Genomics **58,** 281–292 (1999)
Article ID geno.1999.5839, available online at http://www.idealibrary.com on IDEAL®

Characterization of Novel Promoter and Enhancer Elements of the Mouse Homologue of the Dent Disease Gene, *CLCN5,* Implicated in X-Linked Hereditary Nephrolithiasis

Karo Tanaka,*,† Simon E. Fisher,*,‡ and Ian W. Craig*,§,[1]

*Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom; ‡Wellcome Trust Centre for Human Genetics, University of Oxford, Windmill Road, Oxford OX3 7BN, United Kingdom; †Teikyo University School of Medicine, 2-11-1 Kaga, Itabashi-ku, Tokyo 173, Japan; and §The Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5, United Kingdom

Received January 14, 1999; accepted April 6, 1999

The murine homologue of the human chloride channel gene, *CLCN5,* defects in which are responsible for Dent disease, has been cloned and characterized. We isolated the entire coding region of mouse *Clcn5* cDNA and ~45 kb of genomic sequence embracing the gene. To study its transcriptional control, the 5′ upstream sequences of the mouse *Clcn5* gene were cloned into a luciferase reporter vector. Deletion analysis of 1.5 kb of the 5′ flanking sequence defined an active promoter region within 128 bp of the putative transcription start site, which is associated with a TATA motif but lacks a CAAT consensus. Within this sequence, there is a motif with homology to a purine-rich sequence responsible for the kidney-specific promoter activity of the rat CLC-K1 gene, another member of the chloride-channel gene family expressed in kidney. An enhancer element that confers a 10- to 20-fold increase in the promoter activity of the mouse *Clcn5* gene was found within the first intron. The organization of the human *CLCN5* and mouse *Clcn5* gene structures is highly conserved, and the sequence of the murine protein is 98% similar to that of human, with its highest expression seen in the kidney. This study thus provides the first identification of the transcriptional control region of, and the basis for an understanding of the regulatory mechanism that controls, this kidney-specific, chloride-channel gene.

INTRODUCTION

A gene encoding a chloride channel (CLC-5), which maps to the short arm of human X chromosome at Xp11.22, has been isolated by positional cloning as a candidate for Dent disease, an X-linked hereditary nephrolithiasis (Fisher *et al.,* 1994). The protein product comprises 746 amino acids, and it is predominantly expressed in human kidney. The currently proposed topology of the protein envisages about 12 transmembrane domains, which are commonly conserved within other members of the chloride channel family (Jentsch *et al.,* 1995). Various mutations in *CLCN5* have been reported in affected members of families with Dent disease. Moreover, different mutations within the gene have been shown to be responsible for three other clinically distinguishable diseases, X-linked recessive nephrolithiasis, X-linked recessive hypophosphatemic rickets, and the idiopathic low-molecular-weight proteinuria of Japanese children (Lloyd *et al.,* 1996, 1997). All of these four diseases have features in common and are characterized by various combinations of low-molecular-weight proteinuria, hypercalciuria, nephrocalcinosis, nephrolithiasis, and renal failure. Mutations found in Dent disease and other disorders were demonstrated by electrophysiological analysis to abolish the chloride channel function of CLC-5 when expressed in *Xenopus* oocytes, providing evidence of correlation between the molecular lesions and their altered function.

Nine members of voltage-gated chloride channels have been found in mammals (reviewed by Jentsch and Günther, 1997). Among them, three members, CLC-1, CLC-Kb, and CLC-5, are reported to be implicated in human hereditary diseases [myotonia (Koch *et al.,* 1992; George *et al.,* 1993), Bartter syndrome type 3 (Simon *et al.,* 1997). and Dent disease, respectively]. Interestingly, the expression patterns of these three genes are tissue specific, whereas the rest of the gene family is expressed ubiquitously. The transcript of human *CLCN5* is 9.5 kb in size (Fisher *et al.,* 1994), suggesting the existence of a large untranslated region(s), UTR, which may play a role in the expression and regulation of the gene. The mechanism of the kidney-specific expression of *CLCN5* is of considerable interest and provides the motivation for an examination of the nature of the UTR(s) and the promoter region.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AF134117 and AF134118.

[1] To whom correspondence should be addressed. Telephone: +44 1865 275326. Fax: +44 1865 275318.

0888-7543/99 $30.00

An alternative approach to analyzing the relationship between gene structure and function of its product is by comparison with the homologous locus in other species. Next to human, the mouse is genetically the best characterized mammal and the one most commonly used to produce models for human genetic disorders. The characterization of the mouse homologue of a gene responsible for human disease therefore represents an important step in the investigation of disease etiology, and isolation of the murine equivalents of both coding and regulatory regions is essential for this purpose. In this study, we report the isolation and characterization of the mouse homologue of *CLCN5* and the genomic organization of the locus, including the identification of the promoter and a putative enhancer region. Mapping of the transcription start site and identification of the *Clcn5* promoter region represent the first step in the elucidation of the mechanisms regulating this gene. These data are significant for the comparative mapping of human and mouse genomes and are an important prerequisite for the future genetic modification of the gene in mouse to produce an animal model for the human kidney disorder.

MATERIALS AND METHODS

Isolation and characterization of cDNA clones. To obtain the *Clcn5* cDNA sequence, a mouse kidney cDNA library (Clontech) was screened using human *CLCN5* cDNA fragments spanning the coding exons and flanking untranslated regions. A total of 2–5 $\times$ 10^5 recombinants propagated in C600 bacteria were plated on five 20 $\times$ 20-cm agar plates and adsorbed to Hybond-N^+ (Amersham) nylon membranes in duplicate, according to the manufacturer's instructions. The human cDNA fragments were purified and [α-^{32}P]dCTP labeled by random priming (Feinberg and Vogelstein, 1983) before the hybridization at 65°C in Church buffer (Church and Gilbert, 1984). The filters were washed sequentially with increasing stringency and exposed to X-ray film overnight. Phage plaques giving rise to duplicate positive signals were isolated as single plaques after tertiary screening. Phage DNA was isolated using Wizard Lambda Minipreps (Promega). The cDNA inserts were subcloned into plasmids using standard techniques (Sambrook *et al.,* 1989) and sequenced on both strands by the dideoxy method (Sanger *et al.,* 1977) with Sequenase v2.0 (USB) using universal and reverse primers, as well as primers derived from the obtained sequence.

5′ RACE analysis. 5′ RACE was carried out using Marathon-Ready cDNA (Clontech) from mouse kidney mRNA. The anchor-ligated cDNA library was PCR amplified using the anchor primer AP1 (5′-CCATCCTAATACGACTCACTATAGGGC-3′) and a gene-specific antisense primer, C11R1 (5′-TCTCGGTGCCTATCTCGGT-3′), from the most 5′ sequence of the newly isolated cDNA clones. Conditions were as follows: 350 μM dNTPs, 50 mM Tris–Cl (pH 9.2), 160 mM $(NH_4)_2SO_4$, 1.75 mM $MgCl_2$, 1.0 μM each primer, 2.5 U *Taq* and *Pwo* DNA polymerase mix (Boehringer). Cycling parameters were 94°C for 2 min, 94°C for 30 s, 72°C for 1 min, 4 cycles; 94°C for 30 s, 70°C for 1 min, 4 cycles; 94°C for 20 s, 68°C for 1 min, 32 cycles; 68°C for 5 min. PCR-amplified products were purified using Wizard PCR preps (Promega) and directly cloned into the pCRII vector (Invitrogen). Several clones were sequenced using Sequenase v2.0 (USB).

Northern analysis. A mouse multiple-tissue Northern blot containing poly(A)-enriched RNA was obtained from Clontech. Hybridization and washing were according to manufacturer's instructions. Two cDNA fragments from mouse *Clcn5* clones, C1 and C2 (1.6 and 1 kb in size), were gel purified, [α-^{32}P]dCTP-labeled, and hybridized to the blot. The filter was exposed to an X-ray film for several days. After the hybridization with the mouse *Clcn5* cDNA, a β-actin cDNA fragment was reprobed to the same blot as an internal control to normalize levels of mRNAs.

Isolation and mapping of genomic clones. A mouse 129 genomic library in the EMBL3 vector was screened using mouse cDNA clones as a probe. The 5′ and the 3′ ends of the resulting clones were isolated and used to rescreen the library, until a contig spanning the entire coding and the upstream regions of *Clcn5* was assembled. A restriction enzyme map was characterized by consecutive hybridization with cDNA fragments to genomic DNA. The phage DNA was completely digested with *Bam*HI, *Eco*RI, *Hin*dIII, and *Sal*I, separated on 1% agarose gel, and transferred onto the Hybond-N^+ nylon membrane. Each cDNA clone was hybridized individually and exons were mapped onto genomic fragments. Alternatively, the phage DNA was partially digested with *Bam*HI, *Eco*RI, or *Hin*dIII at different concentrations of the enzymes and hybridized with an end clone from the vector to ascertain the order of each genomic fragment. Intron–exon boundaries were identified by comparing the sizes of PCR products from cDNA and genomic DNA amplified with exonic primers. Various sets of primers were selected from different regions of the transcript. The sequences of the intron–exon boundaries were determined using exonic primers flanking introns to sequence genomic fragments subcloned in plasmids.

Construction of reporter gene plasmids. To localize the transcriptional control region of the mouse *Clcn5* gene, four overlapping genomic fragments, SBB6kb, BBH2.4kb, BB5kb, and HB5kb, from the 5′ flanking sequence were subcloned from phage genomic clones into a promoterless, firefly luciferase expression vector, pXP2 (Nordeen, 1988), in the forward orientation. The fragment HB5kb, containing exons 1 and 2, was subcloned into the same vector, but in the reverse orientation to investigate the orientation dependence of its promoter activity.

Deletion analysis of the Clcn5 promoter. The complete sequence of the *Hin*dIII/*Xho*I 4-kb fragment (pHX4kb) was obtained using dye terminator cycle sequencing (Prober *et al.,* 1987) with Thermo Sequenase (Amersham). Primers H1 (5′-TTTAAGCTTGCCTGCCATTAGG-3′), H2 (5′-GTAATTCTAAAGCTTCTGGATGT-3′), H3(5′-CCTGAGCT-GACAAAGCTTCC-3′) and H4 (5′-CTGACAGGAAGCTTTTGCAGAG-3′), containing internal *Hin*dIII sites (underlined), were synthesized, and the corresponding genomic fragment was PCR amplified with each of these "sense" primers and an "antisense" primer, C11R1, localized 30 bp downstream of *Xho*I in exon 2. The *Xho*I site at the 3′ end of exon 2 was utilized for cloning fragments into the reporter vector pXP2 in the forward orientation. High-fidelity PCR *Taq* and *Pwo* mixed DNA polymerases (Boehringer) were used for amplification reactions. The corresponding 3-, 2.7-, 2.6-, and 2-kb PCR products were double-digested with *Hin*dIII and *Xho*I endonucleases and cloned into the luciferase expression vector, pXP2, to create the constructs designated pH1X, pH2X, pH3X, and pH4X. pBgX was prepared by subcloning *Bgl*II/*Xho*I 874-bp fragment into the vector. pBgBg and pH1Bg were generated by omitting the *Bgl*II/*Xho*I 874-bp fragment from the constructs pHX4kb and pH1X. Promoter activities were measured for at least two independent clones in duplicate.

Transfections and luciferase assays. Monkey kidney COS-6 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5% fetal calf serum (FCS). Cells (5 $\times$ 10^4) were seeded in 24-well plates and grown to 80% confluency. Transient transfections were carried out with about 500 ng of the luciferase reporter construct and pSV40 β-galactosidase plasmid using the calcium phosphate method (Graham and van der Eb, 1973). DNA amounts were normalized by comparison, under UV transillumination, of standard quantities of an *Eco*RI fragment in the luciferase vector after staining with ethidium bromide in agarose gels. After 16 h, cells were rinsed with phosphate-buffered saline (PBS) and allowed to recover in DMEM–5% FCS for further 24 h. Cells were washed with PBS and harvested into 50 μl of Lysis buffer (Promega). Five microliters of cell extract was mixed with 50 μl of luciferase assay reagent (Promega) and light production quantified using a

Turner Designs Model 20 luminometer. β-Galactosidase activity was assayed using an enzyme assay system (Promega), and then absorbance was measured at 420 nm.

Primer extension. Primer extension was performed to identify the putative transcription start site. Total RNA was isolated from mouse kidney according to the guanidium isothiocyanate method (Chomczynski and Sacchi, 1987), and 20 μg was annealed at 55°C to an antisense primer ExR3 (5′-AGCTTTAGAAGGGCACTTGCTGC-3′) that had been end labeled with [γ-^{32}P]ATP. Primer extension was then performed in the presence of 25 mM Tris–Cl (pH 8.3), 5 mM DTT, 37.5 mM KCl, 15 mM $MgCl_2$, 0.5 mM dNTP, and 200 U MMLV reverse transcriptase for 60 min at 37°C. The extension products generated were compared to a sequencing ladder produced from the same primer on genomic DNA. Samples were resolved on 8% polyacrylamide–urea sequencing gels.

Sequence analysis. Sequence analysis by computer was performed using the program of the Wisconsin GCG package (Devereux *et al.*, 1984). For prediction of potential regulatory elements, the Signal Scan program (Prestridge, 1991) and the Transcription Element Search Software TESS (Schug and Overton, 1997) were employed and the results compared.

RESULTS

A cDNA Contig Spanning the Mouse Clcn5 Open Reading Frame

Thirteen cDNA clones were isolated from a mouse kidney cDNA library by screening 2–5 × 10^5 recombinants with human *CLCN5* cDNA fragments. Each clone was sequenced and the three overlapping cDNA clones, C1, C2, and C11, were found to cover 90% of the open reading frame (ORF) of mouse *Clcn5*, but the contig lacked the 5′ end of the transcript (Fig. 1a). One of the other clones, C7, was a mouse homologue of chloride channel 4, *Clcn4* (Rugarli *et al.*, 1995), which was presumably detected as a result of cross hybridization to the probe.

To obtain the complete sequence of the mouse *Clcn5* ORF and the upstream cDNA sequence, 5′ RACE was performed using an anchor-ligated cDNA library originating from adult mouse kidney mRNA. A gene-specific antisense primer, C11R1, was therefore designed from the 5′ sequence of the cDNA clone, C11, and used together with an anchor primer, AP1, to amplify further the 5′ end of the transcript. The products were purified and subcloned directly into a TA cloning vector, pCRII. A portion of the products was also analyzed by Southern blot hybridization using a nested cDNA probe that was localized upstream of the C11R1 primer. A signal was observed at around 350 bp (data not shown), indicating a single transcription start site. The subcloned PCR products ranged from 300 to 350 bp resulting from the truncation of 5′ ends. A clone with the longest fragment of 30 independently isolated was selected (termed C0) and sequenced. It contained a translation start and further upstream cDNA sequence (Fig. 1a). This allowed the completion of a contig of four cDNA clones, C0, C11, C2, and C1, covering the entire coding region of mouse *Clcn5*.

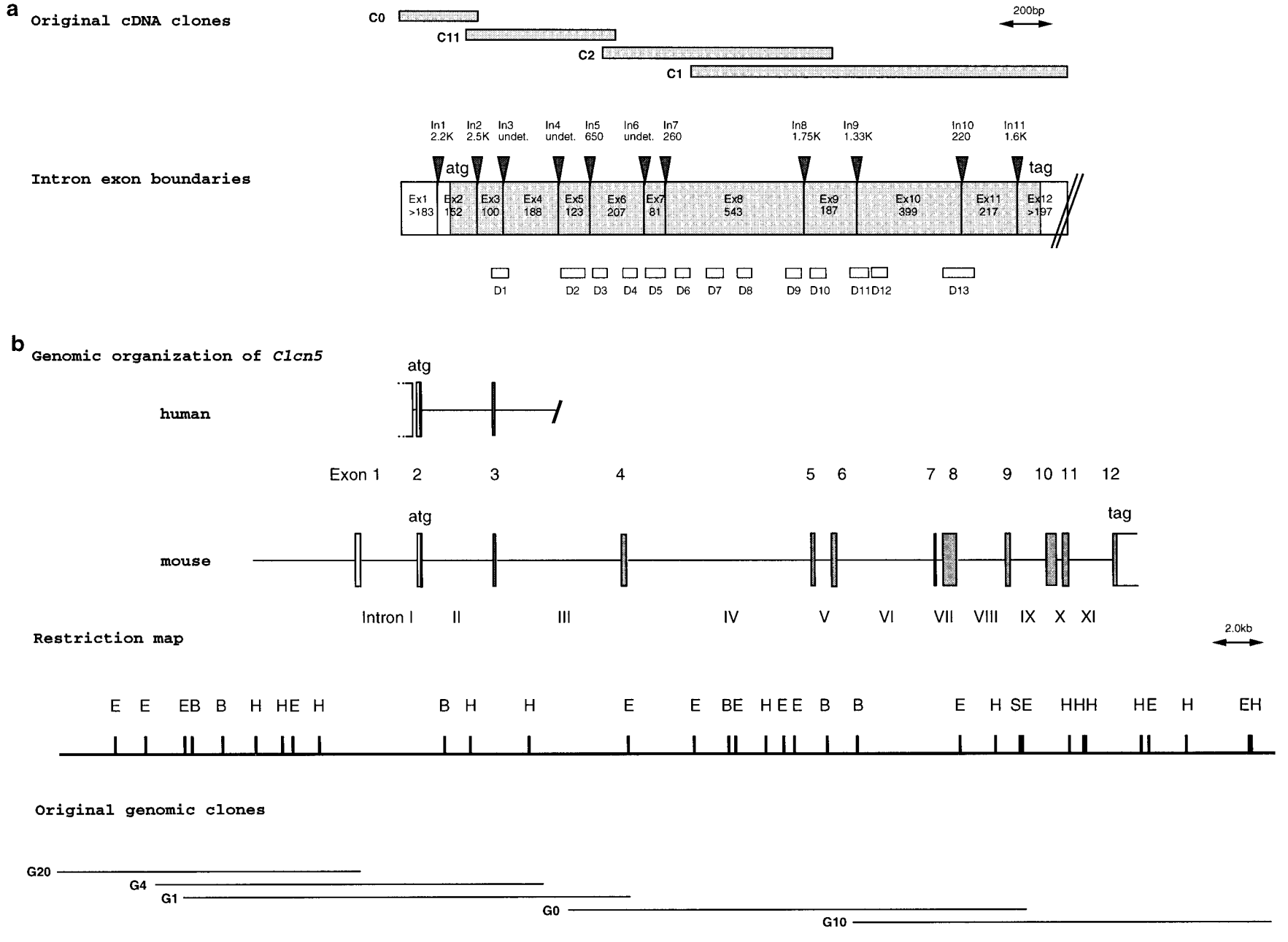
Sequencing of clones C0, C1, C2, and C11 provided a continuous string comprising 2577 bp of the cDNA sequence containing 2238 bp of the ORF. The nucleotide sequence of the ORF is >90% homologous to that in human. Interestingly, the first exon, as reported for the human cDNA clone (Fisher *et al.*, 1995), is intronic (removed by splicing) in mouse, and the second exon is linked to a novel, untranslated exonic sequence of 183 bp. The sequences obtained contain very little (109 bp) of the 3′ UTR and no polyadenylation consensus signal.

The predicted protein of 746 amino acids shares 98% homology with its human counterpart (Fig. 2). A hydropathy plot (Kyte and Doolittle, 1982) revealed a pattern of transmembrane domains similar to that observed in all the other reported members of the gene family (Jentsch *et al.*, 1995). Fifteen amino acid substitutions were found between the human CLC-5 protein and its mouse equivalent, most of which are located outside the transmembrane domains, suggesting that these substitutions do not alter the overall topology. Two potential N-linked glycosylation sites were found at the same sites as observed for the human amino acid sequence.

Genomic Organization of the Mouse Clcn5 Coding and 5′ Upstream Regions

Recombinants (3 × 10^5) from a 129 mouse strain genomic library were screened using mouse cDNA fragments, and five overlapping clones, G20, G4, G1, G0, and G10, were isolated. These were analyzed by restriction endonuclease digest mapping and Southern blot hybridization (Fig. 1b). Since the genomic library is assembled in the phage vector EMBL3, which contains *Sal*I, *Eco*RI, and *Bam*HI as cloning sites, these restriction enzymes were first chosen to characterize a detailed restriction map of the *Clcn5* locus. *Hin*dIII sites were also determined. Mapping with probes derived from different regions of the mouse *Clcn5* cDNA demonstrated that the entire coding region and the 5′ end of the gene are encompassed in an ~30-kb genomic region. This coverage is extended to include the immediate 5′ flanking sequence, by a fifth clone, G20. In total, we have isolated ~45 kb of genomic DNA from the mouse *Clcn5* locus. Intron–exon boundaries were identified by a PCR strategy as previously described (Fisher *et al.*, 1995). When a set of primers designed from the cDNA embrace one intron, or several exons and introns, a larger sized product, or no product, will be detected after the genomic template amplification. By comparing the PCR products from genomic template versus cDNA template, 11 introns were identified and their sizes were estimated. All intron–exon boundaries were sequenced in subcloned genomic fragments using flanking primers, and the sizes of exons were determined (Fig. 1a).

The genomic organization of the gene encoding CLC-5 is highly conserved between human and mouse (Fig. 1b). The intron–exon boundaries in the mouse *Clcn5* coding region were found at positions equivalent to those observed for the human cDNA sequence (Fisher *et al.*,


a
Original cDNA clones
C0
C11
C2
C1
200bp
Intron exon boundaries
In1 2.2K
In2 2.5K
In3 undet.
In4 undet.
In5 650
In6 undet.
In7 260
In8 1.75K
In9 1.33K
In10 220
In11 1.6K
atg
tag
Ex1 >183
Ex2 152
Ex3 100
Ex4 188
Ex5 123
Ex6 207
Ex7 81
Ex8 543
Ex9 187
Ex10 399
Ex11 217
Ex12 >197
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
D11D12
D13
b
Genomic organization of Clcn5
human
atg
mouse
Exon 1
2
3
4
5
6
7
8
9
10
11
12
Intron I
II
III
IV
V
VI
VII
VIII
IX
X
XI
Restriction map
2.0kb
E E EB B H HE H B H H E E BE H E E B B E H SE HHH HE H EH
Original genomic clones
G20
G4
G1
G0
G10

```
                                                *
mouse   1 MDFLEEPIPGVGTYDDFNTIDWVREKSRDRDRHREITNKSKESTWALIHS 50
          ||||||||||||||||||||||||||||||||||||||||||||||||||
human   1 MDFLEEPIPGVGTYDDFNTIDWVREKSRDRDRHREITNKSKESTWALIHS 50
                    .         .         .         *         .
       51 VSDAFSGWLLMLLIGLLSGSLAGLIDISAHWMTDLKEGICTGGFWFNHEH 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||
       51 VSDAFSGWLLMLLIGLLSGSLAGLIDISAHWMTDLKEGICTGGFWFNHEH 100
                    .         .         .         .         .
                         D1
      101 CCWNSEHVTFEHRDKCPEWNSWAQLIINTDQGAFAYIVNYFMYVLWALLF 150
          ||||||||||.||||||||||.||||.|:|||||||||||||||||||||
      101 CCWNSEHVTFEERDKCPEWNSWSQLIISTDEGAFAYIVNYFMYVLWALLF 150
                    .         .         .         .         .
                                                        D2
      151 AFLAVSLVKAFAPYACGSGIPEIKTILSGFIIRGYLGKWTLVIKTITLVL 200
          |||||||||.|||||||||||||||||||||||||||||||||||||||||
      151 AFLAVSLVKVFAPYACGSGIPEIKTILSGFIIRGYLGKWTLVIKTITLVL 200
                    .         .         .         .  D3     .
      201 AVSSGLSLGKEGPLVHVACCCGNILCHCFNKYRKNEAKRREVLSAAAAAG 250
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      201 AVSSGLSLGKEGPLVHVACCCGNILCHCFNKYRKNEAKRREVLSAAAAAG 250
                    .         .(D4)     .         .         .
      251 VSVAFGAPIGGVLFSLEEVSYYFPLKTLWRSFFAALVAAFTLRSINPFGN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      251 VSVAFGAPIGGVLFSLEEVSYYFPLKTLWRSFFAALVAAFTLRSINPFGN 300
        D5  .         .         .         .  D6     .
      301 SRLVLFYVEFHTPWHLFELVPFIVLGIFGGLWGALFIRTNIAWCRKRKTT 350
          ||||||||||||||||||||:|||||||||||||||||||||||||||||
      301 SRLVLFYVEFHTPWHLFELVPFILLGIFGGLWGALFIRTNIAWCRKRKTT 350
                    .         .         .  D7     .         .
      351 QLGKYPVVEVLIVTAITAILAFPNEYTRMSTSELISELFNDCGLLDSSKL 400
          |||||||:|||:||||||||||||||||||||||||||||||||||||||
      351 QLGKYPVIEVLVVTAITAILAFPNEYTRMSTSELISELFNDCGLLDSSKL 400
                *   .    D8   .         .         .         .
      401 CDYENHFNTSKGGELPDRPAGVGIYSAMWQLALTLILKIVITIFTFGMKI 450
          |||||:|||||||||||||||||:|||||||||||||||||||||||||||
      401 CDYENRFNTSKGGELPDRPAGVGVYSAMWQLALTLILKIVITIFTFGMKI 450
                *   .         .         .  D9     .         .
      451 PSGLFIPSMAVGAIAGRLLGVGMEQLAYYHHDWGIFNSWCSQGADCITPG 500
          |||||||||||||||||||||||||||||||::|.:|||||||||||||||
      451 PSGLFIPSMAVGAIAGRLLGVGMEQLAYYHQEWTVFNSWCSQGADCITPG 500
                  D10     .         .         .         .
      501 LYAMVGAAACLGGVTRMTVSLVVIMFELTGGLEYIVPLMAAAMTSKWVAD 550
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      501 LYAMVGAAACLGGVTRMTVSLVVIMFELTGGLEYIVPLMAAAMTSKWVAD 550
                    .  D11    .         .         D12       .
      551 ALGREGIYDAHIRLNGYPFLEAKEEFAHKTLAMDVMKPRRNDPLLTVLTQ 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      551 ALGREGIYDAHIRLNGYPFLEAKEEFAHKTLAMDVMKPRRNDPLLTVLTQ 600
                    .         .         .         .         .
      601 DSMTVEDVETIISETTYSGFPVVVSRESQRLVGFVLRRDLIISIENARKK 650
          ||||||||||||||||||||||||||||||||||||||||||||||||||
      601 DSMTVEDVETIISETTYSGFPVVVSRESQRLVGFVLRRDLIISIENARKK 650
                    .         .         .         .         .
      651 QDGVVSTSIIYFTEHSPPMPPYTPPTLKLRNILDLSPFTVTDLTPMEIVV 700
          ||||||||||||||||||:|||||||||||||||||||||||||||||||
      651 QDGVVSTSIIYFTEHSPPLPPYTPPTLKLRNILDLSPFTVTDLTPMEIVV 700
                    .         .         .         .         .
      701 DIFRKLGLRQCLVTHNGRLLGIITKKDVLKHIAQMANQDPDSILFN 746
          |||||||||||||||||||||||||||||||||||||||||||||||
      701 DIFRKLGLRQCLVTHNGRLLGIITKKDVLKHIAQMANQDPDSILFN 746
                     (D13)
```

FIG. 2. Translation of the nucleotide sequence of mouse *Clcn5* into amino acid sequence (**top**) and its comparison to the human counterpart (**bottom**) from Fisher *et al.* (1995). Fifteen amino acids that differ between human and mouse are indicated in bold. Vertical bars indicate identical amino acids. Amino acids whose comparison value is ≥0.5 are marked with a colon and those with ≥0.1 are indicated with a period. Thirteen hydrophobic domains are underlined (D1–D13). Two potential N-linked glycosylation sites (*) are marked at the same conserved position.

1995). The intron sizes separating coding exons are also similar to each other in these two species. Nucleotide sequences immediately flanking the splice sites show close agreement with the consensus splice donor and

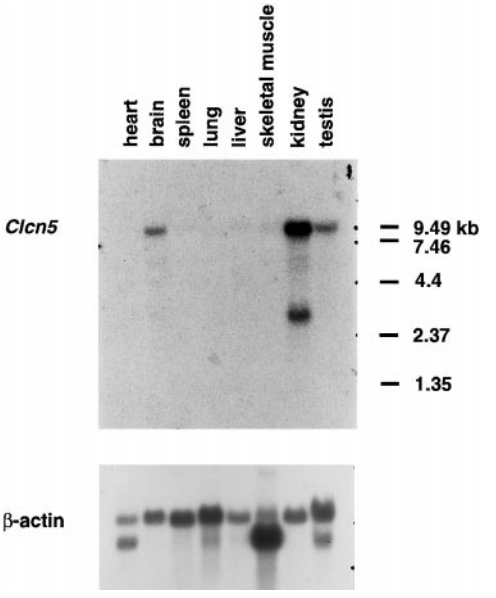


FIG. 3. Northern blot analysis of poly(A)-enriched mRNA from various mouse tissues. The blot was probed with mouse cDNA clones C1 and C2, spanning almost the entire open reading frame. The 9.5- and 3-kb transcripts of *Clcn5* are expressed predominantly in kidney. Expression of *β*-actin is also shown as an internal control. Positions of RNA molecular weight markers are indicated.

acceptor sequences (Senapathy *et al.*, 1990). All *Clcn5* introns belong to the predominant vertebrate splicing GT–AG class of introns (data not shown). The first exons in human *CLCN5* and mouse *Clcn5*, both of which are untranslated, however, revealed features different from each other. Unlike the first untranslated exon in human, which is separated by a 136-bp intron from the adjacent coding exon, the first exon in mouse is localized 2.2 kb upstream of the second exon (Fig. 1b). Although the mouse genomic sequence at the position corresponding to the human 5′ splice site of intron 1 retains the features of a consensus donor splice motif, the surrounding intronic sequences are less homologous to each other and appear to result in the different splicing in the two species.

Tissue-Specific Expression of Mouse Clcn5

Northern blot analysis for expression of mouse *Clcn5* revealed a transcript of 9.5 kb in size, predominantly expressed in kidney, as well as in brain and testis (Fig. 3). Very low expression was also detected in liver. In kidney, a distinct transcript of 3 kb was also detected. The most closely related members of the gene family, CLC-3 and CLC-4, have transcripts of 3 and 5.5 kb, and

FIG. 1. (**a**) Schematic representation of mouse *Clcn5* cDNA. Four overlapping cDNA clones, C0, C11, C2, and C1, are drawn to scale (top). Intron–exon boundaries (inverted triangles) were identified by PCR and sequencing strategies (middle). The initiator methionine is represented by "atg" and the stop codon by "tag," with the open reading frame shaded. The numbers represent the sizes (in basepairs) of exons and introns. Thirteen hydrophobic domains, D1–D13, according to the original nomenclature for CLC-0 (Jentsch *et al.*, 1990), are also shown at the bottom. D4 and D13 (open boxes) are believed not to span the membrane. (**b**) Genomic organization of mouse *Clcn5* was determined by hybridization analysis and PCR studies. The divergent structure in human *CLCN5* at the 5′ upstream region is adapted from Fisher *et al.* (1995). Intron–exon structure correlates with the restriction map, which was assembled from five overlapping lambda clones, G20, G4, G1, G0, and G10 (bottom). The boxes represent coding exons (shaded) and untranslated region (open), and the line represents introns. Restriction sites shown are B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sal*I.

5 and 7.5 kb in rat (Kawasaki *et al.*, 1994; Jentsch *et al.*, 1995), and their tissue distributions are broad. It is, therefore unlikely that the 3-kb transcript of *Clcn5* observed in mouse kidney is due to a signal cross-hybridized to *Clcn3* or *Clcn4*. The expression of the 9.5-kb transcript is highly dominant in kidney, which reflects the tissue specificity of the gene as observed in our previous study in human (Fisher *et al.*, 1994) and recent reports in rats (Steinmeyer *et al.*, 1995; Sakamoto *et al.*, 1996).

Functional Analysis of the 5′ Upstream Region of Mouse Clcn5

To identify the control region of the mouse *Clcn5* gene, four overlapping genomic fragments, encompassing 15 kb of the 5′ flanking region, were cloned into the promoterless vector pXP2 carrying the luciferase reporter gene in the forward orientation (Fig. 4). Constructs pSBB6kb and pBBH2.4kb contain only the 5′ flanking sequences, lacking the newly identified exon 1. Construct pBB5kb contains 5 kb of the 5′ flanking sequence and exon 1, while pHB5kb contains ~1.3 kb of the 5′ flanking sequence, exon 1, exon 2, and 1 kb of the adjacent sequence extending into intron 2. These four constructs were transiently transfected into monkey kidney COS-6 cells to analyze their promoter activity. An SV40-driven β-galactosidase reporter plasmid was cotransfected with the luciferase test constructs, and the values obtained were normalized using the β-galactosidase activity to correct for transfection efficiency.

Luciferase activity was absent in the constructs pSBB6kb and pBBH2.4kb, which showed the same level of light emission as the basal, promoterless pXP2 vector. Construct pBB5kb showed 50–100 times higher luciferase activity than did pXP2. When ~3.5 kb of the 5′ sequence was omitted from this construct by digestion at *Hin*dIII, the remaining 1.6-kb *Hin*dIII/*Bam*HI fragment sustained the same level of luciferase activity (data not shown). This confirms the location of a putative promoter within ~1.3 kb of upstream sequence from the first exon. The highest luciferase activity was achieved by the construct pHB5kb, which contains the 5-kb *Hin*dIII/*Bam*HI fragment. Its activity was $\sim 10^3$ times higher than that of pXP2 and 10–20 times higher than that of pBB5kb. When the 5-kb *Hin*dIII/*Bam*HI fragment was linked to the luciferase gene in the reverse orientation, the luciferase activity was abolished to the basal level. This orientation-dependent regulation of luciferase expression is consistent with the existence of a promoter element within the fragment.

Further Analysis of the Promoter Region

Further characterization of the promoter region was performed by deletion analysis of the *Hin*dIII/*Xho*I 4-kb fragment (Fig. 5). The complete sequence of the fragment encompasses 1.3 kb of the upstream sequence and the first exon, in addition to 2.6 kb of its downstream sequence containing the entire intron 1 and the first 132 bp of exon 2. This construct, pHX4kb, lacks 1 kb from the 3′ end of pHB5kb, but retains the same level of activity, demonstrating $\sim 10^3$ times higher light emission in comparison to the promoterless pXP2 vector (data not shown). We engineered luciferase reporter constructs containing various lengths of the 5′ flanking sequences extending from the *Xho*I site in exon 2. Three forward primers, H1, H2, and H4, were designed to produce various deletions at the 5′ end in PCR products amplified with an antisense primer C11R1 (see Materials and Methods). The promoter activity of each construct is shown in Fig. 5 relative to the activity of pHX4kb (containing sequences from −1319 to +2610 relative to the transcription start site; see below). Consecutive truncations downstream of the *Hin*dIII site caused their promoter activities to progressively decrease; the activity reduced to 54% in pH1X (−399/+2610), and pH2X (−128/+2610) showed a further reduction to 25%. This activity in pH2X (−128/+2610) is, however, still 10- to 20-fold higher than that of pH4X (+497/+2610), which lacks exon 1, and indeed is more than 10^2-fold higher compared to the promoterless pXP2 vector. We therefore concluded that the construct pH2X (−128/+2610) retains the promoter activity and that the sequence from primer H2 to the first untranslated exon contains the minimal promoter region. A consensus TATA for a potential element was identified at −69 to −64 within this region (Fig. 6). To examine if this TATA sequence is functional, we generated another construct, pH3X (−65/+2610), using primer H3, which introduced a further deletion including the putative TATA box. This resulted in a substantial decrease in the luciferase activity, although the activity was higher than those of pH4X (+497/+2610) and pBgX (+1737/+2610), suggesting that the sequences around the TATA motif are important, but not absolutely necessary, for the activity of the promoter.

Another characteristic of this region is that the promoter activity was abolished in pBgBg (−1041/+1736) and pH1Bg (−399/+1736), derived by deletion of the 874-bp *Bgl*II/*Xho*I fragment (from +1737 to +2610) containing parts of the first intron (742 bp) and exon 2 (132 bp) from two active constructs, pHX4kb (−1319/+2610) and pH1X (−399/+2610), respectively (Fig. 5). This indicated the presence of the elements responsible for strong enhancer activity within the 874-bp sequence. The construct pBgX (+1737/+2610) itself represents 1–2% of the promoter activity relative to pHX4kb (−1319/+2610). The levels of activity in pBgX (+1737/+2610) and pH4X (+497/+2610) are presumed to represent endogenous background induced by this enhancer element. The maximal expression of the reporter gene was observed only when the promoter was brought together with the enhancer element in this 874-bp *Bgl*II/*Xho*I sequence.

Transcription Start Site

Primer extension was performed to identify the transcription start site(s) using an antisense primer, ExR3,

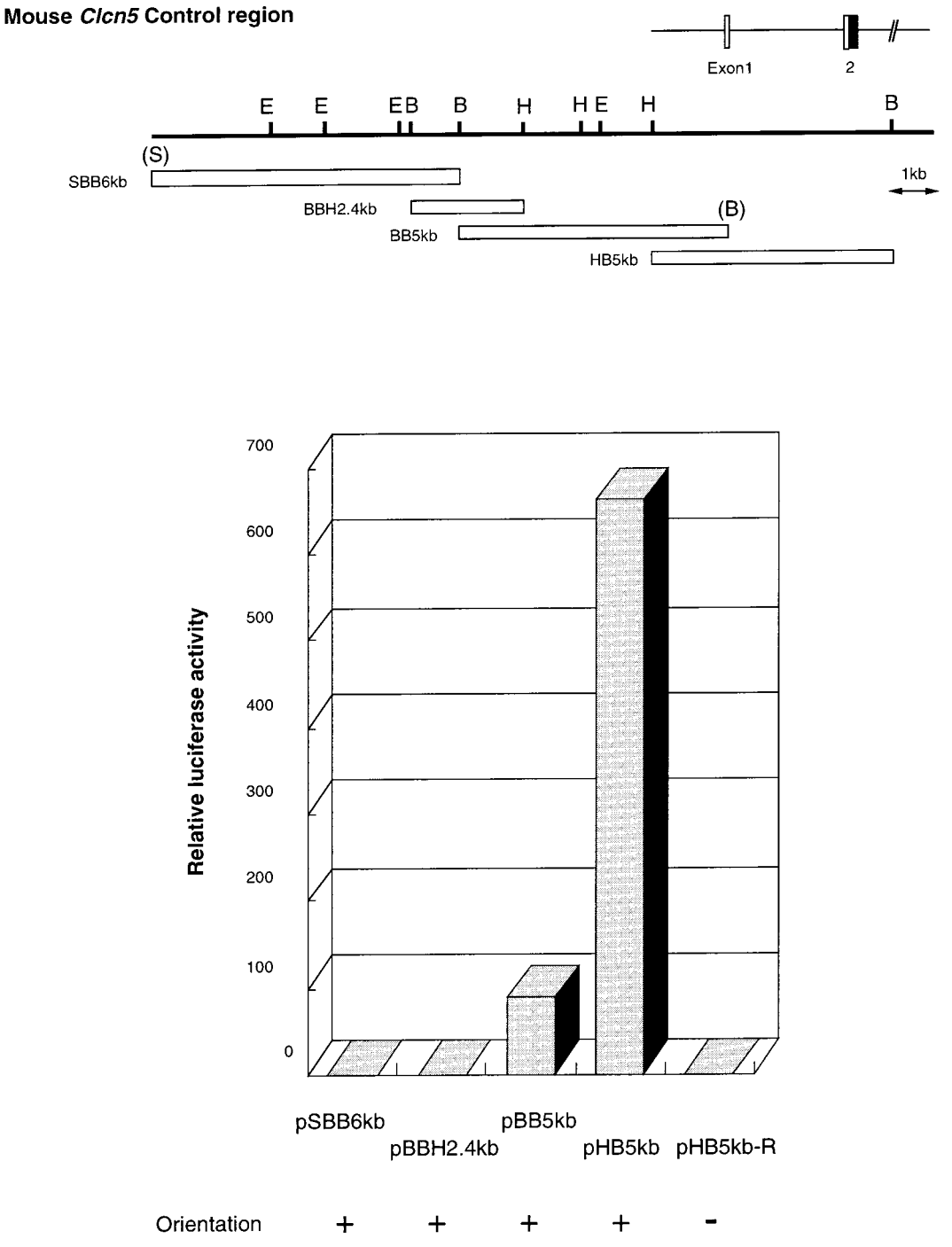


FIG. 4. Functional analysis of the 5′ flanking sequences. Four genomic fragments (SBB6kb, BBH2.4kb, BB5kb, and HB5kb) were subcloned upstream of the promoterless luciferase gene in the pXP2 vector in the forward orientation and their promoter activities were assayed in COS-6 cells. The highest promoter activity was identified in pHB5kb, which contains the region from ~1.3 kb 5′ of exon 1 to intron 2. The construct pHB5kb-R contains the same fragment in the reverse orientation. The luciferase activity from each construct was normalized to β-galactosidase activity from a cotransfected internal control plasmid pSV40-β-galactosidase. The results are expressed as values relative to that of the pXP2 promoterless vector. The histogram shows the average of duplicate results from three independent transfection experiments. Restriction sites shown are B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sal*I. (S) and (B) are from the polylinker of the λ phage vector.

derived from the sequence of the 5′ RACE product. Extension with ExR3 revealed a product of 180 bp that mapped a single transcription start site to a position 103 bp upstream of the beginning of the 5′ RACE clone, C0 (Fig. 6). Computer analysis predicted one putative CAP site at CATCTT, which overlaps the transcription start site determined by primer extension. The sequence partially matched with the transcription initiator (Inr) element consensus, $YCA_{+1}NWYY$ (Bucher, 1990; Javahery *et al.*, 1994). This site is followed by a

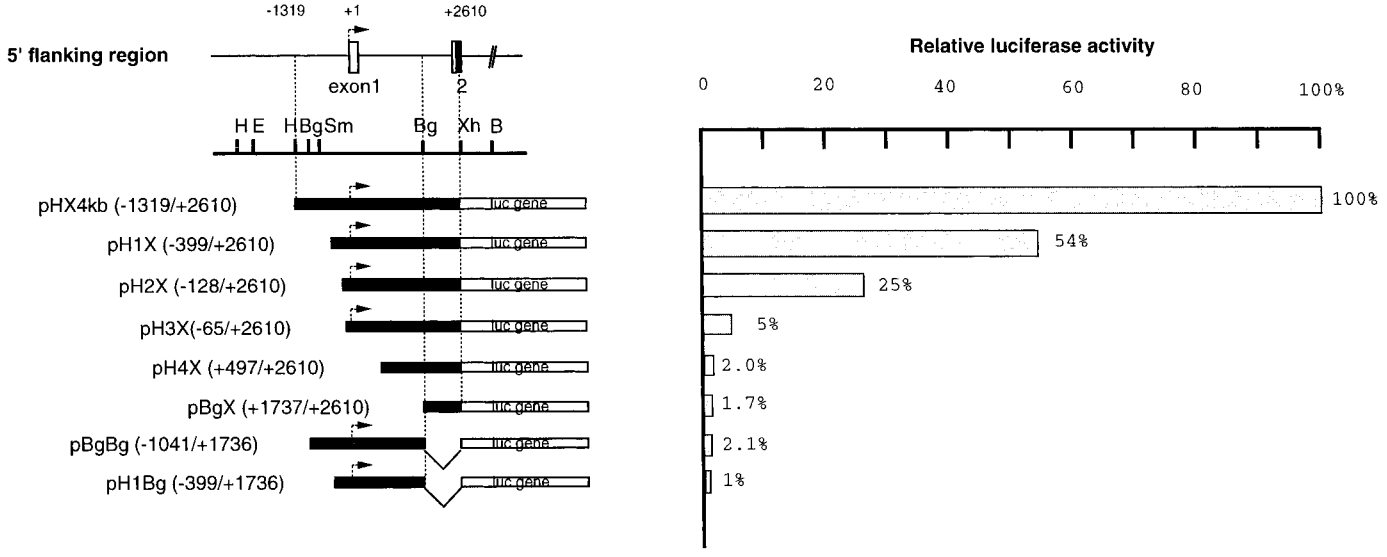


FIG. 5. Deletion analysis of the putative mouse *Clcn5* promoter. Promoter activity was measured for various constructs containing the 5′ flanking sequence of the gene (**left**). A putative transcription start site determined by the primer extension assay is marked with an arrow. Numbers indicate 5′/3′ ends relative to the transcription start site (+1). Restriction sites shown are B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; Bg, *Bgl*II; Sm, *Sma*I; Xh, *Xho*I. The relative luciferase activity of each construct is shown on the **right.** The promoter activity initially identified in pHX4kb (−1319/+2610) was eventually narrowed down within pH2X (−128/+2610). The high activity in pHX4kb (−1319/+2610) and pH1X (−399/+2610) was reduced dramatically when the 874-bp *Bgl*II/*Xho*I fragment (+1737/+2610) was omitted (as shown in pBgBg and pH1Bg), suggesting the existence of enhancer elements within the fragment. The histogram shows the average of the results from four independent transfection experiments, expressed as percentages of the activity of the pHX4kb construct (100%).

CT signal (consensus CTNCNG), at CTTCTG, which is commonly found, on average, 7 nucleotides downstream of the transcription start site (Larsen *et al.*, 1995). These results imply the existence of a single putative transcription start site at the A residue localized at −333 bp from the translation start codon.

Sequence Analysis of the Mouse Clcn5 Promoter

Computer analysis of the mouse *Clcn5* promoter sequence revealed several potential regulatory elements (Fig. 6). In particular, a TATA motif was found at −69 from the transcription start site; however, the region

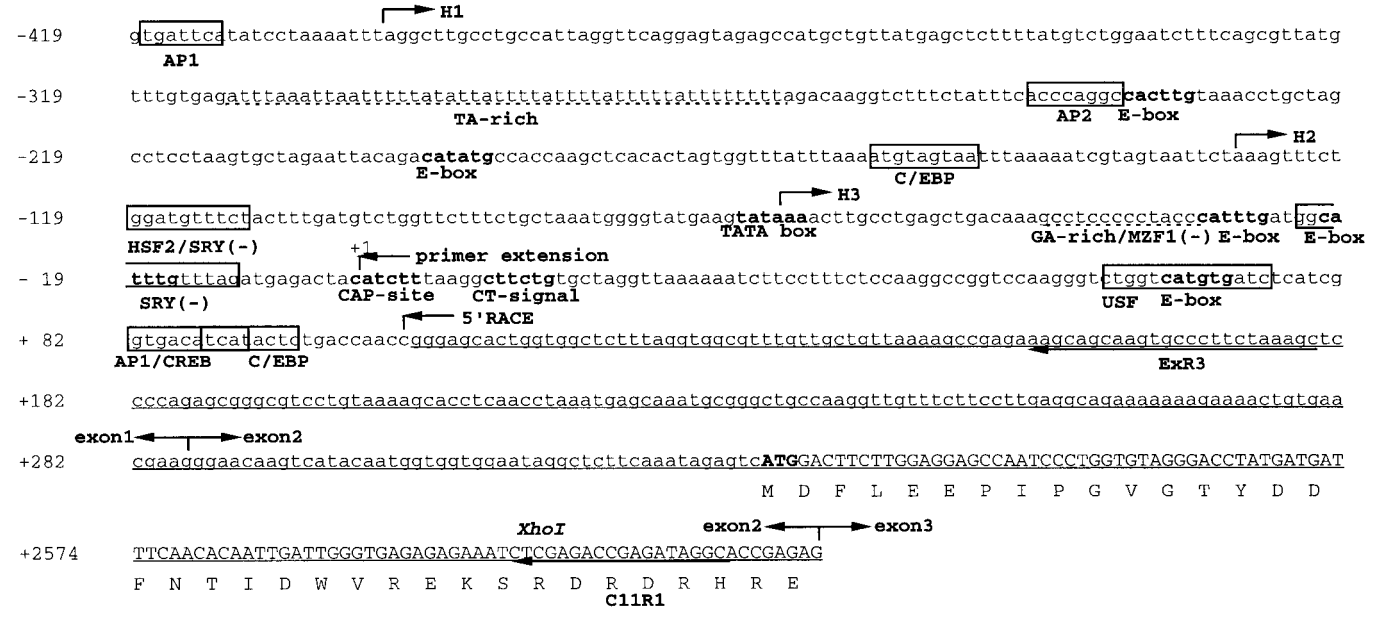


FIG. 6. The 5′ genomic sequence and cDNA sequence from the promoter region of mouse *Clcn5*. cDNA sequence of the 5′ RACE product is shown linked directly to the 5′ flanking genomic sequence. The sequence of the first intron from +287 to +2478 is omitted. The translated region is shown in uppercase letters. Exonic sequence from the 5′ RACE product is underlined. Positions of the primers used for preparation of luciferase test constructs (H1, H2, and H3) and for the primer extension (ExR3) are designated by arrows. A putative transcription start site has been mapped by primer extension and is indicated by an arrow at +1. The minimal promoter region has been localized at the 128-bp sequence between the transcription start site and the H2 primer position. CAP site, CT signal, TATA box, and E boxes are shown in bold. Purine-rich sequence and TA-rich sequence are indicated by dashed lines. Several binding sites for characteristic transcription factors are shown in boxes.

appears to lack a CAAT motif, frequently encountered in eukaryotic promoters. There is a purine-rich motif, GGGTAGGGGGAGG, at −31/−43 in the reverse orientation, which is homologous to a core sequence, GGGGAGGGGGAGG, for the kidney-specific expression of CLC-K1 (Uchida *et al.*, 1998), another member of the chloride channel family. Adjacent to this motif are two identical E boxes followed by a sequence of the SRY binding consensus, AAACWAM (Giese *et al.*, 1994; Harley and Goodfellow, 1994), at −12. The 1.3-kb 5′ flanking sequence farther upstream contains a number of elements; these include AP1, AP2, AP3, Octa-family, PU.1/Ets-family, SRY family, GATA family, HNF-1, HNF-5, and E box as well as a characteristic A/T-rich sequence at −311/−265 and C/EBP motifs at −150. All, or a combination, of these could potentially be involved in regulation of transcription of this tissue-specific gene.

DISCUSSION

We have isolated and characterized the murine homologue of the human *CLCN5* gene, at both cDNA and genomic levels. A cDNA contig with four overlapping clones provided a complete sequence of 2238 bp for the coding region of mouse *Clcn5*. The predicted amino acid sequence shows 98% homology to that of human CLC-5 and its topology is consistent with the currently proposed model for the voltage-gated chloride channels. Experimental evidence in other members of the CLC family demonstrated that domain D4 is extracellular (Schmidt-Rose and Jentsch, 1997) and that D13 is located within the cytoplasmic tail (Gründer *et al.*, 1992). One of the glycosylation sites at Asn408 between D8 and D9 is highly conserved among the CLC family and has been shown to be glycosylated *in vivo* and therefore is believed to reside outside of the cell membrane (Kieferle *et al.*, 1994).

Isolation of the 5′ RACE clones and mapping of a transcription start site characterized the 5′ end of the *Clcn5* mRNA. The sequence of the 5′ RACE product is almost identical to the human cDNA sequence within exon 2, including the translation start site; but examination of farther upstream sequence revealed a novel exon whose equivalent has not yet been recognized in human. The first exon represented in the human cDNA was revealed to be alternatively spliced and intronic in mouse, even though the genomic sequence retains a significant level of homology. The length of the 5′ RACE product is consistent with the position of a single putative transcription start site mapped by primer extension. This site falls within the minimal promoter region and is in agreement with the computer prediction for the transcription initiator consensus. These results suggest that a single transcription start site is employed for the expression of *Clcn5* in mouse kidney. All the evidence taken together, the 5′ UTR of mouse *Clcn5* mRNA has been characterized to comprise 333 bp from the translation start site, which includes the 286 bp of the first novel exon.

A 9.5-kb transcript of mouse *Clcn5* that is predominantly expressed in kidney, and is also observed to a lesser extent in brain and testis, was detected by Northern blot analysis. Given the relatively small size of the 5′ UTR, the predicted open reading frame of 2.2 kb implies the existence of a large (~7 kb) 3′ UTR in the 9.5-kb transcript. Although we have used human probes that cover ~1 kb of the 3′ UTR to screen a mouse kidney cDNA library, none of the clones obtained extended beyond the material already isolated, nor did any of them contain the equivalent 3′ UTR, suggesting that the 3′ untranslated sequences of human and mouse diverge considerably. The nature of an additional 3-kb transcript detected in mouse kidney is still unclear. This 3-kb mRNA was undetectable in human and rat kidney; however, a *Clcn5* mRNA of this size is expressed at about the same low level as the 9.5-kb transcript in rat testis (Sakamoto *et al.*, 1996). Moreover, Northern analysis of various cultured cell lines has provided evidence for the expression of this second transcript (Steinmeyer *et al.*, 1995), i.e., significant expression of the smaller transcript has been observed in LLC-PK1 (pig kidney), and it has also been found at low levels in HIT-T15 (Syrian hamster pancreatic β-cell). Notably, there is no case in which the 3-kb transcript has been observed in the absence of the 9.5-kb transcript. The mechanism of generation of the 3-kb transcript is unknown and needs to be investigated further. Nevertheless, given that the kidney-specific transcript has a single initiation site, we hypothesize that the sporadic expression of a 3-kb transcript may be due to the variable presence of the 3′ UTR sequence as a result of an alternative polyadenylation or alternative splicing event. The importance of the 3′ UTR in the overall regulation of gene expression has been reported in various genes (reviewed by Jackson, 1993). It is involved in regulation of mRNA stability, determination of intracellular localization of mRNA, and control of translation initiation. Certainly, the extensive 3′ UTR of the 9.5-kb transcript could be responsible for the specific localization of the CLC-5 protein in renal proximal tubular cells, as shown by Günther *et al.* (1998), and also for the stability of transcripts and their turnover.

The genomic organization of the open reading frame is highly conserved in mouse and human and manifests intron–exon boundaries at identical positions in the cDNA sequences. Most of the predicted transmembrane domains correspond to separate exons, a common finding in the genomic organization of CLC-1 (Lorenz *et al.*, 1994). Compared to CLC-1, whose open reading frame consists of 23 exons, CLC-5 is organized into 11 exons with a long exon 8, providing structural diversity in the CLC gene family. Whereas the first untranslated exon of *Clcn5* is separated from exon 2 by an intron of 2.2 kb, exon 1 in human cDNA is only 136 bp upstream of exon 2 (Fig. 1b). The existence of dif-

ferent splice variants in the untranslated exons in mouse and human could be a part of the mechanism for transcriptional control of the gene. In preliminary studies, we have investigated the equivalent sequence from the human 5′ flanking region, which revealed no enhancer activity compared to that observed in mouse (manuscript in preparation). Noncoding exons and adjacent introns often have functions, as shown for the *mdr-1b* gene, which contains downstream transcriptional signals in the 5′ UTR (Cohen *et al.*, 1991), or as demonstrated for the *CFTR* gene, which harbors a regulatory element in the first intron (Smith *et al.*, 1996). Our identification of a powerful transcriptional control element within the first intron of mouse *Clcn5* provides the basis for further speculation that the untranslated exon 1 and its adjacent intronic sequence may have significance in the regulation and distinguish some aspects of this regulation in the two species.

Deletion analysis of the ~1.3-kb upstream sequence adjacent to the untranslated exon 1 and mapping of the unique transcription start site at the +1 position delineated the minimal promoter within −128 and −1. This was confirmed by a construct containing −128 to +142 that showed 5- to 10-fold more of the activity relative to the pXP2 in the orientation-dependent manner (data not shown). Further deletion including a putative TATA motif (−69/−64) significantly reduced, but did not obliterate, the promoter activity, implying the functional importance of this element. The Inr found at the putative transcription start site might be sufficient for determining the start location in this promoter, but the activity could be enhanced when the upstream sequence containing the TATA motif is included. Several promoters of other kidney specific genes have canonical TATA box elements (Uchida *et al.*, 1994; Igarashi *et al.*, 1996; Nicholas *et al.*, 1998). On the other hand, the recently identified promoter of rat CLC-K1, another member of the voltage-gated chloride channel family expressed in kidney, lacks both TATA element and CAAT elements (Uchida *et al.*, 1998). In this promoter, a purine-rich sequence was found at about −30 and was shown to be essential for binding of a protein that enhanced the promoter activity in a kidney-specific manner. Scanning homology in the ~500 bp of promoter sequences between mouse *Clcn5* and rat CLC-K1 showed no identity apart from this sequence. The location of the homologous element in mouse *Clcn5* is −31 to −43, which is identical to that of the rat CLC-K1. The sequence overlapped with the predicted binding site for the transcription factor MZF1, which was shown to bind to G-rich core sequences (Morris *et al.*, 1994). Purine-rich sequences are one of the characteristics of promoters for certain genes expressed in kidney (Uchida *et al.*, 1994; Igarashi *et al.*, 1996). Identification of similar elements in the specific promoters of two closely related members of the CLC gene family, both of which are expressed predominantly in kidney, provides evidence for a possible common control mechanism.

The expression of CLC-5 in testis is observed in rat and mouse adult tissues. The testis-determining gene, SRY (Gubbay *et al.*, 1990), is expressed in a wide range of tissues, including kidney, and the SRY-box-related gene SOX9 is expressed in both sexes in developing collecting ducts of the metanephric kidney (Kent *et al.*, 1996), suggesting that the SRY and SRY-like factors may have a wider role in the development of the genitourinary system. Given the colocalization of *Sry* and *Clcn5* transcripts in testis and kidney, it would be of interest to determine if SRY, or SRY-like, HMG transcription factors have a role in regulating the expression of *Clcn5* through the DNA motif predicted in the promoter region.

Incremental deletions of the promoter from the *Hin*dIII site at −1319 to −128 manifested a progressive decrease in its activity, which may be due to other *cis*-acting elements predicted in the upstream sequence. These include sites for the transcription factors that regulate kidney-specific expression of several genes. HNF1b is a tissue-specific transcription factor highly expressed in kidney cells (De Simone *et al.*, 1991) and shown to be required in cell-specific cAMP induction of urokinase-type plasminogen activator gene in kidney cells (Marksitzer *et al.*, 1995). HNF1 was reported to be essential for the kidney-specific expression of aromatic L-amino acid decarboxylase (Aguanno *et al.*, 1996), and phosphoenolpyruvate carboxykinase in the renal proximal tubule was shown to be controlled by HNF1 in conjunction with low levels of C/EBP (Beale *et al.*, 1992; Cassuto *et al.*, 1997). Strikingly, the HNF1-mutant mice manifested hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome caused by renal proximal tubular dysfunction (Pontoglio *et al.*, 1996). The existence of two putative binding sites of HNF1 at −940 and at −1330 in the 5′ control region of mouse *Clcn5* suggests that this tissue-restricted transcription factor might play a significant role for the regulation of the chloride channel, expressed in the renal proximal tubule.

A number of other DNA motifs were predicted within the newly identified enhancer region of 874 bp in the first intron (data not shown). Among them, the existence of the SV40 enhancer core sequence TGGAAAG provided us an intriguing feature of this enhancer. This consensus was first identified in the eponymous and other eukaryotic viral enhancers (Weiher *et al.*, 1983) and was later found in enhancer regions of mammalian genes (Walker *et al.*, 1983; Prochownik, 1985; Gimble *et al.*, 1987). DNA motifs including CCAAT boxes, GATA elements, palindromic sequences, and small tandem repeats are also present around this core sequence. Further functional analysis using gel shift assays and mutagenesis experiments together with an investigation of the tissue specificity will be our next step to ascertain and to characterize the role of these

putative regulatory elements found in both promoter and enhancer regions.

In summary, our cloning and analysis of the mouse version of the gene encoding the chloride-channel protein, CLC-5, revealed its significant conservation with the homologous locus in human and other mammals. This detailed characterization provides the basis for subsequent manipulation of the gene and the construction of a mouse model for human kidney disorders. Our identification of a single putative transcription start site leads to the prediction of the existence of long 3′ UTR sequences in the mRNA, which may play an important role in its stability and turnover. In addition, an investigation of the regulation of the locus achieved by functional analysis of the 5′ UTR and upstream regions has provided evidence for promoter and enhancer elements and the basis for an in-depth examination of the protein/DNA interactions necessary for the kidney-specific gene expression.

ACKNOWLEDGMENTS

We thank Dr. Zheng-Yi Chen for providing the mouse genomic library and Dr. Elena Klenova for critically reading the manuscript and helpful suggestions. This work is supported by an Okinaga Scholarship from Teikyo University, Japan, and the UK National Kidney Research Fund.

REFERENCES

Aguanno, A., Afar, R., and Albert, V. R. (1996). Tissue-specific expression of the nonneuronal promoter of the aromatic L-amino acid decarboxylase gene is regulated by hepatocyte nuclear factor 1. *J. Biol. Chem.* **271:** 4528–4538.

Beale, E. G., Clouthier, D. E., and Hammer, R. E. (1992). Cell-specific expression of cytosolic phosphoenolpyruvate carboxykinase in transgenic mice. *FASEB J.* **6:** 3330–3337.

Bucher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *J. Mol. Biol.* **212:** 563–578.

Cassuto, H., Olswang, Y., Livoff, A. F., Nechushtan, H., Hanson, R. W., and Reshef, L. (1997). Involvement of HNF-1 in the regulation of phosphoenolpyruvate carboxykinase gene expression in the kidney. *FEBS Lett.* **412:** 597–602.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162:** 156–159.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81:** 1991–1995.

Cohen, D., Piekarz, R. L., Hsu, S. I., DePinho, R. A., Carrasco, N., and Horwitz, S. B. (1991). Structural and functional analysis of the mouse mdr1b gene promoter. *J. Biol. Chem.* **266:** 2239–2244.

De Simone, V., De Magistris, L., Lazzaro, D., Gerstner, J., Monaci, P., Nicosia, A., and Cortese, R. (1991). LFB3, a heterodimer-forming homeoprotein of the LFB1 family, is expressed in specialized epithelia. *EMBO J.* **10:** 1435–1443.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12:** 387–395.

Feinberg, A. P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–13.

Fisher, S. E., Black, G. C., Lloyd, S. E., Hatchwell, E., Wrong, O., Thakker, R. V., and Craig, I. W. (1994). Isolation and partial characterization of a chloride channel gene which is expressed in kidney and is a candidate for Dent's disease (an X-linked hereditary nephrolithiasis). *Hum. Mol. Genet.* **3:** 2053–2059.

Fisher, S. E., van Buckel, I., Lloyd, S. E., Pearce, S. H., Thakker, R. V., and Craig, I. W. (1995). Cloning and characterization of CLCN5, the human kidney chloride channel gene implicated in Dent disease (an X-linked hereditary nephrolithiasis). *Genomics* **29:** 598–606.

George, A. J., Crackower, M. A., Abdalla, J. A., Hudson, A. J., and Ebers, G. C. (1993). Molecular basis of Thomsen's disease (autosomal dominant myotonia congenita). *Nat. Genet.* **3:** 305–310.

Giese, K., Pagel, J., and Grosschedl, R. (1994). Distinct DNA-binding properties of the high mobility group domain of murine and human SRY sex-determining factors. *Proc. Natl. Acad. Sci. USA* **91:** 3368–3372.

Gimble, J. M., Levens, D., and Max, E. E. (1987). B-cell nuclear proteins binding in vitro to the human immunoglobulin kappa enhancer: Localization by exonuclease protection. *Mol. Cell. Biol.* **7:** 1815–1822.

Graham, F. L., and van der Eb, A. J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52:** 456–467.

Gründer, S., Thiemann, A., Pusch, M., and Jentsch, T. J. (1992). Regions involved in the opening of CIC-2 chloride channel by voltage and cell volume. *Nature* **360:** 759–762.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., and Lovell, B. R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* **346:** 245–250.

Günther, W., Luchow, A., Cluzeaud, F., Vandewalle, A., and Jentsch, T. J. (1998). ClC-5, the chloride channel mutated in Dent's disease, colocalizes with the proton pump in endocytotically active kidney cells. *Proc. Natl. Acad. Sci. USA* **95:** 8075–8080.

Harley, V. R., and Goodfellow, P. N. (1994). The biochemical role of SRY in sex determination. *Mol. Reprod. Dev.* **39:** 184–193.

Igarashi, P., Whyte, D. A., Li, K., and Nagami, G. T. (1996). Cloning and kidney cell-specific activity of the promoter of the murine renal Na–K–C1 cotransporter gene. *J. Biol. Chem.* **271:** 9666–9674.

Jackson, R. J. (1993). Cytoplasmic regulation of mRNA function: The importance of the 3′ untranslated region. *Cell* **74:** 9–14.

Javahery, R., Khachi, A., Lo, K., Zenzie, G. B., and Smale, S. T. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Mol. Cell. Biol.* **14:** 116–127.

Jentsch, T. J., and Günther, W. (1997). Chloride channels: An emerging molecular picture. *BioEssays* **19:** 117–126.

Jentsch, T. J., Günther, W., Pusch, M., and Schwappach, B. (1995). Properties of voltage-gated chloride channels of the ClC gene family. *J. Physiol. London* **482:** 19S–25S.

Jentsch, T. J., Steinmeyer, K., and Schwarz, G. (1990). Primary structure of *Torpedo marmorata* chloride channel isolated by expression cloning in *Xenopus* oocytes. *Nature* **348:** 510–514.

Kawasaki, M., Uchida, S., Monkawa, T., Miyawaki, A., Mikoshiba, K., Marumo, F., and Sasaki, S. (1994). Cloning and expression of a protein kinase C-regulated chloride channel abundantly expressed in rat brain neuronal cells. *Neuron* **12:** 597–604.

Kent, J., Wheatley, S. C., Andrews, J. E., Sinclair, A. H., and Koopman, P. (1996). A male-specific role for SOX9 in vertebrate sex determination. *Development* **122:** 2813–2822.

Kieferle, S., Fong, P., Bens, M., Vandewalle, A., and Jentsch, T. J. (1994). Two highly homologous members of the ClC chloride channel family in both rat and human kidney. *Proc. Natl. Acad. Sci. USA* **91:** 6943–6947.

Koch, M. C., Steinmeyer, K., Lorenz, C., Ricker, K., Wolf, F., Otto, M., Zoll, B., Lehmann, H. F., Grzeschik, K. H., and Jentsch, T. J. (1992). The skeletal muscle chloride channel in dominant and recessive human myotonia. *Science* **257:** 797–800.

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157:** 105–132.

Larsen, N. I., Engelbrecht, J., and Brunak, S. (1995). Analysis of eukaryotic promoter sequences reveals a systematically occurring CT-signal. *Nucleic Acids Res.* **23:** 1223–1230.

Lloyd, S. E., Pearce, S. H., Fisher, S. E., Steinmeyer, K., Schwappach, B., Scheinman, S. J., Harding, B., Bolino, A., Devoto, M., Goodyer, P., Rigden, S. P., Wrong, O., Jentsch, T. J., Craig, I. W., and Thakker, R. V. (1996). A common molecular basis for three inherited kidney stone diseases. *Nature* **379:** 445–449.

Lloyd, S. E., Pearce, S. H., Günther, W., Kawaguchi, H., Igarashi, T., Jentsch, T. J., and Thakker, R. V. (1997). Idiopathic low molecular weight proteinuria associated with hypercalciuric nephrocalcinosis in Japanese children is due to mutations of the renal chloride channel (CLCN5). *J. Clin. Invest.* **99:** 967–974.

Lorenz, C., Meyer, K. C., Steinmeyer, K., Koch, M. C., and Jentsch, T. J. (1994). Genomic organization of the human muscle chloride channel CIC-1 and analysis of novel mutations leading to Becker-type myotonia. *Hum. Mol. Genet.* **3:** 941–946.

Marksitzer, R., Stief, A., Menoud, P. A., and Nagamine, Y. (1995). Role of LFB3 in cell-specific cAMP induction of the urokinase-type plasminogen activator gene. *J. Biol. Chem.* **270:** 21833–21838.

Morris, J. F., Hromas, R., and Rauscher, F. R., III (1994). Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: Two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Mol. Cell. Biol.* **14:** 1786–1795.

Nicholas, S. B., Yang, W., Lee, S. L., Zhu, H., Philipson, K. D., and Lytton, J. (1998). Alternative promoters and cardiac muscle cell-specific expression of the Na^+/Ca^{2+} exchanger gene. *Am. J. Physiol.* **274:** H217–H232.

Nordeen, S. K. (1988). Luciferase reporter gene vectors for analysis of promoters and enhancers. *BioTechniques* **6:** 454–458.

Pontoglio, M., Barra, J., Hadchouel, M., Doyen, A., Kress, C., Bach, J. P., Babinet, C., and Yaniv, M. (1996). Hepatocyte nuclear factor 1 inactivation results in hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome. *Cell* **84:** 575–585.

Prestridge, D. S. (1991). SIGNAL SCAN: A computer program that scans DNA sequences for eukaryotic transcriptional elements. *Comput. Appl. Biosci.* **7:** 203–206.

Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A., and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238:** 336–341.

Prochownik, E. V. (1985). Relationship between an enhancer element in the human antithrombin III gene and an immunoglobulin light-chain gene enhancer. *Nature* **316:** 845–848.

Rugarli, E. I., Adler, D. A., Borsani, G., Tsuchiya, K., Franco, B., Hauge, X., Disteche, C., Chapman, V., and Ballabio, A. (1995). Different chromosomal localization of the Clcn4 gene in *Mus spretus* and C57BL/6J mice. *Nat. Genet.* **10:** 466–471.

Sakamoto, H., Kawasaki, M., Uchida, S., Sasaki, S., and Marumo, F. (1996). Identification of a new outwardly rectifying Cl^- channel that belongs to a subfamily of the ClC Cl^- channels. *J. Biol. Chem.* **271:** 10210–10216.

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.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

Schmidt-Rose, T., and Jentsch, T. J. (1997). Transmembrane topology of a CLC chloride channel. *Proc. Natl. Acad. Sci. USA* **94:** 7633–7638.

Schug, J., and Overton, G. C. (1997). TESS: Transcription element search software on the WWW. Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania, Philadelphia. [Technical Rep. CBIL-TR-1997-1001-v0.0]

Senapathy, P., Shapiro, M. B., and Harris, N. L. (1990). Splice junctions, branch point sites, and exons: Sequence statistics, identification, and applications to Genome Project. *Methods Enzymol.* **183:** 252–278.

Simon, D. B., Bindra, R. S., Mansfield, T. A., Nelson, W. C., Mendonca, E., Stone, R., Schurman, S., Nayir, A., Alpay, H., Bakkaloglu, A., Rodriguez, S. J., Morales, J. M., Sanjad, S. A., Taylor, C. M., Pilz, D., Brem, A., Trachtman, H., Griswold, W., Richard, G. A., John, E., and Lifton, R. P. (1997). Mutations in the chloride channel gene, CLCNKB, cause Bartter's syndrome type III. *Nat. Genet.* **17:** 171–178.

Smith, A. N., Barth, M. L., McDowell, T. L., Moulin, D. S., Nuthall, H. N., Hollingsworth, M. A., and Harris, A. (1996). A regulatory element in intron 1 of the cystic fibrosis transmembrane conductance regulator gene. *J. Biol. Chem.* **271:** 9947–9954.

Steinmeyer, K., Schwappach, B., Bens, M., Vandewalle, A., and Jentsch, T. J. (1995). Cloning and functional expression of rat CLC-5, a chloride channel related to kidney disease. *J. Biol. Chem.* **270:** 31172–31177.

Uchida, S., Sasaki, S., Fushimi, K., and Marumo, F. (1994). Isolation of human aquaporin-CD gene. *J. Biol. Chem.* **269:** 23451–23455.

Uchida, S., Rai, T., Yatsushige, H., Matsumura, Y., Kawasaki, M., Sasaki, S., and Marumo, F. (1998). Isolation and characterization of kidney-specific ClC-K1 chloride channel gene promoter. *Am. J. Physiol.* **274:** F602–F610.

Walker, M. D., Edlund, T., Boulet, A. M., and Rutter, W. J. (1983). Cell-specific expression controlled by the 5′-flanking region of insulin and chymotrypsin genes. *Nature* **306:** 557–561.

Weiher, H., Konig, M., and Gruss, P. (1983). Multiple point mutations affecting the simian virus 40 enhancer. *Science* **219:** 626–631.