons and 22 introns, with a pattern of exon-intron boundaries that corresponds well with those in rat and human AE2 genes (30, 31). Moreover, 5′ RACE experiments with mouse RNAs from liver, stomach, and kidney confirmed that the mouse AE2 gene may drive transcription of N-terminal variants from alternative promoters. Thus variants AE2b$_1$ and AE2b$_2$ are driven from overlapping sequences within intron 2 (Fig. 1B), whereas variants AE2c$_1$ and AE2c$_2$ can be driven from overlapping sequences within intron 5 (Fig. 1C). Each of these variants has a specific alternative first exon, while remaining exons are common to the complete form of the message AE2a. Figure 1 shows that this diversity at the 5′ end leads to different N-termini in corresponding encoded proteins. The splice variant

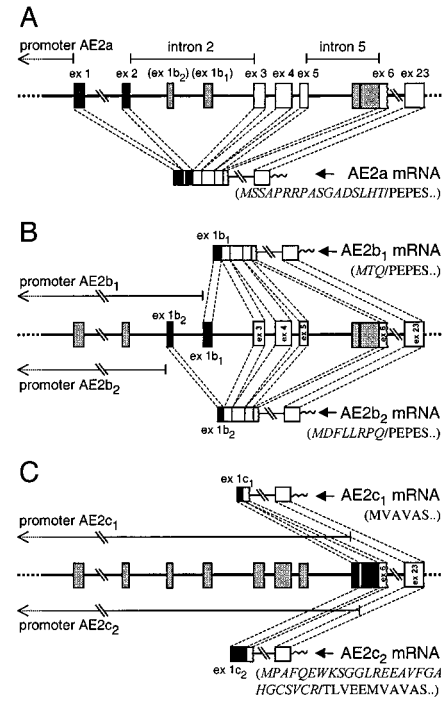

FIG. 1. Splicing patterns for mouse AE2 mRNA isoforms. (A) Exon 1 transcribed from the upstream promoter AE2a is spliced to exon 2 in AE2a mRNA. (B) Alternative exons type "b" transcribed from overlapping sequences of intron 2: above, exon 1b$_1$ transcribed from alternate promoter AE2b$_1$ (i.e., from approx the 5′ half of intron 2 and upstream sequences) is spliced to exon 3 in AE2b$_1$ mRNA (the downstream half of intron 2 results in alternative intron 1b$_1$); below, exon 1b$_2$ transcribed from alternate promoter AE2b$_2$ (approx the first quarter of intron 2 and upstream sequences) is spliced to exon 3 in AE2b$_2$ mRNA (sequences of intron 2 downstream to exon 1b$_2$ results in alternative intron 1b$_2$). (C) Alternative exons type "c" transcribed from overlapping sequences of intron 5: above, exon 1c$_1$ transcribed from alternate promoter AE2c$_1$ (approx the 5′ half of intron 5 and upstream sequences) is spliced to exon 6 in AE2c$_1$ mRNA (the downstream 286-bp long stretch results in alternative intron 1c$_1$); below, exon 1c$_2$ transcribed from alternate promoter AE2c$_2$ (that is the AE2c$_1$ promoter but extended downstream a few more nucleotides, cf. also Fig. 3) proceeds directly with exon 6 in AE2c$_2$ mRNA (with no alternative intron for this case). Open, black, and gray boxes represent common exons, specific exons, and possible exons which are not transcribed, respectively, in the corresponding transcripts. Exons 1b$_1$, 1b$_2$, and 1c$_2$ contain in frame ATG triplets, and may encode a few isoform-specific amino acids (single-letter code) shown in italics, but exon 1c$_1$ does not contain any ATG triplet, and truncated variant AE2c$_1$ starts at amino acid M199 of mouse AE2a protein.

AE2b$_1$ corresponds to the rat AE2b (30) and to the human AE2b$_1$ (22); this variant as well as variants AE2c$_1$ and AE2c$_2$ have already been reported in mouse (including some particularities at the N-terminus of the encoded AE2c$_2$ protein in this species; cf. Ref. (35)). On the other hand, the murine variant AE2b$_2$ corresponds to the novel AE2b$_2$ mRNA recently described in humans (22), and the specific first eight amino acids encoded by its alternative exon 1b$_2$ (cf. Fig. 1B) are the same in both species.

TABLE 2

Exon–Intron Boundaries of the Mouse AE2 Gene

Exon			Intron				Exon	
No.	Size	3′ junction	5′ junction	Size (bp)	No.	3′ junction	5′ junction	No.
1	≥158	CTAGCGG	gtaagc	2088	(1)	actcag	GTTATGC	2
2	114	CAC ACG	gtaagc	1783	(2)	tctcag	CCA GAG	3
$1b_1$	≥133	ACT CAG	gtgggg	587	($1b_1$)	tctcag	CCA GAG	3
$1b_2$	≥72	CCT CAG	gttcga	1142	($1b_2$)	tctcag	CCA GAG	3
3	163	TTT GAA T	gtgagg	125	(3)	ttgcag	AC CAC CGC	4
4	245	GTT CAG	gtcagt	82	(4)	ttgcag	TTC TTT	5
5	119	CAG ACA GG	gtaagt	1059	(5)	ctgcag	A ACT CTG	6
$1c_1$	≥113	TTTCCAG	gttaaa	286	($1c_1$)	ctgcag	AACTCTG	6
$1c_2$	≥293	GTC TGC AG					A ACT CTG	6
6	233	ATG AAA A	gtgagt	81	(6)	ccccag	GT CAC CGA	7
7	143	CAT GAG	gtacac	1097	(7)	cccaag	GTG TTT	8
8	181	GCC CAT G	gtacag	1161	(8)	tggcag	GA GCT GTG	9
9	136	AAG CAC AG	gtatgg	98	(9)	tactag	C CAC CCA	10
10	166	AGA GAG	gtgaag	131	(10)	ccccag	CGT GAG	11
11	115	CTC GTG G	gtatgt	83	(11)	ctctag	GC TGT GTG	12
12	185	GAC AAG	gtcagc	97	(12)	gggcag	CAA TTT	13
13	226	GAT AAG G	gtacgc	204	(13)	ttgcag	CA CTC CTG	14
14	216	CTA CTG G	gtaagg	81	(14)	ctccag	GG GAG AAG	15
15	149	TTC TCG	gtaaga	86	(15)	actcag	TTC TGC	16
16	195	ATC AAG	gtaggc	2422	(16)	ttccag	ATC TTC	17
17	255	GGC CGG	gtatgt	106	(17)	tcttag	ATC CGG	18
18	90	ACC CAG	gtaagg	86	(18)	ttgcag	AAA CTG	19
19	167	ATC ACC AC	gtgagc	338	(19)	tcccag	G CTG ATC	20
20	254	CTT GTG G	gtatgc	90	(20)	tcccag	GC CTC TCC	21
21	170	AAA AAG	gttagt	153	(21)	ctgcag	GTT CGG	22
22	174	AAA TGT	gtaagc	94	(22)	tcacag	CTG GAT	23
23	274							

Table 2 shows the sequences of exon–intron boundaries of the mouse AE2 gene and the sizes of exons and introns, alternative variants included. Splice sites all conform well to the predicted GT/AG rule for splice donors and acceptors (38). Regarding intron sequences of the gene, an adenine-rich stretch (67 adenines out of 100 nt) encountered in the middle region of intron 16 is worthy of note. As for exon sequences, nine changes to the previously reported cDNA sequence (9) were found in the genomic clones. Five of them occur each at the third base of codons 63, 83, 93, 367, and 557 (nt changes C371 → A, A431 → G, A462 → C, T1283 → C, and A1853 → T, respectively). They do not result in any change in the encoded amino acids, and may be considered as single nucleotide polymorphisms. But a change at nt 796–797 leads however to a different amino acid 205: GGC-encoded Gly is changed for GCG-encoded Ala in our sequence; due to this and other reasons recently noted (35), the GCG-encoded Ala reported previously (9), can be considered erroneous. The remaining three changes in our sequence are in untranslated regions: in exon 1 there is a nt change C1 → G, in exon 2 a cytosine is lacking (five instead of six) after nt A88, and in exon 23 there are four cytosines instead of three after nt G4014.

Tissue Expression of the Mouse AE2 mRNA Isoforms

The expression of each AE2 mRNA isoform was analyzed in five different murine tissues (liver, stomach, ileum, kidney, and lung), through a semiquantitative procedure based on RT-PCR. As shown in Fig. 2, AE2a mRNA was encountered in all examined tissues at similar levels, while the N-terminal variants showed some differences in their tissue expression pattern. $AE2b_1$ and $AE2b_2$ isoforms were found to be predominantly expressed in the stomach, although they were also detected at lower levels in the other tissues. $AE2c_2$ mRNA was also found in all examined tissues, being expressed at slightly higher levels in stomach and lung (Fig. 2); this does not concur with earlier reports on this variant as a tissue specific splicing product in mouse kidney (35). Finally, the splicing event leading to $AE2c_1$ mRNA could only been found in the mouse stomach (Fig. 2), in line with recent data (35). Comparisons of all these findings in mouse with those previously described in rat (30), and also with those more recently obtained in humans (22), show some similarities and differences. Thus both in humans and in rodent species, the longest transcript driven from the upstream promoter (AE2a message) appears to be universally expressed in most tissues. The expression of

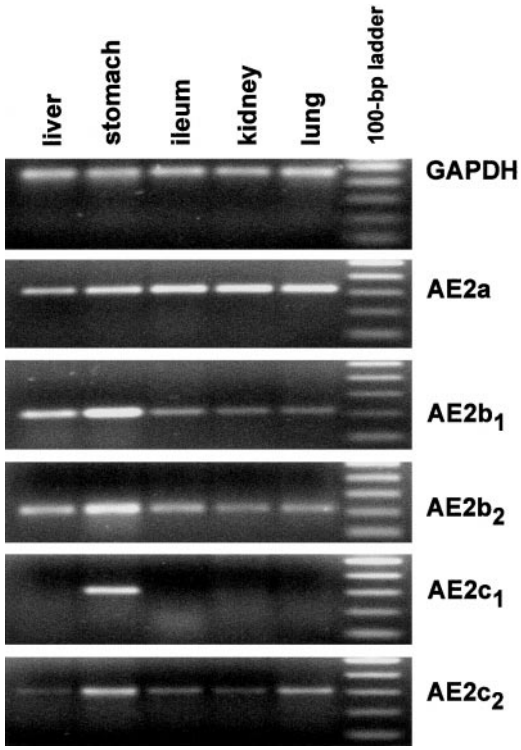

FIG. 2. Semiquantitative RT-PCR for AE2 isoforms in several mouse tissues. RT-PCR for GAPDH mRNA was used as a normalizing control. All bands shown correspond to amplicons which are in the linear phase of amplification at the specified number of cycles. Sizes of each amplicon and respective number of cycles are as follows: AE2a, 319 bp and 30 cycles; $AE2b_1$, 216 bp and 35 cycles; $AE2b_2$, 218 bp and 30 cycles; $AE2c_1$, 288 bp and 40 cycles; $AE2c_2$, 300 bp and 35 cycles; GAPDH, 465 bp and 25 cycles.

variant $AE2c_1$ but not that of variant $AE2c_2$ seems to be tissue specific in rodents (mouse and rat), while expression of both variants $AE2c_1$ and $AE2c_2$ has not been found in humans. On the other hand, mouse and rat AE2 genes express variants $AE2b_1$ and $AE2b_2$ in most tissues, though more intensely in the stomach, while AE2 gene in humans drives expression of variants $AE2b_1$ and $AE2b_2$ in a tissue specific manner, essentially in the liver and kidney but not in the stomach (22).

The RT-PCR procedure used to estimate the relative levels of each AE2 mRNA isoform (normalized to GAPDH mRNA) amplifies corresponding PCR fragments with oligonucleotide pairs that are distinct for each variant (cf. Materials and Methods). Accordingly, the procedure is optimal to compare the levels of each isoform in the various tissues (Fig. 2). However, it is not accurate enough to compare the expression levels among isoforms in a particular tissue, since the efficiencies of isoform-specific PCR amplifications probably diverge when using the different oligonucleotide pairs. To overcome this, an experiment of colony hybridization similar to those carried out for the estimation of the relative abundance of human AE2 variants in HepG2 cells (22), was attempted in mouse stomach, a tissue that expresses all five isoforms in this species. A representative 5′-RACE library was produced in which inserts were nested PCR products amplified from SMART stomach cDNA, using forward primers UPM and NUP (Table 1), for initial and subsequent rounds, respectively, in combination with several reverse primers of mouse AE2 cDNA (all of them downstream from exon 6, and thus common to the five isoforms). After plating the library, nitrocellulose replicas were produced for subsequent screening with ^{32}P-labeled isoform-specific oligonucleotides. While hybridization with the AE2a-specific oligonucleotide yielded over 10 colonies with positive signals, no positive colonies could be found with probes specific for variants "b" and "c." These data suggest that, in the mouse stomach, N-terminal variants each account for less than 10% of the longest transcriptional unit AE2a.

Transcription Initiation Sites of Mouse AE2 Isoforms and Sequence Analyses of the Promoter Regions

Putative initiation sites of the mouse AE2 mRNA isoforms were assumed as the 5′-most nt in corresponding inserts of 5′ RACE subclones obtained through PCR amplifications with particular sets of primers on mouse cDNAs (cf. Materials and Methods). RACE inserts for AE2a and $AE2b_2$ transcripts (obtained from Marathon liver cDNA) each had multiple 5′ends: as shown in Fig. 3, the 5′-most sites encountered for AE2a were positions G −205 (two RACE subclones), C −215 (one subclone), and C −221 (one subclone), while those for $AE2b_2$ were positions C $-32b_2$ (four subclones) and G $-48b_2$ (one subclone). Multiple 5′ ends were also found for the $AE2b_1$ variant (in SMART subclones obtained from mouse stomach cDNA), and they were positions A $-54b_1$ (eight subclones), C $-63b_1$ (two subclones), and G $-124b_1$ (one subclone). It is interesting that, similarly to what was reported for the human and rat orthologs (22, 30, 31), no TATA elements or Inr consensus sequences are locating around these initiation sites presumed for "a" and "b" transcripts. RACE experiments to obtain the 5′ regions for $AE2c_1$ and $AE2c_2$ variants yielded, however, unique 5′ ends in respective RACE inserts, i.e., position A $-129c_1$ for $AE2c_1$ (eight subclones from SMART stomach cDNA), and position T $-213c_2$ for $AE2c_2$ (five subclones from Marathon kidney cDNA). These assumed start sites are very close to each other in the genomic sequence within intron 5 (the exon $1c_1$ start site locates only 106 nt upstream of exon $1c_2$ start site; cf. Fig. 3). Both sites are more downstream than the putative start sites for exon 1c in corresponding rat sequence (30). Thus, although the AT-rich sequence that might serve as a TATA element in rat (30) is conserved in mouse (but not in humans), such a se-

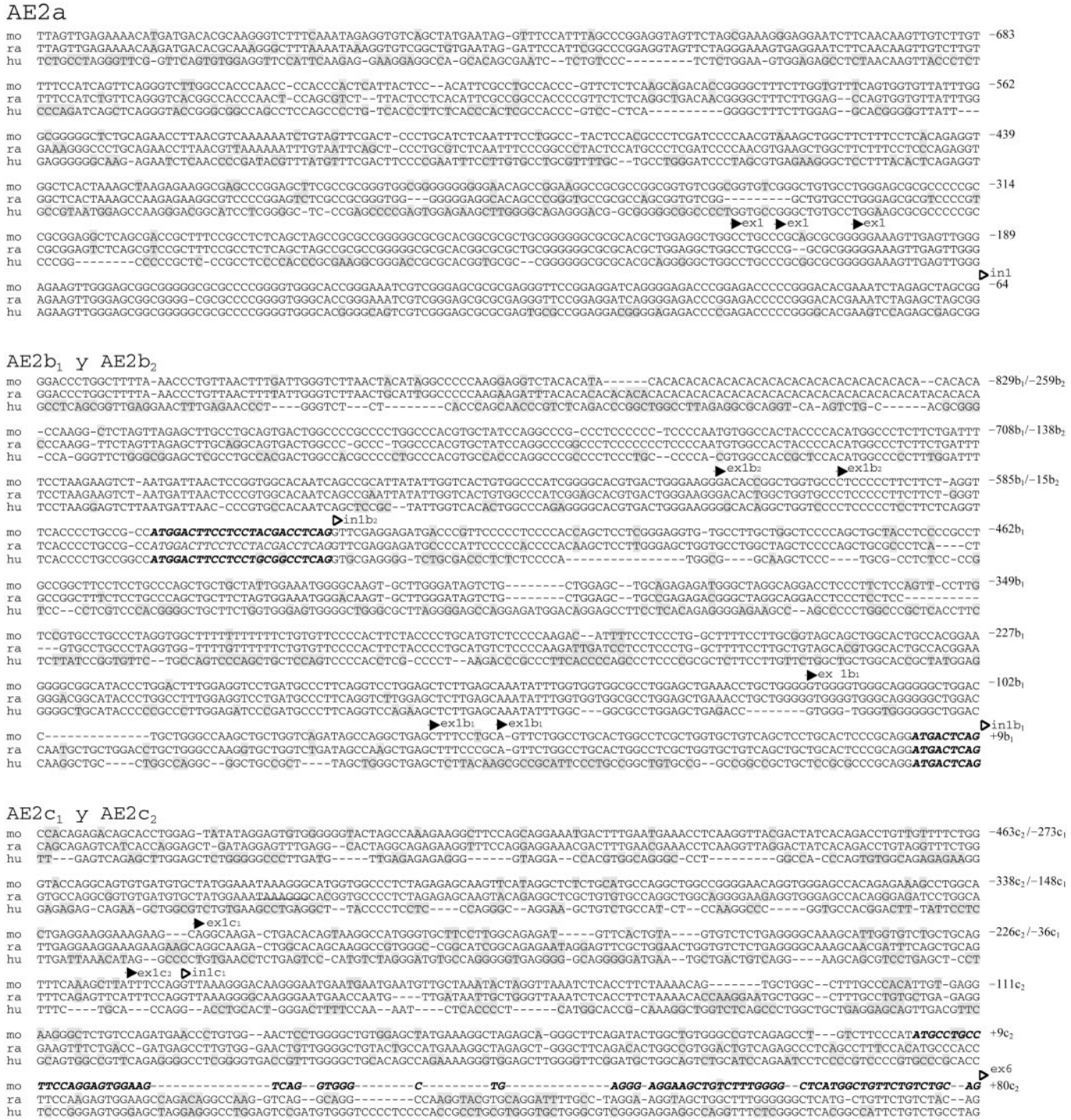

FIG. 3. Alignment of mouse promoter sequences with rat and human counterparts. (Top panel) Sequences of AE2a promoter and exon 1; the first nucleotide shown in the mouse sequence is position T −806 relative to the AE2a translation initiation codon, while the last nt is the donor splice site of exon 1, i.e., position G −64 (intron 1 is excluded for this numbering). (Middle panel) Part of overlapping sequences of $AE2b_1$/$AE2b_2$ promoters within intron 2 (including alternative exons $1b_1$ and $1b_2$); the first guanine shown in the mouse sequence is position ivs2+244 according to intron 2 numbering; the last nt G $+9b_1$ is the donor splice site of exon $1b_1$. Aligned sequences in bold italics indicate starting open reading frames in exons $1b_2$ and $1b_1$ (the rat sequence aligned with mouse coding sequence in exons $1b_2$ is labeled in plain italics to indicate that the complete identity of both sequences, including close upstream nt, strongly suggests that the $AE2b_2$ transcription unit may also be driven in rat). (Bottom panel) Part of overlapping sequences of rodent $AE2c_1$/$AE2c_2$ promoters within intron 5 (the corresponding human sequence is also included in the alignment). The first cytosine shown in the mouse sequence is position ivs5+394 in intron 5; the last nucleotide G $+80c_2$, is the 3′ junction of exon $1c_2$ (as well as of intron $1c_1$), with exon 6. In boldface is the coding sequence of exon $1c_2$ that might lead to a specific N-terminus of mouse $AE2c_2$ protein (cf. Ref. 35). A possible TATA element alleged for the rat AE2c promoter (30) is underlined. In all three panels, nt differing from the other sequences are shaded, and gaps are indicated by dashes; mo, ra, and hu refer to mouse, rat and human sequences; the numbers to the right of mouse sequences are relative to respective murine ATG start codons; ex, exon; in, intron; open arrowheads are at exon/intron junctions; black arrowheads indicate 5′ ends.

quence locates excessively far upstream from the assumed murine start sites (Fig. 3).

When mouse putative promoter sequences were aligned with the corresponding rat and human sequences, they were found to be highly homologous, especially the rat counterpart (Fig. 3). In fact, an almost complete identity between mouse and rat was observed in the proximal region of the AE2a promoter, as well as in extensive regions of overlapping AE2b promoters within intron 2. Thus, it is interesting that an AC-rich stretch located in the 5′quarter of intron 2 is highly conserved in the corresponding rat sequence (30) but not in humans (see Fig. 3, in the upper row of the middle panel). AE2a, $AE2b_1$, and $AE2b_2$ promoter regions are GC-rich sequences with potential binding sites for GCF, ETF, Sp1 and MAZ, as occurs in rat and human counterparts and in other TATA-less promoters associated with multiple start sites. Moreover, other potential *cis*-regulatory elements were also found in these regions. For instance, the mouse AE2a 5′-flanking region shows possible E-boxes, EGR-1/WT1, IL-6 RE, and CACCC elements, etc. Three of the CACCC elements encountered in mouse (positions −391, −615, and −654) are conserved in rat (30), the one at position −615 being conserved in humans as well (22). As shown in Fig. 3, the proximal $AE2b_2$ 5′-flanking region in mouse is highly conserved in rat and in humans, and several motifs alleged in these species (22, 30) are also encountered in the mouse region (e.g., several CCAAT and CACCC elements). A similar consideration can be made for the proximal $AE2b_1$ 5′-flanking region in mouse, especially when compared with the rat counterpart (30). The occurrence of binding motifs for liver-enriched transcription factors in mouse overlapping AE2b promoters (and in rat corresponding sequences) is, however, much lower than in the human counterpart (22). This concurs with our findings that variants "b" of mouse AE2 mRNA are expressed in most tissues, while in humans AE2b variants were only found in liver and kidney (22). Concerning the mouse overlapping AE2c regions, some potential elements have already been suggested (35). In these regions, the homology between mouse and rat sequences is much higher than those between either mouse or rat with humans (Fig. 3). This fact together with the failure to detect AE2c mRNA isoforms in humans might indicate that these isoforms are more specific for rodent species.

The biological relevance of each potential motif mentioned above is merely speculative so far. Additional experiments will be required to determine their significance and to elucidate the overall biological role of the AE2 alternate promoters and resultant N-terminal AE2 isoforms. The molecular cloning of the mouse AE2 gene will certainly facilitate these biological assays in the future.

ACKNOWLEDGMENTS

The authors thank E. Sáez for excellent technical assistance. We are also grateful to Dr. C. D. Funk for his comments on the manuscript. This work was supported by grants from the Spanish FIS (Projects 96/0346 and 99/0647) and the "Fundación Echébano." Sequence data for the mouse AE2 gene have been deposited with the GenBank/EMBL Data Libraries under Accession No. AF255774; the sequence of the mouse AE2 cDNA previously published by Alper *et al.* (9) has Accession No. NM_009207 (formerly J04036); the sequence of the human AE2 gene is available under Accession Nos. U62290 and AH008864.

REFERENCES

1. Alper, S. L. (1994) The band 3-related AE anion exchanger gene family. *Cell. Physiol. Biochem.* **4,** 265–281.
2. Kopito, R. R. (1990) Molecular biology of the anion exchanger gene family. *Int. Rev. Cytol.* **123,** 177–199.
3. Kopito, R. R., and Lodish, H. F. (1985) Structure of the murine anion exchange protein. *J. Cell. Biochem.* **29,** 1–17.
4. Tanner, M. J., Martin, P. G., and High, S. (1988) The complete amino acid sequence of the human erythrocyte membrane anion-transport protein deduced from the cDNA sequence. *Biochem. J.* **256,** 703–712.
5. Lux, S. E., John, K. M., Kopito, R. R., and Lodish, H. F. (1989) Cloning and characterization of band 3, the human erythrocyte anion exchange protein (AE1). *Proc. Natl. Acad. Sci. USA* **86,** 9089–9093.
6. Kudrycki, K. E., and Shull, G. E. (1989) Primary structure of the rat kidney band 3 anion exchange protein deduced from a cDNA. *J. Biol. Chem.* **264,** 8185–8192.
7. Demuth, D. R., Showe, L. C., Ballantine, M., Palumbo, A., Fraser, P. J., Cioe, L., Rovera, G., and Curtis, P. J. (1986) Cloning and structural characterization of a human non-erythroid band 3-like protein. *EMBO J.* **5,** 1205–1214.
8. Gehrig, H., Muller, W., and Appelhans, H. (1992) Complete nucleotide sequence of band 3 related anion transport protein AE2 from human kidney. *Biochim. Biophys. Acta* **1130,** 326–328.
9. Alper, S. L., Kopito, R. R., Libresco, S. M., and Lodish, H. F. (1988) Cloning and characterization of a murine band 3-related cDNA from kidney and from a lymphoid cell line. *J. Biol. Chem.* **263,** 17092–17099.
10. Lindsey, A. E., Schneider, K., Simmons, D. M., Baron, R., Lee, B. S., and Kopito, R. R. (1990) Functional expression and subcellular localization of an anion exchanger cloned from choroid plexus. *Proc. Natl. Acad. Sci. USA* **87,** 5278–5282.
11. Kudrycki, K. E., Newman, P. R., and Shull, G. E. (1990) cDNA cloning and tissue distribution of mRNAs for two proteins that are related to the band 3 Cl^-/HCO_3^- exchanger. *J. Biol. Chem.* **265,** 462–471.
12. Kopito, R. R., Lee, B. S., Simmons, D. M., Lindsey, A. E., Morgans, C. W., and Schneider, K. (1989) Regulation of intracellular pH by a neuronal homolog of the erythrocyte anion exchanger. *Cell* **59,** 927–937.
13. Yannoukakos, D., Stuart-Tilley, A., Fernández, H. A., Fey, P., Duyk, G., and Alper, S. L. (1994) Molecular cloning, expression, and chromosomal localization of two isoforms of the AE3 anion exchanger from human heart. *Circ. Res.* **75,** 603–614.
14. Linn, S. C., Askew, G. R., Menon, A. G., and Shull, G. E. (1995) Conservation of an AE3 Cl^-/HCO_3^- exchanger cardiac-specific exon and promoter region and AE3 mRNA expression patterns in murine and human hearts. *Circ. Res.* **76,** 584–591.
15. Jons, T., Warrings, B., Jons, A., and Drenckhahn, D. (1994) Basolateral localization of anion exchanger 2 (AE2) and acid-

secreting (parietal) cells of the human stomach. *Histochemistry* **102,** 255–263.

16. Brosius, F. C. 3rd, Nguyen, K., Stuart-Tilley, A. K., Haller, C., Briggs, J. P., and Alper, S. L. (1995) Regional and segmental localization of AE2 anion exchanger mRNA and protein in rat kidney. *Am. J. Physiol.* **269,** F461–F468 [published erratum appears in *Am. J. Physiol.*, index to Vol. **270,** 1996, section F following table of contents].
17. Parkkila, S., Rajaniemi, H., and Kellokumpu, S. (1993) Polarized expression of a band 3-related protein in mammalian sperm cells. *Biol. Reprod.* **49,** 326–331.
18. Holappa, K., Mustonen, M., Parvinen, M., Vihko, P., Rajaniemi, H., and Kellokumpu, S. (1999) Primary structure of a sperm cell anion exchanger and its messenger ribonucleic acid expression during spermatogenesis. *Biol. Reprod.* **61,** 981–986.
19. Chow, A., Dobbins, J. W., Aronson, P. S., and Igarashi, P. (1992) cDNA cloning and localization of a band 3-related protein from ileum. *Am. J. Physiol.* **263,** G345–G352.
20. Negrini, C., Rivolta, M. N., Kalinec, F., and Kachar, B. (1995) Cloning of an organ of Corti anion exchanger 2 isoform with a truncated C-terminal domain. *Biochim. Biophys. Acta* **1236,** 207–211.
21. Vázquez, J. J., Vázquez, M., Idoate, M. A., Montuenga, L., Martínez-Ansó, E., Castillo, J. E., García, N., Medina, J. F., and Prieto, J. (1995) Anion exchanger immunoreactivity in human salivary glands in health and Sjogren's syndrome. *Am. J. Pathol.* **146,** 1422–1432.
22. Medina, J. F., Lecanda, J., Acín, P., Ciesielczyk, P., and Prieto, J. (2000) Tissue-specific N-terminal isoforms from overlapping alternate promoters of the human AE2 anion exchanger gene. *Biochem. Biophys. Res. Commun.* **267,** 228–235, doi:10.1006/bbrc.1999.1951.
23. Linn, S. C., Kudrycki, K. E., and Shull, G. E. (1992) The predicted translation product of a cardiac AE3 mRNA contains an N terminus distinct from that of the brain AE3 Cl^-/HCO_3^- exchanger. Cloning of a cardiac AE3 cDNA, organization of the AE3 gene, and identification of an alternative transcription initiation site. *J. Biol. Chem.* **267,** 7927–7935.
24. Simon, J. S., Deshmukh, G., Couch, F. J., Merajver, S. D., Weber, B. L., Van Vooren, P., Tissil, F., Szpirer, J., Szpirer, C., Alper, S. L., Jacob, H. J., and Brosius, F. C., 3rd. (1996) Chromosomal mapping of the rat Slc4a family of anion exchanger genes, AE1, AE2, and AE3. *Mamm. Genome* **7,** 380–382.
25. Kopito, R. R., Andersson, M., and Lodish, H. F. (1987) Structure and organization of the murine band 3 gene. *J. Biol. Chem.* **262,** 8035–8040.
26. Sahr, K. E., Daniels, B. P., and Hanspal, M. (1996) Identification of the proximal erythroid promoter region of the mouse anion exchanger gene. *Blood* **88,** 4500–4509.
27. Kudrycki, K. E., and Shull, G. E. (1993) Rat kidney band 3 Cl^-/HCO_3^- exchanger mRNA is transcribed from an alternative promoter. *Am. J. Physiol.* **264,** F540–F547.
28. Schofield, A. E., Martin, P. G., Spillett, D., and Tanner, M. J. (1994) The structure of the human red blood cell anion exchanger (EPB3, AE1, band 3) gene. *Blood* **84,** 2000–2012.
29. Sahr, K. E., Taylor, W. M., Daniels, B. P., Rubin, H. L., and Jarolim, P. (1994) The structure and organization of the human erythroid anion exchanger (AE1) gene. *Genomics* **24,** 491–501.
30. Wang, Z., Schultheis, P. J., and Shull, G. E. (1996) Three N-terminal variants of the AE2 Cl^-/HCO_3^- exchanger are encoded by mRNAs transcribed from alternative promoters. *J. Biol. Chem.* **271,** 7835–7843.
31. Medina, J. F., Acín, A., and Prieto, J. (1997) Molecular cloning and characterization of the human AE2 anion exchanger (SLC4A2) gene. *Genomics* **39,** 74–85 [published erratum appears in *Genomics* **46,** 535, 1997].
32. Peters, L. L., Shivdasani, R. A., Liu, S.-C., Hanspal, M., John, K. M., González, J. M., Brugnara, C., Gwynn, B., Mohandas, N., Alper, S. L., Orkin, S. H., and Lux, S. E. (1996) Anion exchanger 1 (band 3) is required to prevent erythrocyte membrane surface loss but not to form the membrane skeleton. *Cell* **86,** 917–927.
33. Southgate, C. D., Chishti, A. H., Mitchell, B., Yi, S. J., and Palek, J. (1996) Targeted disruption of the murine erythroid band 3 gene results in spherocytosis and severe haemolytic anaemia despite a normal membrane skeleton. *Nat. Genet.* **14,** 227–230.
34. Hassoun, H., Hanada, T., Lutchman, M., Sahr, K. E., Palek, J., Hanspal, M., and Chishti, A. H. (1998) Complete deficiency of glycophorin A in red blood cells from mice with targeted inactivation of the band 3 (AE1) gene. *Blood* **91,** 2146–2151.
35. Stuart-Tilley, A. K., Shmukler, B. E., Brown, D., and Alper, S. L. (1998) Immunolocalization and tissue-specific splicing of AE2 anion exchanger in mouse kidney. *J. Am. Soc. Nephrol.* **9,** 946–959.
36. Chomczynski, P., and Sacchi, N. (1987) Single-step of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162,** 156–159.
37. Frohman, M. A. (1990) RACE: Rapid amplification of cDNA ends. *In* PCR Protocols. A Guide to Methods and Applications (Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J., Eds.), pp. 28–38, Academic Press, San Diego, CA.
38. Mount, S. M. (1982) A catalogue of splice junction sequences. *Nucleic Acids Res.* **10,** 459–472.