© 1997 Oxford University Press Nucleic Acids Research, 1997, Vol. 25, No. 20 4153–4159

Alternative mRNA splice variants of the rat ClC-2 chloride channel gene are expressed in lung: genomic sequence and organization of ClC-2

Shijian Chu* **and Pamela L. Zeitlin**

Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

Received March 11, 1997; Revised and Accepted August 18, 1997

ABSTRACT

The ClC-2 epithelial cell chloride channel is a voltage-, tonicity- and pH-regulated member of the ClC super family. We have previously shown that rat lung ClC-2 (rClC-2) is down-regulated at birth, and molecular diversity is generated by alternative splicing [Murray *et al.* (1995) *Am. J. Respir. Cell Mol. Biol.* 12, 597–604; Murray *et al.* (1996) *Am. J. Physiol.* 271, L829–L837; Chu *et al.* (1996) *Nucleic Acids Res.* 24, 3453–3457]. To investigate other possible mRNA splice variations, we sequenced the entire rClC-2 gene and found that ClC-2Sa (formerly ClC-2S) results from the deletion of exon 20. The preceding intron 19 has an unusually high CT content and a rare AAG acceptor site. Because both features were also found in intron 13, we next tested the hypothesis that intron 13 would be involved in alternative splicing. As predicted, a second splice product, ClC-2Sb, was found by RT–PCR, but only in lung. When we compared the genomic maps of rClC-2 and human ClC-1 (hClC-1), striking similarities were found in each exon except for rClC-2 exon 20, which is absent in hClC-1. These observations suggest that ClC-1 and ClC-2 may have evolved by gene duplication, mutation and DNA rearrangement.

INTRODUCTION

ClC-2 is a voltage-, tonicity- and pH-regulated chloride channel (4–8), and a member of a chloride channel super family, ClC, which includes more than 10 members found in bacteria, yeast, fish and mammals (9). Whereas mutations in ClC-1 and ClC-5 are responsible in humans for myotonia and Dent's disease respectively (10–12), no disease correlations have yet been found for other gene members. Some ClC proteins are confined to specific tissues in animals, whereas others are ubiquitously expressed. Despite differences in function and tissue distribution, the basic structure of these proteins is well conserved. The predicted proteins have an N-terminal membrane anchoring region with more than 10 putative membrane spanning segments and a C-terminal cytosolic domain.

Among the ClC members, only human ClC-1 has been characterized at the genomic level (11,13), however, no detailed information on introns and alternative splicing was given. Truncated short mRNA transcripts have been observed in at least three ClC family members (3,14,15), raising the possibility that additional molecular diversity in function may be generated by alternative splicing.

We previously identified a short form ClC-2 mRNA transcript, ClC-2Sa, that is produced in multiple tissues by alternative splicing (3). We have also demonstrated that deletion of exon 20 predicts a protein that lacks a 20 amino acid residue peptide sequence in the C-terminal cytosolic region (3). We now report the genomic structure of rat ClC-2. A lung-specific alternative splicing event next to intron 13 is described which is similar to the alternative splicing in ClC-2Sa (3) and predicts a shortened 54 kDa protein ClC-2Sb. Both alternative splicing events occur adjacent to an intron with a high CT content and an AAG sequence at the 3′-end of the acceptor site. There are striking similarities in the genomic structures rClC-2 and hClC-1. One major difference, however, is that hClC-1 is missing the sequence comparable to exon 20 of rClC-2. This suggests that alternative splicing involving the AAG acceptor site may play a role in generating molecular diversity not only in a gene, but in the evolution of a gene family as well.

MATERIALS AND METHODS

Cell lines and culture conditions

L2 cells are derived from type II-like alveolar pneumocytes from normal adult female rat lung and were obtained from the American Type Culture Collection (ATCC, Rockville, MD). The L2 line was maintained in monolayer culture in Ham's F-12K (Biofluids, Rockville, MD), containing 10% fetal calf serum, 2.5 mg/ml fungizone, 100 U/ml penicillin G and 100 U/ml streptomycin. The cell medium was replaced three times a week, and the line was split 1:3 after reaching confluence.

PCR amplification and cloning of PCR fragments

PCR amplifications were performed on a Perkin Elmer DNA Thermal Cycler 480. Before addition of the Taq polymerase (Boehringer Mannheim, Indianapolis, IN), the enzyme was mixed with TaqStart antibody according to the manufacturer's protocol (Clontech, Palo Alto, CA). The primers are listed in Table 1. For amplification of intron 23 from rat genomic DNA using oligos SCA and SCB, 35 PCR cycles were performed and

*To whom correspondence should be addressed. Tel: +1 410 955 2035; Fax: +1 410 955 1030; Email: shijian@welchlink.welch.jhu.edu

each cycle included 1 min at 94°C, 40 s at 59°C and 40 s at 72°C. For amplification of ClC-2Sb from 19 day rat lung cDNA, oligo pair RKRL(T7-3)/B23′ was used. Forty cycles were performed and each consisted of 40 s at 94°C, 40 s at 51°C and 40 s at 72°C. PCR reaction mixes containing amplified products were first cloned into the pCRII plasmid (Invitrogen, San Diego, CA) according to the manufacturer's protocol. The recombinant plasmid DNA was then purified and used as the template in automated DNA sequencing.

Table 1. Oligonucleotides used in PCR experiments

Name	Sequence
SCA	CACCAGCATTGGCAGACTCATTG
SCB	TGACAGAGCCTTCAATGGCCTTC
B23′	TAGGGCAGCTTCTTGATG
RKRL(T7-3)	CCAACGTCTTCCTGACTC

RNA, cDNA and DNA preparation

Total RNA was prepared from rat tissues or cultured cell monolayers using Trizol reagent (GIBCO/BRL, Gaithersburg, MD) following the manufacturer's protocol. The quality and quantity of the total RNA were examined by agarose gel electrophoresis and UV spectrophotometry. The cDNA was prepared from 0.1 µg of total RNA with the cDNA Cycle Kit (Invitrogen), using the oligo dT primer. For subsequent PCR amplification, 20% of the prepared cDNA was used in each 50 µl reaction as the template.

Intron sequences were amplified from genomic DNA by PCR. Genomic DNA was prepared from L2 cells using a Puregene DNA Isolation Kit (Gentra Systems, Inc., Minneapolis, MN) according to the manufacturer's protocol.

Plasmid DNA from rat genomic clone 6369 was prepared as described (16) with modifications. Briefly, 40 ml of L-broth was inoculated and incubated at 37°C for 1.5 h with shaking. IPTG (isopropyl β-D-thiogalactopyranoside) was then added to a final concentration of 0.5 mM. The bacterial cells were harvested 5 h after induction, and the DNA was extracted by the alkaline lysis method with one phenol/chloroform extraction. After isopropanol precipitation, the DNA was dried and resuspended in 0.1 ml of sterile distilled water. The DNA was precipitated again by addition of 26 µl of 30% PEG in 1.5 M NaCl and 3.4 µl of 3 M NaCl and incubation on ice for 30 min. The precipitate was collected by centrifugation and solubilized in 0.8 ml of sterile distilled water. This DNA was then extracted once with chloroform, precipitated with ethanol and resuspended in 60 µl of sterile distilled water for automated sequencing.

DNA sequencing and sequence analysis

DNA sequencing was performed on the ABI DNA Sequencer Model 373 using the genomic clone 6369 DNA as the template. Sequencing primers were synthesized based on the information from each previous sequencing reaction.

The amino acid sequences coded by the corresponding exons in hClC-1 and rClC-2 were paired and sent to Genome at EERIE through the World Wide Web (http://genome.eerie.fr/fasta/lfasta-query.html) for online alignment. The alignments were done with matrix blosum50.mat and ktup=2, and the results are presented in Table 2 as a percentage of identical residues in a stretch of the aligned sequences. If more than one alignment appeared between a pair of amino acid sequences, the data from the longest alignment was presented regardless of the level of the percentage of identical residues.

RESULTS

Genomic organization of the ClC-2 gene

Rat genomic DNA was prepared from L2 cells (rat alveolar Type II cell line) and amplified by PCR to identify genomic sequences. Amplification with primers SCA and SCB (Table 1; 2487–2509 and 2535–2557 in the published ClC-2 cDNA coding sequence, GenBank accession number X64139) yielded a fragment of 195 bp, which is 124 bp longer than the fragment size expected from the cDNA. Sequencing of this PCR product demonstrated that the extra 124 bp is an intron (intron 23) with typical sequence structures (GT at the 5′-end and AG at the 3′-end). Using this fragment as the probe, a genomic clone, 6369, was selected from a rat genomic DNA library constructed in P1 phage (Genome Systems, St. Louis, MO). This clone was then confirmed by Southern hybridization and subsequent DNA sequencing.

Comparison of the full length genomic DNA sequence with the published rat cDNA sequence (6) revealed that the entire ClC-2 gene contains 24 exons separated by 23 introns (Fig. 1). The first 15 bases in the published ClC-2 cDNA sequence were not identified in the 1 kb region upstream from the exon containing the first Met codon. Therefore, this 15 bp sequence was not included as an exon.

Exon 1 contains the first part of the ClC-2 coding region and 153 bp of upstream untranslated sequence with a GC content of 73%. At the 3′-end of the transcript, an untranslated region in the last exon contains 315 bp, ending at the second A in the poly A tail of the published cDNA sequence. However, it is unknown if the last two As in the mRNA were added during polyadenylation or during the transcription of the gene. Therefore, the cleavage site may be 313 or 315 bases downstream from the last codon of the gene. In both cases, the positions of the cleavage sites are consistent with the consensus distance (10–30 bases) from the last cleavage signal AAUAAA.

The lengths of the introns range from 83 to 3489 bp, with three introns (1, 10 and 22) longer than 1 kb (Table 2). The first two residues at the 5′-end in all 23 introns are completely conserved as GT. The third position is either an A (13/23) or a G (10/23). The fourth residue is usually an A (15/23) but is less well conserved. The 3′-end two residues in all 23 introns are AG. The third residue from the 3′-end is also conserved, with C and T as the predominant bases (15/23 and 6/23 respectively). A appears in this position in only two introns, 13 and 19. Since intron 19 is immediately upstream from an unstable 60 bp exon, exon 20 (3), we compared other structural features of these two introns. Both have the highest CT contents among all the introns in the gene (68% and 71%; Table 2). Intron 19 has three potential branch point sequences (3), and intron 13 contains a consensus branch point sequence at –66 to –60 relative to its 3′ splice boundary (CCCTGAC, Fig. 2A).

The transmembrane domains are coded by exons 3–15. Most transmembrane domains, except for D5, D7, D9 and D10, are coded by single exons (Fig. 3). The coding sequences for D5 and D7 are unevenly divided by two exons, while D9 and D10 appear

Table 2. Comparison of genomic organizations of rClC-2 and hClC-1

hClC-1 exons			rClC-2					
			#	exons		introns		
#	%id.*	bp		bp	sequences***	sequences***	bp	CT%
1	-	>180	1	234	CCTGCG...ACCCTG	gtaagt...ctccag	2290	54
2	-	121	2	157	ATGTAT...GTCGCA	gtaaga...ttccag	152	42
3	63	132	3	132	TTTGTT...TACAGG	gtgagg...ccacag	318	56
4	47	129	4	129	CTCAGC...CTGTTG	gtagga...ctgcag	87	42
5	82	134	5	134	GGTCTG...AAGGAG	gtaatt...ctccag	104	49
6	77	78	6	78	GGACCC...TATGAG	gtaaga...tgccag	83	41
7	57	79	7	79	AACGAG...TCGGAG	gtaggc...ctgcag	132	61
8	93	126	8	126	GGGTCC...ATGAAG	gtggga...ctttag	102	54
9	82	85	9	85	AAACCA...CATTGG	gtaagg...ctttag	100	45
10	46	102	10	102	TATTGC...GAAGAA	gtacgt...ctctag	1090	51
11	64	85	11	85	ACGGCT...GGACAG	gtagcc...ccgtag	125	58
12	43	150	12	156	CTCTCA...ATGAAG	gtgaga...ctacag	311	53
13	74	70	13	70	TTCTGG...TCATTG	gtgagt...ctgaag	242	68
14	70	111	14	111	GAGCAG...TGGTCG	gtgagc...ttccag	276	48
15	84	214	15	214	GGGCAG...CCACCA	gtatgg...ctccag	323	53
16	49	134	16	134	GCAGTA...CACCTG	gtgaga...gcccag	183	51
17	40	242	17	173	AGTCCA...TTCCAG	gtaagg...atgtag	103	53
18	44	112	18	115	GTGAAC...CCACGG	gtgagc...acacag	168	58
19	-	80	19	77	GCAACA...TCAGAA	gtgagt...acaaag	208	71
			20	60	TTGGAA...CTGGCG	gtacgt...tcctag	134	51
20	-**	39	21	39	AGTGAC...GAGGAG	gtgaga...cctcag	166	51
21	60	105	22	105	ATCTTA...CACAAG	gtaggc...ctgcag	3489	43
22	62	87	23	87	ACCCAC...AAGGAG	gtgagg...ttccag	124	49
23	43	>371	24	507	CTCCGG...ATAGAA			

*This column compares hClC-1 and rClC-2 amino acid sequences coded by corresponding exons with numbers representing percentages of identical residues in a stretch of aligned sequences and '–' indicating no similarity. The method of alignment and the criteria for presentation are explained in the Materials and Methods section.
**Although the alignment at EERIE did not give any similarity, alignment by hand showed 54% identity.
***In each box, six residues at both ends of an exon (or an intron) are presented (GenBank accession number AF005720).

to be divided evenly by two exons, i.e., D9 is at the junction of exons 12 and 13, and D10 is at the junction of exons 13 and 14 (Fig. 3).

A new rClC-2 short form (ClC-2Sb) mRNA transcript occurs by alternative splicing and a cryptic 3′ splice site

Since the upstream intron next to the unstable exon 20 has a very high CT content and an AAG 3′ splice site, we speculated that these two unique structural features may account for the instability of exon 20. Intron 13 has similar features, a high CT content and AAG at the 3′ splice site. This suggests that intron 13 may also have the same potential to participate in alternative splicing involving adjacent exons. To investigate this possibility, RT–PCR was performed on rat cDNA using primers RKRL(T7-3) and B23′ (Table 1), located in exon 12 and exon 15 respectively (Fig. 2B). In addition to a fragment of the expected length of 406 bp, a shorter fragment was present in rat lung cDNAs (Fig. 4). This fragment was then amplified from 19 day gestation fetal rat lung cDNA, cloned and sequenced. Sequencing data confirmed a missing segment including the entire exon 13 and the first 6 bp in exon 14. This observation suggests that the 3′ splice site AAG in intron 13 is not efficiently recognized. Instead, the splicing machinery recognized a cryptic acceptor site CAG inside exon 14 which is usually more common than AAG in intron acceptor sites (Fig. 2B). Although it is not proven by our

Figure 1. Exon–intron arrangement of the ClC-2 gene.

A.

TCATTGgtgagtctgcctctgcctgcctcccagcctgcccctcccctaccctctggctatggtcttgtggagtccttctctgcccagg
agtgggtgttgacccaactcttgccccaaccccaactcccgttaggggactgctacccctttactaaatgacccaccgggccctccc
ctactc<u>ccctgac</u>cctctctcttcatagtgtcccccactacctctctctcttcctaaaccctccctgaagGAGCAG

Figure 2. (**A**) DNA sequence of intron 13. Bases at the beginning and the end of the sequence in capital letters represent flanking sequences from exons 13 and 14 respectively. A consensus branch-point sequence is underlined. (**B**) Schematic drawing of the alternative splice event involving exons 13 and 14. Exons 12–15 are depicted by boxes and labeled E12–E15. Introns are indicated by straight lines between boxes and labeled i12– i14. The sequence at the junction between intron 13 and exon 14 is shown below. The bent line above the introns and exons indicates the alternative splice event. Two arrows beneath exons 12 and 15 represent primers used in RT–PCR.

Figure 3. Distribution of exons in relation to predicted protein hydrophobicity and putative membrane spanning domains of rClC-2.

studies that the deletions of exons 13 and 20 are due to the conservation of a high CT content and an AAG 3′ splice site in introns 13 and 19, it is evident that both alternative splicing events occurred adjacent to these introns and in both cases splice mechanisms failed to recognize AAG as the 3′ splice site. The deletion of the 76 residues in this alternative splicing event would generate a reading frame shift which would produce a shortened ClC-2Sb protein of 54 kDa, containing 491 amino acid residues up to and including the first half of the D9 membrane spanning domain. The 43 amino acid residues at the C-terminus created by the frame shift would not form any significant hydrophobic membrane spanning domain (Fig. 5).

DISCUSSION

An increasing number of chloride channels have been identified in the past decade. Many of them co-exist in a given mammalian organ. Electrophysiological studies in airway epithelial cells have demonstrated the co-existence of CFTR (cystic fibrosis transmembrane conductance regulator), ORCC (outwardly rectifying chloride channel) (17–19) and a 140 kDa Ca^{2+}/calmodulin kinase II-sensitive chloride channel (20,21). In whole lung, Northern hybridization has demonstrated the presence of several different transcripts from the ClC super family including ClC-2, -3, -5, -6 and -7 (6,22–24). In addition, alternative splicing in these transcripts further increases their potential diversity.

Intron structure and alternative splicing

In our study, introns were identified in the ClC-2 genomic sequence by comparison to the published cDNA sequence. At exon/intron boundaries, the first two bases and the last two bases in all introns are identical. The third base at the 3′-end is usually C or T in 95% of introns (25). Although an A or a G at this position is recognized as a rare event, the consequence and significance of this base change has not been previously emphasized. Our study presents an example where the only two AAG 3′ splice sites in the gene are next to an alternative mRNA splicing site and both of the alternative splicing events fail to recognize AAG as the 3′ splicing sites.

The co-existence of the AAG 3′ splice site and alternative RNA splicing has been observed before (Table 3) although, in most cases, this relationship has not been discussed. One example is the bovine tau gene which encodes the microtubule-associated proteins that display extensive size heterogeneity. Himmler (26) demonstrated that this heterogeneity is due to multiple combinations of unstable exons. Four unstable exons, exons 3, 6, 8 and 10, have been found in the tau 13-exon coding region and three out of the four upstream introns, introns 5, 7 and 9, have AAG as their 3′ splice site. Furthermore, these three are the only introns ending with AAG in the entire gene. Perhaps the simplest example of the inefficient use of AAG as the 3′ splice site by mammalian splice machinery is the alternatively spliced type C atrial natriuretic peptide receptor (27). At the junction of the sixth intron and seventh exon with the sequence AAGCAG, one splice variant recognizes AAG as the 3′ splice site while the other uses CAG, resulting in protein products of 537 and 536 amino acid residues respectively with the longer form being more abundant in the tissues examined.

Table 3. Examples of introns with AAG as their 3′ splice sites which are next to alternative mRNA splicing.

Species	Gene name/disease	Intron with AAG	Alternative splicing event	Ref.
Human	tau	upstream from the unstable exon	A 93 bp exon is deleted in type I tau mRNA, but present in type II	(37)
Human	Central cannabinoid receptor (CB1)	a 167 bp intron	the intron spliced in CB1A mRNA but retained in CB1 mRNA	(38)
Human	β-globin/β-thalassemia	IVS-2	Use of a cryptic 3′ splice site CAG inside the intron, producing a longer RNA	(29,39)
Human	lipoprotein lipase/ hyperchylomicronemia	intron 6	Deletion of exon 7 or exons 6–9, decreased quantity of mRNA	(28)
Human	Wilson disease gene (ATP7B)	intron 5	Deletion of exons 6 and 7, or 6 through 8	(40)
Human	A1 adenosine receptor	intron 3	Deletion of exon 4	(41)
Human	Splicing factor 9G8	a cryptic AAG acceptor site inside intron 3	Use of the AAG acceptor site; retention of intron 3	(42)
Human	Type C atrial natriuretic peptide receptor	intron 6	Use of a cryptic 3′ splice site CAG in exon 7	(27)
Human	Hypoxanthine-phosphoribosyltransferase	intron 7	Deletion of exon 8	(30)
Bovine	tau	introns 5, 7 and 9	Deletion of exon 6, 8 or 10	(26)
Bovine	γ2 subunit of $GABA_A$ receptor	upstream from the unstable exon	Deletion of a 24 bp exon in γ2S mRNA but present in γ2L mRNA	(43)
Rat	AMPA receptor	two introns upstream from the unstable exons	Mutually exclusive expression of two adjacent exons	(44)
Rat	Clathrin light chain B	intron D	Deletion of exon N in many tissues except for brain	(45)

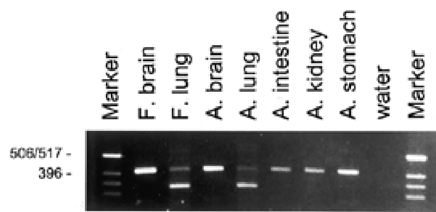

Figure 4. Agarose gel electrophoresis of PCR products amplified with primer pair RKRL(T7-3) and B23′. Fetal (F.) and adult (A.) rat tissues from which the template cDNAs were prepared are indicated. 'Water' represents a control reaction without template DNA. Lanes on both sides are size markers and are labeled on the left (bp).

MAAATAAAATVAGEGMEPRALQYEQTLMYGRYTQELGAFAKEEAARIRLGGPEPWKGSPS
ARATPELLEYGQSRCARCRICSVRCHKFLVSRVGEDWIFLVLLGLLMALVSWAMDYAIAV
CLQAQQWMSRGLNTNILLQYLAWVTYPVVLITFSAGFTQILAPQAVGSGIPEMKTILRGV
VLKEYLTLKTFVAKVIGLTCALGSGMPLGKEGPFVHIASMCAALLSKFLSLFGGIYENES
RNTEMLAAACAVGVGCCFAAPIGGVLFSIEVTSTFFAVRNYWRGFFAATFSAFIFRVLAV
WNRDEETITALFKTRFRLDFPFDLQELPAFAVIGIASGFGGALFVYLNRKIVQVMRKQKT
INRFLMKKRLLFPALVTLLISTLTFPPGFGQFMAGQLSQKETLVTLFDNRTWVRQGLVED
LGAPSTSQAWSPPRANVFLTLVIFILMKHLAGWWARAWLPGSQMGFTRIAVPTGLYLEAM
PWSGQLHLQEQ

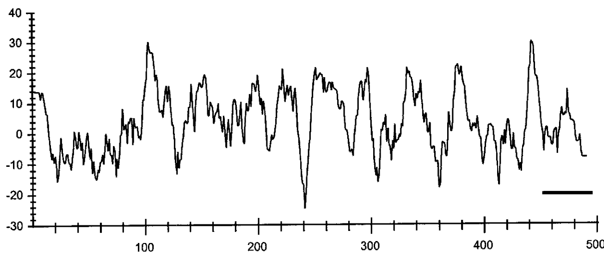

Figure 5. Predicted amino acid sequence and hydropathy plot for ClC-2Sb. The amino acid sequence and the corresponding plot created by the frame shift are underlined.

The possibility exists that the observed association of AAG 3′ splice sites with alternative splicing occurred by chance. However, this is unlikely given the fact that two out of two introns with 3′ AAG in rClC-2 and three out of three in the bovine tau gene were found next to alternative splicing events. In addition, two disease-related mutant genes have been identified in humans where CAG 3′ splice sites were mutated to AAG sites, and alternative splicing was observed in patients carrying these mutations (Table 3) (28,29). Furthermore, *in vivo* mutagenesis of the hprt (hypoxanthine–phosphoribosyltransferase) gene by Recio *et al*. (30) changed the –3 position in intron 7 from T to A and this change resulted in deletion of exon 8 (Table 3). Finally, *in vitro* studies have shown that AAG is a less favorable 3′ splice site relative to CAG and TAG (31). Ultimately, this type of mutation analysis can be applied to ClC-2 *in vitro* to test our hypothesis.

The high CT content in two rClC-2 introns may also be significant. Examples exist demonstrating that CT-rich sequences in introns negatively affect intron splicing in chicken β-tropomyosin (32) and in adenovirus (33). The structure of the intron between the two mutually exclusively expressed exons of the chicken β-tropomyosin gene is similar to the rClC-2 intron 13 where the branch point is unfavorably far away from the 3′-end and is followed by a long stretch of CT-rich sequence (32). These observations are consistent with our results even though the AAG 3′ splice sites in rClC-2 introns may play an additional role in suppressing the usual splicing. The multiple patterns of alternative splicing and the co-existence of 'normally' and alternatively spliced transcripts in the examples listed in Table 3 are likely the result of the interactions of many regulatory factors.

Tissue specificity of alternative splicing

The predominant splice patterns of a gene are often dependent on the tissue studied. Therefore, it is sometimes difficult to tell which 3′ splice site is 'cryptic', or which splice pattern is 'normal' and which is 'alternative'. Either one of the two ClC-2 splice products could be dominant in a given tissue, indicating the participation of other tissue-specific *trans*-acting factors in addition to the AAG 3′ splice sites. In rClC-2, both ClC-2Sa and ClC-2Sb are more abundant relative to the full length transcript in the lung than the brain, suggesting a common splice mechanism in both events which may involve tissue specific *trans*-acting factors as well as the intron structures discussed above. However, these tissue-specific factors are largely unknown in mammalian cells.

It is likely that these splice variants differ functionally as well as structurally. Although ClC-2 mRNA is widespread (6), the chloride conductance characteristics vary significantly. The N-terminal region has been studied by mutational analysis and demonstrated to be sensitive to tonicity (5). We speculate that the C-terminal structural changes observed in the splice variants could affect interactions of ClC-2 with other molecules including other ClC-2 proteins as has been reported for ClC-0 and ClC-1 (34–36). However, since detailed mutagenesis studies have not yet been done, the role of the ClC-2 cytosolic region at the C-terminus is not known.

Similarities and differences in genomic structures of rClC-2 and hClC-1

Comparison of rClC-2 and hClC-1 (11,13) reveals striking similarities in their exon/intron arrangement and the positions of the putative *trans*-membrane domains relative to the exons (Table 2). Although the sizes of the first and the last exons are unknown in the hClC-1 gene and the primary sequences of corresponding exons are clearly different between the two genes, 17 of 23 hClC-1 exons contained identical numbers of base pairs as those in rClC-2. Strikingly, exon 20 of rClC-2 does not have a counterpart in hClC-1 (Table 2). This exon is located in one of the two variable areas of the genes. These areas include a 5′-end region from exons 1–3 and a 3′-end region spanning exons 18–21 (in rClC-2). The amino acid sequences encoded by these exons do not show any apparent homology between the two proteins. The sizes of the two exons in the 5′-end variable area are quite different. The difference between exon 1 in hClC-1 and rClC-2 occurs in the lengths of their coding sequences which are 180 bp in hClC-1 and 81 bp in rClC-2. This 5′-end variable region is also where DNA rearrangement occurred in rabbit ClC-2b (14). However, the truncated structure of rabbit ClC-2b cannot be easily explained by alternative splicing with the available human and rat genomic sequence data. Interestingly, the most variant exon size at the 3′-end is seen in exon 17, whereas the sizes of the exons flanking the 3′-end variable region, i.e., exons 18–22 in hClC-1, are either identical or highly similar to their counterparts in rClC-2. Since many members of the ClC family have been identified in multiple species, and the sequence similarity of a particular member between species (such as ClC-2 in rat and human) is higher than

that between members of an individual species (such as rat ClC-1 and ClC-2), it is likely that the ClC gene duplication occurred before human and rat developed distinct lineages.

These observations suggest that the two genes, which share a high similarity within a sub-branch of the ClC family (22), may have evolved by gene duplication, mutation and DNA rearrangement. Given the instability of exon 20 in rClC-2 (3), this exon in hClC-1 may have been skipped before or after a duplication from ClC-2, or alternatively, exon 20 in rClC-2 may have been an addition before or after a duplication from ClC-1. However, the AAG 3′ splice site, which is present in introns 13 and 19 in rClC-2, is not found in hClC-1. Instead, in hClC-1, TAG 3′ splice sites are found in both introns. The latter intron in hClC-1, intron 19, may be the counterpart of intron 20 in rClC-2 which has a TAG 3′ splice site.

ACKNOWLEDGEMENTS

We thank Dr Carol B. Murray for very helpful discussions on this project. We also thank Minzhi M. Liu for excellent technical assistance. This work was supported by R29 HL48274 to P.L.Z. and a Cystic Fibrosis Foundation Fellowship to S.C.

REFERENCES

1 Murray,C.B., Chu,S. and Zeitlin,P.L. (1996) *Am. J. Physiol*., **271**, 829–837.
2 Murray,C.B., Morales,M.M., Flotte,T.R., McGrath-Morrow,S.A., Guggino,W.B. and Zeitlin,P.L. (1995) *Am. J. Resp. Cell Mol. Biol*., **12**, 597–604.
3 Chu,S., Murray,C.B., Liu,M.M. and Zeitlin,P.L. (1996) *Nucleic Acids Res.*, **24**, 3453–3457.
4 Malinowska,D.H., Kupert,E.Y., Bahinski,A., Sherry,A.M. and Cuppoletti,J. (1995) *Am. J. Physiol*., **268**, 191–200.
5 Grunder,S., Thiemann,A., Pusch,M. and Jentsch,T.J. (1992) *Nature*, **360**, 759–762.
6 Thiemann,A., Grunder,S., Pusch,M. and Jentsch,T.J. (1992) *Nature*, **356**, 57–60.
7 Jordt,S.-E. and Jentsch,T.J. (1997) *EMBO J*., **16**, 1582–1592.
8 Schwiebert,E.M., Cid-Soto,L.P., Carter,M., Stafford,D., Guggino,W.B. and Cutting,G.R. (1996) *Pediatr. Pulmonology*, **13**, 236
9 Jentsch,T.J., Gunther,W., Pusch,M. and Schwappach,B. (1995) *J. Physiol*., **482**, 19-25.
10 Fisher,S.E., Black,G.C.M., Lloyd,S.E., Hatchwell,W., Wrong,O., Thakker,R.V. and Craig,I.W. (1994) *Hum. Mol. Genet*., **3**, 2053–2059.
11 George,A.L.J., Crackower,M.A., Abdalla,J.A., Hudson,A.J. and Ebers,G.C. (1993) *Nature Genet.*, **3**, 305–310.
12 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) *Science*, **257**, 797–800.
13 Lorenz,C., Meyer-Kleine,C., Steinmeyer,K., Koch,M.C. and Jentsch,T.J. (1994) *Hum. Mol. Genet*., **3**, 941–946.
14 Furukawa,T., Horikawa,S., Terai,T., Ogura,T., Katayama,Y. and Hiraoka,M. (1995) *FEBS Lett.*, **375**, 56–62.
15 Adachi,S., Uchida,S., Ito,H., Hata,M., Hiroe,M., Marumo,F. and Sasaki,S. (1994) *J. Biol. Chem*., **269**, 17677–17683.
16 Smoller,D.A., Petrov,D. and Hartl,D.L. (1991) *Chromosoma*, **100**, 487–494.
17 Hwang,T.C., Lu,L., Zeitlin,P.L., Gruenert,D.C., Huganir,R. and Guggino,W.B. (1989) *Science*, **244**, 1351–1353.
18 Li,M., McCann,J.D., Liedtke,C.M., Nairn,A.C., Greengard,P. and Welsh,M.J. (1988) *Nature*, **331**, 358–360.
19 Schoumacher,R.A., Shoemaker,R.L., Halm,D.R., Tallant,E.A., Wallace,R.W. and Frizzell,R.A. (1987) *Nature*, **330**, 752–754.
20 Cunningham,S.A., Awayda,M.S., Bubien,J.K., Ismailov,I.I., Arrate,M.P., Berdiev,B.K., Benos,D.J. and Fuller,C.M. (1995) *J. Biol. Chem.*, **270**, 31016–31026.
21 Fuller,C.M., Ismailov,I.I., Keeton,D.A. and Benos,D.J. (1994) *J. Biol. Chem*., **269**, 26642–26650.
22 Brandt,S. and Jentsch,T.J. (1995) *FEBS Lett.*, **377**, 15–20.
23 Kawasaki,M., Uchida,S., Monkawa,T., Miyawaki,A., Mikoshiba,K., Marumo,F. and Sasaki,S. (1994) *Neuron*, **12**, 597–604.
24 Steinmeyer,K., Schwappach,B., Bens,M., Vandewalle,A. and Jentsch,T.J. (1995) *J. Biol. Chem*., **270**, 31172–31177.
25 Shapiro,M.B. and Senapathy,P. (1987) *Nucleic Acids Res*., **15**, 7155–7174.
26 Himmler,A. (1989) *Mol. Cell. Biol.*, **9**, 1389–1396.
27 Mizuno,T., Iwashina,M., Itakura,M., Hagiwara,H. and Hirose,S. (1993) *J. Biol. Chem*., **268**, 5162–5167.
28 Holzl,B., Huber,R., Paulweber,B., Patsch,J.R. and Sandhofer,F. (1994) *J. Lipid Res*., **35**, 2161–2169.
29 Wong,C., Antonarakis,S.E., Goff,S.C., Orkin,S.H., Forget,B.G., Nathan,D.G., Giardina,P.J. and Kazazian,H.H.J. (1989) *Blood*, **73**, 914–918.
30 Recio,L., Simpson,D., Cochrane,J., Liber,H. and Skopek,T.R. (1990) *Mutat. Res*., **243**, 195–208.
31 Smith,C.W., Chu,T.T. and Nadal-Ginard,B. (1993) *Mol. Cell. Biol*., **13**, 4939–4952.
32 Goux-Pelletan,M., Libri,D., d'Aubenton-Carafa,Y., Fiszman,M., Brody,E. and Marie,J. (1990) *EMBO J.*, **9**, 241–249.
33 Muhlemann,O., Kreivi,J.P. and Akusjarvi,G. (1995) *J. Virol*., **69**, 7324–7327.
34 Middleton,R.E., Pheasant,D.J. and Miller,C. (1996) *Nature*, **383**, 337–340.
35 Ludewig,U., Pusch,M. and Jentsch,T.J. (1996) *Nature*, **383**, 340–343.
36 Fahlke,C., Knittle,T., Gurnett,C.A., Campbell,K.P. and George,A.L.J. (1997) *J. Gen. Physiol*., **109**, 93–104.
37 Goedert,M., Spillantini,M.G., Potier,M.C., Ulrich,J. and Crowther,R.A. (1989) *EMBO J*., **8**, 393–399.
38 Shire,D., Carillon,C., Kaghad,M., Calandra,B., Rinaldi-Carmona,M., Le Fur,G., Caput,D. and Ferrara,P. (1995) *J. Biol. Chem*., **270**, 3726–3731.
39 Treisman,R., Orkin,S.H. and Maniatis,T. (1983) *Nature*, **302**, 591–596.
40 Petrukhin,K., Lutsenko,S., Chernov,I., Ross,B.M., Kaplan,J.H. and Gilliam,T.C. (1994) *Hum. Mol. Genet*., **3**, 1647–1656.
41 Ren,H. and Stiles,G.L. (1994) *J. Biol. Chem*., **269**, 3104–3110.
42 Popielarz,M., Cavaloc,Y., Mattei,M.G., Gattoni,R. and Stevenin,J. (1995) *J. Biol. Chem*., **270**, 17830–17835.
43 Whiting,P., McKernan,R.M. and Iversen,L.L. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9966–9970.
44 Sommer,B., Keinanen,K., Verdoorn,T.A., Wisden,W., Burnashev,N., Herb,A., Kohler,M., Takagi,T., Sakmann,B. and Seeburg,P.H. (1990) *Science*, **249**, 1580–1585.
45 Stamm,S., Casper,D., Dinsmore,J., Kaufmann,C.A., Brosius,J. and Helfman,D.M. (1992) *Nucleic Acids Res.*, **20**, 5097–5103.