Three N-terminal Variants of the AE2 \( \text{Cl}^-/\text{HCO}_3^- \) Exchanger Are Encoded by mRNAs Transcribed from Alternative Promoters*

Zhuo Wang‡, Patrick J. Schultheis, and Gary E. Shull§

From the Department of Molecular Genetics, Biochemistry and Microbiology, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0524

Received for publication, October 30, 1995

Multiple AE2 \( \text{Cl}^-/\text{HCO}_3^- \) exchanger mRNAs have been identified in rat. To determine the genetic basis for these mRNAs and whether they encode different variants of the exchanger, we used both rapid amplification of cDNA ends and S1 nuclease protection protocols and examined the organization of the gene. mRNAs encoding three N-terminal variants of AE2 (AE2a, AE2b, and AE2c) were identified and shown to be transcribed from alternative promoters. The AE2a transcription unit consists of 23 exons, with exons 1 and 2 containing 5'-untranslated sequence and the first 17 codons. The first exon of AE2b is located in intron 2; it contains 5'-untranslated sequence and an alternative 3-amino acid N-terminal coding sequence and is spliced to exon 3. The first exon of AE2c is located in intron 5; it consists of 5'-untranslated sequence and is spliced to exon 6, which contains the translation initiation codon corresponding to Met-200 of AE2a. Northern analysis shows that AE2a is expressed in all tissues, AE2b exhibits a more restricted distribution with highest levels in stomach, and AE2c is expressed only in stomach. Thus, the use of alternative promoters leads to the production of three N-terminal variants of AE2 that exhibit tissue-specific patterns of expression.

Electroneutral \( \text{Cl}^-/\text{HCO}_3^- \) exchange in mammalian tissues is mediated by a family of proteins encoded by multiple mRNAs from at least three genes, termed AE1, AE2, and AE3 (reviewed in Refs. 1–3). The AE1 gene encodes both erythrocyte band 3 (4) and kidney band 3 (5–8), an N-terminal truncated variant of the exchanger. Kidney AE1 mRNA is transcribed from an alternative promoter located in the third intron of the erythrocyte transcription unit (7), and utilizes a Met codon in exon 5 as the translation start site (7, 8). In the case of AE3, a 4.4-kb mRNA encoding a 1227-amino acid variant is expressed in brain and several other tissues (9, 10), and a 3.8-kb mRNA is expressed in heart (9–13). The cardiac AE3 mRNA, which encodes a 1030-amino acid N-terminal variant of the exchanger, is transcribed from an alternative promoter located in intron 6 of the brain transcription unit (11–13). Thus, for both AE1 and AE3, the use of tissue-specific alternative promoters leads to the production of at least two proteins that differ in their N-terminal sequences.

Full-length cDNAs encoding what appears to be a single variant of AE2 have been cloned from multiple tissues and four mammalian species (10, 14–17); however, Northern blot analyses suggest that there are at least three different AE2 mRNAs (10). These include a ubiquitous 4.4-kb mRNA that corresponds to the form that has already been cloned, a 4.2-kb mRNA that is particularly abundant in stomach but is also present at lower levels in some other tissues, and a 3.8-kb mRNA that is expressed only in stomach. Because AE2 mRNAs are expressed in all tissues, it is likely that AE2 plays a housekeeping role such as the regulation of intracellular pH or cell volume. There is evidence, however, that it also serves more specialized functions in polarized epithelial cells. In stomach, AE2 is expressed at high levels on the basolateral membrane of gastric parietal cells (18), where it presumably mediates the exchange of \( \text{Cl}^- \) and \( \text{HCO}_3^- \) that occurs during acid secretion across the apical membrane. In small intestine, it has been localized on the apical membranes of both villus and crypt enterocytes in ileum (16), consistent with a role in both NaCl absorption and \( \text{HCO}_3^- \) secretion. The mechanisms underlying the large variations in AE2 mRNA expression levels in different tissues and the mechanisms by which the protein is localized to either apical or basolateral membranes are unknown.

Interestingly, the size variations among the AE2 mRNAs seem to be due to differences at their 5’ ends (10), consistent with the possibility that they are derived by the use of alternative promoters and encode protein variants that differ in their N-terminal sequences. If the multiple AE2 mRNAs are produced by the use of alternative promoters, it could contribute to the regulation of \( \text{Cl}^-/\text{HCO}_3^- \) exchange in polarized epithelia and other tissues, either as a mechanism for producing variations in the primary structure of the protein, which in turn could lead to differences in membrane location or functional properties, and/or as a mechanism for transcriptional control. Therefore, we considered it important to determine the primary structures of the proteins encoded by the multiple AE2 mRNAs, their genetic basis, and their tissue-specific patterns of expression. The results of this study demonstrate that the AE2 gene contains three distinct promoters, which exhibit significant differences in their tissue specificity, and that the use of these promoters leads to the production of four mRNAs encoding three N-terminal variants of the exchanger.

**EXPERIMENTAL PROCEDURES**

Rapid Amplification of cDNA Ends (RACE) Protocol—The 5’ RACE procedure was performed using the 5’-Amplifier™ RACE Kit (Clon-
First strand cDNAs were synthesized using reverse transcriptase, 2 μg of rat stomach poly(A)^+ RNA, and an AE2-specific primer (P1) complementary to sequences near the 5′ end of the AE2 mRNA to be reverse transcribed. After alkaline hydrolysis of RNA templates, an oligonucleotide anchor supplied with the RACE kit was ligated to the first strand cDNAs using T4 RNA ligase. PCR was performed using anchor ligated first strand cDNAs and primers. The 5′ primer was complementary to the anchor sequence (supplied with the RACE kit), and the other was an AE2-specific primer (P2) that was nested 5′ to the P1 primer. PCR conditions were denaturation at 94°C for 45 s, annealing at either 55°C (experiment shown in Fig. 1) or 60°C (all other experiments) for 45 s, and extension at 72°C for 2 min. Either 35 cycles (experiment shown in Fig. 1) or 40 cycles (all other experiments) were performed with 15 min of final extension following the last cycle.

For amplification of the 5′ ends of AE2a and AE2b described in Fig. 5, the P1 primer was complementary to nt +235 to +255, and the P2 primer was complementary to nt +197 to +214 of the rat AE2a coding sequence (10). For amplification of the 5′ end of AE2c described in Fig. 2, the P1 primer was complementary to nt +1047 to +1065, and the P2 primer was complementary to nt +674 to +699. After fractionation by agarose gel electrophoresis, PCR products were visualized by ethidium bromide staining. Specific PCR products were captured on NA45 paper (Schleicher & Schuell), eluted, and reamplified under the same conditions used for the first round of PCR. The reamplified products were subcloned into the pCR™II plasmid vector (Invitrogen), and sequence analysis was performed by the chain termination method.

For amplification of the 5′ end of AE2a described in Fig. 5, the P1 primer was complementary to nt +32 to +52, and the P2 primer was complementary to nt +17 to +36 (numbered as in top panel of Fig. 7; see also Fig. 5). In this experiment the P2 primer was 5′-end labeled with T4 kinase and [γ-32P]ATP before being used in the PCR reaction. PCR products were fractionated by electrophoresis on a 6% polyacrylamide gel and visualized by autoradiography. Regions of the gel containing PCR products were excised, and the PCR products were eluted, reamplified with unlabeled primer, and subcloned into the pCR™II plasmid vector. Bacterial colonies harboring plasmids with AE2 sequences were identified by colony hybridization using a 32P-labeled oligonucleotide probe corresponding to nt 91 to 117 of the AE2a mRNA (complementary to nt +7 to +22 as numbered in top panel of Fig. 7). Forty-eight positive colonies were isolated and sequenced by the chain termination method.

Cloning and Characterization of the Rat AE2 Gene—The rat cosmid library used in our studies of the AE1 and AE3 genes (7, 11) was screened with a 32P-labeled XbaI-BamHI restriction fragment spanning nt −77 to +163 of the 5′ end of the rat AE2-cDNA (10) under conditions described previously (7, 11). Four clones were identified and colony-purified. Southern blot analysis of EcoRI-digested cosmid DNAs suggested that all four clones contained the full-length AE2 gene. Five EcoRI fragments of done 10-1 that hybridized with the AE2a cDNA, with sizes of 10.2, 4.9, 4.3, 2.1, and 1.8 kb, were subcloned into a pBluescript vector. Each of the plasmids was sequenced by the dideoxy method using the sequencing method and an anchor primer from exon 1 and a primer from exon 4. The sequences of all exons and most of the introns were determined by size analysis of PCR products containing the introns, which were amplified using primers derived from the adjacent exons.

1. **Northern Blot Hybridization Analysis**—For the analysis described in Fig. 3, a blot containing 10 μg of rat stomach poly(A)^+ RNA was prepared, hybridized, and washed as described previously (10). The blot was hybridized with probes specific for each AE2 variant and then stripped before proceeding with the next analysis. For the experiment shown in Fig. 8, we used a blot from a previous study (10), which had been prepared using 5 μg of poly(A)^+ RNA from each of the rat tissues shown. Single-stranded antisense probes were generated from PCR products by extension of the antisense PCR primer using the Klenow fragment of DNA polymerase I in the presence of [α-32P]dCTP. For the specific probes for each mRNA were (i) AE2a, nt −155 to +51 of the AE2a cDNA, corresponding to 155 nt of 5′-untranslated sequence and 51 nt of coding sequence from AE2a-specific exons 1 and 2; (ii) AE2b, nt +1 to +62 of the AE2b transcript unit, corresponding to the first exon of AE2b, plus 48 nt from the RACE anchor primer; (iii) AE2c, nt +1 to +210 of the AE2c transcript unit, corresponding to the first exon of AE2c1 (this probe recognizes both the AE2c1 and AE2c2 mRNAs); and (iv) intron 5, the 330-nt sequence from the 3′ end of intron 5, which is located between the first exon of AE2c1 and exon 6 (this probe recognizes only the AE2c2 mRNA). The probe common to all AE2 mRNAs was a SacI-PvuII fragment, corresponding to nt +1118 to +2001 of the AE2a coding sequence (10), that was labeled with [α-32P]dCTP using random hexamers and the Klenow fragment of DNA polymerase I.

### RESULTS

Cloning of the 5′ Ends of mRNAs Encoding Three Terminal Variants of AE2—Our previous Northern blot analyses (10) demonstrated the existence of at least three AE2 mRNAs. These include a 4.4-kb mRNA corresponding to the AE2 cDNA that has already been cloned, and mRNAs of 4.2 and 3.8 kb that seem to differ at their 5′ ends. Throughout this paper the proteins encoded by the 4.4- and 4.2-kb mRNAs will be referred to as AE2a, AE2b, and AE2c, respectively. To obtain information that would facilitate the cloning of the 5′ ends of the AE2b and AE2c mRNAs, we first characterized the intron/exon organization of the rat gene (discussed below), and then performed Northern blot analysis to determine the points at which the sequences of the mRNAs diverged (data not shown).

A probe corresponding to exons 1a and 2 of the AE2a transcript unit hybridized only with the 4.4-kb AE2a mRNA. Probes corresponding to exon 3 or to exons 3-5 of AE2a hybridized with both the 4.4-kb AE2a mRNA and the 4.2-kb AE2b mRNA, but not with the 3.8-kb AE2c mRNA. A probe corresponding to exon 6 of AE2a hybridized with all three mRNAs. These results suggested that unique sequences in the AE2b and AE2c mRNAs, we first characterized the intron/exon organization of the rat gene (discussed below), and then performed Northern analysis to determine the points at which the sequences of the mRNAs diverged (data not shown).

A probe corresponding to exons 1a and 2 of the AE2a transcript unit hybridized only with the 4.4-kb AE2a mRNA. Probes corresponding to exon 3 or to exons 3-5 of AE2a hybridized with both the 4.4-kb AE2a mRNA and the 4.2-kb AE2b mRNA, but not with the 3.8-kb AE2c mRNA. A probe corresponding to exon 6 of AE2a hybridized with all three mRNAs. These results suggested that unique sequences in the AE2b and AE2c mRNAs would be immediately 5′ to exon 3 and that unique sequences in the AE2c mRNA would be immediately 5′ to exon 6. A probe corresponding to the 5′ end of the AE2b transcript unit, a RACE cloning protocol was employed. First strand cDNA was synthesized using rat stomach poly(A)^+ RNA and a primer from exon 4. The cDNA was then ligated to the anchor oligonucleotide and PCR-amplified using the anchor primer and a primer from exon 3.

Two products, 470 and 290 bp in length, later shown to correspond to the 5′ ends of the AE2a and AE2b mRNAs, respectively, were visible over background when the RACE reaction mixture was analyzed by agarose gel electrophoresis and ethidium bromide staining (Fig. 1, top left). These products were isolated from the gel, reamplified in separate reactions, and again analyzed by agarose gel electrophoresis (Fig. 1, top right). The 470- and 290-bp products were then isolated from the second gel, subcloned, and sequenced.

The two RACE products differed in sequence at their 5′ ends but were identical at their 3′ ends, with the common sequence beginning at the acceptor site of exon 3 (Fig. 1, bottom panel). As expected, the 470-bp product was derived from the AE2a mRNA; it contained sequences from exons 1, 2, and 3 of the AE2a transcript unit, with the first 17 codons occurring in exon 2. The 290-bp product was derived from an mRNA, subsequently shown to be the AE2b mRNA, in which exon 3 was immediately preceded by a unique 5′ sequence. Sequence analysis of seven independent AE2b subclones revealed that they...
AE2 Cl⁻/HCO₃⁻ Exchanger Alternative Promoters

Fig. 1. PCR cloning of the 5' end of the AE2b mRNA. First strand rat stomach cDNA was prepared, ligated with the anchor oligonucleotide, and analyzed by 5' RACE (see "Experimental Procedures"). Upper left, DNA size markers and RACE products were fractionated by agarose gel electrophoresis and visualized by ethidium bromide staining. Note the discrete 470- and 290-bp products, which are derived from the AE2a and AE2b mRNAs, respectively. Upper right, the AE2a and AE2b RACE products from the first set of PCR amplifications were isolated, reamplified in separate reactions, and analyzed by agarose gel electrophoresis. Each product was isolated, subcloned, and sequenced. Bottom panel, nucleotide and deduced amino acid sequences of the 5' ends of the AE2a and AE2b mRNAs. The translation initiation codon is underlined, and the first nucleotide of exon 3, which begins the common sequence is indicated by an arrow.

Extended to a position ranging between 52 and 62 bp 5' to exon 3, with four subclones containing the entire sequence. The unique sequence, termed exon 1b, begins with a 53-nucleotide 5'-untranslated sequence and ends with a 9-nucleotide sequence that is headed by a potential initiation methionine codon. The 3-codon open reading frame at the end of exon 1b is in-frame with the coding sequence of exon 3. If this ATG codon serves as the translation initiation site, then the AE2b mRNA would encode an N-terminal variant of AE2 in which the first 17 amino acids of AE2a are replaced by an alternative 3-amino acid sequence.

The 5' end of the 3.8-kb AE2c mRNA was cloned using a similar RACE protocol. First strand rat stomach cDNA was synthesized using a primer from exon 8, ligated to the anchor oligonucleotide, and then PCR-amplified using the anchor primer and a primer from exon 6. Fractionation of the RACE products by agarose gel electrophoresis revealed a broad smear containing distinct bands at approximately 320 and 380 bp (Fig. 2, upper left), which were within the expected size range of the AE2c product. Each product was isolated and reamplified, analyzed by agarose gel electrophoresis (Fig. 2, upper right), and then subcloned and sequenced. The two products were identical at their 3' ends, with the common sequence corresponding to exon 6, but they differed at their 5' ends (Fig. 2, bottom panel). The 5' end of the 320-bp product terminated within exon 5 and consisted of sequences identical to common regions of AE2a and AE2b. The most likely explanation for this result is that it was due to premature termination of reverse transcriptase activity during first strand cDNA synthesis. The 380-bp product contained a 210-nucleotide sequence at its 5' end, designated exon 1c, that did not correspond to previously characterized AE2 sequences. Exon 1c does not contain an ATG triplet and would therefore serve as a 5'-untranslated sequence. The first ATG triplet of the 3.8-kb AE2c mRNA, which is in an acceptable context for initiation of translation, occurs in exon 6 and corresponds to Met codon 200 of AE2a. Thus, the AE2 variant encoded by this mRNA would lack the first 199 amino acids occurring in AE2a.

Northern Blot Analysis of AE2 mRNAs in Rat Stomach—To confirm that the unique sequences identified above are present in distinct AE2 mRNAs of the expected sizes, Northern hybridization analysis of rat stomach poly(A)⁺ RNA was performed. A probe corresponding to codons 372–666 of the AE2a cDNA, which should be present in all of the AE2 variants, hybridized with multiple mRNAs ranging in size between 4.4 and 3.8 kb (Fig. 3, lane 1). As expected, a probe corresponding to unique sequences from the 5' end of AE2a hybridized with a distinct 4.4-kb mRNA (Fig. 3, lane 2), and a probe corresponding to unique sequences at the 5' end of AE2b hybridized with a distinct 4.2-kb mRNA (Fig. 3, lane 3). However, the unique AE2c sequence hybridized with both the expected 3.8-kb mRNA (termed AE2c1) and an mRNA of about 4.1 kb (termed AE2c2) that was slightly smaller than the AE2b mRNA. It seemed possible that the additional sequence in the larger AE2c mRNA might be due to retention of a 330-nucleotide sequence from the 3' end of intron 5, which is located between the unique AE2c sequences identified by RACE analysis and exon 6. Hybridization analysis confirmed that the 4.1-kb AE2c2 mRNA contained this intron sequence (Fig. 3, lane 5). Inclusion of this sequence introduces six upstream open reading frames, ranging in length from 3 to 24 codons, including one that overlaps the beginning of the open reading frame of the AE2c1 mRNA.

Intron/Exon Organization of the Rat AE2 Gene—A rat cosmid genomic library was screened using an AE2 cDNA probe, and four clones were identified and purified. Because Southern
FIG. 3. Northern blot analysis of alternative AE2 mRNAs in rat stomach. A blot containing 10 μg of rat stomach poly(A)⁺ RNA was hybridized sequentially with probes either common to all transcripts of the AE2 gene or specific for individual transcripts of the AE2 gene. The probes (see "Experimental Procedures") used for each hybridization were: lane 1, a SacI-PvuII CDNA fragment common to all AE2 mRNAs; lane 2, AE2a-specific probe; lane 3, AE2b-specific probe; lane 4, AE2c-specific probe; lane 5, probe containing sequences from the 3' end of intron 5 and located between exon 1c and exon 6. Autoradiographic exposure times (left to right) were: 3, 6, 5, 14, and 2 days. Sizes of the mRNAs (in kb) are indicated on the right.

blot analysis indicated that all four cosmids contained the entire gene, only one clone, which contained an insert of over 40 kb, was further characterized. Five EcoRI restriction endonuclease fragments that hybridized with the cDNA probes were subcloned into a plasmid vector, and then these subclones were used for analysis of the intron/exon organization. The structure of the AE2 gene, which spans 16.5 kb and contains 25 exons, is diagrammed in Fig. 4. All intron/exon boundaries (Table I) conform to consensus donor or acceptor splice sites (20). The fully processed AE2a mRNA contains 23 exons and includes two exons from the 5' end of the gene that are not present in the AE2b and AE2c transcription units. The transcription units for the AE2b and AE2c1 mRNAs consist of 22 and 19 exons, respectively. The alternative first exon for AE2b, designated exon 1b, is located in intron 2 of the AE2a transcription unit, with its donor splice site positioned 396 nt upstream of the acceptor site of exon 3. The alternative first exon of AE2c, designated exon 1c, is located in intron 5 of the AE2a transcription unit, with the donor splice site used in the AE2c1 mRNA positioned 330 nt upstream of the acceptor site of exon 6. This 330-nt sequence is excised in the AE2c1 mRNA but retained in the AE2c2 mRNA.

Identification of Transcription Initiation Sites for AE2a, AE2b, and AE2c—The 5' end of the AE2a cDNA characterized previously (10) terminated at position 195 relative to the translation initiation site of the mature mRNA, and each of the RACE products analyzed terminated just 3' to this position, suggesting that this region contained the AE2a transcription initiation site. S1 nuclease protection analysis (Fig. 5, left and bottom panels) of rat stomach poly(A)⁺ RNA was performed using a single-stranded probe spanning the region thought to contain the transcription initiation site. A number of protected fragments were observed, with a prominent cluster (indicated by a bracket and asterisk in left panel of Fig. 5) corresponding to positions −224 to −221 relative to the translation initiation site of the mature mRNA. Because this cluster seems to represent the major transcription initiation site, the first nucleotide of the cluster is designated as position +1 of the AE2a transcript. However, additional protected fragments 5' to this site were also observed, suggesting that transcription might also initiate 5' to the major cluster. S1 nuclease analyses using probes that extended further 5' were unsuccessful; therefore, we employed a 5' RACE protocol to confirm the location of the transcription initiation site.

In designing a RACE strategy to identify the 5' end of the AE2a mRNA (diagrammed in Fig. 5, bottom panel), we reasoned that a 14-nt palindromic sequence at positions +53 to +66 might interfere with reverse transcriptase activity. Therefore, the initial cDNA synthesis was carried out using a P1 primer complementary to a region just 5' to the palindromic sequence. A 32P-labeled P2 primer that extended 11 nt further 5' was used for the RACE reaction. Analysis of the labeled RACE products on a polyacrylamide gel revealed the presence of bands corresponding to products terminating between posi-
Fig. 5. S1 nuclease and RACE analysis of the AE2a transcription initiation site. Left panel, nuclease protection was performed using a 5' end-labeled probe, 5 μg of rat stomach mRNA or tRNA control, and 50, 100, or 200 units of S1 nuclease (see "Experimental Procedures"). Samples were analyzed by polyacrylamide gel electrophoresis and autoradiography. Markers in first two lanes are purine (A + G) or pyrimidine (C + T) ladders generated by chemical cleavage sequencing of the probe. Undigested probe is shown in the last lane. Nucleotides are numbered on the left, with +1 indicating the beginning of the major transcription initiation site corresponding to a cluster of four protected fragments (delineated by the bracket and asterisk at the right). Right panel, RACE analysis. cDNA was synthesized using rat stomach mRNA and primer 1, ligated to the anchor oligonucleotide, PCR-amplified using 32P-labeled primer 2 and the anchor primer, and aliquots of the RACE reaction mixture were analyzed by polyacrylamide gel electrophoresis and autoradiography (see "Experimental Procedures"). Numbers on the right indicate the estimated termination site of the extension products, with +1 corresponding to the major initiation site determined by S1 analysis. Size estimates were obtained by running sequencing ladders in adjacent lanes and correcting for the size of the anchor sequence. Products in the bracketed regions labeled 1–5 were excised from the gel, reamplified, subcloned, and sequenced. 70% of the RACE products terminated within the region spanning nt +1 to +4. Bottom panel, S1 nuclease and RACE strategies. The region shown is from the 5' end of the AE2a trancription unit, with # indicating the 5' end of the rat AE2a cDNA (10). The S1 nuclease probe and primers used for RACE analysis are labeled and indicated by lines. The asterisks above and the heavy bars below the sequence indicate major initiation sites identified by S1 nuclease and RACE analysis, respectively. Sequences included in the major transcripts are shown in uppercase letters and 5' flanking sequences are shown in lowercase letters.

Fig. 6. S1 nuclease analysis of the AE2b and AE2c transcription initiation sites. 5' end-labeled probes were hybridized with 2.5 or 5 μg of rat stomach mRNA or 5.0 μg of yeast tRNA, digested with 100 units of S1 nuclease, and analyzed by polyacrylamide gel electrophoresis and autoradiography. Left panel, analysis of AE2b. The ladder shown in the first four lanes was generated by chemical cleavage sequencing analysis of the probe. Undigested probe is shown in the last lane. Nucleotides corresponding to protected fragments (sense strand) are indicated on the left. Right panel, analysis of AE2c. The ladder shown in the first four lanes was generated by chain termination sequencing analysis of the region corresponding to the probe. Nucleotides (sense strand) corresponding to ends of the protected fragments are indicated on the left.

On the basis of the gene characterization and RACE experiments, it seemed likely that the transcription initiation sites for AE2b and AE2c were located in introns 2 and 5 of the AE2a transcription unit, respectively. S1 nuclease protection analysis of the AE2b mRNA using a probe derived from sequences in intron 2 revealed a prominent cluster of seven protected fragments (Fig. 6, left panel), with the largest fragment beginning at a G residue located 62 nt upstream of the donor splice site of exon 1b. Because this residue was also the 5'-most nucleotide identified in four of the seven AE2b RACE subclones analyzed, we conclude that it serves as the transcription initiation site. S1 nuclease analysis of the AE2c transcript confirmed that its transcription initiation site is located in intron 5. Two pro-
tected fragments were observed (Fig. 6, right panel), with their 5' ends corresponding to C and A residues located 211 and 209 nt upstream of the donor splice site of exon 1c. This result correlates well with the RACE analysis, in which the longest subclone extended 210 nt upstream of the donor site.

Analysis of Genomic Sequences Flanking the Transcription Initiation Sites—

The sequences of the alternative first exons for each of the AE2 mRNAs and the 5' flanking region for each transcription unit are shown in Fig. 7. The first two exons of the AE2 gene, which are unique to the AE2a mRNA (Fig. 7, upper panel), are separated by a 2-kb intron. The sequence immediately 5' to the AE2a transcription initiation site is GC-rich and contains few sequences resembling known basal promoter elements. It lacks a TATA element (21) but contains several potential CACCC (22, 23) elements. The beginning of the alternative first exon of AE2b is located 1.15 kb downstream of exon 2 (Fig. 7, middle panel). The nucleotide sequence surrounding the AE2b transcription initiation site closely matches a CAP site consensus sequence (21). The promoter region lacks an apparent TATA element, but contains four potential CCAAT boxes (21), a potential Sp1 binding site (21), and six CACCC elements (22, 23). The transcription initiation site for AE2c is located approximately 550 nt downstream of exon 5 (Fig. 7, bottom panel) and is preceded by an AT-rich sequence at nt 230 to 224 that might serve as a TATA element. The 5' flanking sequence also contains a potential Sp1 binding site and several potential CACCC and CCAAT elements. The spliced donor site for exon 1c and the acceptor site of exon 6, which contains the apparent translation start site, are separated by a 330-nt sequence that is excised in the AE2c1 mRNA but retained in the AE2c2 mRNA.

Tissue Distribution of AE2 mRNAs—

When a Northern blot containing poly(A)+ RNAs from rat tissues was hybridized with a cDNA probe corresponding to a common region of the AE2 variants, mRNAs of approximately 4.2–4.4 kb were detected in all tissues, with an additional mRNA of approximately 3.8 kb in stomach (Fig. 8, top panel). When probes specific for each variant were used, the 4.4-kb AE2a mRNA was identified in all tissues, with the highest levels in stomach, large intestine, lung, and uterus. The 4.2-kb AE2b mRNA was detected in a more limited set of tissues, with relatively high levels in stomach, moderate levels in small intestine, large intestine, liver,
Fig. 8. Tissue distribution of AE2 mRNAs. A Northern blot with 5 μg of poly(A)- RNA from the indicated rat tissues was analyzed as described under “Experimental Procedures” with the following probes (from top to bottom): Sac-PvuI fragment common to all AE2 mRNAs, AE2a-specific probe containing sequences from the first two exons of AE2a, AE2b-specific probe containing sequences from exon 1b, and AE2c-specific probe containing sequences from exon 1c. Autoradiographic exposure times (from top to bottom) were 6, 8, 12, and 10 days, respectively. The position of a 4.4-kb size marker is shown on the right.

and kidney, and low levels in lung, brain, and uterus. The 3.8- and 4.1-kb AE2c mRNAs were detected only in stomach.

DISCUSSION

Use of Alternative Promoters Leads to Tissue-specific Expression of mRNAs Encoding N-terminal Variants of AE2—Earlier studies of the AE Cl-/HCO3- exchanger gene family revealed the existence of multiple mRNAs for each of the three known genes (9, 10, 24, 25), and more recent analyses of the AE1 and AE3 genes have shown that some of these mRNAs are transcribed from alternative promoters and encode protein variants that differ in their N-terminal sequences (5–8, 11–13, 25–27).

In the current study we have shown that similar mechanisms, involving alternative exon and promoter usage, are responsible for the tissue-specific expression of mRNAs encoding three N-terminal variants of AE2.

One of these variants is AE2a, a 1234-amino acid protein encoded by a 4.4-kb mRNA that is transcribed from the 5′-most promoter. Mapping of the transcription start site was hampered by palindromic sequences and the GC-richness of the 5′-flanking sequence. Nevertheless, S1 nuclease and RACE analyses show that the major site of initiation occurs at nt +1 to +4 (Figs. 5 and 7) and that a lower level of initiation occurs within the 11-nt sequence preceding this cluster. Some protection of the full-length S1 nuclease probe at nt –15 was observed; however, none of the 48 RACE products analyzed extended to this site. This suggests that the resistance of this region to S1 nuclease may have been due to formation of a hairpin secondary structure in the small portion of the probe not protected by the AE2a mRNA, although the possibility of a low level of transcription initiation occurring at these sites or further 5′ cannot be ruled out. The 5′-flanking region of AE2a lacks the more common basal promoter elements, but does contain several CACCC sequences (22, 23) in both orientations and, in this respect, is similar to the 5′-flanking region of erythrocyte AE1 (7, 28). The AE2a promoter is active in most, if not all, mammalian tissues, in contrast to both the erythrocyte AE1 promoter (28), which is highly tissue-specific, and the brain AE3 promoter, which is moderately tissue-specific (10). The apparently ubiquitous expression of the AE2a variant (Fig. 8) suggests that it serves a housekeeping function in many cell types.

The second variant, AE2b, is a 1220-amino acid protein that contains an alternative 3-amino acid N-terminal sequence that replaces the first 17 amino acids of AE2a. The results of S1 nuclease analysis correlated well with the results of RACE analysis, and demonstrated that the 4.2-kb AE2b mRNA is transcribed from an alternative promoter located in intron 2. The sequence immediately surrounding the transcription initiation site closely matches the consensus CAP signal frequently occurring at initiation sites (21). A potential CCAAT element occurs 65 nt upstream, which is within the preferred region for this element between nt –57 and –212 (21). Additionally, CCAAT sequences occur further upstream, but their functional significance is questionable as they are outside the preferred region for functional CCAAT elements. CACCC sequences and a potential Sp1 binding site were also observed. The AE2b mRNA is expressed in a more limited set of tissues than AE2a (Fig. 8), with the highest levels in stomach, consistent with the possibility that AE2b serves more specialized physiological functions than the ubiquitous AE2a.

The third variant, AE2c, is a 1035-amino acid protein that is identical to AE2a except that it lacks the first 199-amino acids. It is encoded by two mRNAs, 3.8 and 4.1 kb in length, that are transcribed from a promoter located in intron 5. The 4.1-kb mRNA, which retains the intron sequence between the donor splice site of exon 1c and the acceptor site of exon 6, contains multiple upstream open reading frames, which might interfere with translation of the long open reading frame encoding AE2c. Because of this, it is unclear whether it is a functional mRNA or an incompletely processed mRNA. A number of cases have been reported in which incomplete processing of mRNAs serves as a regulatory mechanism (reviewed in Ref. 29), and the existence of the 4.1-kb AE2c2 mRNA raises the possibility that this mechanism might be used in the regulation of AE2c expression. The results of S1 nuclease and RACE analysis showed that the AE2c2 transcription start site is located 209–211 nt upstream of the donor splice site of exon 1c. An AT-rich sequence that might serve as a TATA element begins at position –30 relative to the transcription initiation site. This is preceded by a potential CCAAT element at position –188, which is within the preferred region for such elements, and an inverted CACCC sequence at position –59. The AE2c promoter is highly tissue-specific, with expression of the AE2c mRNAs being observed only in stomach (Fig. 8), suggesting that this variant might serve a cell-type or organ-specific function.

Intron/Exon Organization and Possible Evolutionary Relationships among the AE Genes—The intron/exon organization of the AE2 gene is remarkably similar to that of the other AE genes (Fig. 9), with the greatest similarity between AE2 and AE3. If the alternative exons transcribed from internal promoters are excluded from consideration, both genes have 23 exons. For each gene the first exon consists of 5′-untranslated sequence and the second exon contains the translation start site and first 17 codons. There are 21 additional exons in each gene, with the last exon containing a conserved C-terminal coding sequence and the 3′-untranslated sequence. The positions of the transcription initiation site for the long form of AE3 has not yet been identified. Analysis of sequences extending several hundred bp upstream of the known 5′-most exon of AE3 do not reveal consensus acceptor splice sites (S. Linn and G. E. Shull, unpublished data), and the relative sizes of the brain and cardiac AE3 mRNAs suggest that there is little, if any, additional sequence in the larger mRNA besides that present in the cDNA. Therefore, in the discussion and Fig. 9 we have assumed that the 5′-most exon that was identified in the AE3 gene (11) is exon 1. However, the reader should be aware that the possibility of additional upstream exons has not been rigorously excluded.
a separate evolutionary origin from the corresponding region of the AE2 and AE3 genes. One possibility is that the differences at the 5' ends of these genes might have arisen during the chromosomal rearrangement events that were responsible for dispersing these genes in the mammalian genome. For example, the current structure of either the AE1 gene or the AE2 and AE3 genes might have resulted from a chromosomal rearrangement in which the last 16 exons of an ancestral gene were recombined with the promoter and 5' coding exons of another gene.

The patterns of alternative promoter and exon usage of the AE2 gene are analogous to those of the AE1 and AE3 genes (Fig. 9), although the locations of the internal promoters and the consequent variations in protein structure are different. Transcription from the internal promoters leads to a switch in N-terminal amino acid sequence or to a truncation of the protein, depending on whether the alternative first exon contains coding sequence or consists entirely of untranslated sequence. In the latter case an internal Met codon in a downstream exon serves as the translation initiation site. The internal promoters for AE2b and AE2c are located in exons 2 and 5 of the AE2a transcription unit, respectively, whereas the kidney AE1 promoter is located in intron 3 and the cardiac AE3 promoter is located in intron 6. It is clear from their positions within each gene that the promoters, and associated first exons, for AE2b, AE2c, and cardiac AE3 arose independently of each other during evolution, but it is less apparent whether the promoters and first exons for kidney AE1 and cardiac AE3 arose independently, as they occupy the same positions relative to the 16 highly conserved exons. However, the differences in tissue specificity and the absence of significant similarity between exon 4 of AE1 and exon 7 of AE3, which lie between the alternative first exon and the highly conserved exons, argue against the possibility that these promoters share a common evolutionary origin.

Potential Functions of the Alternative AE2 Promoters—The biological rationale for the evolutionary development of alternative promoters in the AE2 gene has not been established, but a gene regulatory role is suggested by the observed differences in tissue specificity. For example, the high levels of AE2 mRNA in stomach is due, in large part, to expression of the AE2b mRNA at much higher levels in stomach than in other tissues, and to expression of the stomach-specific AE2c mRNAs. Thus, the AE2b and AE2c promoters may have evolved, in part, as a genetic mechanism for producing the high Cl⁻/HCO₃⁻ exchanger levels that are required in stomach.

Because use of the alternative AE2 promoters leads to variations in N-terminal amino acid sequences, it is likely that they also serve as a mechanism for producing protein variants with altered, and biologically relevant, functional characteristics. Generation of the AE2b and cardiac AE3 variants are analogous in that each involves a switch in the N-terminal amino acid sequence relative to the long forms of each exchanger. The extent of the sequence alterations, however, are quite different. The differences between AE2a and AE2b are restricted to the extreme N-terminal sequence encoded by the first coding exon. In contrast, transcription from the cardiac AE3 promoter results in the replacement of a 270-amino acid N-terminal sequence of brain AE3, which is encoded by the consequent variations in protein structure are different.

Fig. 9. Intron/exon organization, promoter usage and splicing patterns of the AE gene family. Patterns of promoter usage and splicing that lead to the generation of multiple AE1, AE2, and AE3 mRNAs are indicated. Exons are numbered below and shown as boxes, with hatched regions indicating untranslated sequences and open regions indicating coding sequences. Exons 1k of AE1 and 1c of AE3 are the alternative first exons for kidney AE1 and cardiac AE3. Exons 1a, 1b, and 1c are alternative first exons for the AE2 variants discussed in text. Transcription initiation sites (see Footnote 3) are indicated with arrows and translation initiation sites with ATG. Homologous exons of the AE2 and AE3 genes are aligned; exons 5–20 of AE1 and exons 8–23 of AE2 and AE3 are also homologous. Left column, AE1e, erythrocyte AE1 mRNA (4, 7, 19); AE1k1, kidney AE1 mRNA with alternative first exon spliced to exon 4, corresponding to mouse and human kidney mRNAs and a minor rat kidney mRNA (5, 7, 8); AE1k2, major rat kidney mRNA in which intron 3 sequences are retained (5, 7). AE3b, mRNA encoding brain form of AE3, which is also expressed in some other tissues (9, 10); AE3c, mRNA encoding cardiac form of AE3 (11–13). AE2 mRNAs are described in text. Right column, number of codons in long open-reading frame of each mRNA.
form of the exchanger in which Met codon 200 serves as the apparent initiation codon. In this respect the AE2c mRNA is analogous to the kidney AE1 mRNA, which also encodes an N-terminal truncated exchanger that utilizes an internal Met codon as a translation start site. At the protein level the differences between AE2c and AE2a are similar to the differences between cardiac AE3 and brain AE3. AE2c is approximately the same size as cardiac AE3, and the truncation of its N-terminal sequence removes the histidine-rich domain (residues 74–88), the proline-rich regions, and the stretches of acidic (residues 122–130) and basic amino acids (residues 94–109) that are homologous to those eliminated in cardiac AE3. Although the functions of these domains have not been determined, it is likely that they play a role in regulating the activity of the exchanger.

Replacement of the 17-amino acid N terminus of AE2a with the 3-amino acid N terminus of AE2b is a relatively limited sequence alteration compared with those seen in the other AE variants. However, the unique N terminus of AE2a contains a potential phosphorylation site for cAMP-dependent protein kinase at serine 10. An increase in cAMP is known to inhibit the absorption of NaCl across the apical membranes of intestinal epithelial cells (30). This process involves the coupled activities of a Cl⁻/HCO₃⁻ exchanger, which appears to be the AE2a variant (16), and a Na⁺/H⁺ exchanger (16) that is thought to be NHE3 (31). Although we are unaware of any evidence demonstrating that cAMP causes a direct inhibition of AE2 in intestine, cAMP has been shown to inhibit Cl⁻/HCO₃⁻ exchange in osteoblasts (32). The identity of the exchanger in osteoblasts was not determined, but it is conceivable that it is the ubiquitous AE2a. In stomach, secretion of acid is stimulated by cAMP (33), but cAMP does not alter the transport capacity of the basolateral Cl⁻/HCO₃⁻ exchanger in the parietal cell (34), which is known to be a variant of AE2 (18). If Ser-10 of AE2a does serve as an inhibitory phosphorylation site, then the elimination of this site in AE2b and AE2c could serve an important function during acid secretion.

A second possibility that should be considered is that the variant N-terminal sequences contain sorting signals or cytoskeletal attachment sites that influence the membrane location of the exchanger in polarized epithelial cells. Such a possibility is suggested by the recent demonstration that some variants of chicken AE1 are sorted to the plasma membrane when expressed in human erythroleukemia cells, whereas variants of chicken AE2 are sorted to skeletal attachment sites that influence the membrane location, as in AE2a and AE2b (18). If Ser-10 of AE2a might localize this variant to the apical membrane in certain polarized epithelial cells, and that elimination of this sequence, as in AE2b and AE2c, might result in sorting to the basolateral membrane. If future experiments prove this hypothesis to be correct, it would not rule out the possibility that the same variant can be sorted to different membranes depending on the cell-type in which it is expressed, but it would provide an explanation of at least one mechanism by which AE2 can be localized to either apical or basolateral membranes. Also, it would implicate AE2b as a possible candidate for the Cl⁻/HCO₃⁻ exchanger that functions on the basolateral membranes of epithelial cells of renal proximal tubules (36), the thick ascending limb (37), and β-intercalated cells of the cortical collecting duct (38).

REFERENCES

1. Alper, S. L. (1994) Cell. Physiol. Biochem. 4, 265–281
2. Jennings, M. L. (1992) in The Kidney: Physiology and Pathophysiology (Seldin, D. W., and Giebisch, G., eds) 2nd Ed., pp. 503–535, Raven Press, Ltd., New York
3. Kopito, R. R. (1990) Int. Rev. Cytol. 123, 177–199
4. Kopito, R. R., and Lodish, H. F. (1985) Nature 316, 234–238
5. Kudrycki, K. E., and Shull, G. E. (1989) J. Biol. Chem. 264, 8185–8192
6. Brosius, F. C., Ill, Alper, S. L., Garcia, A. M., and Lodish, H. F. (1989) J. Biol. Chem. 264, 7784–7787
7. Kudrycki, K. E., and Shull, G. E. (1993) Am. J. Physiol. 264, F540–F547
8. Kollet-tons, A., Wagner, S., Hubner, S., Appelhans, H., and Drendelkahn, D. (1993) Am. J. Physiol. 265, F813–F821
9. Kopito, R. R., Lee, B. S., Simmons, D. M., Lindsey, A. E., Morgan, C. W., and Schneider, K. (1989) Cell 59, 927–937
10. Kudrycki, K. E., Newman, P. R., and Shull, G. E. (1990) J. Biol. Chem. 265, 462–471
11. Linn, S. C., Kudrycki, K. E., and Shull, G. E. (1992) J. Biol. Chem. 267, 7927–7933
12. Yannoukakos, D., Stuart-Tilley, A., Fernandez, H. A., Fey, P., Duyk, G., and Alper, S. L. (1994) Circ. Res. 75, 603–614
13. Linn, S. C., Askew, G. R., Menon, A. G., and Shull, G. E. (1995) Circ. Res. 76, 584–591
14. Alper, S. L., Kopito, R. R., Libresco, S. M., and Lodish, H. F. (1988) J. Biol. Chem. 263, 17092–17099
15. Lindsey, A. E., Schneider, K., Simmons, D. M., Baron, R., Lee, B. S., and Kopito, R. R. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 5278–5282
16. Chow, A., Dobbins, J. W., Aronson, P. S., and Igarashi, P. (1992) Am. J. Physiol. 263, G345–G352
17. Gehrig, H., Muller, W., and Appelhans, H. (1992) Biochim. Biophys. Acta 1130, 326–328
18. Stuart-Tilley, A., Sarstedt, C., Pouysseguer, J., Schwartz, M. A., Brown, D., and Alper, S. L. (1994) J. Biol. Chem. 269, 17792–17796
19. Kudrycki, K. E., and Shull, G. E. (1993) Am. J. Physiol. 264, C559–C568
20. Kudrycki, K. E., Andresson, M., and Lodish, H. F. (1987) J. Biol. Chem. 262, 8035–8040
21. Mount, S. M. (1982) Nucleic Acids Res. 10, 459–472
22. Bucher, P. (1990) J. Biol. Chem. 265, 536–578
23. Schale, R., Muller, M., Otsuka-Murakami, H., and Renkawitz, R. (1988) Nature 332, 87–90
24. Frampton, J., Walker, M., Plumb, M., and Harrison, P. R. (1990) Mol. Cell. Biol. 10, 3836–3842
25. Alper, S. L., Kopito, R. R., and Lodish, H. F. (1987) Kidney Int. 32, (Suppl. 23), 5117–5128
26. Kim, H.-R. C., Kennedy, B. S., and Engel, J. D. (1989) Mol. Cell. Biol. 9, 5198–5206
27. Cox, K. H., and Cox, J. V. (1995) Am. J. Physiol. 268, F503–F513
28. Cox, K. H., Adair-Kirk, T. L., and Cox, J. V. (1995) J. Biol. Chem. 270, 19752–19760
29. Kopito, R. R., Andersson, M., and Lodish, H. F. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 7419–7423
30. Kozak, M. (1988) J. Cell Biol. 107, 1–7
31. Binder, H. J. (1989) in Gastrointestinal Disease: Pathophysiology, Diagnosis, Management (Sliezinger, M. H., and Fordtran, J. S., eds) Vol. 2, 4th Ed., pp. 1022–1045, W. B. Saunders Co., Philadelphia
32. Bookstein, C., Depaoli, A. M., Xie, Y., Nul, P., Musch, M. W., Rao, M. C., and Chang, E. B. (1994) J. Clin. Invest. 93, 106–113
33. Green, J., and Kleeman, C. R. (1992) Am. J. Physiol. 262, C111–C121
34. Malinaowska, D. H., Sachs, G., and Cuppelli, J. (1988) Biochim. Biophys. Acta 972, 95–109
35. Seidler, U., Roithmaier, S., Classen, M., and Silen, W. (1992) Am. J. Physiol. 261, G81–G91
36. Rathjen, P. D., Toth, S., Willis, A., Heath, J. K., and Smith, A. G. (1990) Cell 62, 1105–1114
37. Alpern, R. J., and Chambers, M. (1987) J. Gen. Physiol. 89, 581–598
38. Hebert, S. C. (1986) Am. J. Physiol. 250, C290–C391
39. Weiner, I. D., Weil, A. E., and New, A. R. (1984) Am. J. Physiol. 250, F952–F964
Three N-terminal Variants of the AE2 Cl/HCO Exchanger Are Encoded by mRNAs Transcribed from Alternative Promoters
Zhuo Wang, Patrick J. Schultheis and Gary E. Shull

J. Biol. Chem. 1996, 271:7835-7843.
doi: 10.1074/jbc.271.13.7835

Access the most updated version of this article at http://www.jbc.org/content/271/13/7835

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 37 references, 15 of which can be accessed free at http://www.jbc.org/content/271/13/7835.full.html#ref-list-1