Alternative splicing of ClC-6 (a member of the ClC chloride-channel family) transcripts generates three truncated isoforms one of which, ClC-6c, is kidney-specific

Jan EGGERMONT*‡, Gunnar BUYSE*, Thomas VOETS*, Jan TYTGAT*†, Humbert DE SMEDT*, Guy DROOGMANS* and Bernd NILIUS*

*Laboratorium voor Fysiologie, Katholieke Universiteit Leuven, Campus Gasthuisberg, B-3000 Leuven, Belgium, and †Laboratorium voor Toxicologie, Katholieke Universiteit Leuven, Van Evenstraat 4, B-3000 Leuven, Belgium

ClC-6 is a protein that structurally belongs to the family of ClC-type chloride channels. We now report the identification of three additional ClC-6 isoforms that are truncated because of alternative splicing. We have isolated, from human K562 cells, four types of ClC-6 cDNAs that encode four distinct ClC-6 protein isoforms. ClC-6a (869 amino acids) corresponds to the previously published ClC-6 protein [Brandt and Jentsch (1995) FEBS Lett. **377**, 15–20] and it has a canonical ClC structure. However, ClC-6b (320 amino acids), ClC-6c (353 amino acids) and ClC-6d (308 amino acids) are truncated at their C-termini. Hydropathy-plot analysis indicates that the shortened isoforms contain maximally four (ClC-6b and -6d) or seven (ClC-6c) transmembrane domains. Sequence analysis of a human genomic ClC-6 fragment indicates that the cDNA variability arises from alternative splicing at two different positions: the first alternative site consists of an intron flanked by two alternative donor sites and two alternative acceptor sites, the second being due to an exon that is optionally included or excluded. Reverse-transcription-PCR analysis of ClC-6 expression in human cell lines and tissues shows that the majority (83 %) of ClC-6 mRNAs consists of ClC-6a or ClC-6c messengers. Furthermore, in a mouse tissue panel, the ClC-6a mRNA has a relatively broad tissue expression pattern, since it could be detected in brain, kidney, testis, skeletal muscle, thymus and pancreas. In contrast, expression of ClC-6c is more restricted, since it was only detected in kidney.

INTRODUCTION

ClC-type chloride channels belong to a large family that comprises at least nine different gene products in mammals (ClC-1 to ClC-7, ClC-K1 and ClC-K2) [1,2]. Sequence identity between the different mammalian ClC proteins ranges between 30 and 90 %. Hydropathy analysis together with experimental evidence for a N-linked glycosylation site has led to a common topological model featuring an even number of transmembrane domains (10 or 12) and intracellular N- and C-termini [1,3,4]. It has recently emerged that alternative splicing may further increase the structural diversity within the ClC family. For example, Adachi et al. reported the cDNA cloning of two ClC-K2 isoforms (ClC-K2L and ClC-K2S) expressed in rat kidney [5]. A region of 55 amino acids containing the second hydrophobic domain in ClC-K2L was absent from the ClC-K2S sequence, which suggests the presence of an optional exon in the rat ClC-K2 gene. In addition, Furukawa et al. isolated, from rabbit heart, two ClC-2 isoforms (ClC-2α and ClC-2β) that differed in their N-terminus [6]. However, in the absence of genomic sequence information it is not clear whether the 2α/2β sequence diversity results from alternative splicing or from alternative promoter usage. ClC-2α is identical with the ClC-2G isoform isolated by Malinowska et al. from a rabbit stomach cDNA library [7]. Rabbit ClC-2α/2G and rat ClC-2 are 93 % identical, except for a 151-amino-acid stretch where the identity is decreased to 74 %. At present it cannot be determined whether rabbit ClC-2α/2G and rat ClC-2 are species homologues or, alternatively, whether they represent splice variants.

Heterologous expression in *Xenopus* oocytes or in mammalian cells and reconstitution of purified ClC proteins in lipid bilayers have allowed determination of the functional characteristics of some ClC members such as ClC-0, ClC-1, ClC-2, ClC-5, which have all been identified as 'genuine' chloride channels [3,8–13]. The physiological importance of some of these ClC chloride channels is underscored by the fact that mutations in ClC-1 lead to myotonia congenita (either the autosomal dominant or the autosomal recessive form) [14] and that mutations in ClC-5 are linked to hypercalciuric nephrolithiasis syndromes [15]. In contrast, the functional analysis of the other ClC proteins (ClC-3, -4, -6 and -7, ClC-K1 and ClC-K2) is less advanced and has even generated conflicting results. For example, specific chloride currents have been attributed to ClC-3, ClC-K1 or ClC-K2 when expressed in *Xenopus* oocytes or in mammalian cell lines [5,16–18]. However, attempts by Jentsch and co-workers to functionally express ClC-3, ClC-K1 or ClC-K2 (and also ClC-4, ClC-6 and ClC-7) in *Xenopus* oocytes were negative [1,2,4]. Similarly, we could not attribute a specific chloride current to ClC-6, since expression of ClC-6 (corresponding to the ClC-6a protein described in the present paper) in *Xenopus* oocytes activated an endogenous chloride conductance that was also induced by expressing the structurally unrelated pI_{Cln} protein [19]. It remains therefore an open question whether all ClC proteins are genuine chloride channels located in the plasma membrane.

We now describe the alternative splicing of human ClC-6 transcripts resulting in four different ClC-6 open reading frames. The ClC-6 isoforms are identical in their N-terminus, but they possess different C-terminal regions. We have also identified the

Abbreviations used: RT-, reverse transcription.

‡ To whom correspondence should be addressed.

The nucleotide sequences of ClC-6a–ClC-6d and of a human ClC-6 genomic fragment have been deposited in the EMBL/GenBank Nucleotide Sequence Databases under the accession numbers X96391, X99473, X99474 and X99475 (ClC-6a–ClC-6d) and AC X99472 (human ClC-6 fragment).

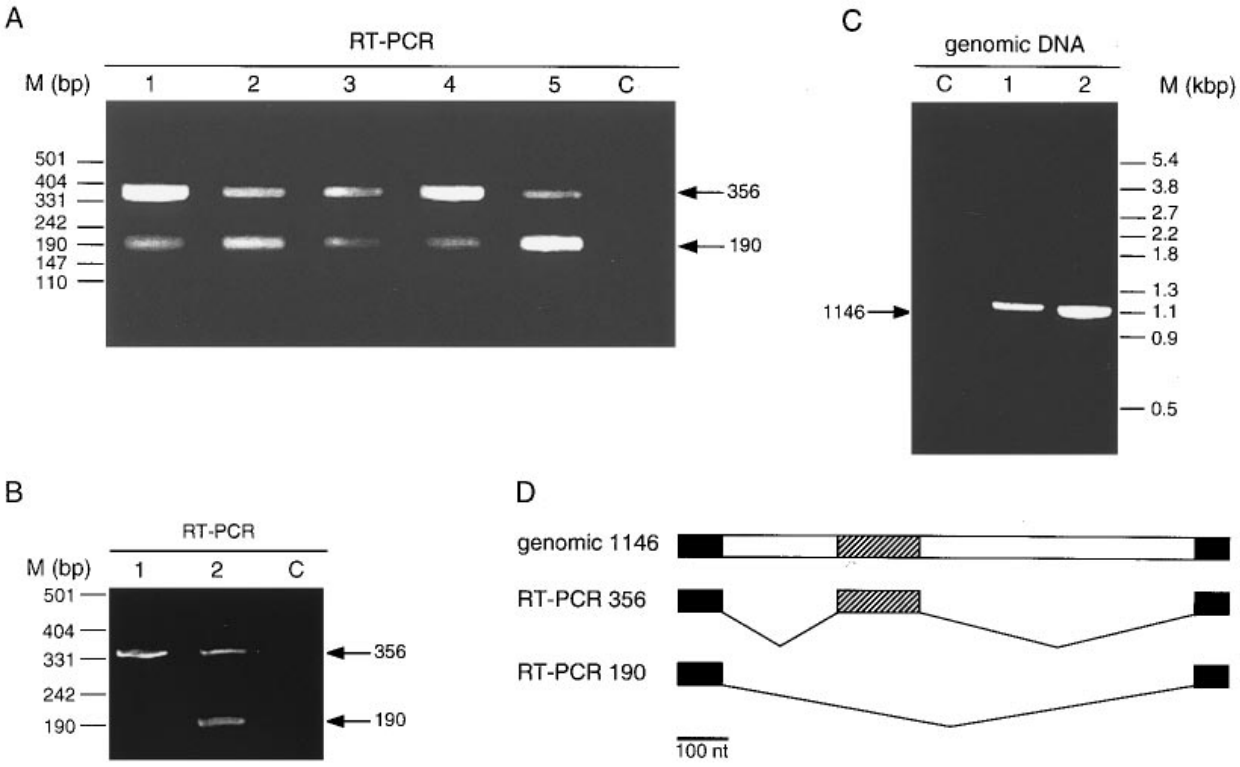

Figure 1 ClC-6 mRNAs are alternatively spliced in human cell lines and tissues

(**A**) RT-PCR analysis of ClC-6 expression in five human cell lines: Jurkat (lane 1), KB3 (lane 2), HeLa (lane 3), EA.hy926 (lane 4) and K562 (lane 5). Lane C corresponds to a control PCR reaction to which water instead of RT mix was added. M are molecular markers. The PCR was performed with the ClCXF and ClCXR primers and PCR products were analysed on a 1.5%-agarose gel. This primer set is predicted to yield a fragment of 356 nt according to the published ClC-6 sequence [2]. Note the appearance of an additional fragment of approx. 190 nt in all cell lines. (**B**) RT-PCR analysis of ClC-6 expression in human brain (lane 1) and human kidney (lane 2). Conditions were as in (**A**). (**C**) PCR analysis of human genomic DNA with the ClCXF and ClCXR primers. Lanes 1 and 2 represent DNA preparations from two different individuals. (**D**) Diagram showing the structural relationship between the two RT-PCR fragments and the genomic PCR fragment. Black boxes represent exons present in both RT-PCR fragments, the hatched box corresponds to an exon only present in the RT-PCR 356 fragment and open boxes represent introns.

sites of sequence divergence at the genomic level. Finally, reverse-transcription (RT)-PCR analysis shows a tissue-specific distribution pattern of the splice variants.

EXPERIMENTAL

RNA preparation

Total RNA of human cell lines was prepared as described by Chomczynski and Sacchi [20] or with the RNeasy kit (Qiagen). The following human cell lines used were used: HeLa (human cervical carcinoma cells), KB3 (human lung epitheloid cancer cells), Jurkat (human acute-T-cell-leukaemia cells), K562 (human chronic myelogenous leukaemia cells) and EA.hy926 (a human endothelial-cell line) [21]. Polyadenylated [poly(A)$^+$] RNA of human fetal brain and human fetal kidney was purchased from Clontech. Poly(A)$^+$ RNA of mouse tissues was prepared with the Fast Track kit (Invitrogen).

RT-PCR detection of ClC-6 splicing

Reverse transcription of 1 μg of total RNA (human cell lines), 20 ng of poly(A)$^+$ RNA (human brain) or 40 ng of poly(A)$^+$ RNA (human kidney) was performed as described in [22], except that random primers were used instead of oligo(dT) primers. Aliquots corresponding to one-tenth of the reverse-transcription reaction were then amplified with *Pfu* polymerase (Stratagene) according to the manufacturer's instructions. The following PCR primers for human ClC-6 were used (nucleotide positions are relative to the ClC-6a open reading frame): ClCXF: ACGTCTAGACCACCTTCACCCTCAACTTC (nt 860–879; preceded by an *Xba*I site); ClCXR: ACTGGATCCGCATTC-TCCTAACACCATCG (reverse of nt 1178–1197; preceded by a *Bam*HI site); ClCX1F: ACTAAGCTTGTTTAACTTCCCCT-ATTTCC (nt 678–697; preceded by a *Hin*dIII site). PCR reactions were carried out for 30 cycles; one cycle consisted of 1 min at 94 °C/1 min at 59 °C/1 min at 72 °C for the ClCXF/CLCXR primer pair and of 1 min at 94 °C/1 min at 57 °C/1 min at 72 °C for the ClCX1F/CLCXR primer pair. The ClCXF/ClCXR PCR product was cut with *Xba*I and *Bam*HI, the ClCX1F/ClCXR PCR product with *Hin*dIII and *Bam*HI. They were subcloned and sequenced bidirectionally using the automated ALF system (Pharmacia, Milwaukee, WI, U.S.A.).

Total RNA of BC_3H1 embryonic-mouse myoblasts was reverse-transcribed and amplified with the human ClCXF and ClCXR primers. The PCR fragments were subcloned and sequenced. This information was used to design the mouse-specific ClC-6 primers ClCXBC3F (TGGGATTCAGTTTGG-AAG) and ClCXBC3R (ACCAAGAGGCTCTCTAAGAC). Poly(A)$^+$ RNA fractions of mouse tissues were reverse-transcribed and PCR amplified with these murine primers using the following conditions: 1 min at 94 °C, 1 min at 55 °C and 1 min at 72 °C for 30 cycles.

RT-PCR cloning and sequence analysis of ClC-6 isoforms

Total RNA of K562 cells was reverse-transcribed and PCR-amplified as described above. PCR primers for human ClC-6 isoforms were designed based on the published sequence [2]. The forward primer (ACTGGATCCGCCATGGATAATAAGAA-AGGTC) starts at the second in-frame ATG codon (nt 202–204) of the published ClC-6 open reading frame and is preceded by a *Bam*HI site for cloning purposes. The reverse primer (ACTAAGCTTTTACCGATTTGACTCGAAG) ends at nt 1229 of the ClC-6a open reading frame and contains a *Hin*dIII site for cloning purposes. The PCR reaction was carried out with *Pfu* polymerase following the manufacturer's instructions. PCR products were digested with *Bam*HI/*Hin*dIII, subcloned in a pBluescript II vector and sequenced bidirectionally using the automated ALF system. The complete sequence of the ClC-6 fragments was derived from overlapping restriction fragments which were sequenced in both directions. Sequence analysis of these PCR products indicated the presence of four cDNAs: ClC-6a–ClC-6d. The first 201 nucleotides of the ClC-6 open reading frame were also isolated by RT-PCR from human K562 RNA.

PCR amplification and sequence analysis of human genomic ClC-6 fragments

Human genomic DNA (kindly given by Professor P. Marynen, Centre of Human Genetics, Catholic University of Leuven, Leuven, Belgium) was PCR-amplified with the ClCXF/ClCXR or the ClCX1F/ClCXR primers using *Pfu* polymerase. PCR with the ClCXF/ClCXR primers resulted in a 1100 nt fragment that was cut with *Xba*I and *Bam*HI and subcloned. The ClCX1F/ClCXR primers generated a 3.0 kb fragment that, after digestion with *Bam*HI and *Hin*dIII, was split into a 0.8 kb *Hin*dIII–*Bam*HI and a 2.2 kb *Bam*HI-*Bam*HI fragment due to the presence of an internal *Bam*HI site. The genomic fragments were subcloned and sequenced bidirectionally using the automated ALF system. A combination of custom-made primers and overlapping restriction fragments was used to determine the sequence in both directions.

RESULTS

Evidence for alternative splicing of human ClC-6 mRNAs

Brandt and Jentsch recently reported the cDNA sequence of human ClC-6, a new ClC-type protein [2]. To investigate the expression pattern of ClC-6 in non-excitable cells, we set up an RT-PCR analysis with a primer pair (CLCXF and CLCXR; see the Experimental section) specific for human ClC-6. We detected the predicted 356 nt fragment in the five human cell lines examined (HeLa, EA.hy926, KB3, Jurkat and K562), but, interestingly, all cell lines also contained to a variable extent a shorter fragment of approx. 190 nt (Figure 1A). Sequence analysis of these two PCR bands indicated that both were derived from ClC-6 mRNAs (results not shown). The 356 nt band was 100% identical with the published human ClC-6 sequence. The short fragment was identical with the 356 nt fragment, except that it lacked a central fragment of 167 nt. To check whether the short PCR fragment was also present in human tissues, we performed a similar RT-PCR on human brain and human kidney RNA (Figure 1B). Both tissues expressed the 356 nt band, but only kidney contained detectable amounts of the short band. Furthermore, analysis of human genomic DNA with these PCR primers resulted in a $\pm$1100 nt fragment (Figure 1C). Sequence analysis and comparison of the cDNA and the genomic PCR fragments indicated that the point of divergence between the short and the long RT-PCR fragments coincided with the presence of exon/intron boundaries in the genomic clone and that the 167 nt fragment was contained in a separate exon (Figure 1D; see below). These data therefore indicate that ClC-6 transcripts can be alternatively spliced in human tissues and cell lines.

```
     678                                                        727
6a   GTTTAACTTC CCCTATTTCC GAAGCGACAG ---------- AGACAAGAGA GACTTTGTAT
6b   GTTTAACTTC CCCTATTTCC GAAGCGACAG GTATGGAAAG AGACAAGAGA GACTTTGTAT
6c   GTTTAACTTC CCCTATTTCC GAAGCGACAG ---------- AGACAAGAGA GACTTTGTAT
6d   GTTTAACTTC CCCTATTTCC GAAGCGACAG ---------- ---------- ----------

     728                                                        787
6a   CAGCAGGAGC GGCTGCTGGA GTTGCTGCAG CTTTCGGGGC GCCAATCGGG GGTACCTTGT
6b   CAGCAGGAGC GGCTGCTGGA GTTGCTGCAG CTTTCGGGGC GCCAATCGGG GGTACCTTGT
6c   CAGCAGGAGC GGCTGCTGGA GTTGCTGCAG CTTTCGGGGC GCCAATCGGG GGTACCTTGT
6d   ------GAGC GGCTGCTGGA GTTGCTGCAG CTTTCGGGGC GCCAATCGGG GGTACCTTGT

     788                                                        847
6a-d TCAGTCTAGA GGAGGGTTCG TCCTTCTGGA ACCAAGGGCT CACGTGGAAA GTGCTCTTTT

     848                                                        907
6a-d GTTCCATGTC TGCCACCTTC ACCCTCAACT TCTTCCGTTC TGGGATTCAG TTTGGAAGCT

     908                                                        967
6a   GGGGTTCCTT CCAGCTCCCT GGATTGCTGA ACTTTGGCGA GTTTAAGTGC TCTGACTCTG
6b-d GGGGTTCCTT CCAGCTCCCT GGATTGCTGA ACTTTGGCGA GTTTAAG--- ----------

     968                                                       1027
6a   ATAAAAAATG TCATCTCTGG ACAGCTATGG ATTTGGGTTT CTTCGTCGTG ATGGGGGTCA
6b-d ---------- ---------- ---------- ---------- ---------- ----------

     1028                                                      1087
6a   TTGGGGGCCT CCTGGGAGCC ACATTCAACT GTCTGAACAA GAGGCTTGCA AAGTACCGTA
6b-d ---------- ---------- ---------- ---------- ---------- ----------

     1088                                                      1147
6a   TGCGAAACGT GCACCCGAAA CCTAAGCTCG TCAGAGTCTT AGAGAGCCTC CTTGTGTCTC
6b-d ---------- ---------- ---------- ----AGTCTT AGAGAGCCTC CTTGTGTCTC
```

Figure 2 Nucleotide sequence alignment of the four different ClC-6 cDNAs isolated from K562 cells

cDNAs corresponding to ClC-6a–ClC-6d mRNAs were isolated via RT-PCR of K562 RNA using PCR primers that span the region between the ATG start codon and nt 1229 of the ClC-6a open reading frame. The alignment is confined to the region where sequence variations were found (nt 678–1147). The nucleotides are numbered above the upper sequence line according to the ClC-6a sequence, with nt 1 being the A of the ATG start codon. Gaps in the nucleotide sequence are indicated by dashes. The first site of divergence occurs at nt 708, the second at nt 955. The prefixes 6a to 6d in the left column refer to the respective ClC-6 isoforms. Only one sequence line (6a-d) is given when the four sequences are identical (nt 788–907). Two sequence lines (6a and 6b-d) are given when the ClC-6b, ClC-6c and ClC-6d sequences are identical (nt 908–1147). The ClC-6a sequence is identical with the sequence published by Brandt and Jentsch [2].

K562 cells contain four different ClC-6 open reading frames

The human ClC-6 sequence (accession number X83378) reported by Brandt and Jentsch [2] contains an open reading frame of 2610 nt that encodes a peptide of 869 amino acids. The 167 nt fragment that is absent in the short RT-PCR fragment corresponds to nt 955–1121 of the ClC-6 open reading frame. Omission of the 167 nt fragment induces a frameshift in the ClC-6 open reading frame such that the TAA trinucleotide at position 1227–1229 of the X83378 open reading frame becomes a stop codon. To verify the expression of mRNAs with such a truncated open reading frame, we isolated ClC-6 cDNA clones from K562 cells by RT-PCR. PCR primers were chosen so that the amplified fragments would start at the ATG start codon of the ClC-6 open reading frame and that they would end at the TAA trinucleotide that is predicted to function as the stop codon in the alternative open reading frame. To our surprise, this analysis yielded not two, but four, types of cDNA clones which differed in sequence and which we designated 6a–6d (Figure 2). The lengths of these sequences were as follows: 6a, 1229 nt; 6b, 1072nt; 6c, 1062 nt; 6d, 1036 nt. The 6a sequence contained the optional exon of 167 nt (nt 955–1121) and it was 100% identical with the first 1229 nt of the published ClC-6 open reading frame. The three other sequences (6b, 6c and 6d) lacked the optional 167 nt exon. A second site of sequence divergence was present at nt 708 of the 6a open reading frame. At this site the 6c sequence was identical

A

```
b/c/d MAGCRGSLCC CCRWCCCCGE RETRTPEELT ILGETQEEED EILPRKDYES LDYDRCINDP 60

                                        1
b/c/d YLEVLETMDN KKGRRYEAVK WMVVFAIGVC TGLVGLFVDF FVRLFTQLKF GVVQTSVEEC 120

                                  2
b/c/d SQKGCLALSL LELLGFNLTF VFLASLLVLI EPVAAGSGIP EVKCYLNGVK VPGIVRLRTL 180

                      3                          4?
c     LCKVLGVLFS VAGGLFVGKE GPMIHSGSVV GAGLPQFQSI SLRKIQFNFP YFRSDRDKRD 240
b                                                                YGKR 240
d                                                                ---- 236

                           5                                  6
c     FVSAGAAAGV AAAFGAPIGG TLFSLEEGSS FWNQGLTWKV LFCSMSATFT LNFFRSGIQF 300
b     QERLCISRSG CWSCCSFRGA NRGYLVQSRG GFVLLEPRAH VESALLFHVC HLHPQLLPFW 300
d     --------SG CWSCCSFRGA NRGYLVQSRG GFVLLEPRAH VESALLFHVC HLHPQLLPFW 288

                  7?
c     GSWGSFQLPG LLNFGEFKSL REPPCVSGNH RGGVCGLDGV RRMPTDVLFE SNR*        353
b     DSVWKLGFLP APWIAELWRV *                                            320
d     DSVWKLGFLP APWIAELWRV *                                            308
```

B

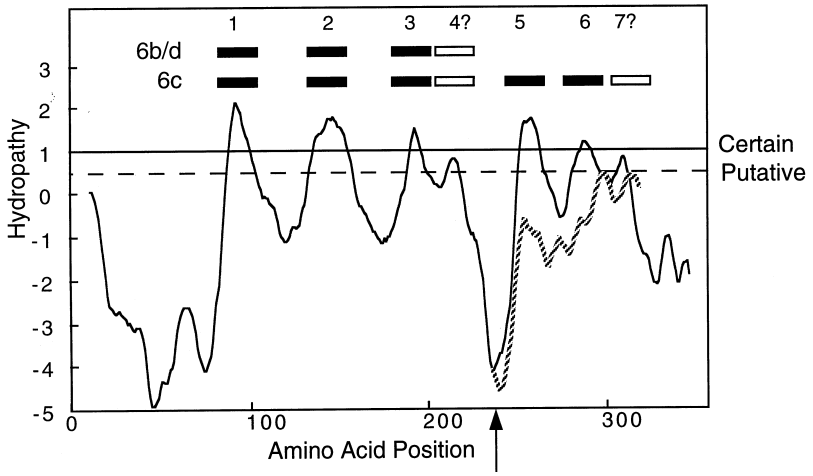

Figure 3 Amino acid sequence of the ClC-6b, -6c and -6d proteins

(**A**) Amino acid sequences (single-letter code) of the truncated ClC-6b, ClC-6c and ClC-6d isoforms. The b, c and d prefixes refer to the respective isoforms. Only one sequence line is given when the three sequences are identical (amino acids 1–236). Candidate transmembrane domains are underlined and numbered according to the hydropathy plot in (**B**). Amino acid sequences that diverge from the ClC-6a sequence are shown in **bold**. Gaps in the amino acid sequence are indicated by dashes. Asterisks denote stop codons. (**B**) Membrane topology prediction of the ClC-6c and ClC-6b/d isoforms using the TopPredII program [23]. The arrow on the *x* axis denotes the point of divergence between ClC-6b/d and ClC-6c at amino acid 236. The *y* axis shows the hydropathy score according to the **G**oldman–**E**ngelman–**S**teitz (GES) scale with a window of 21 residues [30]. The continuous curve represents the hydropathy profile for the ClC-6c isoform. Only the divergent part (downstream of amino acid 236) of the ClC-6b and -6d profile is shown (broken curve). The positions of the predicted transmembrane domains are indicated by black boxes ('certain' transmembrane: hydropathy score > 1) or white boxes ('putative' transmembrane: hydropathy score between 0.5 and 1.0).

with the 6a sequence, the 6b sequence contained an insertion of ten nucleotides and the 6d sequence contained a deletion of 26 nucleotides.

The 1229 nt 6a sequence is completely identical with the first 1229 nt of the published ClC-6 open reading frame. We therefore assume that the 6a sequence represents the human ClC-6 mRNA that encodes the 869-amino-acid protein which we have renamed ClC-6a. The sequence alterations in the 6b, 6c and 6d sequences shift the ClC-6 open reading frame and they introduce premature stop codons. The 6c sequence contains an open reading frame of 1062 nt which codes for a 353-amino-acid protein (Figure 3A). This protein is identical with ClC-6a up to amino acid 318 and has a unique C-terminus of 35 amino acids. We named this protein ClC-6c. Hydropathy-plot analysis of ClC-6c and prediction of its membrane topology were performed with the TopPredII program [23]. This identified five certain and two putative transmembrane α-helical domains (Figure 3B) that are all located in the sequence part that is in common with ClC-6a. They coincide with the transmembrane domains proposed for ClC-6a, except for the two putative domains, 4 and 7, which were not considered as transmembrane domains by Brandt and Jentsch [2]. The 6b sequence contains an open reading frame of 963 nt which encodes a 320-amino-acid protein (Figure 3A). This protein, ClC-6b, is identical with ClC-6a and 6c up to amino acid 236, but it has a unique tail of 84 amino acids. Membrane topology analysis suggests the presence of three or four transmembrane domains (Figure 3B). The 6d open reading frame is 927 nt long, and the encoded ClC-6d protein (308 amino acids)

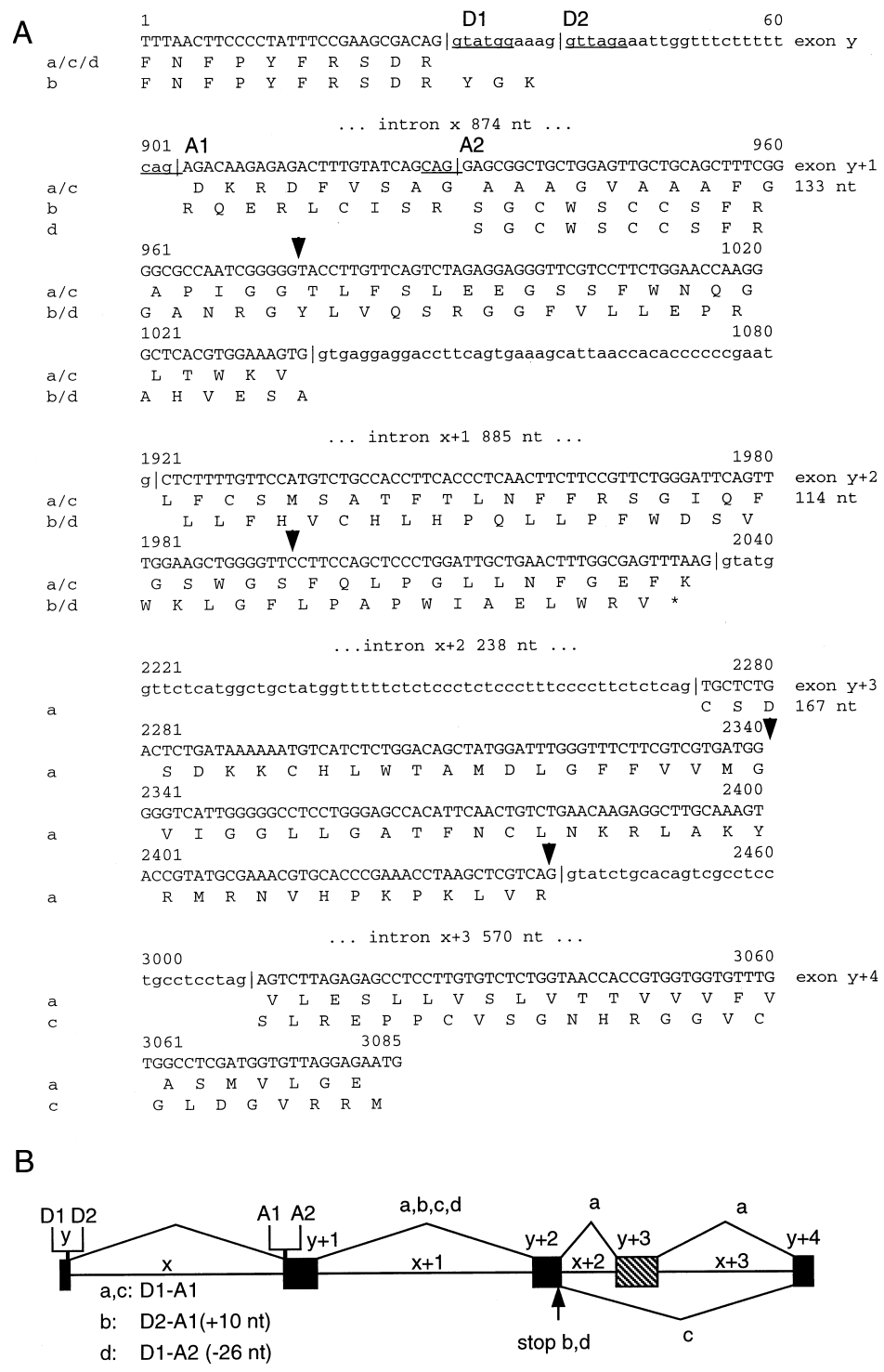

Figure 4 Nucleotide sequence of a human ClC-6 genomic fragment

(**A**) A 3085 nt fragment of the human ClC-6 gene was amplified with the ClCX1F and ClCXR primers, subcloned and sequenced. Exon sequences are represented in CAPITALS, intron sequences in lower case. Exon/intron junctions are denoted by vertical bars (|). Only the beginning and the end of the intron sequences are shown. The nucleotide sequence is continuously numbered and the numbers are given above the nucleotide sequence. Exons are numbered y to y+4 and their length is indicated at the right. Introns are numbered x to x+3 and the total length of each intron is given between the exon sequences. The first intron (intron x) is flanked by two donor (D1 and D2) and two acceptor (A1 and A2) sites. The intronic parts of the optional donor and acceptor sites are underlined. The second site of divergence is due to an optional exon (exon y+3). The inverted solid triangles (▼) indicate the positions of exon/intron boundaries in the corresponding region of the human ClC-1 gene [25]. The lines preceded by the letters a–d (shown below the nucleotide sequence line) represent the amino acid sequences of the ClC-6 isoforms in this region. The ClC-6b and ClC-6d open reading frames terminate in exon y+2 [stop codon indicated by asterisk (*)]. The open reading frames for ClC-6a and ClC-6c continue further downstream of the cloned region. (**B**) Diagram summarizing the different splice options in the genomic region. Exons are represented as black boxes (■); the optional exon (exon y+3) is shown as a hatched box. Introns are depicted as single lines. The splice patterns at intron x that have been detected by RT-PCR in K562 cells, are shown underneath intron x. Exon y+3 is included (ClC-6a) or excluded (ClC-6c). Inclusion or exclusion of exon y+3 does not affect the open reading frame of ClC-6b and ClC-6d as they contain an in-frame stop codon in exon y+2.

is identical with ClC-6b, except that amino acids 237–248 of ClC-6b are absent in the ClC-6d protein. The predicted membrane topology of ClC-6d is identical with that of ClC-6b.

Partial sequence analysis of the human ClC-6 gene: identification of two alternative splice sites

We then wanted to verify that these sequence variations were indeed due to alternative splicing. To this end we amplified a human genomic ClC-6 fragment with a PCR primer pair of which the forward primer (CLCX1F) annealed upstream of the first site of sequence variation and of which the reverse primer (CLCXR) annealed downstream of the second site of sequence variation. This yielded a PCR fragment of 3085 nt, which was subcloned and sequenced. The amplified genomic fragment contained five exons (exons y to y+4) and four introns (introns x to x+3). Importantly, the 6a, 6b, 6c and 6d sequences could all be aligned with the genomic sequence, indicating that the sequence variation was genomically encoded (Figure 4). The first intron (intron x) of the amplified fragment is flanked at its 5′ end by two donor splice sites (D1 and D2), which are separated by 10 nucleotides: CAG|GTATGGAAAG|GTTAGAAATT [vertical bars (|) denote an exon/intron junction]. Both the D1 sequence (AG|GTATGG) and the D2 sequence (AG|GTTAGA) match the mammalian consensus donor site (AG|GTRAGT; see [24]) in six out of eight residues. The 6a, 6c and 6d sequences utilize the D1 site, whereas the 6b sequence uses D2. The extra ten nucleotides in the 6b sequence (GTATGGAAAG) correspond to the ten nucleotides that separate D1 from D2. The 3′ end of this intron contains two splice acceptor sites (A1 and A2) separated by 26 nucleotides. The A1 site (Y_{15}CAG| ...; vertical bar denotes the intron/exon junction) is a close match to the mammalian consensus site ($Y_{>15}$CAG|...; see [24]) and is utilized in the 6a, 6b and 6c sequences. The A2 site (CTTTGTATCAGCAG|...) diverges substantially from the consensus sequence and is only used in the 6d sequence. The presence of two donor sites and two acceptor sites creates in principle four splice options. Splicing of D1–A1 results in a 6a/c open reading frame (the sequence between a and c diverges at the second site of variation; see below). The D2–A1 option leads to an insertion of ten nucleotides and generates the 6b sequence. The D1–A2 option causes a 26 nt deletion and creates the 6d sequence. The D2–A2 option, although theoretically possible, was not detected. Importantly, the 36 nt difference between 6b and 6d implies that these clones will utilize the same reading frame in the downstream sequence.

The second site of variation could be attributed to an optional 167 nt exon at the end of the amplified genomic fragment (exon y+3 in Figure 4). The acceptor site (Y_{31}CAG|...) and the donor site (AG|GTATCT...) that flank this exon are close matches to their respective consensus sites. This exon is included in the 6a open reading frame, but is excluded from the 6c open reading frame (exon skipping). The presence or absence of this exon does not affect the 6b or 6d open reading frame, as they contain a stop codon in the exon y+2 that precedes the optional exon. Hence any sequence variation 3′ of exon y+2 will not affect the amino acid sequence of the 6b and 6d-encoded proteins.

Expression pattern of ClC-6 splice variants

We then examined the relative occurrence of the various ClC-6 isoforms at the mRNA level by sequencing RT-PCR products from Jurkat cells, K562 cells, human kidney and human brain RNA. We used the ClC1F/ClCXR primer pair which spans both alternative splice sites. In RT-PCR analysis this primer pair should yield fragments of approx. 520 or 350 nt, depending on the inclusion or exclusion of the optional exon y+3. The sequence variation at the first optional site (+10 or −26 nucleotides) is too short to be resolved on a 1.5%-agarose gel. Therefore the ±520 nt fragments can represent ClC-6a, -6b or -6d mRNAs, whereas the ±350 nt fragments can represent ClC-6b, -6c or -6d mRNAs depending on which donor and acceptor sites are used at intron x. K562, Jurkat and human kidney contained both ±520 and ±350 nt fragments, whereas the ±520 nt fragment was the only detectable one in human brain (results not shown). We then sequenced the ±350 nt fragments of K562, Jurkat and human kidney and the ±520 nt fragments of human kidney and brain to determine the relative occurrence of the various splice patterns at intron x. The results of this analysis are shown in Table 1. The D1–A1 splice pattern was the most prevalent in both the short PCR fragment (17 out of 22 clones sequenced) and the long PCR fragment (12 out of 13 clones). The other splice patterns were only marginally represented or not detected at all: D2-A1, 4 out of 35 clones; D1-A2, 2 out of 35 clones; D2-A2, 0 out of 35 clones. D1-A1 splicing generates mRNAs that encode either ClC-6a or ClC-6c depending on the presence or absence of exon y+3. Consequently, ClC-6a and ClC-6c isoforms are the main ClC-6 isoforms, with ClC-6b and ClC-6d being less frequently expressed variants.

We then determined the tissue-specific expression of ClC-6 mRNAs in mouse. In a preliminary experiment we performed RT-PCR on mouse BC_3H1 cells with the human ClCXF and ClCXR primers. Similarly to the human cell lines, this yielded a

Table 1 Frequency distribution of ClC-6 splice patterns in human kidney and brain and in K562 and Jurkat cells

The frequency of the different splice patterns was obtained by sequencing a number of PCR products for the different tissues or cell types. RT-PCR was performed with the ClCX1F and ClCXR primers which span both alternative splice sites. This PCR yielded two fragments, one of 350 nt (PCR350) and another of 520 nt (PCR520). These fragments were sequenced to characterize their splice pattern at the first alternative site. The number of sequenced clones corresponding to a specific splice pattern at the first alternative site are given with the results split up between the short (PCR350) and the long (PCR520) fragment.

	No. of sequenced clones							
	PCR350				PCR520			
Splice pattern	Jurkat	K562	Kidney	Brain	Kidney	Brain	Total	Isoform
D1–A1	2	7	8	0	6	6	29	6a or 6c
D2–A1	1	1	1	0	0	1	4	6b
D1–A2	1	1	0	0	0	0	2	6d
D2–A2	0	0	0	0	0	0	0	–
Total	4	9	9	0	6	7	35	

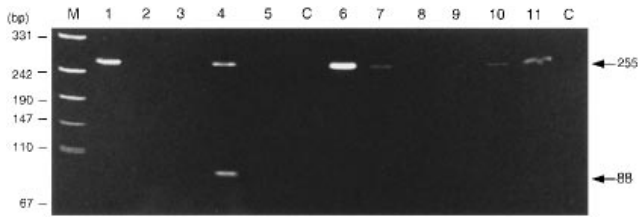

Figure 5 RT-PCR analysis of ClC-6 expression in mouse tissues

Mouse-specific primers that flank the optional exon y+3 were used in a RT-PCR reaction of mouse brain (1), liver (2), lung (3), kidney (4), spleen (5), testis (6), thymus (7), heart (8), stomach (9), pancreas (10) and skeletal muscle (11). Lanes C represent control PCR reactions with water instead of the RT mix. RT-PCR products were separated on a 6% polyacrylamide gel. Inclusion of exon y+3 results in a 255 nt fragment, whereas exclusion results in a 88 nt fragment. The 88 nt fragment was only detected in kidney. The 255 nt fragment has a more widespread expression pattern. The quality of the RT mixes was checked with a PCR for actin. This showed a fragment of comparable intensity in all tissues (results not shown).

fragment of 356 nt and another of 189 nt, thereby confirming the optional inclusion or exclusion of exon y+3 in mouse (results not shown). The sequence of these fragments was used to design murine-specific PCR primers that flanked the optional exon y+3. RT-PCR with these primers should result in a 255 nt RT-PCR fragment in the case of exon inclusion as opposed to a 88 nt fragment in the case of exon exclusion. The 255 nt fragment was clearly present in brain, kidney, testis and skeletal muscle (Figure 5). Weak to very weak signals were obtained in thymus, pancreas, liver and lung. In contrast, the 88 nt fragment was only detectable in kidney and all other tissues were negative. In conjunction with the results of the frequency distribution of the various ClC-6 mRNAs, these data indicate that ClC-6a (represented by the 255 nt fragment) has a more widespread tissue expression pattern than 6c (represented by the 88 nt fragment), which is restricted to the kidney. The relatively broad expression pattern of ClC-6a is consistent with the widespread occurrence of ClC-6 mRNAs (mainly in brain, testis, ovary, intestine and skeletal muscle) as detected by Northern-blot analysis [2].

DISCUSSION

Alternative splicing of human ClC-6: structural implications

We have characterized two alternative splice sites in the human ClC-6 gene. The upstream site of variation consists of an intron (intron x) that is flanked by two splice donor sites (D1 and D2) and two splice acceptor sites (A1 and A2). This creates four possible splice patterns, three of which (D1–A1, D2–A1 and D1–A2) were detected by sequence analysis of RT-PCR products of human tissues (brain and kidney) and cell lines (K562 and Jurkat). The D1–A1 option can be considered as the 'default' pathway, as it was the most prevalent splice pattern both in tissues and in cell lines. Alternative splicing at the second site is due to an optional 167 nt exon (exon y+3) that is either included or excluded from the mRNA (exon skipping). RT-PCR analysis of mouse and human tissues demonstrated that exon inclusion or exclusion was tissue-dependent. The exon was included in every tissue that expressed the ClC-6 gene (brain, kidney, testis, thymus, pancreas, skeletal muscle). Only kidney mRNA contained ClC-6 transcripts in which exon y+3 was excluded. Thus exon inclusion is the preferred splice pathway, with exon exclusion being a kidney-specific event. The presence or absence of exon y+3 affects the open reading frame of ClC-6 transcripts that are spliced along the D1–A1 route: exclusion induces a frameshift and a premature in-frame stop codon. Inclusion or exclusion of exon y+3 does not influence the open reading frame of ClC-6 transcripts that are processed via the D2–A1 or D1–A2 route because these mRNAs contain an in-frame stop codon in the preceding exon y+2.

The combination of splice patterns at both sites leads to four different ClC-6 open reading frames that were detected by RT-PCR: (a) pattern 6a (D1–A1; exon y+3 inclusion) generates an open reading frame that encodes a protein of 869 amino acids that corresponds to the ClC-6 protein previously described by Brandt and Jentsch [2]; we have renamed it ClC-6a to differentiate it from the other splice isoforms; (b) pattern 6b (D2–A1; exon y+3 in- or excluded) creates an open reading frame that codes for ClC-6b, a 320-amino-acid protein; (c) pattern 6c (D1–A1; exon y+3 exclusion) leads to ClC-6c, which is 353 amino acids long; (d) pattern 6d (D1–A2; exon y+3 in- or ex-cluded) gives rise to the ClC-6d isoform which has 308 amino acids. ClC-6d is identical with ClC-6b, except for a deletion of 12 amino acids. Of these four isoforms, ClC-6a and ClC-6c (D1–A1 splice pattern) are the most prevalent isoforms both at the tissue level and in cultured cell types. ClC-6a is expressed in all tissues that were positive for ClC-6. In contrast, the expression of ClC-6c is at the tissue level limited to the kidney. It is not known whether expression of ClC-6c is restricted to specific segments within the nephron. Another uncertainty is whether ClC-6a and ClC-6c are simultaneously expressed within the same renal cell types or whether their expression is mutually exclusive. The kidney-specific expression of ClC-6c suggests a specific function for this isoform in renal chloride transport, but this issue requires additional functional studies. Finally, it is noteworthy that expression of ClC-6c seems to be less stringently controlled in cultured cell types as ClC-6c mRNAs were detected in K562, Jurkat, HeLa, KB3 and EA.hy926 cells.

Although the precise membrane topology of ClC-proteins has not yet been established, it is generally assumed that they contain an even number of transmembrane domains (10 or 12) with intracellular N- and C-termini [1]. The hydropathy plot of ClC-6a is consistent with this general model [2]. However, the alternative splicing has profound structural implications for the ClC-6b, 6c and 6d isoforms. Membrane topology analysis of ClC-6c predicts five certain (domains 1, 2, 3, 5 and 6) and two putative (domains 4 and 7) transmembrane domains. We therefore propose that ClC-6c can have minimally five and maximally seven transmembrane domains with an intra- or extra-cellular C-terminus depending on whether an even (six) or an odd (five or seven) number of transmembrane segments is present. Membrane topology prediction of the ClC-6b and ClC-6d isoforms suggested three certain (domains 1, 2 and 3) and one putative (domain 4) transmembrane domains. The location of the C-terminus is therefore also uncertain depending on whether domain 4 traverses the membrane or not.

Comparison of exon/intron boundaries between ClC genes

The exon/intron composition of the human ClC-1 gene has been completely determined. [25]. Two observations indicate that the architecture of the human ClC-1 and ClC-6 genes is not conserved. First, the exon/intron boundaries of the two ClC genes do not coincide. Exons y+1, y+2 and y+3 of ClC-6 encode a region of 138 amino acids (amino acid residues 237–374 in ClC-6a). The corresponding ClC-1 region (equivalence was determined via multiple sequence analysis of mammalian ClC proteins; results not shown) is encoded by four exons (exons 7–10) [25] the boundaries of which do not coincide with the ClC-6 exon boundaries (Figure 4). Secondly, hydrophobic domain 5

contains a region that is strongly conserved between mammalian ClC proteins: FXXP(F,I,L)(G,A)G(V,T)LF(S,A)(I,L)E. The ClC-6 motif (FGAPIGGTLFSLE; residues 254–266) is contained within a single exon (exon y), whereas the ClC-1 motif (FGTPLGGVLFSIE; residues 279–291) is interrupted by an intron (intron 7 in [25]). The differences in genomic organization are consistent with the rather low degree of similarity between the human ClC-1 and ClC-6 proteins (23% identity) [2]. Thus the different genomic structure and the low degree of identity suggest that ClC-1 and ClC-6 diverged early in the evolution of the ClC proteins.

Do the structural differences between ClC-1 and ClC-6 at the genomic and at the protein level have any functional implications? An interesting comparison may be drawn with another family of transport proteins, the P-type cation-transport ATPases. This family contains cation-transport proteins that are located in the plasma membrane or in intracellular organelles and that transport a variety of cations such as Ca^{2+}, Na^+, K^+, H^+ and Mg^{2+} [26]. All P-type transport ATPases share some common sequence motifs, but the sequence identity between mammalian P-type transport ATPases may be as low as 30% [26]. Exon/intron boundaries are conserved between members with an identical transport function (e.g. the organellar Ca^{2+} pumps SERCA1 and SERCA2), but not necessarily between P-type proteins with a different transport function (e.g. SERCA versus the α subunits of Na^+/K^+-ATPase) [27–29]. Similarly, the ClC family contains proteins some of which are only distantly related; e.g. there is only 23% identity between the amino acid sequences of human ClC-1 and ClC-6 [2]. In addition, the exon/intron architecture of the human ClC-1 and ClC-6 genes is also not conserved. If the analogy with the P-type transport ATPase holds true, this could imply that proteins which share ClC-type sequence motifs are not necessarily plasma-membrane chloride channels. Alternative possibilities are that they are located in intracellular organelles and/or that they transport other substrates such as organic anions.

Note added in proof (received 27 May 1997)

The CIC-2β isoform referred to in the Introduction appears to be a cloning artefact rather than an alternatively spliced isoform [31].

We thank P. Marynen (Centre of Human Genetics, Catholic University of Leuven) for providing human genomic DNA. We thank C.-S.J. Edgell (University of North Carolina) for the gift of the EA.hy926 cell line. We acknowledge the technical expertise of D. Hermans, J. Prenen, A. Florizoone and M. Crabbé. G. B. is a Research Assistant, and J.T. and J.E. are Research Associates of the Belgium National Fund for Scientific Research.

Received 15 January 1997; accepted 14 March 1997

REFERENCES

1 Jentsch, T. J., Gunther, W., Pusch, M. and Schwappach, B. (1995) J. Physiol. (London) **482.P**, S19–S25
2 Brandt, S. and Jentsch, T. J. (1995) FEBS Lett. **377**, 15–20
3 Middleton, R. E., Pheasant, D. J. and Miller, C. (1994) Biochemistry **33**, 13189–13198
4 Kieferle, S., Fong, P. Y., Bens, M., Vandewalle, A. and Jentsch, T. J. (1994) Proc. Natl. Acad. Sci. U.S.A. **91**, 6943–6947
5 Adachi, S., Uchida, S., Ito, H., Hata, M., Hiroe, M., Marumo, F. and Sasaki, S. (1994) J. Biol. Chem. **269**, 17677–17683
6 Furukawa, T., Horikawa, S., Terai, T., Ogura, T., Katayama, Y. and Hiraoka, M. (1995) FEBS Lett. **375**, 56–62
7 Malinowska, D. H., Kupert, E. Y., Bahinski, A., Sherry, A. M. and Cuppoletti, J. (1995) Am. J. Physiol. **268**, C191–C200
8 Steinmeyer, K., Ortland, C. and Jentsch, T. J. (1991) Nature (London) **354**, 301–304
9 Jentsch, T. J., Steinmeyer, K. and Schwarz, G. (1990) Nature (London) **348**, 510–514
10 Thiemann, A., Grunder, S., Pusch, M. and Jentsch, T. J. (1992) Nature (London) **356**, 57–60
11 Sakamoto, H., Kawasaki, M., Uchida, S., Sasaki, S. and Marumo, F. (1996) J. Biol. Chem. **271**, 10210–10216
12 Steinmeyer, K., Schwappach, B., Bens, M., Vandewalle, A. and Jentsch, T. J. (1995) J. Biol. Chem. **270**, 31172–31177
13 Pusch, M., Steinmeyer, K. and Jentsch, T. J. (1994) Biophys. J. **66**, 149–152
14 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) Science **257**, 797–800
15 Lloyd, S. E., Pearce, S. H. S., Fisher, S. E., Steinmeyer, K., Schwappach, B., Scheinman, S. J., Harding, B., Bolino, A., Devoto, M., Goodyer, P. et al. (1996) Nature (London) **379**, 445–449
16 Kawasaki, M., Suzuki, M., Uchida, S., Sasaki, S. and Marumo, F. (1995) Neuron **14**, 1285–1291
17 Kawasaki, M., Uchida, S., Monkawa, T., Miyawaki, A., Mikoshiba, K., Marumo, F. and Sasaki, S. (1994) Neuron **12**, 597–604
18 Uchida, S., Sasaki, S., Furukawa, T., Hiraoka, M., Imai, T., Hirata, Y. and Marumo, F. (1993) J. Biol. Chem. **268**, 3821–3824
19 Buyse, G., Voets, T., Tytgat, J., De Greef, C., Droogmans, G., Nilius, B. and Eggermont, J. (1997) J. Biol. Chem. **272**, 3615–3621
20 Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159
21 Edgell, C.-J. S., McDonald, C. C. and Graham, J. B. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3734–3737
22 De Greef, C., Sehrer, J., Viana, F., Van Acker, K., Eggermont, J., Mertens, L., Raeymaekers, L., Droogmans, G. and Nilius, B. (1995) Biochem. J. **307**, 713–718
23 Claros, M. G. and von Heijne, G. (1994) Comput. Appl. Biosci. **10**, 685–686
24 Lamond, A. I. (1993) BioEssays **15**, 595–603
25 Lorenz, C., Meyer-Kleine, C., Steinmeyer, K., Koch, M. C. and Jentsch, T. J. (1994) Hum. Mol. Genet. **3**, 941–946
26 Green, N. M. (1992) Ann. N. Y. Acad. Sci. **671**, 104–112
27 Eggermont, J., Wuytack, F. and Casteels, R. (1991) Biochim. Biophys. Acta **1088**, 448–451
28 Korczak, B., Zarain-Herzberg, A., Brandl, C. J., Ingles, C. J., Green, N. M. and MacLennan, D. H. (1988) J. Biol. Chem. **263**, 4813–4819
29 Zarain-Herzberg, A., MacLennan, D. H. and Periasamy, M. (1990) J. Biol. Chem. **265**, 4670–4677
30 Engelman, D. M., Steitz, T. A. and Goldman, A. (1986) Annu. Rev. Biophys. Biophys. Chem. **15**, 321–353
31 Furakawa, T., Horikawa, S., Terai, T., Ogura, T., Katayama, Y. and Hiraoka, M. (1997) FEBS Lett. **403**, 11