

Cloning and Characterization of CLCN5, the Human Kidney Chloride Channel Gene Implicated in Dent Disease (an X-Linked Hereditary Nephrolithiasis)

SIMON E. FISHER,* INGE VAN BAKEL,* SARAH E. LLOYD,† SIMON H. S. PEARCE,†
RAJESH V. THAKKER,† AND IAN W. CRAIG*¹

*Genetics Laboratory, Biochemistry Department, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom; and

†MRC Molecular Endocrinology Group, Royal Post Graduate Medical School, Du Cane Road, London W12 0NN, United Kingdom

Received June 7, 1995; accepted July 26, 1995

Dent disease, an X-linked familial renal tubular disorder, is a form of Fanconi syndrome associated with proteinuria, hypercalciuria, nephrocalcinosis, kidney stones, and eventual renal failure. We have previously used positional cloning to identify the 3' part of a novel kidney-specific gene (initially termed hCIC-K2, but now referred to as CLCN5), which is deleted in patients from one pedigree segregating Dent disease. Mutations that disrupt this gene have been identified in other patients with this disorder. Here we describe the isolation and characterization of the complete open reading frame of the human CLCN5 gene, which is predicted to encode a protein of 746 amino acids, with significant homology to all known members of the CIC family of voltage-gated chloride channels. CLCN5 belongs to a distinct branch of this family, which also includes the recently identified genes CLCN3 and CLCN4. We have shown that the coding region of CLCN5 is organized into 12 exons, spanning 25–30 kb of genomic DNA, and have determined the sequence of each exon–intron boundary. The elucidation of the coding sequence and exon–intron organization of CLCN5 will both expedite the evaluation of structure/function relationships of these ion channels and facilitate the screening of other patients with renal tubular dysfunction for mutations at this locus. © 1995 Academic Press, Inc.

INTRODUCTION

Dent disease is a renal tubular disorder characterized by low-molecular-weight proteinuria, hypercalciuria, nephrocalcinosis, nephrolithiasis (kidney stones), and progressive renal failure, which has been described in eight British families (Wrong *et al.*, 1994). Linkage analysis has localized the gene responsible to the short

arm of the X chromosome, and a microdeletion involving the hypervariable locus DXS255 in Xp11.22 was found to be associated with Dent disease in one pedigree (Pook *et al.*, 1993). We have previously used positional cloning to isolate cDNAs from the 3' end of a novel kidney-specific gene mapping within this deletion, in a region 40–80 kb proximal to DXS255 (Fisher *et al.*, 1994). Homology studies indicated that this gene (initially termed hCIC-K2, but now referred to as CLCN5 on the advice of the Genome Database Nomenclature committee) encoded a new member of a family of voltage-gated chloride channel (CIC) proteins (Jentsch *et al.*, 1995) and we proposed it as a strong candidate for Dent disease, on the basis of its expression pattern, putative function, and deletion mapping (Fisher *et al.*, 1994). We have since identified additional deletions and point mutations in CLCN5 in patients from other pedigrees segregating Dent disease (Lloyd *et al.*, submitted).

The CIC chloride channel genes are members of a family, originally defined by the CIC-0 locus, that was isolated from *Torpedo* electric organ by expression cloning (Jentsch *et al.*, 1990). At least six members of this family, in addition to CLCN5, are now known in mammals (Jentsch *et al.*, 1995). Some of these are expressed in a tissue-specific manner such as CLCN1 (the major skeletal muscle chloride channel) (Steinmeyer *et al.*, 1991) and CLCNKa and CLCNKb (both specific to the kidney) (Kieferle *et al.*, 1994) while others, like the swelling-activated CLCN2 gene (Thiemann *et al.*, 1992), are ubiquitously expressed. Initial studies of the 3' end of CLCN5 have suggested that it is a member of the most distant branch of this family, which also includes two other recently identified genes, CLCN3 (Kawasaki *et al.*, 1994) and CLCN4 (van Slegtenhorst *et al.*, 1994).

We have constructed a cDNA contig spanning the entire CLCN5 open reading frame, which is predicted to encode a chloride channel protein of 746 amino acids

¹ To whom correspondence should be addressed. Telephone: 01865 275 327. Fax: 01865 275 318.

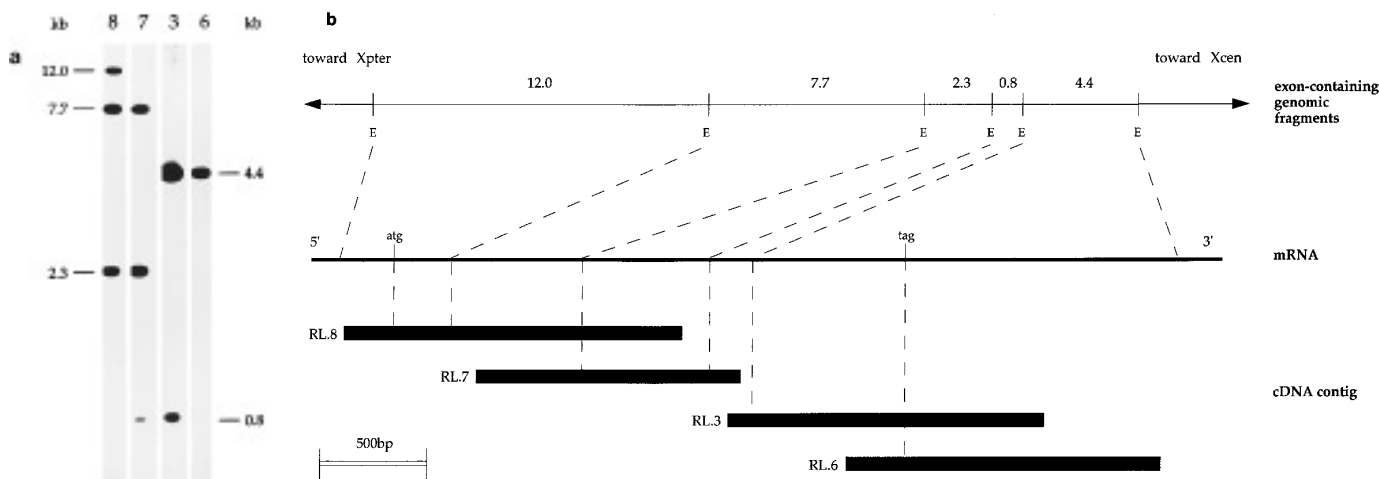


FIG. 1. cDNA contig spanning the CLCN5 coding region. **(a)** Hybridization of inserts from clones RL.8, RL.7, RL.3, and RL.6 to *EcoRI* digests of YAC yWXD6129. Sizes of fragments are indicated. **(b)** Relative overlaps of cDNA clones (bottom) and correlation with genomic *EcoRI* fragments (top) as deduced from hybridization analysis. The scale shown at the bottom applies only to the transcript. The initiator methionine of the ORF is represented by "atg" and the stop codon by "tag." No sites for *EcoRI* are contained in the cDNA contig; all map within introns. Dashed lines indicate how each cDNA relates to the genomic fragments to which it hybridizes. Only exon-containing genomic fragments are shown, and so each E may in fact represent more than one *EcoRI* site. Fragment sizes are given in kilobases. Orientation with respect to the X chromosome is indicated.

with a very similar hydropathy plot to those of other CIC genes. We have also determined all exon-intron boundaries and flanking intron sequences of the CLCN5 coding region, and this information is of significance for structural and functional comparisons between the CIC family members. In addition, it will greatly aid the search for additional mutations in this gene that may be associated with further cases of nephrolithiasis and other hypercalciuric states.

MATERIALS AND METHODS

General procedures. Conventional gel electrophoresis was in 0.8% (w/v) agarose (Sigma Type I-A)/1× TBE gels at 1.5 V cm⁻¹. Gels were transferred to Hybond-N⁺ (Amersham) by alkaline blotting, and filters neutralized in 0.2 M Tris-Cl (pH 7.5), 2× SSC. Plasmids were prepared by standard methods (Sambrook *et al.*, 1989), and inserts purified by the GeneClean (Bio 101) procedure. Agarose plugs containing yWXD6129 YAC DNA were prepared by standard methods (Anand *et al.*, 1990). Labeling was by random priming with [α -³²P]dCTP (Feinberg and Vogelstein, 1983). Hybridization was at 65°C in Church buffer (Church and Gilbert, 1984), and washing was to a final stringency of 40 mM sodium phosphate, pH 7.2/0.1% SDS, at the same temperature. Filters were exposed to Kodak XAR film with Kodak intensifying screens at -70°C.

cDNA walking. A cDNA library from adult renal tissue (Clontech) was used. Recombinants (4×10^5) were plated on four 22 × 22 cm plates, and two replica plaque lifts on Hybond-N⁺ (Amersham) were prepared from each plate according to the manufacturer's instructions. The RL.3 cDNA insert (which corresponds to the 3' end of the CLCN5 gene) was used to screen library filters by hybridization. Phage DNA was isolated using Wizard Lambda Minipreps (Promega). Phage inserts were subcloned into pUC9 using standard techniques. Following characterization, nonchimeric cDNA inserts were used to rescreen the library, until a contig spanning the entire coding region of CLCN5 was assembled.

DNA sequencing from plasmids. Plasmids were sequenced on both strands by the dideoxy method (Sanger *et al.*, 1977) with Sequenase version 2.0 (USB) using universal and reverse primers as well

as primers derived from the obtained sequence. dITP was used to resolve compressions (USB).

DNA sequencing from cosmids. Cosmid UU12 12F16 was isolated from a flow-sorted X chromosome gridded library by hybridization of probe RL.3 to a complete set of library filters (Bentley *et al.*, 1992). Cosmid DNA was prepared using standard techniques. Exon-intron boundaries for introns III, IV, and VI were determined by direct sequencing with primers designed from CLCN5 exonic sequence using a Gibco-BRL cycle sequencing kit according to the manufacturer's instructions.

Analysis of exon-intron structure using PCR. PCR primers were selected from different regions of transcript and used to amplify from cDNA, genomic, and yWXD6129 templates as well as water controls. Primer details are as follows (5' starting position is given with respect to nucleotide sequence in Fig. 2; "reverse" primers are complementary to coding strand): intron I, forward (22-mer, starting at nucleotide 145), reverse (20-mer, 311); intron II, forward (20-mer, 291), reverse (19-mer, 470); intron V, forward (19-mer, 735), reverse (22-mer, 990); intron VII, forward (22-mer, 1037), reverse (20-mer, 1192); intron VIII, forward (18-mer, 1525), reverse (21-mer, 1776); intron IX, forward (22-mer, 1696), reverse (22-mer, 1859); intron X, forward (20-mer, 2050), reverse (20-mer, 2330); intron XI, forward (22-mer, 2224), reverse (20-mer, 2561). Conditions were as follows for all PCRs; 200 μ M dNTPs (Amersham); 10 mM Tris-HCl; 50 mM KCl; 1.5 mM MgCl₂; 1.0 μ M each primer; *Taq* polymerase (Promega). Cycling parameters were 94°C, 5 min ("hot start"); 94°C, 30 s; 55°C, 30 s; 74°C, 48 s; 30 cycles. Products containing introns (see text) were cloned using a TA cloning kit (Invitrogen) according to the manufacturer's instructions. Each intron was amplified, cloned, and sequenced in duplicate to verify exon-intron boundary sequence. Introns were further analyzed by digestion with *EcoRI* and hybridization to an *EcoRI* digest of yWXD6129, enabling their position with respect to genomic fragments to be confirmed.

RESULTS

A cDNA Contig Spanning the CLCN5 Open Reading Frame

Two overlapping cDNA clones from the 3' end of the CLCN5 gene were previously isolated by hybridization

1 t g a t g t g a t a t g g c t g c a a g t g c c t t g a c c c t t t g t c t c c t t c c a t a a a c t g a a a t a c c t a a g c t g c t c c a a c c t c c t t t t g t c t
 91 t t g t t t c a t a a a t c c t t t c c a t t g c a c a t c a a c t c t g t c t c t t t g t a c t g t c a c t c t a t c t g t t c c a t t c a a t c a c c t g c c t
 181 t a g c c a c t c a t c a t t t t g t g c t a c a c c a c a g a a a c c t c t g a a t g t a a t g g a t g t t c c t a c c a g a g g a c a a g t c g t a c a a t g g t g g a g a

 271 a t a g g t t c t t c a a a t a g g a t c a t g g a c t t c t t g g a g g a c c a a t c c c t g g t g t a g g g a c c t a t g a t g a t t t c a a t a c a a t t g a t g g g t g

 1 M D F L E E P I P G V G T Y D D F N T I D W V

 361 a g a g a g a a g t c t c g a g a c c g g g a t a g g c a c c g a g a g a t t a c c a a t a a a a g c a a a g a g t c a a c a t g g g c c t t a a t t c a c a g t g t g a t g a t

 24 R E K S R D R D R H R E I T * N K S K E S T W A L I H S V S D

 451 g c t t t t t c c g g c t g g t t g t g a t g c t c c t t a t t g g g c t t t a t c a g g t t c g t t a g c t g g t t g a t a g a c a t c t c t g c t c a t t g g a t g a c a

 54 A F S G W L L L M L L I G L L S G S L A G L I D I S A H W M T

 D1

 541 g a c t t a a a g a a g g t a t g c a c a g g g g a t t c t g t t t a a c c a t g a a c a t t g t t g c t g g a a c t g a g c a t g c a c c t t t t e a a g a g a g a

 84 D L K E G I A T G G F W F W F N H E H E H C W N S E H E V C T F F E E R

 631 g a c a a a t g t c c a g a g t g g a a t a g t t g g t c c c a g c t t a t c a t c a g c a c a g a t g a g g g a c c t t g c c t a c a t a g t c a a t t a t t t c a t g t a c

 114 D K C P E W N S Q L I I S T D E G A F A Y I V N Y F M Y

 721 g t c c t c t g g c t c t c t a t t t g c c t c c t t g c g t a t c t c t t g t c a a g g t t t g c g c c t a t g c c t g t g g c t g g a a t c c c t g a g a t a

 144 V L W A L L L F A F L A V S L V K V F A P Y A C G S G I P E I

 D2

 811 a a a a c t a t c t t g a g t g g t t c a t t a t t a g g g c t a t t t g g t a a g t g g a c t c t g g t t a t c a a a a c c a t c a c c t t g g t g c t g g c a g t g t c g

 174 K T I L S G F I I R G Y L G K W T L V I K T I T L V L A V S

 D3

 901 t c t g g c t t g a g c t g g g c a a a g a g g c c c t a g t g c a c g t g g c t t g c t g t g g g a a c a t c c t g t g c c a c t g c t c a a c a a a t a c a g g

 204 S G L S L G K E G P L V H V A C C C G N I L C H C F N K Y R

 D4

 991 a a g a a t g a a g c c a a g c g a g a g g t c t t g t c g g t g c a g c a g c a g t g g t g a t c t g a g c c t t t g g a c c c a t a g g t g g a g t a t t a

 234 K N E A K R R E V L S A A A A A G V S V A F G A P I G G V L

 D5

 1081 t t c a g c c t t g a a g a g t c a g t a c t a t t t t c c c c t c a a a a c a t t g t g g c t t c a t t c t t t g c t g c c t t g g t g g c a g c a t t c a c t c t a c g c

 264 F S L E E V S Y Y F P L K T L W R S F F A A L V A A F T L R

 D6

 1171 t c c a t c a a t c c a t t t t g g a a c a c g c c t g t a c t a t t t a t g t g g a g t t t c a c a c c c a t g g c a t c t c t t t g a g c t g t g c c a t t c a t t

 294 S I N P F G N S R L V L F Y V E F H T P W H L F E L V P F I

 1261 c t g c t g g g c a t a t t t g t g g t c t g t g g g a g c a c t g t t a t c g c a c a a a c a t t g c c t g g t g t c g g a a g c a a a g a c c a c c a g t t g g g c

 324 L L G I F G G L W G A L F I R T N I A W C R K R K T T Q L G

 D7

 1351 a a g t a t c c t g t t a t a g a g g t a c t c g t c g t g a c a g c c a t c a c t g c c a t c o t g g c t t c c c c a a t g a a t a c a c t o g g a t g a g c a a a g t g a g

 354 K Y P V I E V L V T A I T A I L A F P N E Y T R M (S) T S E

 D8

 1441 c t c a t t t c t g a g c t g t t a a t g a c t g t g g c c t t c t g g a c t c c t c c a a g c t c t g t a t t a g a a c c g t t t c a c a c a a g c a a a g g g g t

 384 L I S E L F N D C G L L D S S K L C D Y E N R F * N T S K G G

 1531 g a a c t g c c t g a c a g a c c g g t g g c g t g g g a g t c t a c a g t g c a a t g t g c a g c t g g c t t a a c a c t a c a t a g a a a t t g t c a t t a c t a t a

 414 E L P D R P A G V G V Y S A M W O L A L T L L I K I V I T I

 D9

 1621 t t c a c c t t g g c a g a a g t c c c t t c t g g c c t t t a t c c c t a g c a t g g c t g t t g g t g c t a t a g c a g g t c g a c t t c t a g g a g t a g g a a t g

 444 F C T F G M K I P S G L F I P S M A V G A I A G R L L G M

 D10

 1711 g a a c a g c t g g c t a t t a c c a c c a g g a a t g g a c c g t c t t c a a t a g c t g g t a g t c a g g g a g c t g a t t g c a t c a c c c c c g c c t t a t g c a

 474 E Q L A Y Y H Q E W T V F N S W C S Q G A D C I T P G L Y A

 1801 a t g g t t g g g c t g c a g c c t g t a g g t g g g t g a c t c g g a t g a c t g t t t c t t g t t g t c a a t a g t t t g a a c t g a c t g g t g c c t a g a a

 504 M V G A A A C L G G V T R M T V S L V V I M F E E L T G G L E

 D11

 1891 t a c a t c g t c c c t c t a t g g c t g c a g c c a t g a c a a g a a g t g g g t g c a g a t g c t c t t g g g c g g a g g c a t c a t a t g a t g c c c a c a t c c g t

 534 Y I V P L M A A A M T S K W V A D A L G R E G I Y D A H I R

 D12

 1981 c t c a a t g g a t a c c c c t t c t t g a a g c a a a g a a g t t t g c t a t a a g a c c c t g g c a a t g g a t g t a g a a c c c g g a g a a t g a t c c t

 564 L N G Y P F L E A K E E F A H K T L A M D V M K P R R N D P

 2071 t t g t g a c t g t c c t t a c t c a g g a c a g t a g a c t g t g g a a g a t g t a g a g a c c a a t c a g t g a a c c a c t t a c a g t g c c t c c c a g t g g t g

 594 L L T V L T Q D S M T V E D V E T I I S E T T Y S G F P V V

 2161 g t a t c c c g g a g t c c a a a g a c t t g t g g c t t t g c c t c o g a a g a g a t c t c a t t a t t t c a a t t g a a a t g c t o g a a a g a a c a g a t g g g

 624 V S R E S Q R L V G F V L R R D L I I S I E N A R K K Q D G

 2251 g t t g t a g a c c t c c a t c a t t a t t t t c a c g g a g c a t t c t c c a t t g c c a c a t a c a c t c c a c c a c t c t a a a g c t t o g g a a c a t c c t c

 654 V V S T S I I Y F T E H S P P L P P Y T P P T L K L R N I L

 2341 g a t c t a g c c c c t c a c t g t g a c t g a c c t t a c a c c a t g g a g a t c g t a g t g g a t a t t t c c g a a a g c t g g a c t g c g g c a g t g c c t g g t t

 684 D L S P F T V T D L T P M E I V V D I F R K L G L R Q C L V

 2431 a c a c a a c g g g c g a t t g c t t g g a a t c a t t a c a a a a a g a t g t g t t a a g c a t a t a g c a c a g a t g g c g a a c c a a g a t c c t g a t t c c a t t

 714 T H N G R L L G I I T K K D V L K H I A Q M A N Q D P D S I

 2521 c t c t c a a c t a g a a t c a t a g a g t t c t g g a t g t a a a g c g g a a g g a c a t t a c a g a c c a t g g a t a t g t t g t t a a c c g t a c c a a a c a c a t

 744 L F N @

 2611 t t t c a t a t t g g a t g g t g a a g t c a c a t t a g t g t t g t c t c t t c t a c a a g t t a a c c a g t t g c a c t a c a t a a t c t c t g g a a t t a a t t

 2701 t t c t t t a g g a g a a a t a t a g t t a g g c t t c a t g a t g t a c a t t a g g a a g a t a t c a t g a a a g a a a a a t a a g a t t g c t a t g g t t a a t t

 2791 a t a t t g t c t t t t a a a g a t t t t t t a a c t t a a a a g t a g t t a g c c a a t a t g c a a t c a c t g a a a a c t a t g c a a g a g a a a t t c a a c c g t c

 2881 c t g a c c t a a a c c t g a g g a a c c g a c g a a a a a g t c a c t c t t t g g a t c t a a c t g t t g t t a c t g g a a g c g a a g g t a a a c t a a g g g g c

 2971 t t g c t t t c a a a c c a g a a a g g a a a g c a a g a a g a a g a a g t a a g t a t t t c t a g a c t g t a a g a t t c a a t t c a a t g t a t c c t t g

 3061 t t c c t g t t a c a a t a t t a g c a t t a t a g t t t g t a t g t g t a t g t t a t g t t a a t t t t a a t t t c t g a t t a a g a c a a t g c t g c t t g g

 3151 t t a a t c t c t c t a a a g a a t t a

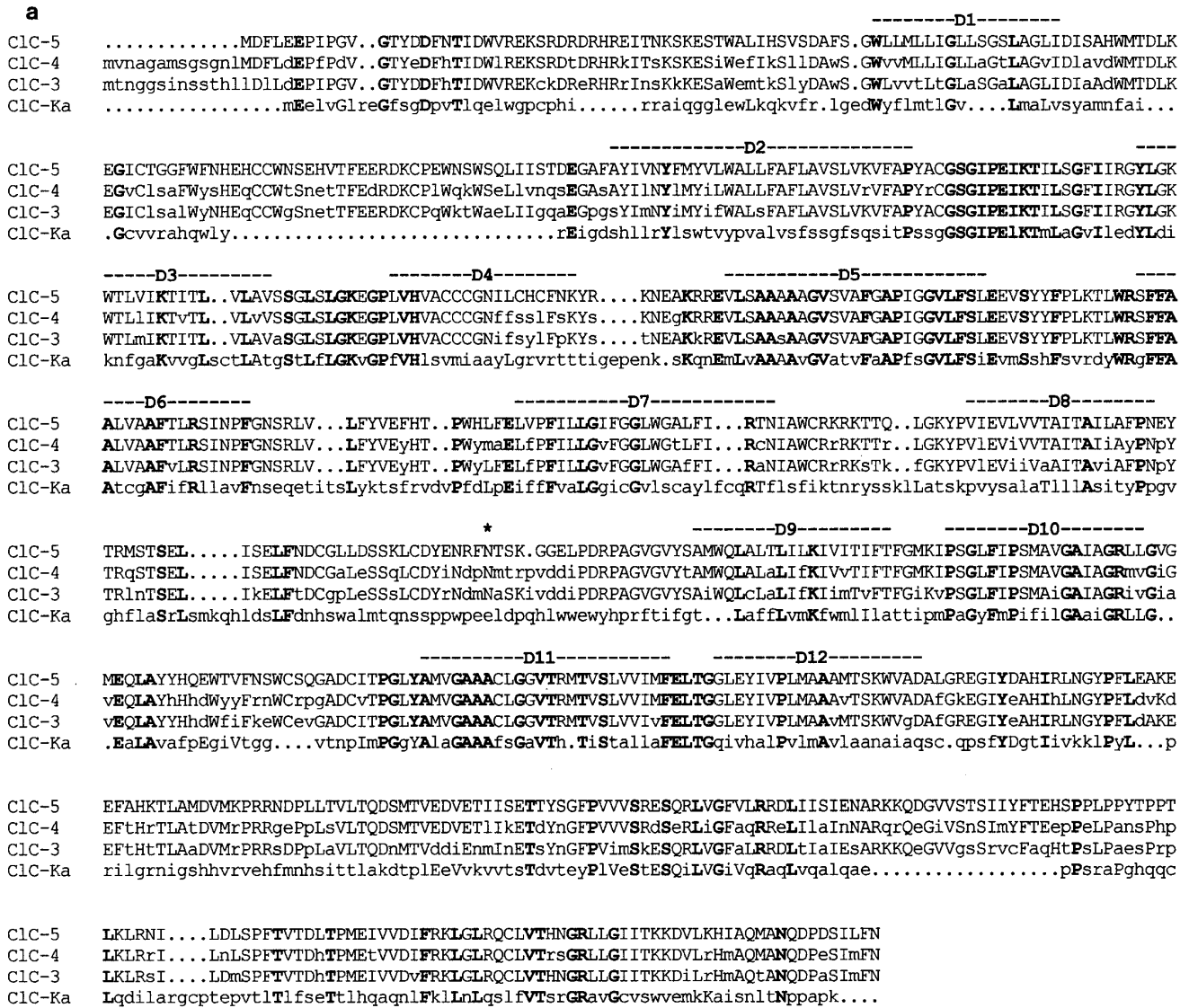


FIG. 3. (a) Comparison of protein sequence encoded by CLCN5 with those of other chloride channel genes. Alignment is shown for CIC-5 (human), its two closest relatives, CIC-4 (human) (van Slegtenhorst *et al.*, 1994) and CIC-3 (rat) (Kawasaki *et al.*, 1994), and one of the other kidney chloride channels, CIC-Ka (human) (Kieferle *et al.*, 1994). Uppercase letters indicate identity with CIC-5. Letters in boldface represent amino acids that are identical in all four peptide chains. The positions of putative hydrophobic domains are shown. The asterisk indicates an N-linked glycosylation site, the position of which is conserved among CIC-5, CIC-4, and CIC-3. This diagram is adapted from an alignment of all known CIC genes obtained using the "Pileup" program of the GCG software package. **(b)** Dendrogram (also obtained using Pileup) showing the relationship between CIC-5 and other members of the CIC family. The chloride channels shown are from human (Cid *et al.*, 1995; Kieferle *et al.*, 1994; Steinmeyer *et al.*, 1991; van Slegtenhorst *et al.*, 1994), except for CIC-3, which has been cloned only in rat (Kawasaki *et al.*, 1994), and CIC-0, the original member of this family, which is from the *Torpedo marmorata* electric organ (Jentsch *et al.*, 1990).

of a YAC containing DXS255 (yW XD6129) to a human cDNA library made from adult kidney (Fisher *et al.*, 1994), and these provided the starting point for this study. The clones, termed RL.3 and RL.6, spanned 1.9

kb of cDNA sequence and included the terminal 780 bp of a putative open reading frame with significant homology to members of the CIC family. Beginning with the rescreening of the kidney library using RL.3,

FIG. 2. Complete nucleotide and predicted amino acid sequence of cDNA clones spanning the coding region of CLCN5. The initiator methionine was assigned to the second ATG downstream of a stop codon in frame (see text). Stop codons are indicated by @. The predicted hydrophobic domains D1 to D12 are underlined. Potential N-linked glycosylation sites are indicated by asterisks. A consensus phosphorylation site for cAMP-dependent protein kinase at Ser-380 is represented by (S). The RL.3 cDNA sequence from nucleotides 1750 to 3173 was previously reported (Fisher *et al.*, 1994).

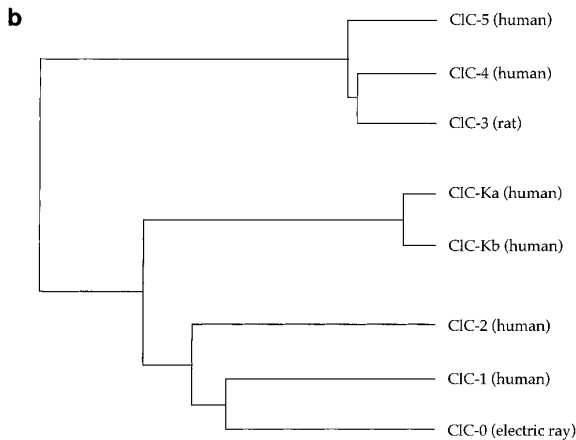


FIG. 3—Continued

several steps of a cDNA walk were required to assemble clones covering the entire coding region of *CLCN5*. It proved impossible to isolate a single cDNA containing the complete ORF from this library, and a significant proportion of clones were found to be chimeric. Figure 1 shows four nonchimeric overlapping clones, RL.8, RL.7, RL.3, and RL.6, which is the minimum number needed to span the 3.7-kb region of the *CLCN5* transcript covered by our cDNA contig. On hybridization to *EcoRI* digests of the yWXD6129 YAC, each clone gave a different pattern of bands (Fig. 1a), and from this we were able to determine the size and order of the exon-containing genomic *EcoRI* fragments (Fig. 1b). Mapping of clones within a rare-cutter restriction map of yWXD6129 indicates that this 3.7-kb region of the *CLCN5* transcript spans a genomic distance of between 25 and 30 kb (data not shown).

Sequence Analysis of *CLCN5*

Sequencing of the new cDNA clones combined with previously obtained sequence data from RL.3 and RL.6 enabled the entire coding sequence of *CLCN5* to be determined (Fig. 2). The open reading frame is predicted to encode a protein of 746 amino acids (CIC-5). The initiation methionine was assigned to the second ATG triplet (nucleotides 292–294) that appears downstream of a stop codon (nucleotides 226–228) in the same reading frame. This start site seems more likely than the first in-frame ATG (at nucleotides 232–234), since its surrounding sequence context fits much more strongly the expectations for a eukaryotic translation initiation site (Kozak, 1987).

The predicted amino acid sequence encoded by the complete ORF shows significant similarity to those of all other known members of the CIC family (Fig. 3). Sequence comparisons indicate that CIC-5 is more closely related to CIC-3 and CIC-4 (to which it shows 77–78% amino acid identity) than to other members of the family (28–32% identity). The hydropathy profile of CIC-5 is very similar to that of other CIC proteins

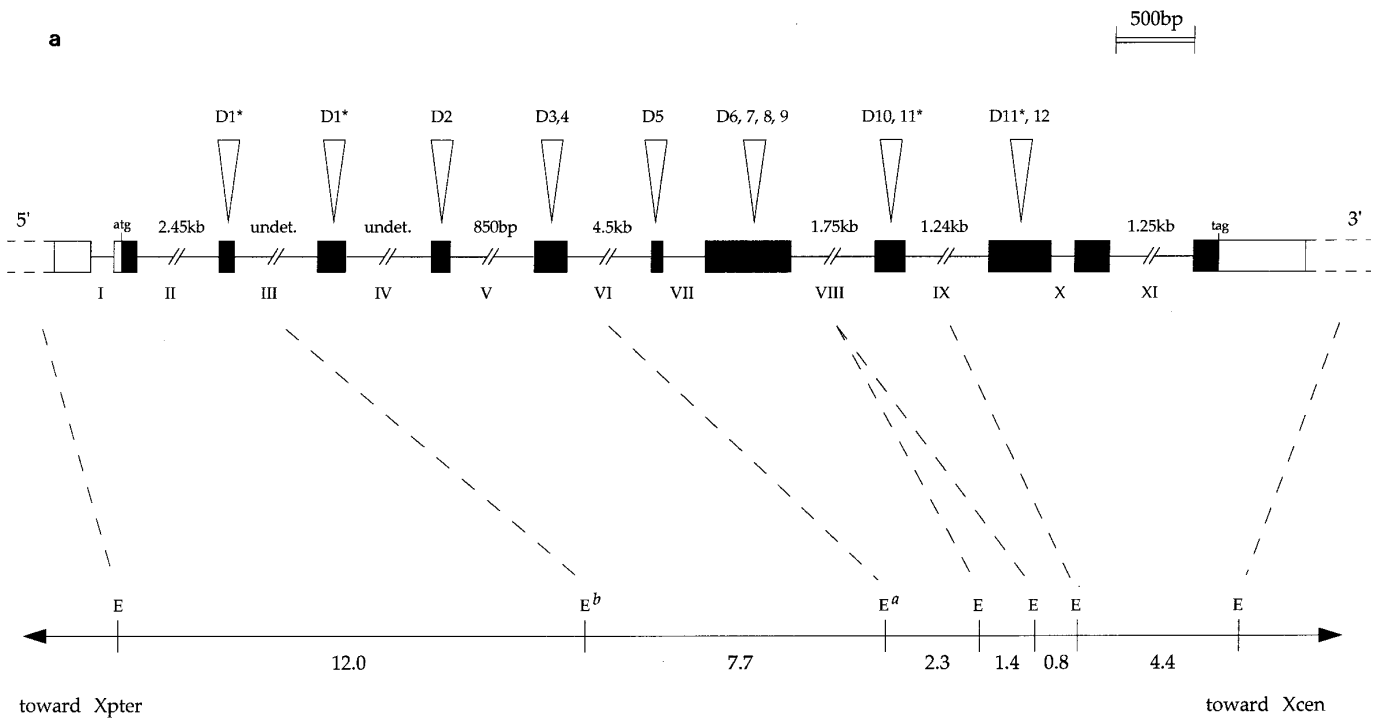
and predicts a topology involving 12 putative transmembrane domains (Fig. 2). Homology between *CLCN5* and other CIC genes is particularly high in these putative membrane spans. We have identified two potential N-linked glycosylation sites, one at position 38, close to the amino terminus, the other at position 408, in the loop between putative domains D8 and D9. There is a consensus phosphorylation site for cAMP-dependent protein kinase (Pearson and Kemp, 1991) at Ser-380, also in the D8–D9 loop (Fig. 2).

We have previously demonstrated that the entire *CLCN5* transcript is ~9.5 kb (Fisher *et al.*, 1994), but have shown here that the region encoding the chloride channel is confined to only ~2.2 kb. Similar observations have been made for the *CLCN4* gene, in which the 2.2-kb open reading frame is found at the 3' end of a 7.5-kb transcript, suggesting the presence of a large untranslated 5' region (van Slegtenhorst *et al.*, 1994). The nature of the untranslated regions of these transcripts, and their possible role in control of CIC expression, has not yet been determined.

Genomic Organization of the *CLCN5* Coding Region

While the region of the *CLCN5* transcript covered by our cDNA contig contains no *EcoRI* sites, hybridization results indicate that it spans five exon-containing *EcoRI* fragments at the genomic level (Fig. 1), suggesting a minimum of five exons for this part of the gene. We adopted a polymerase chain reaction strategy to investigate further the genomic organization of the *CLCN5* coding region. Primers designed from different parts of the transcript were used to amplify from cDNA, human genomic, and yWXD6129 YAC templates. A larger band or a complete absence of product amplified from genomic and YAC templates (as compared to a product of expected size from cDNA template) was taken as evidence for one or more introns in the region bounded by the primers. Products expected to contain introns were cloned and sequenced to determine precise exon–intron boundaries. The difference between product size from genomic template versus cDNA template gave an estimate of the size of the intron.

Using this approach we have established that the coding region of *CLCN5* consists of 12 exons (Fig. 4a) and have determined the exon–intron boundaries of each (Fig. 4b). The acceptor and donor sequences of all exons correspond well to consensus splice site sequences. We were able to estimate the sizes of all introns except III and IV, which were too large to amplify using conventional PCR. (Precise boundaries for introns III, IV, and VI were determined by direct sequencing from a cosmid, as described under Materials and Methods.) Hybridization analysis enabled us to correlate the exon–intron boundary information with the map of genomic *EcoRI* fragments of the gene (Fig. 4a). Intron VIII was found to con-

**b**

Intron	Position	5'-3' sequence at intron/exon boundary
I	244-245	CCTACCAGgtgactgtatttctttatcaacccaactgtagtc.....ctgaagtactaatgtctatttcttgttcaatacagAGGACAAG
II	396-397	ACCGAGAGgtaagacaaaagatggcacatgggtaagtgttagg.....gtagagtgtaccaatgtttctcattttccocctagATTACCAA
III	496-497	TTTATCAGgtatggtaaaactgttagttttcaaaacaaatctcc.....taactttggcctttccctccctcccnnaaatcagGTTTCGTTA
IV	684-685	CAGATGAGgtaacatgtagtgatgttttatgagccattgactt.....taggctaaaacaagcttctttttccocctgttccagGGAGCCTT
V	807-808	TCCCTGAGgtgagtctcttaaaatggtttataaaatggttaca.....agtctatattcaatctttctgtgtttaacctgcagATAAAAAC
VI	1014-1015	GCAGAGAGgtaataatgaatggccttaatagtctcttttggg.....gtcactaattctgagtttggattttgatttttagGTCTTGTC
VII	1095-1096	TTGAAGAGgtaacaacttttcattgtcacagcatgtgcatgctt.....tgactgagtttgctttctcaccctctttctctctagGTCAGCTA
VIII	1638-1639	GCATGAAGgtgaggaattcttttgggactcagtggtgcatgc.....tttctcactaacctctattggtttctctttgcagATCCCTTC
IX	1825-1826	CTGCTTAGgtgagtagtgtttgcattaatttcaagttgctacc.....gcctacctgagtagactgtgtctatttctttgcagGTGGGGTG
X	2224-2225	TTCAATTGgtaagatttcagaaagggatagtggaatccact.....catcttcaattgtttttctctctgtttgaaatagAAAATGCT
XI	2441-2442	CACAACGGgtaagaagcttctgagtggaagtcaaattgaattgtg.....tatttttgttttttattgtgtttgtcttttagGCGATTGT

FIG. 4. Genomic organization of CLCN5. **(a) (Top)** Exon-intron structure. The relative sizes of all exons and introns I, VII, and X are drawn to scale. Sizes of other introns are indicated, except for III and IV, which are undetermined. The initiator methionine is represented by "atg" and the stop codon by "tag." Untranslated regions are indicated by open boxes. The predicted hydrophobic domains contained in each exon are indicated by arrowheads. Asterisks denote domains that are disrupted by introns. **(Bottom)** Correlation between exon-intron boundaries and genomic *EcoRI* fragments, as deduced from PCR studies and hybridization analyses. Each E represents a single *EcoRI* site. Sizes of fragments are given in kilobases. Intron VIII contains a 1.4-kb internal *EcoRI* fragment. Intron VI contains two internal *EcoRI* fragments of 0.7 and 1.75 kb; therefore, E^a represents more than one *EcoRI* site. E^b may also represent more than one *EcoRI* site. Orientation with respect to the X chromosome is indicated. **(b)** Sequences of each exon-intron boundary. Position is given with respect to the nucleotide sequence shown in Fig. 2. Uppercase letters are exonic, lowercase intronic.

tain a 1.4-kb internal *EcoRI* fragment that is not detected by the CLCN5 transcript (Fig. 4a). Intron VI contains two such internal *EcoRI* fragments, of 1.75 and 0.7 kb.

DISCUSSION

We have constructed a cDNA contig spanning the coding region of CLCN5, the kidney-specific chloride

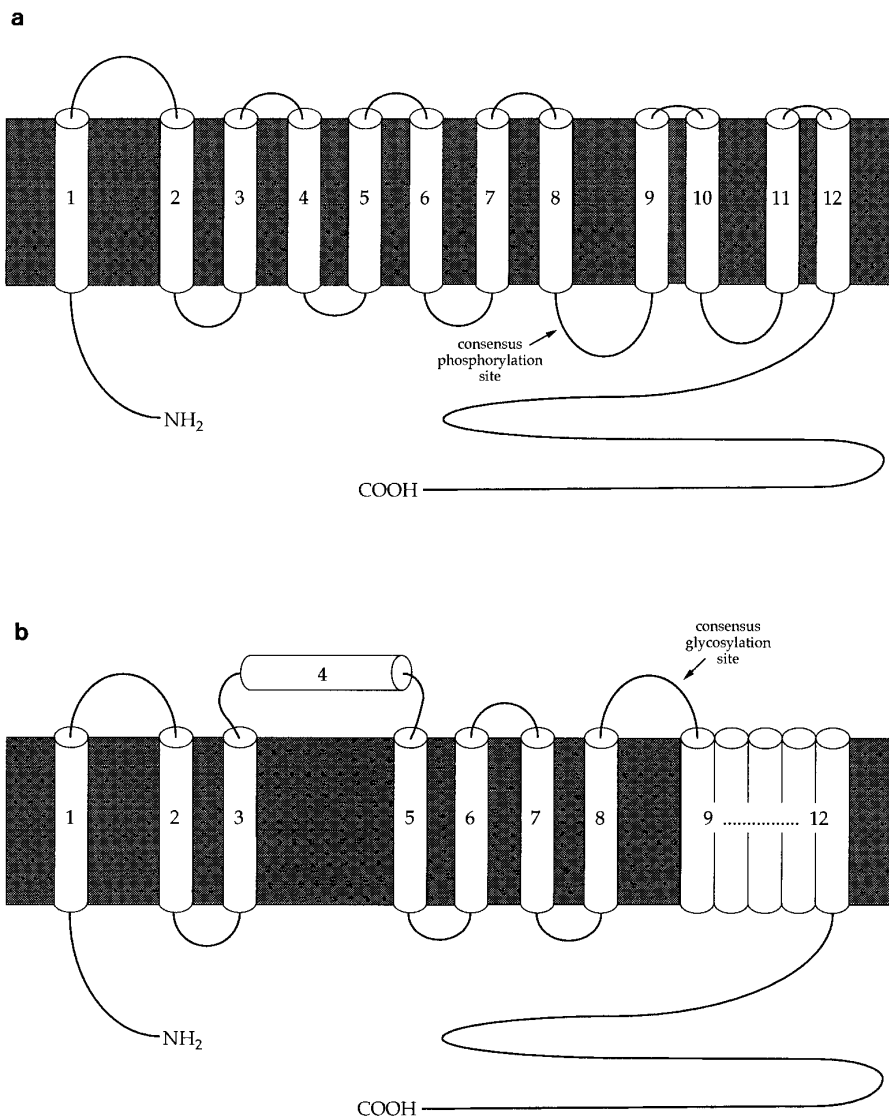


FIG. 5. Two alternative models for transmembrane topology of ClC-5, based on hydropathy plots. **(a)** Conventional representation, involving 12 transmembrane domains and an intracellular C-terminus. Renal tubular chloride channels are regulated by cAMP-dependent protein kinase (Bae and Verkman, 1990; Gesek and Friedman, 1992, 1993) and there is a consensus site for phosphorylation by this enzyme in the D8–D9 loop, which would be intracellular using this model. **(b)** Preliminary revised model, adapted from Jentsch *et al.* (1995), which allows for glycosylation within the D8–D9 loop, while maintaining an intracellular C-terminus. In this representation D4, the least hydrophobic of the conventional domains, does not cross the membrane and the D8–D9 loop becomes extracellular. The precise topology of the D9–D12 region is unclear, but involves either 3 or 5 transmembrane spans. Note that, in this model, the consensus phosphorylation site cannot be used.

channel gene implicated in Dent disease. This is the second member of the ClC family to be associated with a human genetic disorder. Mutations in the skeletal muscle chloride channel gene, CLCN1, are responsible for two forms of human myotonia (George *et al.*, 1993; Koch *et al.*, 1992). Homology studies indicate that CLCN5 belongs to a recently identified distant branch of the voltage-gated chloride channel family which also includes CLCN3, a gene abundantly expressed in brain (Kawasaki *et al.*, 1994), and CLCN4, which was isolated from Xp22.3 (van Slegtenhorst *et al.*, 1994). Two additional members of this family are specific to human kidney, CLCNKa and CLCNKb (along with their rat

homologues rClC-K1 and rClC-K2) (Kieferle *et al.*, 1994), but while these are very similar to each other (over 90% amino acid identity), they have only 28% identity with the ClC-5 protein. It has been suggested that rClC-K1 is involved in urinary concentrating mechanisms (Uchida *et al.*, 1993).

The distribution of CLCN5 expression within the kidney has not yet been investigated, but the association of low-molecular-weight proteinuria and hypercalciuria with disruption of this gene (Fisher *et al.*, 1994; Lloyd *et al.*, submitted) suggests that it is important for renal tubular function. Physiological studies have shown that chloride channels are present in both the

proximal tubule, where they are involved in endosomal acidification (Bae and Verkman, 1990; Reeves and Andreoli, 1992), and the distal tubule, where changes in epithelial chloride conductance play an important role in the regulation of calcium resorption (Gesek and Friedman, 1992, 1993).

Knowledge of the transmembrane structure of ClC-5 may be important in assessing the effects of different mutations on chloride channel function. Hydrophathy analysis of the ClC-5 amino acid sequence identifies 12 putative transmembrane domains (D1–D12) corresponding very closely to those domains predicted in all other known voltage-gated chloride channels. However, recent studies on glycosylation have indicated that the conventional topological representation of these proteins (Fig. 5a) may need to be revised. Consensus sites for N-linked glycosylation have been identified in the D8 and D9 linker of all ClC proteins so far isolated, including ClC-5 (see Figs. 2 and 3). The use of such sites was previously thought to be unlikely, since the conventional model predicts that this loop should be cytoplasmic, but it has recently been shown that the ClC-K channels are indeed glycosylated in the D8–D9 segment, when translated *in vitro* in the presence of pancreatic microsomes (Kieferle *et al.*, 1994). An alternative model for ClC structure placing the D8–D9 loop on the outside has therefore been proposed (Jentsch *et al.*, 1995), in which D4, the least hydrophobic of the putative domains, does not cross the membrane, and the broad hydrophobic D9–D12 region forms an odd number (3 or 5) of transmembrane spans (Fig. 5b). Glycosylation of the ClC-3/ClC-4/ClC-5 branch of the family has not yet been studied, and further analysis is required to establish which model most accurately represents the topology of ClC-5.

On investigation of exon–intron boundaries in CLCN5, we discovered that of the 12 predicted hydrophobic domains, only two, D1 and D11, are disrupted by introns (Fig. 4a). The genomic organization of CLCN5 differs greatly from that of CLCN1, the only other voltage-gated chloride channel for which the exon–intron structure is known (Lorenz *et al.*, 1994). While the open reading frame of CLCN1 is interrupted by 22 introns, we have shown that the corresponding region of CLCN5 contains only 11. Furthermore, none of the introns found in CLCN5 coincide in position with any of the CLCN1 introns.

Nephrolithiasis affects 12% of males and 5% of females by the age of 70 years (Smith, 1989), occurring as a familial disorder in up to 45% of cases (Favus, 1989; Smith, 1989), and is commonly associated with hypercalciuria (Coe *et al.*, 1992; Favus, 1989; Lemann and Gray, 1989). The elucidation of the exon–intron boundaries of CLCN5 will facilitate the screening of further patients with renal tubular dysfunction for mutations at this locus. Such studies will be necessary to establish whether the disruption of this X-linked gene

is a contributing factor to the male predisposition toward kidney stones.

ACKNOWLEDGMENTS

We are grateful to Val Cooper for manufacture of oligonucleotides. This work was supported by the HGMP (S.E.F./I.W.C.) and the MRC (S.E.L./S.H.S.P/R.V.T.).

REFERENCES

- Anand, R., Riley, J. H., Butler, R., Smith, J. C., and Markham, A. F. (1990). A 3.5 genome equivalent multi access YAC library: Construction, characterization, screening and storage. *Nucleic Acids Res.* **18**: 1951–1956.
- Bae, H.-R., and Verkman, A. S. (1990). Protein kinase A regulates chloride conductance in endocytic vesicles from proximal tubule. *Nature* **348**: 637–639.
- Bentley, D. R., Todd, C., Collins, J., Holland, J., Dunham, I., Hassock, S., Bankier, A., and Giannelli, F. (1992). The development and application of automated gridding for efficient screening of yeast and bacterial ordered libraries. *Genomics* **12**: 534–541.
- Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81**: 1991–1995.
- Cid, L. P., Montrose-Rafizadeh, C., Smith, D. I., Guggino, W. B., and Cutting, G. R. (1995). Cloning of a putative human voltage-gated chloride channel (ClC-2) cDNA widely expressed in human tissues. *Hum. Mol. Genet.* **4**: 407–413.
- Coe, F. L., Parks, J. H., and Asplin, J. R. (1992). The pathogenesis and treatment of kidney stones. *N. Engl. J. Med.* **327**: 1141–1152.
- Favus, M. J. (1989). Familial forms of hypercalciuria. *J. Urol.* **141**: 719–722.
- 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. M., 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.
- George, A. L., Crackower, M. A., Abdalla, J. A., Hudson, A. J., and Ebers, G. C. (1993). Molecular basis of Thomsen's disease (autosomal dominant myotonia congenita). *Nature Genet.* **3**: 305–309.
- Gesek, F. A., and Friedman, P. A. (1992). On the mechanism of parathyroid hormone stimulation of calcium uptake by mouse distal convoluted tubule cells. *J. Clin. Invest.* **90**: 749–758.
- Gesek, F. A., and Friedman, P. A. (1993). Calcitonin stimulates calcium transport in distal convoluted tubule cells. *Am. J. Physiol.* **264**: F744–F751.
- Jentsch, T. J., Gunther, W., Pusch, M., and Schwappach, B. (1995). Properties of voltage-gated chloride channels of the ClC gene family. *J. Physiol.* **482**: 19–25.
- 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.
- 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-Horn, F., Grzeschik, K.-H., and Jentsch,

- T. J. (1992). The skeletal muscle chloride channel in dominant and recessive human myotonia. *Science* **257**: 797–800.
- Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**: 8125–8148.
- Lemann, J., Jr., and Gray, R. W. (1989). Idiopathic hypercalciuria. *J. Urol.* **141**: 715–718.
- Lorenz, C., Meyer-Kleine, C., Steinmeyer, K., Koch, M. C., and Jentsch, T. J. (1994). Genomic organization of the human muscle chloride channel *ClC-1* and analysis of novel mutations leading to Becker-type myotonia. *Hum. Mol. Genet.* **3**: 941–946.
- Pearson, R. B., and Kemp, B. E. (1991). Protein kinase phosphorylation site sequences and consensus specificity motifs: Tabulations. *Methods Enzymol.* **200**: 62–77.
- Pook, M. A., Wrong, O., Wooding, C., Norden, A. G. W., Feest, T. G., and Thakker, R. V. (1993). Dent's disease, a renal Fanconi syndrome with nephrocalcinosis and kidney stones, is associated with a microdeletion involving DXS255 and maps to Xp11.22. *Hum. Mol. Genet.* **2**: 2129–2134.
- Reeves, W. B., and Andreoli, T. E. (1992). Renal epithelial chloride channels. *Annu. Rev. Physiol.* **54**: 29–50.
- 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.
- Smith, L. H. (1989). The medical aspects of urolithiasis: An overview. *J. Urol.* **141**: 707–710.
- Steinmeyer, K., Ortlund, C., and Jentsch, T. J. (1991). Primary structure and functional expression of a developmentally regulated skeletal muscle chloride channel. *Nature* **354**: 301–304.
- Thiemann, A., Grunder, S., Pusch, M., and Jentsch, T. J. (1992). A chloride channel widely expressed in epithelial and non-epithelial cells. *Nature* **356**: 57–60.
- Uchida, S., Sasaki, S., Furukawa, T., Hiraoka, M., Imai, T., Hirata, Y., and Marumo, F. (1993). Molecular cloning of a chloride channel that is regulated by dehydration and expressed predominantly in kidney medulla. *J. Biol. Chem.* **268**: 3821–3824.
- van Slegtenhorst, M. A., Bassi, M. T., Borsani, G., Wapenaar, M. C., Ferrero, G. B., de Conciliis, L., Rugarli, E. I., Grillo, A., Franco, B., Zoghbi, H. Y., and Ballabio, A. (1994). A gene from the Xp22.3 region shares homology with voltage-gated chloride channels. *Hum. Mol. Genet.* **3**: 547–552.
- Wrong, O. M., Norden, A. G. W., and Feest, T. G. (1994). Dent's disease: A familial proximal renal tubular syndrome with low-molecular-weight proteinuria, hypercalciuria, nephrocalcinosis, metabolic bone disease, progressive renal failure and a marked male predominance. *Quart. J. Med.* **87**: 473–493.