FEBS Letters 585 (2011) 606–612

journal homepage: www.FEBSLetters.org

Predicted expansion of the claudin multigene family

Katsuhiko Mineta [a,1], Yasuko Yamamoto [b,c], Yuji Yamazaki [b], Hiroo Tanaka [b], Yukiyo Tada [d], Kuniaki Saito [c], Atsushi Tamura [b], Michihiro Igarashi [d], Toshinori Endo [a], Kosei Takeuchi [d,1], Sachiko Tsukita [b,*,1]

[a] Graduate School of Information Science and Technology, Hokkaido University, Kita 14 Nishi 9, Kita-ku, Sapporo, Hokkaido 060-0814, Japan
[b] Graduate School of Medicine/Frontier Bioscience, Osaka University, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan
[c] Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan
[d] Graduate School of Medical and Dental Sciences, Niigata University, 1-757 Asahi-machi, Chuo-ku, Niigata, Niigata 951-8510, Japan

ARTICLE INFO

Article history:
Received 22 November 2010
Revised 12 January 2011
Accepted 17 January 2011
Available online 26 January 2011

Edited by Takashi Gojobori

Keywords:
Claudin
Tight junction
Epithelial cells
Paracellular barrier
Cell adhesion
PSI-BLAST

ABSTRACT

Claudins (Cldn) are essential membrane proteins of tight junctions (TJs), which form the paracellular permselective barrier. They are produced by a multi-gene family of 24 reported members in mouse and human. Based on a comprehensive search combined with phylogenetic analyses, we identified three novel claudins (claudin-25, -26, and -27). Quantitative RT-PCR revealed that the three novel claudins were expressed in a tissue- and/or developmental stage-dependent manner. Claudins-25 and -26, but not claudin-27, were immunofluorescently localized to TJs when exogenously expressed in cultured MDCK and Eph epithelial cell lines. These findings expand the claudin family to include at least 27 members.

Structured summary:
Claudin-25 and **ZO-1** colocalize by fluorescence microscopy (View interaction)
ZO-1 and **Claudin-26** colocalize by fluorescence microscopy (View interaction)

1. Introduction

An important milestone in elucidating the molecular basis of tight junction (TJ)-based paracellular barriers was the identification in 1998 of the claudins (Cldn), a family of proteins with four transmembrane domains [1]. This finding established the field of "Barriology", the science of barriers in multicellular organisms [2,3]. The 24 known claudin genes in human and mouse have highly variable tissue-specific expression patterns. Claudin proteins appear to be required for TJ formation and paracellular barrier functions [2–4]. The claudin-dependent selective permeabilities for ions and solutes differ, and the finely tuned expression ratio of particular claudins helps to control the homeostasis of specific tissues [5–9]. Thus, determining the cell- and tissue-specific combination of claudin-family members is critical for analyzing the function of TJs. Two problems need to be overcome: one is confusion in the nomenclature used to designate claudins 21–24 in human and mouse, and the other is that recently updated gene databases contain sequences that may reveal additional mouse and human claudin family members.

Claudin-1 and -2 were identified in 1998 [1]. Six more claudins, which localized to TJs when exogenously expressed in cultured epithelial cells, were identified in 1999 [10]. Subsequent research focused on database analyses for comprehensive surveys of this multi-gene protein family, with less regard for their subcellular localization; however, there is confusion in the assignment of numbers to certain claudin sequences. In a database search in 2001, 24 sequences for claudin family members were found [2]. At that time, claudins 17–24 had not been characterized well. In 2003, the sequences of human claudins 21–24 were reported by Katoh and Katoh [11]. Then, in 2009, Lal-Nag and Morin [12] assigned claudin-21, -22, and -24 to different sequences, in a paper designating all 24 claudin members and their homologues in human, chimpanzee, rat, and mouse. Consequently, there are inconsistencies in the number assignments of these three claudins between Katoh and Katoh [11] and Lal-Nag and Morin [12].

Moreover, new claudin-family members have been predicted. Wu et al. [13] predicted three additional putative claudin proteins by an iterated PSI-BLAST (Position Specific Iterated Basic Local Alignment Search Tool) search [14] that started with the PMP22_Claudin domain (pfam; PF00822) in 2006, although they did not name them as numbered members of the claudin family. These predicted proteins have been replaced from the database updated.

Abbreviations: TJ, tight junction; AJ, adherens junction; Cldn, claudin

* Corresponding author.
E-mail address: atsukita@biosci.med.osaka-u.ac.jp (S. Tsukita).

[1] These authors co-correspond to the present study.

0014-5793/$36.00
doi:10.1016/j.febslet.2011.01.028

In the present manuscript, to update the claudin family and search for unknown claudin-family members from the latest genomic resources, we looked for distinctions between the claudin family and the PMP-22/MP20/EMP family, which share a basic structural domain. To search for any unknown members of the claudin family from a new perspective, we also considered evolutionarily conserved domains in the sequences. By combining bioinformatics and evolutionary analyses, we identified three novel members of the claudin family and conclude that the claudin multi-gene family consists of at least 27 members in mouse and human.

2. Materials and methods

2.1. Multiple alignment and phylogenetic analysis

The amino acid sequences of known claudins and predicted claudin candidates were aligned by CLUSTAL X [15] under default parameters. Poisson-corrected amino acid distances were used as an amino acid substitution model based on their positions without any gaps in the sequences. The phylogenetic tree of the claudins was reconstructed using the neighbor-joining (NJ) method [16]. Reliability of the topology was examined by the bootstrap method [17], which generated the bootstrap probability by 1000 pseudo-replications at each interior branch in the tree. As an outgroup sequence for the phylogenetic tree, we used the amino acid sequence of human and mouse PMP20 (RefSeq ID: NP_000295 and NP_032911, respectively), a member of the PMP-22/EMP/MP20/Claudin family (Interpro ID: IPR004031).

2.2. Determination of mRNA expression levels using quantitative real-time RT-PCR

To analyze the developmental stage- and/or tissue-specific expression of novel claudin-family members in mouse, the first-strand cDNAs from mouse adult tissues (Mouse MTC™ panels I and III) (Clontech) and four mouse fetal tissues were used as templates. The detailed methods are described in the Supplementary materials and methods.

2.3. Expression vector constructs and cell culture

According to the procedure described by Yamazaki et al. [18], the cDNAs for novel mouse claudins were amplified by PCR from the mouse testis cDNA, and cloned into the T-easy vector (Promega). They were subcloned into the CAGGS-N-Venus-tagged expression vector. For transfection, MDCK I, II and Eph4 cells were cultured in DMEM containing 10% FCS. Transfection was performed using Lipofectamine Plus (Gibco BRL) following the manufacturer's protocol. To establish stable transfectants, the transfected cells were replated and incubated in the same medium containing 500 μg/ml Geneticin (Gibco BRL).

2.4. Immunofluorescence microscopy

Cells were cultured for 3 days on Transwell filters, 12 mm in diameter, to full confluence. The cells were fixed in ice-cold methanol for 5 min and processed for immunofluorescence microscopy [19].

3. Results and discussion

3.1. Basic strategy for finding claudin candidates by sequence identity and evolutionary relationship

We first found some inconsistencies among the amino acid sequences of reported claudins [11,12] and/or those described in the RefSeq database [20] (Table S1). To clarify the nomenclature for claudin-21 to -24 in the present manuscript, we assigned their sequences according to the report of Lal-Nag and Morin [12]. Then, the amino acid sequences of all the known human and mouse claudins were obtained from the NCBI RefSeq database (Table S2). The list provided in Lal-Nag and Morin's paper [12] showed only the ENTREZ gene IDs [21], so we connected these gene IDs and the RefSeq protein IDs manually. For genes that showed transcriptional variants in the database, we chose the longest ones for subsequent analyses.

Considering the sequence identities and their evolutionary conservation, we used the following strategy to identify candidates. First, the amino acid sequence of human claudin-1 was used as a query, and a homology search was performed against the RefSeq database release 32 (as of December, 2008), restricted to human by the PSI-BLAST program with default parameters. We then picked up claudins, epithelial membrane proteins, and unidentified proteins from the PSI-BLAST output and performed an iterative search four times until no novel sequences were included in the output. We constructed a multiple alignment and phylogenetic tree for the sequences obtained by PSI-BLAST. From the phylogenetic tree, we selected only the sequences that clustered with the known claudins. In this process, we excluded transcriptional variants, isoforms, and alleles by examining the alignments and their annotations. The remaining sequences were considered claudin-family candidates. By this procedure, we obtained three candidates for novel claudins.

3.2. Phylogenetic relationships between known claudins and the new candidates

We reconstructed the phylogenetic tree for claudins to include the previously identified and novel human and mouse claudins (Fig. 1). Most of the human and mouse orthologous pairs clustered together (except for mouse claudin-13, which does not have a human homolog). In addition, claudin-22 and -24 were unusually clustered, suggesting that their relationship may be paralogous rather than orthologous. Since the bootstrap values were relatively low in some of the ancestral nodes, it is difficult to discuss the phylogenetic relationships in detail. Nevertheless, the phylogenetic tree indicated that the proteins in the claudin cluster were closely related to each other compared with the outgroup protein, PMP22. Furthermore, the candidates for the novel claudins were positioned at the outer portion of the clade of previously identified claudin family members but closer to the claudin clade than to the outgroup protein, in the phylogenetic tree. Hence, it was most likely that these three novel candidates were evolutionarily related to the claudin family. We designated the three candidates for novel claudins as A, B, and C, in the order of their phylogenetic bifurcation (Fig. 1 and Table 1).

3.3. Designations of the three novel claudin candidates

The three candidates for novel claudins were identified by the PSI-BLAST search combined with phylogenetic tree reconstruction (Fig. 1 and Table 2). A search against the mouse RefSeq database by BLASTP revealed potential mouse orthologues for all the human claudin candidates. Candidate A contained 276 aa and was designated, "claudin domain-containing 1 protein isoform b" in the RefSeq database (Table 2).

Candidate B originally consisted of 420 aa and was designated as a "hypothetical" protein, but after the replacement of the predicted gene model, it was 223 aa long and was called, "transmembrane protein 114." The protein identification number for candidate B was changed in the refined RefSeq database, from XP_001716111 to NP_001139808 (see Table 1) as of August

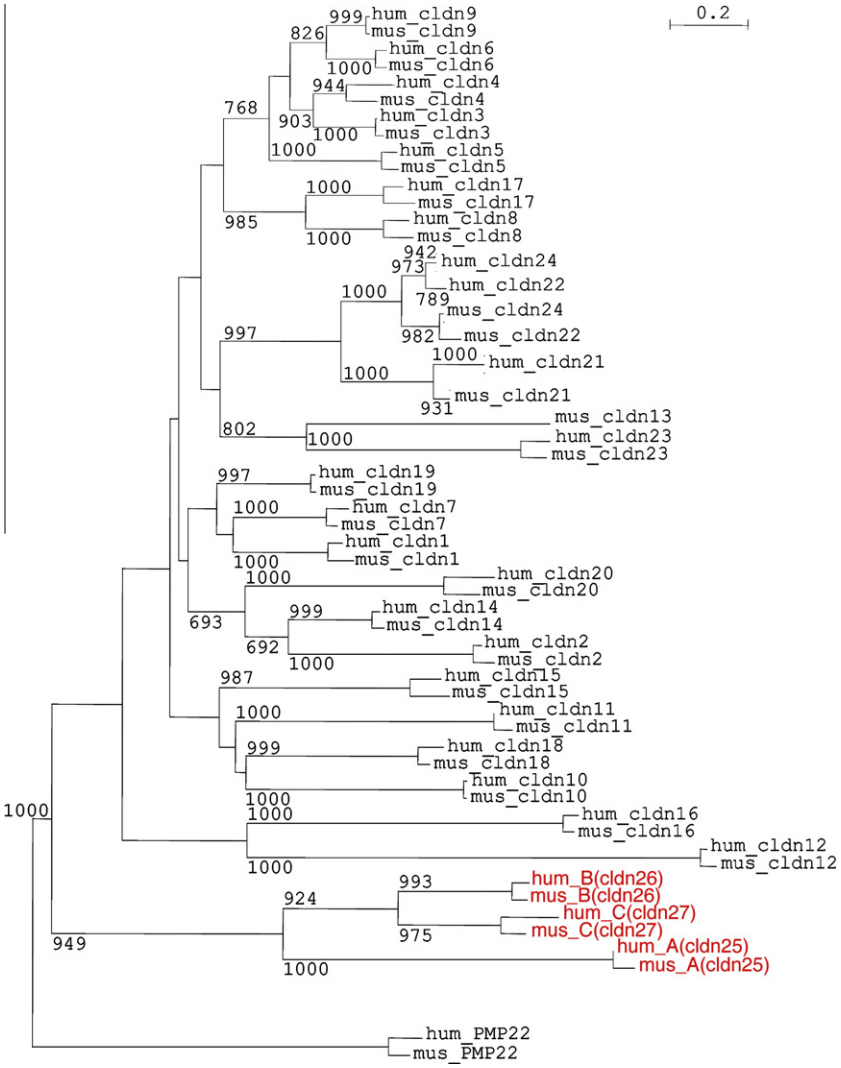

Fig. 1. Phylogenetic tree of the human and mouse claudin family. The tree was constructed by the neighbor-joining method. Numbers on branches represent the bootstrap confidence values (based on 1000 replications) supporting that branch; only values ⩾60% are presented here. The PMP22 sequence was used as an outgroup. The scale bar indicates the rate of amino acid substitutions per site. The taxon label hum_ or mus_ indicates the human and mouse sequence, respectively.

Table 1
Claudin candidates in human and mouse.

Candidate ID	Human	Ortholog in mouse
A	NP_001035272	NP_741968
B	NP_001139808	NP_083346
C	XP_946151	NP_001079004

2010, and in a previous study, part of it was described as showing a possible claudin-like characteristic but lacked a typical claudin-motif [13].

Candidate C originally contained 223 aa and was designated as, "similar to hCG1776376," but after replacement of the predicted gene model, its length was extended to 320 aa, and it was called, "transmembrane protein ENSP00000364084." The protein identification number was also recently changed in the refined RefSeq database as of August 2010, from XP_946151 to NP_001139001. However, the replacement protein (NP_001139001) did not show any homologous sequences in other organisms by a BLASTP search against the nr database of NCBI (data not shown). On the other hand, the original protein model (XP_946151) had homologous sequences in other organisms, such as mouse, and we confirmed the

Table 2
Assignments of claudin IDs.

Our proposal	Candidate ID	NCBI gene[a]	Lal-Nag and Morin (2009)	Katoh and Katoh (2003)
CLDN21	–	*CLDN25*	*CLDN21*	*CLDN24*
CLDN22	–	*CLDN22*	*CLDN22*	*CLDN21*
CLDN23	–	*CLDN23*	*CLDN23*	*CLDN23*
CLDN24	–	*CLDN24*	*CLDN24*	*CLDN22*
CLDN25	A	CLDND1 (claudin domain containing 1)	–	–
CLDN26	B	TMEM114 (transmembrane protein 114)	–	–
CLDN27	C	LOC283999 (transmembrane protein ENSP00000364084)	–	–

[a] As of August 2010.

expression of the mouse homolog (NP_001079004) of the original protein model, as described below. Moreover, the original protein was very similar to the sequence XP_211287, which was suggested by Wu et al. to be the most likely candidate for a novel claudin [13], showing one amino acid substitution and one deletion of 27 aa including a claudin motif region of our original protein model. Therefore, we used the original protein model, XP_946151, as the sequence for claudin candidate C.

Our findings collectively suggested that these three genes, which have novel full-length sequences, are new members of the claudin family.

3.4. Sequence characteristics and mRNA tissue expression of the novel claudin candidates

To explore the potential claudin function of the three candidates, we analyzed their sequences for typical claudin transmembrane regions and representative motifs. First, all the known claudins have four transmembrane regions [12]. The TMHMM [22] and SOSUI [23] programs both predicted that all three candidates were four-transmembrane proteins with N- and C- terminal tails facing the cytoplasm. Second, claudins share a common motif near the first extracellular loop, which has the amino acid sequence [GN]-L-W-x(2)-C-x(7,9)-[STDENQH]-C (PROSITE ID; PS01346) [4]. By performing multiple alignments, we confirmed that all of the claudin candidates possessed this motif near the first extracellular loop, although the motifs in the candidates were not exact matches for the PROSITE motif.

Next, to examine the mRNA expression of the candidates, we performed conventional and quantitative RT-PCR analyses, using cDNA libraries from various mouse tissues as the template. We detected mRNAs for candidates A, B, and C in E7 mouse embryos (Fig. 2A), and gave them proposed designations as human and mouse claudin-25, -26, and -27, respectively (Fig. 2A and Table 2). We then determined the relative expression levels of the mRNAs for claudin-25, -26, and -27 at different developmental time-points and in specific adult tissues (Fig. 2B and C). We also performed quantitative PCR for claudins 21 and 24, since their expression had not been validated yet [12]. We compared the data with that for claudin-3, which is ubiquitously expressed. The mRNA level for claudin-25 was higher from E7 to E17 than that for claudin-3, and the mRNAs for claudins 21, 24, 26, and 27 were relatively low—less than the mRNA for claudin-3. In adults, the mRNAs for claudins 24 and 25 were expressed at moderate levels in some tissues, and the levels of claudins 21, 26, and 27 were very low. In the comparison between embryonic (Fig. 2B) and adult tissues (Fig. 2C), claudins 21, 26, and 27 showed relatively high expression levels in embryonic stages, suggesting that they play roles during development.

From the informatics point of view, gene prediction methods mainly depend on experimental evidence such as large-scale EST data. Therefore, in the case of genes that are expressed at low levels (Fig. 2B and C), the actual gene region and/or ORF is not always as predicted. In this study, we predicted novel claudins based on informatics and evolutionary analyses. The alignment of novel claudins 21–27 shows high conservation throughout the entire length of the proteins (Fig. 3A). Predicted models of the membrane-spanning compositions are shown in Fig. 3B. These models clearly show that the first extracellular loops are larger than the second, a characteristic of the claudin family. We believe our

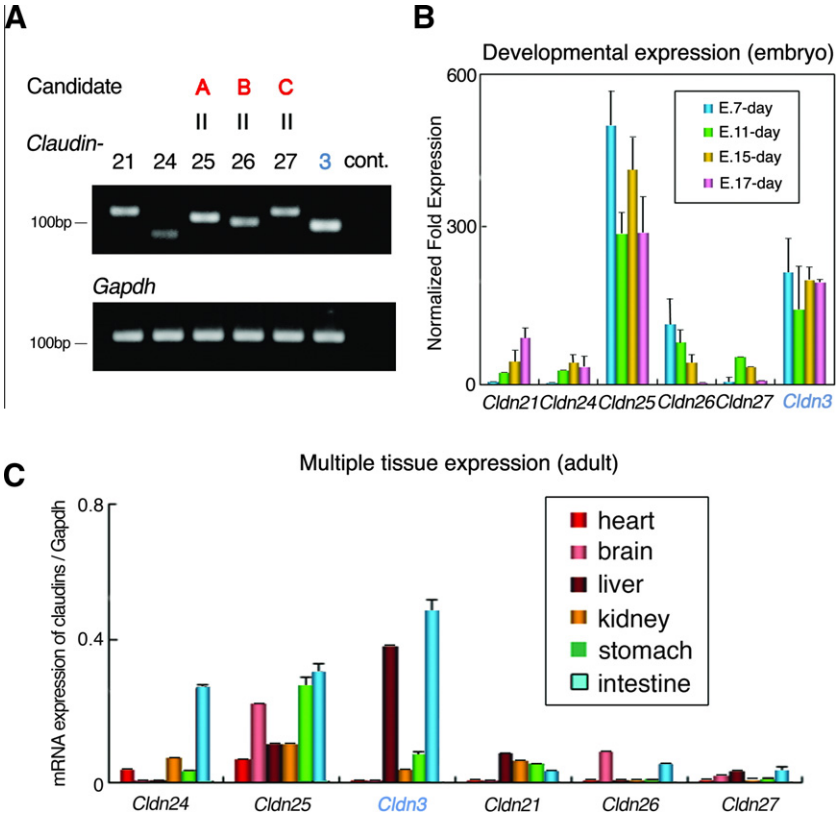

Fig. 2. Expression levels of the mRNAs for the novel claudin candidates as well as claudin-21 and 24 in adult and fetal tissues, determined by quantitative RT-PCR. (A) Gel-electrophoresis patterns of the conventional PCR products from E7 mouse embryo DNA. (B) qPCR results of claudin expression in mouse embryos at various developmental stages. (C) qPCR results of claudins expressed in various organs of adult mice. Cldn: claudin.

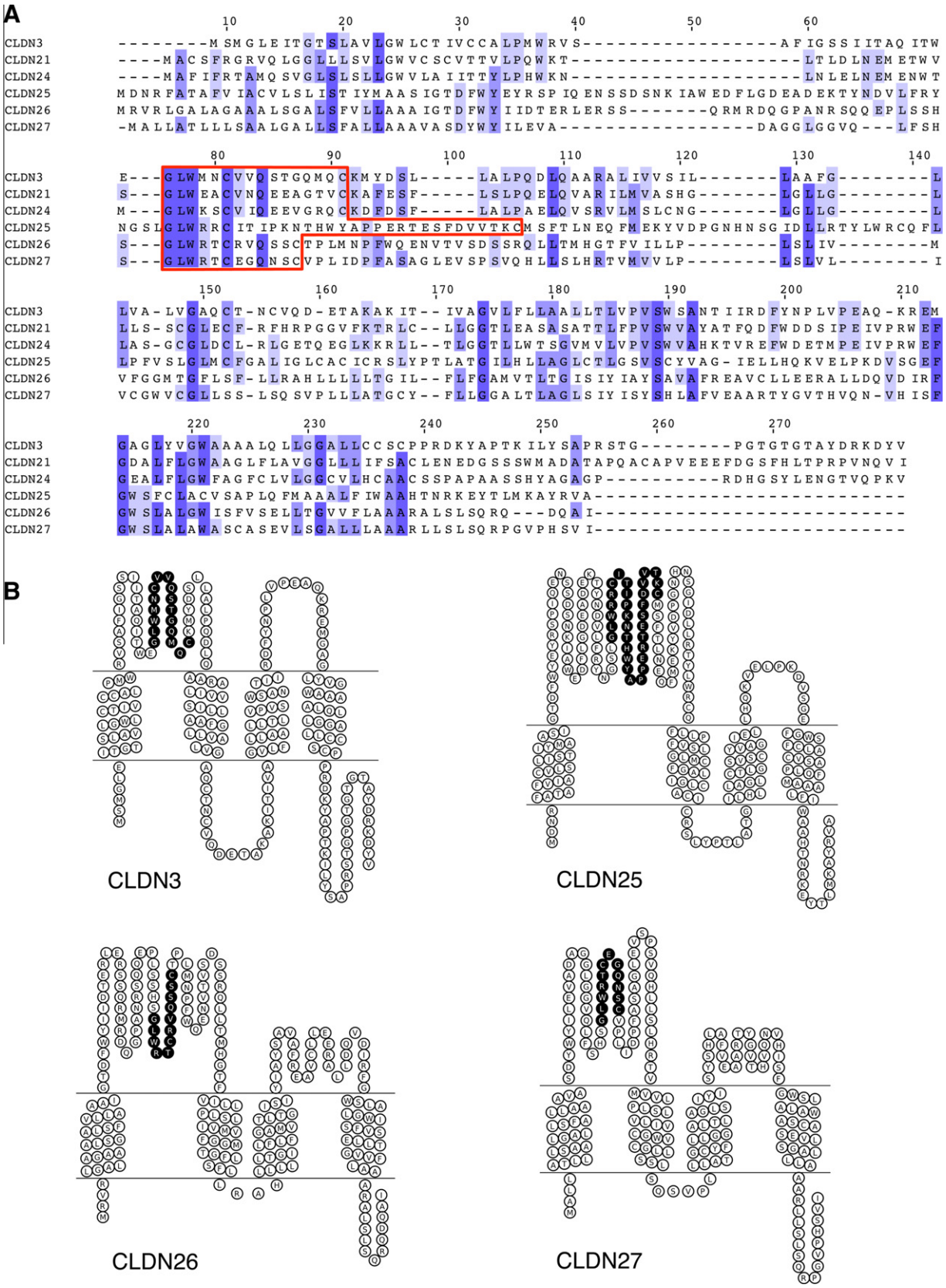

Fig. 3. Sequences of novel claudin members. (A) Multiple alignment of claudins-3, 21, 24, 25, 26, and 27. Blue shading sites indicate more than 40% identity of amino acids, visualized by Jalview [25]. The region in the red box shows the potential PROSITE motif (PS01346). In the case of imperfect motif matching, the potential motif region expanded to the second cysteine residue. The amino acid position is shown at the top. (B) Predicted two-dimensional topology models of claudin-3, 25, 26, and 27, visualized by TOPO2 with manual modifications [26]. The upper part of each figure indicates the extracellular region; the bottom part indicates the cytoplasmic region. The amino acids in the potential motif region are highlighted as black circles. The prediction of transmembrane segments was performed using the TMHMM program [22].

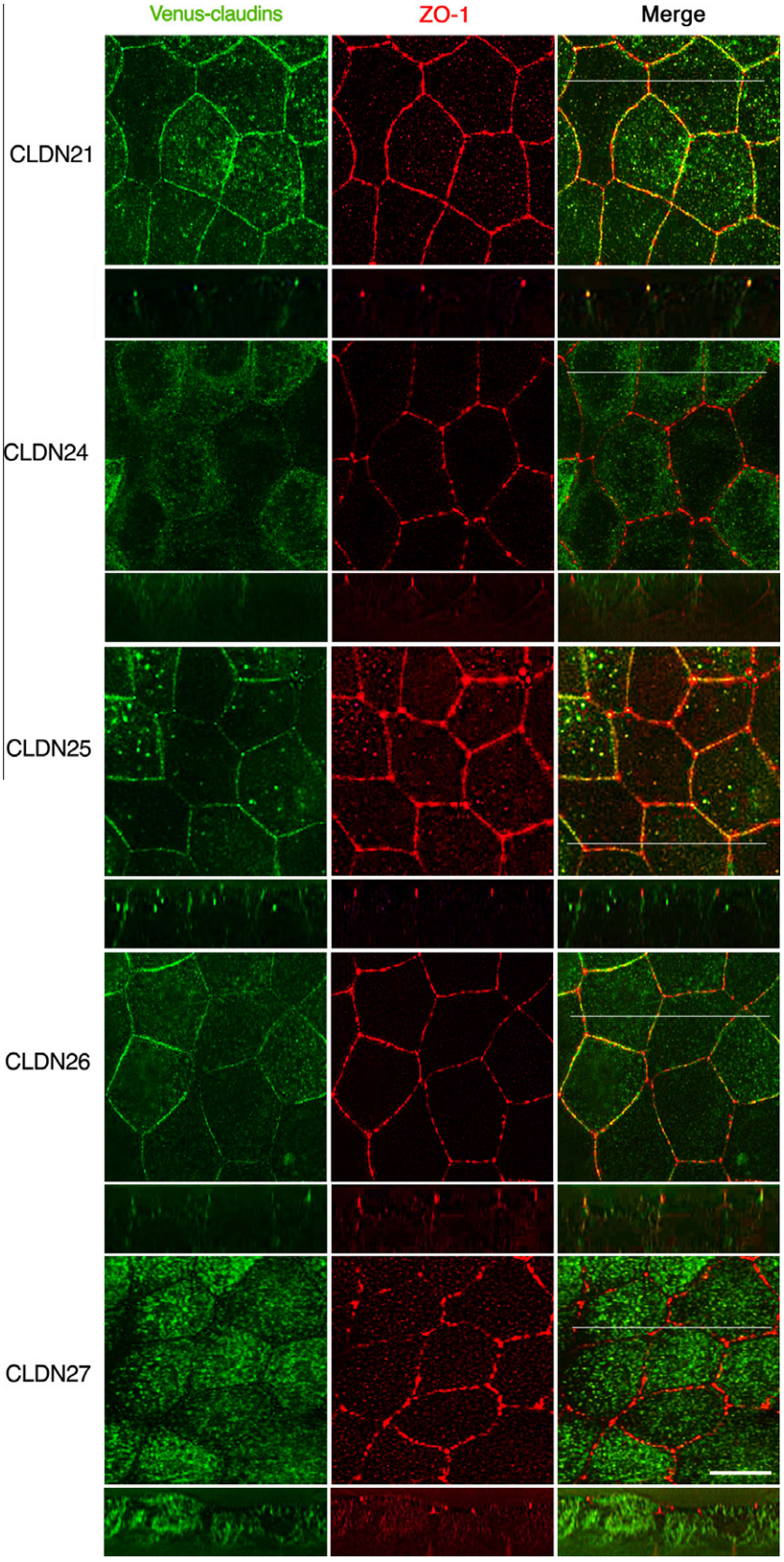

Fig. 4. Immunofluorescent micrographs of novel claudins with ZO-1 as a marker for tight junctions.Transfectants were double stained with an anti-GFP mAb (left panel) for exogenously expressed venus-claudins and an anti-ZO-1 mAb (middle panel). The stainings for claudin-21, -25, and -26 overlapped with that for ZO-1. Claudin-24 and -27 were localized to the apical membranes. Note that substantial amounts of claudin-21 and -25 were localized to cytoplasmic compartments. Z-stacked images are shown below each X-Y image. The position of the Z-stacked image is indicated by the white line in the X-Y image. Green: Venus-claudin. Red: ZO-1. Scale bar: 10 μm.

accomplishment was a result of the continuous effort to improve the gene models in the public databases. From the collective data, it is safe to say that claudins 25, 26, and 27 are informatically identifiable as new members of the claudin family that are expressed in mouse and human.

3.5. Cellular localization of novel claudin-25, -26, and -27

Since the mRNA expressions of the novel claudin-25, -26, and -27 were confirmed by quantitative RT-PCR, we cloned their full-length cDNAs by RT-PCR, using primers optimized for this purpose. Expression vectors were then constructed by inserting the full-length ORF sequences of claudin-25, -26, and -27. Each of the expression vectors, tagged with Venus or HA, was transfected into several lines of epithelial cells, including MDCK I, MDCK II, and Eph4 cells, and stable transfectants were established. After the cells were cultured on a filter cup to reach confluence and were well-polarized, immunofluorescent staining was performed. The fully polarized epithelial cells were singly stained for each claudin (with an anti-GFP antibody) or doubly co-stained for the claudin and ZO-1, a TJ marker (Fig. 4), or E-cadherin, an adherens junction (AJ) marker (Fig. S1). The signals for claudin-25 and -26 colocalized precisely with the ZO1-signal, but were separated from the E-cadherin-signal in MDCK I cells [2,18]. In contrast, claudin-27 remained in the cytoplasm, and was not targeted to the plasma membrane. As shown in Fig. 4 and Fig. S1, a substantial amount of claudin-24 and -27 was immunofluorescently localized to cytoplasmic dot-like structures, most possibly vesicular structures. These observations indicated that claudin-25 and claudin-26, but not claudin-27, are localized to TJs in cultured epithelial cell lines.

The localization of candidate claudins to TJs is important for defining them as claudin-family members. However, the situation is complicated, because TJs include complex combinations of claudins, depending on the cell type (with the exception of Schwann cells, which express only claudin-19 in their TJs) [24]. Furthermore, the TJ localization of claudin is not consistent across all cell types and all claudins. For example, when we immunostained epithelial cells of the small intestine for claudin-4, the signals were detected in the enterocytes in a typical tight-junctional pattern, but in goblet cells, the entire cell membrane, including the tight junctional and lateral membranes, was labeled. Moreover, a previously identified claudin, claudin-24, remained in the cytoplasm rather than being targeted to the plasma membrane, when transfected into cultured epithelial cells (Fig. 4 and Fig. S1).

The subcellular localization of the claudin family members remains to be explored in future studies, along with the claudin-subtype specificities of various cell types and tissues. Nevertheless, it is reasonable to conclude from the database analyses performed in this study that the claudin family is composed of at least 27 members. All of the claudin sequences, including the three novel ones, should be useful for future studies of claudins aimed at understanding their fundamental cellular functions and potential medical applications.

Acknowledgement

This study was supported in part by a Grant-in-Aid for Creative Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), to Sachiko Tsukita, and by a Grants-in-Aid for Scientific Research from MEXT and for Research and Development Program for New Bio-industry Initiatives, to Kosei Takeuchi.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2011.01.028.

References

[1] Furuse, M., Fujita, K., Hiiragi, T., Fujimoto, K. and Tsukita, S. (1998) Claudin-1 and -2: novel integral membrane proteins localizing at tight junctions with no sequence similarity to occludin. J. Cell Biol 141, 1539–1550.
[2] Tsukita, S., Furuse, M. and Itoh, M. (2001) Multifunctional strands in tight junctions. Nat. Rev. Mol. Cell Biol. 2, 285–293.
[3] Tsukita, S., Yamazaki, Y., Katsuno, T., Tamura, A. and Tsukita, S. (2008) Tight junction-based epithelial microenvironment and cell proliferation. Oncogene 27, 6930–6938.
[4] Van Itallie, C.M. and Anderson, J.M. (2006) Claudins and epithelial paracellular transport. Annu. Rev. Physiol. 68, 403–429.
[5] Colegio, O., Van Itallie, C., McCrea, H., Rahner, C. and Anderson, J.M. (2002) Claudins create charge-selective channels in the paracellular pathway between epithelial cells. Am. J. Physiol. Cell Physiol. 283, C142–C147.
[6] Rosenthal, R., Milatz, S., Krug, S.M., Oelrich, B., Schulzke, J.D., Amasheh, S., Gunzel, D. and Fromm, M. (2010) Claudin-2, a component of the tight junction, forms a paracellular water channel. J. Cell Sci. 123, 1913–1921.
[7] Weber, C.R., Raleigh, D.R., Su, L., Shen, L., Sullivan, E.A., Wang, Y. and Turner, J.R. (2010) Epithelial myosin light chain kinase activation induces mucosal interleukin-13 expression to alter tight junction ion selectivity. J. Biol. Chem. 285, 12037–12046.
[8] Van Itallie, C.M., Fanning, A. and Anderson, J.M. (2003) Reversal of charge selectivity in cation or anion-selective epithelial lines by expression of different claudins. Am. J. Physiol. Renal. Physiol. 285, F1078–F1084.
[9] Yu, A. (2009) Molecular basis for cation selectivity in claudin-2-based pores. Ann. N.Y. Acad. Sci. 1165, 53–57.
[10] Morita, K., Furuse, M., Fujimoto, K. and Tsukita, S. (1999) Claudin multigene family encoding four-transmembrane domain protein components of tight junction strands. Proc. Natl. Acad. Sci. USA 96, 511–516.
[11] Katoh, M. and Katoh, M. (2003) CLDN23 gene, frequently down-regulated in intestinal-type gastric cancer, is a novel member of CLAUDIN gene family. Int. J. Mol. Med. 11, 683–689.
[12] Lal-Nag, M. and Morin, P.J. (2009) The claudins. Genome Biol. 10, 235.1–235.7.
[13] Wu, J., Helftenbein, G., Koslowski, M., Sahin, U. and Tureci, O. (2006) Identification of new claudin family members by a novel PSI-BLAST based approach with enhanced specificity. Proteins 65, 808–815.
[14] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
[15] Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.
[16] Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
[17] Felsenstein, J. (1985) Confidence limit on phylogenies: an approach using the bootstrap. Evolution 39, 95–105.
[18] Yamazaki, Y., Okawa, K., Yano, T. and Tsukita, S. (2008) Optimized proteomic analysis on gels of cell–cell adhering junctional membrane proteins. Biochemistry 47, 5378–5386.
[19] Umeda, K. et al. (2006) ZO-1 and ZO-2 independently determine where claudins are polymerized in tight-junction strand formation. Cell 126, 741–754.
[20] Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35, D61–D65.
[21] Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2007) Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res. 35, D26–D31.
[22] Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.
[23] Mitaku, S., Hirokawa, T. and Tsuji, T. (2002) Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane–water interfaces. Bioinformatics 18, 608–616.
[24] Miyamoto, T. et al. (2005) Tight junctions in Schwann cells of peripheral myelinated axons: a lesson from claudin-19-deficient mice. J. Cell Biol. 169, 527–538.
[25] Waterhouse, A.M. et al. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191.
[26] Johns, S.J. TOPO2, Transmembrane protein display software. <http://www.sacs.ucsf.edu/TOPO2/>.