Interspecies Diversity of the Occludin Sequence: cDNA Cloning of Human, Mouse, Dog, and Rat–Kangaroo Homologues

Yuhko Ando-Akatsuka,* Mitinori Saitou,* Tetsuaki Hirase,* Masashi Kishi,‡ Akira Sakakibara,§ Masahiko Itoh,* Shigenobu Yonemura,* Mikio Furuse,* and Shoichiro Tsukita*

Departments of *Cell Biology and ‡Anatomy and Developmental Biology, Faculty of Medicine, Kyoto University, Sakyo-ku, Kyoto 606, Japan; and §Department of Physiological Sciences, School of Life Sciences, The Graduate University of Advanced Studies, Myodaiji-cho, Okazaki, Hichi 444, Japan

Abstract. Occludin has been identified from chick liver as a novel integral membrane protein localizing at tight junctions (Furuse, M., T. Hirase, M. Itoh, A. Nagafuchi, S. Yonemura, Sa. Tsukita, and Sh. Tsukita. 1993. *J. Cell Biol.* 123:1777–1788). To analyze and modulate the functions of tight junctions, it would be advantageous to know the mammalian homologues of occludin and their genes. Here we describe the nucleotide sequences of full length cDNAs encoding occludin of rat–kangaroo (potoroo), human, mouse, and dog. Rat–kangaroo occludin cDNA was prepared from RNA isolated from PtK2 cell culture, using a mAb against chicken occludin, whereas the others were amplified by polymerase chain reaction based on the sequence found around the human neuronal apoptosis inhibitory protein gene. The amino acid sequences of the three mammalian (human, murine, and canine) occludins were very closely related to each other (~90% identity), whereas they diverged considerably from those of chicken and rat–kangaroo (~50% identity). Implications of these data and novel experimental options in cell biological research are discussed.

OCCLUDIN is a ~65-kD integral membrane protein located at tight junctions (TJ).[1] It was first identified in chicken using monoclonal antibodies, and its cDNA was cloned and sequenced (Furuse et al., 1993). The protein comprises four transmembrane domains, a long carboxy-terminal cytoplasmic domain, a short amino-terminal cytoplasmic domain, two extracellular loops and one intracellular turn. One of the most characteristic aspects of its sequence is the high content of tyrosine and glycine residues in the first extracellular loop (~60%).

TJ are thought to play dual roles in the physiological functions of epithelial and endothelial cells by sealing them to create the primary barrier to the diffusion of solutes through the paracellular pathway and by working as a boundary between the apical and basolateral plasma membrane domains to create and maintain cell polarity (for reviews see Schneeberger and Lynch, 1992; Gumbiner, 1987, 1993). To clarify the molecular basis of—and to modulate—these functions, information on TJ proteins such as occludin would be important in cell biological as well as in medical research.

In freeze–fracture electron microscopy, TJ appear as a set of continuous, anastomosing intramembrane strands (Staehelin, 1973, 1974). Considering that occludin is one of the major components of these strands (Furuse et al., 1993, 1996; Fujimoto, 1995), it should provide a good experimental target for modulating TJ functions at the cellular as well as at the whole body level. However, the chicken is not an appropriate species for such studies, mainly because of the lack of a good cell culture system and so far still poor transgenic and gene knock-out animal techniques. As up to now occludin was known only in the chicken, and none of our mAbs and pAbs raised against chicken occludin crossreacted with the murine and human homologues (Furuse et al., 1993), several investigators, including ourselves, have tried to isolate cDNA encoding mammalian homologues, based upon the assumption that evolutionally the occludin amino acid sequence is rather conserved due to its functional importance. However, these experiments have not yet been successful until very recently. Here, we now report the nucleotide sequences of cDNAs encoding rat–kangaroo (potoroo), human, mouse, and dog occludin. The rat–kangaroo cDNA was isolated using one of our mAbs against chicken occludin, and the other cDNAs were amplified by PCR based on the "occlu-

Address all correspondence to Sh. Tsukita, Department of Cell Biology, Kyoto University Faculty of Medicine, Konoe-Yoshida, Sakyo-ku, Kyoto 606, Japan. Ph.: (81) 75-753-4372. Fax: (81) 75-753-4660. E-mail: htsukita@mfour.med.kyoto-u.ac.jp

1. *Abbreviations used in this paper:* TJ, tight junction; NAIP, neuronal apoptosis inhibitory protein.

din-like" sequence found around the human neuronal apoptosis inhibitory protein (NAIP) gene (Roy et al., 1995). We believe that this report will advance research on tight junction formation and function, especially as our data allow gene targeting experiments (in the mouse), the use of well-established functional cell culture systems such as the canine kidney epithelial cell line MDCK, and direct analyses of the corresponding human gene(s).

Materials and Methods

Isolation and Sequencing of Rat–Kangaroo Occludin cDNA

To isolate rat–kangaroo occludin cDNA, a λgt11 expression cDNA library was made from poly(A)$^+$ RNA purified from cultured PtK2 cells, using the TimeSaver™ cDNA synthesis kit (Pharmacia LKB Biotechnology, Uppsala, Sweden) and GIGAPACK II Packaging Extract (Stratagene, La Jolla, CA). The initial cDNA clone, pOc10 (3 kbp), was isolated from the library using chicken occludin mAb, Oc-2, as described (Huynh et al., 1985). The insert was then labeled using a DIG labeling kit (Boehringer-Mannheim Biochemicals, Indianapolis, IN) and used to screen the same cDNA library using a DIG detection kit (Boehringer Mannheim Biochemicals). The cDNA clones pOc8, pOc9-1, pOc9-2, pOc9-3, and pOc9-5 were isolated, and inserts were subcloned into pBluescript SK(−) and sequenced with the 7-deaza Sequenase Version Deoxy™ Terminator Cycle Sequencing Kit (Applied Biosystem, Foster City, CA).

To confirm the identity of this cDNA, pAbs were raised in rabbits against the GST fusion protein with the putative carboxy terminal cytoplasmic domain of rat–kangaroo occludin. The cDNA encoding this domain was obtained by PCR and introduced into the pGEX-2T vector (Pharmacia Fine Chemicals, Piscataway, NJ), to express fusion proteins in *Escherichia coli.*

Isolation and Sequencing of Human, Murine, and Canine Occludin cDNAs

We noted a nucleotide sequence with significant similarity to that of the carboxy terminal region of chicken occludin in deleted versions of the human NAIP gene (Roy et al., 1995). PCR was then performed using two oligonucleotides, TATGAGACAGACTACACAACTGGCGGCGAG-TCC and ATCATAGTCTCCAACCATCTTCTTGATGTG, as primers (see Fig. 1). A λgt11 cDNA library was made from poly(A)$^+$ RNA purified from the cultured human intestinal cell line, T84, using the TimeSaver™ cDNA synthesis kit and GIGAPACK II Packaging Extract, and used as template. A 363-bp cDNA fragment obtained by PCR was then labeled with DIG as described above and used to screen the same cDNA library. Three cDNA clones were isolated, and inserts of these clones were subcloned into pBluescript SK(−). Since inserts of two clones, phOc6 and phOc16, should contain the full ORF, both strands of these clones were sequenced. By the same PCR strategies, we isolated full-length cDNAs encoding murine and canine occludin. For this purpose, λgt 10 and λgt11 cDNA libraries were made from poly(A)$^+$ RNA purified from mouse lung and cultured dog kidney (MDCK) cells, respectively.

To confirm that these cDNAs encode human, murine and canine occludin, a mAb was raised in rats using the GST fusion protein with the putative carboxy terminal cytoplasmic domain of human occludin. To this end the cDNA encoding this domain was obtained by PCR and introduced into the pGEX-3X vector (Pharmacia Fine Chemicals) to express fusion proteins in *E. coli.*

Results

Rat–Kangaroo Occludin

Under our conditions, none of our mAbs and pAbs raised against chicken occludin showed immunofluorescent staining at tight junctions of mammalian cultured cells, but did react with the PtK2 cells, which were established from the kidney of a marsupial, the rat–kangaroo, and are often used to study mitosis and cytokinesis. Our mAb against chicken occludin, Oc-2, stained the cell–cell borders of PtK2 cells in contact, and recognized some bands around 60 kD on immunoblots (data not shown). Using mAb Oc-2, we then screened ~6×10^5 plaques from a λgt11 cDNA library made from PtK2 cells, and finally obtained a full-length cDNA encoding rat–kangaroo occludin, as described in Materials and Methods. Two criteria confirmed that this cDNA encodes the marsupial homologue of occludin. Firstly, the deduced amino acid sequence was similar to that of chicken occludin (Figs. 1–3). Secondly, pAbs raised against GST fusion proteins with the putative carboxy terminal domain produced in *E. coli* showed the same features as mAb Oc-2 in immunofluorescence microscopy and immunoblotting (data not shown).

The complete nucleotide sequence encoded by this cDNA and the deduced amino acid sequences are shown in Fig. 1. The reading frame of the sequence starts at nucleotide 76 and extends until nucleotide 1542, thereby encoding a protein of 489 amino acids with a molecular mass of 54 kD.

Human, Murine, and Canine Occludin

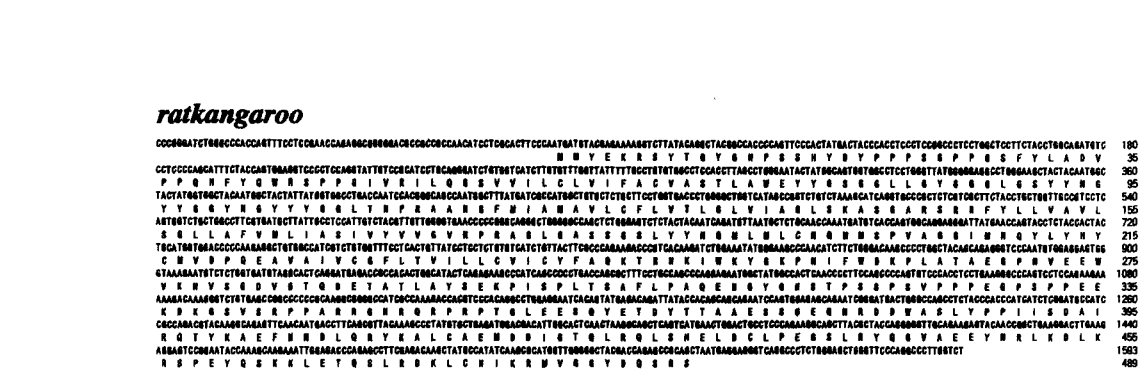
We produced a GST fusion protein with the cytoplasmic domain of rat–kangaroo occludin and raised rabbit pAbs against it. One of these antisera immunofluorescently stained the junctional complex region of the human intestinal epithelial cell line, T84. We then attempted to isolate cDNA clones encoding human occludin, using this pAb. During this study, we learned from the GenBank database, using a biological sequence search program, MPsrch (IntelliGenetics, Inc.), that a 675-nucleotide sequence showing similarity to a part of the carboxy terminal domain of chicken and rat–kangaroo occludin had been found in close proximity to the human neuronal apoptosis inhibitory protein gene (Roy et al., 1995). To determine whether or not this sequence really encodes part of the human homologue of occludin, we then performed PCR with two oligonucleotides (see Fig. 1) as primers, using the λgt11 cDNA library made from T84 cells as templates. We obtained a DNA fragment that allowed us to isolate a full-length cDNA encoding human occludin (see Materials and Methods). Its deduced amino acid sequence showed similarity to those of chicken and rat–kangaroo occludin (Fig. 3). Furthermore, mAbs raised against this gene product specifically stained tight junctions in T84 cells (Fig. 4). We then concluded that this cDNA encodes human homologue of occludin. The cDNAs encoding murine and canine occludin homologues were also isolated and sequenced by the same procedure, using mouse lung and MDCK cell cDNA libraries, respectively.

The complete nucleotide and amino acid sequences encoded by these cDNAs are shown in Fig. 1. They encode polypeptides of 522 (human) and 521 (mouse, dog) amino acids, with a molecular mass of 59 kD.

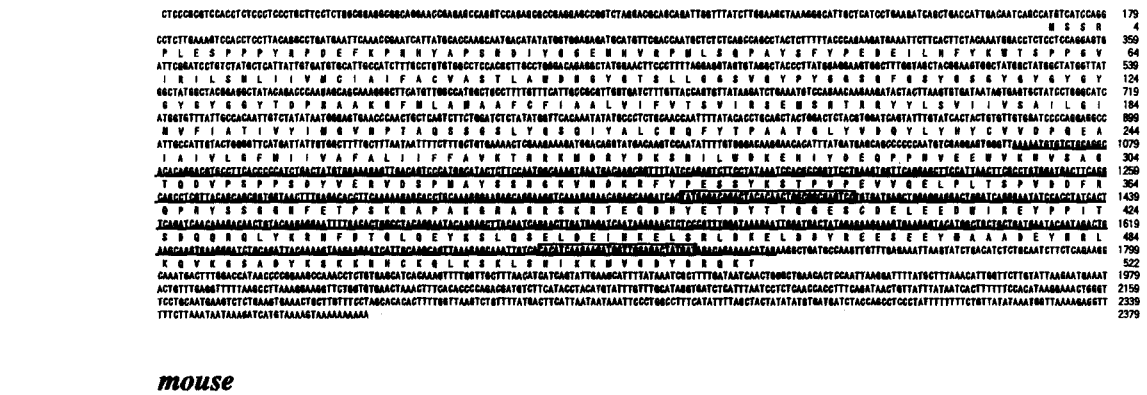
Discussion

In this study, we describe the cDNAs of rat–kangaroo, human, mouse, and dog occludin and the corresponding amino acids deduced therefrom. Hydrophilicity plots show

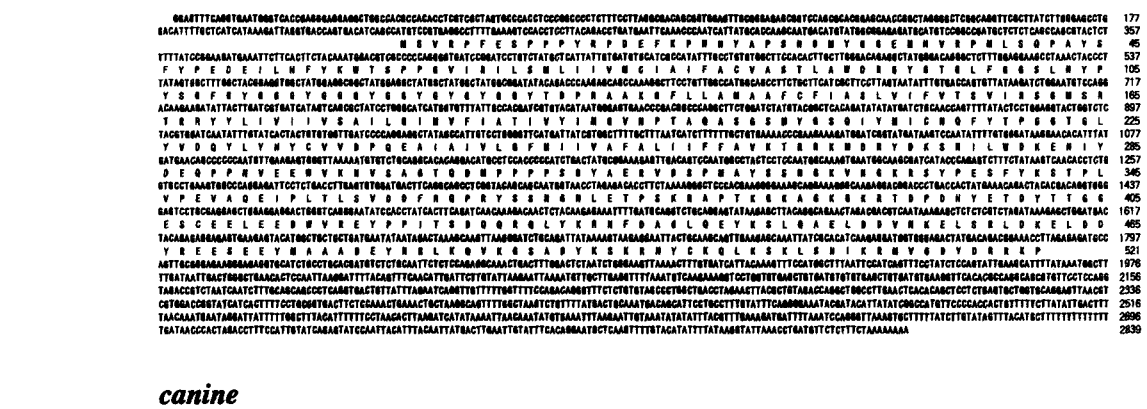
ratkangaroo



human



mouse



canine

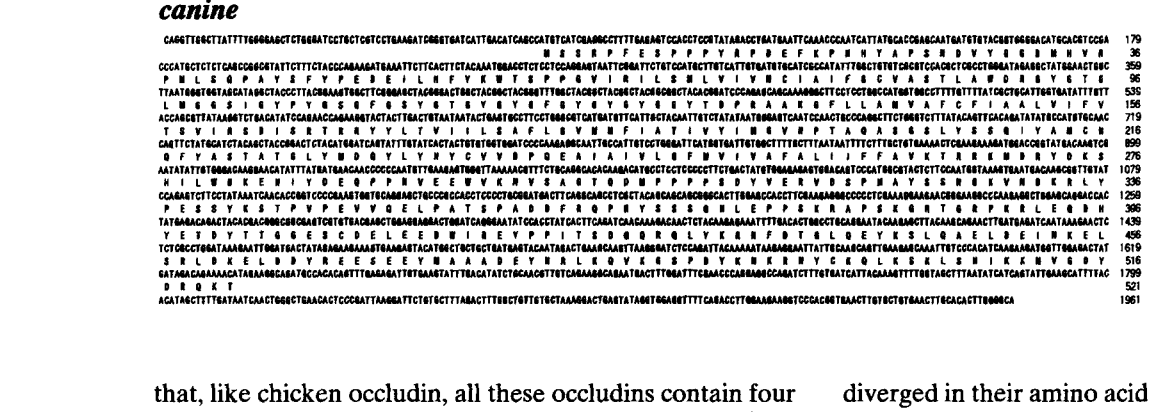


Figure 1. Nucleotide sequences of occludin cDNAs and deduced amino acid sequences of rat–kangaroo, human, murine, and canine occludins. The ORFs of these occludins encode 489, 522, 521, and 521 amino acids, respectively. The 675-nucleotide sequence found in close proximity to the human neuronal apoptosis inhibitory protein gene (Roy et al., 1995) is underlined, and two nucleotide sequences used for PCR are boxed. These sequence data are available from EMBL/GenBank/DDBJ under accession numbers U49183 (rat–kangaroo), U49184 (human), U49185 (mouse), and U49221 (dog).

that, like chicken occludin, all these occludins contain four transmembrane domains in their amino terminal half (Fig. 2). The amino acid sequences of human, murine, and canine occludin were highly homologous (~90% identity), and these sequences had considerably diverged from the avian (chicken) and marsupial sequence (Fig. 3). The percentages of identical amino acids between chicken and rat–kangaroo, rat–kangaroo and human, and human and chicken occludins were 58.4, 45.0, and 45.6%, respectively.

We considered possible reasons why chicken, rat–kangaroo, and human (mouse, dog) occludins have so strongly diverged in their amino acid sequences. Two possible functional domains have so far been identified in chicken occludin. One is the first extracellular domain, which may be involved in cell-cell interaction. This domain is characterized by a high tyrosine and glycine content (~60%) (Furuse et al., 1993), and this feature is conserved among all the occludins identified here (Fig. 3). In this domain of human occludin, for example, 29 out of 46–amino acid residues were tyrosine or glycine, suggesting that at least some glycine and tyrosine residues are involved in the specific function of this domain, i.e., probably tight cell–cell cou-

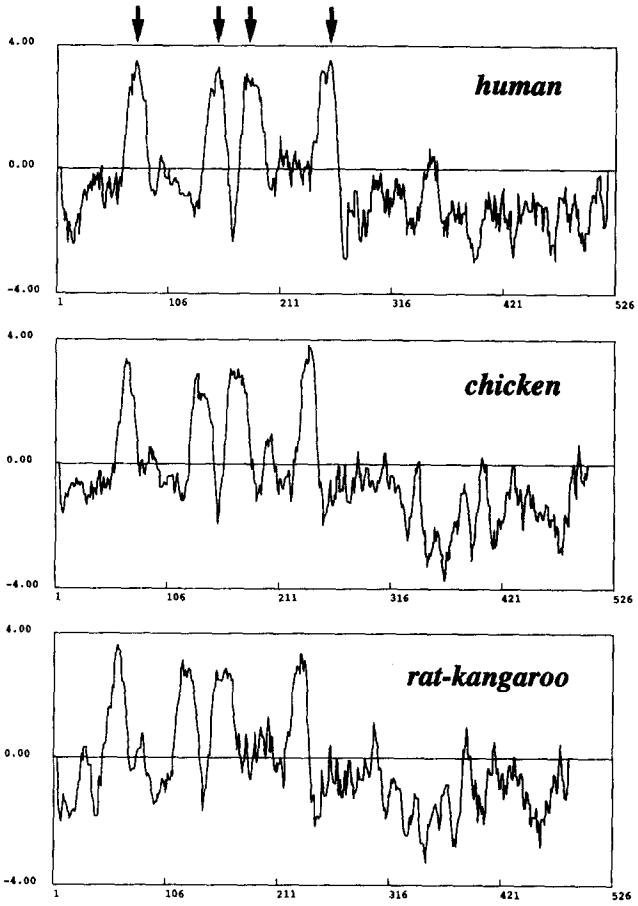


Figure 2. Hydrophilicity plots for occludin from human, chicken, and rat–kangaroo occludins using the Kyte and Doolittle program. The plot records the average hydrophilicity along the sequence over a window of 10 residues. Hydrophilic and hydrophobic residues are in the lower and upper part of the frames, respectively. The axis is numbered in amino acid residues. At the amino terminal half of each occludin, there are four major hydrophobic, potentially membrane-spanning regions (*arrows*).

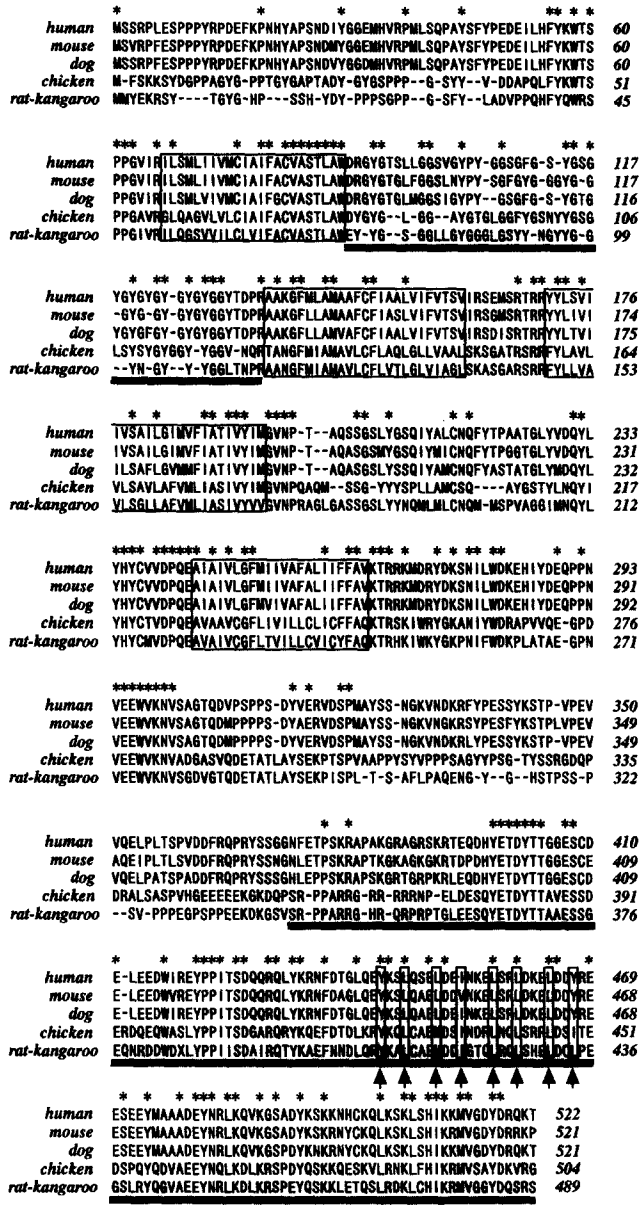


Figure 3. Comparison of amino acid sequences of occludins of human, mouse, dog, chicken, and rat–kangaroo aligned by the GENETYX program. The four transmembrane domains are boxed. Conserved amino acids shared by all occludins are indicated by asterisks; and gaps, introduced to maximize alignment, are indicated by dashes. The first extracellular domain (*serrated line*) is characterized by a high tyrosine and glycine content. The carboxy terminal protein ZO-1 binding domain is underlined. The amino acid sequence of this domain is rather conserved between species, and in the center of this domain hydrophobic amino acid residues are clustered in a pattern to allow the coiled-coil configuration (*arrows*).

pling. The second domain is the carboxy terminal 150 amino acids, which, at least in chicken occludin, is responsible for its association with ZO-1, a major plasmalemmal undercoat protein at TJ (Furuse et al., 1994). As shown in Fig. 3, the amino acid sequence of this domain is also remarkably conserved as compared to other domains. Conformation predictive analysis has further revealed that in all the different species studied the center of this domain can form a typical α-helical coiled–coil structure. This might suggest that the amino acid sequence of this domain has diversified during phylogenetic evolution only with the limitation of keeping an important segment in the coiled-coil structure to interact with ZO-1.

Occludin was first identified by producing mAbs in rats using isolated junctional fractions from the chick liver as the antigen-containing material (Furuse et al., 1993). When we, however, injected isolated rat liver junctional fractions into mice to obtain mAbs, occludin antibodies were not identified (Itoh et al., 1991; Tsukita et al., 1992, 1994). Retrospectively, these successes and failures can perhaps be explained by the diversity of the occludin sequence between chicken and rodents. On the identification of chicken occludin it was expected that this information would soon lead to answers to many of the key questions about the structure and functions of TJ at the molecular level (for review see Gumbiner, 1993). However, as mentioned in the Introduction, the necessary identification of mammalian occludin homologues proved rather difficult.

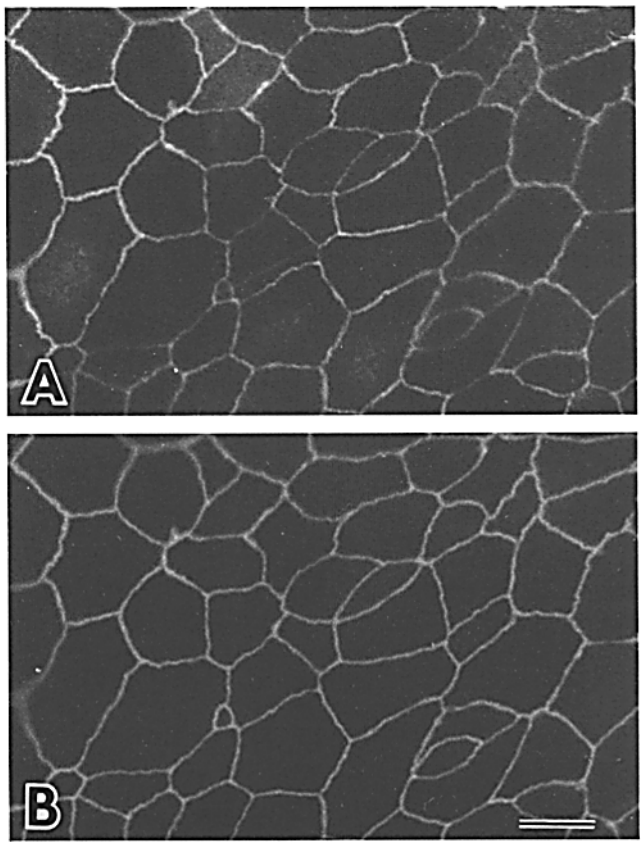


Figure 4. Double immunofluorescence staining of cultured human intestinal epithelial T84 cells with a rat mAb against human occludin HOC 119 (*A*) and a mouse mAb against a TJ-associated protein, ZO-1 (*B*). The human occludin mAb was raised against the GST fusion protein with the putative carboxy terminal cytoplasmic domain of human occludin. Both photographs were taken at the same apical focal plane of polarized T84 cells. Note that occludin and ZO-1 are precisely colocalized. Bar, 10 μm.

Now that this obstacle has been overcome, TJ organization and function can be structurally and functionally examined at the molecular level.

Using various types of cultured human, murine, and canine (MDCK) cells, the barrier and fence functions of TJ and the regulation mechanisms involved can be experimentally analyzed by modulating occludin gene expression or by blocking with anti-sense probes or with antibodies. For example, it can now be determined whether on overexpression of occludin cDNA the number of TJ strands, as seen in freeze–fracture replicas, will increase, with concomitant up-regulation of the barrier function. Through the production of various types of transgenic and occludin gene knock-out mice, we will learn how TJ formation is involved in the morphogenesis of various organs and whether TJ dysfunction is related to various pathological states such as inflammation and tumor metastasis. The possible modulation of TJ functions, especially its barrier function, is also interesting in relation to drug delivery. Thus, it should be possible to modulate the blood–brain barrier through up- or down-regulating occludin synthesis in brain endothelial cells. The modulation of TJ functions in intestinal epithelial cells is required to regulate the absorption of drugs from the intestine. Studies are currently being conducted along these lines in our laboratory.

We thank all the members of our laboratory (Department of Cell Biology, Faculty of Medicine, Kyoto University, Kyoto, Japan), especially Drs. A. Nagafuchi and Sa. Tsukita for their helpful discussions.

This work was supported in part by a Grant-in-Aid for Cancer Research and a Grant-in-Aid for Scientific Research (A) from the Ministry of Education, Science and Culture of Japan, and by research grants from the Yamada Science Foundation, the Mitsubishi Foundation, and the Toray Science Foundation (to Sh. Tsukita).

Received for publication 31 December 1995 and in revised form 22 January 1996.

References

Fujimoto, K. 1995. Freeze-fracture replica electron microscopy combined with SDS digestion for cytochemical labeling of integral membrane proteins. Application to the immunogold labeling of intercellular junctional complexes. *J. Cell Sci.* 108:3443–3449.

Furuse, M., T. Hirase, M. Itoh, A. Nagafuchi, S. Yonemura, Sa. Tsukita, and Sh. Tsukita. 1993. Occludin: a novel integral membrane protein localizing at tight junctions. *J. Cell Biol.* 123:1777–1788.

Furuse, M., M. Itoh, T. Hirase, A. Nagafuchi, S. Yonemura, Sa. Tsukita, and Sh. Tsukita. 1994. Direct association of occludin with ZO-1 and its possible involvement in the localization of occludin at tight junctions. *J. Cell Biol.* 127: 1617–1626.

Furuse, M., K. Fujimoto, N. Sato, T. Hirase, Sa. Tsukita, and Sh. Tsukita. 1996. Overexpression of occludin, a tight junction-associated integral membrane protein, induces the formation of intracellular multilamellar bodies bearing tight junction-like structures. *J. Cell Sci.* 109:429–435.

Gumbiner, B. 1987. Structure, biochemistry and assembly of epithelial tight junctions. *Am. J. Physiol.* 253:C749–C758.

Gumbiner, B. 1993. Breaking through the tight junction barrier. *J. Cell Biol.* 123:1631–1633.

Huynh, T.V., R.A. Young, and R.W. Davis. 1985. Construction and screening cDNA libraries in λgtl0 and λgt11. *In* DNA Cloning: A Practical Approach. Vol. 1. D.M. Glover, editor. IRL Press Limited, Oxford, UK. 49–78.

Itoh, M., S. Yonemura, A. Nagafuchi, Sa. Tsukita, and Sh. Tsukita. 1991. A 220-kD undercoat-constitutive protein: Its specific localization at cadherin-based cell-cell adhesion sites. *J. Cell Biol.* 115:1449–1462.

Roy, N., M.S. Mahadevan, M. McLean, G. Shutler, Z. Yaraghi, R. Farahani, S. Baird, A. Besner-Jonston, C. Lefebvre, X. Kang, M. Salih, H. Aubry, K. Tamai, X. Guan, P. Ioannou, T.O. Crawford, P.J. de Jong, L. Surh, J.-E. Ikeda, R.G. Korneluk, and A. Mackenzie. 1995. The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell.* 13:167–178.

Schneeberger, E.E., and R.D. Lynch. 1992. Structure, function, and regulation of cellular tight junctions. *Am. J. Physiol.* 262:L647–L661.

Staehelin, L.A. 1973. Further observations on the fine structure of freeze-cleaved tight junctions. *J. Cell Sci.* 13:763–786.

Staehelin, L.A. 1974. Structure and function of intercellular junctions. *Int. Rev. Cytol.* 39:191–282.

Tsukita, Sh., Sa. Tsukita, A. Nagafuchi, and S. Yonemura. 1992. Molecular linkage between cadherins and actin filaments in cell-to-cell adherens junctions. *Curr. Opin. Cell Biol.* 4:834–839.

Tsukita, Sh., M. Itoh, A. Nagafuchi, S. Yonemura, and Sa. Tsukita. 1993. Submembranous junctional plaque proteins include potential tumor suppressor molecules. *J. Cell Biol.* 123:1049–1053.