Full Articles

Structure-functional organization of mouse aquaporin 4 gene

A. A. Bondar,[a]★ T. Yu. Alikina,[a] M. N. Zelenina,[b,c] and S. M. Zelenin[a,c]

[a]Novosibirsk Institute of Bioorganic Chemistry, Siberian Branch of the Russian Academy of Sciences, 8 prosp. Akad. Lavrent´eva, 630090 Novosibirsk, Russian Federation, Fax: +7 (383 2) 33 3677. E-mail: alex_bondar@online.sinor.ru
[b]Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 10 prosp. Lavrent´eva, 630090 Novosibirsk, Russian Federation. Fax: +7 (383 2) 34 2474. E-mail: marze@mail.ru
[c]Karolinska Institute, Q2:09 Astrid Lindgren Children´s Hospital, S-17176 Stockholm, Sweden. Fax: +46 8 51777328. E-mail: sergey.zelenin@ks.se*

Full-size mouse aquaporin 4 gene (AQP4) was cloned and a fragment (4964 base pairs, AF219992) covering promoter 0 and exons 0 and 1 was sequenced. The mouse AQP4 gene contains a new exon X between exons 0 and 1. Three forms of mouse AQP4 mRNA were identified. The new AQP4.M23X mRNA has a transcription start in the exon X and encodes the known M23 isoform of AQP4. Using AQP4.M1, AQP4.M23X, and AQP4.M23 cDNAs, recombinant DNA expressing M1 and M23 AQP4 isoforms fused with the GFP fluorescent protein were constructed. In the renal epithelial cell line LLC-PK1, all these fused proteins were localized in the basolateral plasma membrane and demonstrated distinct water channel properties. The obtained data on the structure-functional organization of the mouse AQP4 gene allow more detailed investigation of the mechanisms of the control of its expression at the level of individual mRNAs, which appear to be controlled by different, probably independently functioning, promoters of the AQP4 gene.

Key words: aquaporin 4 (AQP4, MIWC), gene cloning, mRNA, water channel, mouse, brain, lungs, kidney, water homeostasis.

Regulation of the water permeability of the cell membranes is of utmost importance for water and salt homeostasis of the body. Water permeability is extremely high in epithelia where rapid secretion or reabsorption of big volumes of fluid occur, and is restricted in cell types that play barrier role for the fluid movement. The lipid bilayer of the cellular membrane has very limited permeability for water. To increase the permeability dramatically, the cells produce specific protein water channels, aquaporins (AQPs), and incorporate them into the cell membrane.[1,2] These proteins form pores in the lipid bi-

* Karolinska Institutet, Q2:09 Astrid Lindgren Barnsjukhus KS, S-17176 Stockholm, Sweden.

Published in Russian in *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 7, pp. 1035—1046, July, 2002.

1066-5285/02/5107-1122 $27.00

layer, which allow fast migration of water molecules in or out of the cell, depending on the osmotic gradient. To date eleven different AQPs were cloned from mammalian tissues.[2] All of them have highly specific tissue and subcellular distribution indicating that they have distinct roles in different epithelial cell types. Disorders in aquaporin distribution, properties, and expression regulation may be responsible for many known diseases and pathological states of an organism such as neonatal lung dysfunction, hypertension, some types of cardiac decompensation, cirrhosis, diabetes insipidus, brain edema, *etc*.[3,4] Understanding of the molecular mechanisms of regulation, structure, and principles of functioning of aquaporins both at the level of water channel protein and at the level of gene expression may potentially bring researchers to the development of new tools for the therapy of various water homeostasis disorders in children and adults.

Since aquaporin 4 (AQP4) is quite a unique representative of its family, some consideration should be given to the role of this water channel in organism tissues. The AQP4 (which is also called mercurial insensitive water channel, MIWC) is a water channel whose gene is mainly expressed in kidneys, brain, lungs, and skeletal muscles.[5,6]

In lungs, AQP4 is localized in the basolateral membranes of airway epithelium.[7] Presumably, it plays a key role in the rapid clearance of water from the lung immediately after birth.[8] Later in life AQP4 is likely to participate in regulation of the airway hydration, which is necessary for effective gas exchange in lungs.

In skeletal muscles, the AQP4 gene is expressed in the cells of muscle fast fibers and is supposed to play an important role in maintenance of the muscle cell homeostasis during muscle contraction.[9]

In kidney, AQP4 is localized in the basolateral membranes of collecting duct principal cells. There AQP4, possessing the highest permeability for water molecules among proteins of the aquaporin family,[2,10] provides the main pathway for water that has been reabsorbed from collecting duct lumen by principal cells and is forwarded to the interstitium of kidney and then back to blood vessels.[11]

In brain, AQP4 is expressed in glial cells lining the brain ventricular system and the pial surface, and in cerebellum. Aquaporin 4 is assumed to play an important role in the formation of water fluxes that allow buffering of potassium, which is released from neurons during neuronal activity.[12,13] It was also shown that AQP4 is abundant in the osmosensory regions including the supraoptical nucleus and the subfornical organ. This specific distribution of aquaporin 4 may indicate that this protein plays an important role for both the central regulation of the water balance of an organism and the regulation of the water homeostasis of brain. The AQP4 gene expression is known to increase substantially during brain edema that usually occurs after brain trauma or surgery.[14] It is also upregulated in brain in response to ischemic insult and systemic hyponatremia.[15,16] In these pathological states, the blockade of extensive movement through this water channel or a decrease in the AQP4 gene expression would potentially help to prevent the brain edema[4] or, for example, would also play a positive role in kidneys in patients having syndromes of inappropriate water retention, as it would potentially help to decrease excessive water reabsorption. Conversely, it would be helpful to stimulate AQP4 in lungs to resolve lung edema that often occurs in newborn or premature babies having a neonatal lung dysfunction or inflammatory diseases such as pneumonia or bronchitis. Besides, stimulated upregulation of the AQP4 gene expression would potentially improve the maintenance of the water balance in the skeletal muscles of dystrophic patients, since it has been demonstrated that AQP4 expression is decreased in fast twitch fibers during this pathological state.[9] It is therefore highly significant to study the structure-functional organization of the AQP4 gene, the structure and functioning of its promoter regions, and the mechanisms and the pathways of both negative and positive gene expression regulation.

It has been considered so far that the mouse aquaporin 4 gene consists of five exons (0, 1, 2, 3, and 4) separated by four introns. The results obtained in our work allow one to claim that a new exon exists in the 5′-region of the mouse AQP4 gene between exons 0 and 1. Within the framework of this study, we cloned the full-size mouse aquaporin 4 gene including ~8500 base pairs (bp) of the 5′-flanking region (ahead of exon 0) and ~7000 bp of the 3′-flanking region (after exon 4). Gene mapping by restriction endonucleases was carried out and positions of exons 0, 1, 2, 3, and 4 were determined. The nucleotide sequence of the fragment covering exons 0 and 1 of the mouse AQP4 gene and the region between them, promoters 0 and 1 (4964 bp, GenBank entry number AF219992) was determined. Using RT-PCR, sequencing, computer analysis, the primer extension technique, and comparison of the mouse AQP4 gene and three cDNA sequences we determined with mRNA nucleotide sequences (M1, MIWC1, MIWC2, MIWC3) known for the mouse AQP4, we found a new exon X, from which previously unknown AQP4.M23X mRNA is transcribed and encodes the known isoform of mouse aquaporin 4, namely, M23. New data on the size and boundaries of exon 0 were obtained. By the use of confocal microscopy and the recombinant DNA technique for preparation of aquaporin isoforms fused with the green fluorescent protein (GFP), we demonstrated that mouse aquaporin 4 isoforms expressed from AQP4.M23X, AQP4.M23, or AQP4.M1 cDNA are located in the basolateral plasma membrane of transfected pig renal

epithelial cells of the LLC-PK1 line and function as water channels.

Experimental

Cloning of the mouse aquaporin 4 gene. From the mouse genomic library based on lambda bacteriophage (Stratagene), kindly provided by Professor Anita Aperia,** using hybridization with the PCR fragment including 1600 bp and corresponding to the intron between exons 1 and 2 of the mouse AQP4 gene, we selected two clones covering a mouse genomic DNA fragment including 26500 bp and containing all known exons of the AQP4 gene. Using partial hydrolysis of the DNA of these clones by the restriction endonucleases XbaI, SacI, HindIII, and BamHI, specific probe hybridization, and PCR analysis, the exon—intron and restriction gene maps were constructed (Fig. 1). The inserts of the selected λ-clones were subcloned in pBlueScript II series vectors (Stratagene, USA) as a set of fragments completely covering the exon—intron region, ~8500 bp of the 5′-flanking region from exon 0, and 7000 bp of the 3′-flanking region from exon 4. The *E.coli* XL-1 Blue, XL-2 Blue, and SCS-110 strains were used as cloning hosts, and standard procedures described in the literature were employed.[17] The plasmid double-stranded DNA was isolated using the FlexiPrep™ Kit (Amersham Pharmacia Biotech) or by boiling.[18,19] Single-stranded DNA was obtained using phenol extraction.[20]

Sequencing of the mouse aquaporin 4 gene. The set of clones covering the mouse AQP4 gene was obtained by the strategy of unidirected deletions with a predictable size (400—500 bp) using the pBluescript® II Exo/Mung DNA Sequencing System (Stratagene, USA).[20] Sequencing of both DNA strands of the mouse AQP4 gene (4964 bp, GenBank entry AF219992) was done by the Sanger method, as in the classical radioactive version, using the Sequenase Version 2.0 DNA Sequencing Kit (Amersham, UK) and the Thermo Sequenase Cycle Sequencing Kit (Amersham, UK), and also using the ABI Prism Big Dye Terminator Ready Reaction Sequencing Kit (Perkin—Elmer, USA) on an automated DNA sequencer (ABI310, Applied Biosystems), according to the manufacturer protocols. In addition to standard sequencing of the plasmid DNA recommended by the manufacturer, we used 0.25 (2 μL) of the advised amount of Big Dye Terminator Ready Reaction Mix, 0.5 pmol of the DNA matrix, and 0.9 pmol of the primer for sequencing. The sequencing was performed on a Perkin—Elmer 9700 instrument using a temperature profile that included warming-up for 20 s at 96 °C, two cycles including 10 s at 96 °C and 4 min at 64 °C; four cycles including 10 s at 96 °C and 4 min at 60 °C; and twelve cycles including 10 s at 96 °C, 7 s at 50 °C, and 4 min at 60 °C. The products of sequencing reactions were purified from the remaining fluorescence-labeled ddNTP by precipitating with 60% isopropyl alcohol.

Preparation and sequencing of the RT-PCR products. The total RNA was isolated from the brain and lungs of adult mice using phenolic extraction[17] or the RNeasy Total RNA Kit (QIAGEN, Germany). The PCR and reverse transcription were carried out by a previously reported procedure with some modifications.[21] The PCR fragments were separated by electrophoresis in an agarose gel, extracted using the QIAquick Gel Extraction Kit (QIAGEN, Germany), and sequenced as described in the previous section. The products of sequencing were purified from the remaining fluorescence-labeled ddNTP either by gel filtration through columns (Centricep, USA) or by precipitation with 60% isopropyl alcohol with addition of 1—1.5 μg of high-purity pBlueScript II plasmid DNA as the carrier. The PCR products on the cDNA matrix were prepared using the following combinations of sense primers:

AQ4.CF1 5′-TTATGGTTCACGGGTTTGGATG-3′;
AQP4.CFN 5′-GCTTGGTCCCCTTCTCTTTC-3′;
AQP4.001 5′-GGAAGGCTAGGTTGGTGACTTC-3′;
AQP4.1AGFN 5′-GCCACATGGTGCAGAATCTTTC-3′;
AQP4.2AGFN 5′-GTGCCCGTAATCTGACTCCCAG-3′;
AQP4.AGF1 5′-GCCAGGGAAGGCATGAGTGAC-3′;
AQP4.193 5′-GGTGGCTCAGAAAACCCCTTAC-3′

and antisense primers:

AQP4.460 5′-TGGTGACTCCCAATCCTCCAAC-3′;
AQP4.987
5′-CCTCTAGTCATACGGAAGACAATACCTCTC-3′.

Mapping of the mRNA transcription start sites by the primer extension. The start sites for mRNA transcription were mapped as described previously.[22] The transcription initiation site for AQP4.M1 mRNA in exon 0 was determined by using [γ-^{32}P]ATP endlabeled antisense oligonucleotide primer AQP4.AGBPE2 (5′-CTGGGAGTCAGATTACGGGCAC-3′). The labeled primer (1 pmol) was hybridized with 3 μg of the total RNA from the mouse brain and then subjected to reverse transcription. The initiation site of transcription for AQP4.M23X mRNA in exon X was identified in a similar way using the radioactively labeled antisense oligonucleotide primer, AQP4.CBN (5′-CCCTCTCCAGAGACCTAATC-3′). The reverse transcription products were analyzed by electrophoresis in a 4—6% denaturing polyacrylamide gel in the presence of 7 *M* urea. The products of sequencing of the cloned gene fragment with the corresponding primer were used as markers.

Computer analysis of nucleotide sequences. The nucleotide sequences were analyzed using DNASIS program package for DOS (Hitachi), DNASTAR program package (Lasergene), and the software available through the Internet. The translational properties of the mouse AQP4 mRNA were estimated using the expert system at the server of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences http://wwwmgs.bionet.nsc.ru/mgs/gnw/leader/.

Preparation of recombinant of DNA constructs expressing the mouse AQP4 fused with the green fluorescent protein GFP. The DNA copy covering the 5′-nontranslation and protein-encoding regions of the corresponding mouse AQP4 mRNA form was cloned into the reading frame for the green fluorescent protein in the pEGFP-N1 vector (Clontech) under the control of the CMV promoter. The stop codon was removed by site-directed mutagenesis using the USE Mutagenesis Kit (Pharmacia) with simultaneous introduction of an additional PstI restriction site, instead of the stop codon for the subsequent clone selection. The structures of all DNAs obtained at successive stages were confirmed by sequencing. The obtained recombinant DNA was used for transient transfection of pig renal epithelial cell line LLC-PK1 using the CLONfectin reagent (Clontech), according to the manufacturer protocol. Aquaporin 4 fused with the GFP protein at the C-end was

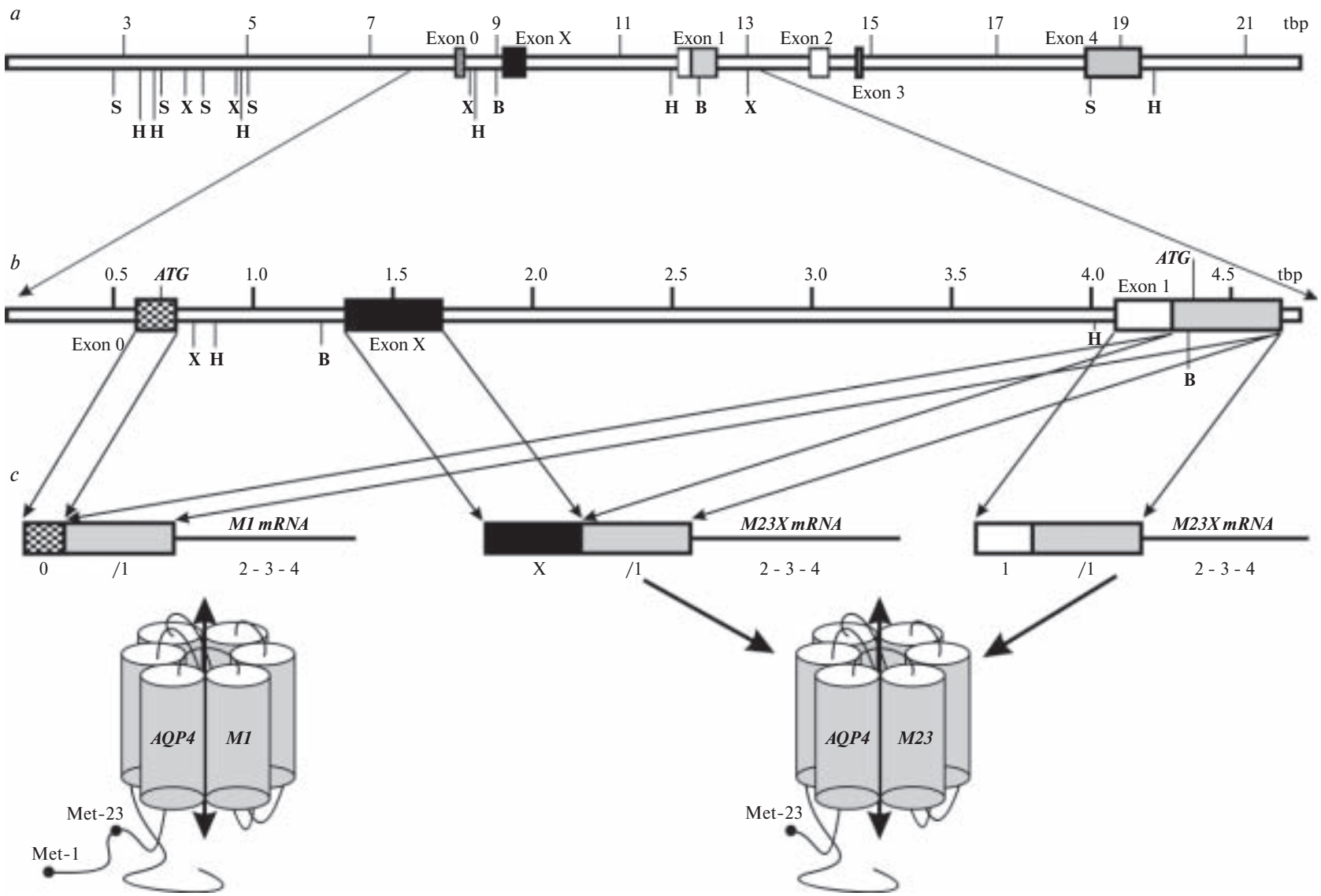

Fig. 1. Structure organization of the mouse aquaporin 4 gene. *a*. Cloned section of the mouse genomic DNA. The positions of exons of the AQP4 gene are shown by rectangles. Designations of the restriction endonuclease sites: S is SacI, X is XbaI, H is HindIII, B is BamHI. The distances are given in thousands of base pairs (tpb). *b*. Map of the fragment of the mouse AQP4 gene sequenced in this study (GenBank AF219992). In exon 1, the alternative splicing site is marked by a boundary between white and gray rectangles. The translation start sites of the AQP4 M1 and M23 isoforms are designated by ATG. *c*. Scheme of the formation of matrix RNA encoding the M1 and M23 isoforms of mouse aquaporin 4. mRNA sections corresponding to exons 0, X, and 1 are shown by rectangles. The general mRNA part transcribed from exons 2—3—4 is shown by a line.

detected by a laser scanning confocal microscope. For the expression of AQP4 fused with the GFP protein at the N-end, the protein-encoding part of the cDNA AQP4 was cloned in a similar way in the GFP reading frame to the pEGFP-C2 vector (Clontech).

Results and Discussion

Although the tissues and organs in which aquaporin 4 is expressed are known quite well, the problem of structure-functional organization of the AQP4 gene has not yet been completely clarified. At least, two aquaporin 4 isoforms exist, called M1 and M23 in accordance with the methionine position at the N-end of the protein.[23] In cell membrane, AQP4 forms heterotetramers of water channels from the M1 and M23 monomers.[6] Presumably, the 22 amino acid sequence at the N-terminus, which consistutes the difference between them, has no influence on the water permeability of the channel but may contribute to membrane trafficking or assembly of heterotetramers. Previously, it was shown that the M1 and M23 isoforms of mouse AQP4 are encoded by, at least, two different mRNAs.[4,23] The M23 isoform is encoded by the mRNA transcribed from exons 1, 2, 3, and 4. The transcription of the mouse M1 AQP4 mRNA starts from an alternative site remote by more than 3500 bp upstream and thus contains an additional nucleotide sequence corresponding to exon 0, partially spliced exon 1, and exons 2—4. However, data on the nucleotide sequences of the AQP4 mRNA and the number of isoforms of the AQP4 protein are rather contradictory.[5,23] The nucleotide sequence of the mouse AQP4 4 gene is known only partially, mainly in the regions of exons 1, 2, 3, and 4.[5] Yang and Verkman[5] have shown the presence of three mouse aquaporin 4 mRNAs called MIWC1 (corresponds to M23), MIWC2 (corresponds to M1), and MIWC3 and suggested that the MIWC3 mRNA encodes a third form of mouse AQP4. Like us, Turdzo *et al.*[23] were also unable to detect the MIWC3 mRNA. They cloned the mRNA containing open reading frames starting from mouse aquaporin 4 Met-1 and Met-23. In the present study, we did find the third mRNA (AQP4.M23X), which, however, differed significantly from the MIWC3 mRNA. For an average size of the mouse AQP4 mRNA cloned by different researchers of not more than 1700 bp, northern-blot analysis revealed also transcripts with sizes of about 3000 and 5500 bp.[5,23] Regarding other mammals, it is interesting to notice a recent study[24] reporting cloning of two AQP4 cDNAs from the bovine brain gene library, namely, bAQP4-A mRNA (5700 bp long, homologous to the human M1 mRNA[22] and mouse MIWC2 mRNA[5]) and bAQP4-B mRNA (5641 bp long, homologous to the mouse MIWC3 mRNA[5]). However, Sobue *et al.*[24] were unable to obtain a single clone homologous to the mouse MIWC1 mRNA[5] or the human M23 mRNA,[22] which is encountered most often among the known AQP4 mRNA forms. Recently, we studied expression of the new mRNA form of mouse aquaporin 4, AQP4.M23X, found in this work, in brain, kidney, and lungs. It was found that AQP4.M23X mRNA, which encodes a known protein isoform, M23, is expressed mainly in brain, whereas the known AQP4.M23 mRNA, encoding the same M23 isoform of mouse aquaporin 4 is expressed predominantly in kidney and lungs. It was also found that the expression level of AQP4.M23X mRNA is developmentally regulated.[25]

The map of the mouse AQP4 gene we cloned with the relative positions of exons and some restriction endonuclease sites is shown in Fig. 1. The nucleotide sequence of the mouse AQP4 gene fragment covering exons 0 and 1 and the region between them, promoters 0 and 1 (4964 bp, GenBank entry AF219992) is sketched in Fig. 1, *b* and shown in more detail in Fig. 2. Comparative analysis of the determined nucleotide sequence of the mouse aquaporin 4 gene with all known nucleotide sequences of the mouse AQP4 mRNA (M1,[23] MIWC2,[5] MIWC3,[5] MIWC1 [5]) revealed several highly homologous regions. Two of them corresponded to exons 0 and 1. The third one was a homologous cluster with a deletion at the center located at a distance of ~1000 bp downstream from exon 0. We suggested that a new exon that we called X may be located in this gene region between exons 0 and 1. To verify this, we isolated the total RNA from the mouse brain and performed PCR analysis with combinations of sense and antisense primers from various gene exons using the total cDNA as the template and the total RNA as the control. We were able to identify three different forms of the mouse aquaporin 4 mRNA (Fig. 3) and to determine their nucleotide sequences. According to the accepted nomenclature, we designated two of them by AQP4.M1 (encodes the M1 isoform) and AQP4.M23 (encodes the M23 isoform). The new mRNA form was called AQP4.M23X, in accordance with the name of exon X, in which transcription starts, and the protein isoform encoded by the given mRNA. According to our results, AQP4.M23X is not a case of alternative splicing of AQP4.M1 mRNA (see Fig. 1, *c*), because none of the RT-PCR reactions using pairs of primers from exon 0 and exon X regions, respectively, gave amplification products. In addition, none of the AQP4.M1 cDNA fragments obtained by RT-PCR using combinations of pairs of primers from exon 0 and exon 1—4 regions contained sections homologous to exon X in the nucleotide sequence. The scheme of formation of mRNA encoding the M1 and M23 isoforms of mouse aquaporin 4 that we identified is shown in Fig. 1, *c*.

Using the primer extension method, we determined the initiation sites of transcription for the mouse AQP4.M1 and AQP4.M23X mRNAs (Fig. 4). The re-

Fig. 2 (*to be continued*)

Fig. 2. Nucleotide sequence of the mouse AQP4 gene fragment (GenBank AF219992) and positions of specific oligonucleotide primers. Exons 0, X, and 1 are framed; the exon X 5′-boundary is indicated according to the data of comparison with the MIWC3 mRNA.[5] The transcription start sites for the AQP4.M1 and AQP4.M23X mRNA were determined by the primer extension technique. The ATG codons, which are the start sites of translation for the AQP4 M1 and M23 isoforms are typed in bold and marked as Met-1 and Met-23.

sults of mapping of the transcription start sites are marked on the nucleotide sequence of the gene (see Fig. 2). According to the results obtained, the transcription of AQP4.M1 mRNA starts in exon 0 from nucleotide G, 127 bp upstream from the translation initiation site of the M1 isoform (see Fig. 2 and 4, *b*). The AQP4.M23X mRNA transcription starts in exon X from nucleotide T separated by 3018 bp from the site of translation initia-

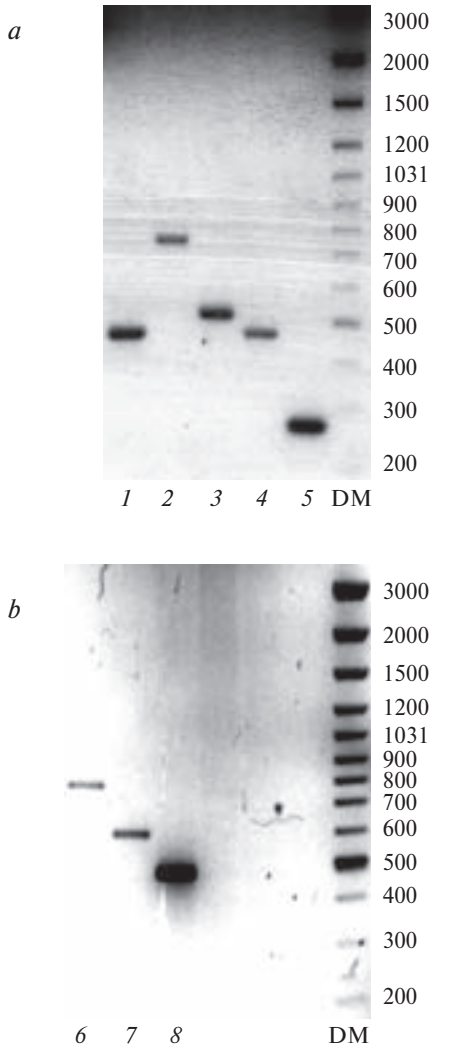

Fig. 3. Electrophoresis pattern for the mouse AQP4 cDNA fragments obtained using the RT-PCR method. The fragments were obtained using the common AQP4.460 antisense primer in combination with the sense primers from different exons of the mouse AQP4 gene. The structure and positions of primers on the gene sequence are presented in Fig. 2 and in Experimental. *a*. Track *1*: M1 cDNA from the mouse brain, AQP4.AGF1 sense primer, corresponds to exon 0 and a part of exon 1. Track *2*: M23X cDNA from the mouse brain, AQP4.CF1 sense primer, corresponds to exon X and a part of exon 1. Track *3*: M23X cDNA from the mouse brain, AQP4.CFN sense primer, corresponds to exon X and a part of exon 1. Track *4*: M23 cDNA from the mouse brain, AQP4.001 sense primer, corresponds to a part of exon 1. Track *5*: M23 cDNA from the mouse brain, AQP4.193 sense primer, corresponds to a part of exon 1. *b*. Track *6*: M23X cDNA from the mouse brain obtained using the AQP4.CF1 sense primer, corresponds to exon X and a part of exon 1. Track *7*: M1 cDNA from the mouse brain obtained using the AQP4.1AGFN sense primer, corresponds to exon 0 and a part of exon 1. Track *8*: M23 cDNA from the mouse lungs obtained using the AQP4.001 sense primer, corresponds to exon 1. Track DM: a DNA length marker, GeneRuler 100bp DNA Ladder Plus (MBI Fermentas, Lithuania).

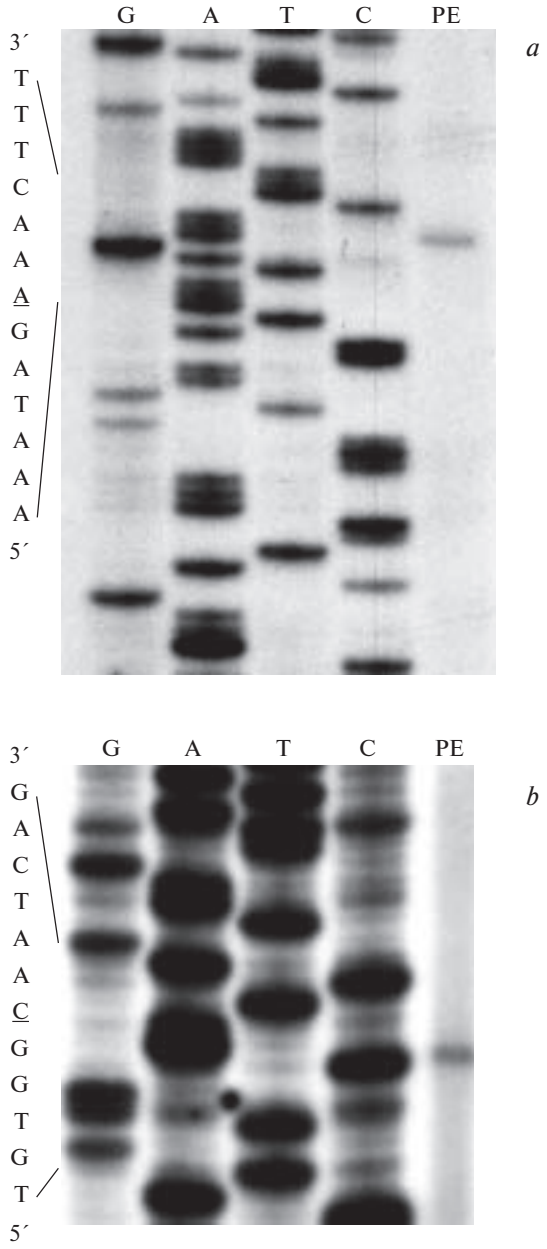

Fig. 4. Mapping of the transcription initiation sites for the M23X and M1 mRNAs of mouse AQP4 by the primer extension technique. Start of transcription for the M23X mRNA (*a*) or M1 mRNA (*b*). ^{32}P labeled AQP4.CBN primer (*a*) or AQP4.AGBPE2 primer (*b*) was hybridized with the total RNA from the mouse brain and reverse transcription was carried out. The product (PE track) was analyzed using electrophoresis in a 6% PAAG—7 *M* urea (*a*) or in a 4% PAAG—7 *M* urea (*b*). The products of sequencing of the cloned gene fragment with the same primer were used as markers (GATC tracks, the antisense sequence is shown) .

tion of the M23 isoform located in exon 1 (see Figs. 2 and 4, *a*). It is noteworthy that our data on mapping of the transcription start site for the AQP4.M23X mRNA

did not coincide with the 5′-end of exon X (3035 bp upstream from the ATG codon Met-23) identified by comparison with the known MIWC3 mRNA [5] (see Fig. 2). Apparently, to clarify this point, additional analysis by other methods is required. We did not carry out mapping of the transcription initiation sites for the AQP4.M23 mRNA, as these data had already been published for MIWC1,[5] and the sequence of the 5′-region of the MIWC1 mRNA was in good agreement with our data on the gene structure.

Using the results obtained (see Figs. 2 and 4) and resorting to published data, we selected appropriate primers, carried out PCR, and applied direct sequencing of the RT-PCR products to determine the nucleotide sequences of the AQP4.M23X (GenBank entry AF469169) and AQP4.M1 (GenBank entry AF469168) mRNAs, covering completely the 5′-nontranslatable and protein-encoding regions. For the AQP4.M23 mRNA, the nucleotide sequence for the fragment from the AQP4.001 primer (see Fig. 2, exon 1) to the end of the protein-encoding mRNA region in exon 4 (primer AQP4.987) was determined.

Analysis of the resulting AQP4.M1, AQP4.M23X, and AQP4.M23 cDNA structures showed that they were fully identical to the nucleotide sequences of exons 0, X, and 1 of the mouse AQP4 gene we determined (Fig. 5). In the region of exons 2—4, the nucleotide sequence of the protein-encoding part of all the AQP4 mRNA forms we determined was identical and coincided with the previous data[23] for the mouse M1 mRNA. Conversely, the first 32 bp of the 5′-end region of the mouse M1 mRNA[23] have no homology with the sequence of our AQP4.M1 mRNA and, correspondingly, of AQP4 exon 0 and represent, very likely, the remains of the vector. As a consequence, the 5′-nontranslation region of the AQP4.M1 mRNA proved to be longer by 117 nucleotides than it is known for MIWC2 mRNA,[5] or longer by 105 bp (if the nonhomologous 32 nucleotides are rejected) than it is known for the M1 [23] mRNA (Fig. 5, *a*). In the region of exon 0 in the M1 [23] and MIWC2 [5] mRNAs, we found two silent nucleotide substitutions and one causing replacement of Gly-4 by Arg-4, which is a known case of polymorphism.[23]

The nucleotide sequence of the AQP4.M23X mRNA in the region of exon X generally coincides with the MIWC3 [5] mRNA, except for the large deletion (see Fig. 5, *b*). Numerous attempts to obtain the RT-PCR fragment fully identical to the corresponding region of MIWC3 and, hence, to confirm the possible presence of an alternative splicing for this mRNA form failed. According to our data, the AQP4.M23X mRNA contains an inset of 206 bp corresponding to the central region of exon X, compared to the MIWC3 [5] mRNA (see Fig. 5, *b*). The splicing sites for the AQP4.M23X and MIWC3 [5] mRNAs between exons X/1 and the splicing sites for the AQP4.M1, M1,[23] and MIWC2 [5] mRNAs between exons 0/1 proved to be identical and located within exon 1 (see Fig. 5). Since MIWC3 was assumed[5] to encode the third isoform of mouse AQP4 4, we analyzed the open reading frames (ORF) and discovered numerous stop codons in all three ORFs in the region corresponding to a deletion of 206 bp in the MIWC3 [5] mRNA.

Analysis of the nucleotide sequence of the AQP4.M23X mRNA showed that the first ATG codon located in the optimal context for translation initiation encodes Met-23 of the mouse AQP4. This means that the third isoform of mouse AQP4, other than M1 and M23 and similar to MIWC3 cannot be translated from the AQP4.M23X mRNA.[5] Recently, we obtained interesting results in a study of tissue- and age-specific expression of the AQP4.M23X and AQP4.M23 mRNAs, which encode, according to the above-described data, the same M23 isoform of mouse AQP4.[25] Using semiquantitative RT-PCR analysis with β-actin as the internal standard, we investigated brain, kidney, and lung tissues of two-days-old and adult mice.[25] Expression of the new AQP4.M23X mRNA took place mainly in brain rather than in kidneys and lungs and was developmentally regulated. The level of AQP4.M23X mRNA expression in the adult mouse brain was almost twice higher than that for the two-days-old one. Conversely, the M23X mRNA expression levels in the kidney and lungs were low in both adult and two-days-old mice. The AQP4.M23 mRNA described previously was mainly expressed in kidneys and lungs. In lungs, expression of this form was higher in infancy than in adulthood.[25]

Using the cDNA obtained from the mouse AQP4.M1, AQP4.M23X, and AQP4.M23 mRNAs identified in the present study, we constructed the recombinant DNA for the expression of the M1 and M23 isoforms of mouse AQP4 fused with the GFP fluorescent protein. The DNA constructs obtained were used for short-term expression in the LLC-PK1 line cells. By use of confocal microscopy and the technique described in our previous study,[26] we showed that all the GFP—AQP4 fused proteins are characterized by AQP4-specific basolateral distribution in the plasma membranes of the LLC-PK1 line cells and possess properties of a water channel.[27] Thus, we demonstrated that all the mRNA forms of the mouse AQP4 we identified having transcription start sites in different gene exons (0, X, and 1) encode functionally active isoforms of mouse aquaporin 4.

The question of organization of promoters 0 and 1 of the AQP4 gene regulating the expression of mRNA of various isoforms of this water channel, like the question of the presence and location in the 5′-flanking exon X region of functionally active regulatory elements activated by particular transcription factors still remains open. New exons might also be discovered in the 3′-region of the mouse AQP4 gene. Usually, investigators clone the

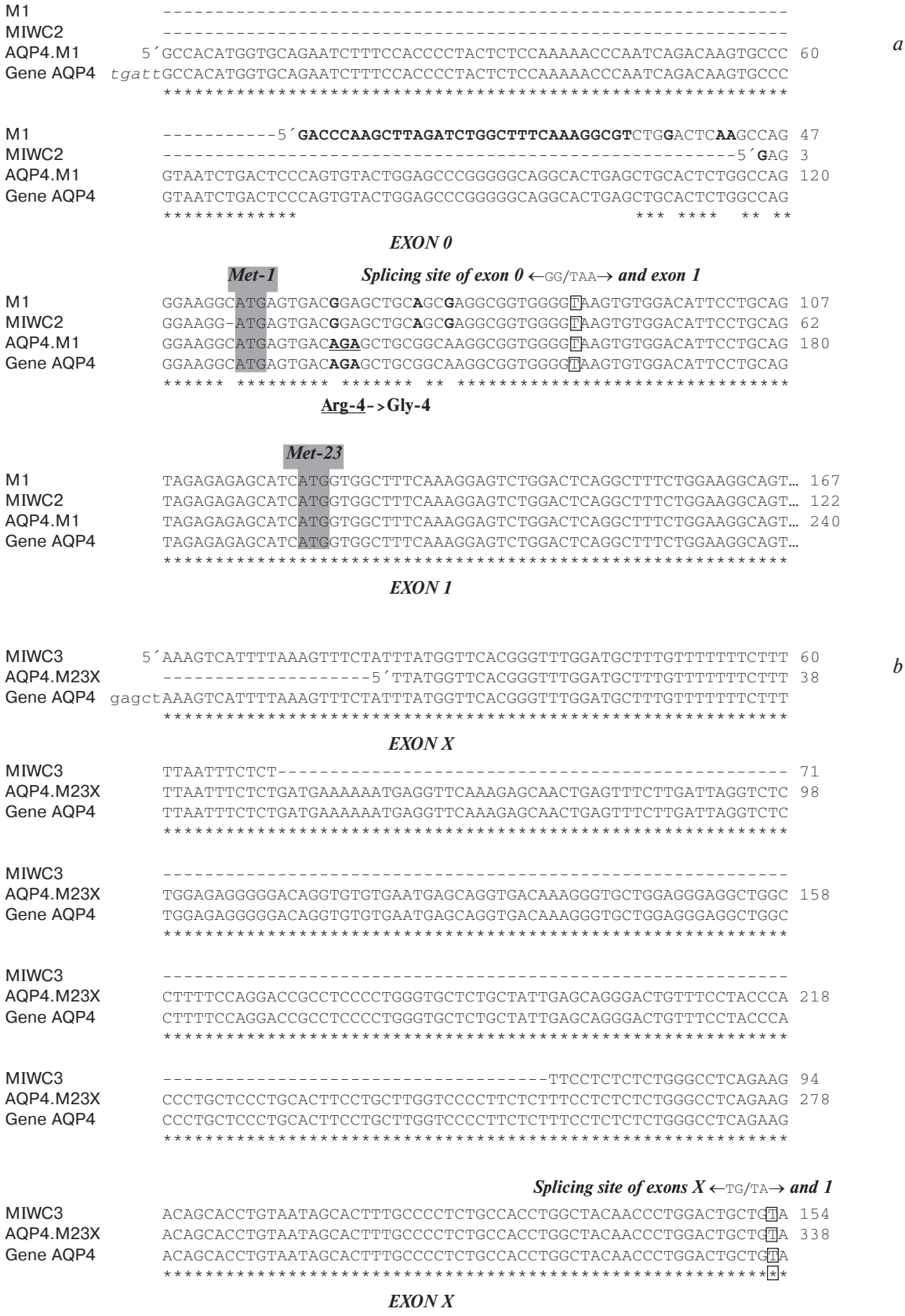

Fig. 5 (*to be continued*)

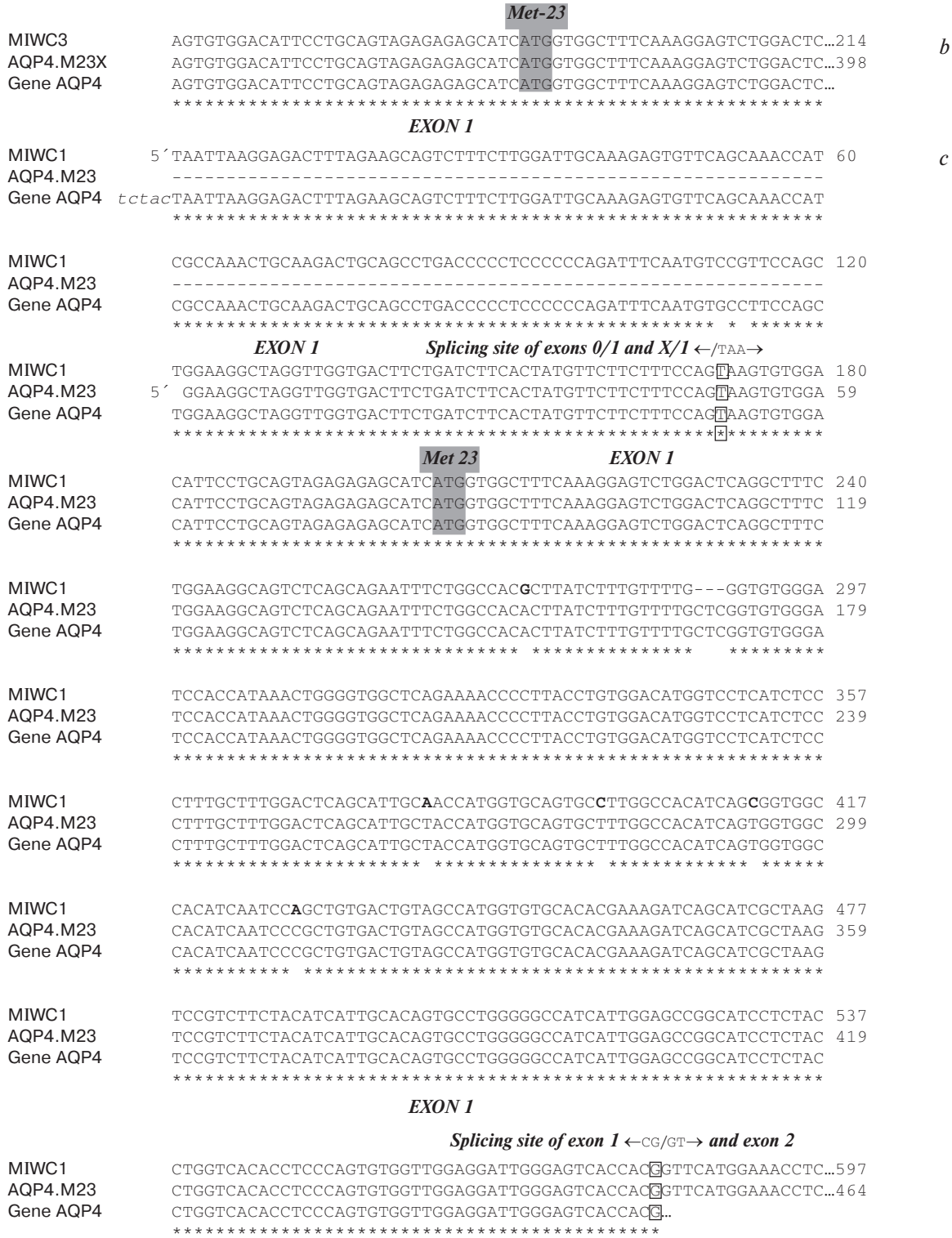

Fig. 5. Comparison of the nucleotide sequences of the 5´-end region of sequenced mouse AQP4.M1 (*a*), AQP4.M23X (*b*), and AQP4.M23 (*c*) mRNAs with previously published mouse M1 (*a*),[23] MIWC1 (*c*),[5] MIWC2 (*a*),[5] and MIWC3 (*b*)[5] mRNA sequences. The cDNA sequences are given. The sequence of the corresponding exons of the mouse AQP4 gene is given. In the region of exons 2—4 of the mouse AQP4 mRNA, our data completely coincided with the results published previously;[23] therefore, the comparison is not given. The missing nucleotides are designated by dashes. Nucleotides differing from the AQP4.M1, AQP4.M23X, AQP4.M23 sequence are typed in bold. Identical nucleotides are marked by asterisks. The ATG codons, which are the start sites of translation for the M1 and M23 isoforms of the mouse AQP4 are marked as Met-1 and Met-23.

AQP4 mRNA with a size of ~1700 bp and identify at the same time the presence of ~5000—6000 bp long AQP4 mRNA forms in northern hybridization experiments. The data reported recently on the nucleotide sequence of the bovine AQP4 mRNA[24] with a size of 5648 bp also points to the possible existence of new exons in the 3´-region of the mouse aquaporin 4 gene.

It is known that parameters of the 5´-nontranslation sequence of mRNA such as the ability to form secondary structures, the length, the content of GC pairs, the presence of false AUG codons, and so on may influence the mRNA translation efficiency. A computer expert system for predicting the translational properties of mammal mRNA based on detailed analysis of the structural characteristics of the 5´-nontranslation region has been designed by Kochetov *et al*.[28,29] The main idea that underlies the expert system is that the leader 5´-nontranslation mRNA sequences with low expression are more difficult to scan by the 40S ribosomal subunit according to the Kozak model.[28—30] Using this system, we analyzed the 5´-nontranslation regions of AQP4.M1 and AQP4.M23X and the MIWC1 mRNA sequence. According to the prediction made, the translational efficiency of AQP4.M1 mRNA is evaluated as low (the expert evaluation coefficient is −0.189765). The translational efficiency of the MIWC1 (M23) mRNA, whose transcription is initiated in exon 1, separated from exon 0 by more than 3500 bp, is evaluated as high (the expert evaluation coefficient is +0.191242). It should be noted that it is still obscure what particular regulatory region of the gene controls the expression of the new AQP4.M23X mRNA. It is unknown whether the region between exons 0 and X, upstream from the transcription initiation site of the given mRNA, contains binding sites for the transcription factors that independently determine the transcription of AQP4.M23X mRNA. In any case, in our opinion, it is fairly probable that promoter 0 located at a distance of somewhat more than 700 bp influences substantially the expression of the AQP4.M23X mRNA. Therefore, it is even more interesting that the translational efficiency of the AQP4.M23X mRNA, whose transcription is initiated in exon X (see Figs. 1, 2, and 4), is evaluated by the expert system as high (the expert evaluation coefficient is +0.182715). Presumably, under certain conditions in an organism or depending on the particular type of cells, in which the AQP4 is expressed simultaneously with AQP4.M1 (having a low translation efficiency), the AQP4.M23X mRNA translated with high efficiency is synthesized.

Thus, we showed that the mouse AQP4 gene consists of at least six exons (0, X, 1, 2, 3, and 4) separated by five introns and contains transcription initiation sites for three mRNA forms of mouse aquaporin 4. Two of them, the new AQP4.M23X mRNA and previously reported AQP4.M23, encode the same water channel isoform, M23. We were also able to demonstrate that all the mRNA forms of mouse AQP4 we identified (M1, M23X, and M23) encode the functionally active isoforms of aquaporin 4. Exon 0 is substantially larger than it has been considered formerly. The additional exon X is located between exons 0 and 1. The new AQP4.M23X mRNA is mainly expressed in brain, the expression being controlled during the development. The data obtained provide the possibility of more detailed research into the the AQP4 gene expression regulation at the level of individual mRNAs, which are obviously controlled by different, probably independently functioning promoters of the AQP4 gene.

We are especially grateful to Professor Anita Aperia, who kindly provided the resources of hers laboratory* for joint experiments within the framework of INTAS grant 97-11404. We also wish to thank Dr. Hjalmar Brismar for technical assistance in experiments on the location of fluorescence-labeled aquaporins in cells by a laser scanning confocal microscope.

This work was supported by the Russian Foundation for Basic Research (Project No. 01-04-49390) and the INTAS (Grant 97-11404).

References

1. M. Borgnia, S. Nielsen, A. Engel, and P. Agre, *Annu. Rev. Biochem.*, 1999, **68**, 425.
2. A. S. Verkman and A. K. Mitra, *Am. J. Physiol. Renal. Physiol.*, 2000, **278**(No. 1), F13.
3. R. W. Schrier, M. A. Cadnapaphornchai, and M. Ohara, *J. R. Soc. Med.*, 2001, **94**(No. 6), 265.
4. G. T. Manley, M. Fujimura, T. Ma, N. Noshita, F. Filiz, A. W. Bollen, P. Chan, and A. S. Verkman, *Nat. Med.*, 2000, **6**(No. 2), 159.
5. T. Ma, B. Yang, and A. S. Verkman, *Genomics*, 1996, **33**, 382.
6. J. D. Neely, B. M. Christensen, S. Nielsen, and P. Agre, *Biochemistry*, 1999, **38**, 11156.
7. S. Nielsen, L. S. King, B. M. Christensen, and P. Agre, *Am. J. Physiol.*, 1997, **273**, C1549.
8. M. Yasui, E. Serlachius, M. Lofgren, R. Belusa, S. Nielsen, and A. Aperia, *J. Physiol.*, 1997, **505**, 3.
9. A. Frigeri, G. P. Nicchia, J. F. Desaphy, S. Pierno, A. De Luca, D. C. Camerino, and M. Svelto, *FASEB J.*, 2001, **15**(No. 7), 1282.
10. B. Yang, A. N. van Hoek, and A. S. Verkman, *Biochemistry*, 1997, **36**, 7625.
11. E. Klussmann, K. Maric, and W. Rosenthal, *Rev. Physiol. Biochem. Pharmacol.*, 2000, **141**, 33.
12. S. Nielsen, E. A. Nagelhus, M. Amiry-Moghaddam, C. Bourque, P. Agre, and O. P. Ottersen, *J. Neurosci.*, 1997, **17**, 171.
13. E. A. Nagelhus, Y. Horio, and A. Inanobe, *Glia*, 1999, **26**, 47.

* Department of Women´s and Child Health, Karolinska Institutet, Q2:09 Astrid Lindgren Barnsjukhus KS, S-17176 Stockholm, Sweden.

14. M. L. Vizuete, J. L. Venero, C. Vargas, A. A. Ilundain, M. Echevarria, A. Machado, and J. Cano, *Neurobiol. Dis.*, 1999, **6**, 245.
15. M. Taniguchi, T. Yamashita, E. Kumura, M. Tamatani, A. Kobayashi, T. Yokawa, M. Maruno, A. Kato, T. Ohnishi, E. Kohmura, M. Tohyama, and T. Yoshimine, *Brain. Res. Mol. Brain. Res.*, 2000, **78**, 131.
16. Z. Vajda, D. Promeneur, T. Doczi, E. Sulyok, J. Frokiaer, O. P. Ottersen, and S. Nielsen, *Biochem. Biophys. Res. Commun.*, 2000, **270**, 495.
17. T. Maniatis, E. Frich, and Dzh. Sembruk, *Molekulyarnoe klonirovanie*, Mir, Moscow, 1984, 480 s. [T. Maniatis, E. E. Fritsch, and J. Sambrook, *Molecular Cloning*, Cold Spring Harbor Laboratory, 1982].
18. A. J. Harwood, *Methods Mol. Biol.*, 1996, **58**, 265.
19. S. A. Ortlepp, *Gene Anal. Tech.*, 1989, **6**, No. 5, 93.
20. pBluescript® II Exo/Mung DNA Sequencing System, Stratagene, Instruction Manual.
21. M. Yasui, D. Marples, R. Belusa, A. C. Eklof, G. Celsi, S. Nielsen, and A. Aperia, *Am. J. Physiol.*, 1996, **271**, F461.
22. M. Lu, M. D. Lee, B. L. Smith, J. S. Jung, P. Agre, M. A. Verdijk, G. Merkx, J. P. Rijss, and P. M. Deen, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, No. 20, 10908.
23. L. C. Turtzo, M. D. Lee, M. Lu, B. L. Smith, N. G. Copeland, D. J. Gilbert, N. A. Jenkins, and P. Agre, *Genomics*, 1997, **41**, No. 2, 267.
24. K. Sobue, N. Yamamoto, K. Yoneda, K. Fujita, Y. Miura, K. Asai, T. Tsuda, H. Katsuya, and T. Kato, *Biochim. Biophys. Acta*, 1999, **1489**, No. 2—3, 393.
25. S. Zelenin, E. Gunnarson, T. Alikina, A. Bondar, and A. Aperia, *Pediatr. Res.*, 2000, **48**, No. 3, 335.
26. M. Zelenina and H. Brismar, *Eur. Biophys. J.*, 2000, **29**, No. 3, 165.
27. M. Zelenina, S. Zelenin, A. Bondar, H. Brismar, and A. Aperia, *J. Am. Soc. Nephrol.*, 2000, **11**, 24A.
28. A. V. Kochetov, M. P. Ponomarenko, A. S. Frolov, L. L. Kisselev, and N. A. Kolchanov, *Bioinformatics*, 1999, **15**, No. 7/8, 704.
29. N. A. Kolchanov, I. I. Titov, and A. V. Kochetov, *Abstrs. Int. Conf. "RNA as Therapeutic and Genomic Target"* (*Novosibirsk, Russia, 30th August—2nd September 2001*), Novosibirsk, 2001, 38.
30. M. Kozak, *Mamm. Genome*, 1996, **7**, 563.

Received October 12, 2001
in revised form December 26, 2001