A systems proteomics view of the endogenous human claudin protein family

LIU, Fei, et al.

Abstract

Claudins are the major transmembrane protein components of tight junctions in human endothelia and epithelia. Tissue-specific expression of claudin members suggests that this protein family is not only essential for sustaining the role of tight junction in cell permeability control but also vital in organizing cell contact signaling by protein-protein interactions. How this family of protein is collectively processed and regulated is key to understanding the role of junctional proteins in preserving cell identity and tissue integrity. The focus of this review is to first provide a brief overview of the functional context, on the basis of the extensive body of claudin biology research that has been thoroughly reviewed, for endogenous human claudin members and then ascertain existing and future proteomics techniques that may be applicable to systematically characterizing the chemical forms and interacting protein partners of this protein family in human. The ability to elucidate claudin-based signaling networks may provide new insight into cell development and differentiation programs that are crucial to tissue stability [...]
A systems proteomics view of the endogenous human claudin protein family
Fei Liu, Michael Koval, Shoba Ranganathan, Susan Fanayan, William S. Hancock, Emma K. Lundberg, Ronald C Beavis, Lydie Lane, Paula Duek, Leon McQuade, Neil L. Kelleher, and Mark S. Baker

J. Proteome Res., Just Accepted Manuscript • DOI: 10.1021/acs.jproteome.5b00769 • Publication Date (Web): 18 Dec 2015

Downloaded from http://pubs.acs.org on January 11, 2016

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.
Title: A systems proteomics view of the endogenous human claudin protein family

Fei Liu, a Michael Koval, b Shoba Ranganathan, a Susan Fanayan, c William S. Hancock, c,d Emma K. Lundberg, e Ronald C. Beavis, f Lydie Lane, g Paula Duek, g Leon McQuade, h Neil L. Kelleher, l Mark S. Baker c

aDepartment of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW 2109, Australia
bDepartment of Medicine, Division of Pulmonary, Allergy, Critical Care and Sleep Medicine and Department of Cell Biology, Emory University School of Medicine, 205 Whitehead Biomedical Research Building, 615 Michael Street, Atlanta, GA 30322
cDepartment of Biomedical Sciences, Macquarie University, Sydney, NSW 2109, Australia
dBarnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, United States
eSciLifeLab, School of Biotechnology, Royal Institute of Technology (KTH), SE-171 21 Solna, Stockholm, Sweden
fUniversity of Manitoba, Department of Biochemistry and Medical Genetics, 744 Bannatyne Avenue, Winnipeg, Manitoba, R3E 0W3 Canada
gSIB-Swiss Institute of Bioinformatics, CMU - Rue Michel-Servet 1, 1211 Geneva, Switzerland
hAustralian Proteome Analysis Facility, Macquarie University, Sydney, NSW 2109, Australia
iDepartments of Chemistry and Molecular Biosciences, and the Proteomics Center of Excellence, Northwestern University, 2145 N. Sheridan Road, Evanston, Illinois 60208, United States

*(F.L.) Tel: (61)-2-9850-8312. Fax: (61)-2-9850-8313. E-mail: fei.liu@mq.edu.au.

Key words: membrane protein complexes, cell-contact signaling, systems proteomics, membrane proteomics, targeted proteomics, top-down proteomics, chemical proteomics
Abstract

Claudins are the major transmembrane protein components of tight junctions in human endothelia and epithelia. Tissue-specific expression of claudin members suggests that this protein family is not only essential for sustaining the role of tight junction in cell permeability control but also vital in organizing cell contact signaling by protein-protein interactions. How this family of protein is collectively processed and regulated is key to understanding the role of junctional proteins in preserving cell identity and tissue integrity. The focus of this review is to first provide a brief overview of the functional context, on the basis of the extensive body of claudin biology research that has been thoroughly reviewed, for endogenous human claudin members and then ascertain existing and future proteomics techniques that may be applicable to systematically characterizing the chemical forms and interacting protein partners of this protein family in human. The ability to elucidate claudin-based signaling networks may provide new insight into cell development and differentiation programs that are crucial to tissue stability and manipulation.
Introduction

Claudins are a family of transmembrane proteins for barrier and pore formation in metazoans, especially for vertebrates and tunicates.\(^1\) Chordate claudins are the essential architectural proteins of tight junction strands in the apical junctional complex of epithelia and endothelia.\(^2\) These junctional strands, containing claudin protein complexes, act mainly as paracellular seals for large molecules and a semipermeable barrier to ions in tissues.\(^5,6,10\) Distinct claudin members are expressed in a tissue to confer tissue specific permeability and barrier characteristics.\(^7,11\) Claudins also serve as protein scaffolds for assembling signaling complexes at cell junctions.\(^3,12\) Understanding how the claudin family of proteins are expressed, organized and regulated in the cell with temporal and spatial resolution is critical to unraveling mechanisms that control paracellular barrier function, tissue integrity and stability.

The seminal discovery of claudins by Furuse and Tsukita in 1998 came 25 years after the initial observation of cell-cell contact ultrastructures.\(^16,17\) Since this seminal work the claudin protein family has been the subject of intense investigations to understand their structure and function in physiological and pathological contexts. The number of claudin genes varies between species. For example, in the case of the puffer fish *Takifugu rubripes*, up to 56 claudin genes are found by genome sequence analysis,\(^18\) whilst in *Homo sapiens*, the claudin gene family has at least 23 and maybe more if all predicted claudin genes are found to be expressed as proteins.\(^19\) Because different claudins are differentially expressed with tissue specificity and temporal regulation, claudins are likely to confer several vital roles particularly in control of paracellular permeability but also in cell differentiation, morphogenesis, and tissue maintenance.

Several reviews have appeared to account in detail advances in claudin physiology and biology of many organisms.\(^1,4,10,14,18,19,20,21\) These reviews represent a vast body of research in claudin
physiology in humans and other organisms. The majority of insights into defining how claudins function have been obtained using epithelial cell lines where the ramifications of claudin expression are measured after manipulation by either cDNA overexpression or RNA silencing. Claudin-deficient\textsuperscript{22} and overexpression\textsuperscript{23,24} transgenic mouse models have also proven informative in identifying functions for specific claudins in regulating epithelial barrier function and other aspects of mammalian physiology relevant to human disease. These are valuable approaches. However it is difficult to employ these models so that claudin expression is manipulated within the normal physiologic range of mRNA and protein expression. The claudin expression profile of many different tissues and cells, including normal and tumor cells, has been assessed at the mRNA and protein level. While these approaches provide some insights into how the claudin proteome is organized in different tissues, they have two major limitations. First, claudin mRNA does not necessarily correlate with levels of protein expression,\textsuperscript{25} since processes such as tight junction turnover are significantly regulated by protein-protein interactions\textsuperscript{26} and post-translational modifications (PTMs) of claudins in both human and animal cell models.\textsuperscript{27-30} Second, direct measurements of claudin protein have been largely restricted to antibody-dependent approaches (e.g. immunoblot) which have significant utility but are difficult to use to measure the stoichiometry of claudin composition, which is a critical parameter needed to determine how differential claudin expression influences epithelial barrier function and other aspects of cell physiology.

The approach of “systems proteomics”, encompassing specialty technologies such as membrane proteomics,\textsuperscript{31,32} targeted proteomics,\textsuperscript{33-35} bottom-up/top-down proteomics,\textsuperscript{36} structural proteomics,\textsuperscript{37,38} and chemical proteomics,\textsuperscript{39,40} has the potential to provide valuable adjuncts to other established methods currently used to define roles for claudins in cell and tissue function. However, the application of systems proteomics has lagged behind our understanding of functional roles of claudins in regulating human cells and tissues. There are several reasons for this. Claudins are highly hydrophobic proteins that are more difficult to isolate and analyze by mass spectrometry
(MS) than hydrophilic proteins.\textsuperscript{41} Also claudin-function is highly dependent on cell localization that, in turn, is regulated by interactions with other proteins. Claudin-protein interactions range from weak, unstable interactions (e.g. claudin-claudin interactions\textsuperscript{14,42}) to interactions that effectively immobilize claudins onto the actin cytoskeleton.\textsuperscript{43} These are classes of protein-protein interactions that are difficult to analyze and will require novel approaches to define.

In this review, human claudin genetics and biology is summarized first for a general outlook while also guiding the reader to other sources for further details on claudin cell and molecular physiology. Here the discussion focuses on the chemical characteristics of human claudins with a particular emphasis placed on MS-centered, systems proteomics approaches to studying human endogenous claudins. Some of the investigations on other organisms such as mice, rats, and canine cell/tissue models will be discussed only for occasional comparisons. Specialty proteomics technologies, such as membrane proteomics, structural proteomics, targeted proteomics, bottom-up/top-down proteomics, and chemical proteomics, are assessed for consideration of their further development and integration into investigating human endogenous claudins. Because endogenous claudins are low in abundance with complex chemical modification and localization profiles, this will most certainly push technology development to new limits. Emerging MS approaches for elucidating membrane protein complexes in the relevant context of claudin biological function is also addressed here. Human claudin PTMs, revealed thus far by MS and/or immuno-detection, are discussed to highlight how claudin function may be regulated. Existing approaches for characterizing endogenous human claudins by MS in conjunction with chemical partition and enrichment techniques are also reviewed to underscore the challenges and also opportunities in unraveling the claudin protein networks holistically with improved temporal and spatial resolution. New chemical delineation of the native claudin protein system may lead to novel hypotheses into claudin biology and new cell assembly and tissue manipulation capabilities. The advanced technologies developed
for claudin systems proteomics could also assist further biological investigations into junctional signaling networks involved in intercellular communication in metazoans.\textsuperscript{44-48}

1. A general overview of the genetic profile and protein family, biological role, and tissue expression of the human claudins

As indicated earlier, recent reviews have already provided insightful accounts of claudin genetics and biology, along with their significant interactions with other important junction proteins. Information on human claudins from various databases, such as UniProt\textsuperscript{49}, Ensembl\textsuperscript{50} and the Human Protein Atlas (HPA)\textsuperscript{51} is compiled in Table 1 to provide a general outlook of human claudins. Much of the recent claudin research is not yet captured in these databases, although they offer an overview of human claudins to help guide navigating the detailed reviews and current articles for in-depth analysis.

Genetic analysis

Including transcript variants, 27 mammalian claudin genes have been reported, and at least 23 are found in humans (Table 1, column 1).\textsuperscript{20,52-54} The mammalian \textit{CLDN13} gene is absent in humans but present in rodents, while \textit{CLDN24}, 25, 26, and 27 are putative claudin genes.\textsuperscript{1} The official gene names for claudins, attributed by the HUGO gene nomenclature committee,\textsuperscript{55} are shown in Table 1 (column 1 and 2) with other synonyms also in use (e.g. in GeneCards\textsuperscript{56}, Table 1, column 3). The human claudin genes are spread across 13 chromosomes (1, 3, 4, 6, 7, 8, 11, 13, 16, 17, 21, 22, and X) with generally few or even no introns found for these genes (Table 1, column 4). Chromosomes 3 and 7 have the highest frequency of claudins (\textit{CLDN1}, 11, 16, and 18 on chromosome 3; \textit{CLDN3}, 4, 12, and 15 on chromosome 7). Some claudins also exhibit high pair-wise sequence homology, such as \textit{CLDN3} and 4, \textit{CLDN6} and 9, \textit{CLDN8} and 17, and \textit{CLDN22} and 24. Coordinated gene expression is possible and has been observed for \textit{CLDN3} and \textit{CLDN4}.\textsuperscript{54} Mammalian claudins are
classified into the “classic” (CLDN1-10, 14, 15, 17, and 19) and “non-classic” (CLDN11-13, 16, 18, 20-24) groups on the basis of their phylogenetic relationship, and it is also possible to split the claudins into eight groups based on their gene structures and phylogenetic distances.

Table 1. An overview of the human claudin family members from databases.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein IDs (UniProt/Ensembl)</th>
<th>Synonyms (GeneCards)</th>
<th>Chromosome Localization</th>
<th>Tissue Frequency with HPA Evidence</th>
<th>Tissue-specific RNA FPKM (max in nonspecific tissue)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLDN1</td>
<td>O95832/163347</td>
<td>SEMP1, ILVASC</td>
<td>3q28</td>
<td>very common</td>
<td>- (skin: 224.1)</td>
</tr>
<tr>
<td>CLDN2</td>
<td>P57739/165376</td>
<td>SP82</td>
<td>Xq22</td>
<td>common</td>
<td>- (kidney: 114.5)</td>
</tr>
<tr>
<td>CLDN3</td>
<td>O15551/165215</td>
<td>CPETR2</td>
<td>7q11</td>
<td>very common</td>
<td>- (colon: 94.8)</td>
</tr>
<tr>
<td>CLDN4</td>
<td>O14493/189143</td>
<td>CPER, CPETR1, WBSCR8</td>
<td>7q11</td>
<td>common</td>
<td>- (colon: 160.9)</td>
</tr>
<tr>
<td>CLDN5</td>
<td>O00501/184113</td>
<td>AWAL, TMVCF</td>
<td>22q11</td>
<td>very common</td>
<td>adipose tissue: 110.8; lung: 32.1; (spleen: 10.4)</td>
</tr>
<tr>
<td>CLDN6</td>
<td>P56747/184697</td>
<td>UNQ757, PRO1488, Skullin</td>
<td>16p13</td>
<td>very common</td>
<td>- (placenta: 1.3)</td>
</tr>
<tr>
<td>CLDN7</td>
<td>O95471/181885</td>
<td>CEPTRL2, CEPTRL2</td>
<td>17p13</td>
<td>very common</td>
<td>- (colon: 297.3)</td>
</tr>
<tr>
<td>CLDN8</td>
<td>P56748/156284</td>
<td>UNQ779, PRO1573</td>
<td>21q22</td>
<td>very common</td>
<td>- (kidney: 63.6)</td>
</tr>
<tr>
<td>CLDN9</td>
<td>O95484/213937</td>
<td>-</td>
<td>16p13</td>
<td>very common</td>
<td>- (pancreas: 2.0)</td>
</tr>
<tr>
<td>CLDN10</td>
<td>P78369/134873</td>
<td>CPETRL3, OSP-L</td>
<td>13q32</td>
<td>less common</td>
<td>- (salivary gland: 150.7)</td>
</tr>
<tr>
<td>CLDN11</td>
<td>O75508/013297</td>
<td>OSP, OTM</td>
<td>3q26</td>
<td>less common</td>
<td>- (testis: 167.0)</td>
</tr>
<tr>
<td>CLDN12</td>
<td>P56749/157224</td>
<td>-</td>
<td>7q21</td>
<td>very common</td>
<td>- (liver: 31.9)</td>
</tr>
<tr>
<td>Gene</td>
<td>Accession Number</td>
<td>Description</td>
<td>Chromosome</td>
<td>Location</td>
<td>Expression</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
<td>-------------------------------</td>
<td>------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>CLDN14</td>
<td>O95500/159261</td>
<td>CLDN14</td>
<td>21q22</td>
<td>NA</td>
<td>kidney: 5.3; liver: 8.4; duodenum: 184.0; small intestine: 14.1; spleen: 12.6; appendix: 0.5</td>
</tr>
<tr>
<td>CLDN15</td>
<td>P56746/106404</td>
<td>CLDN15</td>
<td>7q22</td>
<td>very common</td>
<td>duodenum: 184.0; small intestine: 14.1; spleen: 12.6</td>
</tr>
<tr>
<td>CLDN16</td>
<td>Q9Y517/113946</td>
<td>CLDN16</td>
<td>3q28</td>
<td>rare</td>
<td>kidney: 39.0; (thyroid gland: 6.8)</td>
</tr>
<tr>
<td>CLDN17</td>
<td>P56750/156282</td>
<td>CLDN17</td>
<td>21q22</td>
<td>NA</td>
<td>esophagus: 6.4; (testis: 0.3)</td>
</tr>
<tr>
<td>CLDN18</td>
<td>P56856/066405</td>
<td>CLDN18</td>
<td>3q22</td>
<td>rare</td>
<td>lung: 246.6; stomach: 556.1; heart muscle: 7.3</td>
</tr>
<tr>
<td>CLDN19</td>
<td>Q8N6F1/164007</td>
<td>CLDN19</td>
<td>1p34</td>
<td>NA</td>
<td>kidney: 21.9; placenta: 11.3; adipose tissue: 0.2</td>
</tr>
<tr>
<td>CLDN20</td>
<td>P56880/171217</td>
<td>CLDN20</td>
<td>6q25</td>
<td>NA</td>
<td>- (uterus: 0.3)</td>
</tr>
<tr>
<td>CLDN21</td>
<td>see CLDN24</td>
<td>CLDN21</td>
<td>4q35</td>
<td>NA</td>
<td>- (adrenal gland: 0.1)</td>
</tr>
<tr>
<td>CLDN22</td>
<td>Q8N7P3/177300</td>
<td>CLDN22</td>
<td>4q35</td>
<td>NA</td>
<td>- (stomach: 17.6)</td>
</tr>
<tr>
<td>CLDN23</td>
<td>Q96B33/253958</td>
<td>also named CLDN21</td>
<td>8p23</td>
<td>NA</td>
<td>- (kidney: 0.7)</td>
</tr>
<tr>
<td>CLDN24</td>
<td>A6NM45/185758</td>
<td>also named CLDN21</td>
<td>4q35</td>
<td>NA</td>
<td>- (adipose tissue: 0.0)</td>
</tr>
<tr>
<td>CLDN25</td>
<td>C9JDP6/228607</td>
<td>-</td>
<td>11q23</td>
<td>rare</td>
<td>- (testis: 0.4)</td>
</tr>
<tr>
<td>CLDN26</td>
<td>B3SHH9</td>
<td>TMEM114</td>
<td>16p13</td>
<td>NA</td>
<td>- (testis: 0.1)</td>
</tr>
<tr>
<td>CLDN27</td>
<td>A6NFC5</td>
<td>TMEM235</td>
<td>17q25</td>
<td>NA</td>
<td>- (testis: 0.1)</td>
</tr>
</tbody>
</table>


*b Ensemble code prefix: ENSG00000.

c Not identified as claudin but as TMEM (transmembrane) proteins.

d Tracked from GeneCards synonyms.
CLDN13 is absent in human. CLDN 6 and 9, as well as CLDN 8 and 17, may be considered as paralogs.\textsuperscript{44}

\textsuperscript{1}HPA evidence was calculated based on the manual curation of Western blot, tissue profiling and subcellular location using a limited number of claudin antibodies that have been submitted to HPA and validated by HPA. Many other claudin antibodies are available but not included in the HPA list. Very common = high or medium levels in at least 20 tissues; Common = high or medium levels in at least 10 tissues; Less common = high or medium levels in more than 3 but less than 10 tissues; Rare = high or medium levels in only 1 to 3 tissues; NA = no antibodies.

\textsuperscript{6}In nucleus but not nucleoli, cell junctions by immunofluorescence analysis.

\textsuperscript{b}In plasma membrane by immunofluorescence analysis.

\textsuperscript{1}FPKM value = number of fragments per kilobase gene model and million reads. The FPKM threshold value for detection is > 1.\textsuperscript{58} FPKM measurements reflect measured tissues represented in the database and are not necessarily representative of a complete mRNA expression profile.
Table 2. Peptide evidence for claudin proteins by MS analysis. For each claudin, peptides with good MS evidence are listed (data compiled from www.thegpm.org and www.nextprot.org).

<table>
<thead>
<tr>
<th>Proteins</th>
<th># of Observations</th>
<th>Best Log(E)</th>
<th>Evidence Code</th>
<th>Sequences with Good Evidence (EC = 4)(^b) (amino acid residue numbers in parentheses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLDN1</td>
<td>323</td>
<td>-111</td>
<td>4</td>
<td>VFDSLLNSLSTLQATR(66-81); CLEDEVKQ(107-115); CLEDEVQKMR(107-117); IVQEFYDPMTVPVNR(144-158); FYDPMTPVNYE(148-160); KTTSYTPRPYPK(189-201); TTSYTPRPYPK(190-201); PYPKAPSSGKDYV(198-211); PYPKAPSSGK(198-208)</td>
</tr>
<tr>
<td>CLDN2(^e)</td>
<td>100</td>
<td>-31</td>
<td>4</td>
<td>DFYSPLVPDSMK(146-157); SNYYDAYQAQPLATR(192-206); SEFNSYSLTGYV(219-230)</td>
</tr>
<tr>
<td>CLDN3</td>
<td>1695</td>
<td>-78</td>
<td>4</td>
<td>VYDSLLALPQDLQAAR(65-80); DFYNPVVPEAQK(145-156); DFYNPVVPEAQKR(145-157); VVYSAAPSTGPGASLTGYDR(196-216); STPGASLTGYDR(203-220); STPGASLTGYDR(203-216); STPGASLTGYDR(203-218)</td>
</tr>
<tr>
<td>CLDN4</td>
<td>861</td>
<td>-51</td>
<td>4</td>
<td>VYDSLLALPQDLQAAR(66-81)(^c); CTNCLEDESAK(104-114); CTNCLEDESAKAK(104-116); NIIQDFYNPLVASGQK(142-157)</td>
</tr>
<tr>
<td>CLDN5</td>
<td>169</td>
<td>-85</td>
<td>4</td>
<td>GLWMSCVVQSTGHPCK(134-150); VYDSVLALSTEVQAR(151-166); EFYDSPVPSQK(231-242); RPTATGDYDK(290-300)</td>
</tr>
<tr>
<td>CLDN6</td>
<td>997</td>
<td>-59</td>
<td>4</td>
<td>VYDSLLALPQDLQAAR(66-81)(^c); DFYNPLVAAEQK(146-157); DFYNPLVAAEQKR(146-158); YSTSAPAISRGPSYPTKNY(200-220); YSTSAPAISRGPSYPTKNY(200-217); GPSEYPTK(210-217)</td>
</tr>
<tr>
<td>CLDN7</td>
<td>372</td>
<td>-69</td>
<td>4</td>
<td>SSVPNIK(115-121); AGDNITQAMYK(36-48); GLWMDCTQTSGMMSCK(49-65)(^d); KMYDSVLALSALQATR(65-81); MYDSVLALSAALQATR(66-81); SVLALSALQATRALMV(69-85); VSLVGLFAMFMVATMG(86-101); QIVTDFYNPLIPTNIK(143-158)</td>
</tr>
<tr>
<td>CLDN8</td>
<td>18</td>
<td>-7</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>CLDN9</td>
<td>703</td>
<td>-28</td>
<td>4</td>
<td>VYDSLLALPQDLQAAR(66-81)(^c)</td>
</tr>
<tr>
<td>CLDN10</td>
<td>102</td>
<td>-24</td>
<td>4</td>
<td>ACVTDSTGVSNCK(52-64); ITTLDPPFLFVEQK(142-155); YTYNGATSVMSSR(192-204); YHGGEDFK(207-214)</td>
</tr>
<tr>
<td>CLDN11</td>
<td>336</td>
<td>-362</td>
<td>4</td>
<td>GLWADCMATGLYHCK(51-66); GLWADCMATGLYH(51-64); CKPLVDILILPGYVQACR(65-82); PLVDILILPGYVQACR(67-82); ILILPGYVQACR(71-82); MGQEPGVAK(108-116); FYTAGSSSPTHAK(190-203)(^b); FYTAGSSSPTHAK(190-202); YTAGSSSPTHAK(191-203); YTAGSSSPTHAK(192-203)</td>
</tr>
<tr>
<td>CLDN12</td>
<td>259</td>
<td>-17</td>
<td>4</td>
<td>SSVPNIK(115-121); SRLSAIEIDIPVVSHTTI(228-244); SRLSAIEIDIPVVSHT(228-242); LSAIEIDIPVVSHTTI(230-244)</td>
</tr>
<tr>
<td>CLDN</td>
<td>Evidence</td>
<td>pL</td>
<td>Evidence</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>-----</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>CLDN14</td>
<td>14</td>
<td>-4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CLDN15</td>
<td>11</td>
<td>-8</td>
<td>4</td>
<td>RPYQAPSVMPVATSDQEGDSSFGK(197-221)</td>
</tr>
<tr>
<td>CLDN16</td>
<td>34</td>
<td>-15</td>
<td>4</td>
<td>TCDEYDIAEHLPLK(130-144)</td>
</tr>
<tr>
<td>CLDN17</td>
<td>5</td>
<td>-2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CLDN18</td>
<td>82</td>
<td>-87</td>
<td>4</td>
<td>ESSGFTECR(56-64); GYFTLGLPAMLQAVR(65-80); AVSYHASGHSVAYK(208-221); (SVAYKPGGFK(217-226)); TEDEVQSYPSKH(247-261); TEDEVQSYPSK(247-257)</td>
</tr>
<tr>
<td>CLDN19</td>
<td>18</td>
<td>-11</td>
<td>4</td>
<td>LYDSLLALDGIQSAR(66-81)</td>
</tr>
<tr>
<td>CLDN20</td>
<td>2</td>
<td>Not found</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>CLDN21</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>See CLDN24</td>
</tr>
<tr>
<td>CLDN22</td>
<td>10</td>
<td>-5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CLDN23</td>
<td>91</td>
<td>-15</td>
<td>4</td>
<td>RSSVSTIQVEWPEPDLAPA(202-222)</td>
</tr>
<tr>
<td>CLDN24</td>
<td>54</td>
<td>-9</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CLDN25</td>
<td>2</td>
<td>-3</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

- Evidence code: 4 = good evidence; 3 = moderate evidence; 2 = weak evidence; 1 = no evidence.
- All sequences are at least 8 amino acid residues in length.
- The peptide is common to claudins 3, 4, 6, and 9.
- Single nucleotide polymorphism-associated peptide variants found for this peptide.
- Claudin 2 is considered missing in neXtProt.
- In neXtProt, only peptides by single reaction monitoring were available for claudin 15 and 16.
- In neXtProt, additional peptides have been identified for the claudin although not identified as level 4 evidence by GPM.
From the protein structure and function perspective, the claudin protein family is included in the pfam00822 superfamily of proteins that are characterized by their four helical transmembrane regions.\textsuperscript{1,7,21} The pfam00822 superfamily also includes other subfamilies such as the PMP-22, EMP, MP20 and calcium channel CACNG proteins. Human claudins are in the range of 200-300 amino acids (20–35 kD) and share the common transmembrane regions made of four helices (TM 1–4) with two extracellular loops (ECL1 and ECL2) and one intracellular loop (ICL).

The general molecular architecture of the claudin protein family, determined by an extensive body of genetic, biochemical, and biophysical experiments, is shown in Figure 1a. The four transmembrane regions (TM 1–4) are typically 24 amino acids in length, while the $\alpha$-helix of TM3 extends beyond the membrane by about 10 amino acids.\textsuperscript{53,59} The ECL1 is generally around 50 amino acids with charged amino acids for paracellular pore formation. A Prosite signature sequence for the claudin family has been identified as [GN]-L-W-x(2)-C-x(7, 9)-[STDENQH]-C in ECL1.\textsuperscript{1,60} A consensus sequence, W-x(17-22)-W-x(2)-C-x(8-10)-C is found in ECL1 with the first Trp flanking the end of TM 1.\textsuperscript{61} A signature motif, GLW, is found 2 amino acids away from the N-terminus of the conserved C-x(8-10)-C motif. Most claudins also have an Arg marking the end of the ECL1. Inside the cell, the claudin N-terminus (<10 amino acids), ICL (~12 amino acids), and C-terminus (25 to 55 amino acids) provide interaction sites with other proteins. Most notably, a PDZ\textsuperscript{62} domain binding motif YV at the C-terminus is completely conserved amongst classic claudins and also prevalent amongst the nonclassic ones.\textsuperscript{57,63} The YV motif is the main interaction site for binding other junction associated proteins such as ZO-1, 2 and 3, although other residues in proximity of the C-terminus also contribute to claudin’s protein binding affinity.\textsuperscript{63} The first mammalian claudin structure (mouse claudin-15) at 2.4 Å confirms this general claudin four-helix bundle that spans the full length of the lipid bilayer (Fig. 1b).\textsuperscript{64} Subsequently, a binary structure at
3.7 Å resolution of mouse claudin-19 complexed to the C-terminal region of *Clostridium perfringens* enterotoxin further validates the generality of the transmembrane, four-helix bundle scaffold and identifies the specific binding motifs in both the ECL1 and ECL2 loops to the bacterial toxin.\(^{65}\)

Figure 1. a) A general scheme of the claudin protein structure; b) secondary structural motifs found in the ECL loops of the X-ray structure of mouse claudin-15.\(^{64}\)

The claudin protein family expression level by immunoblotting is not necessarily correlated to its RNA level (Table 1, column 5 and 6). Human claudins 24 to 27 have been tentatively assigned as members of the claudin family, and the evidence for these members remains tenuous even at the RNA level (Table 1, column 6). Many antibodies have been in use for claudin detection in both cells and tissues (Table 1, column 5), although polyclonal antibodies recognize multiple epitopes and antibody cross-reactivity is a significant concern in analyzing claudins. Currently, four human claudin proteins (claudin-8, 9, 20, and 22) are considered “missing”\(^{66}\) at the MS level (as defined by HUPO: the Human Proteome Organization) with evidence for their presence only at the transcription level, although for human claudins 8 and 9, protein presence has been suggested by immunoblotting and likely requires further validation.\(^{67,68}\) In terms of good MS evidence, eleven of the human claudins, namely claudins 1, 2, 3, 4, 5, 6, 7, 10, 11, 12 and 18 have been identified...
multiple times by more than one unique peptide with good log(E) scores (Table 2). Further, four human claudins (15, 16, 19, and 23) have been identified multiple times by just one unique peptide with good log(E) scores. In the case of claudin 9, only one non-unique peptide has been identified multiple times with good log(E) scores, which is also found in claudins 3, 4, and 6. Overall, of the 23 definitive human claudins (not including the disputed claudins 26 and 27), human claudins 8, 9, 14, 17, 20, 22, 24 and 25 do not have definitive MS-based protein evidence. Therefore, only about half of the human claudin protein family can be considered as identified with certainty by MS at the protein level, with the other half either “missing” or uncertain by MS. For human claudins (7, 10, 11, 18, and 19) with isoforms by alternative splicing, no credible MS evidence exists for their presence as protein isoforms. One caveat is that several of these may be enriched in tissues that have not been probed for claudin expression by MS. For instance, claudin-18 has been shown by immunologic techniques to be highly expressed at the protein level in stomach and lung.

2. A general overview of the tissue and functional profile of the human claudins

Claudins are found widely expressed, either as gene transcripts or expressed proteins, in diverse tissues.69 The vast majority of claudin expression is found in barrier forming epithelial and endothelial cells. Not all claudins are concurrently expressed by all tissues, and it is these differences in claudin expression that regulate cell function, most notably, paracellular barrier function. Reflecting a predominant role in regulating barrier function, claudins frequently are found concentrated at the tight junction, however some claudins are also found at other intracellular locations, such as regions along the basolateral plasma membrane.1,14,70 Nuclear localization of claudins has also been observed.1,10,20,71-73 Roles for non-junctional claudins are not well understood, although the literature correlating nuclear claudin localization with tumor metastasis suggest roles for these claudin pools in regulating cell growth, division and migration.61,74-78
The most prevalent roles for claudins are in barrier and pore formation, and the details of how individual claudins influence tight junctions have been extensively reviewed. The majority of claudins are primarily barrier forming, based largely on studies of cell line models where the effects of increasing and decreasing claudin expression on measures of barrier function was assessed, including transepithelial resistance and paracellular flux of small molecular weight tracers. By contrast, claudin-2, 10, 15, and 17 are primarily pore-forming and lead to increased paracellular permeability correlating with increased expression as a result of formation of paracellular ion and water channels. The high-resolution structure of claudin-15 has provided several insights into the structural basis for claudin paracellular ion selectivity. Homology modeling and structure determination of other claudins will likely provide a valuable approach to further define the permeability characteristics of other claudins. Claudins also interact with other junction membrane proteins, such as occludin, tricellulin, and junctional adhesion molecule-A (JAM-A), to stabilize and regulate their membrane retention for improved barrier formation.

Claudin mutations have been implicated in human diseases affecting ion homeostasis. One of the best characterized claudin-associated diseases is familial hypomagnesemia with hypercalciuria and nephrocalcinosis that is caused by mutations in claudin-16 and claudin-19. The majority of the mutant claudin-16 alleles (12 out of 16) show a missense mutation in the highly conserved amino acids in ECL1 or ECL2. Such mutations have been postulated to alter the tertiary structure of the extracellular domains in claudin complexes and thus change the cation specificity of claudin pores. Other mutations affect claudin-16/claudin-19 interactions diminishing the transport and assembly of these claudins into tight junctions. Claudin mutations can be linked to diverse, tissue specific pathologic outcomes, which underscores the need to understand the structural basis that regulates their function and protein-protein interactions.
3. Post-translational modification of claudins.

PTMs greatly expand the claudin chemical diversity found in cells and in turn further diversify claudin function. Potential biological processes that may be differentially regulated through claudin PTM include protein interaction partnerships, trafficking, subcellular localization, oligomer (hetero- or homo-) assembly, and finally net claudin homeostasis. Here we discuss several classes of claudin PTMs: phosphorylation, palmitoylation, ubiquitination and glycosylation (Figure 2).

Proteolysis is another way by which claudins can be subject to PTM. It is not well established whether any proteolytic claudin fragments have biological activity or whether this largely serves to decrease the total claudin protein pool.

Figure 2. Known or hypothetical PTMs of claudins and the regions in which they are predicted to occur. Predicted PTMs include phosphohorylation (red oval), palmitoylation (yellow line), ubiquitination (blue circle), proteolysis (green marks) and glycosylation (grey box).

**Phosphorylation**

Currently, phosphorylation on Ser/Thr/Tyr residues is the most abundantly observed claudin PTM. Theoretical predictions of potential phosphorylation sites indicate that up to 10 phosphorylation sites are possible by a variety of protein kinases (e.g. PKA/C, MAPK, WNK, Src, and Eph). Most of these predicted phosphorylation sites are located in the C-terminal tail. One important feature of claudin phosphorylation is that the Tyr residue found in the highly conserved PDZ binding motif at
the end of the C-terminus can be a phosphorylation site by the Eph receptor tyrosine protein kinase, thus providing a means to regulate its binding activity. For example, in the case of claudin-4 in HT29 colon carcinoma cells, EphA2 activation leads to claudin-4 tyrosine phosphorylation, and this specific Tyr phosphorylation lowers its association with ZO-1 and leads to less claudin-4 integration into cell junctions and higher paracellular permeability in MDCK cell models.\textsuperscript{87} This result indicates that EphA2 Tyr phosphorylation near the claudin C-terminus may alter the function of the tight junction and promote a regime change in signaling at the tight junction. Tyr phosphorylation for claudin-5, detected by a general phospho-Tyr antibody, has also been linked to increased paracellular permeability in cultured human brain capillary endothelial cells.\textsuperscript{88} Similarly, Ser/Thr phosphorylation of claudins is known to influence the barrier strength of cell junctions. Phosphorylation of T192 of claudin-3 by PKA has been found in ovarian cancer cells as a mechanism for disruption of normal junctional function.\textsuperscript{89} In other instances, Ser/Thr phosphorylation of claudin-7 can up-regulate the caspase pathway to increase the chemosensitivity of lung cancer cells toward cisplatin or increase paracellular permeability to chloride anions in kidney cells.\textsuperscript{90,91} Large-scale phosphoproteomics investigations (from PhosphoSitePlus\textsuperscript{92} with specific journal article references and the Global Proteome Machine\textsuperscript{93} have identified multiple Tyr/Ser/Thr sites of claudin-1–12, 15, 16, 18, and 23 in human T cells, embryonic stem cells and cancer cells (summarized in Table 3).\textsuperscript{90,91,94-113} However, the specific roles of these claudin PTMs remain largely unknown without functional validation from biological experiments.

Table 3. Human claudin PTM sites from PhosphoSitePlus or the Global Proteome Machine (GPM) databases (p = phosphorylation; ub = ubiquitination\textsuperscript{108}). Sites in bold are from GPM without a specific article reference from PhosphoSitePlus.

<table>
<thead>
<tr>
<th>Claudin</th>
<th>PTM sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T195-p; Y199-p; S205-p; S206-p; Y210-p</td>
</tr>
<tr>
<td>2</td>
<td>S192-p; Y194-p; Y195-p; Y198-p; S208-p; S219-p</td>
</tr>
<tr>
<td>3</td>
<td>Y198-p; S199-p; S203-p; T204-p; S209-p; T212-p; Y214-p; Y219-p</td>
</tr>
</tbody>
</table>
Y193-p; Y197-p; Y208-p
Y148-p; Y152-p; Y154-p; S155-p; T207-p; Y212-p; Y217-p
Y152-p; S201-p; T202-p; S203-p; S208-p; S212-p; Y214-p; Y219-p
K203-ub; S204-p; S206-p; S207-p; K208-ub; Y210-p
Y67-p; S69-p; S74-p; S215-p
Y200-p
S94-p; T198-p; S199-p; S202-p
Y191-p; Y192-p; T193-p; S196-p; S197-p; S198-p
T111-p; S115-p; S116-p; K121-ub; Y211-p; S216-p; Y220-p; S228-p; S231-p
T84-p; S111-p; S211-p; S217-p; S218-p
S217-p; S278-p
Y206-p; S210-p; Y211-p; S214-p; S217-p; Y220-p; T229-p; Y241-p; T247-p; S253-p; Y254-p; S256-p;
Y260-p
S203-p; S204-p; S206-p; S207-p; Y223-p; Y224-p

* Phosphorylated claudin-8 peptides were reported.94

The effects of phosphorylation are claudin specific, in regards to the specific residue modified, and
vary even amongst highly homologous claudins. For example, S208 phosphorylation of claudin-2 is
required for its membrane retention and reduced trafficking to lysosomes in MDCKII cells;
however, this phosphorylation does not alter its binding to ZO-1 or ZO-2.30 In the case of claudin-5,
PKA phosphorylation at T207 in endothelial cells, while conducive to its translocation into the tight
junction membrane, leads to increased paracellular permeability.114,115 In addition to barrier
function, claudin phosphorylation is also known to control selective ion permeability. For example,
reduction of claudin-16 phosphorylation at S217 results in impaired Mg2+ permeability in renal
MDCK cells.116 As an additional confounding variable, tyrosines are subject to nitrosylation under
conditions of oxidative stress.117 Evidence for tyrosine nitration of claudins has only been found in
rats at this point, where renal claudin-2 nitration was detected in early diabetic rats.118
Palmitoylation

Similar to phosphorylation, palmitoylation can significantly alter claudin localization, protein interactions, trafficking and stability. The TM2 and TM4 of claudins are each flanked by a cysteine residue that has a palmitoylation motif. Some claudins have additional cysteine residues potentially available for palmitoylation in the ICL and C-terminal tail. For example, in MDCK cells two cysteine residues in the ICL and two more in the C-terminal tail of claudin-14 must be palmitoylated as a requirement for efficient plasma membrane localization and tight junction assembly. However, in transfected HEK cells, palmitoylation of claudin-7 is associated with partitioning into glycosphingolipid membrane microdomains and inhibits integration into tight junctions, suggesting a model where palmitoylation controls the relative amounts of claudin-7 in the basolateral vs. tight junction pools. Protein S-acylation has been profiled at the proteome scale, and given the difficulties in measuring palmitoylation by standard biochemical techniques, this is an area that would be ideally suited for state of the art proteomics analysis.

Ubiquitination and SUMOylation

Evidence for human claudin ubiquitination and SUMOylation is scarce but more frequently observed in model cell lines of other organisms. Polyubiquitination of claudin-5 is observed in claudin-5 transfected HeLa cells at K199 to target the protein for proteasomal degradation. This results in claudin-5 being removed from the plasma membrane and loss of the tight junction strands, indicating that with this particular PTM claudin stability can be compromised by a potential mechanism of increased turnover. Claudin ubiquitination has been more consistently observed in canine, mouse, and rat cell models according to PhosphositePlus. For example, polyubiquitination of claudin-1 in MDCK cells appears to be dependent on LNX1p80 overexpression but the sites of ubiquitination are not specified and also do not involve Lys48, a site that is typically ubiquitinated for proteasome-dependent degradation. SUMOylation of claudin-2 at K218 is also reported in MDCK cells in which claudin-2 is removed from the lower lateral membrane but not the tight
junction.\textsuperscript{126} It is noteworthy that not all claudins contain ubiquitination sites, although the classic claudins are thought to be the primary candidates for ubiquitination, and this is supported by Lys ubiquitination sites identified by large scale proteomics (Table 3, claudins 7 and 12). The current understanding of claudin ubiquitination/SUMOylation on claudin localization, function or stability is based primarily on studies of non-human claudins. It is likely that direct evidence for ubiquitination/SUMOylation of human endogenous claudins will be forthcoming.

\textbf{Glycosylation}

Experimental evidence for claudin glycosylation is again scarce, perhaps due to the low abundance level of these proteins that require specific enrichment prior to such analyses. This lack of data also suggests that claudin glycosylation is a rare event. Putative \textit{N}- and \textit{O}-glycosylation sites on some claudins have been identified.\textsuperscript{127,128} Some of these predicted \textit{O}-glycosylation sites overlap with predicted phosphorylation sites. These overlaps are known as the Yin Yang sites, where the two modifications can reciprocally occur on the same Ser/Thr residue or occupy adjacent residues. For example, in the case of human claudin-1, eight potential \textit{O}-\text{GlcNAc} modifications sites were predicted to be located at T190, T191, S24, S56, S185, S192, S205 and S206. Of these eight potential glycosylation sites, four (T191, S192, S205, and S206) were also predicted as Yin Yang sites. However, there is no compelling evidence suggesting a role for this PTM overlap in claudin function. In addition, \textit{N}-glycosylation sites have been predicted to exist in the ECL1 of claudins-1 and 12.\textsuperscript{129} Furthermore, claudin-26 homolog TMEM114 is also predicted to be \textit{N}-glycosylated at position N54 and N88.\textsuperscript{129,130}

In this review, we have undertaken prediction of \textit{N}-glycosylation sites in all claudin proteins using the NetNGlyc 1.0 Server\textsuperscript{131}, where a threshold >0.5 was applied for this analysis. This server predicts \textit{N}-glycosylation sites in human proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr motifs. The predicted \textit{N}-glycosylation sites tend to be much
less abundant compared to those for \( O \)-glycosylation. While potential glycosylation sites were observed for claudins-1, 7, 10, 12, 15, 18, 22, 23 and 24 (all with jury values between 7/9 and 9/9, except for claudin-7 with a jury value of 5/9), most of the predicted \( N \)-glycosylation sites are in the helical regions and therefore not likely to be accessible to glycosylation machinery, except for claudin-12 at 47 NLTV, a segment in the ECL1 region. Claudin-12 expressed by transfected HeLa cells is \( N \)-glycosylated as determined by a complex banding pattern observed by immunoblot that compresses into a single band when treated with Peptide-N-Glycosidase F (Koval, unpublished results), indicating that this is a functional glycosylation site. Claudin-12 transfected COS-7 kidney fibroblasts show a similar banding pattern also suggesting it is glycosylated.\(^{132}\) However, immunoblots of claudin-12 expressed by endothelial cells and in human brain homogenates show predominantly a single band,\(^{133}\) which supports the notion that in an endogenous setting claudin-12 glycosylation is less frequent and instead likely to be a regulated process. Further work is needed to determine the extent of claudin glycosylation in a physiologic setting, although most data in the literature suggests that \( N \)-glycosylation of claudins in situ is an exceptionally rare event.

Another program called NetOglyc,\(^{134}\) which produces neural network predictions of mucin type \( \text{GalNAc} \) \( O \)-glycosylation sites in mammalian proteins, was also used to analyze claudins. Only sites with scores \( >0.5 \) are predicted as glycosylated and marked as positive for \( O \)-glycosylation. Using this approach, all members of the claudins, except 4, 7, 9 and 24 were predicted to be \( O \)-glycosylated (Table 4).

Table 4. Identification of potential \( O \)-glycosylation sites in human claudins, using NetO Glyc 4.0\(^{134}\).

A threshold of 0.5 was applied for this analysis.

<table>
<thead>
<tr>
<th>Gene/Protein</th>
<th>( O )-glycosylation in the C-terminal tail</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLDN1</td>
<td>Y (T195)</td>
</tr>
<tr>
<td>CLDN2</td>
<td>Y (T204, S206, S207)</td>
</tr>
<tr>
<td>CLDN3</td>
<td>Y (S196, S200, T201, S206)</td>
</tr>
<tr>
<td>CLDN4</td>
<td>N</td>
</tr>
<tr>
<td>-------</td>
<td>---</td>
</tr>
<tr>
<td>CLDN5</td>
<td>Y (S201, T207)</td>
</tr>
<tr>
<td>CLDN6</td>
<td>Y (S194, S201, T202, S203, S208, S212)</td>
</tr>
<tr>
<td>CLDN7</td>
<td>N</td>
</tr>
<tr>
<td>CLDN8</td>
<td>Y (S201, T204, T205, S208, S215)</td>
</tr>
<tr>
<td>CLDN9</td>
<td>N</td>
</tr>
<tr>
<td>CLDN10</td>
<td>Y (T198, S199, S202, T215, T216)</td>
</tr>
<tr>
<td>CLDN11</td>
<td>Y (S196)</td>
</tr>
<tr>
<td>CLDN12</td>
<td>Y (S212)</td>
</tr>
<tr>
<td>CLDN14</td>
<td>Y (T201, T202, T203, T204, T207, S224)</td>
</tr>
<tr>
<td>CLDN15</td>
<td>Y (S204, S211)</td>
</tr>
<tr>
<td>CLDN16</td>
<td>Y (S287, S289)</td>
</tr>
<tr>
<td>CLDN17</td>
<td>Y (T207, T213, T214)</td>
</tr>
<tr>
<td>CLDN18</td>
<td>Y (S210, S228)</td>
</tr>
<tr>
<td>CLDN19</td>
<td>Y (S195, S204)</td>
</tr>
<tr>
<td>CLDN20</td>
<td>Y (T198, S203)</td>
</tr>
<tr>
<td>CLDN22</td>
<td>Y (T204)</td>
</tr>
<tr>
<td>CLDN23</td>
<td>Y (S197, S203, S204, S206)</td>
</tr>
<tr>
<td>CLDN24</td>
<td>N</td>
</tr>
<tr>
<td>CLDN25</td>
<td>Y (S207)</td>
</tr>
</tbody>
</table>

Interestingly, all of the predicted O-glycosylation sites are in close-proximity to known or predicted phosphorylation sites of claudins in the C-terminal region.\(^{72,89,135}\) This close proximity of the two PTM types may suggest potential crosstalk between O-GlcNac modification and phosphorylation at these positions. The YinOYang server\(^ {131}\) can calculate the O-glycosylation potential for all Ser, Thr, and Tyr residues in a protein sequence and crosscheck these sites against NetPhos 2.0\(^ {136}\) predictions for potential phosphorylation sites to determine potential Yin Yang sites with a high possibility for both modifications. Currently the experimental evidence for mammalian claudin glycosylation exists only in mice on S241 of claudin-16.\(^ {137}\)

It should be noted that the above predictions are purely sequence-based and do not incorporate the structural information of the individual claudin proteins. Given the importance of glycosylation in protein sorting and endocytosis, a more in-depth investigation of potential claudin glycosylation in human is warranted to help understand whether these modifications play roles in regulating claudin localization and recycling.
Proteolysis

Claudin proteolysis is known to alter the barrier function of intestinal epithelial cells in human and animals.\textsuperscript{138,139} Intestinal barrier integrity is often compromised in autoimmune diseases such as coeliac or inflammatory bowel disease, in which gut/intestinal permeability is higher than normal.\textsuperscript{140} Such barrier disruption is frequently mediated by pro-inflammatory cytokines such as IFN-\(\gamma\), partly through downstream selective cleavage of some claudins. In particular, in T84 (intestinal human epithelial) cells, claudin-2 is specifically cleaved, by yet-to-be identified proteases, in either the ICL or ECL2 after IFN-\(\gamma\) incubation, whilst claudins 1, 3, and 4 in the same cells remain intact.\textsuperscript{141} In addition, the basal expression level of all claudin is reduced by IFN-\(\gamma\) treatment, and a non-specific Ser protease inhibitor (AEBSF) is able to significantly “rescue” this claudin-2 expression loss and cleavage. SwissProt bioinformatics analysis has identified a Ser protease cleavage site in ECL2 of claudin-2 but not in claudins 1, 3 and 4. The transmembrane claudin-2 cleavage is restricted to the Triton-X100 soluble membrane fractions but not in the Triton-X100 insoluble cytoskeletal fraction that contains uncleaved claudin-2. This may suggest that claudin proteolysis is dependent on claudin localization and thus spatially limited, perhaps due to limited access of proteolytic enzymes to the tight junction associated claudin pool. Nonetheless, most evidence to date suggests that the net effect of proteolysis is to decrease overall junction associated claudin protein content, perhaps by inhibiting incorporation of newly synthesized claudin into tight junctions. Claudin proteolysis may also contribute to post-translational control of claudin turnover. For instance, claudin-2 and claudin-4 have half lives of 12 and 4 h respectively, which is determined by the C-terminal cytoplasmic tail as demonstrated using chimeric claudin constructs in MDCK cells.\textsuperscript{142} Claudin-5 turnover is cell dependent, ranging from 70 min in HUVEC\textsuperscript{124} to over 3 hours for bovine retinal endothelial cells.\textsuperscript{143}

4. Mass spectrometry of endogenous claudins
More recently, MS analysis, often coupled with affinity enrichment, has seen wider use in characterizing endogenous claudins at the protein level given the advantage of the technique in identifying low-abundance membrane proteins. An early demonstration of a claudin-targeted enrichment strategy prior to MS employed a GST-fusion affinity column made from *Clostridium perfringens enterotoxin* (CPE) that binds to the ECL2 of claudins 3, 4, and 7.\textsuperscript{144,145} In these reports, affinity enriched claudins from cultured rat cholangiocytes were trypsin digested and subjected to first liquid chromatography (LC) and then MS to identify the enriched claudins. Co-elution of some non-CPE binding claudins was also observed by immunoblotting and MS. Relative MS protein quantification using SILAC (stable isotope labeling by amino acids in cell culture) also identified potential plasma membrane associated claudin-binding proteins as well as claudin-binding proteins in the cytosol, nucleus and mitochondria. Endogenous human claudin-5 in brain endothelial cells has been enriched by cell fractionation for mass spectrometry.\textsuperscript{146} Claudin-5 partners were also identified by co-immunoprecipitation and LC-MS. It was demonstrated with additional siRNA experiments that G-protein subunit αi2 was a new partner of claudin-5 and required for proper claudin-5 assembly in the tight junction.

Recently, a claudin interactome for *Drosophila* has been reported.\textsuperscript{147} Specific antibodies were generated against the claudin Megatrachea in the fly embryos. Subsequent immuoprecipitation, followed by MS analysis, identified 142 protein partners of the claudin Megatrachea, including 10 bona fide members of the septate junctional complex, the invertebrate equivalent to the apical junctional complex. Further validation by RNA interference uncovered new protein partners for Megatrachea interaction including members of the clathrin-mediated vesicle proteins. This is analogous to human claudin-4 internalization by clathrin-mediated endocytosis that depends on a sorting sequence on the ICL, and also suggest that there are significant parallels between invertebrate and vertebrate claudin-based protein complexes and regulation.\textsuperscript{148} A very recent
proteomics investigation on proteins proximal to claudin-4 and occludin in MDCK II cells also significantly expanded the inventory of proteins at the tight junction, particularly in signaling and endocytic trafficking proteins. While a comprehensive analysis of the endogenous human claudins and their interactomes in various cell types remains absent, these studies serve as important foundations to further proteome-wide investigations for human endogenous claudins.

Membrane proteomics of endogenous claudins

Because endogenous claudins are low-abundance proteins in vertebrate epithelial cells, it is difficult to profile their expression comprehensively without protein enrichment techniques. Nonetheless, it is apparent that there are tissue specificities to the expression of claudin members and claudin PTMs and that signalling proteins have substrate preferences for some claudin members over others. Claudin antibodies are widely available and used for specific immuno-detection, however these need to be carefully monitored and tested for cross-reactivity and nonspecificity. Antibodies enable relative claudin expression levels to be monitored by standard immunoblotting methods as well as immunofluorescence imaging. While these approaches are powerful for investigating individual claudins, they are limited in that they do not enable direct calculation of relative claudin stoichiometry and are difficult to employ to measure PTMs.

The accurate and systematic expression and PTM profiling of claudins with spatial and temporal resolution is currently inaccessible without rigorous methods that can isolate, enrich, and distinguish endogenous claudins. Bottom-up membrane proteomics methods have been applied specifically to enrich membrane junction proteins from intestinal human epithelial cells (T84) or mouse liver tissues. A junction-targeting antibody, as validated by immunofluorescence staining, was used to enrich cell junction proteins from T84 that were then immunoblotted to verify the presence of tight junction proteins, stained by silver to verify protein size, and digested for analysis by mass spectrometry. Over nine hundred proteins were identified, half of which were
synaptic or signalling proteins. However, the coverage of known junctional proteins was low.

Nevertheless, this study provided a starting point for more in-depth and adequate analysis of membrane junction proteins such as claudins. A subsequent proteomics investigation utilised membrane protein enrichment protocols involving cell fractionation followed by repeated guanidine treatment and NP-40 extraction. This optimized analysis was able to identify claudin-1 and 3 by MS from the bile canaliculi fraction of mouse livers. More success in claudin protein identification has been reported in rat renal inner medullar collecting duct cells, in which claudin-3, 4, 7, 8, and 10 were found and listed on the IMCD Proteome Database.

The low abundance of the plasma membrane proteome, concomitant with the inadequate proteomic coverage of the full complement of membrane proteins, has been recognized for some time and explains the challenges in proteomic investigations on claudins or other junctional proteins. Hydrophobic proteins are difficult to capture into matrices to enable separation and subsequent identification, which results in hydrophilic proteins being vastly over-represented in proteomic analyses. Methods for the capture and enrichment of hydrophobic membrane proteins addressing this imbalance include isolation in sodium bicarbonate, aqueous two phase partitioning using polyethylene glycol and dextran or non-ionic detergents and sucrose gradients for highly insoluble membrane proteins. Recently, a membrane proteome analysis of human embryonic stem cells (hESCs) identified three claudins: claudin-3, 6, and 7. Claudin-6 was the most abundant of the three, which is consistent with genetic analysis showing claudin-6 as the highest expressed claudin member in hESCs. Although claudin-2, 4, and 12 genes were also reported to express in many hESCs, their protein products were not identified in this analysis. Compared to a prior hESC proteomics study that did not identify any claudin, methods with membrane protein enrichment may help with enhancing claudin and tight junction membrane protein coverage.

**Top-down proteomics of claudins**
Top-down MS has been in decades of development for characterizing intact proteins.\textsuperscript{162-166} Recently, top-down proteomics has emerged as a powerful approach for characterizing proteins in their native and intact form (Figure 3).\textsuperscript{167-171} Unlike bottom-up proteomics, in which proteins are broken into peptide fragments for identification, top-down proteomics reveals the precise complete “proteoform(s)” of a protein without loss of the chemical information embedded in the structure of the protein or proteoform.\textsuperscript{170} While bottom-up proteomics will remain a powerful approach for obtaining a comprehensive claudin overview, the high level of homology between claudin family members requires further validation for reliable protein identification and characterization. Top-down proteomics in this regard holds the definite promise of capturing the exact chemical details of claudins, which are well below 50 kD in size, with high fidelity and thus will be able to uncover authentic biological activities encoded by these claudin protein characteristics. In addition, the coverage bias in bottom-up proteomics toward more abundant proteins is better mitigated in the top-down approach. This is particularly relevant to endogenous claudins because generally low abundance membrane proteins with a high level of homology can be prone to misidentification and irreproducible quantification in proteomics.\textsuperscript{172,173}

Advances in top-down proteomics methodology, in particular with improved protein ion excitation and resolution in mass spectrometry for large protein ion fragmentation, have already enabled large-scale, robust protein identification in human cells.\textsuperscript{174-176} For example, HeLa S3 cells have been fractionated by multiple methods in sequence including solution isoelectric focusing, gel-eluted liquid fraction entrapment electrophoresis, and nanocapillary liquid chromatography.\textsuperscript{174} The resulting fractions were then analyzed by MS in which whole protein molecules could be ionized and fragmented in tandem for protein identification. In total over 3,000 protein products from 1,043 genes were found, displaying a full range of protein diversification by PTMs, RNA splicing and proteolysis. Even proteins as large as over 100 kDa and up to 11 transmembrane helices were mapped. More importantly, unknown isoforms of endogenous proteins related to senescence were
uncovered, thus demonstrating the discovery power of top-down proteomics in unravelling complex biology. This approach has also been expanded to other cell lines such as H1299 with improved proteome coverage of over 5,000 proteoforms, including previously unknown lipid-anchored and hyperphosphorylated proteins. In these HeLa S3 cells, multiple claudin proteoforms have been identified, including claudin-2, 7, 15 and 17. The top-down proteomics approach, combined with specific claudin enrichment protocols, will likely further increase claudin coverage with reliable PTM information in many other cell types. In addition to a combination of bottom-up and top-down proteomics techniques to ensure sensitive detection and extensive coverage of low-abundance membrane proteins in their native states with accurate PTM information, the “middle-down” approach, in which limited digestion is used in conjunction with top-down techniques, is also amenable to extracting chemical information in proteomics. This approach has been in use for understanding histone and ribosome PTMs and may very well be suited for the claudin system.

Figure 3. Top-down and bottom-up proteomics: advantages (in bold) and disadvantages (in grey).

Targeted proteomics of claudins

In targeted proteomics, the mass analysis is focused on one protein or a small set of proteins to reduce the mass bias and interference inherent in complex biological samples for improving sensitivity and quantification. As such, targeted proteomics of the claudin system represents a more quantitative option for claudin stoichiometry and PTM characterization with enhanced accuracy and precision. One earlier example of such approach was demonstrated in quantifying low abundance
proteins, such as growth hormones, in very complex and highly dynamic human plasma samples. Since then, target proteomics has become widely adopted in both top-down and bottom-up proteomics for biomarker validation, human plasma proteome analysis, multiplex protein activity assays, PTM profiling, and mass-linked immuno-selective assays. Great progress has also been made in improving the quantification, throughput and bioinformatics analysis aspects essential for targeted proteomics to scale up and accommodate broader biological applications. Recently, targeted proteomics has enabled quantitation of claudin-5 expression in the plasma membrane fraction of a human hCMEC/D3 cell line (0.879 fmol claudin-5/µg protein) as well as in human brain microvessels (3.39 fmol claudin-5/µg protein). Similar levels of claudin-5 expression (6–8 fmol/ µg protein) are also reported from investigations with mouse brain capillaries. For the claudin system that has many homologous members, these advances in targeted proteomics will continue to facilitate a more reliable analysis of low-abundance claudin expression and PTM profiles with temporal and spatial resolution.

5. Chemical capture and enrichment of endogenous claudins for chemical proteomics

Claudin extraction and subcellular fractionation

Existing subcellular or cell compartment fractionation protocols provide spatial segregation and consequentially also enrichment of proteins. However, cytoplasmic aggregates are often observed and can render fractionation inefficient. Because every cultured cell line has different cytoplasmic and cytoskeletal organisations, optimal conditions for an ideal homogenate before fractionation can be highly variant and difficult to obtain. Nonetheless, cellular fractionation has been used regularly provided that some quality control measures such as marker enzymes or morphological analysis are also employed. Subcellular localization of claudins has been investigated in monolayers of intestinal Caco-2 cells, where knockdown of endogenous protein kinase C theta activity decreased membrane- and cytoskeletal-associated claudin-1 and 4, and
increased the cytosolic pool of these claudins as detected by fractionation and immunoblotting.\textsuperscript{200} In Caco-2 cells, depletion of cholesterol by methyl-\(\beta\)-cyclodextrin leads to displacement of claudin-3, 4, and 7 from the cholesterol rich membrane domains along with other TJ proteins such as JAM-A and occludin. However, claudin-1 is not affected by this depletion.\textsuperscript{201} Overall, claudin extraction methods have been investigated, albeit more consistently in animal cell models such in MDCK cells.\textsuperscript{150} For example, in MDCK II cells, the extractability of claudins is not affected by cholesterol levels, however different detergents, such as Triton X-100 or CHAPS, partition different claudins into different sizes of aggregates.\textsuperscript{202} Also in MDCK cells, sodium caprate is shown to increase solubilities of claudin-4 and 5 in Triton X-100 but not claudin-1, 2, and 3.\textsuperscript{203} As it is widely recognised that the ability of a detergent to extract membrane proteins is cell-type dependent and different detergents differ in ability to solubilize different membrane proteins,\textsuperscript{150,204} extraction or fractionation of claudins or claudin proteoforms from various cell types and cellular compartments will remain extremely challenging and likely require cell-type specific optimisation of isolation protocols prior to mass spectrometry.

**Covalent tagging of endogenous claudins**

Another related issue is claudin enrichment, which can be partially achieved by subcellular fractionation. However, multi-span membrane proteins such as claudins are, in general, more resistant to extraction and identification. While methods on the basis of non-covalent binding are helpful,\textsuperscript{205} approaches for covalent labelling of endogenous claudins are needed for a reproducible capture and enrichment protocol with spatial resolution even at very low expression levels. The approach of chemical proteomics with protein tagging, in which proteins are modified by small molecule probes carrying an affinity anchor, fluorescent reporter, and other functional groups for post-capture analysis, may be suited for system-wide profiling of protein or enzyme activities.\textsuperscript{206-210} Covalent labelling of claudins in the plasma membrane is frequently achieved by using amine or thiol reactive small molecule probes.\textsuperscript{211-215}
The thiol-based labelling strategy is used more widely for labelling cysteines in claudins from both human and other organisms. Reactive epitopes such as non-cell permeable biotinylated NHS esters were used to label HEK293 cells transfected with a mutant claudin-5 expression vector in order to quantify the plasma membrane localization level of the claudin-5 mutant. Cysteine mutants of claudin-2 were generated in MDCK I cells for thiol-labelling in order to determine the protein surface accessibility of claudin-2 in cells. A thiol cross-linker was also successfully utilized to identify, by MS after covalent labelling, the tetraspanin interacting partners in A431 and A549 cells. Amine-reactive labels have also shown the ability to label claudins in cells or animals. New-born mice with or without claudin-1 were subcutaneously injected with an amine-reactive, biotinylation solution in order to assess the extent of barrier function loss in the mouse epidermis lacking claudin-1. Amine-reactive biotinylation agents likewise showed labelling of tight junction claudins in the mouse Eph4 cells junctions for permeability visualization upon treatment of fluorescence labelled-avidin. These precedents suggest that cell surface claudins can be accessible to covalent chemical tagging.

Claudins have also been found in other cellular compartments such as endosomes that likely reflect intracellular compartments active in tight junction turnover. Specifically, surface claudins in MDCK II cells have been differentiated from internal claudins by covalent surface biotinylation labelling of the cell surface. This differential labelling in these cells was able to demonstrate in MDCK cells that the endocytic recycling of claudin-1 and 2 was affected in the presence of an endocytosis inhibitor, resulting in the intracellular accumulation of these claudins, while other claudins were not affected. Chemical tagging of claudins with spatial resolution may be desirable for tracing claudin dynamics in the cell but requires further development of new chemical tags that are specific for endogenous claudins. Such improved chemical tools, in conjunction with subcellular fractionation and high-fidelity methods from advanced proteomics technology, will help
characterize chemically claudin protein forms as potentially the combinatorial protein codes that coordinate cell contact signalling and regulate claudin turnover to control tissue identity and integrity.75

6. Toward molecular identification of claudin-mediated cell-contact signaling platforms

Multicellular organisms rely on coordinated cellular processes to maintain tissue stability and coordinate cell physiology. Cell-cell contact sites represent a critical structure that enables intercellular communication by organizing the formation of large multiprotein complexes (Figure 4). For example, intercellular junctions act as sensors for cell contact that regulate signaling by forming protein complexes that activate kinases and also regulate gene expression by coordinating the transit of transcription factors between the plasma membrane and nucleus.220,221,222,15,44,223 Relevant to claudins, the relatively unstructured intracellular C-terminal region contains several putative and identified binding motifs that promote protein-protein interactions, most prominently the PDZ binding motif.1 Loss of the C-terminal tail leads to intracellular retention of claudins in the ER and proteasome degradation, possibly by improper protein folding or lack of binding to trafficking partners.224-226 These serial truncation experiments, however, demonstrate that the PDZ binding motif is not required for proper claudin trafficking to the plasma membrane and reveal the critical role of the juxtamembrane region of about 20 amino acids in localization control of claudins.

Figure 4. Simplified scheme of the epithelial apical junctional complex, showing a subset of protein constituents, including tetraspan transmembrane proteins (claudins, occludin), single pass transmembrane proteins (JAM-A, cadherin), cytosolic scaffold proteins (ZO-1, ZO-2, catenins), signaling proteins (RhoA, MLCK) and the actin cytoskeleton. The Crumbs/PALS/PATJ polarity
complex is also shown, which defines the apical/basolateral axis and is directly associated with tight
junctions via scaffold protein interactions.

Many excellent reviews and articles have already emphasized that specific claudins, localized to
various subcellular locations, interact with many other proteins to form specific scaffolds for
signaling complexes in different cell types of various organisms. Some recent examples in human cells are highlighted here, while many other related examples are also known in
other mammals such as rats and mice, along with recent proteomic analyses of MDCK-II
cells that significantly expand the list of known and new signaling proteins assembled in and around
the tight junction as discussed earlier.

HEK293 cells are commonly used as nonpolarised, tight junction-free cell models for cell-contact
investigations after transfection with claudins. In transfected HEK-293 cells, classic claudins
such as claudin-1, 3 and 5 are found mainly at the cell-cell contact points and provide multilateral
interaction with other tight junction proteins such as occludin, tricellulin, and MarvelD3 in order to
secure the TJ strand network for stable barrier function. The dynamic aspect of claudin localization in the cell membrane is also demonstrated for claudin-2 and 4 in SKCO 15 or Caco-2 BBE (Brush Border Expressing) cells in response to environmental changes such as cytokine exposure. This is consistent with the existing notion that claudins can mediate junction remodeling and recruit other TJ proteins.

Claudin expression and dynamics are also known to be regulated by epithelial cell adhesion molecule (EpCAM) in epithelial cells. Complexes between claudin-7 and EpCAM have been demonstrated by co-immunoprecipitation experiments and shown to recruit proteins into tetraspanin enriched membrane microdomains. In HEK-293 cells expressing EpCAM, the proliferative activities of EpCAM require formation of a complex with claudin-7 in order to disrupt EpCAM oligomerization and activate mitogenic signaling. Mice injected with HEK-EpCAM-claudin-7 cells developed tumors while HEK cells without the EpCAM-claudin-7 complex did not induce tumors in the animal. In a transgenic mice model expressing human claudin-1, Notch signaling, a critical regulator of intestinal epithelial cell differentiation and lineage determination, is up-regulated by claudin-1 overexpression with concurrent up-regulation of MMP-9 activity and p-ERK signaling. In MDA-MB-231 cells, claudin-5 overexpression increases cell motility and co-immunoprecipitates with N-WASP and ROCK1, suggesting a possible role of claudin-5 in metastasis. In HEK293 cells, claudin-2 expression is regulated by components in Wnt signaling whereby both LEF-1 and β-catenin expressions enhanced the promoter activity of claudin-2. In human colonic cancer cell line SW480, claudin-1 is also identified as a downstream target of Wnt/β-catenin signaling, and increased claudin-1 expression is observed in all 16 primary colorectal cancers investigated with significant claudin-1 localization detected in cell-cell boundaries and the cytoplasm.
The non-junctional localization of some claudins has also been linked to disease progression in various epithelial cancers, which has been extensively reviewed.\textsuperscript{74-77,89,244-246} For example, in craniopharyngioma, claudin-1 expression is reduced in invasive tumors compared to noninvasive ones, and increased cytoplasmic claudin-1 localization is detected by immunostaining in tumor cells that border brain tissues and dura.\textsuperscript{247} In invasive breast cancer, claudin-1 is frequently down-regulated but up-regulated in some aggressive subtypes of basal-like breast cancer.\textsuperscript{248} Higher claudin-1 protein expression is observed in tumors from older patients, accompanied by more frequent cytoplasmic claudin-1 localization.\textsuperscript{249} In melanoma tissue samples, both cytoplasmic and nuclear claudin-1 expression is frequently observed, and nuclear claudin-1 expression is drastically reduced in lymph node metastases.\textsuperscript{250} In vitro models demonstrate a positive link between higher MMP-2 activity and higher cytoplasmic claudin-1 expression level, suggesting a role of claudin-1 in melanoma progression. This is not restricted to claudin-1, since in proliferating human lung adenocarcinoma A549 cells or tissues, claudin-2 is found in the nucleus as part of a complex containing ZO-1, ZONAB, and cyclin D1.\textsuperscript{73} Compared to the wild type, a S208A mutant of claudin-2 exhibits a higher extent of nuclear localization in its dephosphorylated form, suggesting that nuclear claudin-2 may serve to retain ZONAB and cyclin D1 in the nucleus to enhance cell proliferation. Such proliferation promoting effect is also suggested for claudin-1 in osteoblasts.\textsuperscript{251} Likewise, both claudin-7 transcript and protein expression levels are frequently elevated in epithelial ovarian tumors but not in normal ovarian tissues, and in many ovarian cancer cell lines both cytoplasmic and cell membrane claudin-7 expression are detected by immunostaining.\textsuperscript{252} However, the mechanisms and molecular details behind the change of claudin localization remain unclear, particularly when localized to the cytosol or nucleus.

The diverse localization and binding partners identified using standard biochemical techniques suggest that regulation of claudin function is complex. The acquisition of the exact chemical characteristics of the specific claudin protein forms, from different cell locations by MS, in addition
to their levels of expression, may help elucidate the details of the signaling context specific to
cell/tissue types or disease states. Initial attempts at identifying protein partners of claudins are
listed in Table 5. Those interactions were retrieved from Swissprot and IntAct databases. Reflecting
the most prominent function of claudins, namely formation of epithelial barriers via tight junctions,
claudin-claudin and claudin-scaffold protein interactions are well represented in Table 5. This list
will certainly be expanded and refined as more claudin specific protein-protein interactions are
characterized from an expanded range of host tissues.

Table 5. Protein-protein interactions with claudins (compiled from www.swissprot.org and
http://www.ebi.ac.uk/intact/).

<table>
<thead>
<tr>
<th>Gene/Protein</th>
<th>Protein/Protein Interaction Partners (UniProtKB/Swiss-Prot manual annotation)</th>
<th>Binary Interactions (IntAct annotation; in bold when found by two independent experiments)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLDN1*</td>
<td>CLDN3 (but not CLDN2); TJP-ZO*; MPDZ; INADL; HCV E1/E2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>BRD4; DRG1; TACSTD2</td>
</tr>
<tr>
<td>CLDN2*</td>
<td>CLDN3 (but not CLDN1); TJP-ZO*;</td>
<td>KRTAP4-12; KRT31; NOTCH2NL; TJP1</td>
</tr>
<tr>
<td>CLDN3*</td>
<td>CLDN1 and CLDN2; TJP-ZO*;</td>
<td></td>
</tr>
<tr>
<td>CLDN4</td>
<td>TJP-ZO*; EPHA2</td>
<td>TACSTD2</td>
</tr>
<tr>
<td>CLDN5</td>
<td>TJP-ZO*; MPDZ</td>
<td>-</td>
</tr>
<tr>
<td>CLDN6</td>
<td>TJP-ZO*</td>
<td>-</td>
</tr>
<tr>
<td>CLDN7</td>
<td>TJP-ZO*; EPCAM&lt;sup&gt;b&lt;/sup&gt;</td>
<td>HGD; RHOXF2; SYNE4; TACSTD2</td>
</tr>
<tr>
<td>CLDN8</td>
<td>TJP-ZO*</td>
<td>CCDC155; SYNE4</td>
</tr>
<tr>
<td>CLDN11</td>
<td>TSPAN3</td>
<td>-</td>
</tr>
<tr>
<td>CLDN12</td>
<td>-</td>
<td>ECHS1; SEC13; STRN4</td>
</tr>
<tr>
<td>CLDN15</td>
<td>CLDN15 (linear homooligomers)</td>
<td>GEM</td>
</tr>
<tr>
<td>CLDN16</td>
<td>-</td>
<td>CLDN19; Cldn14&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>CLDN19</td>
<td>-</td>
<td>CLDN16; SRPK2; Cldn14&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Can form homo- and heteropolymers with other claudins.
<sup>b</sup>Interaction requires claudin phosphorylation.
<sup>c</sup>Direct interaction with HCV proteins has not been experimentally validated (see text).
<sup>d</sup>Experiment done with mouse proteins.

* TJP1/ZO-1; TJP2/ZO-2; TJP3/ZO-3; these protein partners are common to claudins 1 to 8.

The range of protein partnership for claudins can be illustrated by a protein interaction network
diagram using human claudin-1 as an example (Figure 5). Direct protein interactions with human
claudin-1 were obtained by careful mining of the literature. Extended human protein interactions
between protein partners of claudin-1 and other proteins were retrieved from UniProtKB/Swiss-Prot
and from IntAct (only interactions found by more than one experiment). The domain-specific
annotations of these protein interactions with claudin-1 are listed in Table 6. As not all the
interactions described in literature are yet imported in UniProtKB/SwissProt and Intact, this
interaction network may be more complex.

Figure 5. A diagram of the protein interaction network of claudin-1. The diagram was made using
Cytoscape. In orange are the intrinsic membrane proteins with at least one transmembrane
domain. In yellow are the peripheral membrane proteins. These two groups contain plasma
membrane proteins and endomembrane system proteins (endosome, endoplasmic reticulum, Golgi
apparatus, and cytoplasmic vesicles). Membrane proteins from mitochondria or nucleus are not
included in these two groups. In green are cytosolic, nuclear and mitochondrial proteins. In blue are
secreted proteins (with a signal peptide and no transmembrane domains). The pink border indicates
proteins found in tight junctions. Red edges connect CLDN1 (diamond) with its direct interacting
partners: TJP1/ZO-1, TJP2/ZO-2, TJP3/ZO-3, MPDZ, INADL, TACSTD2, SRC, KRT76, LNX1,
CD81, CD9, OCLN, MARVELD2, MARVELD3, CLDN3, CLDN7, MMP14, MMP2, EFNB1
(Rectangles). Interactions with non-human proteins are not included in the network but briefly
described in the text.
Table 6. Domain-specific annotations of protein interactions with claudin-1.

<table>
<thead>
<tr>
<th>Claudin 1-domains</th>
<th>Protein/Protein Interaction Partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytosolic tail</td>
<td>TJP1/ZO-1\textsuperscript{254}</td>
</tr>
<tr>
<td></td>
<td>TJP2/ZO-2\textsuperscript{254}</td>
</tr>
<tr>
<td></td>
<td>TJP3/ZO-3\textsuperscript{254}</td>
</tr>
<tr>
<td></td>
<td>MPDZ/MUPP1\textsuperscript{255}</td>
</tr>
<tr>
<td></td>
<td>INADL/PATJ\textsuperscript{256}</td>
</tr>
<tr>
<td></td>
<td>TACSTD2\textsuperscript{257}</td>
</tr>
<tr>
<td>cis (transmembrane)</td>
<td>SRC\textsuperscript{258}</td>
</tr>
<tr>
<td></td>
<td>KRT76\textsuperscript{259}</td>
</tr>
<tr>
<td></td>
<td>LNX1\textsuperscript{125,260}</td>
</tr>
<tr>
<td></td>
<td>CD81\textsuperscript{261}</td>
</tr>
<tr>
<td></td>
<td>CD9\textsuperscript{211}</td>
</tr>
<tr>
<td></td>
<td>OCL\textsuperscript{262}</td>
</tr>
<tr>
<td></td>
<td>MARVELD2\textsuperscript{263}</td>
</tr>
<tr>
<td></td>
<td>MARVELD3\textsuperscript{280}</td>
</tr>
<tr>
<td></td>
<td>CLDN3\textsuperscript{233,264}</td>
</tr>
<tr>
<td></td>
<td>CLDN\textsuperscript{265}</td>
</tr>
<tr>
<td></td>
<td>CLDN7\textsuperscript{237}</td>
</tr>
<tr>
<td></td>
<td>CLDN\textsuperscript{265}</td>
</tr>
<tr>
<td></td>
<td>CLDN3\textsuperscript{264}</td>
</tr>
<tr>
<td></td>
<td>MMP1\textsuperscript{266}</td>
</tr>
<tr>
<td></td>
<td>MMP2\textsuperscript{266}</td>
</tr>
<tr>
<td></td>
<td>EFNBI\textsuperscript{267}</td>
</tr>
</tbody>
</table>

While the human claudin-1 diagram does not include its non-human protein partners, pathogenic proteins have been found to interact with human claudin-1. For example, human claudin-1 along with occludin and the tetraspanin receptor CD81, mediate the initial step of hepatitis C virus (HCV) infection.\textsuperscript{268} During this process CD81 directly interacts with the envelope glycoproteins HCVE1 and HCVE2.\textsuperscript{269} The direct physical interaction between claudin-1 and the HCVE1/2 glycoproteins has not been definitively established but inferred by the ability of the particles expressing the HCV glycoproteins to bind claudin-1-expressing cells.\textsuperscript{270,271} A direct interaction between claudin-1 and the pre-membrane (prM) protein from the dengue virus has been suggested to be important for dengue virus entry.\textsuperscript{272} Interestingly, some of the protein partners of human claudin-1 were also shown to interact with viral proteins: MPDZ with the adenovirus type 9 E4 protein;\textsuperscript{273} LNX1 with the endogenous retrovirus K protein Np9;\textsuperscript{274} and SRC with the hepatitis E virus ORF3 protein.\textsuperscript{275}
Human claudin-1 is also involved in cell infection by non-viral pathogens. A recent study suggests that the intestinal epithelial damage caused by the protozoan Entamoeba histolytica is initiated by the interaction of the virulence complex EhCPADH112 (formed by CP112/EhCP112 and ADH112/EhADH112) with claudin-1, occludin, TJP1/ZO-1 and TJP2/ZO-2 followed by their degradation.\textsuperscript{276}

In addition to standard biophysical techniques, MS is a particularly useful technique for acquiring complementary information on stoichiometry and topology of protein complexes.\textsuperscript{277-283} Although lower in resolution as compared with techniques with atomic resolution, MS-based information is crucial to understanding dynamic protein interactions and networks that govern biological function.\textsuperscript{280} Integrated approaches that combine PTM analysis by MS-based proteomics, protein cross-linking, protein complexes characterization by MS, and structural modeling may offer new structural understanding of endogenous protein partnerships or interactomes.\textsuperscript{278,279,282} Examples of complex multi-component systems that have been successfully analyzed by MS-coupled quantitative proteomics include large protein assemblies such as the human initiation factor (eIF)3 complex and the yeast RNA exosome. Recent advances in ionization and dissociation methods, including surface-induced collision and ion-mobility MS procedures, also greatly improve the prospect of mass spectrometry in characterizing intact protein complexes in their native states.\textsuperscript{277,281} These new technological capacities, in conjunction with enrichment techniques, may begin to help shed light on the supramolecular structures of claudin protein complexes that define the signaling context at and beyond cell junctions.

**Conclusion**

The predominant role for claudins in regulating epithelial and endothelial barrier function necessitates that they need to be able to interact with a multiplicity of proteins in diverse ways, including trans (intercellular) interactions with extracellular domains of proteins across intercellular
junctions, cis (in the plane of the plasma membrane) interactions with transmembrane proteins and binding to scaffold proteins that regulate their organization and stability. Thus, claudins have two roles in regulating tight junction permeability, a direct role by forming different classes of paracellular channels and a more indirect role by participating in the formation of the large multi-protein complexes associated with claudin-containing tight junction strands. Although considerable progress has been made to understand how claudin diversity affects paracellular channel selectivity, precisely defining claudin binding partners at a quantitative level is still at an early stage of research. The ensemble of proteins linked to claudins in tight junctions affects paracellular flux through control of strand morphology, localization and stability. Understanding how protein cofactors mediate intercellular interactions, cross-linking of claudins with the actin cytoskeleton and integration with signal transduction pathways will provide insights into how tissue barrier function is regulated in normal and diseased organs, as well as potential therapeutic targets for the manipulation of paracellular flux.

The regulatory role conferred by the claudin family of proteins in tight junctions is complemented by roles in morphogenesis, tissue maintenance, growth control, and cell migration. All of these processes require constant protein processing and modification. Individual claudin members have been investigated biochemically or in cell models/tissues for their expression and localization profile, and some claudin members have been characterized by MS in addition to antibody-based detection. Many claudin PTMs, such as C-terminal phosphorylation and palmitoylation have been identified. There are no doubt many more claudin phosphorylation sites remaining to be identified and future studies are needed to provide added evidence in support of PTMs such as ubiquitination, SUMOylation, palmitoylation, and glycosylation.

More importantly, it remains an enormous challenge to acquire a direct and accurate profile of claudins, PTMs and interacting partners, considering that claudins are low-abundance, membrane-
bound, and heavily modified transmembrane proteins. While the function of tight junction-associated claudins is largely to regulate the paracellular barrier, non-junctional claudins in the lateral plasma membrane, cytosol or nucleus likely have distinct roles in cell regulation. Complementary to antibody-based detection that will continue to be a major method of investigation, advanced chemical capture, enrichment, and analysis tools, tailored for high-precision MS-based proteomic profiling in a multi-pronged and integrated manner, could provide unbiased approaches to identify claudin PTM and interacting proteins. Defining claudin PTMs and interacting proteins in a quantitative and stoichiometric manner has the potential to determine whether the spectrum of claudin expression not only controls paracellular flux but also integrates the flow of information by acting in effect as a cell “identity code” that is recognized across intercellular junctions. Understanding this context in which claudins affect cells could enable a nuanced approach to therapeutic design by either modulating claudin barrier function (e.g. for the purposes of drug delivery) or affecting other non-barrier cell functions regulated by claudins, such as cell proliferation and migration. By defining the different molecular contexts of claudin expression, systems proteomics is a valuable approach that will make it possible to consider specifically targeting junctional vs. non-junctional roles for claudins.

Acknowledgment

We thank the Australian Research Council (LIEF150100161 to F. L. and S. R.), the National Institute of Health (R01-HL116958 to M. K.), the National Health and Medical Research Council of Australia (NHMRC APP1010303 to M. B.), and the New South Wale Cancer Council (RG10-04 & RG08-16 to M. B.) for financial support.


22 Tamura, A. & Tsukita, S. Paracellular barrier and channel functions of TJ claudins in organizing biological systems: advances in the field of barriology revealed in knockout


41 Uchida, Y. *et al.* A study protocol for quantitative targeted absolute proteomics (QTAP) by LC-MS/MS: application for inter-strain differences in protein expression levels of transporters, receptors, claudin-5, and marker proteins at the blood-brain barrier in ddY,


Oku, N., Sasabe, E., Ueta, E., Yamamoto, T. & Osaki, T. Tight junction protein claudin-1 enhances the invasive activity of oral squamous cell carcinoma cells by promoting cleavage...
of laminin-5 gamma2 chain via matrix metalloproteinase (MMP)-2 and membrane-type

245 Tokes, A. M. *et al.* Claudin-1, -3 and -4 proteins and mRNA expression in benign and
malignant breast lesions: a research study. *Breast cancer research : BCR* **7**, R296-305,

246 Swisshelm, K., Macek, R. & Kubbies, M. Role of claudins in tumorigenesis. *Advanced drug

247 Stache, C. *et al.* Tight junction protein claudin-1 is differentially expressed in
cranioopharyngioma subtypes and indicates invasive tumor growth. *Neuro-oncology* **16**, 256-

248 Blanchard, A. A. *et al.* Claudins 1, 3, and 4 protein expression in ER negative breast cancer
correlates with markers of the basal phenotype. *Virchows Archiv : an international journal

249 Blanchard, A. A. *et al.* Claudin 1 expression in basal-like breast cancer is related to patient

250 Leotlela, P. D. *et al.* Claudin-1 overexpression in melanoma is regulated by PKC and

251 Alshbool, F. Z. & Mohan, S. Differential expression of claudin family members during
osteoblast and osteoclast differentiation: Cldn-1 is a novel positive regulator of

252 Dahiya, N., Becker, K. G., Wood, W. H., Zhang, Y. & Morin, P. J. Claudin-7 Is Frequently

253 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular

254 Itoh, M. *et al.* Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2, and

255 Hamazaki, Y., Itoh, M., Sasaki, H., Furuse, M. & Tsukita, S. Multi-PDZ domain protein 1
(MUPP1) is concentrated at tight junctions through its possible interaction with claudin-1 and

256 Roh, M. H., Liu, C. J., Laurinec, S. & Margolis, B. The carboxyl terminus of zona

257 Nakatsukasa, M. *et al.* Tumorb-associated calcium signal transducer 2 is required for the
proper subcellular localization of claudin 1 and 7: implications in the pathogenesis of
gelatinous drop-like corneal dystrophy. *The American journal of pathology* **177**, 1344-1355,

258 Singh, A. B., Sharma, A. & Dhawan, P. Claudin-1 expression confers resistance to anoikis
in colon cancer cells in a Src-dependent manner. *Carcinogenesis* **33**, 2538-2547,

259 DiTommaso, T. *et al.* Keratin 76 is required for tight junction function and maintenance of

260 Guo, Z. *et al.* Proteomics strategy to identify substrates of LNX, a PDZ domain-containing

261 Harris, H. J. *et al.* CD81 and claudin 1 coreceptor association: role in hepatitis C virus entry.

262 Raleigh, D. R. *et al.* Occludin S408 phosphorylation regulates tight junction protein


Tanaka, M., Kamata, R. & Sakai, R. Phosphorylation of ephrin-B1 via the interaction with claudin following cell-cell contact formation. The EMBO journal 24, 3700-3711, doi:10.1038/sj.emboj.7600831 (2005).


