EEG1, a putative transporter expressed during epithelial organogenesis: comparison with embryonic transporter expression during nephrogenesis

ROBERT O. STUART,1,2 ANNA PAVLOVA,4 DAVID BEIER,4 ZHIXING LI,2 YELENA KRIJANOVSKI,4 AND SANJAY K. NIGAM2,3,4
1Veterans Affairs San Diego Healthcare System, 2Division of Nephrology and Hypertension, Department of Medicine and 3Pediatrics, 4Cancer Center, University of California San Diego, La Jolla, California 92093; and 4Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts 02115

Received 9 April 2001; accepted in final form 23 July 2001

EEG1, a putative transporter expressed during epithelial organogenesis: comparison with embryonic transporter expression during nephrogenesis. Am J Physiol Renal Physiol 281: F1148–F1156, 2001.—A screen for genes differentially regulated in a model of kidney development identified the novel gene embryonic epithelia gene 1 (EEG1). EEG1 exists as two transcripts of 2.4 and 3.5 kb that are most highly expressed at embryonic day 7 and later in the fetal liver, lung, placenta, and kidney. The EEG1 gene is composed of 14 exons spanning a 20-kb region at human chromosome 11p12 and the syntenic region of mouse chromosome 2. Six EEG1 exons have previously been assigned to a longer isoform of eosinophil major basic protein termed proteoglycan 2. Another gene distantly related to EEG1, POV1/PB39, is located 88 kb upstream from the EEG1 gene on chromosome 11. Temporal expression of 65 members of the solute carrier (SLC)-class of transport proteins was followed during kidney development using DNA arrays. POV-1 and EEG1, like glucose transporters, displayed very early maximal gene expression. In contrast, other SLC genes, such as organic anion and cation transporters, amino acid permeases, and nucleoside transporters, had maximal expression later in development. Thus, although the bulk of transporters are expressed late in kidney development, a fraction are expressed near the onset of nephrogenesis. The data raise the possibility that EEG1 and POV1 may define a new family of transport proteins involved in the transport of nutrients or metabolites in rapidly growing and/or developing tissues.

microarray; organogenesis; bioinformatics; embryonic epithelia gene 1

DIVERSE EPITHELIAL TISSUES appear to share a core developmental program, which manifests as an ability to form tight sheets of cells that can be organized into hollow tubes that serve as the interface between physiological compartments and that are specialized for the transport of various substances. The search for developmentally important genes in epithelial and other embryonic organs is hampered by the multiplicity of cell types and lack of temporal synchronization. Kidney development is characterized by the interactions of two primordial tissues: the metanephric mesenchyme (MM) and an epithelial component termed the ureteric bud (UB). The UB is induced to undergo many rounds of branching morphogenesis by factor(s) produced by the MM (15). The MM in turn is induced to undergo a mesenchymal to epithelial transformation in response to factor(s) produced by the UB. A well-characterized system utilizing two cultured cell lines derived from embryonic kidney reproduces in vitro certain aspects of this developmental program (13). UB cells (representing the epithelial component of the embryonic kidney) are cultured in a three-dimensional extracellular matrix and subsequently are exposed to the conditioned media from BSN cells (representing their mesenchymal component). The UB cells subsequently enter a morphogenetic program, which proceeds though cellular processes, branching multicellular cords, and eventually branching multicellular tubular structures with lumens (13). The various stages, i.e., processes, cords, and tubules, are associated with distinct patterns of gene expression (7, 12). Here, we describe the cloning, chromosomal localization, and characterization of embryonic epithelia gene 1 (EEG1), a putative transport gene that is differentially expressed in the cell culture model and in a variety of embryonic epithelial tissues. We also compared its developmental expression pattern to a large number of members of the major facilitator class of membrane transporters during kidney development.

METHODS

UB cell tubulogenesis assay. The induction of branching morphogenesis in UB cells has been described in detail (13). Briefly, the “tubulogenic” condition consisted of UB cells suspended in a three-dimensional extracellular matrix consisting of an 80:20% mixture of collagen-I and Matrigel and

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

http://www.ajprenal.org
RESULTS

Identification and cloning. A number of methodologies exist for the identification of differentially expressed sequences. A few such techniques, including differential display (dd) PCR, microarrays, serial analysis of gene expression, and subtractive hybridization, allow for the identification of novel and/or unclassified transcripts. Although the material requirements for DNA array analysis are steadily shrinking, ddPCR is particularly suited for the analysis of very small samples and for the identification of sequences that are unavailable on arrays. We employed ddPCR in a search for up- as well as downregulated transcripts in the UB-BSN cell model of kidney development. Equivalent aliquots of total RNA were isolated from UB cells under various "tubulogenic" and "nontubulogenic" conditions as described in METHODS. A number of bands were observed to increase. However, a small number were observed to decrease in response to BSN-CM (Fig. 1), and one of these, a 279-bp amplicon, was employed as a probe for cDNA isolation using the GeneTrapper kit (Life Technologies). One 2.4 kb cDNA, initially termed 617e1, was isolated from an adult mouse kidney library.

Sequence analysis. Sequence analysis revealed a number of overlapping EST database (dbEST) matches, one of which, AI588018, was identical to the original clone with the exception of an intron deletion (9). Another human EST, AL157431, was the full length cDNA of the human ortholog (18). On the basis of high expression in embryonic epithelial tissues (see below) we termed this gene, EEG1, for expressed in "Embryonic Epithelia Gene 1." Here, the murine form is referred to as mEEG1 and the human ortholog as hEEG1. Homology searches also yielded a single high-scoring match

![Fig. 1. Differential display. Total RNA was obtained from ureteric bud (UB) cells in 3D culture under various conditions of cytokine and/or extracellular matrix exposure, which resulted in varying degrees of branching morphogenesis in the UB/extracellular matrix (BSN) cell system. Equal aliquots were reverse transcribed and used as a template in a ddPCR-based search for genes differentially regulated in this model. Most amplicons either showed no change in response to serum or the conditioned media from the mesenchymal cell line (BSN cells), BSN-stimulated with conditioned media (CM) (band 1) or were uniformly up- or downregulated by addition of serum or BSN-CM (band 2). One amplicon displayed the unusual property of differential downregulation by BSN-cm, an effect not seen with 10% FCS (band 3), and was prioritized for cloning.](http://ajprenal.physiology.org/)

AJP-Renal Physiol • VOL. 281 • DECEMBER 2001 • www.ajprenal.org
in the nonredundant database belonging to proteoglycan-2 (PRG2). PRG2 has been described as a longer isoform of eosinophil major basic protein (MBP) produced from transcription at an alternate upstream promoter (6). Another previously cloned gene of unknown function, POV1/PB39 (1), displays 28% amino acid identity and 44% amino acid similarity to the EEG1 translation. The degree of similarity between EEG1 and POV1 is consistent with that seen between distantly related members of a gene family, e.g., transporters, such as hOAT1 and hOCT1 (32% identity, and 49% similarity), isoforms of the human organic anion, and cation transporters, respectively (4, 8).

The four genes shared some degree of similarity at either the amino acid or nucleotide level. Radiation hybrid mapping of the mEEG1 3'-UTR placed the mEEG1 gene on mouse chromosome 2 [22.56 cR from D2Mit126, lod >10.0], in a region with conservation of syntenic with human chromosome 11p11–12. hPOV1, PRG2, and MBP also map to this human chromosomal region, indicating the possibility that EEG1 shared exons in common with PRG2 (1, 6). It remained extremely unlikely that the EEG1 sequence was a cloning artifact, given the fact that multiple overlapping dbEST entries spanned its entire length. Nevertheless, contiguous overlapping PCR amplicons spanning the entire EEG1 transcript were generated from whole mouse embryonic cDNA and confirmed by sequence (Fig. 2A).

Recent human genome sequencing efforts raised the possibility of resolving confusion regarding the precise relationships of these genes (4a, 17). The Celera chromosome scaffold (GA_x2HTBL4CBQV) contained 500 kb of human chromosome 11 sequence. With the chro-

![Fig. 2](http://ajprenal.physiology.org/)

**Fig. 2.** Embryonic epithelia gene 1 (EEG1) molecular analysis. A: overlapping PCR amplicons corresponding to the entire putative murine (m)EEG1 transcript were amplified from e12 whole mouse embryo cDNA and sequenced. RACE amplicons (5') demonstrated the presence of one 38-bp retained intron in the original sequence. In the absence of this intron sequence, the predicted transcription initiation site occurs at the first ATG codon and conforms to Kozak’s rules (5). B: genomic structure of the human (h)EEG1 gene on human chromosome 11. The mouse EEG1 and human sequence derived from expressed sequence tag (EST) AL157431 shared considerable sequence overlap with a previously characterized gene, proteoglycan 2 (PRG2). PRG2 was, in turn, a putative longer isoform of another nearby gene, major base protein (MBP). To resolve the relationships between potentially 3 genes, the relevant Celera chromosome 11 sequence was investigated for exons via sequence comparisons to the RNA. hEEG1 is constructed from 14 exons spanning 20 kb of genomics on the reverse strand of genomic scaffold GA_x2HTBL4CBQV (shown in reverse orientation for clarity). The canonical MBP gene containing the coding region is located ~36 kb downstream and shares no sequence in common with EEG1. Arrows correspond to coding regions of the EEG1, MBP, and PRG2 transcripts. C: structure of the PRG2 transcript. A chimeric transcript containing 6 exons from EEG1, 2 alternative exons (5 and 8), and the 5 coding exons from MBP have been described in human eosinophils and bone marrow. A transcript spanning the putative EEG1/MBP boundary is not represented in the dbEST. Of more than 100 EST clones representing the MBP sequence, not one extends more 5' than exon 10 “EST cutoff.” Furthermore, the PRG2 transcript as described is predicted to be translated as a protein distinct from both MBP and EEG1 [open reading frame (orf1)]. Plain numbers correspond to EEG1 exons defined here. Prime numbers correspond to PRG2 exons as defined previously (6).
mosome 11 sequence, it was possible to select a 55-kb region containing the hEEG1, PRG2, and MBP gene regions (Fig. 2B). Twenty-two exons and their relationships to the mRNA sequences were defined. hEEG1 is composed of 14 exons covering a 20-kb region. The unambiguous MBP gene containing the coding region and corresponding to more than 100 dbEST entries is located 36 kb downstream and shares no sequence in common with hEEG1. The PRG2 sequence contains the coding exons from MBP and eight alternative exons, as previously described (6).

Several lines of evidence suggested that the PRG2 sequence is extremely rare or is observed only in particular circumstances. Of more than 100 dbEST entries showing high homology to PRG2, not a single entry shows sequence more 5′ than exon 10 when the PRG2 sequence is used as the query (Fig. 2C). Furthermore, the putative PRG2 transcript would not likely code for MBP because a 5′ open reading frame (ORF) exists and would likely be preferentially translated (5). A second open reading frame (ORF) exists that would result in a hybrid EEG1/MBP molecule, which does not exist as an identifiable dbEST entry. Nevertheless, a PRG2 transcript has been reported in immature human eosinophils, a tissue source not well represented in the dbEST (6).

Although EEG1 conceivably shares exons with nearby genes in certain special contexts, it represents a distinct RNA species. Northern analysis revealed the presence of two transcripts of 2.4 and 3.5 kb in the mouse (Fig. 3). The entire 2.4-kb mEEG1 transcript contained two retained introns, one of 350 bp between exons 3 and 4, and another (also present in EST A1588088) of 38 bp. The precise chromosomal location of this intron identified through 5′RACE remains undefined, as the mouse chromosomal sequence is unavailable. The putative, fully processed mouse RNA species contains 2,004 nucleotides and a single ORF of 1,392 nt, specifying a protein of 464 amino acids. The putative transcription start site at position 323 (362 in the Genbank entry) represents the first ATG in the sequence and conforms to Kozak’s rules for translation initiation (CAGACCATGCGAA) (Fig. 2A) (5). The human ortholog specified by AL157431 contains a 1,473-bp ORF specifying a 491-amino acid protein. The mouse and human EEG1 transcripts differed significantly in two regions. In predicted exon 10, mouse EEG1 contains a series of cag-repeats specifying 14 contiguous glutamine residues interrupted by one glutamate residue. In addition, the original mouse 617e1 clone terminated in a poly-A tail after predicted exon 13, thus eliminating exon 14, which encodes the 3′UTR and 33 terminal amino acids.

hEEG1 and hPOV1 share 44% amino acid similarity concentrated in two long sequence intervals representing over 50% of their respective lengths. Their close proximity on human chromosome 11 suggests an ancient gene duplication event. Both hEEG1 and hPOV1 are predicted by hydropathy profile to contain multiple transmembrane domains (Fig. 4, A and B). In the case of hPOV1, 12 transmembrane domains are predicted, and in the case of hEEG1, 10 are predicted. Comparison of the human EEG1 hydropathy profile with those of all known or predicted proteins (http://bioinformatics.weizmann.ac.il/hydrop/) yields a highest scoring match with the human folate-like transporter (Swiss-Prot O60779) with which hEEG1 shares 18% amino acid identity over the length of the sequence (Fig. 4C). The remaining high-scoring matches were all sugar-transporting proteins. The high-scoring match for hPOV1 was a hypothetical C. elegans protein, YSPK (Swiss-Prot Q19425) (Fig. 4D). The PSORT2 program (http://bioweb.pasteur.fr/seqanal/interfaces/psort2.html) predicted a possible NH2-terminal signal peptide from amino acids 1–29, further suggesting that hEEG1 is a membrane protein.

Expression of EEG1. The 2.4- and 3.5-kb transcripts revealed similar patterns of expression on Northern analysis (Fig. 3). Both are highly expressed in the early embryo; expression is very high at embryonic day 7 compared with later times. A 3′-probe was generated that contained only an EEG1-specific sequence whereas the EST probe contained sequence shared with the putative PRG2 transcript. No difference in expression pattern was noted. Expression in the adult was strongest in the heart followed by the lung, liver, spleen, and kidney. In situ hybridization using the 3′-probe revealed intense signal from the placenta in e9.5 mouse embryos (Fig. 5). In addition, the whole of the mesenchymal region representing presumptive liver, spleen, and kidney showed expression of EEG1. By embryonic day 12, EEG1 is found primarily in the liver and lung, a pattern that continues through at least day 16. In addition, expression was noted in the kidney cortex at this time (Fig. 5).

We have also investigated the expression of some 8,740 genes utilizing the Affymetrix RG-U34A GeneChip during rat kidney development (14). Stand-alone blast investigations of the RG-U34A target sequences revealed that the EEG1 sequence is not represented on this DNA array, even as an uncharacterized EST se-
sequence. However, we have already described (Figs. 4 and 5) whole embryonic mouse Northern data and in situ hybridization data, indicating that EEG1 has a high expression in the early embryo including the kidney, followed by a marked decline over the course of development. On the other hand, the rat POV1 ortholog is represented. POV1 expression is highest in the developing rat kidney at the onset of organogenesis.

Fig. 4. EEG1 and POV1 hydropathy. Kyte-Doolittle hydropathy plot and transmembrane segment prediction for hEEG1 (A) and hPOV1 (B). Both hEEG and hPOV1 demonstrate multiple predicted membranes spanning domains according to the TMPred program (http://www.ch.embnet.org/software/TMPRED_form.html). The number and location of the 12 predicted domains for POV1 conform to that expected for major facilitator transport proteins (MFP). The EEG1 sequence has 10 predicted transmembrane domains. In addition, the EEG1 translation conforms to the Pfam definition for “sugar transporter.” Comparisons of the hEEG1 (C) and hPOV1 (D) hydropathy profiles against all known or putative proteins (bioinformatics.weizmann.ac.il/hydroph/) appear. The highest scoring match with hEEG1 was the human folate-like transporter (hFLOH); all other matches were sugar transporters. The highest scoring match with hPOV1 was an uncharacterized C. elegans protein, YSPK. Thus multiple lines of evidence suggest that EEG1 and POV1 are related transmembrane transport proteins.

Fig. 5. Embryonic tissue expression by in situ hybridization (e9.5 mouse embryo). mEEG1 was highly expressed in placenta (P) and in the mesenchymal region (M) containing the presumptive liver, gut, and kidneys (e12.5 mouse embryo). High expression was found in the fetal liver with lesser expression noted in the lung. Faint expression was also widely noted in tissues excluding the central nervous system (e16.5 mouse embryo). Intense expression was again noted in liver and lung with somewhat lesser expression in the kidney cortex. Note the relatively low expression in fetal heart compared with findings on Northern analysis of adult tissue. Probe was the 3’ EEG1-specific sequence (Fig. 2A). Sense controls yielded barely detectable signals (not shown).
(e13.5) and decreases linearly with advancing embryonic and postnatal age (Fig. 6). On the basis of hydropathy similarity, we hypothesized that EEG1 and POV1 are novel members of the major facilitator class of membrane transport proteins (MFP). We therefore sought to compare the temporal expression in the developing kidney of POV1 with other members of this class.

All human MFP were identified, and corresponding probe sets on the RG-U34A GeneChip were identified on the basis of sequence similarity to solute carrier (SLC) genes in the human gene nomenclature database (www.gene.ucl.ac.uk). A priori, it was expected that many transporters would be markers of terminal differentiation, and it was, indeed, the case that a large number of the 65 transporter-specifying RNAs were maximally expressed in the adult kidney (Fig. 6). This group included a heterogeneous collection of organic anion/cation transporters, phosphate transporters, Na/H exchangers, and regulators (though not actual transporters) of amino acid transport. These results are in keeping with previous observations for NKT/OAT1, OAT2, Rcot, OCT1, NaPi, and SGLT1 (11, 16, 19). Several genes peaked in either midembryogenesis or neonatal life. This group consisted almost exclusively of nucleoside transporters (SLC28A1, SLC29A1, SLC29A2) but also included the glycerol-3-phosphate transporter, SLC37A1 (amino acid permeases). The amino acid permeases (SLC group 7) displayed relatively flat expression profiles in the developing kidney. However, their regulators (SLC group 3) display a marked increase in expression toward adulthood. Regulation of amino acid transport in developing through adult kidney may therefore be a function of SLC3 expression. The functional role of SLC group 7 molecules in the absence of their positive regulators remains undefined.

To confirm early embryonic vs. later expression of transporters, we generated electronic “eBlots” based on the source library frequency distribution of corresponding EST sequences in the dbEST. dbEST source libraries are encoded with information as to tissue of origin,
and embryonic vs. adult source. We have previously described a custom computer application, eBlot, which associates source library information available in the dbEST with sequences present on Affymetrix GeneChips using blast-derived homology as a linking field (14). Using eBlot, we were able to determine to what degree the early expression observed here during kidney development was reflected in a much larger database (dbEST). It was found that those MFP genes observed to decrease during kidney development were significantly associated with embryonic source libraries, whereas those MFP genes increasing during kidney development were almost entirely associated with adult source libraries (Fig. 7). Both EEG1 and POV1 are likewise associated with embryonic source libraries. In the case of POV1, of 20 representative EST sequences, 12 were derived from embryonic libraries. In the case of EEG1, 5 of 20 representative entries were derived from embryonic sources. In fact, of 8,740 genes assayed via DNA microarrays, the rat POV gene was one of only eight sequences representing the intersection of 1) significantly high early embryonic kidney expression, 2) unknown function, and 3) association with ESTs derived from embryonic libraries. Data, gene lists, and analytic tools are available at www.organogenesis.ucsd.edu.

DISCUSSION

EEG1 was isolated in a screen for sequences differentially regulated in a cell culture model of kidney development. We have isolated many such sequences chosen on the basis of upregulation in this model, including Timeless, a putative transcription factor that appears to be necessary for early embryonic survival and branching morphogenesis of the UB (3, 7). Nevertheless, interesting sequences may actually decrease during renal development, as many morphoregulatory genes are expressed early in development and decline towards birth. Recent DNA microarray analyses of kidney development confirm this notion (14) and lead us to analyze sequences that were isolated from ddPCR gels on the basis of downregulation in the cell culture model system. Among these sequences was a novel ampiclon representing EEG1. In addition to cloning and analyzing this gene, we define the genomic structure of EEG1, elucidate the complex relationship of EEG1 to other genes in the region, and characterize its expression in embryonic and adult tissues.

Much of the EEG1 nucleotide sequence has previously been assigned to another gene, PRG2, located in the same region of human chromosome 11. The PRG2 sequence is a chimera of six EEG1 exons, two novel exons, and the five coding-region exons of MBP (Fig. 2B). The PRG2 transcript has previously been demonstrated in HL-60 cells (a human leukemic cell line), peripheral blood eosinophils from patients with hypereosinophilic syndrome, and bone marrow. It is not known if the bone marrow was "normal." On the other hand, the human dbEST contains millions of randomly cloned sequences from some 7,000 source libraries. And, while the database contains abundant examples of the short 1-kb MBP transcript, not a single instance of PRG2 is found. Furthermore, the PRG2 transcript would not likely translate as the MBP protein as there are many ORF preceding the putative MBP initiation site (5). The PRG2 transcript is infrequently observed, and the shared exons are more appropriately identified as belonging to the EEG1 gene. Despite the lack of evidence for PRG2 in the dbEST, EEG1 and MBP may, in some cell types, under certain conditions, form a hybrid transcript termed PRG2, though this hypothesis awaits confirmation.

Multiple lines of evidence suggested that EEG1 represents a novel and distinct RNA species predicted to encode a protein with characteristics of a membrane transporter. The closest known EEG1 homolog is POV1/PB39. POV1 is located 88 kb upstream (on the reverse strand) of EEG1 on human chromosome 11. The proximity of the two genes at 11p12, taken together with their significant amount of sequence divergence, suggests an ancient gene duplication event. And, both genes (perhaps distant paralogs of a novel class) have features typical of membrane transport proteins including 12 (POV1) or 10 (EEG1) transmembrane spanning segments. Furthermore, hydropathy pattern matching yielded similar functional associations for both proteins. The known protein most similar to EEG1 (in terms of hydropathy) is the human folate-like transporter with numerous sugar transporter near-matches, whereas POV1 has considerable similarity to several cation and amino acid transporters. Nevertheless, it is important to note that neither EEG1

Fig. 7. eBlot database (db)EST source library associations according to MFP expression profile. The dbEST sequences and associated source library information may be surveyed to gain insight into the tissue distribution of a given gene or group of genes. The eBlot program associates gene sequences grouped by cluster membership with dbEST entries on the basis of sequence similarity and calculates summary statistics for tissue source including derivation from embryonic libraries. Data, gene lists, and analytic tools are available at www.organogenesis.ucsd.edu.
nor POV1 closely resembles members of a major facilitator superfamily at the nucleotide or amino acid level, despite the similarities in hydropathy, suggesting that they may be transporters. Therefore, their assignment into this class remains tentative.

A very limited subset of transport proteins in the kidney appear to have significantly higher embryonic than adult expression. A priori, this circumstance might be hypothesized for transporters of nutrient molecules needed for growth. Indeed, we have found that MFP transporters with the highest embryonic expression were involved in glucose or glutamate uptake. This was noted in the kidney specifically and for tissues generally, as reflected in the dbEST. EEG1 was not present on the DNA arrays employed here; however, in situ hybridization showed that EEG1 is highly expressed in several embryonic epithelial tissues including the kidney, lung, and particularly the liver. No precisely quantitative data regarding the time course of EEG1 expression in the developing kidney could be derived. Nevertheless, it was likely from Northern analysis, taken together with the in situ data, that peak expression in liver, lung, and kidney occurred during embryogenesis. The nearest EEG1 homolog, POV1, was present on the DNA arrays and similarly showed a decreasing temporal pattern of expression coincident with two sugar and one glutamate transporter. At least in the context of present knowledge, EEG1 and POV1 may represent unusual examples of transport proteins with high early embryonic expression. They may serve in transport of nutrients and/or metabolites of particular importance in early development and growth.

R. O. Stuart is supported by the Medical Education and Research Foundation and National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Grant K08-DK-02392. S. K. Nigam is supported by NIDDK Grants PO1 DK-54711 and RO1 DK-49517.

REFERENCES


