

Research article

Identification of the gene encoding Brain Cell Membrane Protein I (BCMPI), a putative four-transmembrane protein distantly related to the Peripheral Myelin Protein 22 / Epithelial Membrane Proteins and the Claudins

Christiane Christophe-Hobertus¹, Claude Szpirer¹, Richard Guyon² and Daniel Christophe^{*1}

Address: ¹Institut de Biologie et de Médecine Moléculaires (IBMM), Université Libre de Bruxelles, rue des Professeurs Jeener et Brachet 12, B-6041 Gosselies, Belgium and ²UMR 6061 CNRS, Faculté de Médecine, 2 av. Professeur Léon Bernard, 35043 Rennes cedex, France

E-mail: Christiane Christophe-Hobertus - chobertu@ulb.ac.be; Claude Szpirer - cszpirer@dbm.ulb.ac.be; Richard Guyon - Richard.guyon@univ-rennes1.fr; Daniel Christophe* - dchristo@ulb.ac.be

*Corresponding author

Published: 5 July 2001

Received: 11 May 2001

BMC Genomics 2001, 2:3

Accepted: 5 July 2001

This article is available from: <http://www.biomedcentral.com/1471-2164/2/3>

© 2001 Christophe-Hobertus et al, licensee BioMed Central Ltd.

Abstract

Background: A partial cDNA clone from dog thyroid presenting a very significant similarity with an uncharacterized mouse EST sequence was isolated fortuitously. We report here the identification of the complete mRNA and of the gene, the product of which was termed "brain cell membrane protein I" (BCMPI).

Results: The 4 kb-long mRNA sequence exhibited an open-reading frame of only 543 b followed by a 3.2 kb-long 3' untranslated region containing several AUUUA instability motifs. Analysis of the encoded protein sequence identified the presence of four putative transmembrane domains. Similarity searches in protein domain databases identified partial sequence conservations with peripheral myelin protein 22 (PMP22)/ epithelial membrane proteins (EMPs) and Claudins, defining the encoded protein as representative of the existence of a novel subclass in this protein family.

Northern-blot analysis of the expression of the corresponding mRNA in adult dog tissues revealed the presence of a huge amount of the 4 kb transcript in the brain. An EGFP-BCMPI fusion protein expressed in transfected COS-7 cells exhibited a membranous localization as expected. The sequences encoding BCMPI were assigned to chromosome X in dog, man and rat using radiation hybrid panels and were partly localized in the currently available human genome sequence.

Conclusions: We have identified the existence in several mammalian species of a gene encoding a putative four-transmembrane protein, BCMPI, which defines a novel subclass in this family of proteins. In dog at least, the corresponding mRNA is highly present in brain cells. The chromosomal localization of the gene in man makes of it a likely candidate gene for X-linked mental retardation.

Background

We recently developed a screening procedure for the selection of sequences encoding proteins targeted to the cell nucleus. Our method relies on the expression in transfected cells of enhanced green fluorescent protein (EGFP) fusion proteins from cDNA library constructs [1]. The selected clones encode EGFP fusion proteins that accumulate in the cell nucleus. Many of them were shown to harbor cDNA sequences corresponding to nuclear proteins that were translated in frame with the EGFP coding sequence. However, in nearly half of the selected clones the production of a fusion protein able to accumulate in the nucleus was shown to result from out of frame translation of the cDNA sequence fused to the EGFP coding region. On the average indeed, only one out of three cDNAs was positioned in frame with the EGFP coding sequence in the starting library. It was not expected that functional nuclear localization sequences would be generated at random (i.e. by out of frame translation of cDNA sequences) as often as was observed.

One clone, called "C60", that was isolated in this approach exhibited a significant DNA sequence similarity with a mouse EST sequence present in the EMBL/GenBank database (clone MNCb-0941, accession #: AU035837) [1]. No open reading frame (ORF) had been identified in this sequence yet, but the comparison of our dog sequence with the one from mouse identified a putative ORF on the basis that in the 385 bp-long region of similarity most of the differences occurred at the third position of base triplets in frame with a starting ATG codon. However, both sequences diverged before the stop codon was reached. Assuming that this was the correct reading frame, the cDNA portion in our EGFP fusion construct was translated out of frame (frame +2). This out of frame translation generated a 201aa-long sequence presenting several neighbouring clusters of arginine residues, which somehow resembled basic type nuclear localization signals. Although it could explain why this cDNA was isolated in the screening, it did not allow us to conclude whether the protein normally encoded by the cDNA is a nuclear protein or not. To further characterize the protein encoded by the cloned sequences we decided to isolate a complete copy of the corresponding mRNA.

Results and Discussion

Identification of the complete dog BCMP1 mRNA

The random primed cDNA insert harbored by clone C60 [1] was used as probe to screen a dog thyroid oligo-dT primed cDNA library in λ ZAPII phage vector [2]. Sixty positive clones were obtained out of the 500,000 cDNA clones screened. The longest insert (from clone C60-1) had a size of 4 kb and was entirely sequenced. Compared to the sequence of the insert of clone C60, this cDNA ex-

hibited a 2 bp extension in 5' and a 2,944 bp extension in 3'. The 3' poly-A tail was preceded by a correctly placed AATAAA motif (fig. 1). The longest ORF corresponded to the putative ORF identified previously by comparing the sequence from clone C60 with that of the mouse EST present in the database (see background section). It extended over 543 bp (181 aa), from position 193 to 735 in the cDNA sequence. The translation initiator codon was located in a suitable sequence context according to Kozak's rule [3]. As in the interval an updated homologous mouse sequence had been deposited in the database (clone MNCb-0941, EMBL/GenBank acc. #: ABO41540), the comparison of both sequences revealed that the coding region was entirely conserved in dog and mouse (fig. 2).

The 3.2 kb-long sequence located in 3' of the TAG codon (3' UTR) in the dog cDNA was distinctly AT-rich and contained 9 ATTTA motifs. These characteristics have been implicated in the rapid decay and restricted translation of mRNA molecules [4,5,6]. This 3' UTR was shorter in the mouse (1.3 kb) but several portions of it exhibited a remarkably high sequence conservation when compared with the dog sequence (fig. 2). Especially, the AT-rich character and the occurrence of multiple ATTTA motifs were preserved. A search in the database also identified a human sequence (DKFZp564E153, EMBL/GenBank acc. #: ALO49257) presenting a very high degree of sequence conservation over 2.5 kb with the 3' part of our dog cDNA (fig. 2). The coding region of the mRNA was not contained in this human sequence and the observation of such an extended conservation of DNA sequence between UTRs from different species was unexpected. During the preparation of this manuscript, a completed human sequence appeared in the database (DKFZp761J17121, EMBL/GenBank acc. #: AL136550). The coding region was entirely conserved between dog, mouse and man, and most of the ATTTA motifs present in the dog sequence were also preserved in man (fig. 2). It may suggest that BCMP1 mRNA is indeed subjected to tight post-transcriptional controls. However, whether the presence of these sequences really confers instability to the mRNA and restricts its translation remains to be determined experimentally.

A number of EST sequences from various species which were clearly homologous to dog BCMP1 could be retrieved from the database by BLAST searches. They indicated that the BCMP1 gene must also exist in the rat (e.g. acc. # BG381247), beef (e.g. acc. # AW352911), pig (e.g. acc. # BF704530) and in the fish *Gillichthys mirabilis* (acc. # AF266205), in addition to the already cited dog, mouse and man.

```

1  CTG GCG GCG AAG GTC GCG GGC GGG GTC GCG GTC GTC GCG GCG 88
2  GAC CCG CGA GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG 90
3  GCG GTC GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG 100
142 GCG GCG GTC GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG 192

1 Met Ala Ser Ala Gly Ser Gly Met Glu Glu Val Asp Val Ser Val Leu 10
193 AGT GCT TCC GCG GCG AGC GGC GCG GCG GCG GCG GCG GCG GCG GCG 240

17 Thr Pro Leu Lys Leu Val Gly Leu Val Cys Ile Phe Leu Ala Leu Cys 32
241 AGC CCG CTC AAG GTC GTC GCG GCG GCG GCG GCG GCG GCG GCG GCG 288

33 Leu Asp Leu Gly Ala Val Leu Ser Pro Ala Trp Val Thr Ala Asp His 68
289 GTC GAC CTC GGG GCG GTC CTC GTC GCG GCG GCG GCG GCG GCG GCG 336

49 Glu Tyr Tyr Leu Ser Leu Trp Glu Ser Cys Arg Lys Pro Ala Ser Leu 64
337 GAG TAC TAC CCG GCG GTC GTC GCG GCG GCG GCG GCG GCG GCG GCG 384

60 Asp Ile Trp His Cys Glu Ser Thr Leu Ser Ser Asp Trp Glu Ile Ala 80
385 GAC ATC TCG CAC TCG GAG TCG AGC GTC AGC AGC AGC GAT TCG CAG ATT GCT 432

81 Thr Leu Ala Leu Leu Leu Gly Gly Ala Ala Ile Ile Leu Ile Ala Phe 96
433 ACT CTG GCT TCA CCG TCG GCG GCG GCT GCG GCG GCG GCG GCG GCG 480

97 Leu Val Gly Leu Ile Ser Ile Cys Val Gly Ser Asp Arg Arg Phe Tyr 112
481 CTG GTC GGT TGG AAT OCT ATC TGC GTC GCA TCT CGA AGC CCG TTC TAC 528

213 Arg Pro Val Ala Val Met Leu Phe Ala Ala Val Val Leu Glu Val Cys 128
529 AGA GCT GTC GTC GCG ATG CTC TTT GCA GCA GTC GTC TTA CAG GTC TGC 576

329 Ser Leu Val Leu Tyr Phe Ile Lys Phe Ile Ile Tyr Val Ser Leu Lys 144
573 AGC CTG GTC CTC GAC CGA ATC AAG GTC ATT GAA AGC GTC AGC TTC AAA 624

395 Ile Tyr Ala Glu Phe Asn Trp Gly Tyr Gly Leu Ala Trp Gly Ala Thr 160
625 AAT TAC CAT GAG TTC AAC TCG GAT TAT GCG CTC GTC TGG GCT GCA ACT 720

363 Ile Phe Ser Phe Gly Gly Ala Ile Leu Tyr Cys Leu Asn Pro Lys Asn 176
673 AFA TTT TCC TTT GGG GAT GGC ATC GTC TAT TGC CTC AAG GCT AAG AAC 720

373 Tyr Glu Asp Tyr Tyr ***
723 TAC GAA GAC TAC TAC TAG AAT CAA CAG TCT CAG ATT TTA AAA ACA ACC 768
749 AAC CAT TCA ACA AAA GAA TTA CTT GAT TAT CTT TTT TAA CTC ACT GTC 816
817 TTT TAA AAG ACT GCT GGA GTA TCG AGC AGT TTT CTC AAT TCA TTT AAG 864
805 AGC TTT GCG GTC TCG GGT GCA AGC AGC AGC ATA ACC AGC TTT TAC AAT CAT 912
912 TTT AGC GAC CCG GAC AAA TTA AGA GAG CTG ATT AGA CAT AGG GAA AAG 960
961 GTC CAA ACC TCC AAT AAG TTA AGC GAA TTT TTT TCG ACA AGT GAA AGG 1008
2009 CTC GCT ATT ACA GCA ACC AAT GCG AGA GAC TAG TTT GAA TGA GAA AGT 1056
2057 CTT GAA TCT CTT AAT GCG TGC AAG GAA GCA GCT GGC ACA AGA AAC ATT 1104
1105 TAG AAG TTC CTT TGC TCT GAC AAG GAT TCC ACT GCT ACA GAG TCC TTA 1152
1153 TTA CCT GCT TAC GCG AOC GAA TGA CTA AAT TGG CGA TTT GTC ATT TCC 1200
1201 TCG AGT GAG CTT TTT AAA GGT GAA CTG CCG TTC AGC ACC TAA TGG GAG 1248
1249 TTC TGA GTC TAA TGG GTT AAA GGT GAA TAG ATT CTA CCT TCA AGA ATA 1296
1291 GTC TTT TTT AAT GGG AAG GAA TGC TTT GCG AGT CAG GAT CTC OCT GGG 1344
1345 AAA GGA GTC AAG ACT TGG AGA GAT GTC CTT TTT ATT ATG TGG TTA GAA 1392
1393 AAT TTT CCT GAG GCG TAT CTA GAA TGA GGA GAA TTT AGA GTC TTT ACT 1440
1441 TTC TCT GGG GCA ACT AGA AAG AAG TAT GCA TGA GTC GCT TTT GTC CTA 1488
1492 AAT AAT CTT TAA AAT ATT TAC CAA ACA TTA CCG CAA GAA ATA AAT GTC CCG 1536
1537 AAT TAA CAA GCG TAA TAG TAG AAG TTA TAG AAG CAT TAC TAC AAT 1584
1585 TGG TAG TAA ACT AAT CCG TAG CCG TAT ACA TGC AAT GCG CTA TTA CAC 1632
1633 AEA AAA ATA ACA TTT GCT CTA GTC AAG TGA TTT GAA CAC TAA OCT TGT 1680
1681 ACA GAT TAA GAG GCT TAG ATT GGT AAT CAC ACT TAC TTA TCG TAC AAA 1728
1729 AGT ACA TAT AAT TTA AAG CTT TGG CCG TCG GTC GTC TAC TGT TTC ACT 1776
1777 GAA AAG TTT TAA TAG AAT TGT AAA GAA GAA AAA TTT CAA TGG TGT TAA 1824
1825 GCT GAA AAT TGA TGA TTT TTT TGT AAG CAC TAT AAT TCA AGC AAG AAT 1872
1873 AGC AAT TAA AAT AAT CAT TTT GGT AAA AAT TCA GTC AAA TAG AAG GCT 1920
1921 GCT TTT TTT TGC AAG TAC CTT AAA GCT CTT TAT TAA AAA ATA AAA GAA 1968
1969 TAG TGA TAG ATT TAT AAG AAT GTC TTA ATG TCT CAG TAG AAT AAT ACA 2016
2017 AAT CAT TGC TGA TCA CTT GGT GCT CCG AAG TGA AAT AAT GCG CTA ACT 2064
2065 GAT TGG TGC ACA CTA TTA TGA TGG ATT TAT GTC ACA GGA ATC ATT TGT 2112
2113 TGG CTG TTC AAA TGG AAG GAA ATG TCA ATG ATA GGG AAT AAG TTT GCG 2160
2161 ACT AAT CTC AGA CTG CAA ACT TGT GGT AAT CCG GGT TAC TAT ACA AAT 2208
2209 TTT GAT AAT TCA AAT GTA AAT ATG TTT TTA TAT CAA GCA TAC AAT TCT 2256
2257 AAT ATA AAA GTC ATA AAT AAT TAC TTT TTT TAA CAA AGG CAC TAA AAC 2304
2305 AAT GTC CCG GGT TTT GCG CTT TTT GCA GAA GAA AGC AAT GGA AAA ATC 2352
2353 ATT TAA AAC ATG CTT ATG TTC TAG TGT AAT GCA AAA TCC GTC AAG AGC 2400
2401 AGC ACT ACA TGA ATA TTA TTT AAT AAT TTT GTC TAC CTC CCA GAA CAG 2448
2449 AAT TCT GCT TTC AGC AAT TTT TAC GAA CCG TTT GCG AAA ACA ACA GAA 2496
2497 AAA AAA AAT TAT TTT AAT AAT CAA GCG CTT GAA AAA AAA AAA AAA CCG 2544
2545 TCC TAT CTT GCG AAT TTT CCG TAG GAT CAC AAA GCA AAT ATG AAG AAT 2592
2593 GAG GGT GAA TCT TAT TTT AAT AAT GAC AAG CAA AAG TAA TTA AAT GCG 2640
2641 TCC TAG AAG GAA ACC TTT AAT AAA GAA TAT GAT TCT CTG TAG AAT GAT 2688
2689 AAG AAA CTA TGG GTC AAG TTT TCC AAG AAA ATA TCA GAA TTA TAT AAT 2736
2737 CCT TTA AAG CAA TTT TCT TGA GAA GAA GAA AAA TCT GGG CCA GAC TGT 2784
2785 CTT GCA GTC AAT TGC GTA TTT TGA GAG CAG GAT GAA AAA TAC GAT AAT 2832
2833 TTT TGA AAG GGG AAT TGT GGT AAT AAT CAC TCT CTA TTT TGA AAG AAT 2880
2881 CCG TTT AAT CTT GAA TAA TGA AAG AAT AAT AAT TTT GTC GTC GTC GTC 2928
2929 AAG AGC CAC TTT TCT TAG TCT TAG TGG AGG AAT CAG TGT GCT GTC AAG 2976
2977 CCA GCA CCA CAC CTC GAG TCT TTT GAA GTC TTT ATT CTT TGT CAA TAT 3024
3025 GAC AAG CTC GAA AAT GAC AAT AAT TTA TAG GCT AAG GCA CAC ATG TTT 3072
3073 GCA ATA CAC TGA AAA GCA GAT TTT TTT TAT AAT TTT TTT CCG CTC CCG 3120
3121 AAT GAA GAG CCG TAT GTC AAT TTT CTA CCA TCA CTA ATA TTT GAA AAT 3168
3169 AAT TGA TGA CCA ACU AAT TCA GCA CCA CAT GAA AAT TTT TGC TCA TCC 3216
3217 GAA ATA TTT TAT GTC TTA TGA AAA CCA CCA GAT GAA TTT TCA GCT AAT 3264
3265 TTA GTC CTT TGC CCG TTT AAT TTT GGT TGA TTT TTT CTT ACA AAT AAA 3312
3313 ACT GTC AAT TAC AAA CCG AAT GGT TAT CTT AAT TAT GCA CAT GTC CAC 3360
3361 ACA TAG AAT AAT AAA TAC TTA GCA GCG CTA GTC AAT TTT AAT TTT TCT TGC 3408
3409 ACA ACT ATT TAC CTT TTT GTA AAT TCA ACA TGT AAT TTT TAA AGA CAA 3456
3457 GAA TAG AAG GAC CTA TGT GTC TGA AAT ATA AAT GAT ATA TAT GGA TTA 3504
3505 GCA TGT ACC TGT AAT TTA TTA AAT GCG CAA TGA AAT GAC TGG TAA GTC 3552
3553 AAT TGT AAT TGT AAT GCT AAT AAT ATT TAT TCA GAT TGT AAT TTT 3600
3601 GAA CAG AAT AAT GCA CAC TAA CCG GCG CTT CTA TGT CTT CTT TAA TGC 3648
3649 CTA AAA CTC TGC CTA AAT ATT TCA TCT GTC TTA AAT AAT AAA AAT AAT 3696
3697 CTT GAT AAT GTC TAT GAT AAT GTC CTA CTA CCG TGC AAT ATT TAT 3744
3745 TAT TTT TAA AAT TAT CAC AAT TTT ACT ACT TAA ATA TGA ATC CAT GGT 3792
3793 AAT CTG GTC AAT TTT TTT AAT AAT CTT CTG GCG GCA ACC TGT TTC TCA 3840
3841 CCG CAC TGC TTT TCA GTC TGC AAT TTT ACA ATC AAT TCT CCA TTT CAT 3888
3889 GAT TTT GGT AAT TCA CAT ATG AAT TTA TCT ATG TGG AAT AAA TAA AAA 3936
3937 TAA AAT TGC TTT CAC TGA AAT AAA AAA AAA AAA AAA AAA AAA AA

```

Figure 1
Nucleotide sequence of dog BCMP1 cDNA. The amino acid sequence encoded by the ORF appears above the corresponding DNA sequence. The underlined sequence corresponds to the insert of the original clone C60 (see text). ATTTA motifs appear in bold and the polyadenylation signal is highlighted in blue.

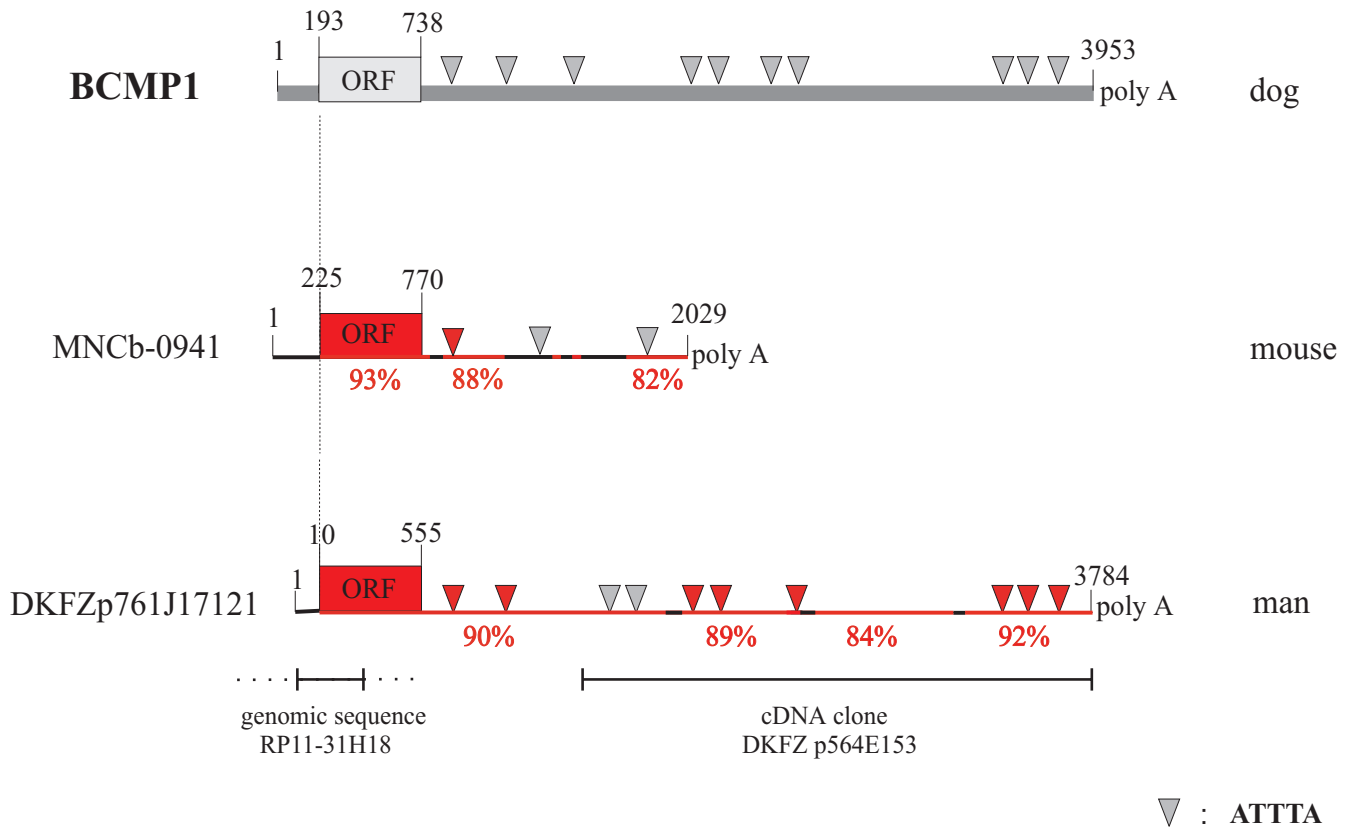


Figure 2
Structure of BCMP1 mRNA. Schematic representation of known sequences from mouse and man that are homologous to dog BCMP1 sequence (short EST sequences were not considered). Coordinates are given for each sequence individually. Highly conserved regions are highlighted in red with indication of the percentage of identity relative to the dog sequence. Inverted triangles symbolize ATTTA motifs. They appear in red when the motif is conserved in the dog sequence.

Analysis of BCMP1 mRNA expression in the dog

Originally, the cDNA had been isolated from a dog thyroid cDNA library. In order to investigate whether the corresponding mRNA was also present in other cell types, a northern blot experiment was performed using poly-A+ RNA preparations from various dog tissues (fig. 3). Huge amounts of the 4 kb transcript were detected in brain cells. The presence of the mRNA was also detectable in most of the other RNA preparations but to lesser extents as compared to that found in brain RNA. The encoded protein is thus expected to be particularly abundant in the brain, unless the peculiar 3' UTR of the mRNA mediates a deep control on its translation (see above).

Prediction of BCMP1 protein structure and subcellular localization of an EGFP-BCMP1 fusion protein

The 181 aa-long protein sequence encoded by the mRNA did not present any significant resemblance with sequences present in protein databases. A search for the presence of protein family signatures (PfamHMM on Ex-pasy server) revealed the occurrence in the novel protein

of sequence motifs resembling significantly to one of the two motifs specific to the peripheral myelin protein 22 (PMP22) family of proteins and to the motif specific to the claudins (fig. 4). The two identified signatures overlapped partially in the novel protein sequence. PMP22 and the related epithelial membrane proteins (EMPs) [7], as well as the claudins [8], all belong to the super-family of four-transmembrane domain (4TM) proteins. As could be expected, the search for the existence of putative transmembrane domains in the novel protein (HMMTOP on Ex-pasy server) identified the presence of four of such domains (fig. 5). The protein thus appeared to be a novel member of this large family of proteins, somehow related to PMP22/EMPs and the claudins. As these are integral membrane proteins, and as the mRNA encoding the novel protein was predominantly found in brain, the newly identified protein was termed "brain cell membrane protein 1" (BCMP1).

According to the putative BCMP1 structure, the extracellular loop between TM1 and TM2 would be larger than the intracellular loop between TM2 and TM3 and the ex-

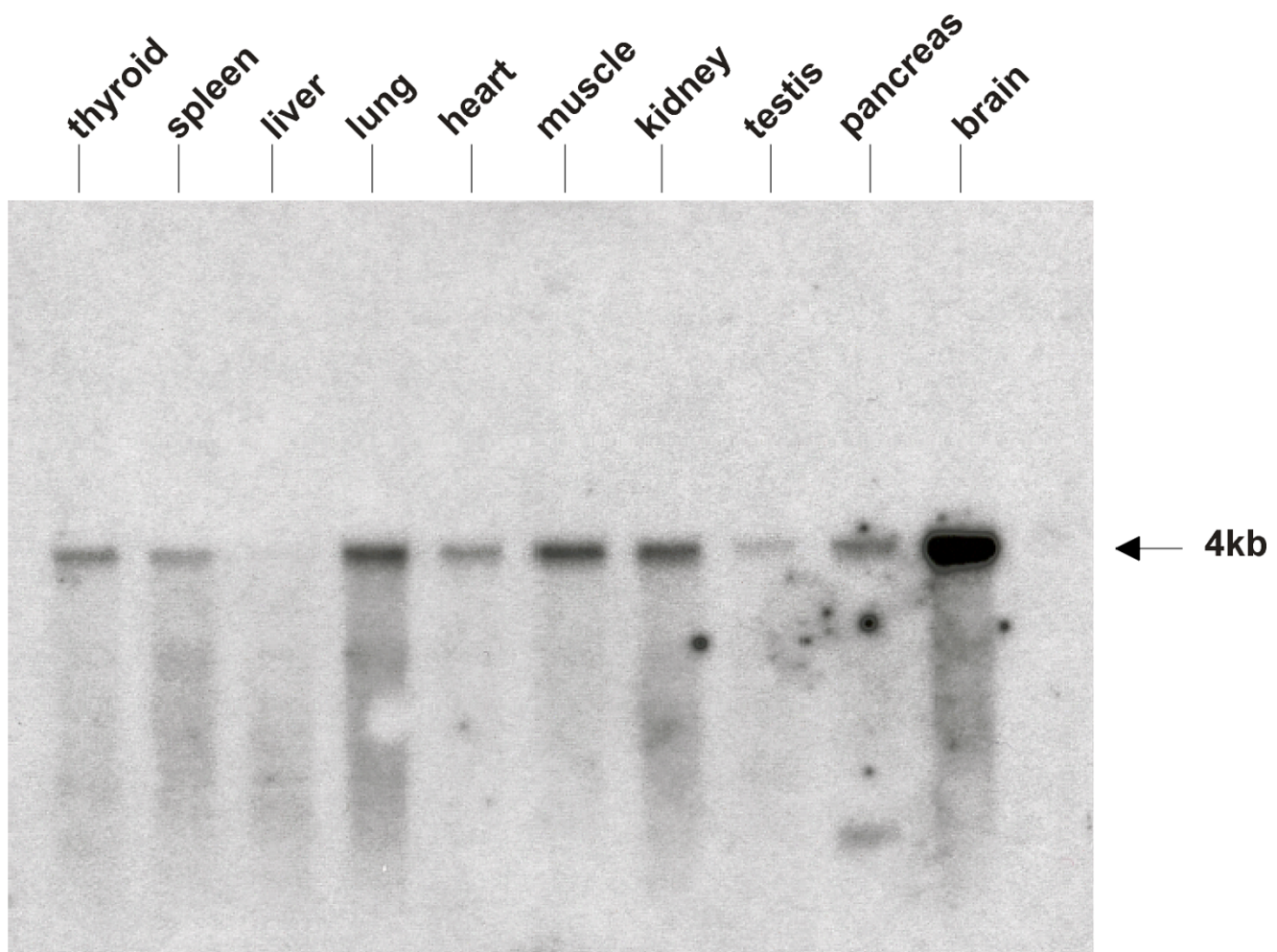


Figure 3
Northern-blot analysis of BCMP1 mRNA expression. PolyA⁺ mRNA preparations from various dog tissues were probed with the coding region of BCMP1 cDNA. The arrow points to the signal corresponding to the expected 4 kb-long transcript.

tracellular loop between TM3 and TM4, as it was supposed to be also the case in PMP22/EMPs and claudins. However, the intracellular amino-terminal arm proceeding the first transmembrane domain appeared to be much longer in BCMP1 than in its relatives.

In order to refine the classification of BCMP1 within the four-transmembrane domain protein family, a phylogenetic tree was constructed on the basis of the alignment of the available protein sequences related to the PMP22/EMPs and claudins (fig. 6). Dog BCMP1 and its mouse and human orthologs segregated as a distinct subgroup in the tree. Their closest relatives were the recently identified mouse PERP gene product [9] and the protein encoded by the CG6982 gene in drosophila (EMBL/GenBank acc. #: AAF56054). This group of proteins thus

shared primary structure determinants which defined a distinct subclass in the protein family.

In order to assess experimentally the postulated membranous localization of BCMP1, an EGFP-BCMP1 fusion protein was expressed in transiently transfected COS-7 cells and the subcellular localization of the hybrid protein was observed by fluorescence microscopy (fig. 7A and 7B). A fine granular fluorescence was observed all over the surface of cells expressing the EGFP-BCMP1 fusion protein, consistent with a plasma membrane localization of the tagged protein. A stronger fluorescence surrounding the cell nucleus was also observed. It indicated that a significant part of the expressed fusion protein accumulated in the endoplasmic reticulum. The pattern of EGFP fluorescence remained almost un-

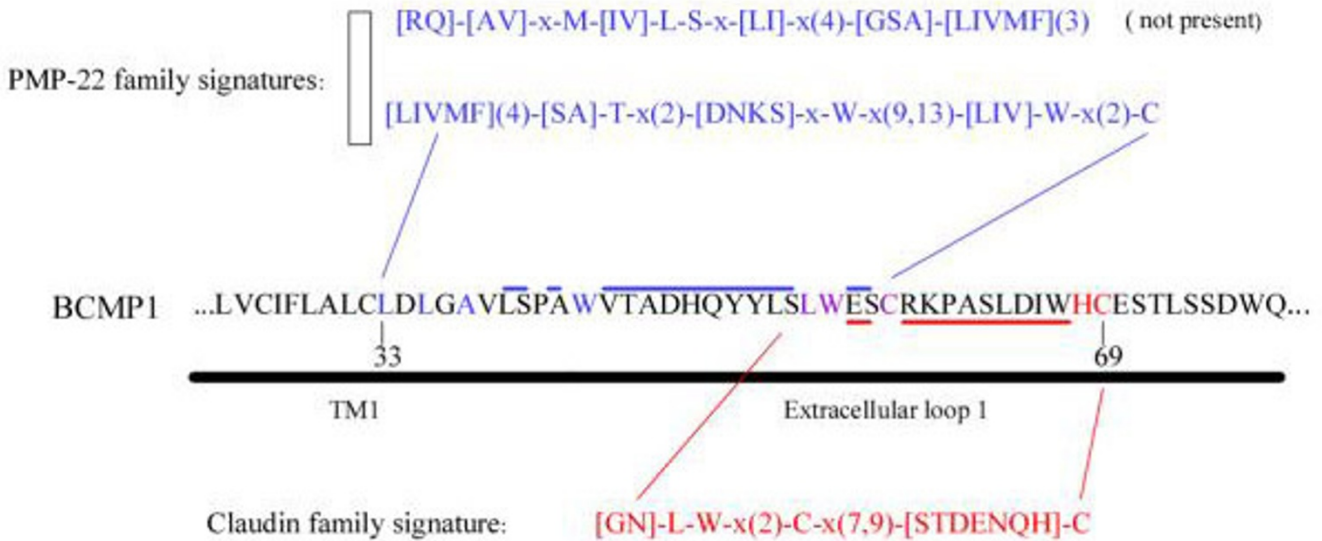


Figure 4
Identification of PMP22 and Claudin family signatures in BCMP1 primary structure. Conserved residues are coloured (blue: PMP22 signature, red: claudin signature, violet: overlapping PMP22 and claudin signatures) and conserved spaces between specific residues are over - or underlined. The part of the predicted BCMP1 structure (see fig. 5) to which the primary sequence corresponds is shown (TM1 = first transmembrane domain).

changed when the cells were permeabilized with saponin in order to stain the nuclear DNA (fig. 7C and 7D). This indicated that the fusion protein was embedded in the membranes and that it was not able to readily diffuse out of the cell.

Localization of the BCMP1 gene in dog, man and rat

In dog, the BCMP1 coding sequence was typed in duplicate on the 118 cell lines of the RHDF5000-2 radiation hybrid panel [10] on the latest version of the RH map [11]. The BCMP1 gene was linked to chromosome X close to FH2548 with a Lod score of 11.88. Marker FH2548 is located close to the DMD locus in dog (distance: 4.4 cR₅₀₀₀, approx. 500 kb). More informations about dog RH maps can be found at <http://www-recomgen.univ-lyon1.fr/doggy.html>.

The human EST sequence DKFZp564E153 (EMBL/GenBank acc. #: AL049257) that corresponds to the 3' UTR of dog BCMP1 mRNA had been localized on chromosome X. The corresponding human genomic sequence could not be found by BLAST searches against sequences available in the database. However, by using the coding region of dog BCMP1 a significant match was identified with genomic sequences assigned to human chromosome 8 (clone RP11-31H18, EMBL/GenBank acc. #: ACO41003). The similarity extended from position 1 to 418 in the human cDNA sequence (fig. 2), which corresponded to the amino-terminal part of the protein up to

the first extracellular loop. As an intron was found at this same position in PMP22, EMP-1 and EMP-3 genes, it appeared likely that we had identified the first coding exon of the human BCMP1 gene. PMP22, EMP-1 and EMP-3 genes all contain an additional intron separating the sequences encoding the first transmembrane domain and the first extracellular loop into two exons [7]. This intron is clearly not present in the human BCMP1 gene. In order to clarify the location of the gene in the human genome (chromosome X or chromosome 8 ?), the GeneBridge-4 WGRH panel was used to map the sequences encoding human BCMP1 using a pair of primers directing the amplification of a 666 bp-long fragment encompassing the entire first coding exon and the exon-intron junction. It revealed that the amplified segment was located on chromosome X, 0.20 cR₃₀₀₀ from marker W1-7096 and 6.51 cR₃₀₀₀ from marker DXS1214. This location agreed with the previous assignment of the EST sequence DKFZp564E153. It also corresponds to the cytogenetic location Xp11.4. As the DMD gene maps at Xp21.2 in man, it is thus also close to the BCMP1 gene in this species. The chromosomal localization result revealed unambiguously the existence of a single BCMP1 locus in the human genome. As a consequence, it indicated that the sequences of the genomic clone RP11-31H18 had been inappropriately assigned to chromosome 8 instead of chromosome X in the database.

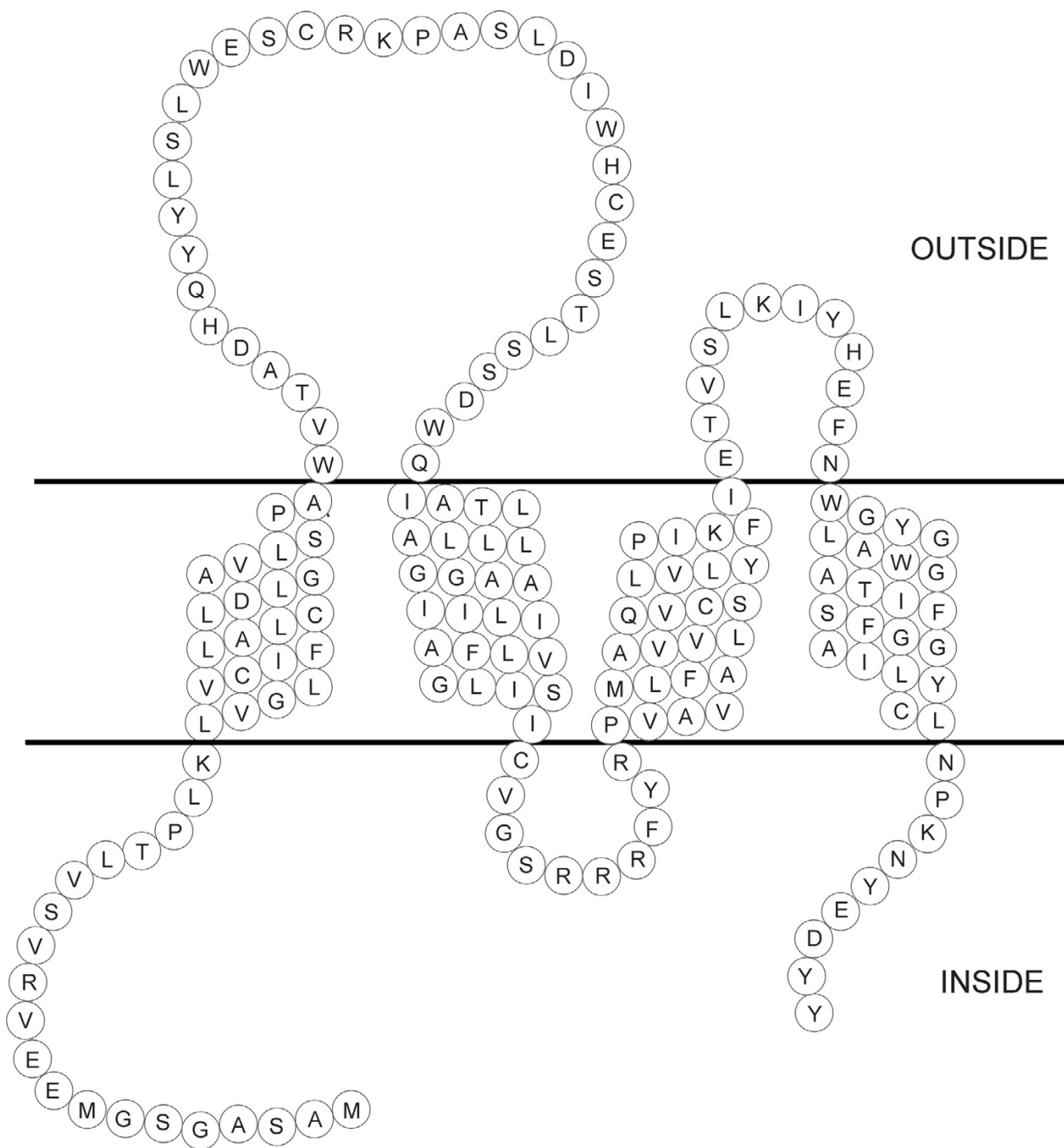


Figure 5
Predicted structure of BCMP1 in the plasma membrane. The structure was drawn on the basis of the predictions obtained from HMMTOP on the ExPASy server.

In the annotated human genome sequence available on the Ensembl server, the first coding exon of the human BCMP1 gene (gene ID:ENSG00000101959) is present in the chromosome X sequence (the sequences of clone

RP11-31H18 have now been properly reassigned to chromosome X; see ContigView on Ensembl server). Part of the coding region of human BCMP1 and the whole 3'

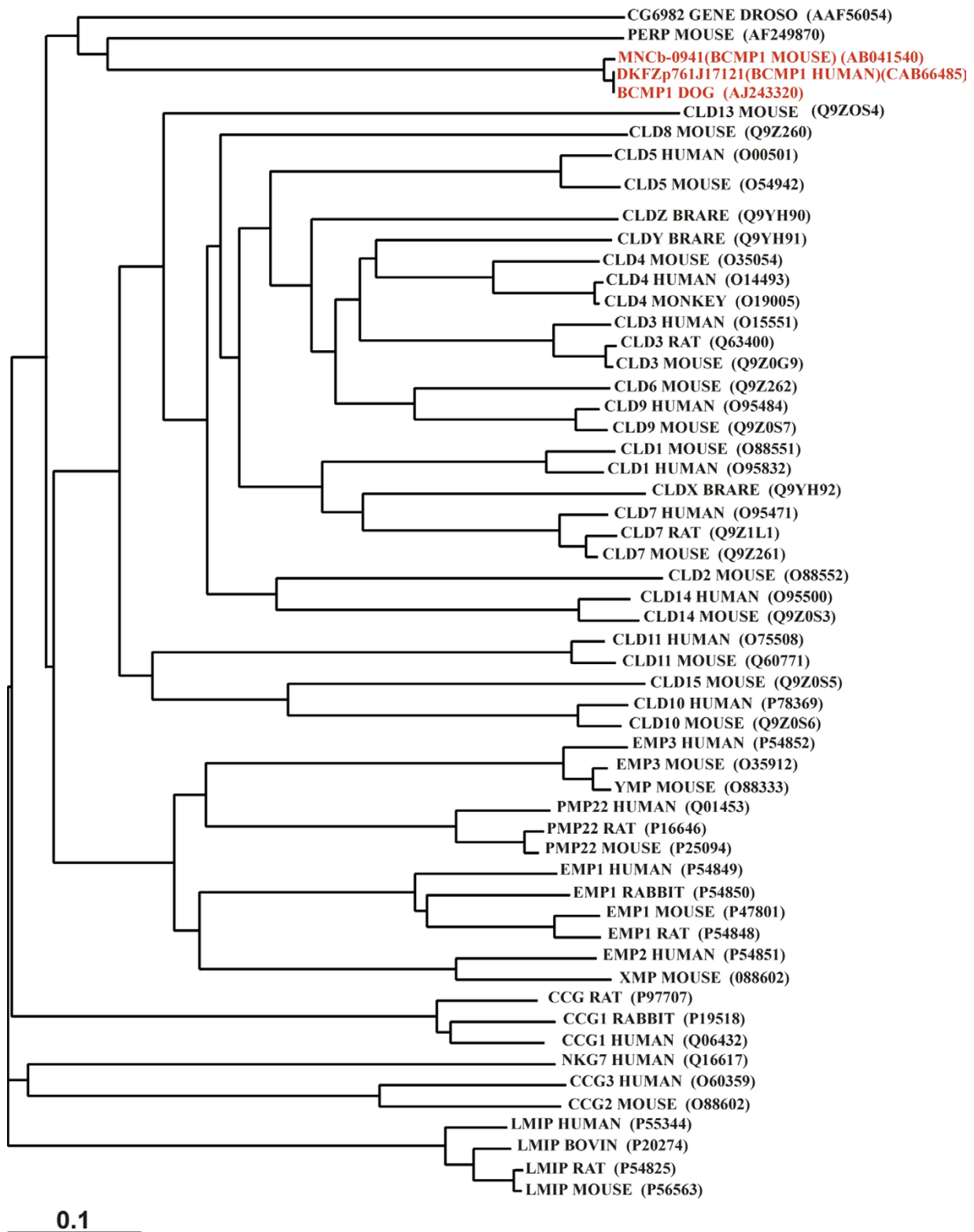
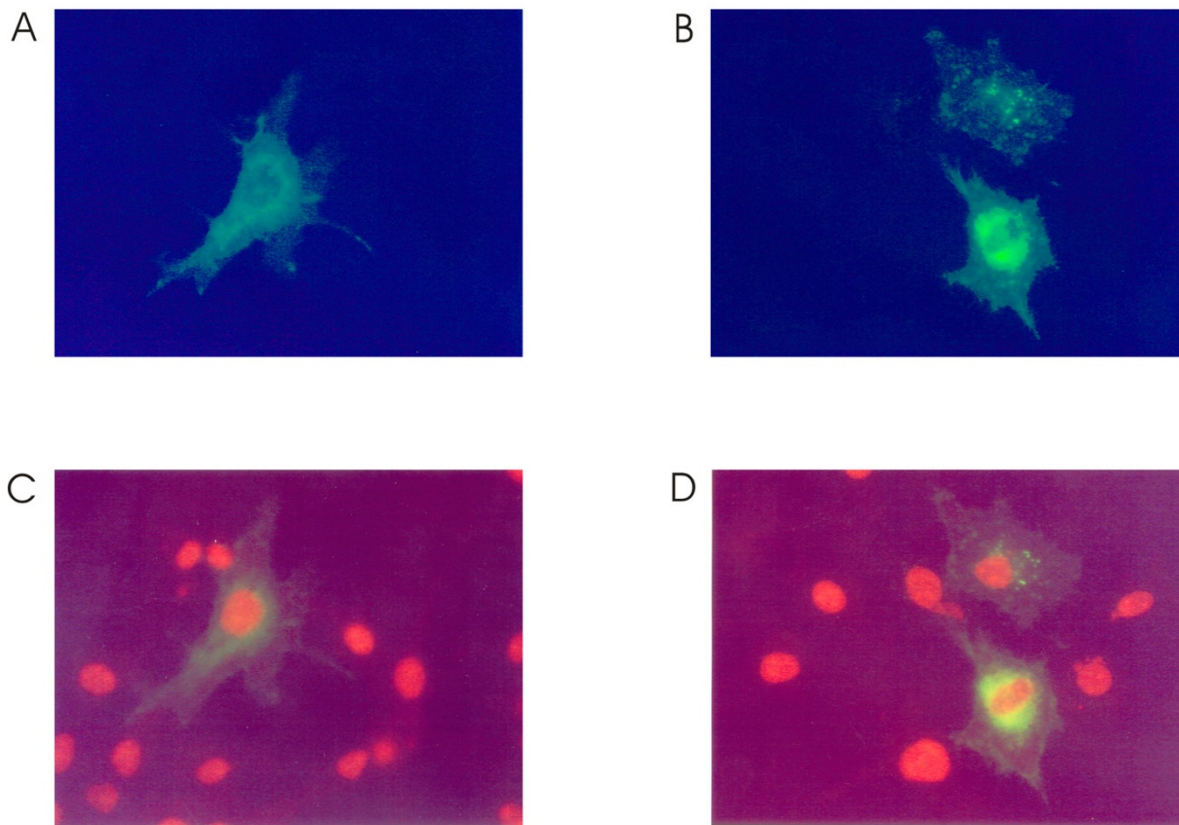


Figure 6
Phylogenetic tree of four-transmembrane proteins related to PMP22/EMPs and the claudins. See materials and methods section for the description of the tools used for the construction of the tree. Database accession numbers of individual sequences are given in parentheses. Dog BCMP1 and its mouse and human orthologs appear in red.

**Figure 7**

Subcellular localization of the EGFP-BCMP1 fusion protein. The fusion protein was expressed in COS-7 cells. Parts A and B: observation of EGFP fluorescence in the perinuclear region and at the cell membrane. Parts C and D: cells were permeabilized and the nuclear DNA was stained with ethidium bromide in order to visualize the cell nucleus.

UTR region corresponding to DKFZp564E153, are still missing in the currently available human genome sequence.

Using primers derived from the rat EST236642 (EMBL/GenBank acc. # AI408352), which is 91% identical to the segment 2028-1434 of the mouse brain cDNA MNCb-0941, itself similar to dog BCMP1 cDNA, the rat gene (symbol: *Bcmp1*) was assigned to the chromosome X, between DXRat67 and DXRat28, at 497.9 cR along the MCW map (LOD score: 9.0; the local map is: DXRat67 - 29.9 CR - *Bcmp1* - 0.4 cR - DXRat28). The marker DXRat67 co-localizes with the gene *Dmd* [12], itself cytogenetically assigned to Xq22 [13]. The rat genes *Bcmp1* and *Dmd* are thus closely linked, as was already observed in dog and man.

Conclusions

We have described here the identification of the gene encoding a novel protein, called Brain Cell Membrane Protein 1 (BCMP1), which belongs to the large family of four-transmembrane proteins and appears to be highly expressed in the brain. The gene seems to be conserved on chromosome X within mammals, in close association with the DMD locus in man, rat and dog at least. The encoded BCMP1 protein shares significant resemblances with both PMP22/EMPs [7] and the claudins [8], but exhibits distinct features, notably a predicted intracellular amino-terminal extension, which distinguishes it from the other known members of the family.

PMP22/EMPs are integral membrane proteins that seems to be implicated in various cellular processes, such as cellular differentiation, control of proliferation, and apoptosis [7]. PMP22 has been shown to play a critical role in peripheral nerves, where it is involved in the as-

sembly of peripheral nerve myelin and in the regulation of proliferation and differentiation of Schwann cells. The claudins also constitute integral membrane proteins which are localized exclusively at tight junctions [8]. Claudin-1, -2 and -3 have been shown to present calcium-independent cell-adhesion activity [14].

Alterations in the PMP22 gene are responsible for hereditary motor and sensory neuropathies in human and rodents, known as Charcot-Marie-Tooth type 1A (CMT1A) disease and *Trembler (Tr)* mouse respectively [7]. Individuals presenting nonsyndromic recessive deafness (autosomal recessive deafness DFNB29) were recently shown to harbor mutations in the gene encoding claudin-14 [15]. The Xp11.4 region of the human genome which comprises the BCMP1 gene has been linked to several forms of syndromic X-linked mental retardation, such as MRXS-2, -4, -6 and -10, and to a number of non-syndromic MRX cases [16]. The TM4SF2 gene which apparently encodes another member of the superfamily of four transmembrane proteins, a tetraspanin [17] more distantly related to BCMP1 than are PMP22/EMPs and the claudins, is located very close to the BCMP1 gene in man. Mutations in the TM4SF2 gene and gene inactivation resulting from chromosomal translocation have been shown to be involved in several cases of X-linked mental retardation [18]. Whether the BCMP1 gene is also involved in such genetic disorders and what is the function of the encoded protein thus constitute the obvious questions which will support our future investigations.

Materials and methods

DNA constructions

Standard DNA manipulations were conducted according to published procedures [19]. The full length BCMP1 cDNA clone was obtained by screening a dog thyroid cDNA library in λ ZAPII phage vector [2] using the original clone C60 [1] as probe. The DNA sequences corresponding to the cDNA insert in clone C60 were amplified by PCR using primers complementary to the sequences flanking the insert in the construct, 5'CAGATCTCGACCCACGCG^{3'} and 5'TACCTGCGGCCGCGATAT^{3'} respectively, and were labeled with digoxigenin (DIG labeling and detection kit, Boehringer Mannheim). Hybridization, washing and signal detection were performed as recommended by the supplier of the labeling system. The cloned DNA was sequenced on both strands using the Big Dye Terminator methodology and a model 377 DNA sequencer (Applied Biosystems). The construct encoding the EGFP-BCMP1 fusion protein was obtained by inserting a PCR fragment corresponding to the BCMP1 ORF between the EcoRI and BamHI sites in the pEGFP-C1 vector (Clontech). The following primers were used to amplify these sequences from the DNA of clone

C60: 5'TTCGAATTCGGCGGGCAGCGGC^{3'} and 5'TGTGGATCCTAGTAGTAGTCTTC^{3'}.

RNA analysis

Northern blot analysis was performed on 4 μ g of polyA+ mRNA from various dog tissues. Acridine orange staining of the gel confirmed that each lane contained identical amounts of RNA. A ³²P-labeled PCR fragment corresponding to the BCMP1 ORF was used as probe (see above for preparation of the DNA fragment). Hybridization and washes were conducted in standard conditions in the presence of 50 % formamide [19].

Cell transfection

Transfection of COS-7 cells was performed using the DEAE-dextran method [20]. About 200 ng of a crude plasmid DNA preparation was engaged per dish (diameter: 3 cm). The subcellular localization of EGFP fluorescence was observed 48 hours after transfection using an Eclipse TE300 inverted microscope (Nikon) equipped with NB-2A and NG-2A filter blocks. The transfected cells were permeabilized using saponin (0.075% final concentration) and nuclear DNA was stained with ethidium bromide (1 μ g/ml final concentration) in order to visualize the cell nucleus.

Chromosomal localization

Dog BCMP-1 could be readily typed on the dog x hamster radiation hybrid panel RHDF5000-2 composed of 118 cell lines from panel RHDF5000 [10]. The following pair of primers, 5'TCTGGAGTGAACAAATGGGCTAA^{3'} and 5'GCAGTCTGAGATTAGTGGCAAA^{3'} generated a PCR product of 137 bp on dog genomic DNA. PCR results were scored in terms of present, absent or ambiguous in the 118 hybrid cell lines. The typing data were incorporated into the latest radiation hybrid map [11], using the Multimap package [21]. The GeneBridge 4 human x hamster radiation hybrid panel DNA (Research Genetics Inc.) was screened by PCR using the following primers: 5'GGCAGCGGCATCCAGGAA^{3'} and 5'TGGGGAAGACCAACAGAGAACC^{3'}. The PCR results were analyzed according to the prescription of the supplier of the panel DNA.

The panel of rat x hamster radiation cell hybrids [12] was typed by PCR with the following primers: 5'-AACTGTGAATACCAATCTAAGT-3' and 5'-GTTTTTCATTATGCAGTTACAG-3'. The mapping results were obtained from the rat radiation hybrid map server at the Medical College Wisconsin [(http://rgd.mcw.edu/RH-MAPSERVER/)].

Bioinformatics

Sequences comparisons were performed using the BLAST tool ([http://www.nc-

bi.nlm.nih.gov/BLAST]). Protein sequences alignments and phylogenetic tree were constructed using Clustal X (Ver. 1.8) and TreeView (Ver. 1.6.1) respectively. Structural predictions based on protein sequences were obtained using programs available on the ExPasy server ([http://www.expasy.ch]). Localization of the BCMP1 gene in the human draft genome sequence was achieved using data and tools available on the Ensembl server ([http://www.ensembl.org]).

Acknowledgements

We thank Drs J.E. Dumont and G. Vassart for their continuous support and critical interest in our work. We are indebted to C. Govaerts for sequence alignments and phylogenetic tree construction, to the technicians of the Service de Génétique Médicale de l'Hôpital Erasme for the realization of DNA sequencing runs and to P. Van Vooren for technical help. This work was supported by the Belgian program "Poles d'Attraction Interuniversitaire" (Prime Minister's Office, Science Policy Programming) and the Fonds National de la Recherche Scientifique (FNRS, FRSM). R.G. is supported by funds from the CNRS and American Kennel Club. C.S. and D.C. are Research Directors of the FNRS.

References

- Pichon B, Mercan D, Pouillon V, Christophe-Hobertus C, Christophe D: **A method for the large-scale cloning of nuclear proteins and nuclear targeting sequences on a functional basis** *Anal Biochem* 2000, **284**:231-239
- Wilkin F, Savonet V, Radulescu A, Petermans J, Dumont JE, Maenhaut C: **Identification and characterization of novel genes modulated in the thyroid of dogs treated with methimazole and propylthiouracil** *Biol Chem* 1996, **271**:28451-28457
- Kozak M: **Determinants of translational fidelity and efficiency in vertebrate mRNAs** *Biochimie* 1994, **76**:815-821
- Chen C-YA, Shyu A-B: **AU-rich elements: characterization and importance in mRNA degradation** *TIBS* 1995, **20**:465-470
- Xu N, Chen C-YA, Shyu A-B: **Modulation of the fate of cytoplasmic mRNA by AU-rich elements: key sequence features controlling mRNA deadenylation and decay** *Mol Cell Biol* 1997, **17**:4611-4621
- Kruys V, Marinx O, Shaw G, Deschamps J, Huez G: **Translational blockade imposed by cytokine-derived UA-rich sequences** *Science* 1989, **245**:852-855
- Jetten AM, Suter U: **The peripheral myelin protein 22 and epithelial membrane protein family** *Prog Nucleic Acid Res Mol Biol* 2000, **64**:97-128
- Morita K, Furuse M, Fujimoto K, Tsukita S: **Claudin multigene family encoding four-transmembrane domain protein components of tight junction strands** *Proc Natl Acad Sci USA* 1999, **96**:511-516
- Attardi LD, Reczek EE, Cosmas C, Demicco EG, McCurrach ME, Lowe SW, Jacks T: **PERP, an apoptosis-associated target of p53, is a novel member of the PMP-22/gas3 family** *Genes & Dev* 2000, **14**:704-718
- Vignaux F, Hitte C, Priat C, Chuat J-C, Andre C, Galibert F: **Construction and optimization of a dog whole-genome radiation hybrid panel** *Mamm Genome* 1999, **10**:888-894
- Mellersh CS, Hitte C, Richman M, Vignaux F, Priat C, Jouquand S, Werner P, Andre C, DeRose S, Patterson DF, et al: **An integrated linkage-radiation hybrid map of the canine genome** *Mamm Genome* 2000, **11**:120-130
- Watanabe TK, Bihoreau MT, McMarthy LC, Kiguwa SL, Hishigaki H, Tsuji A, Browne J, Yamasaki Y, Mizoguchi-Miyakita A, Oga K, et al: **Aradiation hybrid map of the rat genome containing 5,255 markers** *Nature Genet* 1999, **22**:27-36
- Millwood I, Bihoreau MT, Gauguier D, Hyne G, Levy E, Kreutz R, Lathrop GM, Monaco A: **A gene-based genetic linkage and comparative map of rat X chromosome** *Genomics* 1997, **40**:253-261
- Kubota K, Furuse M, Sasaki H, Sonoda N, Fujita K, Nagafuchi A, Tsukita S: **Ca²⁺-independent cell-adhesion activity of claudins, a family of integral membrane proteins localized at tight junctions** *Curr Biology* 1999, **9**:1035-1038
- Wilcox ER, Burton QL, Naz S, Riazuddin S, Smith TN, Ploplis B, Belyantseva I, Ben-Yosef T, Liburd NA, Morell RJ, et al: **Mutations in the gene encoding tight junction claudin-14 cause autosomal recessive deafness DFNB29** *Cell* 2001, **104**:165-172
- Chiurazzi P, Hamel BCJ, Neri G: **XLMR genes: update 2000** *Eur J Hum Genet* 2001, **9**:71-81
- Maecker HT, Todd SC, Levy S: **The tetraspanin superfamily: molecular facilitators** *FASEB J* 1999, **11**:428-442
- Zemni R, Bienvenu T, Vinet MC, Sefiani A, Carrie A, Billuart P, McDonnell N, Couvert P, Francis F, Chafey P, et al: **A new gene involved in X-linked mental retardation identified by analysis of an X;2 balanced translocation** *Nature Genet* 2000, **24**:167-170
- Sambrook J, Fritsch EF, Maniatis T: **Molecular cloning: a laboratory manual** *Cold Spring Harbor, Cold Spring Harbor Laboratory Press* 1989
- German C: **High efficiency gene transfer into mammalian cells** *In: DNA Cloning, A Practical Approach vol 2 (Edited by Glover DM) Oxford, IRL Press* 1985:143-190
- Matisse TC, Perlin M, Chakravarti A: **Automated construction of genetic linkage maps using an expert system (MultiMap): a human genome linkage map** *Nature Genet* 1994, **6**:384-390

Publish with **BioMedcentral** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

editorial@biomedcentral.com