Multiscale modelling of claudin-based assemblies: A magnifying glass for novel structures of biological interfaces

Alessandro Berselli a,b,*, Fabio Benfenati a,c,*, Luca Maragliano a,d,*, Giulio Alberini a,c,*

a Center for Synaptic Neuroscience and Technology (NSYN@Unige), Istituto Italiano di Tecnologia, Largo Rosanna Benzi, 10, 16132 Genova, Italy
b Department of Experimental Medicine, Università degli Studi di Genova, Viale Benedetto XV 3, 16132 Genova, Italy
c IRCCS Ospedale Policlinico San Martino, Largo Rosanna Benzi 10, 16132 Genova, Italy
d Department of Life and Environmental Sciences, Polytechnic University of Marche, Via Brece Bianche, 60131 Ancona, Italy

A R T I C L E   I N F O

Article history:
Received 1 July 2022
Received in revised form 24 October 2022
Accepted 24 October 2022
Available online 28 October 2022

Keywords:
Claudin
Tight junctions
Molecular dynamics simulations
Coarse grained molecular dynamics simulations
Free energy calculations
Protein-protein molecular docking

A B S T R A C T

Claudins (Cldns) define a family of transmembrane proteins that are the major determinants of the tight junction integrity and tissue selectivity. They promote the formation of either barriers or ion-selective channels at the interface between two facing cells, across the paracellular space. Multiple Cldn subunits form complexes that include cis- (intracellular) interactions along the membrane of a single cell and trans- (intercellular) interactions across adjacent cells. The first description of Cldn assemblies was provided by electron microscopy, while electrophysiology, mutagenesis and cell biology experiments addressed the functional role of different Cldn homologues. However, the investigation of the molecular details of Cldn subunits and complexes are hampered by the lack of experimental native structures, currently limited to Cldn15. The recent implementation of computer-based techniques greatly contributed to the elucidation of Cldn properties. Molecular dynamics simulations and docking calculations were extensively used to refine the first Cldn multimeric model postulated from the crystal structure of Cldn15, and contributed to the introduction of a novel, alternative, arrangement. While both these multimeric assemblies were found to account for the physiological properties of some family members, they gave conflicting results for others. In this review, we illustrate the major findings on Cldn-based systems that were achieved by using state-of-the-art computational methodologies. The information provided by these results could be useful to improve the characterization of the Cldn properties and help the design of new efficient strategies to control the paracellular transport of drugs or other molecules.

© 2022 Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

The exchange of ions and molecules across epithelia and endothelia is finely tuned by highly selective barriers that preserve the homeostasis of each compartment, such as the blood–brain, the renal and the intestinal barriers [1–5]. The first microscopy images of epithelia identified a tripartite junction between opposing cells, including the zonula occludens (tight junctions), zonula adherens (intermediary junctions) and macula adherens (desmosome) complexes [6]. Further results, based on transmission electron microscopy (TEM), showed that tight junctions (TJs) are organized in multimeric networks, named strands, making contacts between the membranes of adjacent epithelial cells. The complex arrangement of these TJ strands controls the efflux of ions and other permeants through the paracellular spaces between neighboring cells [7].

In a preliminary description, epithelia were categorized in "leaky" or "tight" according to their permeability, with transepithelial electric resistance (TEER) ranging from 101 to 105 Ω cm² and characterized by a variable number of strands and depth in the apical-basal direction [8–10]. Originally, a linear correlation was thought to exist between junctional electric resistance and strand number, resulting from the circuitual representation of TJs as series resistors. In contrast, the first electrophysiology measurements revealed a logarithmic relationship between TEER and strand number [9], which was explained by the postulation of a dynamic model for the TJ structure. Accordingly, the strand was represented with regions, assimilated to pores, that could fluctuate between an "open" and a "closed" state based on specific probabilities that correlate the strand number to the measured TEER [9]. Nevertheless, the subsequent discovery of TJ’s main components occludins [11] and claudins (Cldsns) [12], showed that the TEER measurements of different epithelia did not depend only on the strand number, but were also crucially affected by the TJ strand composition [13–15]. For this reason, the dynamic model [9] was found to be inadequate to reproduce the differences in the TJ tissue-specificity. A schematic representation of the main TJ components is illustrated in Fig. 1. Notably, Cldn proteins were observed to spontaneously assemble into strands at cell–cell contacts [16,17], suggesting a major role in preserving the integrity of the TJ backbone and regulating the paracellular traffic [2,18]. Specifically, they arrange into the TJ meshwork as a result of either homophilic interactions between identical Cldns or heterophilic interactions between different isoforms [19,20].

The strict selectivity of Cldn strands to the passage of permeants such as ions, water and small molecules, hampers the treatment of several pathologies [21] and promotes the development of innovative, effective TJ modulators to enable direct drug delivery to the various tissues [22–25]. Despite the relevant biological role and the interest as potential targets in the rational design of new therapeutic agents, the structural details of Cldn-based complexes remain elusive [5,26]. Recently, the investigation of the Cldn molecular features has been integrated with the use of computer-based techniques able to provide a detailed description of these systems up to the atomistic resolution [27–29]. Several works were published reviewing the progress achieved in the last few decades concerning the investigation of the structural and functional features of Cldns [2,5,15,18,20,30–35]. Nevertheless, most of them are mainly focused on the Cldn physiology [2,5,18,20,30–34], while only few articles emphasize the results achieved by in-silico investigations [15,18,35]. Here, we aim at providing a detailed analysis of the knowledge of Cldn-based systems obtained through the implementation of both all-atom [36–41] and coarse-grained (CG) [42–49] computational methods, and their correlation with live cell studies. General information on the Cldn protein family are provided in the first part of the manuscript (Section 1 General features of claudin structures), followed by the discussion of the Cldn high-order multimeric models (Section 3 Structural models of claudin assemblies from computational approaches) and their transferability among the most investigated homologs, which are treated with dedicated paragraphs (Section 4 Computational insights into the features of specific claudins). Finally, some considerations on the accuracy and limitations of the current computational methods (Section 5 Accuracy and limitations in the study of biological structures with computational modelling methodologies) precede the conclusions, in which a summary of the
2. General features of claudin structures

Claudins are widespread in all compartments of the organism. Despite most of them share very similar structural properties, such as the overall folding, crucial differences in the sequence and three-dimensional structure are found. These changes determine the tissue-specificity of each TJ strand and also affect the capability of Cldns to form intermolecular interactions. In this section, the general features of the Cldn monomers are introduced, followed by the discussion of the available experimental structures and the homophilic or heterophilic cis- and trans-interactions established by distinct homologs to seal the paracellular spaces.

2.1. Structural and functional properties of claudin monomers

Mammalian Cldns form a family of 27 membrane proteins that fold in a four-helix bundle (TM1-4) spanning the membrane bilayer, defining the transmembrane domain (TM1-TM4) [50]. In addition, Cldn topology includes two extracellular loop domains, named ECL1 and ECL2, respectively. ECL1 links TM1 and TM2 and is arranged in a four-stranded β-sheet ([β1–4]) and a short α-helix (ECH), while ECL2 connects TM3 and TM4, providing an additional strand ([β5]) to the ECL β-sheet domain. Two conserved Cys residues in ECL1 stabilize the β-sheet forming a disulfide bond [31,51]. The two loops between the β1 and β2 and between TM3 and β5, named V1 and V2, respectively, are the least conserved across Cldn isoforms [52]. TM2 and TM3 are connected by a loop (ICL) that is intracellular, along with the N/C termini (Fig. 2A). The cytoplasmic C-terminal region is characterized by a conserved PDZ domain which is targeted by zonula occludens (ZO)-1 and –2, to favor the correct Cldn localization within the membrane and regulating the TJ strand motility [53]. Indeed, ZO proteins belong to the membrane-associated guanylate kinase (MAGUK) family [54] and act as linkers between TJ proteins and the actin cytoskeleton [55] (Fig. 1B). This process occurs via the formation of intermitting and dynamic interactions that allow actin filament remodeling, thus permitting cell movement while preserving the correct coupling between cells [53]. To form TJs, Cldns associate via cis-interactions within the same cell membrane and via trans-interactions between facing monomers in adjacent cells [31,56]. While residues in ECL1 are typically the determinants of the TJ ion-selectivity and size of the paracellular cleft, those belonging to ECL2 are responsible for the trans-interactions [31,32,51,57].

From a phylogenetic point of view, Cldns can be grouped into classic (Cldn 1–10, 14, 15, 17 and 19) and non-classic (Cldn 11–13, 16, 18 and 20–27) members. The elements of the first group share three motifs (named Motif I-III, introduced in Fig. 2B) in the ECL domain [5,18] that are fundamental for the strand formation [20,31,33,52–61]. Motif I (residues F, F/Y, N/D/S, P/s, x(5–6), K/R,Y/F/R, E, L/I/M/F) is found in ECL2 and, together with the ECH Motif II (S/T, L/M/I/V, L/F, X, L/V) [18], generates the hydrophobic interface [62]. Motif III (V/I, V/I/L, V/T/I) is identified in the β1-β2 connecting loop, and its contribution to the strand formation was investigated by Cys-scanning mutagenesis and fluorescence resonance energy transfer (FRET) experiments [59].

In addition to the aforementioned classifications, today there is a wide consensus in classifying Cldns according to their tissue-specific ability to form paracellular channels [63] or barriers to the passage of ions and small molecules [7,18]. Specifically, it is known that Cldn2 [64], Cldn10b [65,66], Cldn15 [67], Cldn21 [68] form cation-selective pores, while Cldn10a [69] and Cldn17 [70] allow the permeation of anions. In contrast, a barrier function was demonstrated for other Cldns, such as Cldn1 [71], Cldn5 [72], Cldn11 [73] and Cldn14 [74]. A role in tightening paracellular spaces was also suggested for Cldn3 [75], Cldn8 [76] and Cldn19 [77], but a selective permeability towards divalent cations was detected for the latter when heterotopically associated to the Cldn16 [77–79]. For other Cldn members, the physiological features...
are not fully ascertained. Examples include Cldn4, whose role as barrier [18,31,80] or anion-selective channel [81–85] is still debated, and is dependent on the specific cell type [57,80,84,86].

In addition, Cldn4 was found to interfere with high-order structures of other cation-selective homologs (Cldn2, Cldn7, Cldn15, Cldn19) in MDCK cells, inhibiting the channel activity and augmenting the paracellular barrier. This process was named inter-claudin interference and was observed to take place via the progressive collapse of the Cldn meshworks, strand disassembly and consequent channel disassembly. This phenomenon could be due to changes in the physiological remodeling and spontaneous dynamical rearrangement of TJ strands occurring upon co-expression with Cldn4 [87]. In general, the co-localization of different Cldns within a given TJ meshwork strictly affects its physiological and morphological features. Indeed, super-resolution STED microscopy combined with FRET experiments revealed that barrier-forming Cldns can homogeneously intermix in TJ strands, while channel-forming Cldns are segregated in heterogeneous TJ meshwork to generate balanced, alternated and interspaced ion-specific fluxes providing finely tuned paracellular permeability [88].

### 2.2. The intermolecular association of claudins: homophilic and heterophilic interactions

The attitude of different Cldns to arrange into strands through the association with either identical (homophilic) or different (heterophilic) homologs was extensively explored in the literature [18–20,30,60,78,81,88–90]. Specifically, the Cldn composition affects not only the physiology, but also the morphology of the TJ strand [91], which can be ascribed to classes named A, B or C according to the mesh features described in Ref. [88]. Remarkably, 11 out of the 14 classic Cldns were found to be able to arrange into homophilic strands, while only the homologs Cldn11 and Cldn20 were observed among non-classic members [88].

As a representative example, heterotypic associations between Cldns expressed in the cerebral barriers (Cldn1, Cldn2, Cldn3, Cldn5, Cldn12) were systematically investigated [20]. Results provided by FRET analysis revealed that the heterophilic trans-interactions among Cldn1/Cldn5 and Cldn1/Cldn3 couples occur with a frequency comparable to the respective homophilic interactions of Cldn5 and Cldn3 [20]. In contrast, the non-classic Cldn12 is not capable to generate homophilic trans-interactions due to the lack of aromatic residues (such as Tyr148 in Cldn5 or Tyr147 in Cldn3) in ECL2.

In another work, it was reported that Cldn4 localizes at the TJs in the kidney collecting ducts only in presence of Cldn8 [81]. Similarly, Cldn16 requires the heterophilic cis-interaction with Cldn19 in the thick ascending limb of the Henle’s loop, where it is involved in the reabsorption of divalent cations [77–79]. Cldn3 and Cldn10b are also co-expressed in the thick ascending limb of the Henle’s loop, but only Cldn16, Cldn19 and Cldn3 generate heteropolymeric TJ strands. In contrast, Cldn10b is segregated in separated homophilic multimers arranged in a specific cell subpopulation deputed to Na⁺ reabsorption [78].

Similarly to Cldn10b, also Cldn15 and Cldn11 exclusively form homophilic strands [18]. This feature is associated with slight differences and non-conserved patterns in the ECL sequence domain with respect to those of the other homologs able to generate heterogeneous strands [18]. In particular, Cldn10b and Cldn15 lack one residue in the β1–β2 loop and one in the ECL2 Motif I region. Additionally, the Gly-Leu-Trp motif, typical of other classic Cldns, is replaced by the Asn-Leu-Trp pattern in the homologs Cldn10b and Cldn15 (corresponding to the residues 48–49–50 and 47–48–49, respectively, Fig. 2B). Even further differences are found for Cldn11, which lacks the three motifs typical of classical members, hence explaining the very limited hetero-compatibility [18].

Another study showed that the replacement of the Cldn3 Asn44 residue with Thr, mimicking the Cldn4 sequence, allows the formation of trans-interactions between these two homologs (Cldn3 Asn44Thr and Cldn4), still retaining the ability of Cldn3 to bind Cldn1 and Cldn5 [92]. In addition, also the Cldn4 chimera provided with either ECL1 or ECL2 of Cldn3 gained the compatibility towards Cldn1, 3 and 5. The segregation process of channel-forming Cldns within the TJ meshwork was recently addressed [88]. The authors showed that, while Cldn2 and Cldn10a are spontaneously organized in distinct homophilic regions when coexpressed in COS-7 cells, they are able to interact if a chimeric form of Cldn2 contains the Cldn10a ECL domain.
All these studies clearly emphasize the ability of different Cldns to generate either homophilic or heterophilic strands as a consequence of the conserved patterns located in the ECL domains. This underlines the complexity of the TJ morphology and its dependence on the Cldn composition.

2.3. Experimentally determined claudin structures

Currently, a comprehensive description of the Cldn-based TJ structure is still lacking due to the limited availability of experimentally determined structures of Cldn monomers and the complete absence of Cldn assemblies. X-ray diffraction (X-RD) crystallography is the leading technique adopted to determine the 3D structure of membrane proteins [93]. However, several challenges are encountered to obtain high-resolution diffracting crystals, such as protein overexpression [94,95], membrane extraction, purification and crystallization [93]. In particular, the latter step can take years and the resulting crystallized proteins are very fragile and sensitive to X-ray damage [96]. For this reason, in the last decade, the cryo-electron microscopy (EM) technology has progressively become more popular for the determination of protein structures [97]. This technique overcomes the issue of the synthesis of ordered and large crystals that limits the applicability of X-RD methods through the direct observation of the biological system at the EM, which is maintained in a frozen hydrated environment to preserve sample integrity from high vacuum conditions [98]. From the very first low resolution images (~15 Å, on average, in 2010 [99]), this technology is progressively improving at an impressive pace year after year, as a result of advances in both hardware and software components [100].

The obstacles and the limitations illustrated for these techniques justify why only recently crystal and cryo-EM based struc-

Table 1

| Claudin | Organism     | Method   | Resolution (Å) | PDB ID | Publication year | Notes                                      | Ref. |
|---------|--------------|----------|----------------|--------|------------------|--------------------------------------------|------|
| Cldn15  | Mus musculus | X-RD     | 2.4            | 4P79   | 2014             | Lack of residues 34−41 in the ECL1       | [52] |
| Cldn19  | Mus musculus | X-RD     | 3.7            | 3X29   | 2015             | Lack of residues 104−113 in the ICL and 70 − 74 and 56 − 59 in the ECL1 | [102] |
| Cldn4   | Homo Sapiens | X-RD     | 3.5            | 5B2G   | 2016             | Lack of residues 98−105 in the ICL        | [103] |
| Cldn4   | Homo Sapiens | X-RD     | 3.37           | 7KP4   | to be published  | Bound to cCPE                              | [104] |
| Cldn4   | Homo Sapiens | Cryo EM  | 5.0            | 7TDN   | to be published  | Bound to cCPE and sFab COP-3              | [105] |
| Cldn4   | Homo Sapiens | X-RD     | 6.9            | 6OV3   | 2019             | Bound to cCPE and sFab COP-2              | [106] |
| Cldn9   | Homo Sapiens | X-RD     | 3.25           | 6OV2   | 2019             | Bound to cCPE in open form and additional disulfide bond Cys25-Cys84 | [106] |
| Cldn3 WT| Mus musculus | X-RD     | 3.6            | 6AKE   | 2019             | Bound to cCPE in closed form and additional disulfide bond Cys25-Cys84 | [107] |
| Cldn3 P134A | Mus musculus | X-RD | 3.9            | 6AKF   | 2019             | Bound to cCPE TM3 bending                  | [107] |
| Cldn3 P134G | Mus musculus | X-RD | 4.3            | 6AKG   | 2019             | Bound to cCPE                             | [107] |

Fig. 3. The claudin-15 cis-linear interface. (A) Structure of the Cldn15 monomer. The negatively charged surface is shown in the region including Asp55, Asp64 and Glu46. The hydrogen bond between Pro149 and Lys155 in ECL2 is reported. (B) cis-linear arrangement of Cldn15 monomers stabilized by the interaction between the Met68 residue in each subunit and the crevice formed by Phe146, Phe147 and Leu158 in the ECL of the adjacent protein. Electrostatic potential surface was computed with the Adaptive Poisson-Boltzmann Solver software [108]. Isosurfaces are drawn with a red-white-blue color scale ranging from −5.0 (red) to +5.0 (blue) kT/e. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
tures of single Cldn monomers have become available. In 2014, the mouse Cldn15 (mCldn15) crystal structure was solved (PDB ID: 4P79 [52]), and to date it still remains the only Cldn structure in the isolated apo form. Other experimentally-solved Cldn arrangements have later been obtained, albeit capturing the Cldn monomers bound to the C-terminal fragment of the bacterial Clostridium perfringens enterotoxin (cCPE) [101]. These structures allow to grasp the fine details of Cldn folding at the molecular resolution. A list of the available 3D Cldn crystals is reported in Table 1.

In Cldn15, the length of the TM domains is consistent with the thickness of a lipid bilayer, except for the longer TM3, whose terminal region spans into the extracellular space. Notably, the ECL2 is one amino acid shorter than those of other classic Cldns due to the lack of the residue 152, according to Cldn15 numbering (Fig. 2B), which is in general poorly conserved among the different Cldn homologs [5]. In addition, the Cldn15 ECL2 is stabilized by a hydrogen bond (HB) between the Pro149 backbone carbonyl oxygen atom and the Lys155 sidechain amine group. Interestingly, this interaction is found in wild type (WT) Cldn15 and Cldn3, but it is lost in the other homologs and in mutants of Cldn3 [18]. The negatively charged Glu46, Asp55 and Asp64 residues are located between the β3 and β4 strands (Fig. 3A), providing a negatively charged “palm” determinant for the ion-selectivity of the Cldn15-based TJs [52].

The Cldn15 crystal lattice revealed a linear arrangement of monomers interacting via highly conserved hydrophobic residues. Accordingly, the ECH Met68 residue in one protomer protrudes from the ECL domain, fitting the hydrophobic pocket arranged by the residues Phe146, Phe147 and Leu158 of the adjacent subunit [52] (Fig. 3B). This interaction is defined as cis-linear arrangement [18], and freeze-fracture electron microscopy (FFEM) images showed that it is pivotal for the Cldn15 strand formation. Specifically, Cldn15 strand was abolished upon mutation of the Phe146 and Phe147 residues, while it was not affected by Met68 replacement [52].

When the structures of the different Cldns are compared, they all exhibit a similar folding (Fig. 4). In addition to the highly conserved disulfide bond in ECL1 [52], Cldn9 (and Cldn6, despite the crystal structure is not available), show a further disulfide bridge formed by the Cys25 and Cys84 sidechains, joining TM1 with TM2 [106] (Fig. 4D). Overall, the Cldn structures bound to the cCPE reveal a TM domain which fits the one in Cldn15, with only an important difference observed in Cldn3 and Cldn4. Indeed, in the Cldn3-WT crystal structure (PDB ID: 6AKE), the presence of Pro134 in the TM3 region, close to the extracellular space, induces

![Fig. 4. Experimental claudin structures.](image-url)
an inflection of the helix that moves the ECL domain closer to the outer leaflet of the membrane bilayer [107] (Fig. 4B). While a similar conformation is observed in Cldn4 (PDB ID: 5B2G), due to the conservation of the Pro134 residue, Cldn15 (PDB ID: 4P79) and Cldn19 (PDB ID: 3X29) display a straight TM3 region thanks to the presence of an Ala residue in the same position.

Conversely, the interactions between the bacterial enterotoxin and Cldn ECLs induce remarkable variations between the cCPE-bound Cldn4 crystal (PDB ID: 5B2G), because of the disruption of the intramolecular interactions occurring among Tyr67, Leu71 and Leu77 (Fig. 4C). Nevertheless, the ECH is supposed to exist in the physiologically folded conformation of all classic Cldns thanks to the conservation of Motif II in ECL1 [104]. In addition, notwithstanding the availability of Cldn15 structure in isolated form, the ECL1 loop connecting the /1 and the /2 strand (including residues from 34 to 41) is missing, while it is resolved in Cldn crystals bound to the cCPE. Thus, the structural differences observed for the cCPE-based structures with respect to the Cldn15 arrangement (PDB ID: 4P79) suggest that the Cldn structures bound to the toxin are not fully representative of the native conformation at the level of the ECL domains [102,103,107].

In conclusion, the currently available 3D crystal structures of Cldns are instrumental to assemble putative multimeric arrangements to reproduce the TJ features, as described in the following section. However, to date, the refinement and validation of these models require the application of computational techniques able to capture the dynamic rearrangement of the multimers during multiscale molecular dynamics (MD) simulations. The future breakthroughs in TJ investigation will be provided by the determination of high-resolution Cldn-based assemblies, which is currently lacking in the literature. Nevertheless, both single particle cryo-EM (PDB ID: 6MHQ, 6MHY [109]) and X-RD crystallography (PDB ID: 7JJP, 7JKC, 7JLW, 7JM, 7JMC, 7JMD, 7JNO, 7JN1 [110]) were recently used to solve the structures of connexin-46 and connexin-50 at the resolution of 3.5 Å and 1.9 Å, respectively. These proteins belong to the gap junction family and, similarly to the TJs, they also form intercellular channels involved in the complex junctional meshwork (Fig. 1B). For this reason, the publication of these structures at such a high resolution, with respect to the typical accuracy of the two techniques [100], brings hope for the determination of the closely related Cldn-based multimers within the next few years.

3.3. Tetrameric claudin model based on crystallographic claudin-15 structure

After the introduction of the Cldn15 crystal structure [52], the same authors proposed the first Cldn-based model of paracellular strands, based on a putative arrangement of multiple Cldn15 subunits [44]. The global configuration is showed in Fig. 5A. In this model, monomers in the same cell membrane are arranged in an antiparallel double-row characterized by the presence of two cis-interfaces. The first one, named face-to-face in Ref. [5], occurs between the ECL1 regions of opposing monomers, and is supported by Cys-scanning mutagenesis experiments and FFEM imaging [52,62,113] (Fig. 3B). The second interface is based on the cis-linear arrangement observed in the Cldn15 crystal lattice (Fig. 5C) in each of the two rows. The trans-interactions of Cldns belonging to the two opposing cells form additional head-to-head contacts between residues belonging to the ECL2 domains and to the variable regions (V1, V2). Remarkably, the model shows a sequence of paracellular pores with a diameter smaller than 10 Å, each formed by four Cldn15 monomers, as illustrated in Fig. 5D. Nevertheless, the original multi-pore architecture does not include the ECL1 residues from 34 to 42 and the ECL2 149 - 150 belonging to each subunit. After the publication of this model, often referred to as the Suzuki model, its general validity has been debated because of concerns regarding the missing residues and the steric clashes that would have destabilized the architecture after the insertion of the missing parts [5,59]. In addition, other inconsistencies were found concerning the arrangement of the cis-interactions that do not involve previously observed interactions among the TM helices belonging to different Cldns [114]. However, subsequent independent works based on MD and protein–protein docking simulations refined the structure of this Cldn15-based paracellular model, including the missing regions of the starting configuration [115–118]. The details of these works will be illustrated in Section 4 Computational insights into the features of specific claudins (4.1 Claudins-15).

3.2. Tetrameric claudin model from coarse-grained molecular dynamics simulations

In a different study, CG MD simulations [42–44] were used to investigate the self-assembly of Cldn5 monomers embedded in a lipid membrane during a 10μs-long trajectory [112] (Fig. 6A). From the analysis of the interacting domains among Cldn5 proteins, the authors identified five dimers, labelled from A to E and showed in Fig. 6B. Interestingly, the dimer A structure fitted the cis-linear configuration observed in the Cldn15 crystal arrangement, while dimer C was defined by a TM3-ECL2-TM4 interaction between adjacent Cldn5 monomers. However, in the subsequent work of the same authors [111], each of these dimers was identified as a non-pore forming structure, due to the occluded configuration resulting from the trans-association of the two copies of each system.

In contrast, separate docking simulations for dimer D and dimer B configurations revealed two different pore-like tetrameric arrangements, named Pore I and Pore II, respectively. Dimer D, generating the Pore I, was characterized by the same cis-face-to-face interface introduced in the Cldn15-based strand model [62] and previously illustrated in Fig. 5D. On the other hand, dimer B was characterized by a leucine zipper motif among the residues Leu83, Leu90, Leu124 and Leu131 belonging to the TM2 and TM3 helices and supported by the aromatic interactions between the opposing pairs of Trp138 and Phe127. This arrangement, illus-
Table 2
Phylogenetic, tissue-expression, functional, structural and pathological features of the most relevant claudins studied in the literature.

| Claudin | Phylogenetic classification | Tissue expression | Function | Intermolecular association | Pathologies | Predicted cis-interactions | Predicted trans-interactions |
|---------|----------------------------|------------------|----------|----------------------------|-------------|---------------------------|-----------------------------|
| 1       | classic                    | Skin, gall bladder, human ovarian epithelium, inner ear. [133–137] | barrier. [71] | homophilic, cis-heterophilic with Cldn3. [16,19,138] | Hepatitis C virus (HCV) co-receptor, ILVASC/NISCH syndrome. Psoriasis, ulcers, dermatitis. [133,139–144] | face-to-face and cis-linear interfaces. Closers proximity of respective TM1s rather than TM4s in dimeric association. [145] | ECL2 Phe,Tyr(Phe, x(9–10), Glu, Leu/Ile/Met/Phe motif in trans-interface. [5,33] |
| 2       | classic                    | Proximal tubules of the kidneys, liver, epididymis, rat brain. [146–148] | barrier. [75] | homophilic, cis-heterophilic with Cldn1 and Cldn3. [16,19] | Obstructive azoospermia, upregulation in Crohn’s disease. [149,150] | face-to-face and cis-linear interactions supported. [113–115,151] | ECL2 Phe,Tyr(Phe, x(9–10), Glu, Leu/Ile/Met/Phe motif in trans-interface. [5,33] |
| 3       | classic                    | Brain capillary endothelium, inner ear, human gallbladder, liver, intestinal epithelial cells, nephrons, mouse prostate. [135,137,152–155] | barrier. [72] | homophilic, cis-heterophilic with Cldn1, 2, 4, 5, 19 trans-heterophilic with Cldn1, 2, 5, [19,92,156] | Receptor for cCPE, BBB TJ disruption in glioblastoma and encephalomyelitis. [153,157] | face-to-face and cis-linear interactions. [59,145] | ECL2 Phe,Tyr(Phe, x(9–10), Glu, Leu/Ile/Met/Phe motif in trans-interface. [5,33] |
| 4       | classic                    | Collecting ducts, pancreas, stomach, colon. [157–164] | barrier. [72] | homophilic, cis- or trans- heterophilic with Cldn3. [156,169] | Crohn’s disease, DiGeorge/Velo cardio-facial syndrome, hyperplastic vessels in the inner ear. [137,178,179] | face-to-face and back-to-back interfaces. Stable TM3-TM4-mediated cis-interface was also predicted [18,111,119,177] | ECL2 Phe147, Tyr148, Tyr158, Glu159 residues. [33] |
| 5       | classic                    | Brain endothelial cells, the ovarian surface, colon epithelium, the pancreatic acinar cells, the autotroph TJ in the Schwann cells, in the chick retinal pigment epithelial cells during their embryonic development and in the alveolar epithelial cells. [72,136,168–173] | barrier. [72] | homophilic, cis- or trans- heterophilic with Cldn3. [156,169] | Crohn’s disease, DiGeorge/Velo cardio-facial syndrome, hyperplastic vessels in the inner ear. [137,178,179] | face-to-face and back-to-back interfaces. Stable TM3-TM4-mediated cis-interface was also predicted [18,111,119,177] | ECL2 Phe147, Tyr148, Tyr158, Glu159 residues. [33] |
| 8       | classic                    | Distal nephrons, thin ascending limbs, intestine, inner ear. [137,178,179] | barrier, down-regulation of Cldn2 expression. [76] | cis-heterophilic with Cldn4 but also Cldn3 and Cldn7. [81] | Crohn’s disease. [150] | Does not form TJ strand alone. [81,88] | ECL2 Phe,Tyr(Phe, x(9–10), Glu, Leu/Ile/Met/Phe motif in trans-interface. [5,33] |
| 9       | classic                    | Inner ear, liver, kidneys. [180,181] | barrier, maturation of the epithelms, barrier to K+ and Na+. [181] | homophilic. [182] | Mediates HCV entry into hepatoma cell lines, autosomal recessive non-syndromic hearing loss. [180,183] | ECH-ECL2-mediated cis-interaction. [106] | ECL2 Phe,Tyr(Phe, x(9–10), Glu, Leu/Ile/Met/Phe motif in trans-interface. [5,33] |
| 10a      | classic                    | Kidney. [66] | barrier, anion channel. [69] | homophilic, trans-heterophilic with Cldn1b. cis-heterophilic with Cldn19. [18,30,59,60] | HELIX syndrome, anhidrosis, alacrima, dry mouth, kidney failure. [184–186] | cis-interactions affected by TM1 and ECL1. [60] | ECL2 Phe,Tyr(Phe, x(9–10), Glu, Leu/Ile/Met/Phe motif in trans-interface. [5,33] |
| 10b      | classic                    | Brain, lung, salivary and mammary gland. [187,188] | barrier, anion channel. [65,66] | homophilic, trans-heterophilic with Cldn10a. cis-heterophilic with Cldn19. [18,30,59,60] | HELIX syndrome, anhidrosis, alacrima, dry mouth, kidney failure, hypermagnesemia, low urinary Ca2+ and Mg2+. [184–186,189,190] | face-to-face and cis-linear arrangement. [59] | ECL2 Phe,Tyr(Phe, x(9–10), Glu, Leu/Ile/Met/Phe motif in head-to-head trans-interface, supported by hydrophilic region of residues 56–67. [5,33,59] |
| 15       | classic                    | Duodenum, jejunum, ileum, colon. Kidney endothelial cells. [152,178] | barrier. [67] | homophilic. [18] | Malabsorption. Up-regulated in celiac disease. [191–194] | face-to-face interaction between [4-strands, cis-linear between ECH and ECL2. [52,62] | ECL2 Phe146, Phe147 Glu157, Leu158 involved in head-to-head interactions homodimer. [33,62] |
No trans-hetero-interaction with Cldn19 was detected.\[89\]

FHHNC syndrome, eyes heterophilic.\[5,33\]

Table 2 (continued)

| Claudin | Phylaorphic classification\[31\] | Tissue expression | Function | Intermolecular interactions Predicted trans-interactions Predicted cis-interactions |
|---------|--------------------------------|------------------|---------|-------------------------------------|--------------------------------|
| 16      | non-classic                   | kidney, taste receptor cells, inner ear. | reabsorption of divalent cations with Cldn17.\[77–79\] | homo/homophilic, cis-homophilic, cis-homophilic with Cldn15 and Cldn16.\[38,86\] |
| 17      | classic                      | kidney, taste receptor cells, inner ear. | water barrier, anion channel.\[70\] | homo/homophilic, homo/homophilic, homo/homophilic with Cldn15 and Cldn16.\[38,86\] |
| 19      | classic                      | kidney, taste receptor cells, inner ear. | barrier for monovalent ions, reabsorption of divalent cations with Cldn17.\[77–79\] | homo/homophilic, homophilic with Cldn16, 3, 12, 15, 18, Cldn11.\[30,89\] |

Effect of palmitoylation on claudin – claudin interactions

Post-translational addition of palmitic acid to Cys residues of membrane proteins is a typical modification adopted to favor the dynamic localization and delocalization within the cell membrane.\[122–124\] This process allows to anchor the protein to the lipid microdomains named membrane rafts\[123,125\], along with other regulatory tasks\[126–128\]. Various studies investigated the role of the palmitoylation in Cldn localization and strands formation\[122,129–131\]. Palmitoylation was demonstrated to be necessary for the proper delivery of Cldns from the endoplasmic reticulum to the plasma membrane and the highly conserved Cys residues in the TM domain could be targets of post-translational modifications\[129\]. Immunoblotting and fluorography experiments showed that Cldn2, Cldn4 and Cldn14 incorporate palmitic acid.

In a subsequent study of the same group, the investigation of Cldn self-assembly was extended to other classic Cldns (-1, -2, -4, -15, -19)\[115\]. In that work, CG MD simulations\[42–44\] showed different distributions of the dimeric A-E sequences for all the homologs. Specifically, the results suggested that the dimer-o-based Pore I configuration was ubiquitous among each protein, while the Pore I tetramer previously described for Cldn5 was also observed in Cldn4 self-assembly. The transferability of the cis-face-to-face interface among all the classic Cldns was also experimentally predicted, due to the conservation of the ECL2 Motif I and the ECH Motif II [53][Table 2]. Recently, the same authors introduced a novel method named Protein Association Energy Landscape (PANEL) to investigate the association of membrane proteins.\[120\] This algorithm maps the energy profile of the interacting interfaces occurring between two membrane proteins, using an extensive sampling of their rotational space. The resulting putative configurations are stochastically extracted from a uniform distribution and refined with short MD simulations. The PANEL method was adopted to study Cldn5 dimers\[120\] and the energy landscape suggested a relative stability of the symmetric dimers ordered from the most stable interface C, followed by the interface B and lastly the dimer D. This evidence was due to the different extension of the interacting interface, that involves both the TM and the ECL domains in the dimer C, while it is limited exclusively to the TM and the ECL domains for dimers B and D, respectively. Remarkably, this dualism between pore-forming (dimer B, D) and non-pore-forming (dimer C, A) configurations agrees with the bimodal “open” and “close” states postulated for the original TJ dynamic model\[9\].

Although the two aforementioned pore-like arrangements exhibit relevant structural differences, several experimental [59,80,119] and computational\[112,115,120,121\] studies suggested that Cldns could form multiple binding configurations and, therefore, a possible coexistence of the two architectures cannot be excluded.
Fig. 5. Multimeric claudin-15 arrangement introduced in Ref. [62]. (A) Paracellular space with a sequence of identical pore-like structures, viewed from the apical side. (B) cis face-to-face interface. (C) cis-linear configuration. The key interaction between Met68 of a subunit and the Phe146, Phe147 and Leu15 residues belonging to the adjacent monomer is highlighted. (D) Single-pore model. Missing residues are indicated.

Fig. 6. Self-assembly claudin-5 coarse-grained molecular dynamics simulations performed in Ref. [112]. (A) Snapshots of Cldn5 (red) cis self-assembly process in a lipid bilayer (blue) at different time frames. (B) Frontal and apical view of the five putative Cldn5 dimers isolated from the CG MD simulations. Reprinted with permission from Ref. [112]. Copyright 2016, American Chemical Society. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
and, specifically, the latter undergoes to S-acylation in Cys sites in both the TM2 and TM4 helices. Nevertheless, the Cldn14 ability to form strands is not apparently influenced by the presence of palmitoylation binding sites. In contrast, mutagenesis experiments showed that Cldn7 palmitoylation fosters membrane recruitment, but hinders the TJ formation [131]. In another study, mass spectroscopy analysis suggested that palmitoylation of multiple putative sites in Cldn3 is a non-specific, stochastic and weakly cooperative process, with low palmitoylation probability for the Cys residues located in TM3 [122]. Additional proteomic analysis of platelet-enriched plasma showed that also Cldn5 is a target of palmitoylation [132]. Indeed, Cldn5 exhibits four Cys residues that could be targeted by thio-esterification: Cys104/Cys107 on the ICL and Cys182/Cys183 in the TM domain. To further investigate from a computational point of view the role of palmitoylation in Cldn5, the authors of Ref. [130] studied the effect of the post-translational modification on Cldn5 dimerization, replicating the cis self-assembly simulations on the WT and palmitoylated (Cldn5P) variants of the protein. CG simulations were performed with different membrane compositions mimicking the lipid environment of the organelles of the secretory pathway. Results revealed that the oligomerization rate of Cldn5 decreases with the increasing heterogeneity of the lipid membrane, suggesting a predominance of the Cldn – lipid contacts over the Cldn – Cldn interactions in both Cldn5 and Cldn5P. Notably, the palmitoylation affected the dimeric arrangement distribution, favoring the symmetrical configuration of dimer B (Fig. 6) over the other cis-interactions occurring at the level of the TM region (dimer C, Fig. 6), particularly in low-complexity membranes. Overall, all these studies reveal variable effects of the addition of the lipid chain in the TM region on the process of Cldn-Cldn dimerization, suggesting that different homologs can arrange into cis- and trans-complexes with multiple binding configurations.

4. Computational insights into the features of specific claudins

After the introduction of the Suzuki model [62] as putative linear arrangement of Cldn15 proteins, several computational studies investigating the features of different Cldn multimers at the atomistic detail were published. Nevertheless, the in-silico structural modelling of Cldn assemblies is almost totally limited to the homoophilic representation of the cis- and trans- interactions, with the only exception for the heterodimerization occurring between Cldn16 and Cldn19 [205]. In that study, results based on mutagenesis and electrophysiology were used to investigate an alternative putative heterophilic Cldn16/Cldn19 multimer to generate paracellular channels for divalent ions. Any further investigation on the structural stability or predicted physiology of this putative TJ model was provided with MD simulations. Here, we aim at sum-

![Fig. 7. The back-to-back interface and the Pore II configuration. (A) Dimer B structure from frontal and apical views. The leucine zipper between TM2 and TM3 is shown. (B) Pore II configuration resulting from the trans-interactions of two dimers B.](image-url)
marizing the methodologies and the results for the homophilic Cldn assemblies, including the major outcomes for the homologs Cldn15, Cldn2, Cldn4 and Cldn5, which are reported in the following sub-sections.

4.1. Claudin-15

Highly expressed in the intestine [178], Cldn15 forms channels for cations, while it generates barriers to the passage of anions [67,84]. The isolated configuration of the Cldn15 structure was solved at a 2.4 Å resolution (PDB ID: 4P79 [52]), and adopted to build the aforementioned TJ model introduced in Ref. [62] and named the Suzuki model. Due to the lack of various ECL1 (34 to 42) and ECL2 (149 and 150) residues in the crystal structure, an initial modelling of the missing parts was required to generate the head-to-head interactions and to assemble the Suzuki model [62].

To this aim, different and independent works focused on the refinement of this putative Cldn15 paracellular assembly and shaped the missing residues with diverse strategies [116–118]. In addition, the Cldn15 multimeric models were inspected at several dimensional scales, ranging from the tetrameric single-pore architecture [117] to complex filaments [58,116,118] composed by up to 300 monomers [206].

4.1.1. Single and double pore formation and ion flux

In the work of Ref. [116], we modelled the missing Cldn15 segments using the RCD + loop closure tool [207,208], which is based on an ab-initio algorithm. Multiple loop configurations have been tested to achieve a non-overlapping arrangement among the various subunits and, starting from the Cldn15 monomer, a single-pore tetramer and the a double-pore octamer were assembled, respectively (Fig. 8A,B). Further MD simulations [37,209] of the tetrameric model revealed a stable configuration along the 250-ns-long trajectory and a maximal constriction localized at the center of the channel, with a radius of ~ 2.5 Å. A representative profile of the pore cavity is reported in Fig. 9. The four negatively charged Asp55 residues, each belonging to a Cldn15 subunit, are located in the inner segment of the cavity, defining a cage that should promote the cation permeation [84,210]. The double-pore simulation reported a convergence of the minimal pore radius of the two cavities towards the value of ~ 3 Å, in agreement with the single-pore representation. The hydrophobic interactions that define the cis-linear interface and the trans-association, involving both ECL1 and ECL2 residues, were stable during a short, 35 ns-long, simulation. To further study the structural features of the refined tetrameric model, in a subsequent work, we performed thermodynamic and kinetic calculations to investigate the properties of ion transport through the pore [117].

Umbrella sampling (US) simulations [211] provided an estimation of the free energy (FE, or potential of mean force, PMF) profile of the ion permeation events, and Markovian Voronoi-tessellated milestoning method [212,213] computed the mean first passage time (MFPT) for each ion. FE profiles for the passage of monovalent ions were found to be symmetric with respect to the central constriction of the cavity along the pore axis, consistently with the structure of the model (Fig. 8C). Specifically, Na⁺ and K⁺ permeations were favored by

![Fig. 8. Claudin-15 single-pore and double-pore models.](image-url)
energetic minima of ~ 4 kcal/mol, while the Cl\textsuperscript{-} passage was hindered by a barrier of ~ 8 kcal/mol located in the same region of the pore. Remarkably, the values of the Na\textsuperscript{+} and K\textsuperscript{+} FE minima and of the Cl\textsuperscript{-} maximum profiles were found at the cavity center, in correspondence to the position of the Asp55 residues, while the Asp64 sidechains point out of the pore lumen and they are supposed to be partially shielded by the Trp63 residues [214]. As a further confirmation of the pivotal role of the Asp55 residues in the ion selectivity, the number of coordinating water molecules of the single permeating Na\textsuperscript{+} ion was calculated during the passage through the pore. Dehydration events were observed in the cavity center, where, on average, one water molecule is replaced by an oxygen atom of the Asp55 sidechains. Results from the kinetic calculations indicated that cation permeation occurred in the microsecond timescale, while the anion passage was characterized by two additional orders of magnitude, as reported in Fig. 8D.

An alternative strategy to study the ionic permeation through Cldn15 pores has been applied by another group in Ref. [118]. In this work, a multi-pore configuration was assembled starting from 12 Cldn15 monomers and leveraging the periodicity of the simulation box. Missing loops were modelled using MODELLER [215, 216] and relaxed by turning off the intermolecular interactions occurring within the modelled sequences during energy minimization. Then, the simulation box was edged at the extremities of the dodecameric system along the membrane plane, and replicated via periodic boundary conditions to generate a continuous and uniform strand of Cldns. A 250-ns-long trajectory revealed a stable configuration of the paracellular scaffold, where the hydrophobic head-to-head interfaces forced the water molecules to flux within the channel vestibule [217].

To assess the ion selectivity of the model, the authors simulated Na\textsuperscript{+} and Cl\textsuperscript{-} currents through the pores in response to the application of various electric fields [218, 219]. In addition, the size selectivity threshold was computed from the currents of a group of congeners positively-charged organic molecules of increasing size, such as methylammonium (MA\textsuperscript{+}), ethylammonium (EA\textsuperscript{+}), tetramethylammonium (TMA\textsuperscript{+}) and tetraethylammonium (TEA\textsuperscript{+}). By applying increasing external potentials [219] and measuring the ion current, a linear relationship typical of passive paracellular channels was found [220–223]. A conductance of 121 pS and 29 pS was detected for Na\textsuperscript{+} and Cl\textsuperscript{-}, respectively, corresponding to a relative selectivity \( P_{Na^+} / P_{Cl^-} \sim 4.2 \) and correlating with the TEER values measured on MDCK I monolayers expressing Cldn15 [118]. The conductance, calculated on the alkylammonium cations, moderately decreased with the increase of the size of the molecule, but revealed a significantly lower value for the passage of TEA\textsuperscript{+}, suggesting a radius cutoff for cation translocation between 3.0 and 3.5 Å. The major role was ascribed to the interaction between the cations with the Asp55 residues belonging to the various subunits. In particular, it was observed that ~ 35 \% of the entire contact time of Na\textsuperscript{+} with the protein during the passage through the cavity was associated with these pore-lining residues. A smaller effect was found for the other negatively charged pore-lining residues Glu46 and Asp64. To further confirm this evidence, ion current simulations were performed on mutated versions of the Cldn15 model. In particular, the replacement of Asp55, Glu46 or Asp64 with neutral amino acids abolished the cation selectivity, while the substitution with a Lys residue induced a reversal of the charge selectivity towards anions.

In addition to the aforementioned strategies, the Cldn15 pore-like assembly was also investigated with the PANEL method [120, 121]. The position of the charged-pore lining residues was analyzed to predict the determinants of the cation selectivity, suggesting that Asp55 and Asp148 could affect the passage of charged particles. However, while the major role of Asp55 in favoring the permeation of cations is ascertained, the influence of the second amino acid was not discussed in detail.

4.1.2. Cis- and trans- interactions of multimeric claudin-15 models

Despite the different methodologies involved in these two works, both revealed that MD simulations can provide refined and stable configurations of both single- and double-pore arrangements obtained from the Suzuki model [116–118]. However, the linear shape of the Cldn15 paracellular structure proposed in the forming structure and in further MD simulations is not fully representative of the Cldn-based TJ arrangements. Indeed, both live-imaging confocal spectroscopy of Rat1 cells expressing mCldn2-GFP and FFEM of mCldn15 strands expressed in HEK293T cells showed lateral movement and remodeling of the intermolecular architecture [58], mainly via arching and bending deformation [53, 224], subjecting the strain to compression and tensile forces. In this context, Zhao et al. [58] performed standard MD simulations of two multimeric structures arranged with 8 and 16 protomers of Cldn15 based on the Suzuki model. Results showed a stable evolution of the octamer, consistent with the results illustrated in Refs. [116, 118], while the longest assembly revealed a less stable configuration within the first 40 ns, because of the rotation of the single proteins along the axis perpendicular to the membrane plane. The decrease of the \( \beta \)-sheet structure after 100 ns of MD simulation suggested that the cis-linear interface could be unstable for extended strands, and a redefinition of the interactions involving Met68, Phe146 and Phe147 should have been considered. To define a new model for cis-configurations, coherent with experimental outcomes, the authors performed blind Cldn15-Cldn15 molecular docking simulations with HADDOCK [225], generating three putative low-energy cis-interfaces, named cis-1, cis-2 and cis-3, all reported in Fig. 10. Specifically, the first model exhibited a relative rotation of the two monomers along the axis perpendicular to the membrane plane of ~ 17\(^\circ\) with respect to the cis-linear arrangement, while larger angles were found for cis-2 and cis-3. The analysis of the residue contacts between the ECH and the ECL2 revealed that the two configurations shared the interaction between Ser67 and Glu157. Specifically, site-directed mutagenesis suggested that the Cldn15 mutants Ser67Ala and Glu157Ala could form discontinuous strands, with a significant decrease in the capability to form the TJ backbone for the second variant. However, the FFEM images of the strand generated by the double mutant Ser67Glu/Glu157Ser showed a morphology similar to that of WT Cldn15, confirming the
presence of their mutual interaction in the cis-interface. Moreover, the cis-1 configuration was stabilized by other redundant interactions in which Arg79 had a central role, favoring the strand proliferation. Several single-point mutations of Arg79 with either bulky (Trp, His) or negatively charged (Glu) residues prevented the Cldn15 strand formation. In addition, cis-2 and cis-3 arrangements were supposed to favor the strand branching, despite this occurrence was not demonstrated. Overall, these data suggest that the Cldn15 cis-multimeric network is mainly stabilized by the interactions between Ser67, Glu157 and Arg79, with the latter residue also involved in additional contacts that stabilize the multimeric architecture.

The role of the Cldn15 strand flexibility was also investigated in a more recent study [206], where the previously refined paracellular pore structure of Ref. [118] was used to build extended multimeric arrangements ranging from 36 to 300 monomers. MD simulations were carried out adopting a hybrid-resolution model, named PACE, where protein atoms are represented at the atomic scale, while lipids and solvent are described at a CG resolution [226–228]. The 36-monomer long system was studied with both all-atom and hybrid method, revealing a comparable stability and preservation of the secondary structure domains. Due to the longer timescale explored, the hybrid model exhibited a dynamic curvature in the multimeric configuration, that was further analyzed in three systems arranged with 84, 180 and 300 monomers. In order to compute the mechanical properties of the strand, the persistence length ($l_p$) and the bending modulus were calculated at 300 K. In this system, $l_p$ described the length over which segments of the Cldn filament assumed uncorrelated directions. The parameter was calculated from the equilibrated structures of four strands of difference lengths and the resulting $l_p = 174 \pm 42$ nm was in agreement with the experimental value extracted from FFEM images [58]. This analysis of the bending mechanism revealed that cis-interactions are highly flexible and able to adapt to a wide range of curvatures. The relative angles of deformation between adjacent monomers ranged from $-70^\circ$ to $+70^\circ$ and were collected within a gaussian distribution centered at 0°, which corresponded to the cis-linear interface [52]. Detailed investigation of the interactions supported the curved cis-configurations of adjacent Cldn15 monomers and suggested that the lateral shift of adjacent monomers occurred adopting the ECH as a pivot. In particular, the ECH domain preserved the intermolecular contact between flanking proteins, alternating interactions with TM3, ECL1 and ECL2. These results identified three dominating patterns. The first one is defined by the ECH Met68 which interacts with the TM3 Phe146 and Phe147 of the neighboring monomer, as originally described for the cis-linear arrangement [52,62]. The second interaction was associated with the contact between Glu157 and Ser67, which rearranged and rotated the proteins around a highly preserved interface between the ECH and the ECL2 domains. Remarkably, the relevance of this interaction for the integrity of the paracellular pore and the strand formation was also investigated in previous works [58,116,118]. Lastly, the third region included the V1 domain, forming hydrophobic seals [31,62,222]. These simulations revealed that the movement of filaments on opposing membranes are correlated, preserving the head-to-head interactions and maintaining the tetrameric structures of the Cldn pores. Additional investigations on the strand stability were performed inspecting the effects of the mutation Ala134Pro on the rearrangement of the Cldn15 assembly [229]. Indeed, replacement of the Ala134 residue belonging to the TM3 with a Pro in Cldn3 (PDB ID: 6AKE) native structure is responsible for the TM3 tilted configuration, bending the entire ECL domain of approximately 8° towards the membrane and conferring rigidity to the monomer [107]. In that study, strands of both WT Cldn15 and Ala134Pro Cldn15 were simulated with the procedure of the previous work [206], and results showed that the mutation significantly limits the lateral flexibility of the strands, hindering the dynamic rearrangement of the cis-interactions based on the pivotal role of ECH. Therefore, the bent conformation assumed by the Ala134Pro mutation did not induce neither the formation nor the abolishment of cis-interactions and did not modify the entity of the trans-contacts, revealing the indirect role of the TM domain in the modulation of the dynamic properties of the Cldn strand [229].

4.2. Claudin-2

Cldn2 was the first member of the Cldn family to be discovered together with the homolog Cldn1 [12] and is the most characterized member forming an ion-selective channel [12,14,16,30,64,83,114,147,222,230–232]. Cldn2 is highly expressed in proximal tubule of kidney nephrons [147,152], where it favors the paracellular permeation of Na⁺ and small positively-charged particles [64,83,84].

4.2.1. Ultrastructural claudin-2-based tight junction model

Cldn2 has been thoroughly investigated from an experimental point of view. In one of the first works focused on Cldn2 physiology at the level of single channel [231], the authors measured a conductance of Cldn2-based TJs of ~ 90 pS using patch-clamp recordings on MDCK I and Caco-2 cell monolayers expressing this protein. In addition, kinetic analysis suggested a bimodal configuration of the paracellular pores, which could exist in a single open plus two closed states. Based on these observations, the same group
proposed a circuit model as the first in-silico ultrastructural TJ representation [233], reproducing the ion-selectivity of the global Cldn2-based strand. In that work, the paracellular environment was modelled as an array of three resistive strands replicating the arrangement observed in FFEM images [8,9] (Fig. 11). Each strand was defined with a constant conductance baseline and a variable number of resistors, representing Cldn2 channels, arranged in parallel and able to switch between open and closed states, with a conductance of 222 ps and 10 ps, respectively. The transition between states occurred with frequencies properly chosen to reproduce the dwell times recorded in-vitro, hence the time spent by the Cldn2 channel in each of the possible states. Results revealed that the gating process of Cldn2 pores fitted the experimental data only for branching strands. Despite the lack of molecular details, this study represents the first model of the functional ultrastructure of a Cldn-based strand. In addition, this bimodal representation of one “open” and two “closed” states found support in the recent CG simulation on classic-Cldns self-assembly processes, suggesting the compatibility of one pore-forming (dimer D) and two non-pore-forming (dimer A, C) cis-configurations for Cldn2 [115].

4.2.2. Pore formation and ion flux

In the work of Ref. [115], the tetrameric Cldn2 pore was assembled by joining two dimer D complexes via trans-interactions occurring between the respective ECL domains. In this structure, the Gln63 residue is located in the central part of the cavity, as shown in Fig. 12A. Close to this polar amino acid, the Asp65 residues belonging to the four subunits are also found. The average pore diameter during a 300-ns-long trajectory was 7.4 Å, with values of 5.4 Å and 10.4 Å at the narrowest and widest points, respectively (Fig. 12B). This occurrence was in agreement with experimental data [222,234] and with the model proposed for Cldn2 [115].

During MD simulations, water molecules filled the paracellular channel and the passages of 13 Na+ and only 1 Cl− were observed over the explored timescale. In a subsequent work [235], the same authors investigated the molecular mechanism of ion permeation by performing FE calculations based on replica exchange metadynamics [236–238]. The FE landscapes of permeating water, Na+, K+, Cl−, Mg2+ and Ca2+ were computed, by adopting the pore axis as a collective variable (CV). Moreover, to demonstrate the major role of Asp65 in charge selectivity, both WT Cldn2 and its Asp65Ala variant (Cldn2M) were studied. Results revealed that water could freely pass through the pore in both systems. Moreover, the permeation of monovalent cations showed low barriers in the external segments of the pore axis and an energy minimum at the center of the Cldn2-WT pore, correlating with the positions of the Asp65 sidechains. The mutation switched the FE minima into barriers for Na+ and K+ also in the inner region. As expected, the presence of four Asp65 residues generated a high barrier for Cl−, which was only in part damped by the mutation. The FE profile of divalent cations showed maxima located at the entrances of the cavity, while minima correlated with the favorable interactions with the Asp65 sidechains. In addition, the Pore 1 configuration with the charged residues belonging to the ECLs was also studied with the aforementioned PANEL method [121], confirming that the position of the pore-lining residues Asp65, Asp146 and Lys157 should affect the ion selectivity. However, while Asp65 was demonstrated to determine the permeation of cations [57,210,222,235], the role of Asp146, located at the pore entrances, was not further discussed.

4.2.3. cis- and trans- interactions in the claudin-2 multimeric model

In a recent work [239], the authors investigated the role of the Cldn2 mutation Gly161Arg in the onset of obstructive azoospermia, a disease causing the obstruction of the male genital tract. Gly161 is located at the end of ECL2, close to the aforementioned Motif I, and the substitution with Arg was found to exert a dominant negative effect on TJ formation, with the consequent disruption of the blood-epididymis barrier. To inspect the effect of this mutation on the multimeric arrangement, the authors performed docking analysis of both cis- and the trans- interactions responsible for the arrangement into the Suzuki model [62]. Multiple configurations were generated and ranked with HADDOCK [225], producing a tetramer with lower stability and a wider pore in the mutated form, suggesting a loss in ion selectivity and structural integrity. In particular, the insertion of the Arg sidechain was supposed to hinder the correct trans-association deputed to the ECL2, hence generating brittle strands. In this work, the dimeric arrangements were generated based on an ECL-mediated cis-organization despite the initial suggestion that the formation of stable homodimers is mediated by the TM2 domain, given the position of Cys104/Cys108 revealed by Cys-scanning mutagenesis and PAGE experiments [114]. Subsequent studies, using CG MD simulations [42,43], revealed that the main TM2-mediated association represents a minor distribution in the population of the Cldn2 dimers [115]. The same authors observed that the ECL-mediated cis-conformation is characterized by the coupling of Cys104 and
Cys108 belonging to neighboring protomers, at a distance of \( \frac{10}{24} \) Å. This is consistent with the positions mapped experimentally and it fits the results provided by PAGE with the face-to-face cis-configuration \([114,232]\). Thus, the tetrameric Pore I configuration is still the best representative model of the homophilic aggregation of Cldn2 monomers.

4.3. Claudin-4

Cldn4 is an important component of renal tubules, but is expressed also in other organs \([157–164]\). The details of the Cldn4 features are debated due to divergent in-vitro-based results, which are significantly influenced by the type of cells in which the proteins are expressed \([18,30,63,84]\). A barrier function is supposed when Cldn4 is expressed in some tissues \([18,30,31,80]\), such as the urinary bladder epithelium. On the contrary, other studies suggest that Cldn4 is responsible for Cl⁻ reabsorption in the kidney collecting ducts \([81–85]\) due to the presence of the Lys65 residue in ECL1 \([57,81,240]\) and a regulatory role on cation exchange can also occur via interclaudin interference \([87]\). In this context, computational investigations of Cldn4 multimeric models could help to understand the molecular determinants of its physiological function.

4.3.1. Pore formation and ion flux

Modelling based on the PANEL method \([121]\) proposed the formation of two pore-like structures, that recapitulate the architecture of Pore I (Fig. 5D) and Pore II (Fig. 7B). Both models suggested that the ion-selectivity is driven by the Lys65 and Arg158 residues, thus predicting an anion-selective function for both configurations. However, these assumptions were not supported by explicit MD simulations. Recently, we investigated these two putative Cldn4 configurations, as shown in Fig. 13, to analyze and compare their ion selectivity \([241]\). Umbrella sampling (US) simulations \([211]\) demonstrated that the FE of water and ions permeation through the two pores (Fig. 13C,D) was significantly
affected by the different pore-lining residue orientation. In particular, Pore I showed barriers of \(~\sim\)7–8 kcal/mol to the passage of monovalent and divalent cations, respectively, with maxima at the cavity center correlating with the positions of the Lys65 sidechains. On the contrary, Cl\(^-\) permeation was characterized by an energetic minimum of \(~\sim\)2.5 kcal/mol in the center of the model (Fig. 13C), due to the stabilizing interactions with the positively charged pore-lining residues. Notably, the same analysis for Pore II suggested a different outcome with FE barriers for all the ions. Specifically, each cation revealed two peaks symmetrically located near the pore entrances and an energetic minimum in the central segment correlating with the position of the Asp76 and Asp146 sidechains. In the same position, the passage of the anion was hampered by a FE barrier (Fig. 13D). In addition, a partial dehydration of the Cl\(^-\) ion was observed during the passage through the pore, as a consequence of the interactions with the positively charged residues. Globally, FE calculations revealed opposite behaviors to the passage of Cl\(^-\) between the two pore models, suggesting that they are mutually exclusive, at least in the case of the Cldn4 protein. This difference was mainly due to the reverse orientation of the Asp76 and Asp146 residues belonging to the four Cldn4 subunits (Fig. 13A,B).

4.3.2. Cis- and trans- interactions in claudin-4 multimeric models

In our work, the two pore models were assembled with different strategies [241]. While Pore I was arranged relying on the template available for the Cldn15 [52], the assembly of the Pore II configuration required the application of a detailed docking protocol performed with the MEMDOCK [242] and Rosetta DOCKING2 [243–245] platforms. Remarkably, in the Cldn 4 dimer B cis-interface, the leucine zipper motif described for Cldn5 in Ref. [111] (Fig. 7A) resulted from the arrangement of Leu93, Leu130, Leu93 and Leu127. In contrast, a hydrophilic cis-interface characterized the dimer D configuration, involving Glu61, Gin63 and Lys65 belonging to the ECL1 \(/\beta\)-strands of the two opposing monomers. Similar hydrophilic trans-interactions between the ECH Leu70, Leu71, Leu73 and the ECL2 Pro150, Leu151, Val152 of opposing Cldn4 proteins were observed in both tetramers. In addition, further contacts between the ECL1 \(/\beta\)-1, \(/\beta\)-2 and \(/\beta\)-3, \(/\beta\)-4 domains belonging to opposite Cldn4 dimers were found in Pore I, supporting the formation of the paracellular \(/\beta\)-barrel architecture.

4.4. Claudin-5

Cldn5 is the major component of the TJ belonging to the blood–brain barrier (BBB) [24], but it is also found in other tissues such as human colon [168], blood-testis barrier [170], pancreatic acinar cells [155], Schwann cells [173], chick retinal pigmented cells during development [171] and alveolar epithelial cells [172]. Cldn5 is typically considered a barrier-forming protein in each tissue where is expressed [168], with very few exception as observed in the human cystic fibrosis airway epithelia, where Cldn5 was associated to an increase in the paracellular permeability, probably as a consequence of changes in the heterophilic interactions with other Cldns [156].

4.4.1. Strategies for the enhanced permeabilization of the blood–brain barrier

The major interest for Cldn5 consists in its role in the BBB, where Cldn5 strands seal the paracellular space to preserve the chemical homeostasis of the central nervous system (CNS) and to guarantee its impermeability to ions and molecules with mass greater than 800 Da [51,247–250]. This regulation of the diffusion process through the paracellular space contributes to hamper the delivery of drugs to the brain for the treatment of various neurological disorders [24,249–251]. Therefore, novel strategies are necessary to allow a safe and transient weakening of the paracellular interactions and to improve drug delivery to the brain. In this context, a detailed description of the structural features of Cldn5-based assemblies is of major interest [35]. Because of the lack of experimental structures of Cldn5-based complexes, this investigation could be significantly promoted by the use of various computational methods [252–254]. A possible strategy to enhance the permeability of the BBB via reversible and localized disruption of the TJ architecture adopts physical stimuli [24], such as focused ultrasounds [255–258]. The effects of these waves were originally observed in Cldn15-based systems with CG MD simulations [44,259–262] and were extended to the BBB Cldn5 paracellular complexes in Ref. [263]. In this last work, the instantaneous pressure variation induced by ultrasound waves was modelled as a periodic function defined by amplitude and frequency of the perturbation and included in the barostat regulating the pressure of the system. Results revealed that low intensity ultrasounds induced negligible modifications in Cldn assembly. By increasing ultrasound intensity, openings of the paracellular spaces were observed, as predicted by MD simulations. Another approach to modulate TJ interactions consists in targeting Cldn5 with binding molecules [22–24,26,264], and the proper design of these ligands could be significantly boosted by the adoption of computer-aided techniques, starting from the description of putative paracellular complexes [26].

4.4.2. Pore formation and ion flux

In previous studies, cis-dimers of Cldn5 proteins were observed in self-assembly CG simulations [112] and used to generate two paracellular pores, named Pore I and Pore II, respectively [111]. In a subsequent work [235], the same authors studied the thermodynamic features of water and ion permeation through Pore I using metadynamics [236–238]. The radius profile of the starting configuration is characterized by a symmetrical shape with respect to the cavity center, consistently with the Pore I architectures of other works [116,265], with a minimal value of \(~\sim\)3.3 Å located at the center of the cavity, where the sidechains of the four Cldn5 form an uncharged cage. To determine charge and size selectivity of this configuration, MD simulations were performed on both the WT and the Cln63Asn variants. The replacement of Gln63 with Asn63 did not alter the polarity of the inner region of the cavity, but the lack of the C\(_{\gamma}\) atom increased the pore radius up to \(~\sim\)4.0 Å. FE calculations suggested that, while water passage was allowed, barriers were associated with ionic permeation for both systems. In particular, symmetrical barriers with maxima at the cavity center were observed, correlating with the position of the positively charged pore-lining Lys65 residues. Conversely, the study of the Pore II configuration was limited to the calculation of the FE associated with the paracellular permeation of water and \(/\alpha\)-D glucose [111]. Because of the lack of a comparison in terms of ion translocation between the two pores, we computed the FE profiles of permeating ions and water molecules through both these Cldn5 complexes [265]. In analogy with the procedure described for Cldn4 [241], Cldn5 was homology modelled from the Cldn15 structure (PDB ID: 4P79 [52]) and FE calculations were performed using the US method [211]. Results were consistent with those reported in Ref. [235]. Indeed, energetic landscapes of the cation passage through the Pore I cavity revealed symmetrical barriers with maxima at the center, and barrier heights proportional to the ion charge. The passage of Cl\(^-\) was similarly hindered by two barriers of \(~\sim\)2 kcal/mol at the cavity entrances, because of the unfavorable interactions between the anion and Glu146/Asp149 (Fig. 14C). The same analysis for the Pore II complex revealed different profiles, but still compatible with the barrier function of Cldn5 (Fig. 14D). A central barrier of \(~\sim\)5 kcal/mol was shown for Cl\(^-\), with two peaks corresponding Glu146/Asp149. FE walls.
of ~2 kcal/mol and ~4 kcal/mol were found at the entrances of the cavity for monovalent and divalent cations, respectively, consistently with the presence of the Lys65/Lys48 sidechains. Energetic minima were observed at the cavity center, where the interaction between the cation and the four Asp149 residues overcome the unfavorable contact with the Arg145 residues. In this context, the presence of energy barriers for ion permeation in the two models and the heterogeneity of the strand suggests the possible coexistence of these pores in Cldn5-based TJ strands.

4.4.3. Cis- and trans-interactions of the claudin-5 multimeric models

The two Cldn5 pore models [265] were assembled with the same strategy described previously for Cldn4 [241]. A dense hydrophilic interface formed by Gln57, His61, Gln63 and Lys65 is observed in Pore I, while the leucine zipper motif stabilizes the cis-association in Pore II. Similar trans-interactions between the ECH and the ECL2 fragments were found in both configurations, but additional contacts between the ECL1 $\beta_{1-2}$ and $\beta_{3-4}$ loops belonging to opposing Cldn5 dimers provided a stable conformation of the $\beta$-barrel structure along the entire 1-µs-long MD simulation. In addition, a conserved intramolecular HB between the Lys157 sidechain and the Asp149 backbone stabilized the ECL2 conformation in the Pore I model, as previously described for Cldn15 [5,52]. On the opposite, in the very recent work of Ref. [177], the PANEL algorithm was adopted to generate putative cis-configurations of Cldn5 dimers. Their results revealed a stable arrangement involving the interactions of TM3 and TM4, similar to the previously introduced dimer C configuration [112]. These dimeric intracellular configurations were used to reconstruct the putative TJ scaffold, by adding facing dimers through the respective ECL domains, as refined using the VINA local docking tool implemented in YASARA [266,267]. The resulting multimeric assembly revealed a totally sealing interface spanning the paracellular space, in which ECL regions of opposing filaments did not leave space for pore formation. Despite this configuration clearly accounts for the barrier function of Cldn5-based BBB TJs [51,246–249], it also

Fig. 14. Claudin-5 tetrameric models and free energy profiles. Lateral views of Pore I (A) and Pore II (B). Monomers are indicated with different coloring. Polar residues of two subunits are shown and colored according to their charge: blue for basic residues, orange for acidic residues and green for neutral residues. FE profiles of ions and water translocation through Pore I (C) and Pore II (D). FE landscapes were reproduced with the Matplotlib library and colored with the 'PRGn' colormap, ranging from −10.0 (purple) to +10.0 (green) kcal/mol. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
implies a complete impermeability to water, at variance with the Cldn5 pore-like tetrameric architectures of Ref. [265].

**4.5. Comparison of the claudin-4 and claudin-5 transport properties**

In the context of the structural comparison among different Cldn members, our FE calculations for Cldn5 [265] and Cldn4 [241] revealed new insights into the regulation of ion transport through Cldn-based paracellular models. Both these homologs generate barriers to the passage of each cation independently of the adopted model. While Cldn5 provided energetic walls to the Cl− permeation in both arrangements (Fig. 14C,D), Cldn4 was permissive for anions in Pore I, but not in Pore II (Fig. 13C,D). The difference between the two homologs is attributable to variations in the charged and polar pore-lining residues belonging to the ECL domains (Fig. 2). Specifically, Arg31, Glu48, Arg158 and Asn149 in Cldn4 are replaced by Gln31, Lys48, Tyr158 and Asp149 in Cldn5 (Fig. 15), differentiating the energetic landscapes of Cl− translocation only in the Cldn4/5 Pore I models.

The effect of these variations are also reflected by the electrostatic potential surface shown in Fig. 16, where an overall neutral environment is observed at the entrances of the Cldn4 Pore I model (Fig. 16A), while an acidic surface is found in the corresponding region of Cldn5 (Fig. 16B). On the other hand, the aforementioned variations of the sequences between the two homologs do not induce significant differences in the electrostatic potential surfaces Pore II tetramers (Fig. 16C,D), as also confirmed by their similar response to the ion permeation described by the FE profiles (Fig. 13D, 14D). These results reveal that Pore I, but not Pore II, discerns between the differences in anion permeation for Cldn4 and Cldn5 as a consequence of the different pore-lining residues, while the ability of the second model to preserve the tissue specificity of these two Cldn homologs is more limited [32,63].

**5. Accuracy and limitations in the study of biological structures with computational modelling methodologies**

In the various sections discussed in this review, it emerges that the introduction of computer-aided methodologies has significantly enhanced the knowledge on Cldn-based systems, providing molecular and atomistic explanations to the experimental outcomes. However, the accuracy and the predictivity of these techniques are limited by different factors that are listed below.

(i) The literature lacks experimental data on Cldn structures. Few X-RD crystals are available, with only Cldn15 resolved in the isolated form (PDB ID: 4P79 [52]), and the others determined in the cCPE bound form (Table 1). No high-resolution structure of high-order Cldn multimers is currently available, so that a direct comparison of computationally predicted models and experimental references is still missing.
(ii) A property limiting the accuracy of the predictions of MD simulations concerns the intrinsic approximation of the force-fields adopted to compute the inter-atomic forces [268,269]. The most popular force-fields for all-atom simulations (CHARMM36m [36], AMBER ff14SB [38], OPLS-AA/M [41] and GROMOS 54A8 [40]) are defined as a sum of pairwise-additive terms, in which the parameters derive from ab-initio quantum calculations and experimental nuclear magnetic resonance (NMR) data. For these models, the performance has been typically benchmarked on simple systems, such as small protein or isolated protein domains [36,38,270–273] because of the high-throughput of the simulations and the large availability of experimental data [274]. Nevertheless, the resources available nowadays allow to compute systems counting up to millions of atoms, also thanks to the support provided by the implementation of the GPU-accelerating technology, defining a gap between the systems adopted for the validation procedure of the force-fields and those in which the force-fields are effectively used [275–277]. Another important issue of the most popular force-fields currently used in the literature concerns the absence of polarizability in the atomic parameters [268], leading to significant inaccuracies in the description of the interactions between divalent cations and anions [278–281].

(iii) Structural remodelling, such as the rearrangement of the TJ strands [206,229,259,261,262] or the Cldn dimerization process [111,112,115], generally occurs on timescales that are not explored by classical all-atom MD simulations. The standard MD simulation set-up adopts the canonical 2-fs-long time step and trajectories longer than 1μs are rarely computed, and the uniform sampling of each degree of freedom often leaves many regions of the system configurational space unexplored [282]. Different strategies are used to overcome these limitations, as described above. Most phenomena involving the spontaneous formation or rearrangement of long Cldn assemblies are studied using CG methods [47]. This approach is based on the reduction of the system degrees of freedom by merging groups of atoms into representative fragments of the system, named beads, to enable the use of longer time steps [47,254]. While these methods increase the temporal horizon of the simulation, they are not able to finely describe the molecular processes of the system, which is reverse mapped to the all-atom structure in a post-production analysis phase [283]. In addition, the approximations introduced in the CG representation alter the correlation between the simulations and experimental time scales, thus requiring a specific rescaling depending on the system under study [284]. A different approach to study events requiring an extended exploration of the system configurational space with all-atom MD simulations consists in the introduction of enhanced sampling methods [282,285–287]. These techniques were developed to map the rough energy landscapes that characterize biological systems [288], in which barriers among different possible stable and metastable states limit the capacity of the system to escape from local minima and visiting different configurations in the time scale of a typical MD simulation [282]. A variety of strategies were developed to introduce biases aiming at facilitating the sampling of the less visited regions in the configurational space and at recovering the ergodicity of the system. Some of them require the identification of a proper set of CVs to follow the event of interest during the simulation. These methods generally involve the addition of external terms in the potential energy function, such as umbrella sampling (US) [211,289] and the metadynamics [290] methods, or the direct computation of the free energy gradient, as prescribed by thermodynamic integration [282] and the related variants, including adaptive-biasing force [291], blue-moon [292], temperature-accelerated MD (TAMD) [293] and others [236,294]. As a representative example, with the implementation of metadynamics [290], the system is provided with a sort of “memory” through the introduction of additional potential energy to the visited states during the simulation to discourage re-sampling of the same regions of the conformational space, in favor of those yet unexplored. Conversely, among the free energy gradient-based methods, the TAMD algorithm [293] is based on accelerating the sampling along the given CV through the introduction of an artificial temperature in the equation of motions. A different approach with respect to biasing a chosen set of CVs to overcome the energy barriers separating different configurations of the system consists in increasing the temperature, as prescribed by the parallel tempering methods, such as the replica-exchange protocol [238,295]. However, both the CV-based and tempering approaches have some drawbacks: while the first strictly depends on the proper choice of the CV to describe the event of interest, the second has a higher computational cost because of the acceleration of all degrees of freedom to the same extent [282]. For this reason, the application of combinatorial methods is becoming progressively more popular [296–299].

(iv) Another limitation of the computational modelling consists in the approximated characterization of the cellular environment and, in particular, of the impact of membrane composition on protein function. While most all-atom MD simulations of multimeric Cldn systems were performed in the presence of a homogenous composition of the membrane [58,116–118,241,265], the CG approach [112,130] could investigate the role of the heterogenous lipid phase on Cldn self-assembly with the protocols adopted for other systems [300–304]. It is worth mentioning that studies including all-atom MD simulation of Cldn monomers embedded in heterogenous membrane are also found [130], facilitated by the implementation of the different lipid components in the CHARMM-GUI input generator interface [305–309].

(v) A common procedure to reproduce protein–protein interactions at an affordable computational cost consists in the use of molecular docking simulations [243,310–312]. These algorithms explore a huge number of putative complex conformations and adopt a scoring function to estimate their probability to occur [225,242,243,313–319]. Despite this strategy is extremely useful to provide an extensive screening, by dropping the computational burden that typically characterizes the MD simulations, several factors affect the docking accuracy [320]. Among these, an initial stable configuration of the protein structures is fundamental, which can be obtained from the Protein Data Bank (PDB), or generated by homology or ab-initio modelling methods. Additionally, another important issue is represented by the choice of a rigid-body or a flexible docking algorithm, depending on the availability of a reliable initial guess for the protein complex and the computational resources to explicitly address the proteins flexibility during the simulation [321]. The majority of docking methods are designed and benchmarked for water soluble proteins [322], while only a limited number of algorithms are specifically customized for membrane proteins [242,319], such as Cldns. Furthermore, the accurate representation of the binding energy is still a major problem, due to the approximate description of various contributions, such as entropic effects, solvent, protonation of the charged
residues, etc. [321]. Nevertheless, the periodic assessment of the performances and benchmark testing with initiatives such as the Critical Assessment of Prediction of Interactions (CAPRI) [322,323] make possible to constantly improve the software and overcome these limitations, increasing the accuracy year after year.

In conclusion, the predictive potential of MD simulations significantly boosted the investigation of Cldn-based systems at distinct structural levels thanks to the synergistic integration of molecular docking and all-atom/CG simulations. However, limitations in the accuracy are still present and crucial improvements can be provided by the ongoing refinement of the force-fields. For example, the explicit treatment of the polarizability with methods such as POSSIM [324], AMOEBA [325] or the Drude oscillator model [326] is currently possible, although these implementations increase the computational cost [268], despite the optimization of the GPU-based technology [326–329]. Additionally, the future improvements and construction of force-fields will be driven by the adoption of machine learning algorithms [330,331] that will allow to capture features for which the classical models fail [332]. A similar direction is also followed to improve the performances of the docking algorithms, implementing deep learning approaches to generate more reliable scoring functions [330,333].

Finally, in the context of Cldn-based systems, the comprehensive description of the multimeric arrangements will be enabled by the acquisition of high-resolution structures of Cldn complexes, which will represent the breakthrough required to validate or confute the Cldn-based TJ models postulated and simulated in the last decade.

6. Summary and outlook

The investigation of the Cldn-based systems is essential for a comprehensive description of the physiology of TJ barriers. In the last years, the application of computational methodologies to this topic enhanced the predictivity on the Cldn structural assemblies and the details of their characterization. The combination of MD simulations and molecular docking calulations allowed the study of multimeric arrangements of various dimensions, investigating both the local phenomena, such as single-ion translocation or Cldn-Cldn association, and the global properties, including the dynamic rearrangement and the mechanical features of TJ strands. In this context, various studies suggest that the building blocks of strands are Cldn-based tetrameric units arranging in a pore-like architecture with a network of cis- and trans-interactions. To date, two main configurations of paracellular pores have been proposed. The first one is based on an antiparallel double row of cis-Cldn15 subunits characterized by two distinct intracellular interaction patterns. The association between two copies of this assembly belonging to opposite membranes contribute to the formation of ECL-based trans-interactions, defining multiple β-barrel pore-like tetramers (Pore I). In contrast, the second model (Pore II) proposes an alternative configuration resulting from the trans-interactions of two identical cis-dimers, each stabilized by a hydrophobic cisleucine zipper TM motif. Despite some experimental results corroborated both arrangements, others provided criticisms about their feasibility. In this context, the validation was of primary interest for independent computational investigations published afterwards. Specifically, FE calculations of water and ion permeation events were used to inspect the transferability of the two pore models among different classic Cldns. However, only recently the systematic comparison of the thermodynamic properties was shown for Cldn5 and Cldn4, revealing that the two pore-like multimers provide analogous ion selectivity for the first homolog, while being mutually exclusive for the second protein. In conclusion, despite the molecular description of the Cldn-based multimers is becoming progressively more accurate thanks to the synergy between experimental and computational investigations, several questions remain unanswered. In particular, the possible coexistence of these and others models is still to be probed, along with their arrangement in real TJ systems. The increasing number of studies focused on the properties of different Cldn multimeric models, in conjunction with the efforts made to release experimental structures of monomers and higher order assemblies, will contribute to elucidate the physiology of these systems. Future work should also be directed to the development of protocols to modulate the paracellular permeability in a safe and reversible manner, in order to deliver drugs across different compartments of the human body.

Funding sources

The research was supported by IRCCS Ospedale Policlinico San Martino (Ricerca Corrente and 5x1000 grants to Fabio Benfenati and Luca Maragliano) and by Telethon/Glut-1 Onlus Foundations (GSP19002_PAsGlut009 and CS22A002 projects to FB).

Credit authorship contribution statement

Alessandro Berselli: Conceptualization, Writing – review & editing, Visualization. Fabio Benfenati: Supervision, Writing – review & editing. Luca Maragliano: Supervision, Project administration, Writing – review & editing. Giulio Alberini: Supervision, Writing – review & editing, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Jörg Piontek for useful discussions. We sincerely thank all six Reviewers for carefully reading our work and for their useful suggestions. We acknowledge the HPC Support Team at Fondazione Istituto Italiano di Tecnologia (IIT). Computing time allocations were granted by the CINECA supercomputing center under the ISCRA initiative and by the HPC center at IIT.

References

[1] Zihni C, Mills C, Matter K, et al. Tight junctions: from simple barriers to multifunctional molecular gates. Nat Rev Mol Cell Biol 2016;17:564–80.
[2] Günzel D, Yu ASL. Claudins and the modulation of tight junction permeability. Physiol Rev 2013;93:525–69.
[3] Meoli L, Günzel D. Channel functions of claudins in the organization of biological systems. Biochim Biophys Acta Biomembr 2020;1862:183344.
[4] Ellouby-Non L, Ben-Yosef T, Chapter 1 – Functions of Claudin Tight Junction Proteins and Their Complex Interactions in Various Physiological Systems. Academic Press, International Review of Cell and Molecular Biology; 2010. p. 1–32.
[5] Krause G, Piotze J, Piontek J. Assembly and function of claudins: Structure-function relationships based on homology models and crystal structures. Semin Cell Dev Biol 2015;42:3–12.
[6] Farquhar MG, Palade GE. Junctional complexes in various epithelia. J Cell Biol 1963;17:375–412.
[7] Günzel D. Claudins: vital partners in transcellular and paracellular transport coupling. Pflügers Arch 2017;469:35–44.
[8] Claude P, Goodenough DA. Fracture faces of zonulae occludentes from “tight” and “leaky” epithelia. J Cell Biol 1973;58:390–400.
[9] Claude P. Morphological factors influencing transepithelial permeability: A model for the resistance of theZonula Occludens. J Membr Biol 1978;39:219–32.

A. Berselli, F. Benfenati, L. Maragliano et al. Computational and Structural Biotechnology Journal 20 (2022) 5984–6010
Angelow S, El-Husseini R, Kanzawa SA, et al. Renal localization and function
Carpenter EP, Beis K, Cameron AD, et al. Overcoming the challenges of
Le Moellic C, Boulkroun S, González-Nunez D, et al. Aldosterone and tight
Rajagopal N, Irudayanathan FJ, Nangia S. Palmitoylation of claudin-5 proteins
Van Itallie CM, Gambling TM, Carson JL, et al. Palmitoylation of claudins is
Gong Y, Yu M, Yang J, et al. The Cap1–claudin-4 regulatory pathway is
Samanta P, Wang Y, Fuladi S, et al. Molecular determination of
Dowal L, Yang W, Freeman MR, et al. Proteomic analysis of palmitoylated
Zorko MS, Veranic P, Leskovec NK, et al. Expression of tight-junction proteins
Irudayanathan FJ, Trasatti JP, Karande P, et al. Molecular architecture of the
Zhu Y, Brännström M, Janson P-O, et al. Differences in expression patterns of
Hou J, Renigunga A, Konrad M, et al. Claudin-16 and claudin-19 interact and
Baala L, Hadj-Rabia S, Hamel-Teillac D, et al. Homozygosity mapping of a
Fromm M, Piontek J, Rosenthal R, et al. Tight junctions of the proximal tubule
Levental I, Lingwood D, Grzybek M, et al. Palmitoylation regulates raft affinity
Hou J, Gomes AS, Paul DL, et al. Study of claudin function by RNA interference.
Danev R, Yanagisawa H, Kikkawa M. Cryo-electron microscopy methodology:
Cheng Y. Single-particle cryo-EM-How did it get here and where will it go.
Rajagopal N, Durand AJ, Nangia S. Predicting selectivity of paracellular pores
Irudayanathan FJ, Wang X, Wang N, et al. Architecture of the paracellular
Samanta P, Wang Y, Fuladi S, et al. Molecular determination of claudin-15 organization and channel selectivity. J Gen Physiol 2018; 150:949–68,
Rossaj, Ploeger C, Vorreiter F, et al. Claudin-3 and claudin-5 protein folding and assembly into the tight junction are controlled by non-conserved residues in the transmembrane loop 2 (TML2) segments. J Biol Chem 2014;289:7641–53,
Rajagopal N, Nangia S. Obtaining protein association energy landscape for integral membrane proteins. J Chem Theory Comput 2019;15:6444–53,
Rajagopal N, Durand AJ, Nangia S. Predicting selectivity of paracellular pores for biomimetic applications. Mol Syst Des Eng 2020;5:866–96,
Rodenburg RNP, Snieder J, van de Waterbeemd M, et al. Stochastic palmitoylation of accessible cysteines in membrane proteins revealed by native mass spectrometry. Nat Commun 2017;8:1280,
Levental I, Lingwood D, Gzybek M, et al. Palmitoylation regulates raft affinity for the majority of integral raft proteins. Proc Natl Acad Sci 2010;107:22050–4,
Cheng Y, Vivero-Escalante A, Hepler PK. The importance of claudin-7 palmitoylation on epithelial permeability or heterotypic claudin compatibility. Eur J Cell Biol 2006;85:633–44,
Baala L, Hadj-Rabia S, Hamel-Teillac D, et al. Homozygosity mapping of a
Fromm M, Piontek J, Rosenthal R, et al. Tight junctions of the proximal tubule
Levental I, Lingwood D, Grzybek M, et al. Palmitoylation regulates raft affinity
Hou J, Gomes AS, Paul DL, et al. Study of claudin function by RNA interference.
Danev R, Yanagisawa H, Kikkawa M. Cryo-electron microscopy methodology:
Cheng Y. Single-particle cryo-EM-How did it get here and where will it go.
Rajagopal N, Durand AJ, Nangia S. Predicting selectivity of paracellular pores
Irudayanathan FJ, Wang X, Wang N, et al. Architecture of the paracellular
Samanta P, Wang Y, Fuladi S, et al. Molecular determination of claudin-15 organization and channel selectivity. J Gen Physiol 2018; 150:949–68,
Rossaj, Ploeger C, Vorreiter F, et al. Claudin-3 and claudin-5 protein folding and assembly into the tight junction are controlled by non-conserved residues in the transmembrane loop 2 (TML2) segments. J Biol Chem 2014;289:7641–53,
Rajagopal N, Nangia S. Obtaining protein association energy landscape for integral membrane proteins. J Chem Theory Comput 2019;15:6444–53,
Rajagopal N, Durand AJ, Nangia S. Predicting selectivity of paracellular pores for biomimetic applications. Mol Syst Des Eng 2020;5:866–96,
Rodenburg RNP, Snieder J, van de Waterbeemd M, et al. Stochastic palmitoylation of accessible cysteines in membrane proteins revealed by native mass spectrometry. Nat Commun 2017;8:1280,
Levental I, Lingwood D, Gzybek M, et al. Palmitoylation regulates raft affinity for the majority of integral raft proteins. Proc Natl Acad Sci 2010;107:22050–4,
Cheng Y, Vivero-Escalante A, Hepler PK. The importance of claudin-7 palmitoylation on epithelial permeability or heterotypic claudin compatibility. Eur J Cell Biol 2006;85:633–44,
Baala L, Hadj-Rabia S, Hamel-Teillac D, et al. Homozygosity mapping of a
Fromm M, Piontek J, Rosenthal R, et al. Tight junctions of the proximal tubule
Levental I, Lingwood D, Grzybek M, et al. Palmitoylation regulates raft affinity
Hou J, Gomes AS, Paul DL, et al. Study of claudin function by RNA interference.
Danev R, Yanagisawa H, Kikkawa M. Cryo-electron microscopy methodology:
Cheng Y. Single-particle cryo-EM-How did it get here and where will it go.
Rajagopal N, Durand AJ, Nangia S. Predicting selectivity of paracellular pores
Irudayanathan FJ, Wang X, Wang N, et al. Architecture of the paracellular
Samanta P, Wang Y, Fuladi S, et al. Molecular determination of claudin-15 organization and channel selectivity. J Gen Physiol 2018; 150:949–68,
[329] Huang J, Lemkul JA, Eastman PK, et al. Molecular dynamics simulations using the drude polarizable force field on GPUs with OpenMM: Implementation, validation, and benchmarks. J Comput Chem 2018;39:1682–9.

[330] Befort BJ, DeFever RS, Tow GM, et al. Machine learning directed optimization of classical molecular modeling force fields. J Chem Inf Model 2021;61:4400–14.

[331] Unke OT, Chmiela S, Sauceda HE, et al. Machine learning force fields. Chem Rev 2021;121:10142–86.

[332] Wang J, Olsson S, Wehmeyer C, et al. Machine learning of coarse-grained molecular dynamics force fields. ACS Cent Sci 2019;5:755–67.

[333] Crampion K, Giorkallos A, Deldossi M, et al. Machine-learning methods for ligand-protein molecular docking. Drug Discov Today 2022;27:151–64.