THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1993 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 268, No. 21, Issue of July 25, pp. 15772-15778, 1993
Printed in U.S.A.

The Human Aquaporin-CHIP Gene

STRUCTURE, ORGANIZATION, AND CHROMOSOMAL LOCALIZATION*

(Received for publication, January 28, 1993, and in revised form, April 8, 1993)

Chulso Moon‡, Gregory M. Preston‡§, Constance A. Griffin¶, Ethylin Wang Jabs‖, and Peter Agre‡**

From the ‡Departments of Medicine and Cell Biology/Anatomy, ¶Oncology, ‖Pediatrics and Medicine, and the Center for Medical Genetics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Aquaporin-CHIP is the first known molecular water channel. Originally identified in red cells and renal tubules, transcripts and proteins related to AQP-CHIP are also expressed in diverse epithelia with distinct developmental patterns. Northern analyses of RNA from several tissues revealed transcripts of 3.1 kilobases and other sizes. The nucleotide sequences of human kidney AQP-CHIP cDNAs are identical to the human bone marrow AQP-CHIP cDNA. The 17-kilobase human AQP-CHIP structural gene was isolated, and restriction maps were constructed and partially sequenced. The TATA consensus sequence is located 87 bp 5′ to the translation initiation site, and sequences surrounding the polyadenylation consensus were determined. Four exons were identified corresponding to amino acids 1–128, 129–183, 184–210, and 211–269, separated by introns of 9.6, 0.43, and 0.80 kilobases. Genomic Southern analyses indicated the existence of a single AQP-CHIP gene which was located at human chromosome 7p14 by *in situ* hybridization. Sequence comparisons of AQP-CHIP and cDNAs of similar proteins from diverse species suggested a common evolutionary origin. At least three of these proteins are now known to function as membrane water pores and are referred to as the "Aquaporins." These genomic AQP-CHIP DNA sequences should permit molecular characterization of the complex patterns of AQP-CHIP expression.

Aquaporin-CHIP[1] is a 28-kDa integral protein purified from the plasma membranes of red cells and renal tubules where it was referred to as "CHIP28" (Denker *et al.*, 1988; Preston and Agre, 1991). The protein exists as a homotetramer which physically resembles certain channel proteins (Smith and Agre, 1991). AQP-CHIP was the first molecular water channel identified, a discovery made by expression of AQP-CHIP in oocytes (Preston *et al.*, 1992) and verified by reconstitution of highly purified AQP-CHIP into proteoliposomes (Zeidel *et al.*, 1992). Detailed immunohistochemistry and immunoelectron microscopy with affinity-purified antibodies documented the presence of AQP-CHIP in water-resorptive epithelia in kidney (Nielsen *et al.*, 1993a). These studies were subsequently confirmed by other investigators (van Hoek and Verkman, 1992; Sabolic *et al.*, 1992; Zhang *et al.*, 1993).

The AQP-CHIP cDNA isolated from a human bone marrow cDNA library (Preston and Agre, 1991) was identified as a relative of MIP26, major intrinsic protein of lens (Gorin *et al.*, 1984) and other members of a recently recognized family of proteins from diverse animals, plants, and microbes (Baker and Saier, 1990; Pao *et al.*, 1991). The function of most MIP26 homologs is unknown. Subsequent to AQP-CHIP, two other related proteins were found to be water transporters (Fushimi *et al.*, 1993; Maurel *et al.*, 1993), and these three proteins are now referred to as the "Aquaporins" (Agre *et al.*, 1993).

AQP-CHIP is widely expressed throughout the body as a result of metabolic and developmental signals. AQP-CHIP mRNA was identified among the growth factor-induced delayed early response genes by Lanahan *et al.* (1992) who also identified AQP-CHIP transcripts in other tissues, as did other investigators (Deen *et al.*, 1992). AQP-CHIP protein exists in many polarized epithelia with high water permeability and in nonfenestrated capillary endothelia (Nielsen *et al.*, 1993b). *In situ* hybridizations with antisense AQP-CHIP RNA probes revealed transcripts in diverse rat tissues with different developmental patterns of expression (Bondy *et al.*, 1993). For example, AQP-CHIP was expressed in choroid plexus throughout fetal development and maturity, but AQP-CHIP was expressed in red cells and kidney only after birth. It is uncertain if the AQP-CHIP transcripts and proteins in these diverse tissues are products of a single gene or if they represent tissue-specific products of distinct genes which are highly related to AQP-CHIP. Nor is it clear how the developmental expression patterns are regulated. The human AQP-CHIP structural gene was therefore isolated to provide the genetic probes needed to address these questions.

EXPERIMENTAL PROCEDURES

Materials—Human bone marrow AQP-CHIP cDNA probes were previously described (Preston and Agre, 1991): i) the nearly full-length 2.9-kb cDNA beginning at nucleotide +15 of the coding sequence and including approximately 2 kb of 3′-untranslated sequence; ii) the 1.3-kb *Hin*dIII fragment containing the coding sequence (above) and approximately 500 bp of 3′-untranslated sequence (to +1302); iii) the 850-bp polymerase chain reaction amplification product, PCR-3, containing sequences corresponding to exons 1–3 and

* This work was supported by National Institutes of Health Grants HL33991, HL48268, AA09012, HG00373, and CA06973. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L11320–L11327.

§ Recipient of a Fellowship Award from the American Heart Association, Maryland Affiliate.

** To whom correspondence and reprint requests should be addressed: Department of Biological Chemistry, The Johns Hopkins University School of Medicine, 725 North Wolfe St., Baltimore, MD 21205.

[1] The abbreviations used are: CHIP, CHannel forming Integral Protein of 28 kDa (CHIP28); AQP, Aquaporin; 1 × SSC, 0.15 M NaCl, 15 mM sodium citrate; kb, kilobase(s); bp, base pair(s); TIP, tonoplast intrinsic protein.

190 bp from the 5′ end of intron C; or iv) other cDNA probes described below.

General Methods—Except where listed, standard methods were employed (Sambrook *et al.*, 1989). cDNA probes were gel-purified and labeled with [α-^{32}P]dCTP (3,000 Ci/mmol; Amersham) using a random priming kit (Pharmacia LKB Biotechnology Inc.). DNA sequencing was performed using the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). Except where listed, Northern blots, library plaque lift filters, and genomic and chromosomal assignment Southern blots were washed under conditions of moderate stringency (0.5 × SSC, 0.1% SDS, 55 °C) or high stringency (0.1 × SSC, 0.1% SDS, 65 °C).

Northern Analyses—RNA from rat tissues, human bone marrow, and anemic mouse spleen (Chomczynski and Sacchi, 1987) was transferred to Hybond nylon membrane (Amersham) and probed with ^{32}P-labeled AQP-CHIP cDNAs under conditions of high stringency using Rapid-Hyb buffer (Amersham) and otherwise as described (Preston and Agre, 1991). RNA concentrations were estimated by absorbance at 260 nm and by ethidium bromide staining of gels to visualize 28 S and 18 S ribosomal RNAs.

Isolation of Renal cDNA Clones—An adult human kidney cDNA library in λgt10 was obtained from Clontech; 250,000 recombinants were screened under conditions of moderate stringency with two nonoverlapping ^{32}P-labeled probes used separately. The cDNA probes corresponded to nucleotides +15 to +272 and +483 to +822 of the human bone marrow AQP-CHIP cDNA (Fig. 2). Two positive plaques were purified; one plaque hybridized with both probes, and the other hyridized only to the 3′ probe. DNA extracted from the phage was digested with *Eco*RI and electrophoresed into a 1% agarose gel. The gel-purified inserts were ligated into pBluescriptII (Stratagene), and the nucleotide sequences were determined.

Isolation and Characterization of Genomic DNA Clones—A human genomic DNA library in λEMBL-3 was provided by Jeremy Nathans, Johns Hopkins University (Wang *et al.*, 1992). The ^{32}P-labeled 2.9-kb AQP-CHIP cDNA was used to screen 500,000 plaques with moderate stringency. DNA from positive recombinants was digested with *Sal*I. The DNA fragments were electrophoresed into a 0.8% agarose gel, purified, and ligated into pBluescriptII.

Restriction Mapping and Determination of Flanking Nucleotide Sequences and Exon-Intron Boundaries—The genomic clones and subclones were digested with single restriction enzymes or combinations of enzymes. The digests were analyzed by Southern hybridization with ^{32}P-labeled probes obtained from the human bone marrow AQP-CHIP cDNA (listed above). Nucleotide sequences corresponding to flanking regions and exon-intron boundaries were determined with oligonucleotide primers (sense or antisense, nucleotides are numbered).

Proximal Promoter (antisense, 18–41), 5′-ACTGCCCTCCAGAAGAGCTTCTTC-3′; Exon 1-Intron A (sense, 343–362), 5′-ATCACCTCCTCCCTGACTGG-3′; Intron A-Exon 2 (antisense, 438–455), 5′-CATAGCACCAGCTGGAGG-3′; Exon 2-Intron B (sense, 483–500), 5′-CCGTGACCTTGGTGGCTC-3′; Intron B-Exon 3 (antisense, 612–628), 5′-AGTGGTTGCTGAAGTTG-3′; Exon 3-Intron C (sense, 551–568), 5′-TTGACTACACTGGCTGTG-3′; Intron C-Exon 4 (antisense, 785–804), 5′-GGGCTTCATCTCCACCCTGG-3′; Exon 4–3′-untranslated (sense, 608–627), 5′-CACACAACTTCAGCAACCAC-3′; Polyadenylation (sense, 155′-172′), 5′-AGGCTGCTGCACACCAAG-3′.

The sizes of the second and third introns were determined by the polymerase chain reaction using flanking oligonucleotide primers (above) with the following conditions (25 cycles at 94 °C, 1 min; 42 °C, 30 s; 72 °C, 2 min). The amplification products were slowly electrophoresed into an 0.8% agarose gel with adjoining 1-kb DNA ladder standards. The products were also cut with several restriction enzymes to facilitate restriction mapping.

Genomic Southern Analyses—Total genomic DNA from the leukocytes of two unrelated individuals was digested with single restriction enzymes or combinations of enzymes, electrophoresed into a 1% agarose gel, and transferred to a nylon membrane (GeneScreen, Du Pont-New England Nuclear). The blot was prehybridized in 1 M NaCl, 1% SDS, 50 mM sodium phosphate (pH 7.0) for 3 h at 65 °C and then hybridized with ^{32}P-labeled 1.3-kb human AQP-CHIP cDNA (2×10^6 cpm/ml) overnight at 65 °C in 1 M NaCl, 1% SDS, 100 μg/ml salmon sperm DNA, 10% dextran sulfate. Blots were then washed with moderate stringency at 65 °C and exposed to Kodak XAR film with intensifying screens for 2 weeks.

Chromosomal Assignment, Somatic Cell Hybrid Panel—The complete NIGMS rodent/human somatic cell hybrid panel 1 consisting of 17 mouse/human hybrids and 1 Chinese hamster ovary (CHO)/human hybrid was from Coriell Institute for Medical Research (Camden, NJ). The hybrid cell DNAs and their cytogenetic characterization were provided by the manufacturer. The hybrid cell DNAs were digested with *Hin*dIII, electrophoresed using 1% agarose gel, Southern-transferred onto Zetabind (CUNO Inc., Meriden, CT), hybridized with the 850-bp ^{32}P-labeled AQP-CHIP cDNA, and washed with high stringency.

Chromosomal Localization and Chromosomal in Situ Hybridizations—The plasmid containing the 14-kb genomic human AQP-CHIP gene (Fig. 3) was nick-translated with biotin-14 dATP (Life Technologies Inc.) with 11% incorporation determined by tritium tracer incorporation. Slides with human chromosome spreads were made from normal male lymphocytes cultured with 5-bromodeoxyuridine (Bhatt *et al.*, 1988). Fluorescence *in situ* methods were adapted from Lichter *et al.* (1990). The probe was denatured at 70 °C for 5 min in a mixture containing 2 × SSC, 50% formamide, 10% dextran, 20 ng/μl biotinylated probe, 500 × salmon sperm DNA, and 250 × C_0t DNA, preannealed at 37 °C for 30 min, placed on slides, and hybridized at 37 °C overnight. Slides were washed in 50% formamide, 2 × SSC at 37 °C for 5 min, and followed with two changes of 2 × SSC at 37 °C for 5 min each. The biotinylated probe was detected with fluorescein isothiocyanate-avidin and biotinylated anti-avidin, using reagents from an *in situ* hybridization kit (Oncor Inc., Gaithersburg, MD) following manufacturer's instructions.

RESULTS

Aquaporin-CHIP mRNAs in Diverse Tissues—Northern analyses of human bone marrow, anemic mouse spleen, and mouse kidney were previously shown to contain major transcripts of approximately 3.1 kb and faint signals slightly below 1.4 kb, whereas circulating red cells were devoid of AQP-CHIP RNA (Preston and Agre, 1991). The nucleotide and deduced amino acid sequences of rat AQP-CHIP cDNAs were 88% and 93% identical with human (Deen *et al.*, 1992). When total RNA from several rat tissues were analyzed at high stringency with a ^{32}P-labeled human AQP-CHIP cDNA, no signal was detected in RNA from brain, but distinct transcripts of approximately 3.1 kb and faint transcripts of approximately 1.4 kb were detected in each of several other tissues (Fig. 1*A*). The predominant signal in skeletal muscle was 1.4 kb. RNA from small intestine and lung each contained a transcript of 4.2 kb, whereas lung also contained the 3.1-kb signal. The 3.1-kb transcript most likely corresponds to the standard polyadenylation consensus sequence AATAAA located at about nucleotide +3,000 in the human AQP-CHIP cDNA; the 1.4-kb transcript probably corresponds to the weak consensus sequence AAGAAA at nucleotide +1,068 (Preston and Agre, 1991). Corresponding consensus sequences were noted in the rat AQP-CHIP cDNA (Deen *et al.*, 1992), but it is not known if other polyadenylation consensus signals exist still further 3′ in either species.

The microanatomy of kidney is complex, so the distribution and the nucleotide sequence of the renal AQP-CHIP cDNA was determined. RNA isolated from distinct segments of dissected rat kidney was analyzed by Northern blot with an AQP-CHIP cDNA probe. A single transcript of approximately 3.1 kb was seen in cortex, outer medulla, and inner medulla (Fig. 1*B*), confirming the lack of transcript heterogeneity in these functionally distinct levels of the nephron. Two recombinants were isolated from an adult human kidney cDNA library probed with AQP-CHIP cDNAs for determination of the nucleotide sequences without the severe selective restrictions of the polymerase chain reaction (Fig. 2). The nucleotide sequences of the coding regions beginning at +84 and + 707bp and the sequences of the 3′-untranslated regions of the kidney recombinants were identical with the human bone marrow AQP-CHIP cDNA (Preston and Agre, 1991). Identical nucleotide sequences indicate that AQP-CHIP in kidney and red cells are the products of the same gene and not homologous

genes as proposed by others (Zhang *et al.*, 1993).

Structure of the Human Aquaporin-CHIP Genomic Locus—Two recombinants were isolated from a partially *Sal*I-digested human genomic DNA library which reacted with the ^{32}P-labeled AQP-CHIP cDNA probe. The isolates contained overlapping inserts of 17 and 14 kb which were comprised of a 9.5-kb DNA fragment and overlapping 7.5- and 4.5-kb DNA fragments which were released by *Sal*I because of an apparent polymorphism at one *Sal*I site (Fig. 3). Restriction maps and the locations of four exons were established.

Nucleotide sequences of the 5′- and 3′-flanking regions and the intron-exon boundaries were identified using sequencing primers corresponding to the AQP-CHIP cDNA (Fig. 4). The 5′-flanking region contained the sequence TATAAA located at nucleotides −87 to −82 with respect to the translation initiation site. Sp1 binding sites and a presumed transcription initiation site were also noted (Fig. 4). No other potential ATG translation initiation signals were identified. The nucleotide sequences surrounding the polyadenylation consensus sequence and 3′ to the termination of the kidney and bone marrow cDNAs were also determined (Fig. 4).

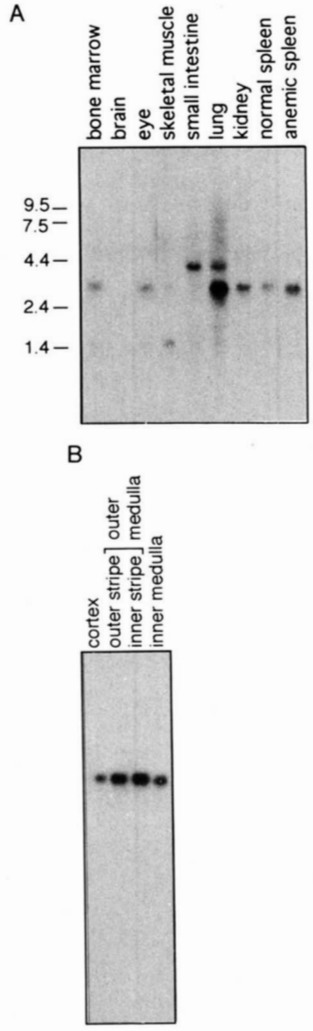


FIG. 1. **Northern analyses with Aquaporin-CHIP cDNA probes.** *A*, Northern blot of 20 μg of total RNA from the indicated rat organs or 5 μg of total RNA from skeletal muscle, human bone marrow, and anemic mouse spleen were probed with ^{32}P-labeled 1.3-kb AQP-CHIP cDNA (see "Experimental Procedures"). *B*, aliquots of 5 μg of total RNA from dissected regions of rat kidney were analyzed similarly with ^{32}P-labeled 850-bp AQP-CHIP cDNA probe.

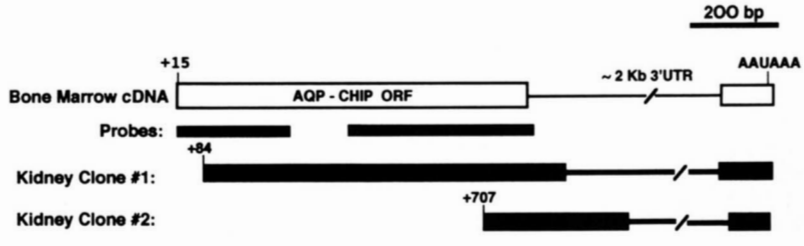
Four exons were identified corresponding to amino acids 1–128, 129–183, 184–210, and 211–269 of the amino acid sequence deduced from the human bone marrow AQP-CHIP cDNA. The nucleotide sequences of these exons were identical with the cDNA sequence deduced from human bone marrow (Preston and Agre, 1991) and kidney clones (Fig. 2). Clear exon-intron class 0 boundaries were identified for each of the four exons (Fig. 4). The size of the first intron (A) was determined to be 9.6 kb by mapping restriction fragments (Fig. 3). Exon-specific sense and antisense oligonucleotide primers were used for polymerase chain reaction amplification of the 4.5-kb genomic DNA restriction fragment template. Based upon the electrophoretic mobility of the amplification products and restriction digests of the 4.5-kb genomic DNA fragment, intron B was estimated to be 0.43 kb and intron C to be 0.8 kb.

Genomic Southern analyses were performed looking for multiple genes related to AQP-CHIP. Human leukocyte DNA from two unrelated individuals was cut with restriction enzymes and analyzed with medium stringency. Only one or two distinct hybridization signals were detected, suggesting the existence of a single AQP-CHIP gene locus (Fig. 5), and the patterns of the two individuals were essentially identical. Moreover, the observed patterns were consistent with the known restriction sites in the AQP-CHIP genomic clones (Fig. 3). For example, *Hin*dIII was expected to release 4.4- and 1.3-kb fragments and a flanking fragment. Hybridizations were observed with DNA fragments estimated to be 4.5 and and 1.2 kb (Fig. 5, *left*). The 4.5-kb signal was resolved into fragments of 4.6 and 4.4 kb (Fig. 6 *top*), and the 4.6-kb fragment probably corresponds to 5′-flanking DNA. *Pst*I digestion yielded DNA fragments of 1.9 and 1.6 kb as expected (Fig. 5, *left*).

Chromosomal Assignment and Localization of the Aquaporin-CHIP Gene—After digestion with *Hin*dIII, Southern analysis of DNA from a complete panel revealed human-specific DNA fragments of 4.6 and 4.4 kb (Fig. 6, *top*) and 1.3 kb (not shown). The presence of these bands was 100% concordant with hybrid cell lines known to contain human chromosome 7 in >4% of the cells. Moreover, all the other human chromosomes were 22–61% discordant for the presence of these bands. The homologous fragment from mouse was 16 kb and fragments from Chinese hamster were 10 and 5 kb.

Analysis of 51 metaphase cells from chromosomal *in situ* hybridizations demonstrated 25 (49%) which had at least one paired signal involving both chromatids of a single chromo-

FIG. 2. **Diagram of Aquaporin-CHIP cDNAs.** Representation of the structural relationship between the human bone marrow AQP-CHIP cDNA (*ORF*, open reading frame), the cDNA probes, and recombinants from a human kidney cDNA library. *Open boxes* in the bone marrow cDNA and the *black boxes* in the kidney cDNAs denote nucleotide sequences which were determined.



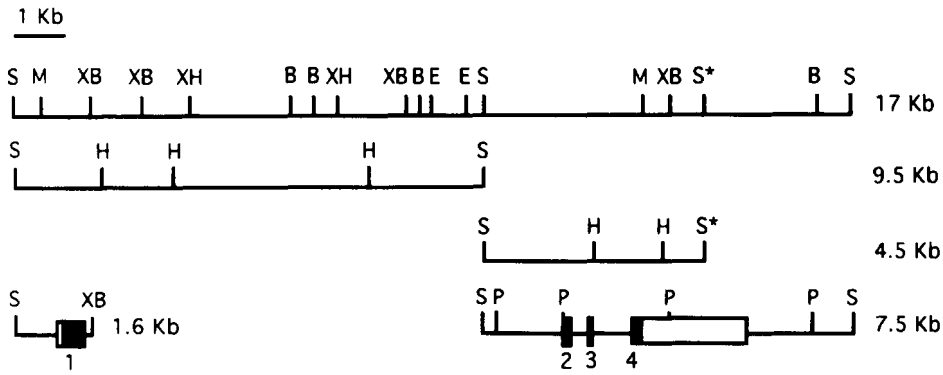


FIG. 3. **Restriction map and intron-exon organization of human Aquaporin-CHIP genomic locus.** Two overlapping recombinants were isolated from a genomic DNA library due to an apparent polymorphism at one *SalI* site (*S**). Generation of subclones was achieved by digestion with single restriction enzymes or combinations of enzymes yielding the composite restriction map. Locations of the exons were established by Southern analysis with ^{32}P-labeled probes corresponding to the cDNA sequence. Abbreviations: *S* = *Sal*I, *M* = *Sma*I, *XB* = *Xba*I, *XH* = *Xho*I, *B* = *Bam*HI, *E* = *Eco*RI, *H* = *Hin*dIII, *P* = *Pst*I.

5'-gccgagtctgcatccatccagaggaggtcgtggt gtgggggcgggccaggagcgaagagaggccttcctccctttgtgctccccc cgcccccggcccctataaataggcccagcccaggctgtggctcagctctc agagggaattgagcacccggcagcggtctcaggccaagccccctgccagc ATG GCC CAG M A S 1 2 3	Proximal promoter
126 127 128 R N D CGC AAT GAC gtgagtggggtgtccctgggcttgggggggttccagaatgatgctgaaag	Exon 1/ Intron A
agtccctgcctaccctcctcaccagtcctcaccacctctctcccctgcag CTG GCT GAT L A D 129 130 131	Intron A/ Exon 2
181 182 183 L L A CTC CTG GAT gtgagtcaggggccctcccagatggaggtgggggaagggagggcgggggc	Exon 2/ Intron B
cacctatgactctctgccttcgcccctccctctgtttctttccctcacag ATT GAC TAC I D Y 184 185 186	Intron B/ Exon 3
208 209 210 N H W AAC CAC TGG gtaggagacccacgggggggtggggtgggaagctttggtgtcccatggtaa	Exon 3/ Intron C
cccaggggcttttgagtggagccctctgaacacacctgctctgttcctag ATT TTC TGG I F W 211 212 213	Intron C/ Exon 4
tgtctctttggagttggaatttcattatatgttaagaaaataaaggaaaatgacttgtaagg tccttcaggcttcctgtagtttttcttcttttgatgctgtgtccccacaaatgtttctggat tactccacaaggaaatcaggccactgctagagacccagtctgaggaggagccactacctccc	Poly A site

FIG. 4. **Nucleotide sequences of the 5′-flanking region, intron-exon boundaries, and surrounding the 3′ polyadenylation site.** Nucleotide sequences were obtained by dideoxynucleotide sequencing-selected restriction fragments (Fig. 3) with oligonucleotide sequencing primers designed from the human AQP-CHIP cDNA. The 5′-flanking sequence contained a tataaa element (*underlined*), a presumed transcription initiation site (*bold underlined*), and Sp1 sites (*underlined*). Nucleotide sequences surrounding the introns (A, B, and C) were similarly established. The flanking sequence surrounding the 3′ end of the human bone marrow and kidney AQP-CHIP cDNAs (*double underlined c*) contained the polyadenylation consensus sequence aataaa (*underlined*).

some. Thirty-six paired signals were seen; thirty-four (94%) were located near the midpoint of the short arm of a large C-group chromosome. To determine the specific chromosome and band, slides were G-banded by fluorescence plus Giemsa (Bhatt *et al.*, 1988), photographed, and aligned with the color slides to determine sub-band location. Thirty-two metaphase signals were analyzable after banding, and thirty (94%) were between 7p13 and 7p15. To confirm and refine this localization, a second experiment was performed. Fifty-seven metaphases were G-banded and photographed prior to hybridization with the probe with the same conditions; 12 of 16 analyzable paired signals were located at 7p14 (Fig. 6, *bottom*).

Evolutionary Comparisons of Aquaporin-CHIP-related Genes and Gene Products—AQP-CHIP contains an amino acid sequence related to MIP26 protein of lens fiber cells (Gorin *et al.*, 1984) and several recently identified proteins from diverse animal, plant, and microbial species. Each of the homologs contains internal tandem repeats corresponding to approximately the first half and second half of the proteins, and each repeat contains a highly conserved 3-residue motif Asn-Pro-Ala, NPA (Pao *et al.*, 1991; Wistow *et al.*, 1991). The membrane topology of none of these proteins has been rigorously established. The loop connecting the tandem repeats (Fig. 7*A*) contains amino acid sequences which are markedly variable, and the location of this connecting loop has been shown be extracellular in AQP-CHIP.[2]

Intron-exon boundaries have also been established for human MIP26 (Pisano and Chepelinsky, 1991) and two plant homologs found in water containing tonoplasts, α-TIP and γ-TIP (Höfte *et al.*, 1992). All of the intron-exon boundaries are class 0, and all four proteins contain an intron located in the variable loop which separates the sequences encoding the first and second repeats (Fig. 7*A*). Although the amino acid sequences of MIP26 and AQP-CHIP are only 44% identical overall, the intron-exon boundaries were located at three identical points within their coding sequences clearly establishing that they evolved from a common origin (Fig. 4). The sizes of the introns were not conserved: intron A = 9.6/0.5 kb, B = 0.43/0.44 kb; C = 0.80/1.6 kb (AQP-CHIP/MIP26). Deduced amino acid sequences of AQP-CHIP and eight related proteins were compared by PILEUP which revealed clustering of proteins along phylogenetic lines (Fig. 7*B*). The overall sequence identities ranged from 23–44%. When the first or second repeat of AQP-CHIP were aligned with the

[2] G. M. Preston and P. Agre, unpublished results.

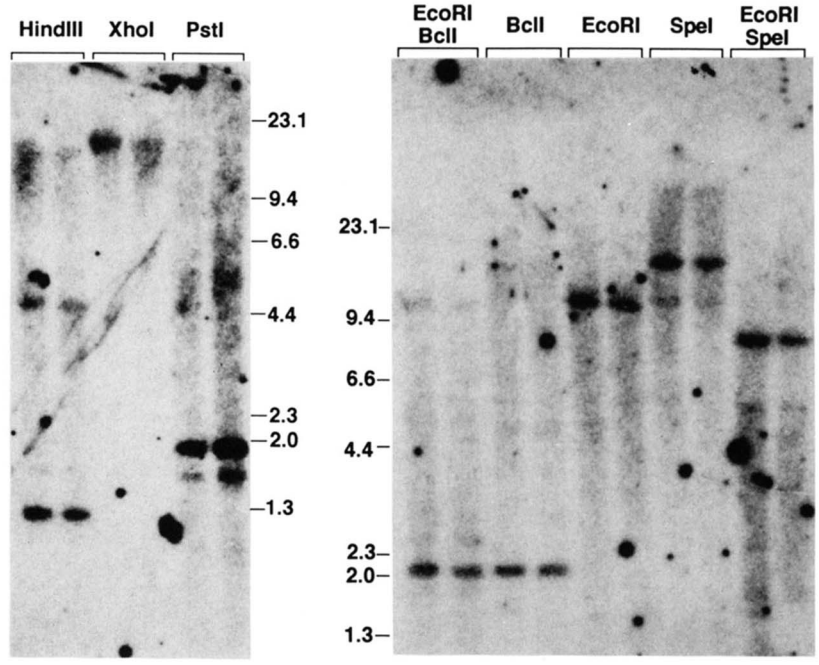


FIG. 5. **Genomic Southern analyses with Aquaporin-CHIP cDNA probe.** DNA from two unrelated humans was digested with single restriction enzymes or combinations of enzymes as indicated. Each lane contains 7.5 μg of DNA. Southern analyses were undertaken with ^{32}P-labeled 1.3-kb AQP-CHIP cDNA probe under conditions of medium stringency.

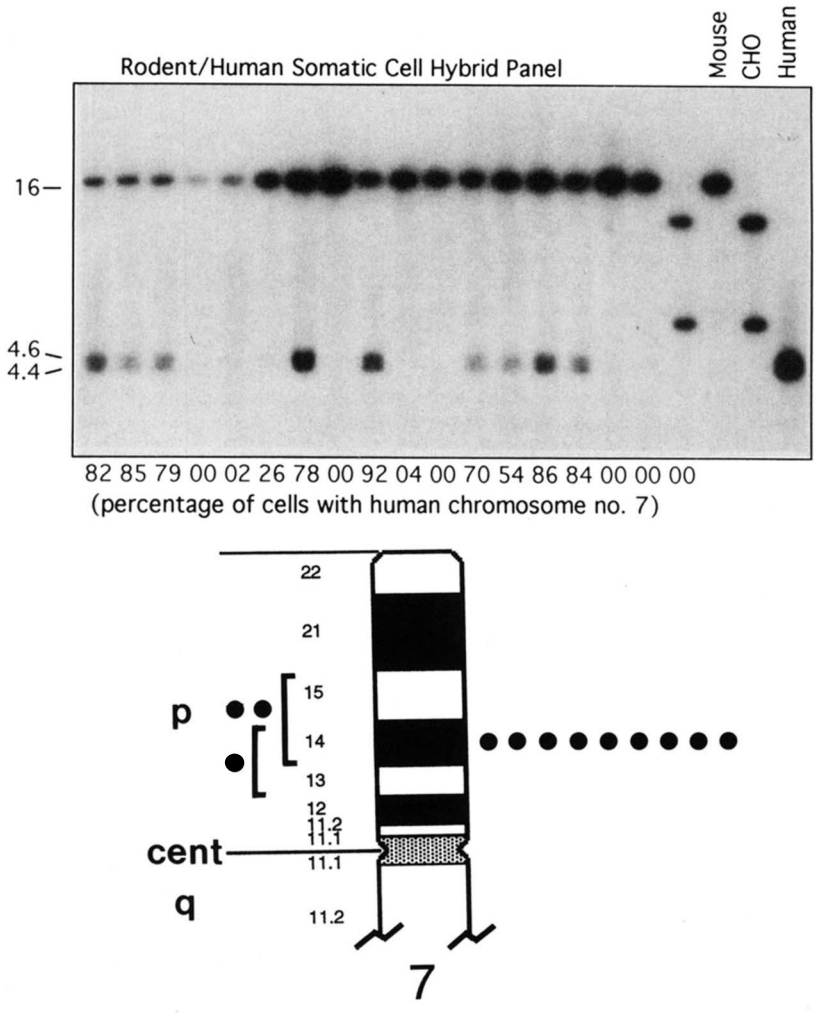


FIG. 6. **Chromosomal assignment and localization of human Aquaporin-CHIP genomic locus.** *Top*, Southern analysis of NIGMS panel 1 probed with ^{32}P-labeled 850-bp AQP-CHIP cDNA. The first 17 lanes are *Hind*III digests of mouse/human hybrid DNAs, and the 18th is Chinese hamster ovary (CHO)/human hybrid DNA. The hybrid cell DNAs were GM/NA09925, 26, 27, 28, 29, 30A, 31, 32, 33, 34, 35A, 36, 37, 38, 40, and GM/NA10324, 567, and 611 (*left to right*). The control lanes contain mouse, CHO, and human genomic DNA (*far right*). Note that the 4.6- and 4.4-kb *Hind*III human fragments were present in hybrids containing human chromosome 7 in >4% of their cells. *Bottom* ideogram of the short arm of human chromosome 7 showing hybridization of AQP-CHIP to previously G-banded metaphase chromosomes from 12 cells. Each of the *9 dots* to the *right* of the ideogram represents a paired signal localized to the indicated band. The *3 dots* to the *left* represent signals which overlapped the bands indicated by the *brackets*. The AQP-CHIP gene maps to 7p14.

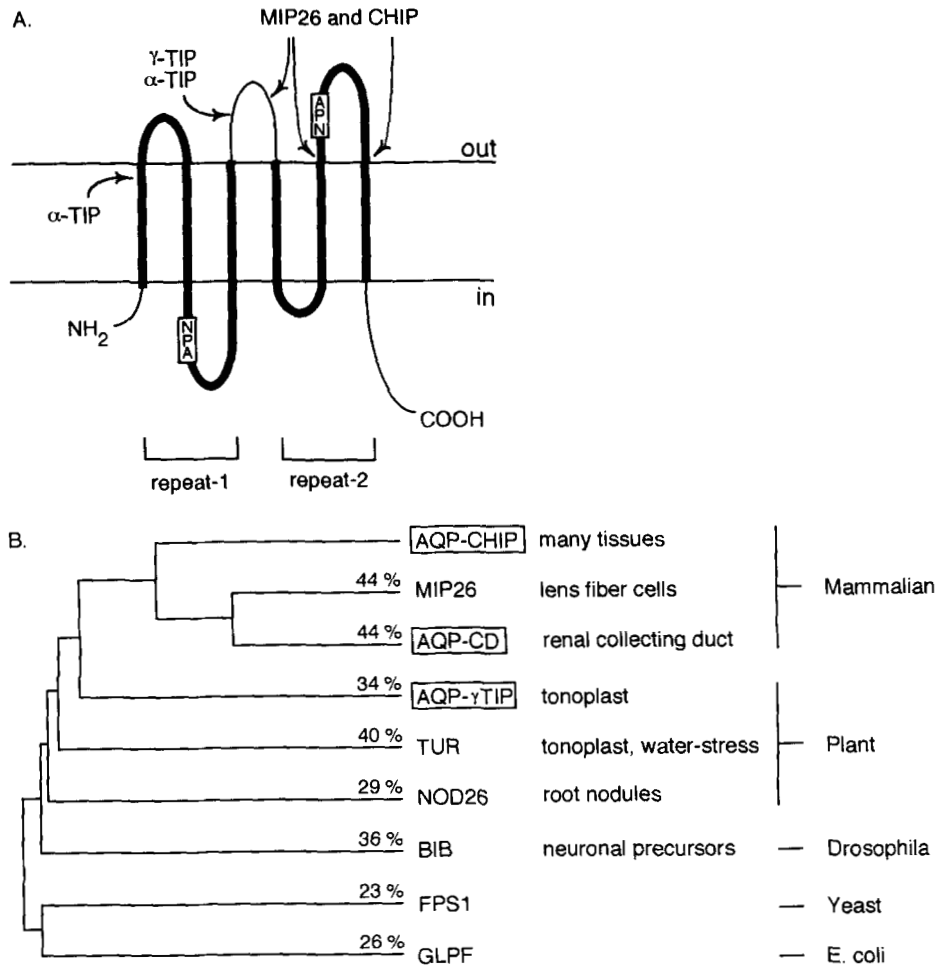


FIG. 7. **Comparison of gene structures and evolutionary distances between Aquaporin-CHIP and related proteins.** *A*, model of proposed membrane topology. Internal tandem repeats contain amino acid sequences which are highly conserved (*heavy lines*) and predicted polypeptide loops containing the highly conserved repeats Asn-Pro-Ala (NPA). The NH_2 and COOH termini and connecting loop contain amino acid sequences which are most divergent (*thin lines*). The exon-intron boundaries (*arrows*) are known for two plant proteins, α-TIP and γ-TIP (Höfte *et al.*, 1992), and two mammalian proteins, MIP26 (Pisano and Chepelinsky, 1991) and AQP-CHIP. *B*, cluster analysis showing phylogenetic relationships between AQP-CHIP and representative members of the MIP26 protein family. The horizontal branch lengths are inversely proportional to the degree of similarity between the sequences. Sequences were obtained from AQP-CHIP (Preston and Agre, 1991), MIP26 (Gorin *et al.*, 1984), AQP-CD (Fushimi *et al.*, 1993), AQP-γ-TIP (Höfte *et al.*, 1992), TUR (referred to as 7A, Guerrero *et al.*, 1990), NOD26 (Sandal and Marcker, 1988), BIB (Rao *et al.*, 1990), FPS1 (Van Aelst *et al.*, 1991), and GLPF (Muramatsu and Mizuno, 1989). The entire deduced amino acid sequences of the indicated proteins were aligned with the PILEUP program, version 7.1 (Genetics Computer Group, Madison WI) of progressive alignment (Feng and Doolittle, 1990) using a gap weight of 3.0 and a gap length weight of 0.1 and The Johns Hopkins University VAX Computer System. Percentages indicate number of residues identical with AQP-CHIP when aligned. Similar clustering was achieved when subsets of the sequences were analyzed.

corresponding first or second repeats of the homologous proteins, higher degrees of identity were noted (Preston and Agre, 1991).

DISCUSSION

The studies reported here document the existence of a single AQP-CHIP structural gene. Consistent with this, the erythroid and renal AQP-CHIP cDNAs were identical. AQP-CHIP mRNA and protein were identified in epithelia of several organs including choroid plexus, ciliary body of the eye, hepatobiliary ducts, and nonfenestrated capillary endothelia (Bondy *et al.*, 1993; Nielsen *et al.*, 1993b). Northern analysis revealed transcripts of 3.1 kb and other sizes (Fig. 1*A*). These studies do not resolve whether differences in transcript size result from use of alternative transcription initiation or polyadenylation signal sequences or alternative splicing of yet unidentified exons. Use of probes designed from the 5′- and 3′-untranslated sequences and intron-exon boundaries reported here should permit clarification.

The site of the AQP-CHIP gene locus at human chromosome 7p14 does not coincide with a suspicious disease phenotype, and the broad distribution of AQP-CHIP protein suggests that the null phenotype may be lethal (Nielsen *et al.*, 1993b). Nevertheless, AQP-CHIP may not be essential in all tissues, since the protein is not expressed in rat red cells and renal tubules until after birth (Bondy *et al.*, 1993). Surprisingly, AQP-CHIP was identified among delayed early response genes in fibroblasts following stimulation with growth factors (Lanahan *et al.*, 1992). Analysis of the 5′-untranslated sequence may provide molecular explanations for the developmental or metabolic regulators of AQP-CHIP expression. Moreover, there exists the possibility of tissue-restricted AQP-CHIP mutants resulting from defects in tissue-specific AQP-CHIP expression, and the potential clinical phenotypes would therefore reflect the tissue affected.

First recognized in 1990 (Baker and Saier), more than 12

members of the MIP26 protein family are now known. It is therefore likely that many additional members of this protein family will be identified. Some of these homologs are not water-selective pores. MIP26 may be an ion channel, since it conducts current when reconstituted into planar lipid bilayers (Ehring *et al.*, 1990). The product of the *E. coli* gene *GlpF* facilitates glycerol transport (Heller *et al.*, 1980), and the product of the *S. cerevisiae* gene *FPS1* suppresses defective growth on fermentable sugars of the *fdp1* mutant (Van Aelst *et al.*, 1991).

It is likely that at least some homologous proteins will be water pores. A homologous water-transporting protein, AQP-CD, has already been identified in renal collecting ducts and is probably the vasopressin-regulated water channel (Fushimi *et al.*, 1993). Another water-transporting homolog, AQP-γ-TIP, has been identified in plant vegetative bodies (Maurel *et al.*, 1993). AQP-CHIP was not identified in lacrimal, mammary, and salivary glands and therefore cannot explain osmotic water movement through those epithelia (Nielsen *et al.*, 1993b); AQP-CHIP also cannot entirely explain transcellular water movement through the choroid plexus where the protein is localized only in the apical membranes. A likely explanation for these observations is that other water channels may also exist in these tissues. Therefore, the list of "Aquaporins" is very likely to lengthen.

Acknowledgments—We thank Jeremy Nathans for the human genomic DNA library, Mark Knepper for the kidney Northern blot, John B. Williams for valuable suggestions, Paul Bray and Lu Li for generous instruction and assistance with genomic Southern blotting, and Anita Hawkins for assistance with *in situ* chromosomal localization studies.

REFERENCES

Agre, P., Preston, G. M., Smith, B. L., Jung, J. L., Raina, S., Moon, C., Guggino, W. B., and Nielsen, S. (1993) *Am. J. Physiol.*, in press
Baker, M. E., and Saier, M. H. (1990) *Cell* **60,** 185–186
Bhatt, B., Burns, J., Flannery, D., and McGee, J. (1988) *Nucleic Acids Res.* **16,** 3951–3961
Bondy, C., Chin, E., Smith, B. L., Preston, G. M., and Agre, P. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 4500–4504
Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159
Deen, P. M. T., Dempster, J. A., Wieringa, B., and Van Os, C. H. (1992) *Biochem. Biophys. Res. Commun.* **188,** 1267–1273
Denker, B. M., Smith, B. L., Kuhajda, F. P., and Agre, P. (1988) *J. Biol. Chem.* **263,** 15634–15642
Ehring, G. R., Zampighi, G. A., Horwitz, J., Bok, D., and Hall, J. E. (1990) *J. Gen. Physiol.* **96,** 631–664
Feng, D. F., and Doolittle, R. F. (1990) *Methods Enzymol.* **183,** 375–387
Fushimi, K., Uchida, S., Hara, Y., Hirata, Y., Marumo, F., and Sasaki, S. (1993) *Nature* **361,** 549–552
Gorin, M. B., Yancey, S. B., Cline, J., Revel, J.-P., and Horwitz, J. (1984) *Cell* **39,** 49–59
Guerrero, F. D., Jones, J. T., and Mullet, J. E. (1990) *Plant Mol. Biol.* **15,** 11–26
Heller, K. B., Lin, E. C. C., and Wilson, T. H. (1980) *J. Bacteriol.* **144,** 274–278
Höfte, H., Hubbard, L., Reizer, J., Ludevid, D., Herman, E. M., and Chrispeels, M. J. (1992) *Plant Physiol.* **99,** 561–570
Lanahan, A., Williams, J. B., Sanders, L. K., and Nathans, D. M. (1992) *Mol. Cell. Biol.* **12,** 3919–3929
Lichter, P., Tang, C., Call, K., Hermanson, G., Evans, G., Housman, D., and Ward, D. (1990) *Science* **247,** 64–69
Maurel, C., Reizer, J., Schroeder, J. I., and Chrispeels, M. (1993) *EMBO J.*, in press
Muramatsu, S., and Mizuno, T. (1989) *Nucleic Acids Res.* **17,** 4378
Nielsen, S., Smith, B. L., Christensen, E. I., Knepper, M. A., and Agre, P. (1993a) *J. Cell Biol.* **120,** 371–383
Nielsen, S., Smith, B. L., Christensen, E. I., and Agre, P. (1993b) *Proc. Natl. Acad. Sci. U. S. A.*, in press
Pao, G. M., Wu, L.-F., Johnson, K. D., Höfte, G., Chrispeels, M. J., Sweet, G., Sandal, N. N., and Saier, M. H. (1991) *Mol. Microbiol.* **5,** 33–37
Pisano, M. M., and Chepelinsky, A. B. (1991) *Genomics* **11,** 981–990
Preston, G. M., and Agre, P. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 11110–11114
Preston, G. M., Carroll, T. P., Guggino, W. B., and Agre, P. (1992) *Science* **256,** 385–387
Rao, Y., Jan, L. Y., and Jan, Y. N. (1990) *Nature* **345,** 163–167
Sabolic, I., Valenti, G., Verbavatz, J.-M., Van Hoek, A. N., Verkman, A. S., Ausiello, D. A., and Brown, D. (1992) *Am. J. Phys.* **263,** C1225-C1233
Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Sandal, N. N., and Marcker K. A. (1988) *Nucleic Acids Res.* **16,** 9347
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
Smith, B. L., and Agre, P. (1991) *J. Biol. Chem.* **266,** 6407–6415
Van Aelst, L., Hohmann, S., Zimmermann, G. K., Jans, A. W. H., and Thevelin, J. M. (1991) *EMBO J.* **10,** 2095–2104
van Hoek, A. N., and Verkman, A. S. (1992) *J. Biol. Chem.* **267,** 18267–18269
Wang, Y., Macke, J. P., Merbs, S. L., Zack, D. J., Klaunberg, B., Bennett, J., Gearhart, J., and Nathans, J. (1992) *Neuron* **9,** 429–440
Wistow, G. J., Pisano, M. M., and Chepelinsky, A. B. (1991) *Trends Biol. Sci.* **16,** 170–171
Zeidel, M. L., Ambudkar, S. V., Smith, B. L. and Agre, P. (1992) *Biochemistry* **31,** 7436- 7440
Zhang, R., Skach, W., Hasegawa, H., van Hoek, A. N., and Verkman, A. S., (1993) *J. Cell. Biol.* **120,** 359–369