Molecular Dynamics Simulations of Tight Junction Proteins

Eleni Fitsiou

This dissertation is submitted for the degree of

Doctor of Philosophy

August 2020

Department of Chemistry

I would like to dedicate this dissertation to my husband Antonios and sons

Dimitrios and Ioannis.

“Wisdom begins in wonder”

Socrates

Declaration

I, Eleni Fitsiou, declare that this thesis titled 'Molecular Dynamics Simulations of Tight Junction Proteins' has not been submitted in support of an application for another degree at this or any other university. It is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated. Where I have quoted from the work of others, the source is always given.

Lancaster University, UK

Abstract

Tight junctions (TJs) are specialised cell-cell structures that serve primarily as a barrier to molecular transport through the intercellular space between the cells. The claudin family of proteins are the main structural and functional components of the TJ strands that circumscribe the cells. The detailed molecular organisation at the TJs is not entirely resolved, being relatively inaccessible by current experimental methods. Here, we have employed molecular dynamics simulations using both atomistic and coarse-grained models to investigate the TJ structure formed by claudin-1 using self-assembly coupled with free energy calculations and enhanced sampling techniques. A feature of the studies is that the self-assembly simulations have been carried out using atomistic detail (a first) by simulating only the extracellular domains of claudin-1 in an implied membrane.

The results show that the cis-interaction can occur in the absence of trans-interacting partners and that a claudin dimer is the smallest stable unit. The dimers further form higher-order aggregates with a plethora of interacting dimeric interfaces. The trans-interaction of claudins resulted in a compact structure with a minimal pore size confirming the barrier properties of claudin-1. The simulations also enabled the identification of the key regions of the claudin responsible for the trans-interaction, with the identified important amino acids being in agreement with experimental studies. The role of the lipid environment, with a focus on the skin lipids in the stratum granulosum, was also investigated, along with the effect of single-point mutations in claudin-1. The single-point mutation studies were consistent with experimental results.

The simulation studies have enhanced our understanding of the assembly and structure of claudin-1 TJs, a notable finding being that kinetic locking is likely to be important in determining the TJ structure. The single-point mutation studies suggest that simulations could serve as screens towards defining potential gene-therapy strategies.

Acknowledgements

First of all, I would like to sincerely thank my supervisor Prof. Jamshed Anwar for his support throughout the project. His expertise has been invaluable and his guidance and help has been much appreciated. I would also like to thank Dr Massimo Noro (STFC) for his advice, support and offered opportunities; Dr Gareth Tribello (Queen's University Belfast) for his invaluable help with the metadynamics simulations; and Dr Mike Pacey for his assistance in resolving issues with HEC.

I would like to acknowledge Unilever for funding this project and promoting research. In particular, I would like to thank Scott Singleton and Patrick Warren for organising my visit to Unilever and their valuable feedback on my work.

Lastly, I would like to thank my colleagues and friends from the Chemistry Department at Lancaster University, for a very friendly and inclusive environment; especially, the Chemical Theory and Computation group, for their collaboration and assistance over these years.

Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

BBB	Blood-brain barrier
CG	Coarse-grained
CV	Collective variable
DOPC	1,2-dioleoyl-*sn*-glycero-3-phosphocholine
ECD	Extracellular domain
ECL	Extracellular loop
EM	Electron microscopy
FES	Free energy surface
LJ	Lennard-Jones
MD	Molecular dynamics
MM/PBSA	Molecular mechanics Poisson Boltzmann surface area
NMR	Nuclear magnetic resonance
NPT	Isothermal -isobaric ensemble
NVT	Canonical ensemble
PBC	Periodic boundary conditions
PMF	Potential of mean force
POPC	1-palmitoyl-2-oleoyl-glycero-3-phosphocholine
RMSD	Root mean square deviation

RMSF	Root mean square fluctuation
SASA	Solvent accessible surface area
SC	Stratum corneum
SG	Stratum granulosum
TEER	Transepithelial electrical resistance
TJ	Tight junction
vdW	Van der Waals
WHAM	Weight histogram method

1 Introduction

1.1 Overview

A multicellular organism is separated into compartments and its ability to regulate the various conditions in each compartment is called homeostasis. Epithelial cells together with endothelial and mesothelial cells are types of cells that line the external surfaces of our body (i.e. skin) and internal organs or cavities (e.g. intestinal lumens, blood vessels, nasal cavities), and help organisms to regulate different microenvironments (Tsukita et al., 2001). These cells are closely packed and arranged in one or more layers to form cellular sheets that provide boundaries between different compartments.

In order for cellular sheets to exert their role, the cells must be mechanically linked to each other and function as an effective barrier to prevent molecules from passing between and through them. Solutes can diffuse through the intercellular space (paracellular route) or through the intracellular pathway (transcellular route) in which case their transport is regulated by the cell (Tsukita et al., 2001).

There are four distinct types of cell-cell contact sites, namely tight junctions (TJs), adherens junctions, desmosomes, and gap junctions. TJs are considered responsible for creating the seal between adjacent cells that limit the uncontrolled passage of ions and molecules through the paracellular route (Alberts et al., 2015). Adherens junctions and desmosomes are anchoring junctions connecting neighbouring cells, while gap junctions form communication ‘bridges’ between cells.

The distinctive arrangement of cell-cell junctions was first detailed by Farquhar and Palade in 1963, who described the junctional complexes in various epithelia (Farquhar and Palade, 1963). Recently, there is an increased interest in the area of TJs due to their importance in novel drug delivery systems, their involvement in cancer propagation and homeostasis. Initially, the morphology of cellular junctions was investigated with the help of electron microscopy (EM) where the TJs appeared as a series of membrane fusions called 'kissing points'. Subsequently, freeze fracture EM studies showed that

TJs are a continuous, anastomosing cross-linked network of intramembranous particle strands (Lee et al., 2008).

Two models have been proposed for the structure of functional TJs, i.e. the lipid and protein models (Lee et al., 2008), which would seal the gap between the cells. In the past, the lipid model was more favourable, but in recent years due to the discovery of the key transmembrane TJ proteins, the protein model prevailed. In a recently published review paper, the significance of the presence of lipids in any functional model is highlighted, and a hybrid model was proposed as more appropriate (Zihni et al., 2016).

The discovery of integral membrane proteins has been a breakthrough in this field, contributing significantly to our understanding of the molecular architecture of TJs. In 1986, the first TJ-associated protein (ZO-1) was identified (Stevenson et al., 1986) and in 1993, the first integral protein was identified (occludin) (Furuse et al., 1993). However, it was found that occludin could not be reconstituted by itself to form TJ strands. Furthermore, due to lack of charged amino acids in its extracellular loops (ECLs), occludin failed to explain the charge selectivity shown by the epithelial barrier (Furuse, 2010a). Thus, the research for functionally significant proteins led to the identification of the claudin family, which are transmembrane proteins with two extracellular loops. Claudins are believed to be the fundamental unit that forms oligomers and lead to the formation of a seal between adjacent cells (Günzel and Alan, 2013). Claudins together with occludin and possibly other transmembrane proteins compose the backbone of TJ strands or fibrils that function as a gate, controlling the permeation of solutes between the cells. It is believed that their ECLs create the wall or the holes on the wall in the case of leaky epithelia. Specifically, some studies point out the importance of specific amino acids present in the second half of the first ECL, which is thought to be more important for lining the TJ pore and the ion selectivity shown by some epithelia (Furuse, 2010a). Claudin's second ECL was suggested to be important for the interaction between opposing cells (Krause et al., 2008). It is emphasised however, that both claudin loops are indispensable in order for TJ to exert their functional role. The permeation pathway is believed to consist of pores of variable diameters depending on the proteins expressed. A definitive size will be obtained when the structure of TJs will be ascertained at an atomistic level resolution.

The first perception of TJs was that they are stable inactive barriers, but it is now recognised that they are dynamic structures (Weber, 2012). The dynamics of their molecular structure is correlated with their barrier function. However, the fundamental questions regarding TJs' architecture and molecular composition have still not yet been fully answered.

The research progress has been slow compared to the other cell-cell junctions due to the lack of structural data (X-ray or nuclear magnetic resonance (NMR) studies) concerning the backbone of the TJ strands – the proteins of the claudin family, as it is now emphasised. An important milestone in the progress was the discovery of the crystal structures of claudin-15 at high resolution (2.4 angstroms) (Suzuki et al., 2014) and subsequently of claudins-19 and -4 in complex with *Clostridium perfringens* enterotoxin (Saitoh et al., 2015, Shinoda et al., 2016). Also, the crystal structure of human claudin-9 in complex with the C-terminal of the above mentioned enterotoxin was recently resolved and suggests possible mechanisms for the disruption of the TJ assembly (Vecchio and Stroud, 2019). Recently, the crystal structure of claudin-3 at 3.6 Å resolution was also resolved, and it revealed that the differences in the tertiary structure of each claudin changes the position of the interacting residues, thus, affects the morphology of the TJ strands (Nakamura et al., 2019). It is noted however, that the molecular organisation of TJ claudins in their native environment i.e., when embedded in lipid bilayers is currently inaccessible by experiment. The mechanism of claudins assembly into strands is still unclear.

The need to elucidate the interactions between TJ proteins that create the 10 nm particle seen by freeze-fracture EM is urgent, since TJs are involved in many diseases. The question is how do claudins interact at an atomistic level to create charge and size selective pores? It is hypothesised that claudins interact in the plane of the same plasma membrane (cis-interaction) and create a strand that spatially associates with a strand from the neighbouring cell (trans-interaction) leading to a 'paired strand' structure. The knowledge regarding the interactions that lead to higher order structures is fragmentary. There is insufficient evidence to support models that speculate the mechanism of oligomerisation.

Moreover, in a given tissue there are different combinations of TJ proteins. For instance, it is stated that a combination of more than three different claudins is the usual case in

epithelial tissues (Tamura and Tsukita, 2014). Furthermore, two distinct pathways have been proposed to explain the permeability of small molecules, ions and macromolecules through the TJs (Liang and Weber, 2014). The different combinations and the two distinct pathways mentioned above, result in increasing the complexity of the pore structure and thus, renders it difficult to clarify its molecular architecture.

It must be emphasised that the need to characterise the barriers in terms of the pore or no-pore pathways is important. Defining the nature of these structures will help us understand how the barrier is created, regulated and altered in diseases. Also, the ability to modify the barrier properties by design is of vital importance since it could facilitate therapy significantly. The repair and maintenance of a dysfunctional paracellular barrier could help the treatment of inflammatory diseases or even cancer metastasis (Förster, 2008). In addition, the design of effective skin care products could enhance the wellbeing of many patients suffering from skin diseases or allergies.

To summarise, although the newly identified crystal structures of claudins and the identification of antibodies specific for claudins have enhanced the progress in the research field, there are still unanswered questions that need to be addressed. The atomistic resolution offered by molecular dynamics (MD) simulations can help us understand the physical basis of the structure and function of these macromolecules. MD simulations is a mature technique that can provide information that cannot be obtained by experiments, and hence, help us clarify the basic principles of TJs assembly and function (Hospital et al., 2015). An MD simulation provides detailed information about the individual motions of the particles as a function of time. The generated trajectories will shed light in the way TJ proteins interact and assemble, forming the barrier between the neighbouring cells. Moreover, apart from sampling the configuration space of a system providing us with possible structural pathways, it characterises the properties of the model system at equilibrium.

1.2 Epithelial Cells, their Junctions and Cellular Transport

The existence of separate compartments with different composition is essential for the development, maintenance and function of multicellular organisms. Epithelial cells provide boundaries between compartments by coming together and forming semi-

permeable cellular sheets that line the surfaces and cavities of organs in multicellular organisms,. Therefore, they are crucial for many biological processes.

Epithelial sheets cover outer surfaces of our body or line internal cavities except blood vessels, the heart and serous cavities, which are lined by another type of cells called endothelial cells (Tsukita et al., 2001). Endothelial cells are of great importance for the blood-brain barrier (BBB), which hinders the delivery of drugs that can treat neurological diseases (Luissint et al., 2012). Additionally, there are flat mesoderm derived cells that line some body cavities called mesothelial cells. All of these cells are crucial for the maintenance of different compartments and the selective diffusion of nutrients, ions and solutes through their sheets.

Epithelial cells have different surface domains separating their surface into: a basal surface that faces the extracellular matrix or the underlying tissue; a lateral surface (side surface) that faces adjoining cells; and, an apical surface that faces the epithelial lumen or outside environment (Giepmans and van IJzendoorn, 2009). Therefore, epithelial cells show a remarkable polarisation of their plasma membrane, i.e. they show distinct surface domains with specific composition and properties. The polarisation of epithelial cells is a property that is attributed to the intercellular junctions and specifically to the TJs (Günzel and Alan, 2013).

1.2.1 Types of epithelial cells

Cells are closely packed to form cellular sheets and can be arranged in one or more layers, called simple or stratified epithelium, respectively. There is also the case of pseudostratified epithelium, where cells are arranged in one layer, but since their nuclei are in different levels they appear as having more layers (Histology, 2011).

Epithelial cells can also be categorised according to their shape and function. Their fundamental function is protection because they act as a physical and chemical barrier. For instance, epithelial cellular sheets, found in our skin's epidermis, are components of the complex epidermal permeability barrier. In the case of intestine, epithelial cells block the movement of toxins to the gut lumen. Additionally, the urinary bladder that is composed of epithelial cells has to act as a barrier to water and electrolytes in urine (Günzel and Alan, 2013). Apart from their protection role, they exert other functions such as secretion of substances, regulated exchange of molecules between an organism and its surroundings and also sensation.

With regards to their shape, epithelial cells can be classified as squamous, cuboidal and columnar. For example, the outermost layer of our skin, namely the epidermis, is composed of squamous stratified epithelial cells, which are flat-like cells that pile-up and function primarily as a shield to protect our organism from the outside environment (Morita and Miyachi, 2003).

1.2.2 Cellular transport pathways

Epithelial tissues not only discriminate compartments in our body, but are able to regulate the passage of solutes through them. There are two pathways that a solute can follow: either through the cell itself where the passage is regulated by membrane-bound transporters or channels; or through the space between the cells (intercellular space) (see Figure 1.1). The former is called transcellular pathway and the latter is the paracellular pathway; it's the latter which is controlled by intercellular TJ structures (Capaldo and Nusrat, 2015).

In the case of molecules going through the cell (transcellular pathway), they move perpendicular to the cell membrane and through the membrane lipid bilayer. While in the case of paracellular flux, the molecules diffuse through the intercellular space and do not enter the cell's bilayer (Alan, 2010). With regards to the paracellular pathway in epithelial cellular sheets, it is believed that any paracellular pores that exist are created by the TJ proteins (mainly claudins). The pores have varying diameter ranging from approximately 4-8 Å (Alan, 2010), but the structure of the pore and its molecular basis is not clarified. The permeation of ions and molecules through the cellular sheets is typically measured by assessing electrical resistance or conductivity (Zihni et al., 2016). A very common measure of the transepithelial ion permeability is the transepithelial electrical resistance (TEER), which characterises the permeability of epithelial sheets.

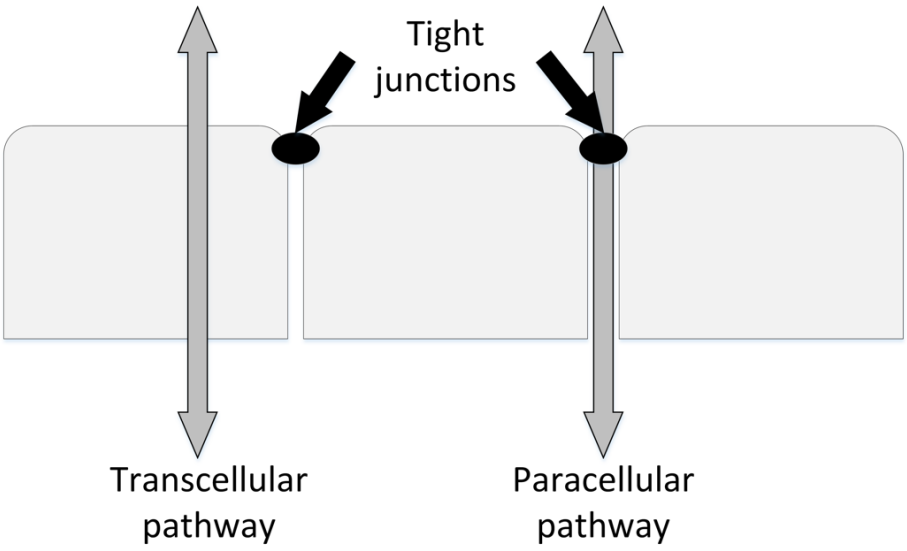


Figure 1.1: The two distinct cellular transport pathways: transcellular and paracellular.

1.2.3 Cell – cell junctions

The surface cells of the body are joined together by specific junctions. Some keep the cells together (adherens and desmosomes) and others prevent or regulate flow of molecules between the cells. The latter are the so-called TJs. There are also communication junctions called gap junctions that facilitate the movement of small molecules via channels between cells and connect their cytoplasm. A cell junction (or intercellular bridge) is a type of structure that consists of multiprotein complexes that provide contact between neighbouring cells or between a cell and the extracellular matrix.

It is believed that cell-cell junctions mediate and control the segregation of the apical from basal plasma membrane surfaces. As a consequence, these junctions are considered a primary landmark of cell polarity (Giepmans and van IJzendoorn, 2009). Cell-cell junctions are located on the side (lateral) surface and can fix the localisation of one cell relative to another. These junctions are proposed to have the ability to form 'fences' by blocking the movement of molecules in the plasma membrane and separate distinct environments. The apical and basolateral surface domains are two distinguishable domains of the cell surface with regards to cell polarity. However, the role of cell-cell junctions such as TJs and adherens in the establishment of surface polarity is not yet clear (Giepmans and van IJzendoorn, 2009). For instance, there is

data contradicting the role of TJs as a fence, showing that they are not vital for the asymmetrical distribution of proteins in the cell membrane. This observation is in disagreement with the once predominant lipid-based model of the TJs, where lipid and protein diffusion between the outer leaflets of apical and basolateral membranes was blocked by TJs. Furthermore, there have been observations that the formation of TJs follow cell surface polarity (Madden and Sarras, 1985), but it was noted that their contribution to the maintenance of cell polarity may be important.

The TJs, the adherens junctions and the desmosomes, compose the junctional complex, which is located at the most apical part of the cell lateral membrane (Tsukita et al., 2001). One of the differences between the two types of anchoring junctions is that adherens junctions are anchorage sites for actin filaments while desmosomes are anchorage sites for intermediate filaments (Alberts et al., 2015). The anchoring junctions differ in the cytoskeletal protein anchor, as well as, the transmembrane linker protein that extends through the membrane. Each junction is characterised from the transmembrane adhesion proteins that span the plasma membrane with one end linking to the cytoskeleton inside the cell and the other end linking to other structures. There are two types of external attachments, so the transmembrane proteins fall into two protein families. When cell-cell attachment is considered, proteins of the cadherins family mediate the anchoring, while in the case of cell to matrix attachment, proteins of the integrin family mediate the connection (Alberts et al., 2015).

Adherens junctions are found below the TJs and often form a continuous adhesion belt, especially in epithelial cells. The interacting plasma membranes are held together by cadherins, which interact to 'zipper up' the two adjacent cells. Classical cadherins are transmembrane proteins with five characteristic extracellular domains (ECDs), that initiate cell contacts through trans-interaction, that is through interaction of their ECDs on adjacent cells (Hartsock and Nelson, 2008). Cadherins bind in multiple conformations and mediate strong adhesion between cells. A known adhesive state is called the X-dimer and in a recent study by Manibog et al., the role of calcium ions in the catch bond formation was demonstrated (Manibog et al., 2014). The ions enhance the rigidity of the ECD and promote adhesion under tensile force that flexes the X-dimers and induces long –lived hydrogen bonds formation. The study used single-molecule force-clamp spectroscopy, atomic force microscopy, coupled with MD

simulations and steered MD simulations in order to characterise the mechanism of catch bond formation that locks cadherins into closer contact.

Desmosomes appear as spot-like particles, riveting the cells together, providing mechanical strength (Garrod and Chidgey, 2008). A bundle of intermediate keratin filaments is attached to the surface of a plaque, which is located on the cytoplasmic surface of each interacting plasma membranes. These dense plaques are composed of mixtures of intracellular adaptor proteins. Their transmembrane linkers are desmosomal cadherins, which bind to the plaques and interact through their extracellular domains (ECDs) to hold the adjacent membranes together.

Gap junctions are located near the basal side of the cell. They are clusters of intercellular channels that allow direct communication between cells through their cytoplasm (Goodenough and Paul, 2009). In electron micrographs, they appear as close membrane appositions leaving a narrow gap of about 2-4 nm (Alberts et al., 2015). There are two families of proteins that can be found in gap junctions, namely the connexins and the innexins. Both protein families have similar structures composed of four transmembrane segments and two ECLs. Moreover, they can both form hexamers that interact with hexamers from the adjacent cell and assemble into axial channels, forming the gap junctions (Abascal and Zardoya, 2013). The connexins hexamers called connexons, can interact with identical connexons from an adjacent cell and form homotypic channels; or with different ones and form heterotypic channels (Goodenough and Paul, 2009). These channels facilitate the passage of small molecules and ions between neighbouring cells. The fundamental unit of gap junctions is a hexamer (connexon) and it is has been suggested than a hexameric assembly is also the fundamental unit of the TJs proteins (Mitic et al., 2003).

Lastly, the TJs (or zonula occludens), which are of interest here, seal the gap between cells near their apical (upper) side and at the same time prevent the leaking across the epithelium.

Within the epithelial sheets, cells are attached to each other directly by cell-cell junctions and at their basal side their cytoskeleton is attached to the basal lamina through cell-matrix junctions. These attachments control the orientation and behaviour of the cell's cytoskeleton, thus, allow cells to sense and respond to changes of the environment. Note that in special cases, there are others types of cell- cell adhesion

molecules that provide transient cell-cell attachments weaker than anchoring junctions (Alberts et al., 2015).

1.3 Tight Junctions

1.3.1 Introduction - Tight junctions as multifunctional structures

TJs prevent or regulate the flow of molecules between the cells, through the intercellular gap. They act as either barriers or selective gates, allowing the permeation of molecules in a size and charge selective way and help multicellular organisms maintain homeostasis (Cereijido and Anderson, 2001). By forming a permeability barrier, TJs separate the microenvironment between two compartments, the one found on the side that is connected to another tissue (basal side), from the microenvironment with different chemical composition that flows on their opposite side (apical side) (Alberts et al., 2015). Apart from their fundamental gate function, they are believed to serve as a fence by blocking the movement of proteins and lipids within the cell membrane, thus, maintaining cell polarity. TJs are multifunctional complexes because they also coordinate signalling and trafficking molecules involved in cell differentiation, polarity and proliferation (Chiba et al., 2008).

TJs are located towards the apical part (upper side) of the lateral cell membranes and together with the adherens junctions and the desmosomes they constitute the junctional complex. When the complex was visualized with electron micrographs, the extracellular space in the case of the TJs seemed completely obliterated ('kissing points') while in the other two components of the junctional complex the apposing membranes were 15-20 nm apart (Tsukita et al., 2001). Initially, techniques such as ultrathin section EM and freeze-fracture replica EM were used to investigate the morphology of the junctions between cells. In the micrographs, TJs appeared as membrane fusions (described as linear fusion or series of focal fusions in different tissues) and as a belt-like network of anastomosing strands of intramembranous particles, respectively (Farquhar and Palade, 1963, Staehelin, 1974, Lee et al., 2008, Günzel and Alan, 2013). The formed strands or fibrils are viewed as 10 nm transmembrane particles (Anderson, 2001), initially assumed to be rigid, but now are believed to be dynamic particles undergoing changes – regularly being broken and annealed (Tsukita et al., 2001).

In the 1960s, the study across epithelial tissues led to the conclusion that epithelial cells display a wide range of electrical resistance leading to the discrimination between 'leaky' (e.g. small intestine, colon) and 'tight' (e.g. urinary bladder, skin) depending on their barrier properties (Turner et al., 2014, Anderson and Van Itallie, 2009). Furthermore, it was revealed that the so-called ‘leaky’ epithelia show size and ionic charge selectivity. Usually, in order to investigate the barrier properties of TJs we rely on electrical conductance measurements or flux assays of tracers (Liang and Weber, 2014). However, it is suggested that the barrier posed to macromolecules is different than that posed for small ions leading to the need for implementing additional assays.

A high permeability has been observed for small molecules (radius smaller than 4 Angstroms) through the extracellular space, while a relatively low permeability has been shown for macromolecules and it is hypothesised that this low permeability is due to barriers dysfunction or even temporary strand breaks (Anderson and Van Itallie, 2009). It is believed that at least two distinct pathways exist called pore and leak (Liang and Weber, 2014). To date, there is still fragmentary knowledge of how the proteins located in the TJs interact and form the strands. This knowledge would enhance our understanding of the relative contributions of the pore and leak pathway and the way they are altered in many diseases.

1.3.2 Tight junctions: description and proposed models

As aforementioned, with the help of ultrathin-section EM, TJs were described as linear fusions or a series of fusions of the outer leaflets of the lipid bilayers (Furuse, 2010b). However, freeze-fracture replica EM showed that the TJ is a network of dynamic strands that are intramembranous particles. These particles polymerise and the formed strands are associated with the strands of the adjacent cell leading to a paired strand. Anderson described the architecture of TJs as a row of particles that have a diameter of 10 nm and a distance of 18 nm from centre to centre, in all tissues (Anderson, 2001).

For many years, the nature of these particles (strands) was a subject of discussion. It is essential to elucidate their nature because it will help us understand how TJs are assembled and subsequently exert their function. Two models have been proposed, the lipid (da Silva and Kachar, 1982) and the protein model (Staehelin, 1974). In essence, the lipid model is a lipid-protein hybrid model because the role of proteins was always

accepted (Zihni et al., 2016). It should be emphasised that the protein model is widely accepted nowadays.

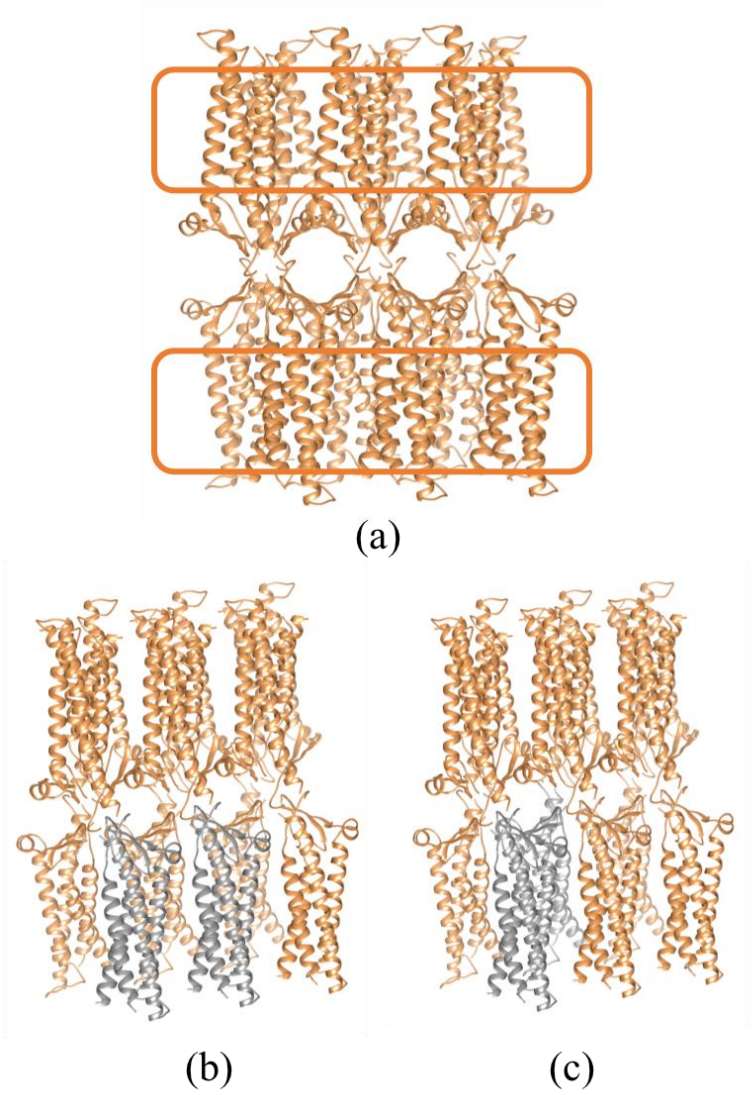


Figure 1.2: (a) The proposed Suzuki model for the arrangement of claudins at the paracellular TJ channels shown in cartoon representation. The rectangular frame is indicative of the two cell membranes where claudins are embedded, and the "β-sheet" pores are viewed from the top (apical) side, perpendicular to the elongation direction of the TJ strand. The linear (b) and 'face to face' (c) proposed interfaces are also highlighted in grey colour (Suzuki et al., 2015).

According to the protein model, in the standard lipid bilayer of the cell membrane, there are numerous integral transmembrane proteins (claudins, occludin, tricellulin etc.)

located at TJs and the barrier is formed by the part of these proteins that is extracellular. Thus, the barrier is regulated by the extracellular domains (ECDs) of the transmembrane proteins (Zihni et al., 2016). This model has gained support because the crystal structures of a number of TJ proteins were resolved at a high resolution (Suzuki et al., 2014, Shinoda et al., 2016, Saitoh et al., 2015, Nakamura et al., 2019, Vecchio and Stroud, 2019) and subsequently a model for the architecture of TJs has been proposed that fits to the ultrastructure of the strands (the so-called Suzuki model) (Suzuki et al., 2015). This model proposes that TJ strands consist of claudins protomers arranged in an antiparallel, double-row arrangement. Each contact site is composed of four claudins, two in each cell, which interact through their extracellular and transmembrane domains leading to a configuration that resembles a 'pipe' (see Figure 1.2). Therefore, the claudins interact in cis- (in the same cell membrane) and trans- way (from opposing cell membranes) leading to the formation of the TJ barrier or pore. In the Suzuki model the protein particles are thought to meet in the paracellular space in a 'head to head' manner and do not interdigitate like two parts of a zipper (Suzuki et al., 2015). The width of the TJ strand as observed through freeze-fracture EM images, is consistent with the dimensions of a claudin dimer. Note that the Suzuki model is based on cysteine crosslinking experiments that further support a claudin-15 dimer formed by interactions between the edges of their ECD (face-to-face interface). The model further suggests that, two adjacent cells could associate through the claudin strands (antiparallel rows) and form β-barrel like structures. This arrangement consists of tightly packed claudin monomers. The proposed model offers a possible structural explanation of the way claudins polymerise in a cell membrane, but the proposed linear arrangement does not agree with the observed branching network. Moreover, it fails to explain how other TJ proteins such as occludin integrate into the strands. The validity of the so-called Suzuki model is still debated (Alberini et al., 2017, Zhao et al., 2018, Samanta et al., 2018).

On the contrary, the lipid model proposes that the contact sites of the adjacent cells are membrane hemifusions (Zihni et al., 2016). A fundamental process during hemifusion is the transition of lipids from a lamellar configuration to an inverted hexagonal configuration where the acyl chains of the lipids are oriented outwards (da Silva and Kachar, 1982, Lee et al., 2008). Therefore, the observed TJ strands that pose the barrier at TJs are, according to the lipid model, cylinders of inverted lipid micelles. The

hemifusion state is unfavourable and so the proteins are proposed to stabilise the structure.

Although the protein model has gained support, there are issues that could not be satisfied solely by the protein model and the need for another model that could address these issues led to the consideration of a hybrid model. Specifically, there are observations that the outer leaflets are functionally linked across the TJs and this observation cannot be satisfied by the protein model because the diffusion of the lipids located in the exoplasmic leaflets cannot take place. Also, fluorescence recovery after photobleaching studies have demonstrated that lipid probes diffused from one cell to another (Grebenkämper and Galla, 1994). Additionally, in another study a lipid soluble anion (dipicrylamine) has been translocated from one cell to another across the TJs (Turin et al., 1991). Another observation is that the major constituents of the TJ strands, specifically the proteins of the claudin family, show weak adhesive activity compared to E-cadherins (Lee et al., 2008). It is also emphasised, that in order to obliterate the extracellular space the proteins should depict a membrane-bridging activity, but this has not been reported yet (Lee et al., 2008).

The exact nature of the TJ strands is unknown, however, both phospholipids and claudins have been localized at TJ strands (Kan, 1993). It is also argued that claudins can reconstitute TJ-like strands and not TJ strands, because claudins-1 and -2 ‘did not surround individual cells continuously’ as the authors pointed out (Furuse et al., 1998). Thus, the strands in the hybrid model are proposed to be lipidic particles where the proteins function in the lipid structure and in this way the lipids can fill the gaps and there is no need for continuous proteinaceous polymers (Zihni et al., 2016). However, based on this description it would be difficult for the TJ pore to show the charge selectivity that is now widely accepted. In many cases the charge selectivity is proved to be strongly related to the charge of specific residues of claudin’s ECD (mutation studies).

To summarise, one can state that the hybrid lipid-protein model can provide answers to questions like how do the intramembranous strands interconnect by branching (strands are not only linear connected) and how do other proteins such as occludin fit in the claudin based model (Zihni et al., 2016). However, there is now sufficient evidence that claudins are the main structural and functional elements of the TJs. Specifically, the

charges found in their ECD can rationalise the charge selectivity that the TJ pore shows, meaning that these charges are in direct contact to the solutes in the intercellular gap (Colegio et al., 2003, Piehl et al., 2010, Piontek et al., 2017). The fact that claudins are the main components of TJ strands is based on loss-of-function and gain-of-function experiments as well as on overexpression of these proteins in cells that normally lack TJs such as L. fibroblasts (Furuse, 2010b). Furthermore, cell biological analysis have revealed that tissues express more than one member of the claudin family (currently 27 members in total are identified) and therefore the diversity of the different barrier properties are mirrored in this mosaic of different combinations and proportions of claudins (Günzel and Alan, 2013).

To conclude, co-expression studies reveal the hetero- and homotypic way different claudins subtypes can co-polymerise into strands and further combinations can be expected based on the structure of their ECLs (Markov et al., 2015). Thus, it is now acknowledged that single claudins co-localise with other claudins within the TJ strand particles. Next the formed strand interacts with an opposing strand (from another lipid membrane) resulting as mentioned above in a paired-strand conformation. Claudins therefore, as the main components of the strands, are contributing to the epithelial barrier and transport in tissues. This is supported by evidence that correlates claudin expression patterns with functional parameters like selective paracellular permeability, analysis of diseases and knockout studies (Markov et al., 2015). All these observations strongly favour the support of the protein model which now prevails.

1.3.3 Claudins are the gate–keepers of tight junctions

The exact nature of TJs is complicated since they are multiprotein complexes comprised of transmembrane proteins (i.e. claudins, TAMPs and JAM-A) linked to scaffolding proteins through their PDZ-motif (e.g. ZO-1), which are in turn linked to the actin cytoskeleton by linkers such as cingulin (Van Itallie and Anderson, 2014). Numerous TJ proteins have been identified, including the 27 mammal members of the claudin family, after the recent expansion of the family to accommodate three more members (Mineta et al., 2011). Their identification, as previously noted, has supported greatly the protein model because claudins are considered responsible for forming barriers and channels, that selectively regulate the permeation of molecules through TJs based on their charge and size (Alan, 2010, Furuse, 2010a, Günzel and Alan, 2013, Krause et al.,

2009, Markov et al., 2015, Suzuki et al., 2015, Tamura and Tsukita, 2014, Tsukita and Furuse, 2000).

Claudins are named from the Latin word ‘claudere’ which means to close, and belong to the bigger PMP-22/EMP/MP20/claudin family (pfam 00822). They are further categorised into classic and non-classic claudins based on their sequence similarity and function (Günzel and Alan, 2013). The crystal structures of claudins have revealed that the ECLs are folded into a characteristic β-sheet motif while the transmembrane region of the protein composes a left hand, four-helix bundle (Suzuki et al., 2014). The extracellular part of claudin-15 is composed of two loops that are connected by a β-sheet domain (see Figure 1.3) (Suzuki et al., 2014). The first bigger loop (49-52 residues) has four β strands and the fifth strand is contributed by the second shorter loop (16-33 residues) (Anderson and Van Itallie, 2009). Also, this overall conformation is supported by the known crystal structures of other claudins, which are in complex with an enterotoxin (Saitoh et al., 2015, Shinoda et al., 2016, Vecchio and Stroud, 2019) and the more recently resolved crystal structure of claudin-3 (Nakamura et al., 2019). However, in the crystal structures of claudin-19 and claudin-4, the small helix present in the first loop is either replaced by a coil or is missing (see Figure 1.4 and Figure 1.5). This helix is considered to be important for the cis-interaction of claudins (linear interface) in the Suzuki model (Suzuki et al., 2015). The newly resolved crystal structure of claudin-3 has also pointed out the importance of the bend angle of the long helix (third transmembrane helix), which can be straight or bent depending on the residues present (see Figure 1.6) (Nakamura et al., 2019). This angle is important because it could determine the claudin –claudin interactions and affect the morphology and adhesiveness of the strand.

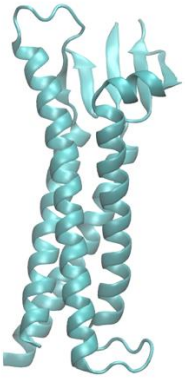


Figure 1.3: Crystal structure of mouse claudin-15 (ID: 4P79.pdb retrieved from the RCSB protein data bank and visualised with VMD)(Humphrey et al., 1996, Rose et al., 2010, Rose et al., 2016).

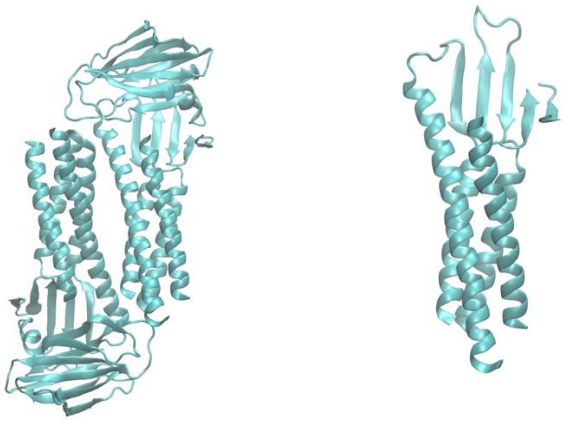


Figure 1.4: Crystal structure of mouse claudin-19 in complex with C-terminal fragment of Clostridium Perfringens enterotoxin (ID: 3X29.pdb retrieved from the RCSB protein data bank and visualised with VMD)(Rose et al., 2010, Rose et al., 2016, Humphrey et al., 1996). The left –hand side picture displays the whole pdb file, while the right-hand side only the chain A which is a single claudin -19.

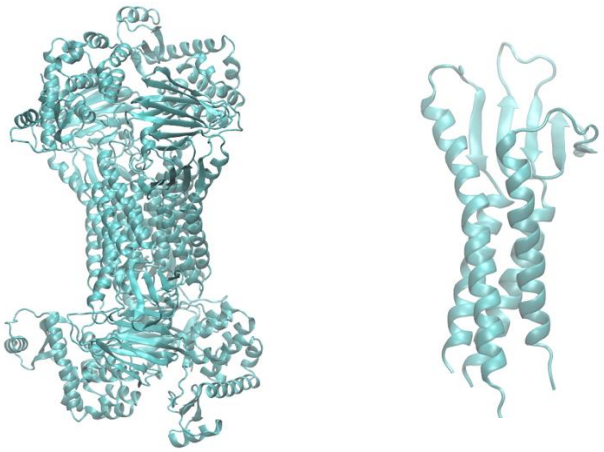


Figure 1.5: Crystal structure of human claudin-4 in complex with C-terminal fragment of Clostridium Perfringens enterotoxin (ID: 5B2G.pdb retrieved from the RCSB protein data bank and visualised with VMD)(Humphrey et al., 1996, Rose et al., 2010, Rose et al., 2016). The left –hand side picture displays the whole pdb file, while the right-hand side only the chain that has a single claudin-4.

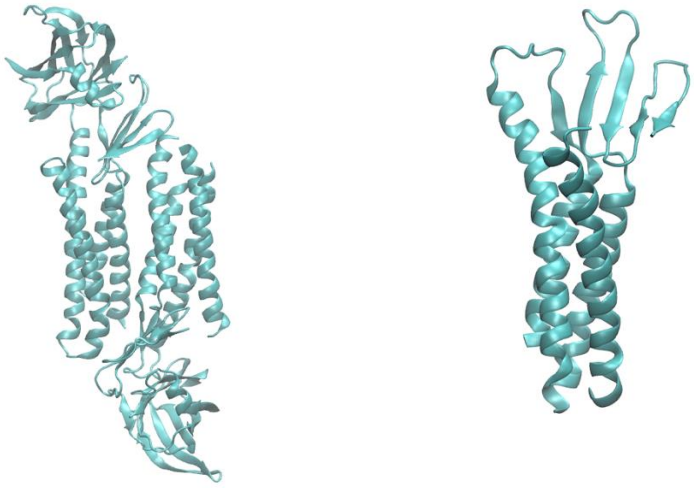


Figure 1.6: Crystal structure of mouse claudin-3 in complex with C-terminal fragment of Clostridium Perfringens enterotoxin (ID: 6AKE.pdb retrieved from the RCSB protein data bank and visualised with VMD)(Rose et al., 2010, Rose et al., 2016, Humphrey et al., 1996). The left –hand side picture displays the whole pdb file, while the right-hand side one, only the chain that has a single claudin -3.

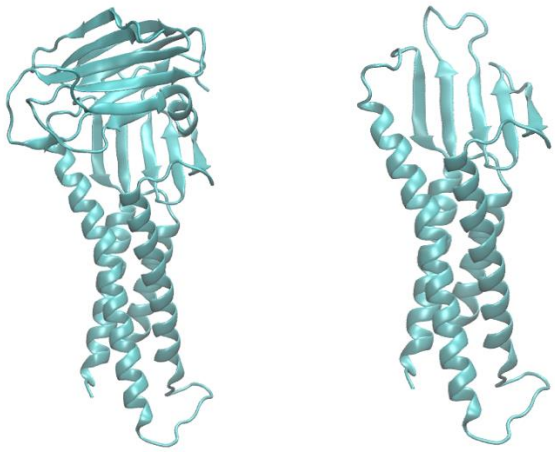


Figure 1.7: Crystal structure of human claudin-9 in complex with C-terminal domain of Clostridium Perfringens enterotoxin (ID:6OV3 retrieved from RCSB protein data bank and visualised with VMD ((Rose et al., 2010, Rose et al., 2016, Humphrey et al., 1996). The left-hand side picture displays the whole pdb file while the right-hand side one, only the chain that has a single claudin-9.

To summarise, TJs are multiprotein complexes because there are more than 50 proteins located there. Their functional and structural components are the family of claudins that are the gate-keepers, and there is now a consensus regarding their tertiary structure. Since the structure of a biological assembly is closely related to its function, the resolved crystal structures are a very important finding because they help to clarify claudins self-assembly and the characterisation of the TJ pore.

1.4 Structure and Function of Skin

1.4.1 Layers of the skin

Our interest is in claudin-1, which although quite ubiquitously expressed in the body, plays a key role in the skin. Given that the TJ proteins are embedded in lipid cell membranes, the nature and composition of the appropriate membrane is also crucial. In this section, we discuss the basic structure of skin because the TJs have only recently been found in the skin and it is now recognised that they play an important barrier role. Also, we examine the types of lipids found in stratum granulosum, a specific layer of our skin where the TJs are located.

The integument or skin is considered the largest organ of the body and forms a physical barrier to the environment. It is also considered an important route to deliver drugs into the body. Its primary function is to protect the body from physical, chemical, immune, and pathogen dangers and the damaging UV radiation and free radicals (Menon, 2002). It also regulates body temperature, stores water, fat and performs endocrine functions (synthesises vitamin D and peripheral conversion of pheromones) (Menon, 2002). The skin is a dynamic organ constantly repairing and renewing itself.

The human skin is a complex tissue consisting of several distinct layers, namely the epidermis, the dermis and the subcutaneous fat layer, or hypodermis (Shahzad et al., 2015). The outermost layer of the skin is the epidermis which serves as the physical and chemical barrier and contains keratinocytes, melanocytes, Langerhans' cells and Merkel cells (Has and Sitaru, 2013). The dermis, which is located underneath the epidermis, provides structural support and cushions the body from stress and strain. The dermis is much thicker than the epidermis, ranging from 1-4 mm (Marks et al., 2006). Strength, elasticity and extensibility are properties of the dermis, which can be attributed to the interlacing fibres (mostly collagen I and III and some elastin) located there (Foldvari, 2000). Additionally, blood and lymph vessels, hair follicles and glands (which produce sweat and sebum) are located in the dermis. Underneath the dermis is a loose connective tissue layer, the subcutis or hypodermis which attaches the skin to underlying bone and muscle and supplies the skin with blood vessels and nerves. Hypodermis is an irregular layer of adipose and connective tissue immediately deep to the skin, mainly used for fat storage.

1.4.2 Epidermis layers

The epidermis is the thinnest part of the skin and varies in thickness from 0.05 mm on the eyelids to 0.8-1.5 mm on the soles of the feet and palms of hands (Bensouilah, 2012). It provides protection to water vapour loss and helps the skin to regulate body temperature. It is made up of stratified squamous epithelial cells with an underlying basal lamina (basement membrane). The epidermis consists of 4-5 layers that are formed by the differing stages of keratin maturation. The *stratum corneum* (SC), the outermost layer of skin, is a layer of non-viable cornified cells (corneocytes) which are constantly shed and renewed. Under the SC is the viable epidermis which is further subdivided into the following strata: *stratum granulosum* (SG), *stratum spinosum* and

stratum basale (*stratum germinativum* or basal layer) (Shahzad et al., 2015). SG and *stratum spinosum* are sometimes referred to as the Malphigian layer. It is worth mentioning that in thick epidermis there is a variably thin layer of cells present which is called *stratum lucidum*.

1.4.3 The terminal differentiation process

The epidermal cells are constantly renewing themselves by going through the same differentiation process. The most abundant cell type located in the epidermis are the keratinocytes which as they differentiate from the basement membrane they generate cells for the superficial layers (Brandner et al., 2015).

The last layer of the epidermis is the SC (10 -20 μm thick) where the terminal differentiated keratinocytes-corneocytes (enucleated cells) are surrounded by a complex extracellular matrix ("bricks and mortar" model). This matrix (the 'cement' or 'mortar') is composed of the contents of lamellar bodies (secretory organelles) that have fused with the plasma membrane and consequently released their contents in the extracellular space, as well as lipids that exist in the space between cells (Brandner et al., 2015). The lipid matrix is composed mainly by a heterogeneous mixture of free fatty acids, cholesterol and saturated long-chain ceramides in an approximate 1:1:1 molar ratio (Wertz and Norlén, 2004). Recently, a study revealed more about the molecular organisation of SC where the molecular structure and function of the human skin's permeability barrier was analysed using MD simulations validated against cryo-electron microscopy data (Lundborg et al., 2018). Such findings could improve transdermal strategies for delivering drugs and vaccines to the body (Notman and Anwar, 2013).

1.4.4 Epidermal barrier and the importance of TJs

The epidermal permeability barrier is crucial not only for blocking substances going in our organism, but for keeping nutrients and water in our organism as well. This two-way consideration led to the discrimination between the inside-out and outside-in barrier. In the past, the main physical barrier was considered to be the one posed by the outermost layer of the epidermis, the SC, and specifically the lipids present there. However, this consideration fails to justify phenomena such as the tremendous water loss that patients face after burn injury which is more severe than in cases where only the SC is lost (e.g. by tape stripping) (Brandner, 2009). Moreover, loss-of-function and

gain-of-function studies support the functional role of other components in the epidermal permeability barrier, such as the TJs (Turksen and Troy, 2010).

The presence of TJs in the epidermis layer was a subject of discussion for decades although in 1971 the existence of TJ-like structures with EM was observed and a tracer was stopped at the SG layer of the epidermis (Hashimoto, 1971). In general, it was thought that typical TJ do not exist in mammalian epidermis although they were found in lower vertebrates.

However, with the advances in immunoelectron microscopy and the identification of the numerous proteins located in TJ, scientists were able to locate characteristic TJ-like structures in the SG of human epidermis (Brandner et al., 2002). Specifically, functional TJs were localised in the second layer of the SG of human epidermis which has three SG cell layers (Yoshida et al., 2013). These findings together with the localisation of critical TJ proteins (such as claudin-1, occludin and ZO-1) at the epidermis layers provided a strong indication that zonula occludens exist in epidermis.

It is well established that TJs control the paracellular diffusion of ions, molecules and even inflammatory cells in simple epithelial and endothelial cells (Kirschner et al., 2010). In the case of epidermis (a stratified epithelium) the role of TJ in the inside-out barrier is also well established. This is because there have been assays that show how a tracer is blocked in the granular layer of human epidermis (Hashimoto, 1971). Furthermore, mice that are claudin-1 deficient suffer from extensive water loss and die shortly after birth due to dehydration (Furuse et al., 2002). Whichever is the explanation of the dehydration (leaky TJ is the cause for the water diffusing out of the body or is it an effect that alters the composition of the SC) the role of TJs is important.

It is natural to assume that the impairment of the SC barrier is closely related to the TJ-deficiency which leads to an altered SC composition. This is because in the 'bricks and mortar' model the corneocytes – terminal differentiated keratinocytes–are the bricks and the mortar is the lipid lammellae that fill the gaps between the cells. TJs play a role in sealing the extracellular space between cells therefore function as an effective barrier that could potentially alter the SC's composition (Morita and Miyachi, 2003).

To conclude, the epidermal permeability barrier is a complex barrier system that comprises of not only the outermost layer of the epidermis, i.e. the SC, which is of great

importance. But it composes of the TJs, the skin microbiome, the chemical barrier and the immunological barrier which act together and not separately (Bäsler et al., 2016).

1.4.5 The diversity in lipid composition of plasma membranes

The cell has a 'wall' primarily to protect its interior from the outside ever changing environment. However, this wall is selectively permeable because an interaction between the two compartments (cytoplasm and extracellular region) is essential. The plasma membrane or cell membrane as it is called, is a lipid bilayer in which many proteins are either embedded or peripheral. It is mainly composed of phospholipids, but its composition differs not only amongst different organisms and cellular types but also between the outer and inner leaflet of the plasma membrane of a specific cell (Bennett and Tieleman, 2013). Also, there have been reports of specific areas called lipid rafts that are microdomains which are detergent-insoluble, sphingolipid and cholesterol enriched areas and act as a platform for proteins to exert their roles (e.g. signalling and trafficking) (Lee et al., 2008). TJs are believed to be assemblies of such microdomains (Lee et al., 2008).

An important and widely studied type of lipid is cholesterol, which is also a sterol, and can be conceived as a molecule devoid of a bulky 'head' or 'tail' so its structure resembles a cylinder. It is a highly unpolar- hydrophobic molecule, so when cholesterol is incorporated into phospholipid membranes it alters its properties, namely its thickness, rigidity, fluidity, curvature and compressibility leading most of the times to a more stiff –rigid structure (Krause and Regen, 2014). The latter phenomenon however, strongly depends on the bilayers initial composition. A well-known effect induced by the insertion of cholesterol in bilayers is the condensing effect i.e. the thickening of phospholipid bilayers. Cholesterol tends to accumulate in the hydrophobic area (lipid tails) of the plasma membrane increasing the organisation of the hydrocarbon chains. This results in a more perpendicular to the plane of the bilayer orientation of the chains. The distance between the phosphate groups of the phospholipids is also increased. Consequently, the lipid bilayer becomes thicker due to the increased organisation that cholesterol induces. Another important property is that cholesterol alters the secondary structure of the embedded proteins and the depth that they can reach. It also helps the organisation of the lipid microdomains, modulates the activity of the associated proteins and alters the physical properties of plasma membranes. Finally, cholesterol is

considered an important factor in many biological phenomena that lipid membranes participate (e.g. endocytosis and exocytosis)(Quinn, 2010).

1.4.6 Lipids found in skin epidermis

It is difficult to examine the composition of the cell membrane lipids at the different epidermal layers because their composition changes and specifically the transition of SG to SC results in abrupt changes in the lipid composition of the cell membranes. The SG is the last layer of viable epidermis and thus, has living cells in contrast to SC where cells are terminally differentiated and are flattened without nucleus and cytoplasmic organelles (corneocytes). The SG cells are expected to have a much more complex lipid membrane composition compared to corneocytes. The ratio of lipids found in SC has been extensively studied due to its importance for the epidermal barrier, and it is roughly a 1:1:1 molar ratio of free fatty acids : cholesterol: ceramides which are believed to form stacked bilayers (Lundborg et al., 2018). However, the present lipid analysis technology does not allow for localised, organelle specific determination of lipid composition and moreover, a good quantitative lipid analysis tool is missing. This is a problem when considering the composition and structure of the cell membranes in SG where TJs have been found.

However, some guidelines of the changes in the lipid composition during the differentiation process exist, and this helps the attempt to model the lipid composition of the other epidermis strata. Additionally, there are some experimental studies on human and other mammalian skin with regards to their lipid composition that examine and assess the total lipid concentration and the distribution of all major lipid species (Elias et al., 1979, Lampe et al., 1983). These studies give an overview of the distribution of lipid species in different epidermal layers and help the attempt to elucidate each individual layer's lipid composition of the cell membrane.

Because we have an interest in the SG layer, we focus on the composition of lipids found there. Biochemical studies have demonstrated that in mammalian SG neutral lipids (free sterols, free fatty acids, triglycerides, wax esters, and n-alkanes), polar lipids - phospholipids (phosphatidylethanolamine, phosphatidylcholine, phosphatidylserine and sphingomyelin) and glycosphingolipids are found (Lampe et al., 1983). Phospholipids have two hydrophobic fatty acid chains -'tails' and a choline group as

the ‘hydrophilic head’. Those two components are joined together by a glycerol moiety (see Figure 1.8).

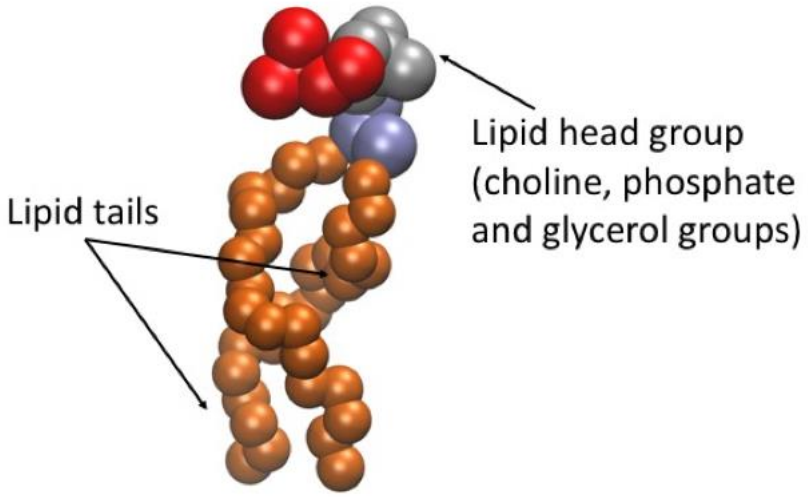


Figure 1.8: Representation of a DOPC molecule as vdW spheres visualised with VMD (Humphrey et al., 1996). The head group and lipid tails are also displayed.

In SG, the predominant lipid in the polar fraction is oleic acid (C18:1), which is an unsaturated fatty acid with 18 carbon atoms and a double bond. The 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) is a phospholipid that has two oleic acids as its hydrophobic tails and has a phase transition temperature equal to -17 ^{0}C, which is an important factor to consider because it influences the properties of a bilayer (see Figure 1.8). Cholesterol sulfate (the sulfate ester of cholesterol) is present in substantial amounts in all epidermal layers, and specifically in higher percentage in SG (5.5%). Lastly, all the epidermal layers contain sphingolipids which are a mixture of ceramides and glycosphingolipids. In the SG sphingolipids hold approximately 11% of the lipid composition and are found mostly as ceramides I (5%).

To summarise, although we do not know the exact lipid composition of the cell membrane in all the epidermis layers, there are guidelines of how the different lipid species change during the terminal differentiation process.

1.5 Critical Review of Recent Literature and Outstanding Research Questions

In this section, we conduct a critical review of recent research advancements, looking at the broader and specific issues and identifying the outstanding research questions that naturally lead onto the aims and objectives of the thesis. Please note that the review is not too detailed, as we elaborate on the specific literature in depth in the individual results chapters, being guided by the specific research question being addressed.

Freeze-fracture replica EM shows the hydrophobic plane of membranes, where TJs have been observed as particles on the protoplasmic face (P-face, inner leaflet viewed from the outside) and complementary grooves on the exoplasmic face (E-face, outer leaflet viewed from inside) (Tsukita et al., 2001). Claudin-1 forms continuous strands largely associated with the P-face, while claudin-2 induces discontinuous strands at the P-face with complementary grooves at the E-face, which are found to be chains of particles (Furuse et al., 1998). Therefore, the morphology of the observed strands is closely related to the proteins expressed. The strands, approximately 10 nm in diameter (Anderson, 2001), were more recently viewed as having a double-stranded morphology (width of 6.9 ± 0.8nm) (Krystofiak et al., 2019). However, there are still difficulties in dynamically visualising these structures in high resolution (Gonschior et al., 2020).

The antiparallel double-row putative model for the architecture of TJs (see Figure 1.2) was proposed by Suzuki et al., after the group resolved the first crystal structure of a claudin (Figure 1.3) (Suzuki et al., 2015, Suzuki et al., 2014) and there are now more crystal structures determined that provide clues about the cis- and trans-interaction sites (Saitoh et al., 2015, Nakamura et al., 2019, Shinoda et al., 2016, Vecchio and Stroud, 2019). The so-called Suzuki model is characterised by two cis-interfaces (see Chapter 3 and 4 for details), but validity of the model still needs to be examined since the possibility of multiple cis-interfaces has been demonstrated by other studies (Zhao et al., 2018, Irudayanathan et al., 2015, Irudayanathan et al., 2017, Irudayanathan et al., 2018). This possibility for claudins is also supported by a cryo-EM study of a protein that belongs to the same bigger family, namely Euglena IP39 (Suzuki et al., 2013). Also, the Suzuki model does not have transmembrane contacts that have been proposed for claudins-3 and -5 (Rossa et al., 2014) and claudin-10 (Milatz et al., 2017). Hence, the question of how specific claudins interact side-by-side and form the TJ strand with

varying density and branching is still unresolved. It is possible that the formed interfaces are numerous, but the highest affinity ones would dominate (defined functional states). A few claudins have been studied and the results to date stress the need to examine each individual member, since it is noted that even when they have significant sequence and structure similarities they do not form the same (cis-)interfaces (Irudayanathan et al., 2017, Irudayanathan et al., 2018). The experimental findings for each claudin cannot be explained by the Suzuki model and therefore further studies are required to identify the dimeric configurations of each claudin subtype.

The strands from neighbouring cells meet and form barriers/pores that are believed to function as the 'gatekeepers' allowing in some tissues specific small molecules to go through the paracellular route (Anderson and Van Itallie, 2009). Claude hypothesised that the number of the strands may be related to the tightness of the junction, but this hypothesis is not considered valid (Claude, 1978). Stevenson et al., for example, demonstrated that two different type of cells, although having the same number of strands, show different trans-epithelial resistance values (Stevenson et al., 1988). It is suggested that it is also a matter of quality (tightness) meaning how compact the seal is and not only the total number of strands. Also, in some tissues there should be redundancy in the barrier where more than one strands are required (multiple strands). The overall tightness of the barrier determines the electrical resistance across the epithelium. Additionally, it appears that the strands might undergo a kind of maturation process because images of TJs typically reveal "one continuous apical-most strand, variably cross-linked medial strands and looser, less well organised and sometimes discontinuous basal strands" (Van Itallie and Anderson, 2014). The strands are also cross-linked in varying degrees in different tissues, hence the question of what regulates the density, the continuity and the branching of strands needs to be addressed. Furthermore, the strands appear to be dynamic, forming and remodelling even in steady state (Weber, 2012). The possibility of different states (open or closed pores) has also been proposed but we cannot directly measure this (Weber, 2012). The TJ strands need to be flexible structures to overcome the stresses and strains of osmotic pressure from the different sides of the cell membrane but also during different cell processes (e.g. endocytosis, cell division).

Claudins are categorised into barrier and channel forming (Günzel and Alan, 2013) and only claudin-2 has been found to allow water molecules through the paracellular route (Rosenthal et al., 2010). Expression of claudin-2 in cultured cells increased the permeability for solutes smaller than 4 Å (Van Itallie et al., 2008). The pore structure has been investigated for claudin-15 (Alberini et al., 2017, Alberini et al., 2018, Samanta et al., 2018) for claudins important for the blood brain barrier (Irudayanathan et al., 2017) and other classic claudins with in silico studies (Irudayanathan et al., 2018). The selectivity filter in the case of claudin-15 is due to four negatively charged aspartic acids (D55), and it was demonstrated that water molecules and small ions can pass through the pore (Samanta et al., 2018) and that the passage of chloride is almost 30 times slower of that of sodium (Alberini et al., 2018). The charge selectivity that some channels show has been investigated with charge-reversing mutations, and it has been demonstrated that the charged residues of the first extracellular loop are responsible for this selectivity (Günzel and Alan, 2013, Weber, 2012, Angelow and Alan, 2009b) and specifically of the C-terminal half of ECL1 for claudins -4 and-15. It is also believed that there are two pathways: a pore and leak pathway that a solute can follow making the attempt to elucidate the channel function even more complicated (Shen et al., 2011). The combination of claudin types is expected to determine the barrier/channel property of TJs.

More than two distinct claudins are co-expressed in single cells and it is believed that they can interact in a homotypic (same subtype) or heterotypic (different subtypes) way to form homotypic or heterotypic strands. Claudin-1, which is indispensable for the epidermal permeability barrier and of interest here (Furuse et al., 2002), was found to interact with claudin-3, but not with claudin-2 strands (Furuse et al., 1999). This begs the question of claudin compatibility, and since there are at least 27 members identified in mammals (Mineta et al., 2011) which are the members that form heteromeric strands? There is a lack of information with regards to specific motifs that control heterotypic binding of claudins and it is believed that small changes in specific amino acids of the ECD control this compatibility (Daugherty et al., 2007). This combination is expected to regulate the TJ pore structure and optimise organ-specific strands that control the perm-selectivity of a given tissue.

Therefore, it is concluded that claudins contribute to fibril formation and barrier function but how exactly they do this remains to be resolved for each claudin subtype and their combinations. For example, a peptide mimicking the second half of ECL1 (residues 53-80) of claudin-1, consisting of the area between the second transmembrane helix and β3-β4 strands (see Figure 3.1), was found to reversibly interfere with the structure and function of TJs both in cultured cells and in vivo but whether it interferes with the cis- or trans- interfaces was not established (Mrsny et al., 2008). Also, Milatz et al., found that the cytoplasmic ends of the first transmembrane helices are closer than the average distance of the corresponding fourth helices for claudin-1/claudin-3 heteropolymers (Milatz et al., 2015). For claudin-5, it was demonstrated that ECL2 is important for trans-interaction but not cis-polymerisation, highlighting the aromatic residues PHE147, TYR148 and TYR158 as well as the hydrophilic residues GLN156 and GLU159 (Piontek et al., 2008). These studies provide invaluable information but more studies are needed to provide compelling evidence of the 'key' amino acids for each protein.

Another important area is the strength and kinetics of the adhesion between the proteins of the cell-cell junctions. For example, do cadherins from adherens junction bind stronger than claudins? Using single molecule force spectroscopy, Lim et al., demonstrated that homophilic claudin-1 interactions form weak and short-lived interactions that have a dissociation rate 100-fold greater than that of E-cadherin (Lim et al., 2008b). Moreover, the ECL1 of claudin-2 was found sufficient for trans-interaction and that the dissociation follows a two-step energy barrier model (Lim et al., 2008a). Note that N-cadherin and E-cadherin molecules show different kinetics and binding energies (Panorchan et al., 2006). Cadherins are also known to bind in multiple conformations with binding intermediate structures (Manibog et al., 2014). It appears that both junctions provide cell-cell adhesion, but the dynamic nature of TJs requires them to be more dynamic by nature, maintaining at the same time their barrier/pore structure. It is possible that claudins also form (trans-) intermediate adhesion structures that we still do not know about.

TJs have received attention from the research community due to their implication in hereditary diseases, inflammation, cancer and also because several TJ proteins are targets of viruses, bacteria and other pathogens (Brandner et al., 2015, Förster, 2008,

Krug et al., 2014, Sawada, 2013). For example, the crystal structures of claudins -3, -4 and -19 in complex with the C-terminal of Clostridium Perfringens enterotoxin (see Figure 1.4, Figure 1.5 and Figure 1.6) were resolved and show that the extracellular domain (ECD) undergoes conformational changes (Shinoda et al., 2016, Saitoh et al., 2015, Nakamura et al., 2019). Earlier the toxin was believed to bind only to the second ECL, but it is now recognised that the whole ECD is involved in the binding, with the toxin fragment covering the entire ECD region, forming extensive hydrophobic contacts. Also the middle strands (β1 and β2) take part in the binding interface. In the case of claudin-3 the toxin was more distant from the β-sheet domain and mostly interacted through the variable regions (V1 and V2). Furthermore, a trimeric toxin unit was suggested to interact with trimers of claudin in an in silico study (Irudayanathan et al., 2018). It is suggested that the kink of the longer third transmembrane region can provide clues about the CPE-sensitive claudin subtypes (Suzuki et al., 2017). However, it is worthy to note that these claudins are not in the TJs. Additionally, mutations of small/hydrophobic residues of the transmembrane regions of claudin-14 and -16 are reported to cause inherited human diseases (Suzuki et al., 2017) and there are seven human diseases known to be caused by mutations of TJ proteins (Anderson and Van Itallie, 2009). Furthermore, the possibility of reversibly disrupting the TJ barrier to administrate drugs has been examined (Deli, 2009, Hashimoto et al., 2016, Tscheik et al., 2013). All these factors point to the need to obtain a molecular level understanding of this complex structure.

To summarise, claudins contribute to TJ strand formation, but the important residues for cis- and trans- interaction of each claudin and their tissue-specific combinations still need to be established. The distinct amino acid sequence of the ECDs in different claudins contribute not only to the formation of the strands and head-to-head (trans-) interaction but also to ion-selective channels with various sizes. The different lipid environment might also play a role in the self-assembly of claudins changing their hydrophobic depth and tilt angle, thus promoting different TJ strand morphologies. The role of lipids remains to be elucidated because to the best of our knowledge only a couple in silico studies examined the effect of different lipid environments on the self-assembly of claudins (Irudayanathan et al., 2015, Rajagopal et al., 2019). Moreover, how and whether post-translational modifications affect the organisation of claudins still remains to be answered (Rajagopal et al., 2019).

1.6 Aims of the Thesis

The aim of the thesis is to elucidate the molecular architecture and mechanism of assembly of TJ proteins and to characterise their role in regulating the permeation of solutes through the intercellular space using molecular simulations.

Objectives include, to:

1. Investigate the *cis*-interactions (within the same membrane) between the extracellular domain particles of claudins using an *implied* lipid membrane. The ECDs are restrained on a 2-d plane to mimic the placement of claudins in a lipid bilayer. The loops are free to move on the plane. The *implied* bilayer reduces the complexity of the system.
2. Examine the interactions between TJ proteins of neighbouring cells – the *trans*-interaction. It is the *trans*-interaction between the opposing loop regions that serves as the barrier and regulates the transport of solutes through the intercellular-space.
3. Explore the assembly of the whole claudin molecules embedded in explicit lipid bilayers of different lipid composition and characterise the morphology of the resulting TJ strand. To access the longer timescales required to observe this phenomenon, we employ coarse-grained (CG) simulations. We start by examining the properties of the lipids and then evaluate the stability of a single protein whilst embedded in the bilayers. Furthermore, we employ advanced methods to enhance sampling and increase the probability of observing the equilibrium configurations (i.e. metadynamics).
4. Study the effects of point mutations on the structure of a single claudin by employing replica exchange and metadynamics simulations. The protein is embedded in a pure lipid bilayer, and we focus on how much the structure changes and which areas are affected the most. We also note the differences in the achieved results from the two different enhanced sampling methods.

1.7 Thesis Outline

In this chapter, we have stated the research problem and objectives of this thesis, as well as reported on previous research in the field of TJs. There was a focus on the structure and function of TJs with a view to understanding their molecular architecture and barrier

function. TJs are found in our skin's epidermis, and therefore, the different layers of skin were presented as well as the factors that compose the epidermal permeability barrier. The next chapter contains an overview of molecular simulations, force field methods, enhanced sampling methodologies and free energy calculation methods, which have been employed in the thesis.

In Chapters 3 and 4, an emphasis is given on the benefits of atomistic representation in computer simulations. This relates to the studies since we present atomistic MD simulations of the extracellular domain of the TJ protein claudin-1. A specific part of the protein, its extracellular domain, has been isolated because we hypothesise that this is the key determinant of the self-assembly and barrier structure of the TJs. This enabled the study of the protein's behaviour in atomistic representation, thus, in detail. The critical residues that define the cis- and trans-interaction between the domains are highlighted and characterised. To the best of our knowledge, these are the first studies that present in atomistic detail the interactions between claudins in large systems.

Claudins are tetraspan proteins which means that they have four transmembrane regions (see Figure 3.1). Although their extracellular domain is relatively large and is expected to dominate their self-assembly, it might be the case that the transmembrane region stabilises their aggregation. Thus, in Chapter 5 simulations of systems composed of the whole protein while embedded in lipid bilayers are presented. Next, CG models were used to characterise the aggregates of claudin-1 and compare them to the atomistic ones found in Chapter 3. Furthermore, we analysed the properties of the bilayers and compared them to experimental values, as well as characterised the relative position of claudin-1 whilst embedded in the various model bilayers. The morphology of the resulting TJ strand is also described.

In Chapter 6, we focus on describing the effects that mutations have on the structure and function of claudin-1. This is reviewed using two different methods, namely metadynamics and replica exchange simulations. There is also a discussion on the possible impact the mutation has on the self-assembly of claudin-1. The results are compared to the ones calculated for the wild-type protein and the agreement between the two previously mentioned methods is discussed.

Lastly, in Chapter 7 the conclusions from all the studies and their significance are presented. The limitations of the studies are discussed as well as suggestions for future work on the research area.

2 Methodology: Molecular Simulation

2.1 Molecular Mechanics (Force Field Methods)

Molecular dynamics (MD) simulations have evolved into a valuable tool and a mature technique, which can effectively provide answers about the structure, function and mechanism which underlie the physical phenomena. MD simulations generate trajectories of molecules based on the forces that result from interactions between atoms. The forces can be calculated using quantum mechanics (first principles MD) or for large systems using the molecular mechanics approximation. In the molecular mechanics approximation, atoms are represented as spheres and bonds as springs and the evolution of the system is based on Newtonian mechanics.

MD simulations can be carried out on simple or complex systems (e.g. large proteins) challenging computational resources. For small systems and for short time scales (of the order of a few tens of picoseconds) quantum mechanics methods can be employed. For larger systems, comprising tens of thousands of atoms up to tens of millions of atoms or more and for time scales extending to microseconds, one has to resort to classical MD simulations employing the molecular mechanics approximation. If we want to move to even longer time and length scales the description of the system needs to be further simplified. A way to do so is by using mesoscopic or coarse-grained (CG) methods where a group of atoms are represented by a single CG particle (Hug, 2013). Thus, depending on the size of the system and the phenomena of interest one wants to observe, the appropriate method needs to be selected after considering the benefits of the offered resolution while balancing the computational cost (Hospital et al., 2015).

An atomistic MD simulation represents explicitly all the atoms of the system and offers greater reproduction of the actual system and its properties. However, the computational cost can be large depending on the system's size. CG representations have become very popular when large systems or long simulations are required to access biological

relevant timescales. CG is a simplified representation of the physical system that uses mapping, where usually four atoms are mapped to a single entity, a bead (4 to 1 mapping), and this enables the study of bigger systems. This is because in a classical molecular mechanics approach, the time step of the method depends upon the motion that depicts the highest frequency usually bond vibrations and angle bending. Since hydrogens and other light particles or groups of atoms are not explicitly represented in a CG model, their (fast) dynamics do not need to be captured and thus the time step can be further increased.

Classical mechanics cannot describe very light particles such as electrons while nuclei are sufficiently heavy and can be described. Additionally, the difference in their mass indicates that electrons, which move faster, can easily adjust to the nuclei changes (Jensen, 2007). In force field methods the electronic distribution is ignored and the energy calculation is based on the positions of the nuclei. Therefore, the Born-Oppenheimer approximation is accepted, which states that the motions between the nuclei and electrons can be separated because of the different time scales of their motion. Thus, it is clear that molecular mechanics cannot provide properties of a system that depends upon the electronic distribution in a molecule such as the formation of a bond (Leach, 2001).

It is worth mentioning that apart from MD simulations, there is another type of simulations called Monte Carlo simulations that generate random configurations. These configurations depend only on their predecessor and the algorithm uses criteria to decide upon the acceptance of every new configuration with the advantage of 'visiting' higher energy states (Leach, 2001). Monte Carlo simulations aim to generate an ensemble (group) of representative configurations under specific conditions which can be considered as a collection of points, thus an ensemble (Paquet and Viktor, 2015). But Monte Carlo simulations do not provide information about the time evolution of the examined system as MD do. Another difference between MD and Monte Carlo is that while the total energy of a system has two contributions, specifically the kinetic and potential energy, in the case of the Monte Carlo method the kinetic energy distribution is not considered.

2.2 Statistical Mechanics

The link between the MD trajectories and the macroscopic world is the domain of statistical mechanics. Statistical mechanics provides a connection between the microscopic properties observed e.g. in an MD simulation such as velocities and coordinates, with the macroscopic quantities of a system such as temperature or pressure. Note that the averages of the macroscopic properties are of interest rather than the atomic information calculated. One way to obtain the average of a property, is to calculate its time average during an MD simulation. However, an important concept in statistical mechanics is the statistical ensemble, which is a collection of independent copies of a system in different states. This means that many identical systems are in the same macrostate but each one is in a different microstate (statistical ensemble). Thus, instead of analysing the behaviour of a single system, in statistical mechanics we analyse an ensemble of states based on a probability distribution over all the possible states of a system. The averages calculated from the two approaches are connected with the ergodic hypothesis (more details in 2.2.2).

2.2.1 Phase space

The physical system is described by a set of the positions and velocities of all atoms at any moment in time. So, the concept of phase space is introduced where in order to define a system containing N atoms, 6N values are required, specifically 3N for the positions and 3N for the momenta. A point in phase space is unique being defined by a set of atomic coordinates and positions at a given time. All possible states of a system are represented in phase space. Note that the dynamics of the system yields a 'trajectory' in phase space.

For a closed system its total energy is the sum of the potential energy due to its position and its kinetic energy due its motion. This is also referred to as the Hamiltonian *H(p,q)* of the system and it is expressed by:

$$H(\boldsymbol{p}, \boldsymbol{q}) = T + V \qquad (1)$$

where p is the momentum, q the coordinates of an atom and T and V the kinetic and potential energy of the system, respectively. The kinetic energy is a sum of contributions of all the individual particles momenta while the potential energy is a sum of all pairs,

triplets etc, and determines the complexity of the function that describes the total energy of a system.

2.2.2 Ergodic hypothesis

An important concept in MD simulations is the ergodicity of a system. We can calculate time averages from an MD simulation and a system is ergodic when the two averages, namely the ensemble average and the time average, are equal. For instance, imagine a system evolving in time. While the system evolves in time the value of the property changes and one can assume that its value (measured experimentally) is really the time average of the property over a long time-interval. However, the ensemble average is calculated over many replicas of a system at the same macrostate. Because of the complexity of the time evolution of the property for many atoms, one can replace the time average with the ensemble average (Leach, 2001). Hence, in order to calculate the 'real' average value of a property, we have to replace many replicas of the system that are considered simultaneously (statistical mechanics) with a single system evolving in time and so the ensemble average is equivalent to the time average. If during an MD the system gets ‘locked’ in the same place of phase space, it does not sample all states, therefore it doesn’t converge to equilibrium and the ergodic hypothesis does not stand.

2.2.3 Common statistical ensembles

MD simulations are usually performed under specific conditions referred to as statistical ensembles (these ensembles are the macroscopic constraints on the system). For example, a closed and isolated system would exist at a fixed volume, number of particles and energy (called the microcanonical ensemble, constant NVE) (Leach, 2001). If we want to reproduce or compare the calculated averages from the simulation to experimental data, apart from the constant number of particles both the temperature and pressure of the system must be constant and preserved which gives the isothermal-isobaric ensemble (NPT). The Monte Carlo method usually samples from the canonical ensemble in which the number of particles, the temperature and the volume of the system is kept constant (NVT). Another known ensemble is the grand canonical where the number of particles can change while the chemical potential is kept constant. Note however that both simulation methods can be modified in order to sample from the desired ensemble with the use of barostats and thermostats.

Each statistical ensemble (e.g. NPT or NVT) can be represented by a partition function that can be related to thermodynamic properties. The partition function describes the statistical properties of a system because it is a function of temperature and the microstate energies. It is an important quantity in statistical mechanics, because it provides the base to connect the microstate energies, which are determined by other thermodynamic variables, with other desired thermodynamic properties of a system (e.g. heat capacities). The partition function Q, gives the total number of accessible energy states at a given temperature and it is small when there are few available states. The fraction of molecules with a particular energy depends on the number of available states (density of states) and the probability of a state being occupied.

To summarise, simulations are performed under statistical ensembles that have a specific partition function to connect the microscopic properties observed during the simulations, with the macroscopic quantities of interest.

2.3 General Form of Potential Energy Function in Molecular Mechanics

In MD simulations, the evolution of a system can be predicted by calculating the forces acting on the atoms based on the potential energy function (the force that acts on each atom is the derivative of the energy with respect to coordinates). Force fields are complex equations that describe the potential energy of the system based on the molecular structure and characterise the interactions between atoms. The potential energy function in its most general form describes the intra- and inter- molecular forces and can be divided in at least three terms, as described by Equation 2:

$$V = V_{bonded} + V_{non-bonded} + V_{special} \quad (2)$$

where V_{bonded} describes the bonded potential terms, $V_{non-bonded}$ the non-bonded potential terms and $V_{special}$ an optional special term that represents the imposed bias to the behaviour of the system so that it will perform in a particular way.

The bonded interaction energy function is a sum of terms that describe bond stretching, the angle bending, the proper and improper dihedrals between atoms (see Equation 3) (Leach, 2001).

$$V_{bonded} = V_{bonds} + V_{angles} + V_{improper} + V_{proper} \quad (3)$$

The first two terms are represented as harmonic potentials and even when they are described by Hooke's law they perform relatively well compared to more complex forms. Bond rotations (torsions) are periodic so the associated potential energy term has a periodic form (cosine) (Cramer, 2013). In most cases, a term that gives the potential energy of improper dihedrals is added, and this refers to four atoms that are not successively bonded. This is a way to model planar structures such as the peptide bond in the amino acid side chain or to model a specific flat geometry (e.g. aromatic rings).

The non-bonded term refers to all 1-4 and above interactions, which considers atoms that are not connected via a bond or involved in an angle comprising 3 atoms or a dihedral angle. It takes into account the 'through space' interactions which are often modelled as a function of an inverse power of the distance between the atoms (pair-wise sum) (Leach, 2001). This term is broken down to the Lennard-Jones (LJ) and coulombic contributions arising from van der Waals (vdW) and electrostatic interactions respectively as described by Equation 4:

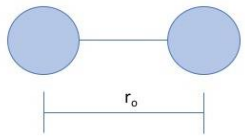
$$V_{non-bonded} = V_{LJ} + V_{coulomb} \quad (4)$$

If one calculates all the above mentioned terms, it is evident that force fields are complex mathematical descriptions of the interactions that occur in a system of atoms. Moreover, not only the functional form of a force field can differ but also its parameters (more details in the next section). Generally, the force fields are designed to reproduce specific properties and as a consequence are parametrised accordingly (Leach, 2001). In most cases, they are used to reproduce structural properties and they perform this with very good accuracy. It is important to emphasise, that force fields are empirical and this means that there is not a correct or incorrect force field. However, one force field can perform better than the other in terms of reproducing desired properties.

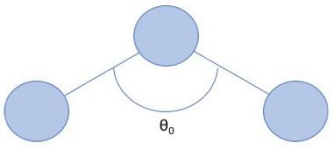
2.3.1 Force field terms

The bonded terms in the force field can be considered to comprise energy penalties, which contribute in varying degrees to the overall potential energy of a system. For

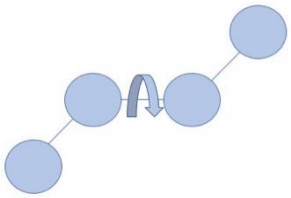
example, the bond length (r_i) should not deviate significantly from its reference value (r_0) (see Equation 5), so the penalty is high when atoms violate their equilibrium bond length. The associated force constant K_i in Hook's law that describes the stiffness of a bond, is usually several hundreds of kcal $mol^{-1}Å^{-2}$ and most of the time there is no need for the use of more complex functions, such as the Morse potential (Leach, 2001). Similarly, a harmonic potential function is also used for the term that describes angle bending in which case the energy penalty is even smaller. For example, the corresponding K_j value for the angle bending is less than 1 kcal $mol^{-1}deg^{-1}$ for some common C-C-C angles (Leach, 2001). Equation 6 describes the angle potential energy where the reference angle is θ_0, the angle between 3 atoms at time t is θ_i and K_j is the force constant.



$$V_{bonds} = \sum_{bonds} \frac{K_i}{2}(r_{i-} r_0)^2 \quad (5)$$



$$V_{angles} = \sum_{angles} \frac{K_j}{2}(\theta_i - \theta_0)^2 \quad (6)$$



$$V_{torsions} = \sum \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) \quad (7)$$

It should be noted that the accuracy of the force field could be improved with the incorporation of higher order terms to both components but the associated computational cost is high. The energy penalty for bond stretching and angle bending

is significant, while the corresponding energy penalty for the torsional term is relatively lower (see Equations 5, 6 and 7).

The torsional potentials are usually expressed as a cosine series expansion or in an equivalent form (Equation 7). The value of V_n in Equation 7 gives an indication of the barriers associated with the rotation around a bond where ω is the torsion angle, n is the multiplicity which gives the number of minimum points in the function at a full 360 degrees rotation, and γ defines where ω passes through its minimum value. Lastly, in many force fields, terms such as out-of-plane bending terms are used to represent a desired geometry, such as to ensure the planarity of an aromatic ring. This out-of-plane motion is usually added as an improper dihedral term where the four atoms are not bonded in the sequence 1-2-3-4.

The non-bonded interactions are the sum of van der Waals (Lennard-Jones, LJ) and electrostatic interactions as mentioned above (see Equation 4). In the case where two non-bonded atoms with no permanent charge approach each other because their electronic wave functions are correlated they interact. The associated force is referred to as London dispersion force or vdW and it is a dipole- induced dipole interaction (Cramer, 2013). The associated potential energy term in this case is usually proportional to the inverse sixth power of the distance between the atoms and it is increasingly negatively when they approach each other while it is negligible when they are far away (Cramer, 2013). As the atoms continue to come closer, their electronic densities begin to interpenetrate and so a repulsive force is generated in the case where a bond cannot be formed. In this case, the potential energy function is proportional to r^{-12}, where r is the distance between the two atoms. The LJ potential is a functional form that is used in force field methods to model the non-bonded interactions and a common form is given by Equation 8:

$$V_{LJ} = \sum_{i}^{N} \sum_{j \neq i}^{N} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \quad (8)$$

where r_{ij} is the distance between the atoms. The LJ potential has a strongly repulsive (r^{-12}) potential at short distances (r), an attractive term at large distances (proportional

to r^{-6}) and is zero when the distance between the atoms (r_{ij}) is sigma (σ) (see Equation 8 and Figure 2.1).

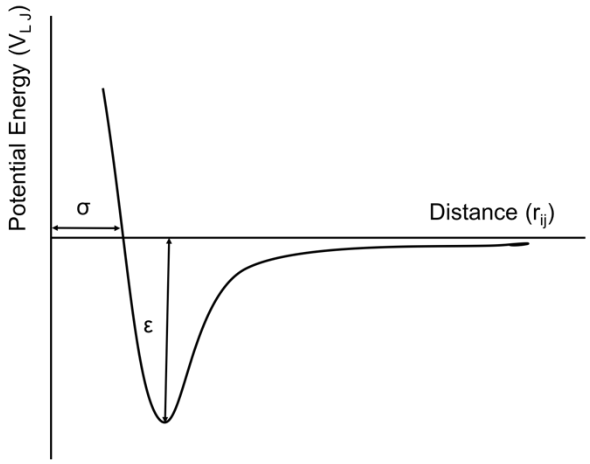


Figure 2.1: The Lennard-Jones potential energy which shows how the potential energy of two non-bonded atoms changes with respect to their distance. The epsilon (ε) and sigma (σ) values are also displayed.

The LJ potential energy function has two parameters, namely sigma (σ) and epsilon (ε) (see Figure 2.1). Sigma σ defines a length scale (sum of the van der Waals radii) while ε is the well depth and controls the strength of the interaction. Note that there are force fields where the repulsion term of the LJ potential is replaced by either another exponential expression such as the Buckingham potential, or the force field has a sum of small to medium size more complicated potentials such as the Morse or the Hill potentials (Leach, 2001, Cramer, 2013).

When the non-bonded atoms have 'permanent' non zero electrical moments their interaction is usually modelled by applying a partial charge to every atom and using Coulomb's law described in Equation 9:

$$V_{coulomb} = \sum_{i} \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \frac{1}{4\pi\ \varepsilon_0} \tag{9}$$

where q_i and q_j are the partial charges of atoms i and j respectively, rij is the distance between them and ε_0 is the permittivity of free space.

It is worth noting that atoms in reality are polarisable, so they do not have static moments as their moments are influenced from the presence of others or an external electric field. In some cases, it is beneficial to take this property into account when parametrising a force field but at a high computational cost.

In MD simulations, atoms are represented as spheres of specific mass but the representation of the distribution of charge can vary. It is common to use fractional point charges. These partial atomic charges do not only represent the atomic number of the atom but also contain information about its environment and state and are usually located at the nucleus centre (Leach, 2001). Sometimes, the electrostatic properties of a molecule are better described when point charges are placed along a bond. There are also other approaches used to represent the unequal distribution of charge in a molecule, such as the central multipole expansion which is based upon multipoles but they make the description of a system more complicated (Leach, 2001).

Cross-terms can also be added to force fields and describe the coupling between some internal variables (such as bond lengths and angles) with the aim to improve the ability of the force field to predict the system's properties (usually structural properties). For example, to describe how adjacent bonds stretch in order to accommodate the fact that a bond angle has decreased. One important cross term is the Urey-Bradley in CHARMM force field (Best et al., 2012). This term is referred to as an angle bending term between 1,3 non bonded atoms and it is modelled by a harmonic function of the distance between the outer atoms.

2.3.2 Parametrisation of force fields

The quality of each force field depends on the quality of the associated parameters in the potential energy function and the parameters that are optimised are inherent components of a force field. The methods used to generate the parameters are based on quantum mechanics methods as well as fitting to experimental data (e.g. geometries, vibrations, heat of vaporisation, free energies of aqueous solvation). The parameters are optimised in an iterative process until convergence is achieved. The main idea is to examine smaller model systems at the quantum mechanical level or experimentally and then transfer the parameters to the larger system, for example, break down the proteins into their building blocks, the peptides, and study them. Transferability (scale-up) is a

key property of force field parametrisation as well as accurate optimisation strategy. For example, in the general form of a potential energy function the constants as well as the properties at equilibrium, such as a bond length (r_0) or a reference angle (θ_0) are parameters that are optimised based on the aforementioned data.

To summarise, force fields are a description of the way atoms interact and they are differently parameterised which means that they use different functional forms for the potential energy as well as different parameters (e.g. k_b, k_θ, r_0). As previously stated, they restrain the system not to deviate from the reference value of bond lengths and angles (which are usually taken from experimental data e.g. crystal structures). Thus, most of the variations in structure and relative energies are derived from the terms referring to the torsions and non-bonded interactions and these are the parameters that mostly differ between various force fields (Leach, 2001). It must be pointed out that parameters such as those referred to the torsion angles, van der Waals and electrostatic interactions are not transferable from one force field to another and it is not advised to mix force field parameters (Ponder and Case, 2003). The most commonly used force fields for biological systems are CHARMM, GROMOS, AMBER and OPLS (Hug, 2013).

2.3.3 Classical mechanics: the time evolution of a system

Classical mechanics and quantum mechanics are the two major sub-fields of mechanics. The former considers the physical laws that describe the motion of particles under the influence of forces. In classical MD simulations, the behaviour of particles is predicted by solving Newton's equation of motion. The force that each atom 'feels' due to its environment is described by the derivative of the force field (potential energy function) with respect to atomic coordinates and by solving Newton's equation we can define the new atomic positions and velocities after a short period of time (time step). The time step is usually very small, typically 0.5-2 femtoseconds (order of 10^{-15} second) for atomistic systems in order to gain a more realistic potential (Leach, 2001). This allows one to observe the time evolution of a system, the so-called trajectory.

Newton's second law is described by Equation 10:

$$\vec{F} = \frac{d\boldsymbol{p}}{dt} = m \times \frac{d\boldsymbol{v}}{dt} = m \times \boldsymbol{a} \qquad (10)$$

and relates the force ($\vec{F}$) acting on a particle with its mass (m) and acceleration ($\vec{a}$) which is the rate of change of velocity with respect to time. Additionally, the force acting on an atom is the negative derivative of the potential energy function (V) and so the above equation can be rewritten to Equation 11 :

$$\frac{-\partial V}{\partial r} = m \times \frac{\partial^2 \boldsymbol{r}}{\partial t^2} \tag{11}$$

where the potential energy is related to the position of an atom, $\vec{r}$.

For an MD simulation, the initial configuration is required so that we know the starting positions of every atom. The velocities of the atoms are generated from a Boltzmann distribution based on the target temperature. Since the positions and velocities of the atoms are known, we combine them with the knowledge of the forces acting upon them and hence, we can deduce the new positions and velocities that the atoms will possess after a very short time (time step). By solving the same equations iteratively, we can predict the evolution of the system and characterise its properties. Thus, a trajectory of the particles can be generated with the hypothesis that the force is constant during the time interval. It is emphasised that MD yields trajectories, so there is a connection through time of the successive configurations of the system.

Additionally, it is important to choose an appropriate algorithm for the integration of the equations of motion (e.g. leap–frog, velocity Verlet) that will result in sampling from the correct ensemble (e.g. NPT). The equation of motion for a system composed of N particles is a set of 3N second order differential equations or 6N first order differential equations. In order to transform these equations to difference equations an integrator is chosen and it must possess some important properties such as accuracy and consistency but equally importantly it must be time-reversible and symplectic (the volume in phase space should be preserved) (Hug, 2013). The most famous second order algorithm used in MD simulations is the Verlet algorithm and some algorithms derived from it such as the velocity Verlet and the leap frog algorithm. Also, there are some algorithms that are called predictor- corrector algorithms that are more accurate but they are not time-reversible and symplectic.

2.3.4 Thermostats and barostats

As previously mentioned, MD and Monte Carlo simulations can sample from various statistical ensembles (e.g. NPT). In order to do so, the use of thermostats and barostats is implemented in simulations because the direct integration of Newton's equations of motion leads to the microcanonical ensemble (NVE). Thus, to perform MD simulations in constant temperature and/or constant pressure the use of local or global thermostats and barostats is essential (e.g. Berendsen, Nose-Hoover, Parrinello-Rahman). There are several different methods that can be implemented with advantages as well as disadvantages, but their main feature is that they alter the Newton's equation or the Hamilton's equations in order to sample from the desired ensemble.

To fix the temperature of the system a simple way is to scale the velocities of the atoms, since the temperature is related to the kinetic energy of the system (Woodcock, 1971). An alternative more rigorous way is to couple the system to an external 'heat bath' which provides or absorbs thermal energy from the system to maintain the desired temperature (Berendsen et al., 1984). The scaling of velocities at each step is such that the difference between the temperature of the system and the heat bath is proportional to the rate of temperature variation as described by Equation 12:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}\left(T_{bath} - T(t)\right) \tag{12}$$

where $T(t)$ is the temperature of the system at time t, T_{bath} is the reference temperature of the external bath and τ is a coupling constant, which is an important parameter because it defines whether the coupling is weak or strong. A problem that arises from the scaling is that the components tend to have different temperatures ('hot solvent, cold solute' phenomenon) and a suggested solution is to apply the coupling separately to these components. For example, in a simulation of a solvated protein we can have two coupling groups, the protein atoms and the solvent atoms separately.

The pressure of the system is calculated via the virial theorem, which relates the average over time of the kinetic energy of a system bound by a potential force, with the average over time of its potential energy (Leach, 2001). Pressure fluctuations are greater than energy or temperature fluctuations and these are linked to the change of the system's

volume as in macroscopic systems. Based on the isothermal compressibility of each substance the change in the volume can be large. The pressure could be controlled based on the rescaling of the system's volume analogous to the temperature coupling, or by coupling the system to a pressure 'bath', analogous to the 'heat bath'. The rate of change in pressure is given by Equation 13:

$$\frac{dp(t)}{dt} = \frac{1}{\tau_p}\left(p_{bath} - p(t)\right) \tag{13}$$

where (p_{bath}) is the pressure of the bath, *p(t)* the actual pressure of the system and τ_p is a coupling constant. Both in temperature and pressure coupling, we calculate a scaling factor to adjust the velocities and coordinates of the atoms to maintain the temperature and pressure respectively at the desired values.

2.4 Periodic Boundary Conditions

Most of the simulations use periodic boundary conditions (PBC) because an open system where particles are able to move freely does not maintain constant particle density while in the case of a closed boundaries one, the surface effects will dominate the physical behaviour of the particles (Hug, 2013). We can overcome these problems by choosing to apply PBC during a simulation.

When PBC are applied, the image of the central simulation box is replicated throughout space to form an infinite lattice (Allen and Tildesley, 1989) and so the central box is surrounded by 26 identical boxes (Jensen, 2007). There are no walls or surfaces at the boundaries and the particles density is kept constant because the periodic images of the particles will move in the same way as the particles in the central box. This means that if the trajectory of an atom takes it outside the box from one side, its image will simultaneously enter the box from the other side (lattice symmetry) (Cramer, 2013). Thus, it is sufficient to monitor the behaviour of the particles located in the central box. PBC apart from maintaining a constant particle number also help to conserve the systems linear but not the angular momentum. PBC aims to make the space isotropic and to model bulk phenomena especially when the system is large. Note that there are different types of simulation boxes that can be used during a computer simulation such as a rectangular box, a rhombic dodecahedron and a truncated octahedron (because they

are considered more efficient since they need fewer solvent molecules), but the cubic box due to its geometrical simplicity is a common choice (Allen and Tildesley, 1989).

2.4.1 Calculating interactions when PBC are applied

Because of the application of PBC, each particle in a box can interact not only with the other particles in the box but their images as well. To avoid these atoms seeing their images, we apply the so called minimum image convention where each atom interacts with its closest image (Allen, 2004). The minimum image convention uses a cut-off distance to approximate the short-range forces, so the interactions of non-bonded pairs only within this cut-off distance are calculated. As a result, a particle should interact only with the closest image of itself and neglect the rest.

However, using a cut-off distance does not necessarily save significant computational time because all distances must be calculated. Thus, neighbour lists over atom pairs within the cut-off distance (and a buffer size) can be prepared according to the starting geometry (Jensen, 2007). Therefore, during the simulation the interactions between the atoms in the list are regularly evaluated and so all distances do not have to be calculated. Neighbour lists are updated when the list changes.

While the energy of the vdW interactions can be negligible outside the cut-off distance (since they decay as r^{-6}) this does not apply for electrostatic interactions because they vary with the inverse power of the distance between the atoms (r^{-1}). In the case of long-range interactions between charged atoms, applying PBC in a simulation can result in artificial periodicity because a particle can interact with its image (Hug, 2013). These electrostatic interactions are modelled as a force depending on r^{-d} where d is the dimensionality of the system (Allen and Tildesley, 1989) and their range is often greater than half the box length (Leach, 2001). Particularly, the most important are the charge-charge interactions between ions and dipole- dipole interactions between molecules (Allen and Tildesley, 1989). One approach to solve this problem would be to increase the box length but this would be impractical and the computational effort great. Hence, different methods have been developed for treating electrostatic interactions.

The most common method used to treat long-range interactions is Ewald summation where the basic concept is to use two terms, one for the near (real space) and one for

the far (reciprocal space) contribution. For the former a cut-off distance is used while for the latter Fourier series are used to reproduce the charge distribution (the simulation boxes have the same distribution pattern).

2.4.2 Sampling phase space

Choosing the correct time step for a simulation is not trivial because if it is too small for a fixed number of timesteps, the generated trajectory will cover a small portion of phase space and hence, the average properties will not be representative of their true values; if the time step is too big, instabilities due to energy overlaps between atoms can arise.

However, there is a restriction that helps to define the time step used in a simulation. It states that the appropriate time step should be an order of magnitude smaller than the fastest motion, which is usually the vibration of bonds involving light atoms such as hydrogens. To access phenomena of interest, it is usual to integrate the equation of motions iteratively for millions of time steps and this is time consuming. A way to overcome this is to choose to freeze these vibrations to their equilibrium values, and thus the times step could be made larger (use of constraints) or alternatively use CG models.

2.5 Thermodynamic Properties

One way to validate the accuracy of a computer simulation is to compare the values obtained from the simulation with experimental data. Apart from structural data, we can also calculate thermodynamic properties of a system. Some common properties that can be determined through a computer simulation are the internal energy, the heat capacity (partial derivative of the internal energy with respect to temperature), the pressure (calculated via the virial theorem of Clausius and the forces acting on the atoms), the temperature (connected to the kinetic energy) and the radial distribution function (Leach, 2001). Thermodynamic concepts such as temperature and pressure can only be described as ensemble averages or as parameters that define the ensemble in computer simulations and should be distinguished from mechanical properties that are calculated at each time step (such as energy and the instantaneous pressure) (Allen and Tildesley, 1989).

2.5.1 Free energy surface

Free energy is one of the most important thermodynamic quantities and it is a state function meaning it does not depend on the path taken by the system. It also defines whether a system is in equilibrium or how stable it is. There are two forms of free energy, namely Gibb's and Helmholtz free energy, but we are mostly interested in the former which can be converted into work. Gibb's free energy applies to systems under the NPT ensemble, the results from which can be used for direct comparison with experiments. The Gibb's free energy (G) is described by Equation 14:

$$G = U + pV - TS = H - TS \tag{14}$$

where p is pressure, T is temperature, V is volume, U is the internal energy, H is the enthalpy and S is the entropy of the system. We are mostly interested in the change in the free energy (ΔG) to help us determine the direction of a spontaneous reaction and to evaluate the maximum work that can be associated with a thermodynamic process.

Chemical and biological processes can have complicated free energy surfaces (free energy landscapes) with many local energy minima and maxima. The free energy of a system can be reconstructed based on some collective variables that are chosen to describe the system. A collective variable (CV) is usually a function of the atoms positions and is an efficient way to describe the behaviour of the system by reducing its complexity (reducing the high dimensional space of the system). A system can encounter high energy barriers in the attempt to cross from one state to the other. The next state could be more favourable than the previous one and therefore, enhanced sampling methods are used to help the system cross the high energy barriers and reach a more favourable conformation (enhanced sampling methods are discussed in more details in section 2.6).

Free energy is important because it is also a way to express probability P(s) of finding the system in a given microstate s (see Equation 15). At thermodynamic equilibrium, the probability of finding a given value of position and velocities is the Boltzmann distribution. Thus, the value of the free energy associated with the value of a CV(s) can be calculated by accumulating a histogram of s and take its logarithm times the

Boltzmann constant K_B and the temperature T . The probability is given by Equation 15:

$$P(s) \propto e^{\frac{-F(s)}{k_B T}} \propto e^{\frac{-K(u)+V(q)}{k_B T}} \quad (15)$$

where *P(s)* is the probability of finding the system at a given state *s*, while *F(s)*, *K(u)* and *V(q)* is the total, kinetic and potential energy of the system and K_B is the Boltzmann constant. Another important equation that relates the free energy with the partition function *Q* is Equation 16.

$$F = -K_B T \ln Q \quad (16)$$

The partition function is the sum of all possible states in phase space in the relevant ensemble multiplied by their probability. However, it is not feasible to sample the entire phase space that is why we look at ratios of the partition function Q.

To summarise, under constant temperature and pressure the free energy of a system is the Gibb's free energy and the corresponding partition function is given by Equation 16. The logarithm of the partition function is a way to express free energy (probability of finding the system at a given state) and this is a way to include information about the energy and entropy associated with the system. MD simulations do not sample the relevant high energy states and that is why it is difficult to calculate free energies out of simulations. To overcome this, we need to make the system sample high energy states by adding a bias such as during the umbrella sampling technique and calculate the difference between the free energy of the two states.

2.6 Enhanced Sampling Methods

2.6.1 Potential of mean force – umbrella sampling

The free energy landscape of a system might be complex, but it contains the invaluable information of how the free energy changes as a function of some degrees of freedom usually functions of coordinates. This change in the free energy with respect to a chosen coordinate is known as potential of mean force (PMF) (Roux, 1995). It should be noted that the PMF can describe physically achievable processes unlike free energy

perturbation methods where unphysical pathways - transformations occur (Leach, 2001). The simplest way to calculate PMF is to monitor the change in the free energy when the distance between the particles changes. For example, if we are interested in the PMF of two interacting peptides, we monitor how the free energy changes along their intermolecular distance. However, because the PMF can vary a lot (several K_BT values) over the range of the selected degree of freedom (e.g. distance) and this could lead to significant errors, a method called umbrella sampling is used to overcome this problem during MD and Monte Carlo simulations. The problem arises from the fact that during a computer simulation the system does not adequately sample regions where the free energy is high.

Umbrella sampling aims to modify the potential function in order to sample the aforementioned insufficient sampled regions (Leach, 2001). These regions are unfavourable due to their high energy that is why they are not sampled enough. The modified potential function has the form of Equation 17:

$$V'(\boldsymbol{r}^N) = V(\boldsymbol{r}^N) + W(\boldsymbol{r}^N) \tag{17}$$

where $V(r^N)$ is the potential energy of the system and $W(r^N)$ is the weighting function that has usually a quadratic form with respect to distance. The modified potential enhances sampling in the lower probability regions of phase space bridging the two states like an umbrella. The reaction is divided into windows with specific *r* assigned to each window and the bias potential keeps the system in the vicinity of *r*. Therefore, binding free energies for peptides or proteins can be calculated provided that there is sufficient sampling of the two states and overlapping between the windows.

2.6.2 Metadynamics

Physical and chemical phenomena of interest often take place on longer timescales that are not accessible with existing computing resources and algorithms. This timescale problem as well as the high dimensionality of the description of a system is addressed by enhanced sampling techniques such as metadynamics, replica exchange molecular dynamics and steered MD.

Chemical reactions, crystal structure transitions, protein folding and protein-protein interactions are all considered rare events in the currently accessible simulation times. Metadynamics is an enhanced sampling technique that adds a history dependent bias potential, in the form of Gaussians, to accelerate the dynamics of a system and simulate rare events (Laio and Gervasio, 2008). Metadynamics discourages the system from revisiting configurations that have already been sampled while pushing it away from local energy minima (Barducci et al., 2011). Hence, it encourages the system to explore the whole free energy surface (FES) and explore new reaction pathways, equilibrium configurations and metastable states.

The reconstruction of the FES is based on a few CVs which are selected degrees of freedom where the added bias potential acts on. CVs offer a simplified or CG representation of complicated systems and their choice is not trivial. They are functions of the coordinates and optionally of the momentum of atoms and are used not only for analysing the simulations but for biasing them too. Common CVs are distances, angles and dihedral angles between atoms or group of atoms as well as coordination numbers, root mean square deviation values, radius of gyration etc. It is important to note that the value of the CV should be invariant of the translation of the whole system. For example when considering a translocation of an ion through a membrane channel, the distance between the ion should be relative to the bilayer, because the position of the ion solely is not very informative. A CV should be continuous because it adds a bias to the Hamiltonian of the system and when we compute the force, integrating non continuous quantities could lead to problems.

CVs are degrees of freedom that can efficiently describe the process of interest and are chosen before the simulation to bias it or after the simulation for post-processing of the data. They are usually defined after trial and error attempts. Sometimes the choice of a CV is easy, like the distance between two atoms interacting when forming a bond but most of the times the choice is not obvious and usually more than one CV is required to describe the whole process.

To conclude, CVs are important parameters used in metadynamics to describe the free energy landscape of complex systems by reducing the high dimensional space (6N) of MD simulations to a low dimensional one (1 or 2 depending on how many CVs are

chosen). However, the search of appropriate CVs that describe adequately a system is not a trivial process.

2.6.3 Replica exchange method

Replica exchange is another enhanced sampling method used to overcome the problems that brute force MD simulations face; such as a system being trapped in local energy minima unable to cross high energy barriers due to the timescale problem. In replica exchange or parallel tempering, replicas of a system are simulated in parallel (at the same time) at a range of different temperatures starting from near ambient temperature leading to higher ones (Earl and Deem, 2005). Then configurations of the system from one temperature can be swapped with other systems from adjacent temperature based on a Metropolis criterion (or a swapping probability Δ) (Shea and Levine, 2016). The swapping probability is described by Equation 18:

$$\Delta = \left(\beta_i - \beta_j\right) * (V_i - V_j) \tag{18}$$

where $\beta_i = \frac{1}{K_B T}$, K_B is the Boltzmann constant, T is temperature and V_i is the potential energy of replica i. A swap between the replicas (i and j) will take place either when the swapping probability Δ is smaller or equal to zero or in the case of a positive Δ the swap will take place with a probability of $P = \exp(-\Delta)$. The set of temperatures used in a replica exchange simulation is critical for correct and efficient sampling.

The choice of the range of temperatures is important, because the resulting transition probabilities of neighbouring systems should be approximately the ideal value of 0.25. Furthermore, we want to adequately heat up the system so that we will observe energetically unfavourable behaviour. In this way, the system can cross high energy barriers and explore other energy states that might otherwise have been difficult to access. It is also essential to maintain the original room temperature replica of the system for analysis of physiological relevant systems. Thus, this computational technique helps systems escape from local energy regions enhancing sampling.

It should be noted that the differentials of replicas (differences in temperature) at higher temperatures are smaller compared to the lower ones (due to the exponential relation

with energy) and hence some can be omitted. If the high temperatures are close, the swapping probabilities will be high. However, care should be taken to avoid large gaps between replicas because this will reduce the probabilities of swapping between adjacent replicas. The probability should be close to the ideal value of about 0.25 which could serve as a good criterion to select the range of temperatures with trial and error i.e. if the probability is close to the ideal value then the neighbour temperatures are chosen. While if it is around 0.4 then it means that one could get away with a bigger difference in temperature between the adjacent replicas to reduce the probability of transition. An interesting implementation of the technique it to generate a number of configurations for each simulation rather than using the same initial one (note that they should have the same volume). All of the replicas should be equilibrated at the same ensemble that we wish to sample from.

A good convergence indicator is for the energy of the replica near the temperature of interest to continually decrease. After the replica exchange simulations finish we can analyse the data. First, the generated low energy configurations can be characterised. Also, we can use algorithms which compare protein backbone atoms and based on their root mean square value groups them together. The resulting cluster of configurations can shed light in the search of the most dominant protein morphologies in a given environment. Thus, in the case of peptides or small proteins, one can investigate their folding while in the case of dimers or larger oligomers one can obtain information regarding their aggregation propensity. However, it should be noted that the time evolution of the system has been lost and so we cannot obtain averages of the system's properties as well as kinetics of molecules.

In the MD simulation package GROMACS (used in the studies of the thesis) the replicas exchange coordinates, other software packages could exchange temperatures (Abraham et al., 2015). So, when viewing the trajectory it is noticeable that it is no longer continuous. This means that the time evolution of observables like root mean square deviation will show abrupt jumps. There are guidelines on how to produce a 'continuous' trajectory for all the examined replicas. The latter trajectory would be useful if one wants to identify the structural transitions that occur at higher temperatures as energy barriers are surmounted. Usually, however the ensemble at the lowest temperature is of interest. The latter part of the trajectory of the lowest temperature

replica can be used for analysis and the most favourable structure is considered to be the one after convergence from the replica at the target temperature.

2.7 Free Energy Using Molecular Mechanics Poisson Boltzmann Surface Area Method

There are other methods to calculate interaction energies that do not have the same accuracy as e.g. umbrella sampling but are very attractive because they are quick and provide a good estimate of the interaction free energies. These methods are often used to evaluate relative stabilities of different biomolecular structures or binding energies between a protein and different ligands. A very popular method is the Molecular Mechanics Poisson Boltzmann Surface Area method (MM/PBSA) largely used in biomolecular simulations (Genheden and Ryde, 2015). The basic concept is to produce an estimate of the free energy using an ensemble of structures at the initial and final state of the system and combined with MD it can incorporate conformation fluctuations and entropic contributions to the binding free energy.

The MM/PBSA method combines three energetic terms to account for the change in free energy upon binding, namely the change in the potential energy in vacuum, the polar and non polar solvation energy as well as the entropic contribution (Genheden and Ryde, 2015). The first term has contributions from the bonded and non-bonded terms of the potential energy function (such as bonds, angles and torsional terms) in vacuum. The second term comes from the use of an implicit solvent model to calculate solvation energies which is the energy required to transfer a solute from vacuum into the solvent. This term has electrostatic and non-electrostatic contributions where the electrostatic one comes from solving the Poisson – Boltzmann equation while the other term includes repulsive and attractive forces. The repulsive forces are due to the cavity formation and the attractive ones due to vdW interaction between the solute and solvent. It is noted that the attractive vdW energy term is not accounted when using the solvent accessible surface area (SASA) model. In this case, an assumption is made that SASA is linearly dependent on the nonpolar term with the help of a fitting parameter. Moreover, it should be noted that when using g_mmpbsa a tool incorporated into the GROMACS package the entropic contribution which is the last energetic term is also discarded (Kumari et al., 2014).

To summarise, MM/PBSA is a method that has found many applications mostly because it requires a modest computational effort and shows intermediate performance. It contains several crude approximations that are questionable such as the lack of conformational entropy or the lack of information regarding the number and free energy of water molecules found in the binding site. It should be noted that although it ignores structural changes upon binding and conformational entropy, it has been found to show acceptable prediction in terms of ranking when used for the study of protein-ligand complexes and its accuracy is better than docking methods (Genheden and Ryde, 2015).

3 The Mutual Interaction of the Extracellular Domain Particles of Claudin-1

3.1 Introduction

In the previous two chapters of this thesis, the relevant background with regards to what epithelial cells are, why they are important and what types of junctions they form was presented. There was a focus on the structure and function of the tight junctions (TJs) and the structure of skin, which are of interest here. There was also an overview of the methodology used in the thesis. In this chapter, we focus on the cis-interaction between the TJs proteins, i.e. the interaction between the proteins while they are embedded in the same lipid bilayer (side by side). We have employed atomistic molecular dynamics (MD) simulations to investigate the molecular organisation of the main TJs proteins, the claudins. The self-assembly simulations of claudin-1 particles within a membrane, focuses on the extracellular domain (ECD) which is larger in diameter and laterally overhangs much of the transmembrane domain (see Figure 3.1 and Figure 3.2). The length and timescale issues of atomistic-resolution MD simulations have been overcome in a creative way by simulating the self-assembly of the ECDs in an *implied* membrane. The concept of an implied membrane is introduced and it is further employed in Chapter 4, where the trans-interaction between the ECDs is examined.

As previously mentioned, TJs are cell-cell contact structures found in epithelial and endothelial tissues, located at the contact region between neighbouring cells towards their apical side. They serve as a barrier to molecular transport through the intercellular space (paracellular pathway). In some tissues, TJs totally block the passage of small molecules and ions (barrier function) while in others they allow molecules of specific charge and size to go through (selective-channel property). Therefore, the molecular

permeability of TJs is tissue-dependent with some epithelia being characterised as 'tight' (urinary bladder, skin) while others as 'leaky' (small intestine, colon) (Turner et al., 2014). TJs are also believed to function as ‘fences’ restricting the intermixing of proteins between the apical and basolateral domains of the plasma membrane. A functional TJ barrier is critical to the physiology of the body. Its dysregulation can lead to pathologies such as inflammation, metastasis and oedema (Sawada, 2013). For instance, TJs that are present in skin prevent water loss from the body and block the entrance of pathogens and xenobiotics into the body (Bäsler et al., 2016). TJs are also the target of several viruses including the hepatitis C virus (Meertens et al., 2008) and bacteria such as the bacterium *Clostridium perfringens* that produces the enterotoxin responsible for food poisoning (Shinoda et al., 2016). Moreover, there are numerous hereditary diseases that are linked with mutations of TJ proteins, which include hypomagnesemia, deafness, neonatal sclerosing cholangitis with ichthyosis and familial hypercholanemia. They are also the focus for strategies for enhancing drug delivery of large molecules including proteins across the gastrointestinal tract (Gonzalez-Mariscal et al., 2005, Deli, 2009, Cording et al., 2017, Takahashi et al., 2012). For example, peptides derived from the extracellular domain of TJ proteins or from the C-terminal of the Clostridium Perfringens enterotoxin have been shown to enhance paracellular transport and drug delivery (Deli, 2009, Cording et al., 2017, Takahashi et al., 2012).

An important research question is the nature of the molecular architecture of the proteins that make up the TJs, a better understanding of which would enhance our knowledge of the basic principles of paracellular permeability.

In electron microscopy (EM) images, TJs appear as series of membrane fusion ('kissing') points circumscribing the cell. They appear as a cross-linked network of strands. The observed ‘kissing’ points in EM images correspond to approximately 10 nm discrete transmembrane particles with a centre to centre distance of 18 nm between them (Anderson, 2001). At the molecular level, TJs are a multi-protein complex comprising transmembrane proteins, cytoplasmic plaque proteins, signalling proteins and adaptors that connect them to the actin cytoskeleton (Günzel and Alan, 2013). The transmembrane proteins include primarily the claudins, tight junction associated MARVEL proteins (TAMPs), and JAMs, as discussed in detail in Chapter 1. The claudins are considered to be the fundamental functional and structural elements of TJs

as they alone can reconstitute tight junction-like strands in cells that normally lack these structures (L. fibroblasts) (Furuse et al., 1998). The known crystal structures of claudins (claudin -3, -4, -9, -15 and -19) (Suzuki et al., 2014, Saitoh et al., 2015, Shinoda et al., 2016, Vecchio and Stroud, 2019, Nakamura et al., 2019) reveal a structure that looks like an Olympic torch with a transmembrane region (the body of the torch) comprising a left-handed four-helix bundle (Suzuki et al., 2014) and an extracellular part (the head of the torch) composed of two loops that are integrated via β-sheets (see Figure 3.1). The ECD is comprised of two loops, where the first and larger one exhibits a small helix and four β strands integrated with a fifth β strand of the second smaller loop (Suzuki et al., 2014). The small helix of the first loop is located in the area after the second transmembrane region and before the β4 strand. Additionally, there is a small intracellular loop, while both termini of the protein end up in the form of tails within the cell. On reviewing the 3-dimensional structure of claudins, it is evident that the ECD particle (the 'torch' head) is larger in diameter compared to the rest of the body (Figure 3.1). Closer examination of the ECD reveals a curved 'palm' shaped structure which is believed to line the paracellular pore and hence determines its characteristics.

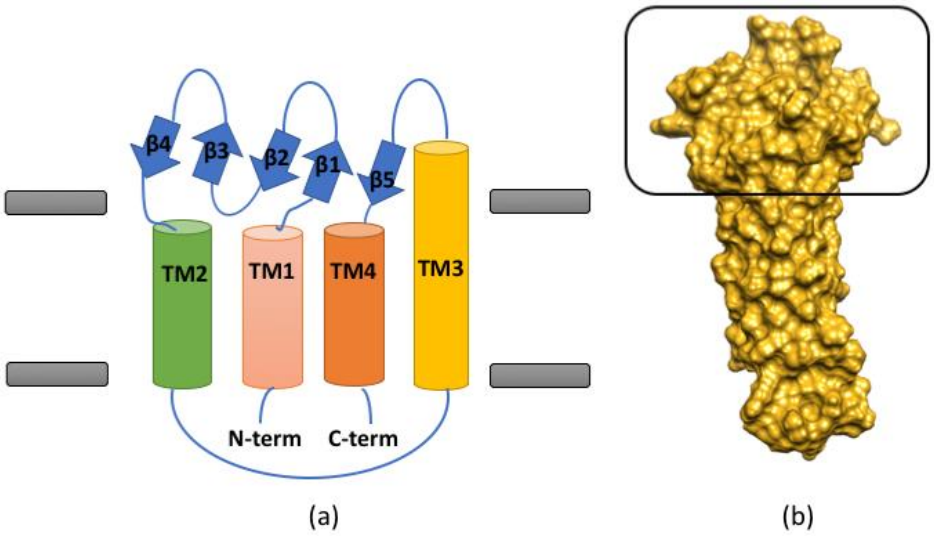


Figure 3.1: (a) Claudin structure in schematic form showing the extracellular domain comprising the β-sheets and loops and the transmembrane domain comprising four α-helices. (b) Structure of claudin-1 in surface representation, with the extracellular domain region framed. Rendering was done with UCSF Chimera (Pettersen et al., 2004).

Whilst we have the aforementioned structural data (crystal structures of claudins), their molecular organisation in TJs is not wholly resolved. There is insufficient data to hypothesise how they form the particles observed in EM images. The crystal packings of the claudins suggest a possible molecular organisation - the so-called Suzuki model (Suzuki et al., 2015). In the Suzuki model the TJ strands or fibrils within an individual membrane consist of claudin monomers arranged in anti-parallel double rows. The proposed formed interfaces are two: the face to face and the linear interface.

Whilst this structure has gained some acceptance, there is also acknowledgement that other dimeric interfaces are likely, indeed must exist, to explain the branching of the TJ strands observed microscopically (Furuse et al., 1998, Gong et al., 2015, Rossa et al., 2014, Zihni et al., 2016). Further, the crystal structures are essentially in an aqueous environment (except for claudin-15 which was crystallised in the lipidic cubic phase) and hence, would not be expected to reflect fully the molecular organisation of these proteins in their native environment, i.e. when embedded in a lipid membrane.

Experiments demonstrate the existence of small pores existing at the interface of the (trans-)interaction between the claudin ECDs that protrude from the membranes of adjacent cells, which allow specific ions and small molecules to pass through the intercellular space (Günzel and Alan, 2013). These pores have specific radii and are charge-selective, the radius and selectivity depending on the particular residues present in the extracellular loop region of the given claudin (Piontek et al., 2017, Rossa et al., 2014, Krause et al., 2009, Krause et al., 2015, Veshnyakova et al., 2012, Milatz et al., 2015). Claudin-1 is considered as predominantly barrier forming along with the claudins -3, -5, -11, -14 and -19 whilst claudins -2, -10a, -10b, -15 and -17 are considered to be predominately pore forming with charge selectivity (Günzel and Alan, 2013, Krug et al., 2012a, Heinemann and Schuetz, 2019). The paracellular pore in the Suzuki model shows up as β-barrel-like structure with pore facing (aspartic acid, ASP55) and pore lining (aspartic acid, ASP64) residues, resulting from the intra-cellular (cis-) and inter-cellular (trans-) interaction of the TJ proteins (the trans-interaction and the formation of the TJ pore are discussed in more detail in Chapter 4 of the thesis).

The definitive molecular organisation of the TJ claudins in their native environment i.e. when embedded in lipid membranes is currently inaccessible by experiment. This question of molecular organisation has also been tackled by molecular dynamics simulation, wherein the molecular trajectories are simulated using Newtonian mechanics driven by inter-molecular forces. Seminal simulations were carried by the Nangia research group with a focus on the architecture of the blood brain barrier TJs (Irudayanathan et al., 2015, Irudayanathan et al., 2017, Irudayanathan et al., 2018). They studied the self-assembly of claudin particles (specifically claudins -3 and -5) embedded in a lipid membrane with MD simulation using coarse grained (CG) models (Irudayanathan et al., 2015, Irudayanathan et al., 2017). It was necessary to resort to CG models that enable large systems to be simulated for longer times. In particular, the CG approach enabled the relatively-slow molecular re-organisation of the claudins in the membranes to be tracked. Recently, they extended their work to other claudins including -1, -2, -4, -15 and -19 (Irudayanathan et al., 2018). The simulations have identified multiple, preferred (cis) dimeric interfaces that showed some variation between specific claudins. The formed dimers were used to produce trans-interacting systems through docking methods and the generated paracellular pore was characterised.

CG models reduce the number of degrees of freedom by representing chemical moieties (comprising a collection of atoms) by single CG particles, enabling the simulation of larger systems for longer timescales (Marrink et al., 2007). Whilst accurate CG models can be developed, the common approach for biomolecules is to employ the 'universal' MARTINI force field (Monticelli et al., 2008), which indeed was used in the simulation studies referenced above. The MARTINI-type models by design are simple representations and hence semi-quantitative (Marrink et al., 2004). Beyond the loss of atomistic resolution, the molecular flexibility in these models, particularly of protein structures, is significantly restricted; indeed, the secondary structure is selected at the outset and remains fixed throughout the simulation.

In this chapter, we employ MD simulation using atomistic resolution to examine the self-assembly of ECD particles of claudin-1 with a view to elucidating the molecular organisation of the TJs. Claudin-1 is expressed rather ubiquitously in the body and plays a critical role in establishing a functional and efficient epidermal barrier (Furuse et al.,

2002, Kirschner et al., 2010). The atomistic molecular simulation approach offers greater accuracy over the CG approach. We revisit the problem of limited accessible timescale in a creative way by focusing on the extracellular domains (the heads of the 'Olympic torches') of the claudin particles and simulating their self-interaction at an atomistic level in an *implied* membrane. Thus, the extracellular domain particles are restrained to lie on a plane that serves as the membrane. Note that lipids are not explicitly represented in the simulations. From Figure 3.1 we notice that the ECDs are in fact larger in diameter and laterally overhang much of the transmembrane domain. Whilst we acknowledge that the cis-interaction of the transmembrane domains may prove to be important for particular orientations, it is clear from the steric picture that the ECD is likely to be critical for a significant part of the cis-interaction orientation space.

3.2 Methodology

The behaviour of a single, full claudin particle embedded in a lipid membrane and of the claudin-1 ECD in an implied membrane were examined in order to characterise their structural stability. We then carried out a large-scale self-assembly simulation of a grid of separated and randomly-oriented ECD particles in an implied membrane. The particles were restrained to lie on a 2-dimensional plane to mimic their placement in a lipid membrane, on which they were free to interact amongst themselves.

The use of an implied lipid bilayer reduces the degrees of freedom in the system, thus eliminating the need for an explicit definition of the membrane lipids along with the claudin transmembrane region for each of the claudin particles. This atomistic approach (in contrast to CG) maintains the accuracy of the simulations whilst enabling a relatively large system to be simulated for a longer timescale. In particular, the uninteresting, slow diffusion of the claudin transmembrane region (the body) within the explicit lipid membrane is eliminated. The assumption being made here is, given that the ECD dimensionally largely overhangs the transmembrane region, the claudin-claudin interaction and hence the overall organisation of the claudin particles is determined largely by the extracellular region. Note that the cross-section diameter of the domain particle is approximately 30 angstroms in the longer dimension and 24 angstroms in the smaller one, while the diameter of the transmembrane helix bundle is approximately 20 angstroms (Figure 3.2 and Figure 3.3). However, the extracellular region does not

surpass the transmembrane region in every direction, and therefore the role of the latter in modulating the interactions is not ruled out.

3.2.1 Model building of claudin-1

A full-sequence atomic model of human claudin-1 was constructed through homology modelling with I-TASSER (Yang et al., 2015) an on-line server used for protein structure and function prediction. I-TASSER used the three known crystal structures of claudins i.e. claudins -4, -15 and -19 (PDB:4P79 (Suzuki et al., 2014), PDB:3X29 (Saitoh et al., 2015) and PDB:5B2G (Shinoda et al., 2016)) as templates to generate the model (see Appendix A2). The disulfide bridge between the conserved cysteines 54 and 64 was maintained, on the basis of the oxidising environment of the extracellular space, and because it is essential for pore function (Li et al., 2013). The homology model selected showed the best prediction based on the significance of threading template alignments and the convergence of parameters of the structure assembly simulations (see Figure 3.2). In the assembly simulations, the model was built from the identified fragments taken from the templates using Monte Carlo simulations. We isolated the ECD comprising the two extracellular loops based on its topology (CLD1_HUMAN) which was retrieved from the UniProtKB database (Table 3.1) (Consortium, 2018). The first loop is composed of 53 residues (amino acids 29-81) while the second loop has 27 residues (amino acids 137-163).

Table 3.1: The amino acid sequence of the two loops of human claudin-1 comprising the extracellular domain in FASTA format (one letter code) as retrieved from the UniProtKB database (Consortium, 2018).

ECL1	QWRIYSYAGDNIVTAQAMYEGLWMSCVSQSTGQIQCKVFDSLLNLSSTL QATR
ECL2	TAWYGNRIVQEFYDPMTPVNARYEFGQ

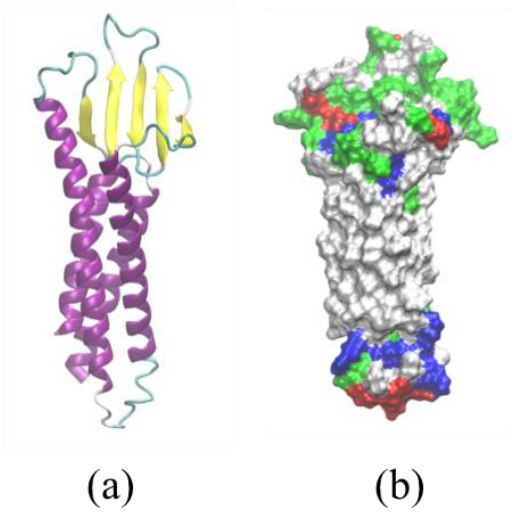


Figure 3.2: Structure of claudin-1 in cartoon (a) and surface (b) representation. The a helices are coloured in purple and the extended β-sheets in yellow. In the surface representation, white regions depict areas with hydrophobic character, green regions have hydrophilic character while blue are positively charged and red are negatively charged areas of the claudin. Rendering was done with VMD (Humphrey et al., 1996).

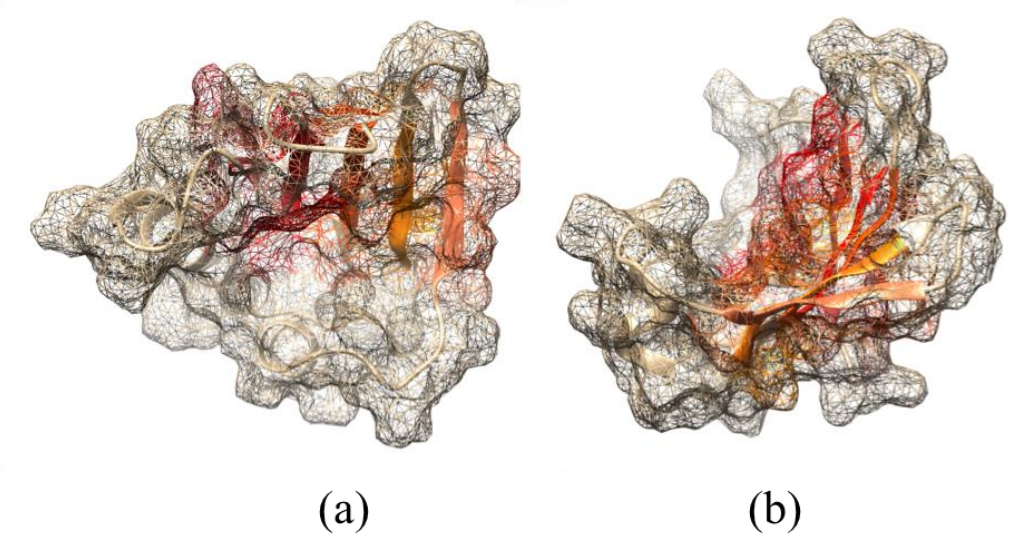


Figure 3.3: The ECD particle of claudin-1 viewed from the top (a) and side (b) in a combined cartoon and surface representation, rendered with UCSF Chimera (Pettersen et al., 2004). The surface representation looks like a dotted network while the beta strands are coloured in the palette of orange and red shades and are in cartoon representation.

The ECD shows an overall acidic character (-1 charge) and resembles a triangle from a top view (see Figure 3.3). The β4 strand is located on the one side of the ECD and the long helix that extends from the third transmembrane domain on the other side. Note that the folding of the ECD is such, that both elements (β4 strand and long helix) together with the areas in the proximity of the small helix and the back of the β sheet are expected to be important for the cis-interaction between the ECDs since they are located at the 'circumference' of the structure. The question is which one of the above mentioned regions is indispensable for the cis-interaction between claudins.

3.2.2 MD simulation of a single extracellular domain particle in an implied membrane

The ECD consists of two loops and hence has 4 terminal alpha-carbons. These terminal atoms were restrained to lie at a particular height on the z-axis (this coordinate represents the surface of the implied membrane) with a harmonic restraint force constant of 1000 KJ mol^{-1} nm^{-2}. Distance and angle restraints were also applied between the terminal atoms to maintain their relative positions. Specifically, a harmonic restraint with a force constant of 1000 KJ $mol^{-1}nm^{-2}$ was applied to maintain the distance between the four terminal atoms while a larger force constant of 10000 KJ $mol^{-1}rad^{-2}$ was used to maintain the angles between the terminal atoms within less than 10° degrees variation. The ECD particle was solvated in water (7543 water molecules) with ions at physiological concentration and simulated for 150 ns (temperature was set to 310K and pressure at 1 bar). The simulation converged with respect to both structure and potential energy in 30 ns into the trajectory.

3.2.3 MD simulation of a full claudin-1 particle in an explicit bilayer membrane

The explicit lipid membrane system comprised a full claudin-1 (without the cytoplasmic tail) embedded in a bilayer of the phospholipid dioleoylphosphatidylcholine (DOPC). DOPC was chosen as it yields a hydrophobic thickness that is commensurate with the length of the trans-membrane region of claudin-1 (see Figure 3.4). The membrane patch consisted of 132 lipids with the water layer thickness being approximately 3.5 nm (generated with CHARMM-GUI (Wu et al., 2014)). There were ions present in the simulation box at the physiological concentration of 0.15M NaCl. The system was

simulated at 310 K and 1 bar for 150 ns being converged after 30 ns in terms of both structure and potential energy. To characterise the structural stability of a claudin we calculated the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) of the backbone atoms. A particular focus was the variation in structure of the ECD.

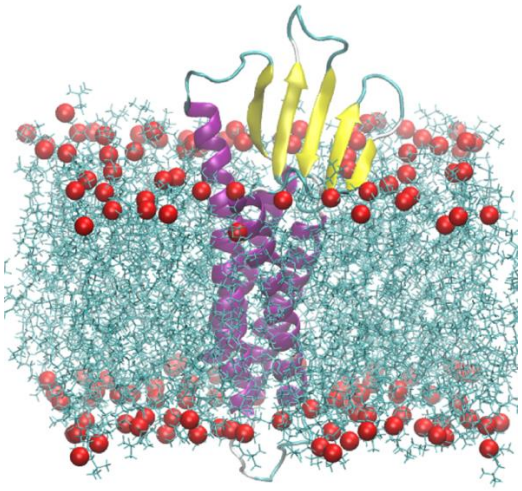


Figure 3.4: Claudin-1 embedded in a DOPC bilayer where the protein is shown in cartoon representation and coloured based on its secondary structure elements, while the DOPC headgroups are shown as red vdW spheres and the lipid tails as opaque bonds (rendered with VMD) (Humphrey et al., 1996).

3.2.4 Self-assembly of ECD particles in an implied membrane starting from a grid

The grid system comprised 8 x 8 separated and randomly oriented ECDs (centre-of-mass separation distance of approximately 5 nm along the x- and the y-plane) immersed in an aqueous environment (see Figure 3.5), in total comprising about 1.26 million atoms. To minimise self-interaction between the domain particles in the z-dimension the simulation box in the z-dimension was extended to approximately 7 nm. Note that this z-dimension was four times larger than the van der Waals interaction cut-off plus the height of an ECD, thus enhancing the water region between the periodic images. To increase sampling, the domain particles were randomly rotated about the z-axis to minimise biasing any particular cis-interaction. The system was simulated in the isothermal-isobaric (*NPT*) ensemble at 310 K and 1 bar for a total simulation period of 220 ns, beyond which the emergent structure showed little further evolution.

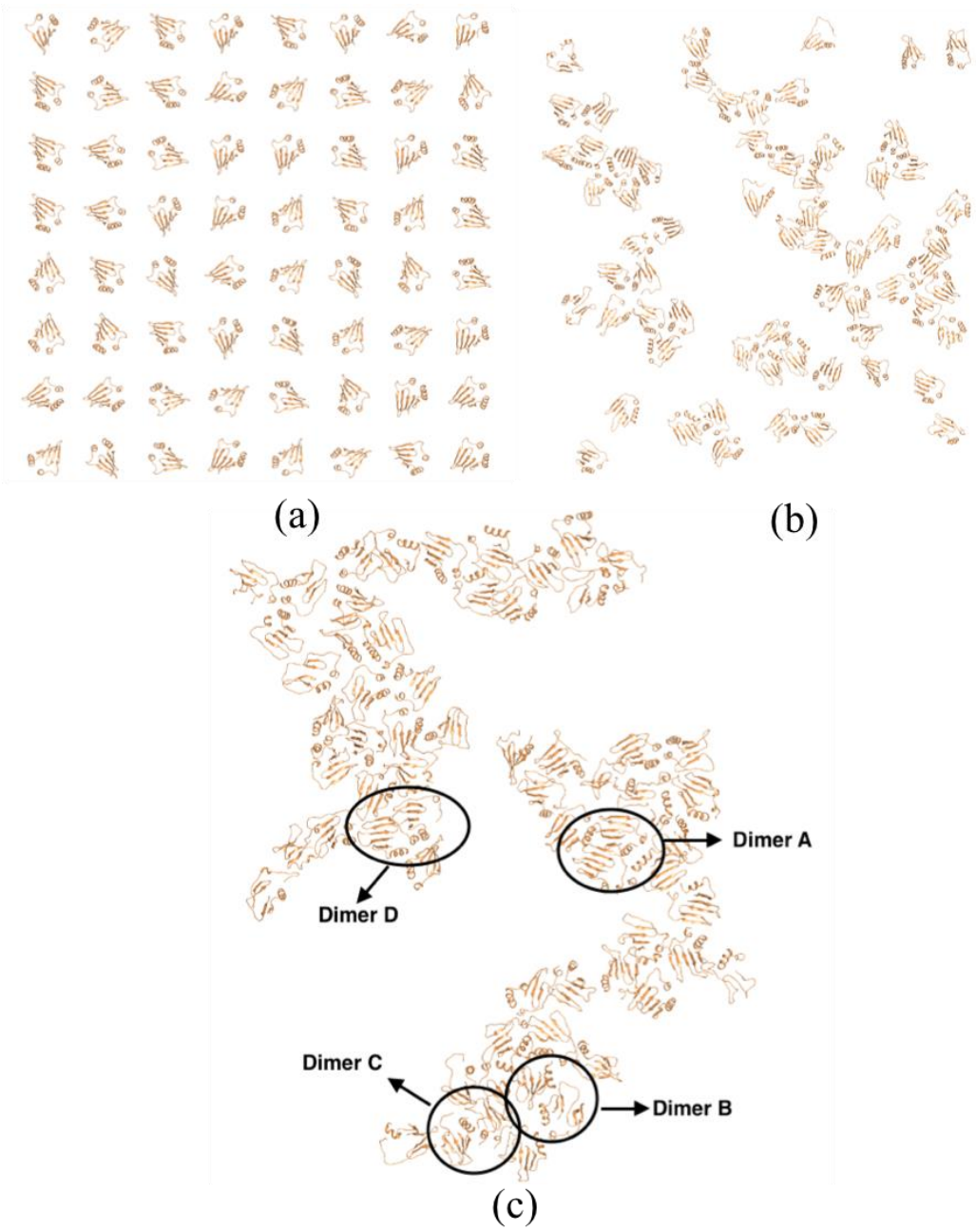


Figure 3.5: Self-assembly of ECDs starting on a grid. (a) The initial arrangement of the domains on a grid; (b) after 20 ns; (c) the final frame of the 220 ns trajectory highlighting the frequent dimers. The domain particles are all part of a single cluster at the end of the trajectory, although due to PBC they appear as if they form two different clusters. The protein atoms are in cartoon representation and coloured gold, while water molecules and ions are not displayed for clarity. The images were rendered with UCSF chimera (Pettersen et al., 2004).

3.2.5 The Suzuki 'face-to-face' interface

The double-row antiparallel arrangement of the proposed Suzuki model is based on the formation of two cis-interfaces between claudin monomers (Suzuki et al., 2015). One of the interfaces has a hydrophobic character and is observed in the crystal lattice (linear interface) and the other, designated as 'face-to-face' interface is formed through extracellular β-sheets interactions. The resulting structure looks like a 'half pipe' and joining this structure with the one from the opposing cell leads to the formation of a pore. In our simulation, we observed a similar linear interface as one of the stable interactions, but did not observe the exact Suzuki symmetric 'face-to-face' interaction between the claudin-1 ECD domains. In view of this, we set up this dimer interface and characterised it in terms of structural stability and free energy (see Figure 3.6). The generated dimer (after equilibration) was simulated in the *NPT* ensemble at 310 K and 1 bar for a simulation time of 50 ns.

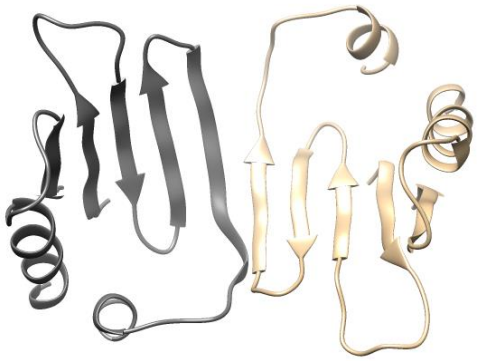


Figure 3.6: The 'face-to-face' interface of the Suzuki model. The claudin monomers are shown in cartoon representation and coloured in silver and gold (rendered with UCSF Chimera).

3.2.6 Potential of mean force profiles using umbrella sampling

The commonly-observed dimer interfaces were isolated from the trajectory and their binding free energy was characterised by means of potential of mean force calculation using umbrella sampling as implemented in GROMACS (Roux, 1995, Hub et al., 2010). The potential of mean force (PMF) is the change in free energy (within the solvated environment) as a function of a reaction coordinate, which in this case is the centre of

mass separation distance ξ between two domain particles whilst being restrained on the x-y plane (the implied membrane). Initial configurations of the dimers at various separation distances were generated by pulling one of the domain particles while restraining the other via its backbone atoms. The pulling rate employed was 0.01 nm/ps and the force constant 1000 KJ mol^{-1} nm^{-2}. Each of the configurations were spaced approximately 0.2 nm, which served to define the umbrella sampling windows (total of 17-20 windows). For each configuration, independent simulations were conducted with the centres of mass of the domains restrained in each window by a biasing potential using a force constant of 1000 KJ mol^{-1} nm^{-2}. After the systems were equilibrated, 10 ns production simulations were carried out in the *NPT* ensemble for each umbrella sampling window. The resulting PMF profiles were generated based on the weight histogram (WHAM) method implemented in the GROMACS package (Roux, 1995, Hub et al., 2010).

3.2.7 MD simulations: technical details

All simulations were carried out using the GROMACS platform (version 5) (Abraham et al., 2015). The force field utilised was CHARMM 36 with explicit TIP3P water model (Best et al., 2012). All simulations were performed at 310 K and 1 bar with periodic boundary conditions and ions at the physiological concentration of 0.15M NaCl. The bonds involving hydrogen atoms were constrained using the LINCS algorithm. The particle mesh Ewald method was used to calculate the long-range electrostatic interactions with a real-space cut-off at 1.2 nm, cubic interpolation, Fourier spacing equal to 0.16, and a precision (parameter ewald-rtol in GROMACS) of 10^{-5}. The same cut-off distance was used for the van der Waal's interactions with a switch function at 1.0 nm.

All initial configurations were energy minimised using the steepest descent algorithm and equilibrated stepwise beginning with 1 ns simulation in the canonical (*NVT*) ensemble, followed by a 2 ns simulation in an isothermal-isobaric (*NPT*) ensemble using the Nose-Hoover thermostat and the Parrinello-Rahman barostat. The backbone atoms of the protein were position-restrained during the equilibration runs. The production runs were carried out in the *NPT* ensemble using the Nose-Hoover thermostat and the Parrinello-Rahman barostat.

We characterised the simulation trajectories by examining the clustering of the ECDs, characterisation of the formed interfaces between dimers, the relative orientation of the interacting monomers, and the binding free energy of the common dimeric interfaces. Much of the analysis was carried out using GROMACS utility programs (Abraham et al., 2015) and UCSF Chimera (Pettersen et al., 2004) while the relative orientation angles were computed with in-house python scripts and the python package MDAnalysis (Gowers et al., 2016, Michaud-Agrawal et al., 2011) (see Appendix A1). We also considered the angles formed between the long helix which is part of the second loop, and the vertical axis, since it is hypothesised to play a crucial role in the formed dimeric interfaces. Additionally, the angles formed between the long helix and each β strand were examined. The binding free energy was also calculated, using the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) method (Genheden and Ryde, 2015) with a view to estimating the contribution of individual residues to the binding free energy. With regards to the MM/PBSA method, we analysed the first 2 ns (11 frames) of the trajectories from the umbrella sampling simulations where the dimers stayed bound (window 0) (Kumari et al., 2014). The dielectric constant for the solvent was set to 80, reflecting an aqueous environment.

3.3 Results and Discussion

3.3.1 Claudin-1 embedded in a phospholipid bilayer maintains its structural stability in MD simulation

The explicit bilayer simulation revealed that the structure of claudin-1 remains stable and that the first loop (the larger one) shows greater flexibility compared to the second loop (see Figure 3.7). The average RMSD value for the backbone protein atoms was 0.24 nm. Focussing our analysis on the ECD, the average RMSD of the backbone atoms of the first and second loop were 0.23 and 0.13 nm respectively. These relatively low RMSD values confirm the quality of the built model of claudin-1 and the accuracy of the CHARMM36 force field parameters, considering that the extracellular region of the protein is regarded its most flexible region.

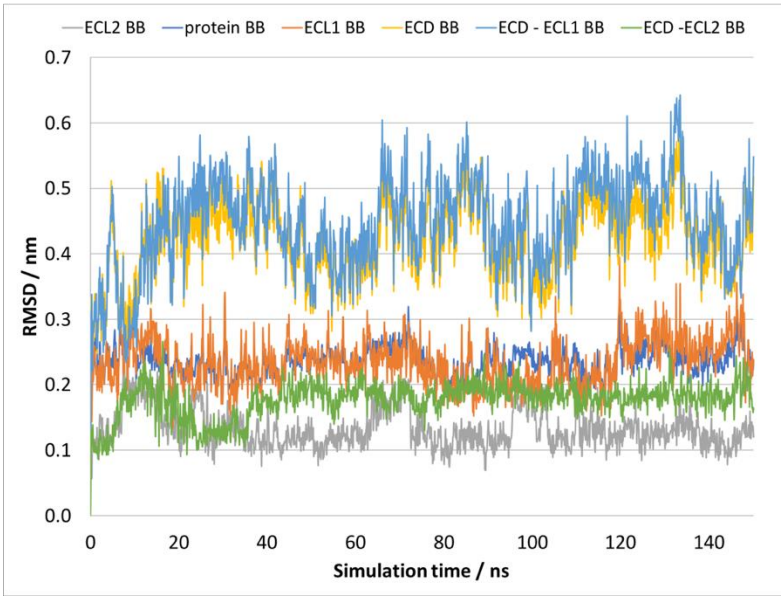


Figure 3.7: Root mean square deviation values for the backbone atoms of the ECD particle in an implied bilayer and for the whole claudin particle embedded in an explicit DOPC bilayer. The analysis is broken down to the backbone atoms of the whole protein and the first and second loops (ECL1 and ECL2) respectively.

The deviation from the starting structure was also characterised in terms of (i) the tilt of the ECD relative to the vertical axis and (ii) the opening up of the ECD. The tilt angle was defined as the angle between a vector defining the axial direction of a particular secondary structure element and the vertical axis. The 'opening' angle was defined as that between the long helix (of the second loop) and the individual β strands of the ECD. The variation in the tilt and opening angles as a function of simulation time are shown in Figure 3.8. Each of the β-strands show a different opening angle and tilt angle that increase in going from β-strand 1 to 4 akin to petals of a flower each opening up more and becoming more horizontal than the previous one. Both sets of angles show a variation of 10-20° degrees depending on the given β-strand, without any significant systematic drift.

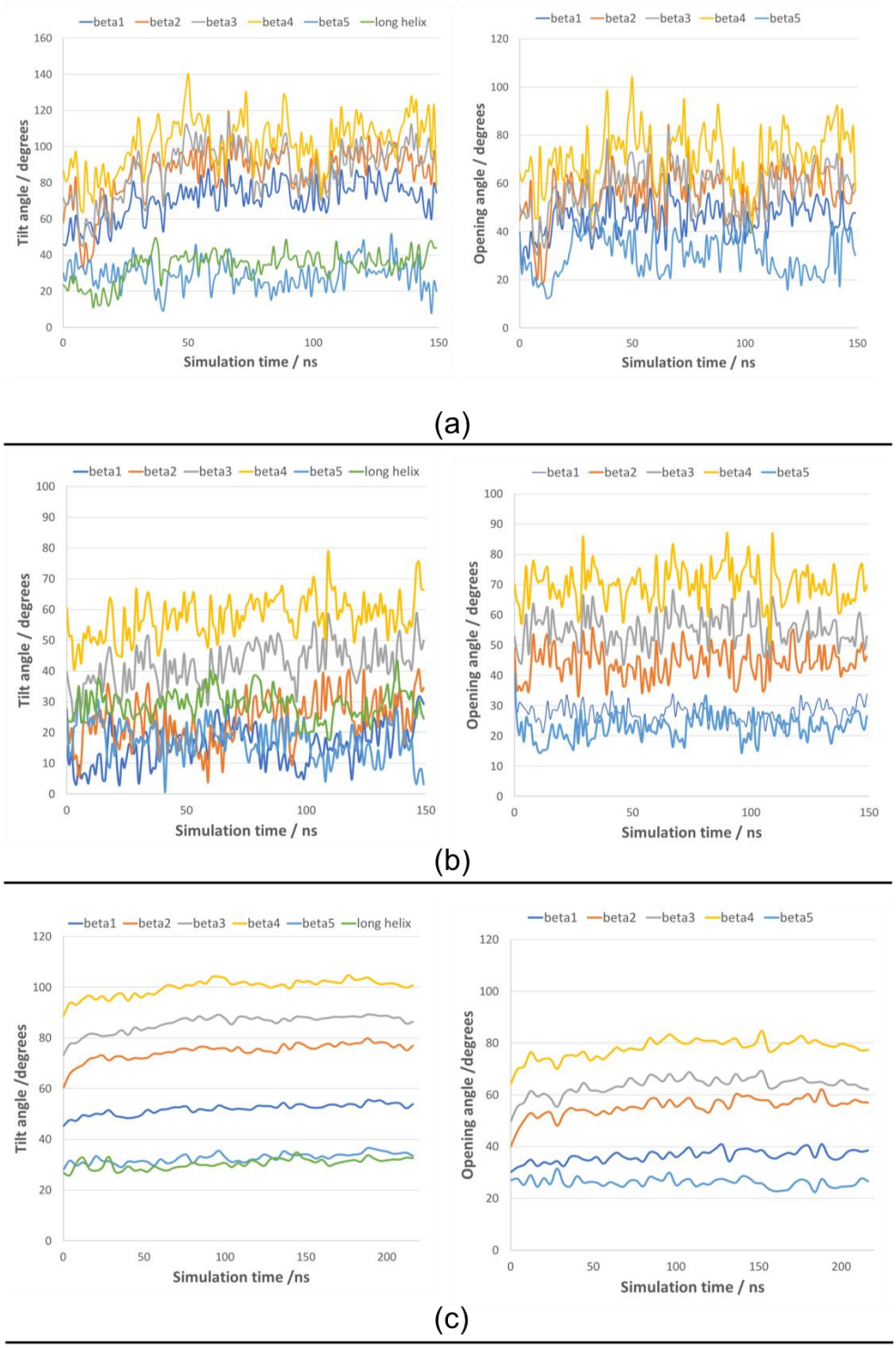


Figure 3.8: The tilt and opening angles as a function of time of the five β strands and the long helix of an ECD for three different systems: (a) Single, solvated ECD particle in an implied bilayer. (b) Full claudin-1 particle embedded in an explicit DOPC bilayer. (c) Self-assembly simulation of ECDs in an implied membrane starting from a grid system. The different colours represent different secondary structure elements as shown in the labels on top of each graph.

3.3.2 Isolated ECD in an implied membrane shows greater flexibility than the claudin-1 particle embedded in a bilayer membrane

Analysis of the simulation trajectory of the isolated ECD in an implied membrane confirmed that the various restraints (position restraints on the terminal atoms restraining them to the x-y plane, and associated distance and angle restraints) conserved the relative positions of the terminal atoms as per design. The implied membrane ECD shows greater flexibility and some structural deviation relative to the ECD of the full claudin particle embedded in an explicit bilayer. The average RMSD for the first and second loop for the implicit-membrane ECD was 0.45 nm (compared with 0.23 nm for the explicit-membrane ECD) and 0.18 nm (0.13 for the explicit-membrane ECD) respectively. For a visual comparison, the structure of the implied-membrane ECD is shown superimposed on the explicit-membrane ECD in Figure 3.9. The figure shows the structure of both ECDs in cartoon representation isolated from the last frame of the corresponding trajectories and superimposed. The RMSD for the superposition of the two structures (1048 atoms) was 0.42 nm.

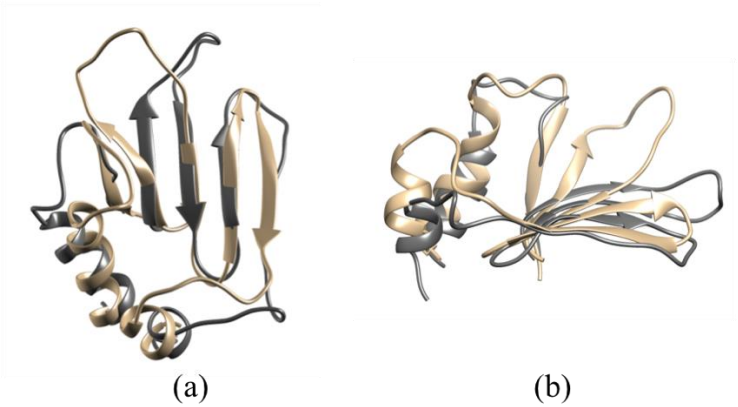


Figure 3.9: The superimposed final frame structures of ECDs from the implied-membrane and the explicit-membrane simulation systems, coloured silver and gold respectively from a (a) top and (b) side view (rendered with UCSF chimera (Pettersen et al., 2004)).

The implied bilayer ECD is generally more open with a higher opening angle. The average tilt and opening angles along with the associated standard deviations are tabulated in Table 3.2. The tilt angle of the long helix and the β5 strand are not too different for the explicit-membrane and the implied-membrane ECDs, but differ

markedly for the other β strands. The implied membrane ECD β strands are more open and inclined more horizontally, towards the (implied) membrane surface.

Table 3.2: Averaged opening and tilt angles of a single ECD particle in the implied membrane, the ECD domain of a single claudin particle in an explicit membrane, and the self-assembly system comprising 64 ECD particles in an implied-membrane. The uncertainties (±) represent 1 standard deviation.

		β1	**β2**	**β3**	**β4**	**β5**	**Long helix**
ECD simulation	Tilt angle / °	74 (±8)	90 (±9)	93 (±11)	106 (±12)	29 (±8)	37 (±5)
	Opening angle / °	46 (±7)	58 (±8)	60 (±9)	73 (±11)	32 (±8)	
Membrane simulation	Tilt angle / °	17 (±6)	25 (±8)	44 (±6)	59 (±6)	16 (±6)	29 (±5)
	Opening angle / °	27 (±3)	44 (±5)	55 (±5)	70 (±6)	23 (±4)	
Self-assembly simulation	Tilt angle / °	52 (±2)	76 (±2)	87 (±2)	101 (±2)	33 (±2)	31 (±2)
	Opening angle / °	37 (±2)	56 (±2)	64 (±2)	79 (±3)	26 (±2)	

Specifically, with respect to the tilt angles, we observe that the long helix and the β5 strand fluctuate around similar values for both the explicit-membrane and the implied-membrane simulations; while the rest of the β strands display a significant difference in their values. In the case of the implied membrane, the tilt angles of the β1-β4 strands are in the range of approximately 74-106 degrees, while for the explicit-membrane

simulation they range from 17 to 59 degrees. However, both depict a tendency to move closer to the x-y plane (bilayer) but in the case of the implied bilayer the domain moves more profoundly towards the implied membrane surface. Overall, we notice that in the implied membrane simulation the two loops maintain coherence (the extended β-sheet conformation is preserved) but the larger first loop is inclined more horizontally. While in the case of the explicit-membrane simulation the ECD exhibits a less dynamic behaviour where the angles fluctuate less, nevertheless again showing a tendency to become more horizontal. It might be the case that for the explicit-membrane system, the residues located close to the lipid bilayer (e.g. the hydrophobic residues isoleucine 32 and 62, phenylalanine 161 and methionine 52) interact with the phospholipids stabilising the structure of the claudin and additionally the presence of the lipids cause a steric hindrance so the loops cannot bend more towards the surface of the bilayer. Note that although the tilt angles for the β-strands in the implied membrane simulation might seem high, the opening angles, which show the relative position of each element, are consistent with the result from the explicit membrane simulation.

Considering the 'opening' angles of the β strands which describe the compactness of the protein loops, we notice that the values are similar within a difference of less than approximately 20 degrees (see Table 3.2) between the two systems studied. We also notice that β1 and β5 strands which are located closer to the long helix slightly reduce their opening angle in the explicit membrane simulation suggesting that they move even closer to the long helix of ECL2. The implied membrane ECD simulation displays a similar broad range of angle values, but more fluctuations, depicting a more dynamic behaviour. For example, the more profound 'opening up' of the first loop in the implied-membrane system that is shown in Figure 3.9 is also reflected in Table 3.2 where the opening value of the β1 strand is 27 degrees for the DOPC system and 46 degrees for the ECD system. Note however that β3 and β4 strands show similar angle values while β2 shows a small variation.

Figure 3.10 shows the radius of gyration (R_g) values of the backbone atoms for both systems- the implied-membrane (blue line) and the explicit-membrane (orange line) simulations. The R_g values are similar, indicating that the ECDs remain folded and largely compact as also described by the opening up angles. For the implied-membrane ECD, the average R_g value was 1.38 nm, indicating some opening up of the structure as

seen in Figure 3.9. For the explicit-membrane, there appears to be only a slight decrease, from 1.46 to 1.42 nm considering the first and final structure, with an average value of 1.44 nm.

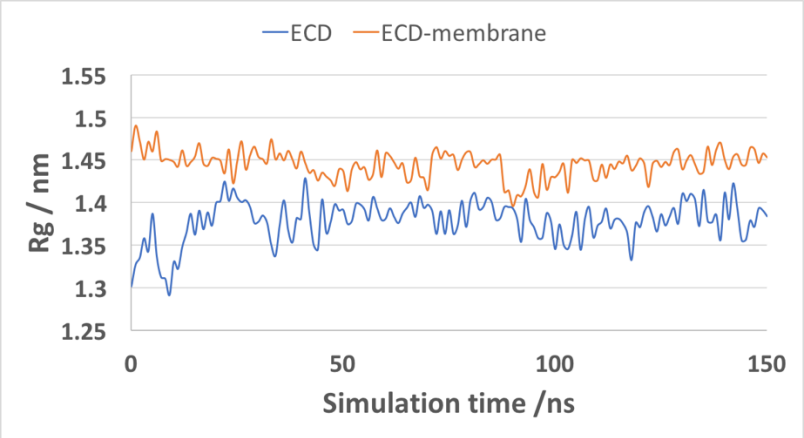


Figure 3.10: Radius of gyration with respect to simulation time for the implied and explicit membrane systems, shown as blue and orange lines respectively.

To further characterise the structural stability, we evaluated the average number of intra-particle hydrogen bonds and salt bridges, β-sheet content, and the RMSF value for both the implied- and explicit-membrane ECDs. The number of hydrogen bonds and salt bridges were marginally higher for the explicit-membrane ECD, 48 compared with 43 hydrogen bonds, and 8 compared with 6 salt bridges. Note that the total number of hydrogen bonds formed between protein atoms for the explicit membrane system was on average 140. Table 3.3 shows the salt bridges formed in the explicit-membrane simulation. The two interactions, glutamic acid 48 -arginine 81 and glutamic acid 147 -arginine 31, were not observed in the implied-membrane simulation. The salt bridges and hydrogen bonds enhance structural stability in both simulated structures.

Table 3.3: Salt bridges formed in the explicit-membrane ECD.

No	**Residues forming salt bridges**	**Domain**
1	GLU160 – ARG158	ECL2-ECL2
2	GLU147 – ARG31	ECL2-ECL1
3	ASP150 – ARG143	ECL2-ECL2
4	GLU147 – ARG143	ECL2-ECL2
5	ASP68 – LYS65	ECL1-ECL1
6	GLU48 – ARG31	ECL1-ECL1
7	GLU48- ARG81	ECL1-ECL1
8	GLU147-ARG31	ECL2-ECL1

The β-sheet content on average was very similar (22 residues for the implied and 21 residues for the explicit membrane) with the implied-membrane ECD showing greater fluctuation (see Figure 3.11). In the explicit membrane system every strand loses one or occasionally 2 residues from the extended conformation. Considering the implied membrane ECD there is a significant change in the β5 strand with the number of residues involved in its structure being reduced occasionally from 5 to 2. Note that β5 strand is part of the second loop, which was reported to have a helix-turn-helix motif (Krause et al., 2009) before the crystal structure of any claudin was resolved, and moreover, in another model of claudin-1 also depicts reduced content (Irudayanathan et al., 2018). Figure 3.11 presents the sum of residues adopting a β-sheet (blue line) or an α helix (orange line) conformation (Kabsch and Sander, 1983) in the implied membrane system and for comparison the β-sheet (black line) content of the explicit-membrane system is also displayed. To conclude, the implied membrane system shows a dynamic behaviour without any systematic drift with regards to the extended β-sheet content.

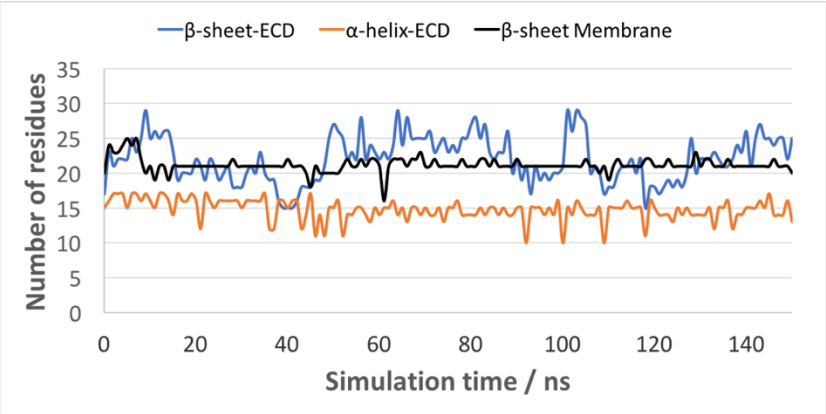


Figure 3.11: The sum of residues adopting a β-sheet (blue line) or an α-helical (orange line) conformation in the implied membrane simulation, as well as the β-sheet content of the explicit membrane system (black line) for comparison.

The RMSF value (which enables identification of highly mobile and flexible regions) as a function of residue index is presented for both the implied and explicit-membrane claudin particles in Figure 3.12. The RMSF value is also related to the crystallographic B-factor (see Equation 19). Comparing the ECDs only, the identified high-mobility regions are identical for both ECD particles, with the implied-membrane ECD exhibiting the greater mobility (Figure 3.13 and Figure 3.12). The high-mobility regions comprise the residues 37-42, 58-60 and 70-71 that correspond to the hairpin loops (connecting β1-β2 and β3-β4 strands) and residue 152 (an important methionine for the trans-interaction) in loop 2 *(Fukasawa et al., 2015)*.

$$B = \frac{8\pi^2}{3}(RMSF)^2 \tag{19}$$

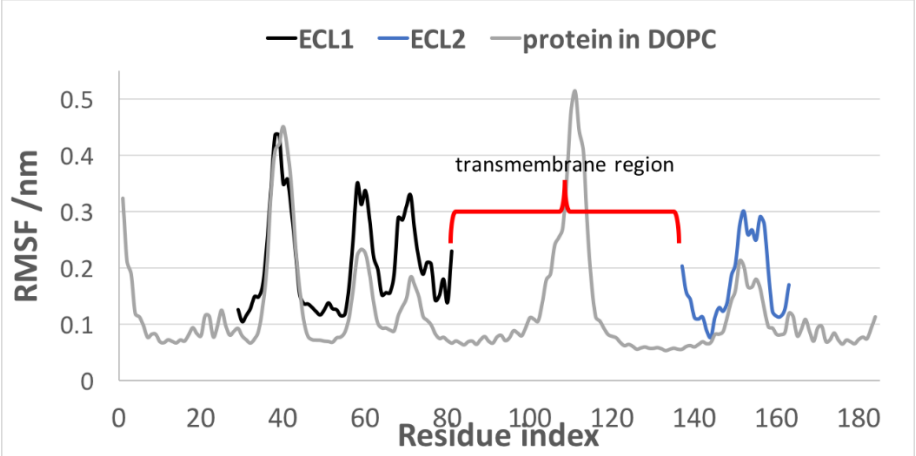


Figure 3.12: Root mean square fluctuation values as a function of residue index for the ECD in an implied membrane (black and blue line) and the full claudin-1 particle in an explicit membrane (grey line).

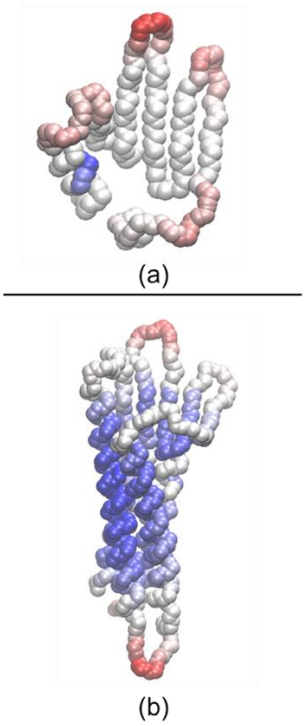


Figure 3.13: (a) The implied-membrane ECD particle and (b) claudin-1 from the explicit membrane simulation. Both structures are colour-coded according to the extent of structural deviation (based on their b-factor) with residues in red showing the highest displacement and those in blue the lowest one (rendered with VMD). The colouring method is BWR, midpoint in VMD is 0.03 nm and offset at 0.1 nm (Humphrey et al., 1996).

Note that the third transmembrane region (the longer helix) displays low RMSF values, probably related to both the high hydrophobicity of this domain and the fact that it does not display hydrophilic residues exposed to the lipid bilayer. The transmembrane regions 1 and 4, have some exposed hydrophilic residues to the lipids therefore, they may act as a driving force to bring the area behind the β-sheet together and hide those exposed hydrophilic residues. It is also noteworthy that claudins have a consensus motif present in their ECD (signature sequence W-GLW-C-C-R) which has been suggested to 'anchor' the β-sheet domain to the lipid bilayer (Suzuki et al., 2014, Suzuki et al., 2017). Here, the conserved residues tryptophan at positions 30 and 51, glycine at position 49, leucine at position 50 and the cysteines at positions 54 and 64 all display small fluctuations and thus, serve as the anchor for about which the β-sheet domain can bend towards the implied membrane (x-y plane) since the 'base' of the domain remains stable.

While one could introduce additional restraints into the implied membrane ECD to align the secondary structure elements to be more akin to the ECD of the full claudin particle and to reduce the opening angles, we decided to proceed without any additional restraints, preferring greater flexibility to arbitrary restraints. The simple lipid bilayer employed here comprising phospholipid DOPC (selected to optimise the commensurability of the hydrophobic region) for the full claudin particle may not be an appropriate choice, compared with the actual biological membrane. Using more restraints to match data that itself may be inappropriate could potentially be misleading. The flexibility of the ECD structure is essential, given that claudin-1's barrier function should withstand cell-to-cell stresses induced by for instance osmotic imbalances.

3.3.3 Self-assembly of ECDs in an implied membrane yields a network of strands

The randomly rotated domain ECD particles, initially located on a grid (Figure 3.5), very quickly come together forming dimers and small clusters, which then aggregate into a morphology that resembles the network of strands as seen in electron microscopic images (Furuse et al., 1998, Zhao et al., 2018, Zihni et al., 2016). A key observation is that stable trimers are also observed as part of the strand network, which are essential

for the formation of any branched morphology; dimer arrangements alone can only result in isolated dimers or a linear strand. The clusters once formed, remain largely unchanged for the rest of the simulation, suggesting strong binding interaction. Some slight rearrangements do occur but these are minor. The process occurs relatively rapidly with the bulk of aggregation occurring within the first 50 ns. The rate of aggregation in terms of number of aggregate clusters formed as a function of simulation time (based on a cut-off distance of 0.35 nm) is shown in Figure 3.14. The initial grid comprised 64 isolated domain particles which by the end of the simulation converged to a single aggregated cluster. The rapid aggregation suggests a high chemical potential (a certain super-saturation) of the domain particles in the system.

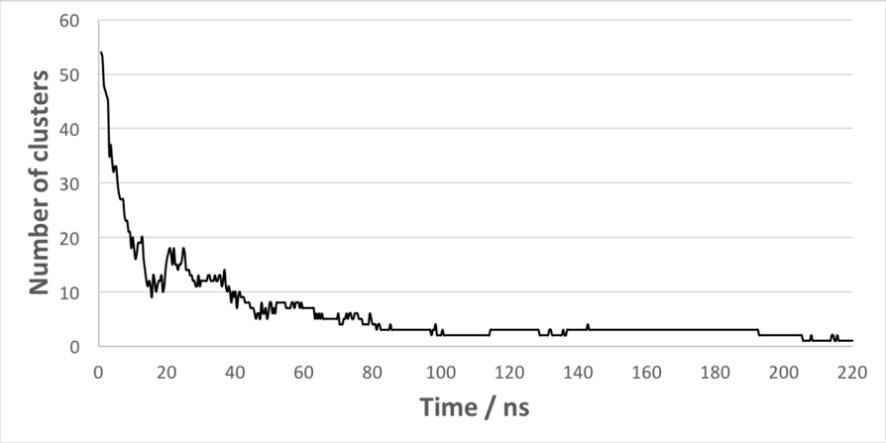


Figure 3.14: Number of aggregate clusters as a function of simulation time for the implied-bilayer grid ECD system.

We examined the variation in the tilt and opening angles (presented earlier in Figure 3.8), RMSD, and RMSF values for the self-assembly of the implied-membrane ECDs averaged over all 64 domains. The primary interest here was to ascertain whether any significant structural change took place on contact between ECDs or during subsequent evolution of the emergent structure.

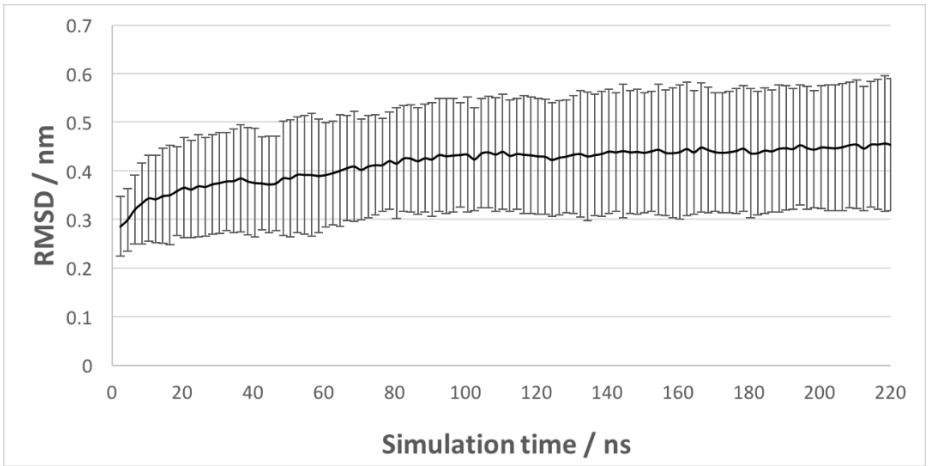


Figure 3.15: RMSD value with respect to simulation time for the grid (8x8) system. The error bars are also displayed (2σ) where sigma is the standard deviation divided by the square root of the population size (64).

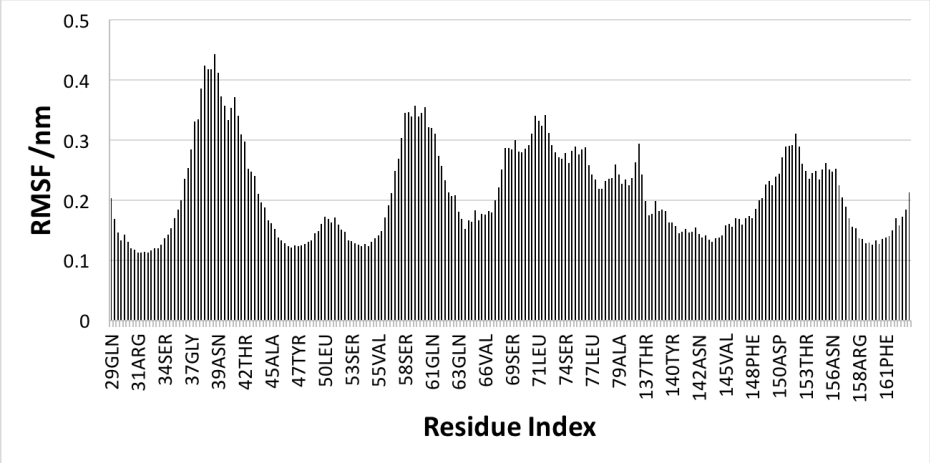


Figure 3.16: RMSF average values for each residue averaged over all 64 domain particles.

The β strands all show significant increase in both the tilt and opening angles, reflecting increased opening up of the ECDs with the β strands becoming more horizontal (parallel to the implied membrane). The opening up occurs relatively quickly within about 100 ns after which both the tilt and opening angles remain relatively stable. Most of the contact between ECDs occurs over the period 50-100 ns. The tilt and opening angle plots do not show any discontinuity over this or any later time period, indicating that interface formation does not induce any significant change in the structure of the ECDs.

The RMSD values converged to about 0.45 nm (presented with the associated error bars in Figure 3.15). The flexible regions as identified using RMSF values as a function of residue index (Figure 3.16) were pretty much identical to that for the single implied-membrane ECD, which were presented earlier in Figure 3.12.

3.3.4 Emergent strands reveal a plethora of interfaces

The ECD particle-particle interfaces within the emergent strands were characterised in terms of the respective rotation angles theta 1 (θ1) and theta 2 (θ2) (about the z-axis) of each of the two domain particles involved in any given dimer interaction, using python code coupled with the MDAnalysis library (Gowers et al., 2016). Characterisation of the interface requires the full 0-360° degrees rotation of each of the two ECD particles. The centre of rotation and the reference vector defining the rotation angle (see Figure 3.17) were chosen to enable direct comparison of the characterised interfaces with those previously reported using CG models of claudins embedded in a lipid bilayer (Irudayanathan et al., 2017, Irudayanathan et al., 2018). The centre of rotation was taken to be the alpha carbon of residue 137 (threonine) as this residue is the closest to the centre of mass in the x-y plane of the whole claudin particle. The reference rotation vector was defined by the vector connecting the centre-of-rotation alpha carbon and the alpha carbon of residue 29 (glutamine). The rotation angle was defined with respect to the vector connecting the centres of rotation of the two interacting ECDs, being zero when the two vectors are aligned.

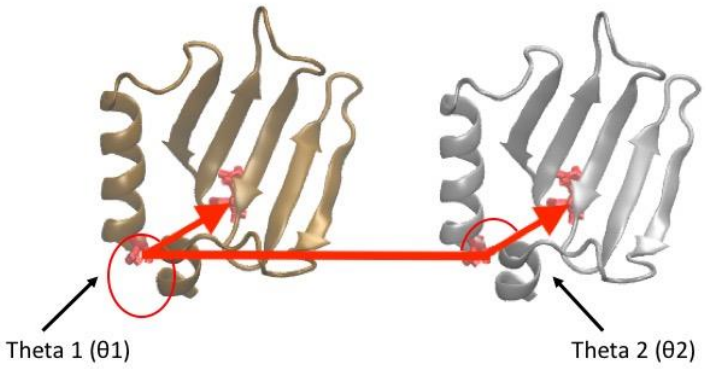


Figure 3.17: Definition of the rotation angles θ1 and θ2 to characterise ECD-ECD dimeric interfaces.

The distribution of rotation angle for pairs of interacting particles when in close proximity (separation distance being less than 3.5 nm) are presented in Figure 3.18. The distribution gives the frequency of a specific combination of θ1 and θ2 being observed. It is notable that the statistics here are poor, as we have averaged over only the last configuration of the trajectory. The interactions between the domain particles are strong and once they come together they become kinetically locked. Hence, one cannot characterise any latter part of the trajectory as having equilibrated. Consequently, averaging over many configurations of the latter part of the trajectory would be inappropriate as these configurations would be highly correlated. The rotation angle distribution shows a wide range of values, reflecting the wide variety of observed interfaces. Whilst the statistics are poor, there are some hotspots. Comparing the distribution with that reported by others (Irudayanathan et al., 2018) reveals that there are regions that are in common, such as the values of (θ1 ~ 130°, θ2 ~ 70° degrees) and of (θ1 ~ 90°, θ2 ~320° degrees). Further, we also concur with respect to 'un-preferred' regions, specifically when (θ1 ~180°, θ2 ~180° degrees).

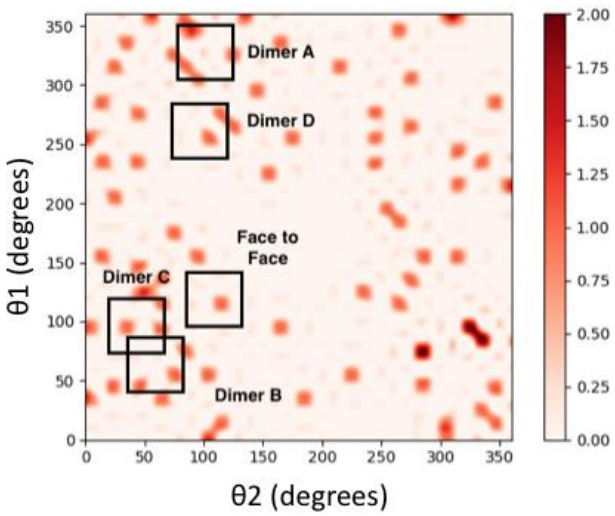


Figure 3.18: Frequency (colour-coded) of observing a specific interface between two ECDs as a function of their respective rotation angles θ1 and θ2 (as defined in Figure 3.17) The colour bar is displayed on the right-hand side of the graph.

The four notable, frequently observed dimers, which have also been observed by others (Zhao et al., 2018, Irudayanathan et al., 2015, Irudayanathan et al., 2017, Irudayanathan et al., 2018), are referred here as A, B, C and D and are shown in Figure 3.19 and in terms of interface angle distributions in Figure 3.18. Dimer A is frequently observed in the simulations and has also been identified as a predominant interface by others (Irudayanathan et al., 2018); energetically it has been considered to be less favourable compared to the others (Irudayanathan et al., 2015). Dimers B and C in gross terms have a similar orientational relationship with the contact region comprising mostly the back area of their second loops (Figure 3.19). Dimer D has a structure similar to the linear interface proposed by the Suzuki model, which was derived from the crystal lattice (Suzuki et al., 2014).

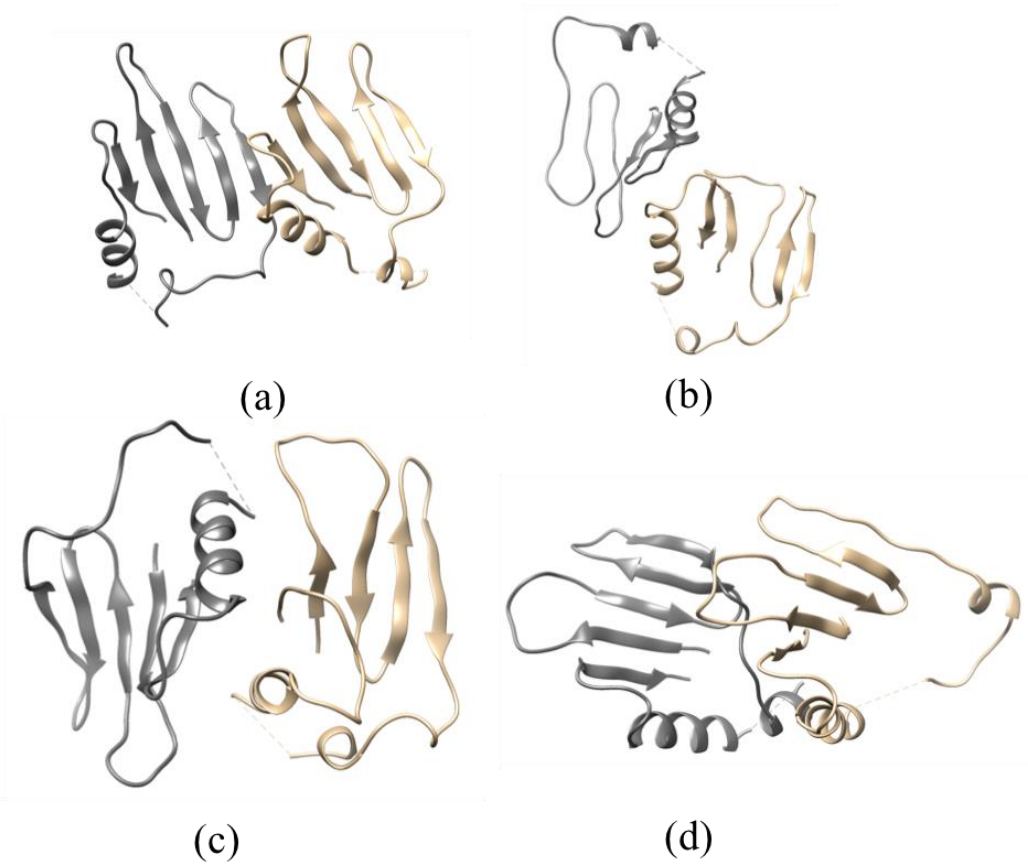


Figure 3.19: The frequently observed dimers A, B, C and D respectively in cartoon representation, with each monomer presented in a different colour.

3.3.5 ECD-ECD interactions are characterised by large binding free energies

The self-assembly simulation reveals that there is little scope for optimisation or dynamic re-arrangement of the emergent strands as the ECDs become kinetically locked in their interaction with neighbouring ECDs. This suggests strong binding interactions. To confirm this, we carried out PMF calculations using umbrella sampling with a view to estimating the binding free energy between any two interacting ECDs. The potential of mean force profiles (the free energy change as a function of separation distance) for each of the predominant interfaces A, B, C and D, and the Suzuki 'face to face' dimer are presented in Figure 3.20. The profiles show a characteristic minimum that defines the local optimum configuration (for the particular interface) and then gradually increase with increase in ECD-ECD separation, eventually plateauing when the particles no longer interact in any significant way (when $\xi \sim 3.5$ nm). The binding free energy is given by the free energy difference between the plateau and the minimum. The minor peaks in the free energy profiles in principle represent barriers that are probably exacerbated by convergence issues given that the ECDs are very large structures and sampling of rotational degrees of freedom is relatively poor. The locations of the minimums (with respect to the separation distance) vary as the ECDs are not rotationally symmetric and the separation distance based on the centre of mass varies depending on the relative orientation.

The estimated binding free energies are tabulated in Table 3.4 and are generally large (-21 to -58 KJ mol^{-1}) which equates to 8-22 k_BT at 310 K, where k_B is Boltzmann's constant. Note that the 'face-to-face' interface is also characterised by a strong interaction of -42.8 KJ mol^{-1}. The significance of the binding free energies being at least an order of magnitude greater than k_BT (the thermal energy) is that once the ECDs come together they are unlikely to unbind again. The implication is that the formed claudin strands are not in a thermodynamic equilibrium state but rather the structures are kinetically trapped. This begs the question why should nature favour strong interactions between the component proteins in tight junctions? What is clear is that TJs based on a claudin-claudin binding free energy of a few k_BT would readily be prone to failure.

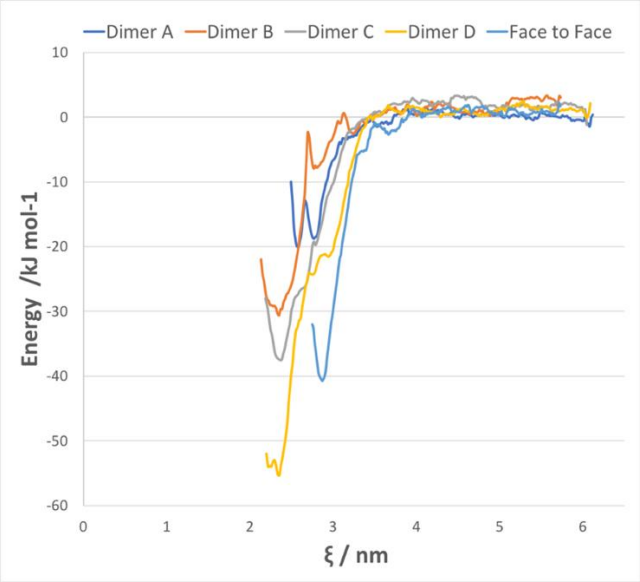


Figure 3.20: Free energy of binding between two ECDs as a function of separation distance ξ (based on centre of masses) for the frequently observed dimers A, B, C and D and for the Suzuki face-to-face model.

Note that the highest binding free energy is depicted by Dimer D (-57.8 KJ mol^{-1}) whilst the lowest is given by Dimer A (-21.4 KJ mol^{-1}). Dimers B and C both show a mid-range value which is consistent in that both of these dimer arrangements are variations relative to the other (see Table 3.4).

Table 3.4: Binding free energies determined using potential of mean force umbrella sampling and the MM/PBSA method for the frequently occurring dimeric interfaces. The table also shows the total number of contacts formed between the monomers and the solvent accessible surface area.

	ΔG / KJ mol^{-1}	**ΔG / k_BT**	**MM/PBSA**	**No Contacts**	**SASA (A^2)**
Dimer A	-21.4	-8.3	-166.4	83	11814
Dimer B	-33.9	-13.2	Unfavourable	86	13067
Dimer C	-40.8	-16.8	-20.9	65	12371
Dimer D	-57.8	-22.4	-48.1	88	12204

In comparison, the binding free energies reported in literature for the whole claudin molecules are reported to be in the range of about -60 to -180 KJ mol^{-1} for some similar dimeric interfaces (Irudayanathan et al., 2015). However, it is pertinent to note that these literature energies are for the binding between two whole claudins that include interactions within the transmembrane region.

Table 3.4 also displays the number of contacts formed between the monomers and the solvent accessible surface area (SASA) as indicators of stability of the dimer configuration. The number of contacts was calculated with the UCSF Chimera tool FindContacts and Dimer D shows the highest number of contacts formed between the monomers.

3.3.6 Certain key residues play a significant role in stabilising dimers

The atomistic resolution enables us to characterise in detail the particular residues that interact strongly and hence play a key role in stabilising the dimers (Figure 3.19 and Figure 3.21). On the basis of a distance criteria, the key residues involved are phenylalanines (67, 148 & 161), valines (145 & 155), leucines (70 & 73), tyrosines (35, 149 & 159) which exert hydrophobic character; lysine (65) and arginine (158) with positively charged side chains, and aspartic acid (150) with its negatively charged side chain. Specifically, Dimer D that was characterised as the most stable, reveals leucine 70 and 73 from one monomer and phenylalanine 161 and arginine 158 from the other monomer as 'hot spots' due to the increased number of contacts formed with multiple atoms from the opposing monomer. Also, Dimer C is the only one that shows a salt bridge formation (see Figure 3.22).

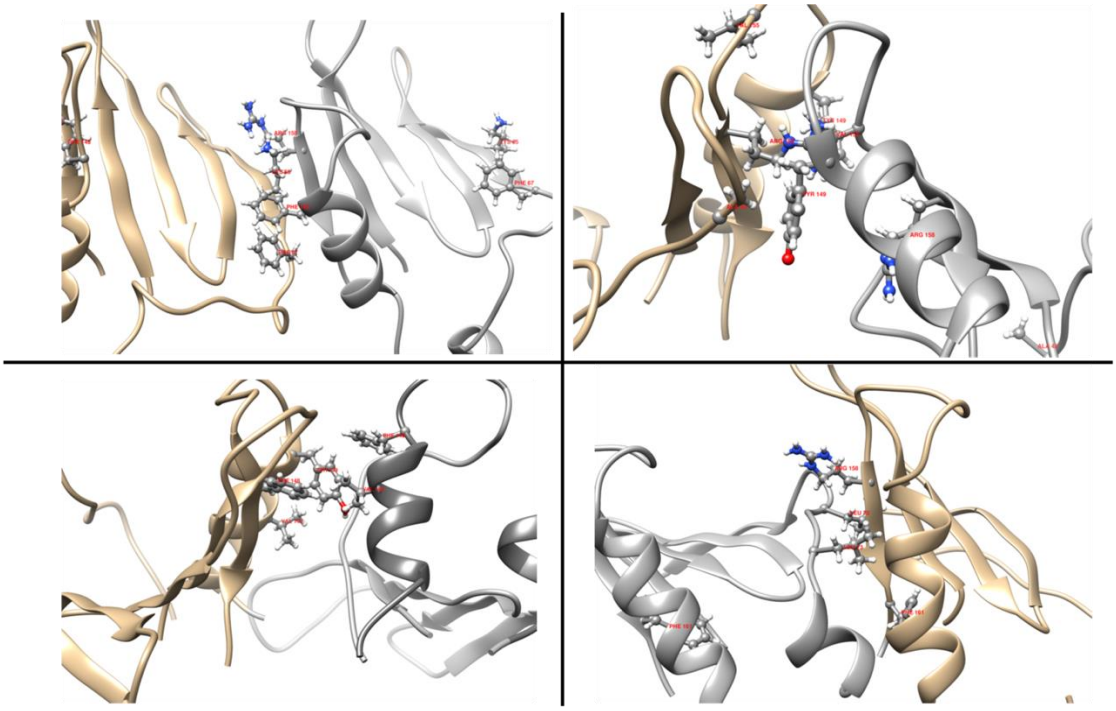


Figure 3.21: The 'key' residues involved in dimers A, B, C and D, illustrated in a ball and stick form and labelled with the name of the residue and its chain specifier (rendered with UCSF chimera (Pettersen et al., 2004)).

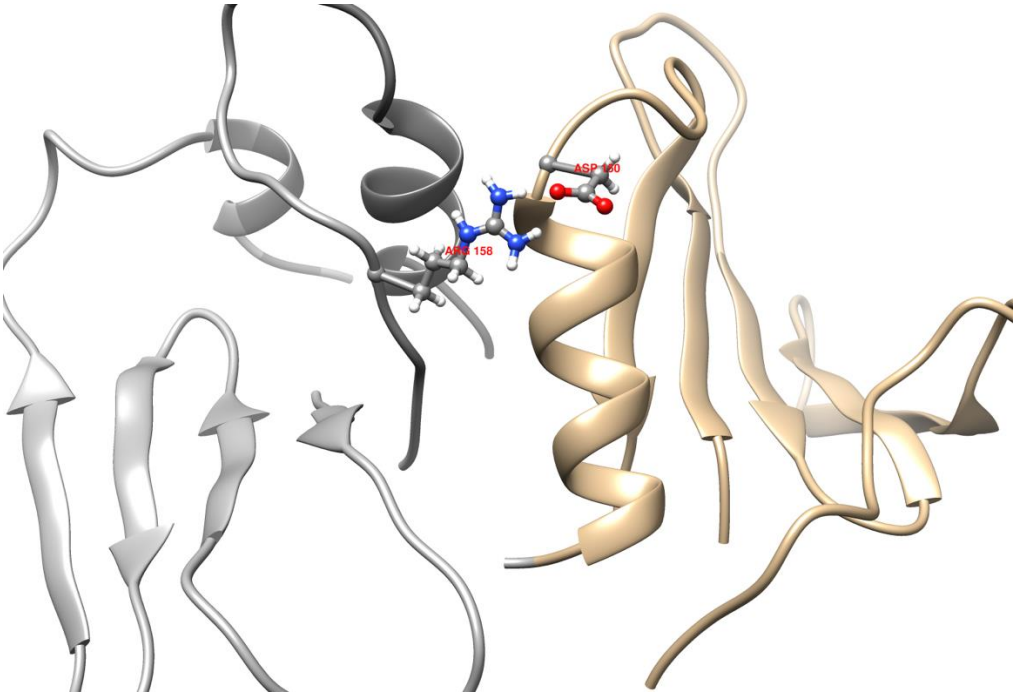


Figure 3.22: The salt bridge formed by aspartic acid 150 and arginine 158 in Dimer C. The residues are shown in ball and stick form while the rest of the loops are depicted in cartoon representation.

To quantify the energetic significance of the interactions, the binding free energy for the identified dimers was estimated with a view to extracting a breakdown of the free energy contribution for each residue using the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) methodology (Kumari et al., 2014). The MM/PBSA calculation is an approximation (unlike the PMF calculations) but is able to partition the free energy to the individual residue contributions. Given this, we utilise the MM/PBSA free energies for the residue-residue interactions (Figure 3.23) with caution making inferences only for gross distinctions.

The MM/PBSA binding free energies are presented in Table 3.4 and it is clear that the values obtained differ from the ones calculated with PMF previously, including their rank order. Given the assumption made from the MM/PBSA method, a good match was not expected but the rank order was.

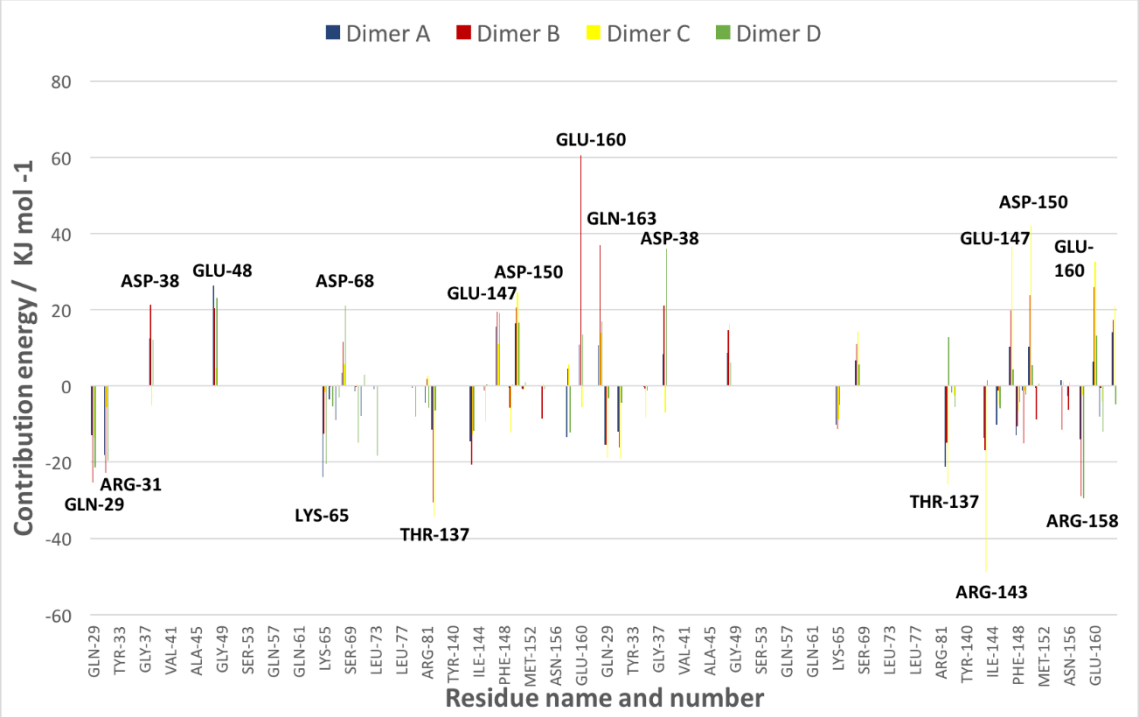


Figure 3.23: Free energy contributions of particular residues to the interface interaction for the frequently observed dimers. The important residues and their indexes are highlighted in bold letters.

The residues characterised by large negative free energy contribution values to the dimer binding include glutamine (29), arginine (31), lysine (65), threonine (137) and arginines (143 & 158) (see Figure 3.23). These residues that act as 'hot spots' have indeed been highlighted as key positions in experiments (Hou et al., 2010, Piontek et al., 2017, Veshnyakova et al., 2012). Conversely, residues that exhibit positive contributions, meaning they did not participate in binding, showed a redundant role in claudin mutation experimental studies (Veshnyakova et al., 2012). Position 65 is emphasised as an important determinant of the charge selectivity of the TJ pore: mutation studies of claudin-17 have shown that the anion selectivity that it conveys, depends on the positive charge at position 65 (Krug et al., 2012a). Also, Hou et al. (Hou et al., 2010), demonstrated that claudin-4 forms anion selective channels because of the lysine (+) residue at the same position. Moreover, Piontek et al. (Piontek et al., 2017), found that the charge of lysine 65 in claudin-1 is necessary for TJ strand formation, because claudin-1 mutants failed to display tight junctions like strands. In the same study, it was concluded that mutating the residues of claudin-1 at positions 48, 53 and 68 does not impede the formation of TJ strands, a result which agrees with the redundant role of these residues in our aggregation study.

In Appendix A2 the alignment between claudins-1, -2, -4, -15 and -17 is presented and more sequence alignments between claudins can be found in literature (Krause et al., 2008, Günzel and Alan, 2013).

3.4 General Comments and Conclusions

Atomistic simulations of a large assembly of claudin particles embedded in a membrane are currently not feasible due to time- and length-scale limitations of unbiased MD simulation. Consequently, previous simulations have had to resort to CG models with non-atomistic resolution to yield a low resolution picture of how claudin particles aggregate (Irudayanathan et al., 2017, Irudayanathan et al., 2018). The atomistic studies carried out prior to the current simulation (Alberini et al., 2017, Samanta et al., 2018) have looked at the stability of the pre-formed Suzuki model rather than self-assembly. In this chapter, we have re-visited the problem by simulating the interaction of isolated ECDs at an atomistic level in an implied membrane. Clearly for this approach to have validity, the interaction between the claudin particles must be primarily through their

ECDs (the 'torch heads') rather than their transmembrane region (the 'torch body'). The ECDs do have a larger diameter and protrude beyond the transmembrane region, although not for the entire circumference. Indeed, the literature does indicate that many of the important pair-wise claudin interactions are dominated by the ECD, especially in complex bilayers (Irudayanathan et al., 2015).

The ECD atomistic simulations, in line with earlier CG simulation studies of full claudin particles in a lipid bilayer, reveal rapid aggregation and assembly of the ECDs, to yield a network strand structure with strong ECD-ECD interactions. The emergent branched, network strand structure, resembles electron micrographs of tight junctions. The cis dimers observed show considerable variety in term of the relative orientation angles of the claudin particles that characterise the interacting interfaces. Free energies of binding indicate strong interactions suggesting that the emerging structures are out of equilibrium, being kinetically locked. The redundancy in interaction interfaces explains (confirming an earlier suggestion (Zhao et al., 2018)) why the claudins form a cross-linked network of strands; a single specific interface would result in either dimers or a linear strand formation. The primary driving force for aggregation is the hydrophobic interaction resulting from the numerous hydrophobic residues on the ECD, with leucine 70 and 73 and phenylalanine 161 being ‘key’ residues in Dimer D. Dimer D was found energetically more favourable according to PMF calculations and also resembles the proposed linear interface based on the Suzuki model. Note that the second proposed interface from the Suzuki model, namely the face-to-face dimer, also showed high binding free energy in the study. Arginine 158 with a positively charged side chain contributed in all characterised dimers. The ECD-ECD binding free energies are large ranging from -21 to -58 kJ mol^{-1}, the implication being that the emergent network strands are likely to be kinetically locked and not at equilibrium. We acknowledge that molecular packing considerations between the transmembrane regions may be important for particular configurations. However, this may be a secondary consideration given that that driving force for aggregation arises mainly from the ECDs and there is little scope for significant re-arrangement because of kinetic trapping. The chemical potential (the driving force for aggregation) of the claudin transmembrane region is likely to be significantly lower (in relative terms) given that the transmembrane region is hydrophobic and is embedded in a hydrophobic lipid environment.

The emergent structure is not linear but rather a branched network, consistent with experimentally observed morphology and previous CG simulations. The network does not reveal any unique orientation angle(s) between dimers but rather a multiplicity of claudin-claudin interactions. A unique branching angle would imply a highly-specific claudin-claudin interaction. A number of the common dimer-interactions observed are near-identical or similar to those observed in the CG simulation of the full claudin particles by others, the rotation angles for the frequently observed interfaces being within the range of previous values.

The atomistic simulations enable focussing on the key residues, particularly in rationalising specific site-directed mutations of residues. For example, the simulations confirm the importance of the key residue lysine at position 65 (Piontek et al., 2017). Further, the simulations were able to confirm the lack of importance of other residues which have been reported to be redundant for the formation of strands in experiments using mutation studies (Piontek et al., 2017, Veshnyakova et al., 2012). Therefore, the results confirm the importance of atomistic resolution is such studies, being particularly relevant in strategies to develop therapeutic strategies wherein one might mutate particular residues (for instance via coded messenger RNA) to alter function by design.

4 Characterisation of the Trans-Interaction of Claudin-1 Particles

4.1 Introduction

The broader aim of this thesis was to elucidate the protein-protein interactions that give rise to the formation of the TJ strands and to further characterise the resulting molecular architecture including the pore structure. This will enhance our understanding of how to regulate and restore dysfunctional epithelial barriers in disease (Förster, 2008), and enable research to facilitate better delivery of drugs across epithelial and endothelial tissues e.g. across the intestinal epithelium and through the blood-brain barrier (BBB) (Deli, 2009, Foldvari, 2000, Gonzalez-Mariscal et al., 2005). In Chapter 3 we investigated how the extracellular domain (ECD) particles of claudins aggregate in the TJs, while embedded in a single *implied* lipid bilayer. This side-by-side (cis-)interaction between claudins resulted in a structure that resembled the electron micrographs of TJs and highlighted the plethora of interfaces that can form between the ECD particles. The atomistic MD simulations of the ECDs, coupled with potential of mean force (PMF) calculations, suggested that the proteins become kinetically locked and are not able to break their initial strong interaction and evolve into lower free energy architecture.

In this chapter, the 'head-to-head' (trans-)interaction between the ECDs was examined, which actually defines the barrier properties of the TJs. Although one could study the trans-interaction by embedding the proteins in two opposite bilayers, this approach has some issues. If the proteins are embedded in their respective bilayers, the water molecules found in the extracellular space (between the bilayers) should be pushed away when the proteins interact; this would require the formation of a water channel, which is challenging. Alternatively, in the case where the proteins are placed closely and hence the water region is minimal, it would imply that the system would be biased to form specific trans-interfaces. Instead of embedding the proteins in lipid bilayers, we

address the trans-interaction using a different approach employing again the concept of the implied bilayer and examining two different studies. First, the self-assembly of big grids of ECDs was investigated, and then in a rotation study, six systems each with a different relative orientation of the top ECD to that at the bottom ECD were systematically examined to sample configuration space.

4.1.1 Epithelial permeability barriers; The TJ sealing strands

The properties of diffusion barriers and channels, especially in the case of the paracellular channels are still poorly defined (Samanta et al., 2018, Alberini et al., 2018, Krug et al., 2012a). Nevertheless, in the case of paracellular permeability through the TJs, it is well documented that claudins constitute the TJ sealing strands by self-interacting within the same cell membrane and between adjacent cells to form channels or pores. The resulting pore structure is thought to facilitate the transport of small molecules through the paracellular route, which is parallel to the cells' membrane. As detailed in Chapter 1, the dysregulation of the barrier and channel properties of TJs have been implicated in several human diseases (Sawada, 2013) and the need to better understand the size and charge selectivity of the TJ pore is imperative.

TJs can physically restrict the movement of solutes through the paracellular space, whilst in some epithelial they allow some molecules of specific size and charge to go through (Anderson and Van Itallie, 2009). Since the morphology is not well defined, it is hard to interpret biophysical and physiological data, hence, the need for the full characterisation of the paracellular pore is crucial.

The work of the Tsukita laboratory was a landmark in the field of the molecular architecture of the TJs (Takeichi, 2006) and has enabled today's research on TJs to flourish. The group identified the first integral protein, the occludin (Furuse et al., 1993), clarified the fundamental role of claudins on the self-assembly of TJs (Furuse et al., 1998), and resolved the first crystal structure of a claudin, namely claudin-15 (see Figure 1.3) (Suzuki et al., 2014). The group also proposed a structural model for the architecture of the TJ pore, the so-called Suzuki model (Suzuki et al., 2015) that has contributed significantly towards the characterisation of the TJ pore (Figure 1.2). However, the detailed spatial arrangement of the proteins in the pore is not resolved and the validity of the proposed model has been questioned (Zhao et al., 2018).

The TJ strands contain claudins, occludin and other proteins (Cereijido and Anderson, 2001, Furuse, 2010b) and it has been suggested that the pores fluctuate between open and closed states; although the latter has not been demonstrated because we cannot directly measure TJ barrier function. The TJ pore structure is very dynamic by nature even in steady state (Shen et al., 2008, Weber, 2012). Fluorescence recovery after photobleaching analysis demonstrated that although claudin-1 was stable in the TJs other TJ proteins exhibit distinct exchange behaviors (Shen et al., 2008). Previously, the TJs were thought to be static structures but now a dynamic model seems to be a more accurate description, where the strands are believed to undergo constant remodelling (Weber, 2012, Shen et al., 2008).

The dynamic nature of the junction needs to be maintained both at the bicellular and tricellular contact sites. A bicellular junction is when two epithelial cells meet and form TJs towards their apical side. While in the case where three cells meet, they form tricellular junctions (vertical oriented strands) and tricellulin is the fundamental protein that regulates the seal between these cells (Ikenouchi et al., 2005). It has also been suggested, that there is a pore and leak pathway, a dynamic duo, which is based on the dynamic regulation of the TJ protein interactions and that the leak pathway is controlled by tricellulin while the pore pathway by claudins at the bicellular junctions (Shen et al., 2011). The pore pathway is a high capacity, charge selective pathway while the leak pathway allows limited passage of large molecules and does not show any charge selectivity (Shen et al., 2011). The former pathway, which is regulated by claudins, has an estimated diameter of 0.4 -0.8 nm while the latter one allows larger solutes and macromolecules to go through with a size limit of approximately 3-6 nm (Zihni et al., 2016).

The multi-protein nature of the TJ strands and the variety of different dynamic interactions that occur is a challenge to the field. Additionally, it is hard to differentiate between the cis- and trans- interaction that occurs in cell-cell contact sites experimentally (examples of methods used are co-localisation of endogenous proteins with EM, co-immunoprecipitation of overexpressed proteins and immunofluorescence staining or direct fusion with fluorescent proteins).

The transepithelial electrical resistance method (TEER) is used to characterise the permeability of TJ pores. It measures the flux of ions across the epithelium and provides

a direct assessment of ion permeability. Based on their TEER values, epithelia sheets are characterised as ‘tight’ or ‘leaky’. To further characterise the charge selectivity of the junction the diffusion potential technique can be used where we measure the effects of unilateral ion substitution or measure the fluxes of tracers through the epithelia cellular sheets (Shen et al., 2011, Cereijido and Anderson, 2001). The diameter of the claudin-15 pore was found recently to be 0.4 nm (4.2 Å) in the narrowest region and 0.8 nm (8.1 Å) in the widest region with MD simulations (Samanta et al., 2018). In the case of claudin-2, which is highly expressed in ‘leaky’ epithelia, the pore diameter was found to be at least 5.8 Å (Angelow and Alan, 2009a) and it has been shown that it allows water molecules to go through. Specific residues of mainly the first extracellular loop were also highlighted as important for the charge specificity that the pore shows (Colegio et al., 2003).

Note that the transport of solutes parallel to the cell membrane does not require energy consumption from the cell because it is driven by a concentration gradient. Therefore, TJs enable the cell to couple the transcellular and paracellular pathways in an energy efficient way (Günzel, 2017).

To conclude, there is limited data on the exact spatial arrangement of the TJ proteins, partly because of the branching and dynamic nature of this cell apparatus (cross-linked strands). We know that the strands regulate the passage of molecules through the TJs possibly because of the pores in or between the strands.

4.1.2 Models for the molecular architecture of the TJ pore

An important step towards the elucidation of the pore structure was the proposed Suzuki model (see Figure 1.2) and a handful of studies have been carried out in order to examine its validity (Suzuki et al., 2015, Alberini et al., 2017, Zhao et al., 2018, Samanta et al., 2018). The model as previously discussed, is based on the formation of two (cis-) interfaces between claudins embedded in the same membrane, namely the linear interface (see Figure 4.1) and the ‘face to face’ interface (see Figure 4.2). The same type of interaction occurs on the adjacent cell and the cells then interact through the extracellular space (trans-interaction) forming a barrier with pores in the paracellular route.

The model, which was supported by cysteine cross-linking data and agrees with electron microscopy (EM) images, has received acceptance but other conformations are indeed perceivable as highlighted by Zhao et al. (Zhao et al., 2018) and demonstrated by MD studies where multiple interfaces between claudins were formed (Irudayanathan et al., 2018, Irudayanathan et al., 2015, Irudayanathan et al., 2017). Most of the work regarding the architecture of the TJ pore and characterisation of the trans-interaction has been related to this proposed model and MD simulations have been carried out to confirm (or otherwise) the proposed arrangement.

The Suzuki model comprises multiple claudin protomers arranged in a double row conformation that interact in both a cis- way, i.e. in the same cell and in a trans-way spanning the extracellular space. The proposed cis-interfaces, as previously mentioned, are two: a linear cis- interface between a hydrophobic conserved residue of the small helix (methionine 68) and a hydrophobic pocket at the back of the β-sheet (conserved residues of the third transmembrane region-TM3 and the extracellular loop 2-ECL2); and an intermolecular interaction between two monomers through their β4 strands at the edge of the extracellular domain (Suzuki et al., 2015). The latter one is called the 'face to face' interface and the former one the linear interface (see Figure 4.1 and Figure 4.2). The resulting structure looks like a 'half pipe' and subsequently, an opposing cell could associate through trans- interaction with the neighboring cell and create β barrel like structures (see Figure 1.2). Note that there are areas of the extracellular loops missing from the model that may cause steric hindrance when introduced.

Finally, there are other models proposed in the literature for the TJ pore. In the first one, the arrangement of claudins is different, because they form different cis-interfaces, and this results in a different pore structure (pore II) that was consistent with biochemical results (Irudayanathan et al., 2017). Another alternative symmetric model was also proposed for the tetrameric claudin unit, with a modified linear interface (Piontek et al., 2017). There was also the hypothesis that claudins form hexamers such as the connexins at the gap junctions. However, the predominate one is the model proposed by the Suzuki group.

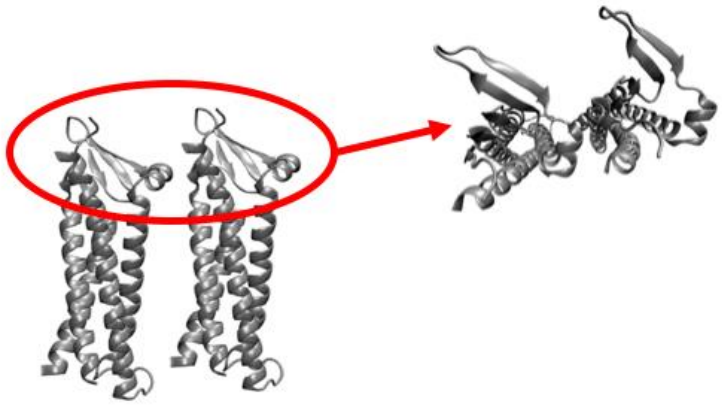


Figure 4.1: The linear interface of the Suzuki model where the monomers are coloured silver and are in cartoon representation. On the right-hand side of the figure there is a focus on the ECLs from a top view.

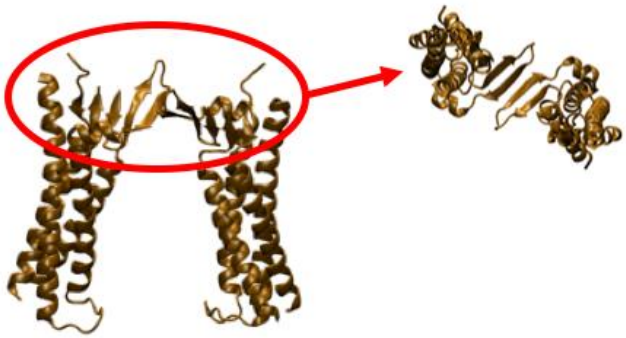


Figure 4.2: The 'face to face' interface of the Suzuki model where the monomers are in cartoon representation and coloured gold. On the right hand side of the figure there is a focus on the ECLs from a top view.

4.1.3 Molecular dynamics simulations to elucidate the trans-interface

The accuracy of the Suzuki model was examined in computational modelling studies and the structure was found stable with minor refinements (Alberini et al., 2017,

Alberini et al., 2018, Samanta et al., 2018). In a different study the model was found to be unstable as the number of added protein monomers increased (Zhao et al., 2018). Alberini et al., found that the complex is overall stable and that the pore size fluctuates, so they strengthen the validity of the Suzuki model; while Zhao et al., identified a new (cis-) interface that showed a 17^0 degrees lateral rotation between the monomers. Note that, the starting configuration could bias the system towards a specific behaviour as the energy barriers need to be overcome to converge towards a stable configuration may be too high for the system to cross within the timescales of MD. Additionally, some studies have examined the trans-interface between claudins using docking methods (Irudayanathan et al., 2018, Piontek et al., 2017, Irudayanathan et al., 2017). Both methods, namely brute force MD and docking, have their limitations, therefore examining the arrangement of claudins in the pore from a random initial organisation is advised.

Here, we explore how claudins arrange themselves to form a barrier with built-in pores without biasing their initial arrangement as in previous studies. This is based on the finding from the study of Zhao et al., that the number of proteins modelled according to the Suzuki model can influence the stability of the simulated strand and moreover due to the cross-linked nature of the observed strands. A single linear formed interface such as the one proposed cannot satisfy the observed nature of the cross-linked strands (Zhao et al., 2018, Alberini G, 2017, Samanta et al., 2018). We employ MD simulations of claudin-1, which is an indispensable protein for the epidermal barrier, with a view to investigating the way these proteins interact between neighbouring cells, the so-called trans-interaction (see Figure 4.3) that creates the TJ barrier.

In characterising the trans-interaction, our primary goal was to identify the preferred interaction sites of the ECDs as they approach each other, their preferred relative orientation, the relative separation distance between them and the relative displacement of the ECDs. Here, we start by carrying out self-assembly simulations of two adjacent layers of claudin-1 embedded in their respective *implied* membrane, with each layer comprising a grid of separated and randomly oriented claudin particles. We then carry out a systematic rotation study wherein one ECD is systematically rotated and placed on top of another ECD, to see how the trans- preferences play out and whether the identified preferred orientations from the grid systems are indeed observed.

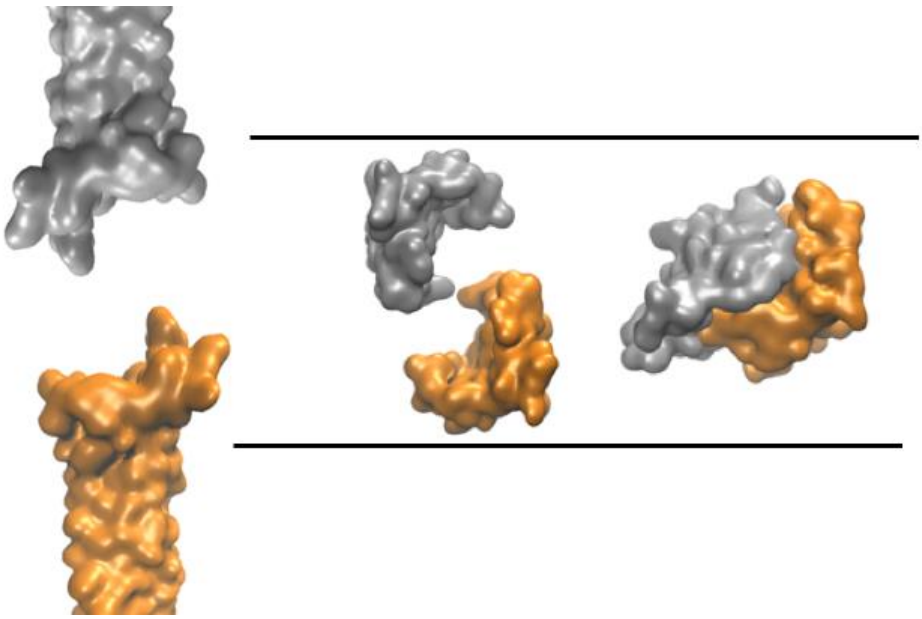


Figure 4.3: The trans-interaction between claudins. Each monomer is differently coloured and the black lines are indicative of the extracellular space boundaries. The figure shows two whole proteins and their respective ECDs truncated. The first truncated ECDs are at the beginning of the trajectory and the second set are the end frame (phi0 system).

The systems studied focused on the trans-interface and only the ECD of the claudins was investigated, since it is the determinant of the trans-interaction and also because it is technically challenging to assess this type of interaction while the two proteins are embedded in a lipid bilayer (see Figure 4.3). This is because the initial distance between the bilayers should be large and the two proteins wouldn't be able to pull the bilayers closer, pushing all the water molecules away; a water channel should be created. Thus, in the case of a system composing two proteins embedded in their respective lipid bilayers with a small separation distance, the starting position would be a strong bias to the resulting trans-interacting structure.

The idea of the ECDs restrained on a x-y plane which serves as an implied cell membrane was introduced in Chapter 3 and the same protocol was followed in this study for both the self-assembly simulations starting from grids of ECDs, as well as, with the considerably smaller systems with different initial arrangement (rotation study). The results from both studies highlighted the unstructured, variable regions of the ECDs (V1

and V2 area) as the ‘key’ players of the trans-interaction, which display mostly hydrophobic character. The second study also demonstrated that another ‘key’ area is the one close to the small helix in the first loop. The binding energies pointed out that the ECDs were more stable in specific orientations compared to others, and that the ECD particles interact stronger when they reduce their penetration depth and form more contacts. Moreover, due to the atomistic detail at the interaction sites, we were able to identify the important residues that interact and form the trans –interfaces of claudin-1.

4.2 Methodology

The ECDs are the decisive factor of the trans-interface. Here, we isolate the ECD of claudin-1 and carry out MD simulations to define their self-assembly in big grids and also look at their interaction more systematically in a rotation study. The second study offers the benefit of more freedom to the ECDs (they rearrange themselves more freely and penetrate deeper in the opposing plane) and is set up in order to explore systematically the orientation sampling.

4.2.1 Atomistic protein model of claudin-1

As in Chapter 3, homology modelling was used to construct an atomic resolution model of the TJ protein claudin-1, using as templates the three known crystal structures of claudins (Suzuki et al., 2014, Saitoh et al., 2015, Shinoda et al., 2016) and the online platform I-TASSER (Roy et al., 2010) (see Appendix A2). Since the focus was on the trans-interaction that forms the TJ barrier structure, we have isolated the part of the protein that protrudes from the cell, i.e. the extracellular domain of claudin-1. There are two loops that comprise the domain and their primary structure was presented in Table 3.1. Briefly the terminal atoms of the two loops were restrained to lie on a surface and the distance and relative angle between them was maintained by applying harmonic restraints. All simulations were carried out in an aqueous environment under physiological ionic concentration (0.15 M NaCl) without any lipids present (implied bilayer). We acknowledge that the lipids could play a role in the cis-interaction of claudins, however the trans-interaction is driven only by the interactions between the residues located at the ECDs as they alone protrude from the membrane and seal the intercellular space.

4.2.2 Self-assembly of two adjacent layers of claudin particles embedded in their respective (implied) bilayers

The initial systems studied were two replicas of claudin grids but differed in the applied position restraints (Figure 4.4). Specifically, each system comprised two layers of grids of 6x6 claudin domains, so in total 72 ECDs, with a centre of mass separation distance of 5 nm on the same plane, while the distance between the layers was also approximately 5 nm. The ECDs were restrained to lie on a x-y plane facing each other in a way two neighbouring cells would interact (see Figure 4.4). The ECDs that lie on the same plane, were randomly rotated on the z-axis to reduce the bias of the system towards a specific behaviour that would indicate a preferred interaction site. The simulation box was filled with water in a way that it prevented the periodic images in the z- direction to interact. The z component of the box was set to approximately 12 nm which is 2-times the ECD height (approximately 6 nm) plus at least four times the cut-off distance for vdW interactions. This resulted in a system of 72 ECDs solvated in 420179 water molecules and 2576 ions giving a total of 1353905 atoms. The difference between the two grid systems was the applied force which kept the terminal alpha carbons of the top grids on the same plane (force constants 500 KJ mol^{-1} nm^{-2} and 200 KJ mol^{-1} nm^{-2} referred to as **grid500** and **grid200** systems respectively from hereafter). The bottom ECDs were restrained on the x-y plane with a 'spring' having a force constant of 1000 KJ mol^{-1} nm^{-2} in both systems. After energy minimisation using the steepest descent algorithm, and equilibration in a constant temperature and pressure, the ECDs were free to interact amongst each other both in the same plane (cis-interaction), as well as between neighbouring planes (trans-interaction).

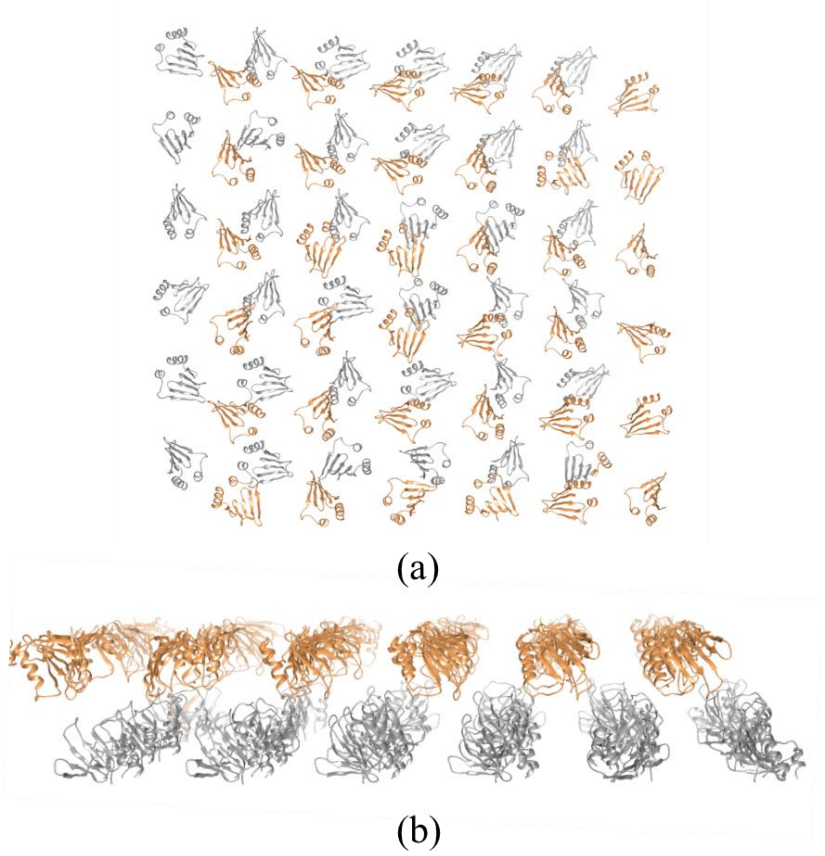


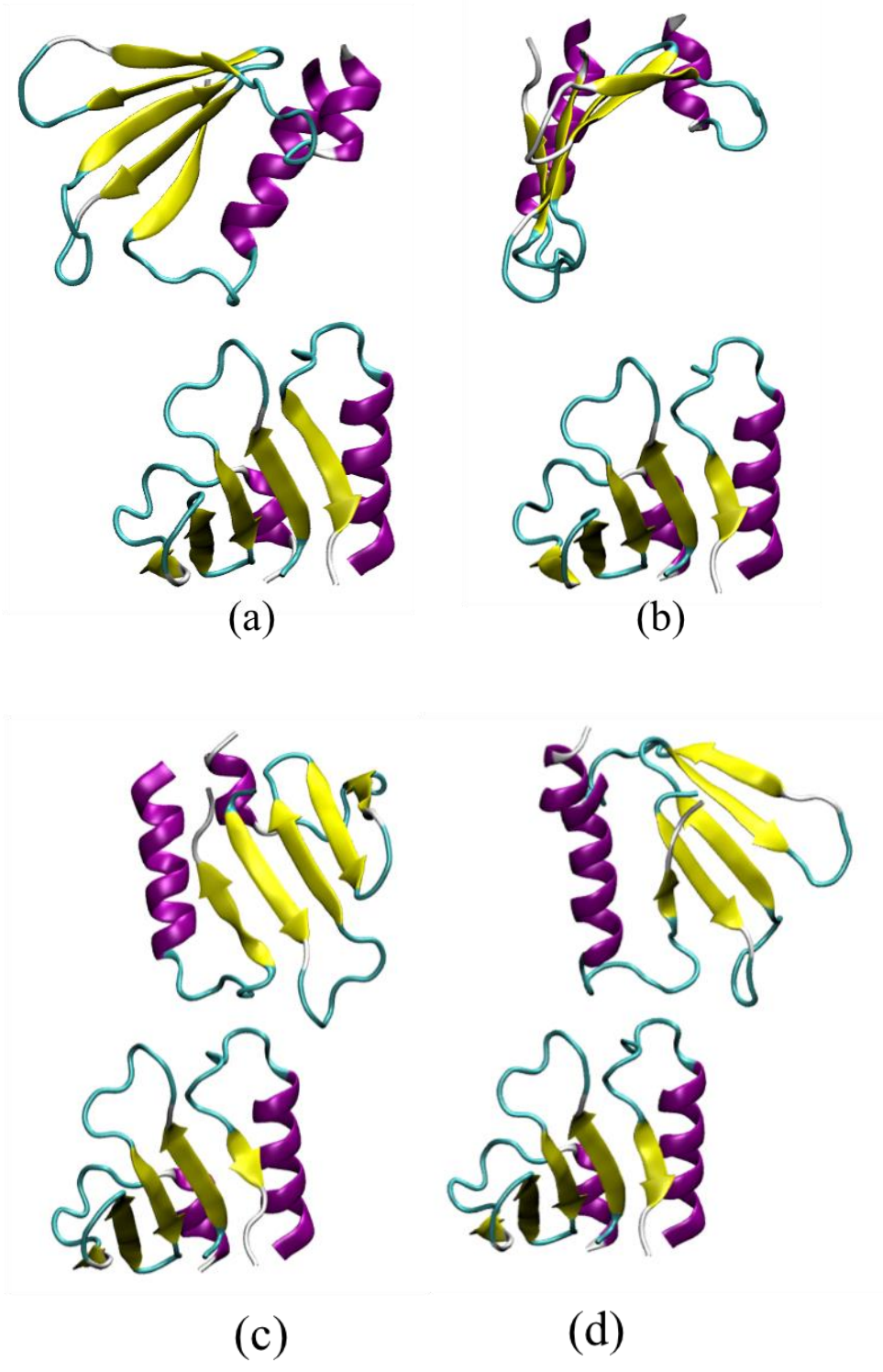
Figure 4.4: The grid of 72 ECDs rendered with UCSF Chimera in cartoon representation (Pettersen et al., 2004). The top layer of ECDs is coloured in gold and the bottom layer in silver. The grid is displayed from the top (a) and side (b) view to enhance the understanding of how the ECDs are placed and face each other. From the top view we notice that the ECDs are not directly on top of each other but rather in the empty spaces between them.

4.2.3 Systematic sampling of the trans-interaction between two claudin-1 particles

The next study, namely the rotation study, was based on the observation that the restraints applied on the big grids of loops were large, and thus, the ECDs did not penetrate deeply in the opposing layer. In the rotation study, we simulated two ECDs in different states of rotation, facing each other in a trans-way. The top domain particle was rotated every 60 degrees before placing it on top of the other, in order to sample more trans-orientations (see Figure 4.5). Each system is named based on the initial rotation angle of the top ECD so, this results in 6 systems in total named phi0 (φ=0), phi60 (φ=60), phi120 (φ=120), phi180 (φ=180), phi240 (φ=240) and phi300 (φ=300)

where the angle phi(φ) defines the rotation angle of the second ECD before it is flipped and translated on top of the first one. Note that the ECDs show a slightly varying initial distance between them in the phi0-phi300 systems, to prevent overlap of protein atoms after the placement of the top ECD (see Figure 4.5). In this case, the force constant applied on the alpha carbons was 250 KJ mol^{-1} nm^{-2} and the two ECDs were solvated in a range of 19-24 thousand water molecules.

It was challenging to evaluate the most efficient position restraints, because the ECDs could jump out of the central simulation box and form undesirable contacts, e.g. interact through the area underneath them which is supposed to be the transmembrane region of the protein. For instance, during the trajectory one ECD would reduce the distance between itself and the periodic image of its opposing ECD and thus, interact with the periodic image and not the ECD that was placed on top of it. To prevent this, piecewise linear/harmonic distance restraints were introduced between the two ECDs in the form of upper distance limits (function 10 of bonds in GROMACS) (Abraham et al., 2015). The potential form of the restraints was approximated with a quadratic (harmonic potential) form under a specified lower distance limit (approximate 6 ± 0.5 nm depending on the initial distance value) but once the atoms moved further away (largest bound was approximately 7 ± 0.5 nm depending on the system) the potential was linear and thus, the force that brought them back strong (force constant 1000 KJ mol^{-1} nm^{-2}). Additionally, the position restraints that kept the ECD particles on the aforementioned line (along the z-coordinate) were implemented on the selected four alpha carbons of the terminal residues and had a force constant of 250 KJ mol^{-1} nm^{-2}. This resulted in the proper trans-interaction between the ECDs, such as between opposing cells, after the system was energy minimised and equilibrated in the *NPT* ensemble.


(a)
(b)
(c)
(d)

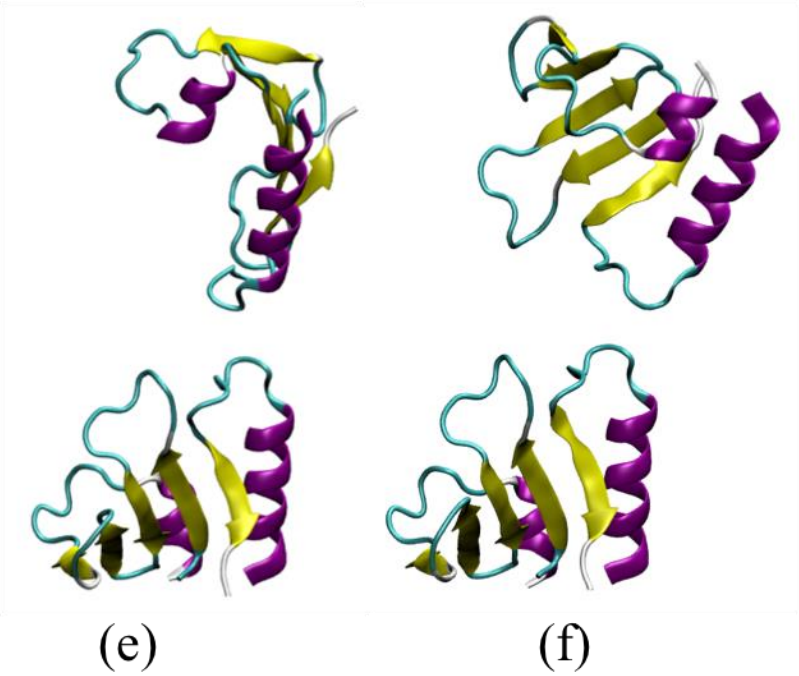


Figure 4.5: The simulation systems that comprise the rotation study; (a) phi0, (b) phi60, (c) phi120, (d) phi180, (e) phi240 and (f) phi300. The ECDs are in cartoon representation and coloured based on their secondary structure with alpha helices shown in purple and β-sheets in yellow. Unstructured areas are shown in cyan. Water molecules and ions are not displayed for clarity. In the pictures the position of the bottom loop was kept constant to better understand the initial placement of both ECDs.

4.2.4 Technical details of the molecular dynamics simulations

The MD simulations were performed using the program GROMACS (Abraham et al., 2015) (version 5.0.5 for the grids and 2016.3 for the rotation study) using the CHARMM 36 force field parameters (Bjelkmar et al., 2010). The water model was TIP3P which is compatible with the force field, while lipids were not present in the simulations because we had an implied bilayer. The timestep was set to 2 femtoseconds and the temperature and pressure were kept constant at 310 K and 1 bar with the Nose-Hoover thermostat and the Parinello-Rahman barostat respectively (*NPT* ensemble). The duration of the simulations was 200 ns for the big grids and 250 ns for each the

rotation study simulations to ensure convergence. Periodic boundary conditions (PBC) were applied and to calculate the long –range electrostatic forces we used the particle mesh Ewald summation method with a real-space cut-off at 1.2 nm, cubic interpolation, fourier spacing equal to 0.16 nm and a precision (Ewald -rtol in GROMACS) of 10^{-5}. The short range electrostatic and van der Waals cut-off distances were set to 1.2 nm, with a switch function for the vdW at 1 nm. Additional restraints that had the aforementioned force constants, were applied in order to restrict the movement of the domains on the initial 2-dimensional planes; by enforcing harmonic potentials on the four selected alpha carbons of the terminal residues (glutamine 29, arginine 81, threonine 137 and glutamine 163) (see Figure 4.6). Prior to the production phase all systems were energy minimised using the steepest descent algorithm and prepared through a two-step process; first an equilibration run in the canonical ensemble, *NVT* and then a continuation run in the isothermal–isobaric ensemble, *NPT*.

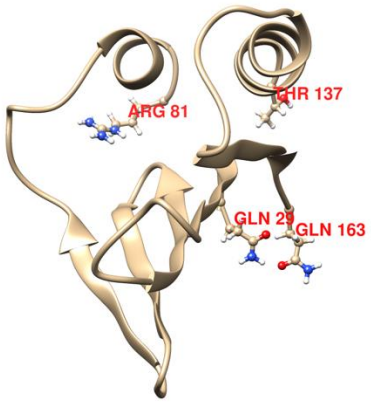


Figure 4.6: The ECD of claudin-1 in cartoon representation (coloured gold) from a top view. The terminal residues on which the force constant acts on are shown in ball & stick representation (atom colours). The name and index of the terminal amino acids is also displayed.

4.3 Results and Discussion

This section is divided into two subsections. First, we review the results with regards to the self-assembly studies of the large grids of ECDs where each system consisted of 72 domains and then the findings of the rotation study where 6 systems in different states were considered each consisting of two ECD particles (phi0, phi60, phi120, phi180, phi240 and phi300).

4.3.1 Self-assembly simulations of opposing ECD particles

4.3.1.1 Does the cis- interaction precede the trans one?

The domains initially placed on grids, quickly interact amongst each other on the same plane (intra-cellularly, cis-interaction) and by the end of the trajectory some have interacted with the opposing domains (inter-cellularly, trans-interaction). The resulting clustered ECDs resemble an elongated cross-linked network of particles as seen in electron micrographs of TJs, although they have not yet formed a single aggregate (Figure 4.7). By visualising the trajectory, it is evident that more ECDs have formed dimeric interfaces by interacting on the same plane and then these aggregates position themselves on top of the ones formed on the opposing layer. The cis-interaction happens quickly within the first 50 ns (see Figure 4.8 and Figure 4.9). Note that there are dense areas in the simulation box where the ECDs have clustered and interacted in both ways. However, it is suggested that the system is kinetically locked in this local free energy minimum and the associated energy barriers prevent the loops from breaking their initial interaction and rearranging themselves into a possibly more favourable arrangement. The ECDs interact quickly and their binding energy is higher than the thermal energy, preventing them from breaking their initial interaction.

The clustering ability of the two systems with the different applied position restraints (grid200 and grid500) is presented in Figure 4.8 and Figure 4.9. The number of atoms that formed each aggregate based on a cut-off distance (0.35 nm) were considered. The cut-off was the largest distance that two protein atoms can be apart and be regarded as part of the same aggregate. Furthermore, each graph breaks down the analysis in three different components: all the protein atoms (whole system) and the protein loops lying on the bottom and top plane. This categorising is merely for our better understanding, since there is no bottom and top layer in a system that is simulated under PBC.

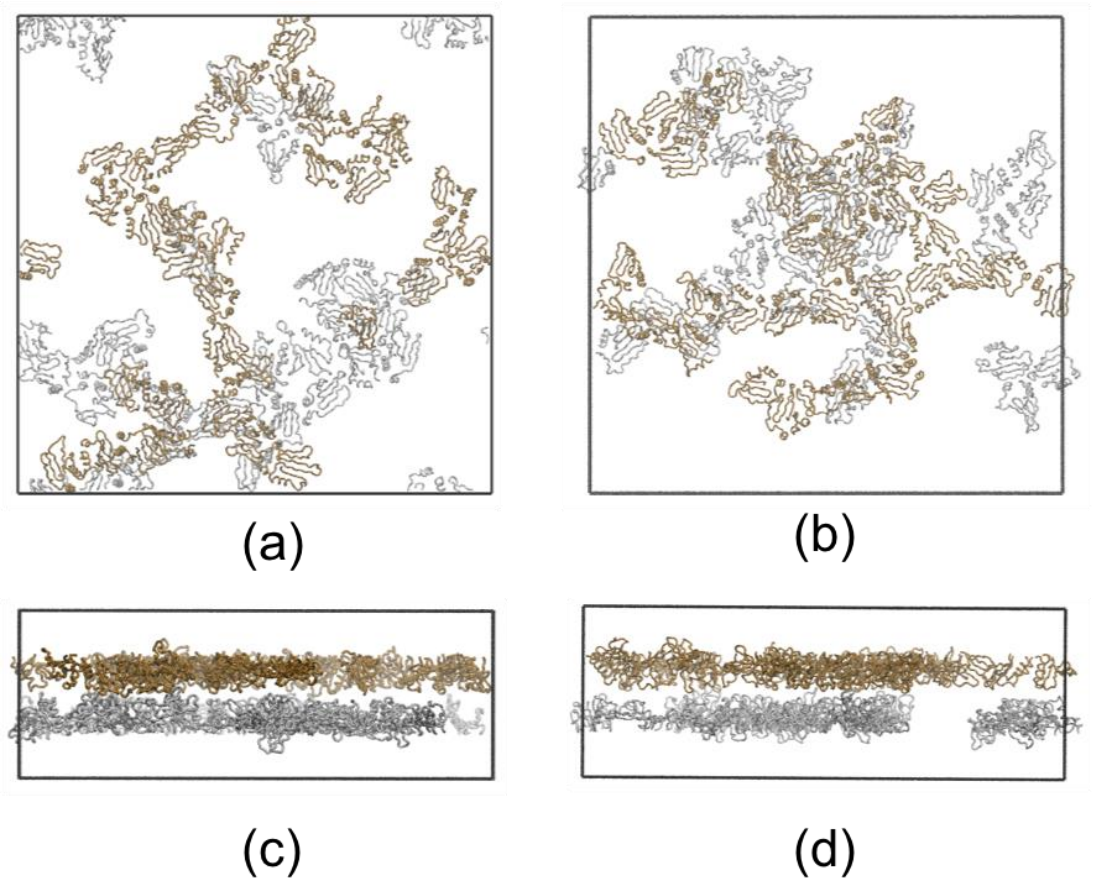


Figure 4.7: The end frame of the two grid systems composed of claudin ECDs. The top (a) & side (c) view of the system that has the weakest position restraints -grid200 and the top (b) & side (d) view of the system with the strongest position restraints - grid500. The end frames were rendered with VMD (Humphrey et al., 1996). The bottom layer ECDs are silver while the top are coloured gold. Ions and water molecules are not shown and the simulation box is displayed in black.

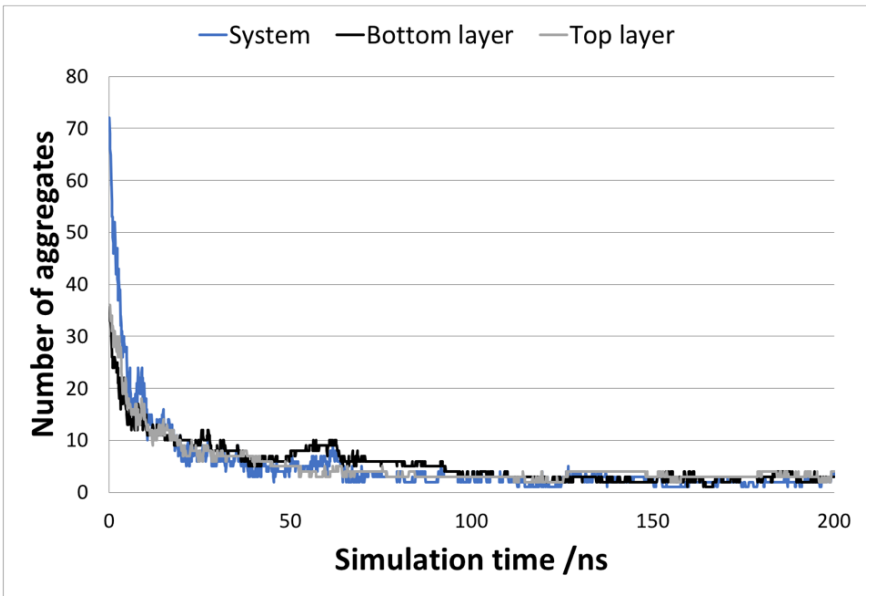


Figure 4.8: The number of aggregates with respect to simulation time for the grid200 system. The blue line represents the whole system (all protein atoms) while the black and grey lines represent the bottom and top layer of domains respectively.

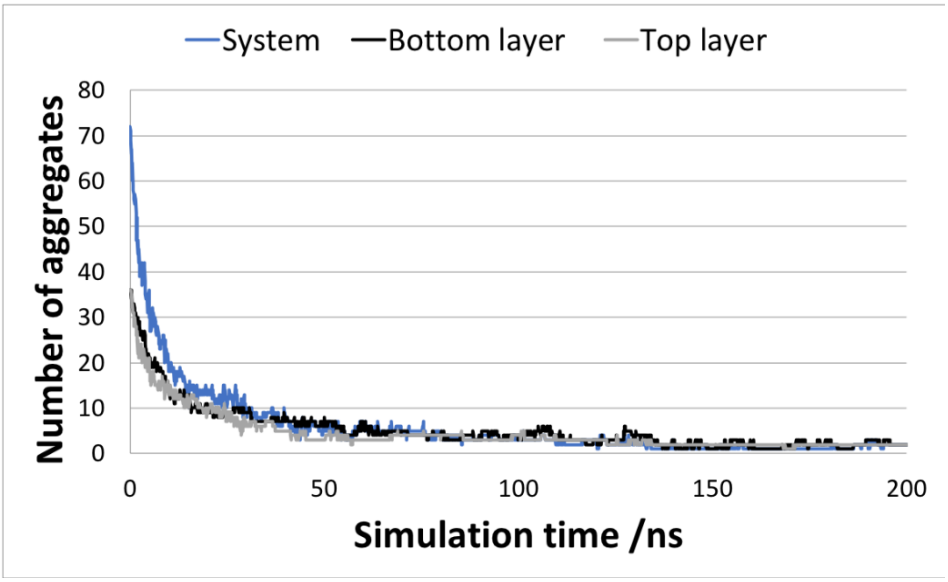


Figure 4.9: The number of aggregates with respect to simulation time for the grid500 system. The blue line represents the whole system (all protein atoms) while the black and grey lines represent the bottom and top layer of domains respectively.

We conclude from the graphs, that the system with the higher force constant (K value) for the position restraints interacts more quickly and that the bulk of aggregation appears within the first 40 ns of the trajectory, unlike the system with the smaller force constant for the position restraints (grid200) which requires slightly more time. This was contrary to expectations, however the differences between the analysis of the two systems were subtle.

The domains quickly interact in both ways (cis- and trans-) and thus, we cannot infer from the graphs if the one precedes the other although upon visualisation of the trajectory it appears that the cis- preceded the trans-interaction in most cases. Maybe, in future studies the chosen distance criterion could be altered, to display more clearly the formed aggregates. If in the secretory pathway claudins form stable oligomers, as has been proposed in the literature (or more probably when they are embedded in the membrane), it is obvious that the trans- interaction occurs between the already formed cis aggregates (Koval, 2013).

4.3.1.2 Relative orientation between ECDs on the same plane

It is important to examine the formed cis-interfaces between ECDs that are on the same plane, and compare them to the formed ones found in the previous study (Chapter 3) where we only examined the cis-interaction between them. This analysis can suggest areas of the protein that are more likely to interact and form dimers. The definition of the theta angles was given in Chapter 3. The specific areas show a characteristic relative angle to the other particle with which they interact.

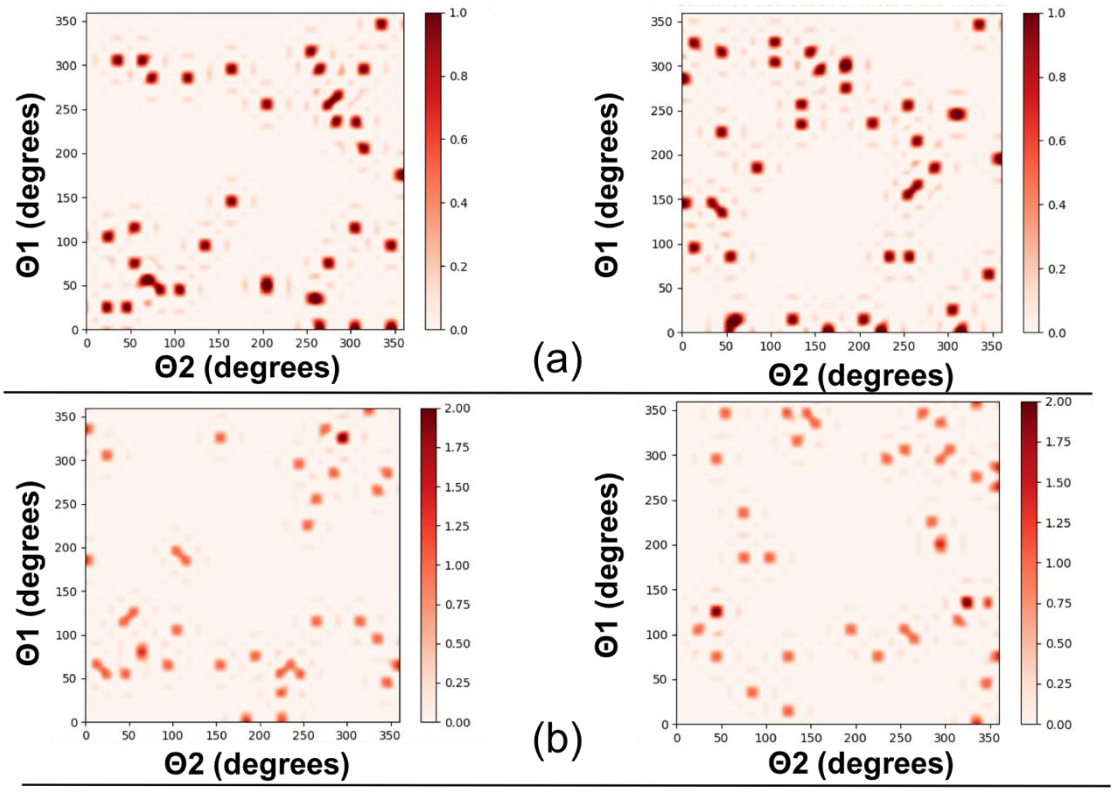


Figure 4.10: The combined angle distribution of the ECDs when in proximity (frequency). Theta 1 (Θ1) and theta 2 (Θ2) are defined based on the vectors connecting specific residues (glutamine 29 and threonine 137 as explained in Chapter 3). (a) The distribution for the top and bottom layer ECDs of the grid200 system and (b) the angle distribution for the top and bottom layer ECDs of the grid500 system respectively.

Briefly, the angles theta1 (Θ1) and theta2 (Θ2) aim to show the relative orientation of the ECD particles when in proximity, showing a distance less than 3.5 nm between their centre of mass. The centre of rotation was the alpha carbon of residue threonine 137, as this lies closer to the centre of mass of the whole claudin particle (x-y plane). The other reference point for rotation to define the vector, was the alpha carbon of residue glutamine 29. The frequency of a specific combination of Θ1 and Θ2 being observed is presented in Figure 4.10 and it is noted that this refers to the last frame only (see Figure 4.11). The angle distribution graphs seem to agree with the previous ones reported in Chapter 3 of this thesis and with an in silico study (Irudayanathan et al., 2018). The centre of the graphs (angle values approximately 180 degrees) is an un-preferred area,

while the left-hand corner side is a common popular region (Θ1~130 and Θ2~70 degrees) in most graphs, showing many formed dimer interfaces.

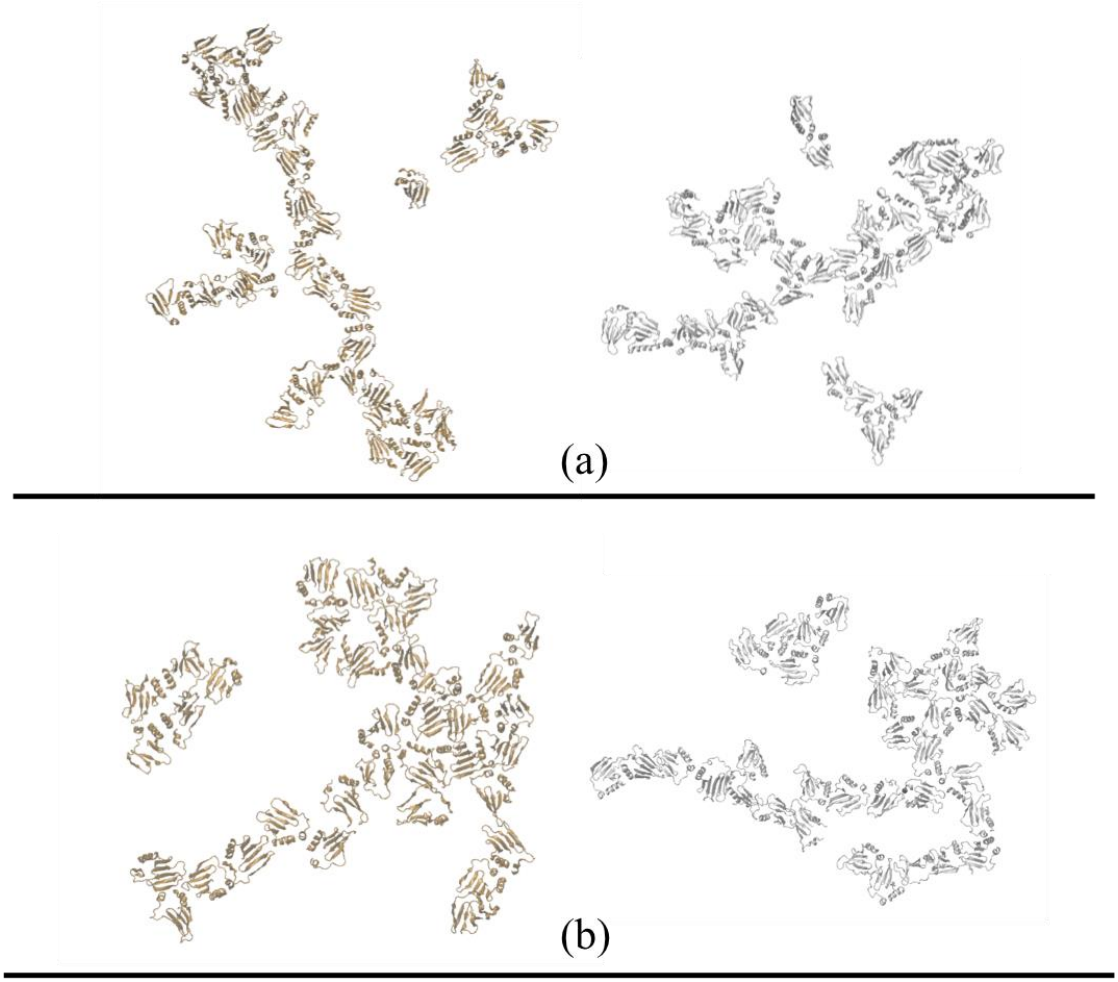


Figure 4.11: The end frames of the 200 ns trajectory of the top and bottom layer respectively for the examined trans grid systems with the applied weak (a) and strong (b) position restraints (grid200 and grid500 respectively). The proteins are in cartoon representation.

The last frame of the top and bottom layer of ECDs for both examined grid systems is presented in Figure 4.11. It is evident that the systems do not display a linear antiparallel double row arrangement of claudins such as the one proposed in the Suzuki model. However, it agrees with the in silico study where the self-assembly of classic claudins were examined with CG models of the proteins embedded in lipid bilayers (Irudayanathan et al., 2018). The overall 'gross' morphology of the strand in the

aforementioned study is very similar to the one found in this Chapter, and it also displays areas where the width of the strand is more than the width of a single claudin molecule. This is noted because a recent study of carbon replicas of claudin expressing cells, revealed that the TJs show a morphology of a double stranded network of fibrils or strands, where the fibril width was found to be 6.9 ± 0.8 nm which approximates to the predicted 6 nm diameter of the Suzuki pore model (Krystofiak et al., 2019). Maybe, to visualise such a fibril *in silico*, the ECDs need to break their initial interaction and rearrange themselves into a more favourable morphology. This would require extensive sampling and thus, enhanced sampling techniques are advised (see Chapter 6 where the most favourable conformation of claudin-1 and some claudin-1 mutants with metadynamics and replica exchange simulations is presented).

4.3.1.3 The penetration depth in the grid systems is low due to the applied restraints

The starting configuration of the ECDs placed on a grid, was such that the loops were randomly rotated and arranged to lie on a 2-dimensional plane, the implied bilayer. The initial distance between the layers was approximately 5 nm. Figure 4.12 and Figure 4.13 show the fluctuations of the average z coordinate of the alpha carbon of residue 29 on which the position restraint force acted on, for both systems. It is clear that the force was strong in both cases and the range of the z coordinate was small, namely approximately 0.3 nm and 0.2 nm for the system with the weak and strong position restraints (grid200 and grid500) respectively. Thus, the force has kept the penetration depth between the trans- interacting loops small, less than 0.9 nm as observed in Figure 4.14. Figure 4.14 demonstrates that the grid500 system has resulted in more trans-interacting pairs compared to the grid200 system, namely 28 and 22 pairs respectively. The pairs were calculated using python code (see Appendix A1).

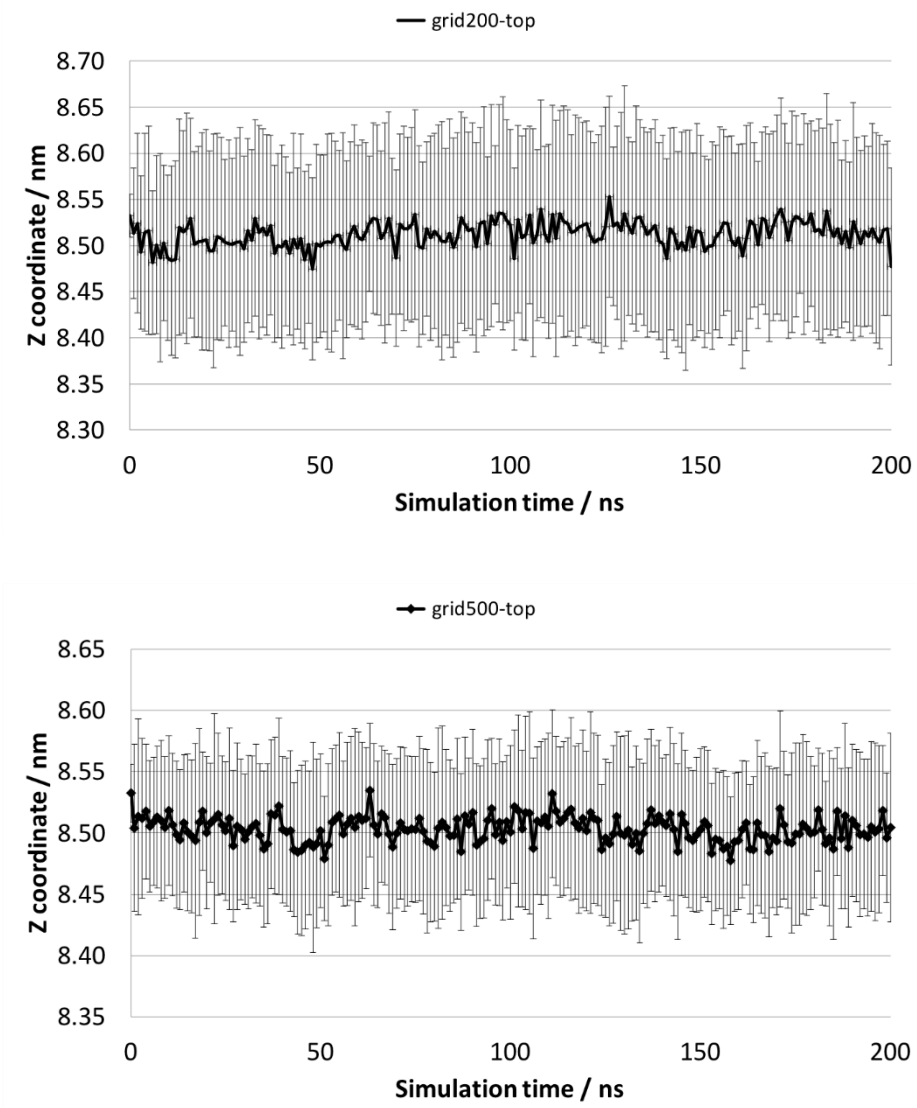


Figure 4.12: The first graph shows the average value of the z coordinate (Ca of residue29) of the system with the weak position restraints (top layer ECDs, grid200) and the second the corresponding values for the system with the strong position restraints (grid500). The value is averaged over all 36 loops and the standard deviation is also displayed (2σ).

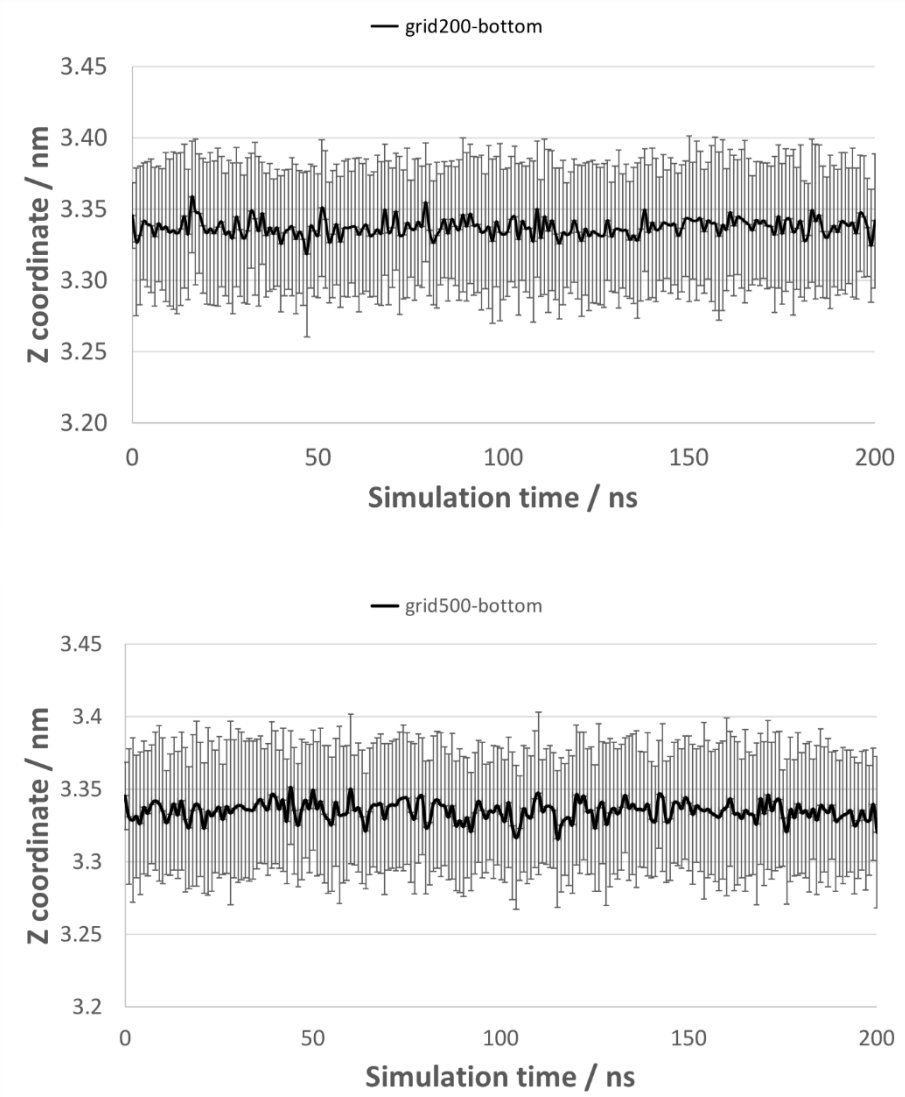


Figure 4.13: The first graph shows the average value of the z coordinate (Ca of residue 29) of the grid200 system (bottom layer ECDs) and the second the corresponding values for the grid500 system. The value is averaged over all 36 loops and the standard deviation is also displayed (2σ).

The python script aimed to examine how 'deeply' each loop had immersed into the opposing one. We assumed that a trans-interaction has occurred when the opposing loops were in close proximity defined as their centre of mass having a certain separation distance (dx =0.2 dy=0.2 and dz=0.7 nm). The penetration depth was then calculated for the trans-interacting loops based on their coordinates, specifically the bigger z coordinate of the atoms of the bottom ECD was subtracted from the smaller z coordinate

of the atoms of the top ECD. Figure 4.14 shows the number of formed interacting pairs and the corresponding penetration depth (z2-z1) for both grids.

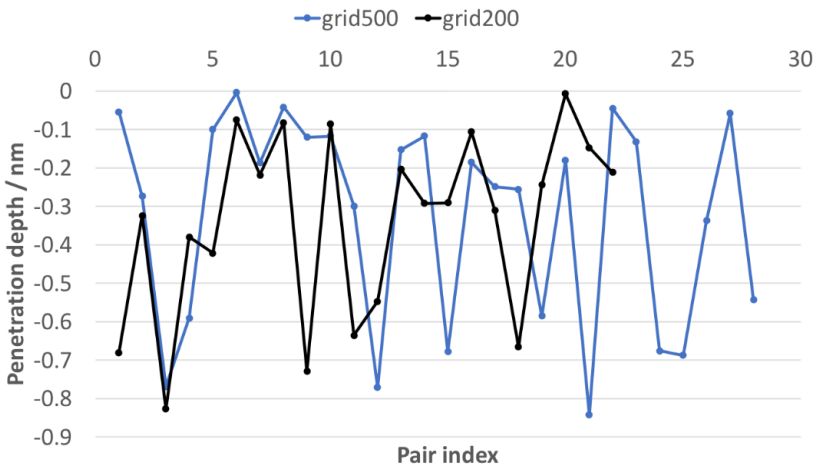


Figure 4.14: Penetration depth for both grid systems (strong position restraints-grid500 and weak position restraints -grid200 systems). The figure shows the number of trans-interacting pairs (pair index numbered sequentially) and the corresponding penetration depth which is less than 0.9 nm in all isolated pairs.

To conclude, we noticed that the formed trans-interaction pairs were not plenty, because of the large force that acted on the ECDs to keep them on the same x-y plane. When visualising the formed pairs, it was evident that the higher areas of the ECDs formed initial contacts and the ECDs were unable to penetrate deeper into the opposing layer.

4.3.1.4 Characterisation of the structural deviation of the domains

The root mean square displacement (RMSD) of the backbone atoms of each domain with respect to simulation time was examined which is indicative of how much the structure has deviated from the initial conformation. Also the root mean square fluctuation (RMSF) of the backbone atoms of the domains was examined which highlights the protein areas that fluctuate more. The RMSD value shows how much the structure of a single ECD has changed with respect to simulation time. This change can be caused by various factors, one being the interaction between the domains or the different environment that the ECDs were (solvated with ions). We started with a

homology model based on crystal structures, which represents a single frame of the conformation of the protein, while the simulation was for solvated molecules, and thus, some deviation was expected. Figure 4.15 depicts the mean values of RMSD for the top and bottom layer of both systems (grid200 and grid500) with the corresponding standard deviation value (2σ). We conclude that all systems display less than 0.5 nm (5 Å) structural deviation, which means that the ECDs changed their structure within acceptable limits.

When we consider the amount of variation in the average RMSD values (sigma value), we find that the only system which displays a value of slightly more than 0.6 nm (6 Å) deviation are the bottom loops of the grid200 system. It might be that the differing position restraints between the top and bottom ECDs cause a slightly more deviation of the structure compared to the grid500 system where the force constants acting on the springs of the bottom and top layer were both strong (top layer 500 KJ mol^{-1} nm^{-2} and bottom 1000 KJ mol^{-1} nm^{-2}). Of course, the difference in RMSD values is small, so we can conclude that the structural deviation of the ECDs is similar in all cases. These values might seem large but considering that the ECDs are the most flexible area of the protein, a slightly higher RMSD value is accepted since it could accommodate the free movement of the ECDs in an physiological environment (aqueous with ions). Moreover, when the loops interact we expect that the structure will undergo some changes.

Note that the RMSD values are consistent with the respective values of the cis-interacting ECDs that were calculated in the previous study (Chapter 3) and suggest the good quality of both the homology model and the CHARMM force field parameters. Moreover, the values found in this study are only slightly higher than previous values reported in literature (namely 0.3-0.4 nm for the ECLs) while the whole protein was embedded in a membrane (Alberini G, 2017, Samanta et al., 2018).

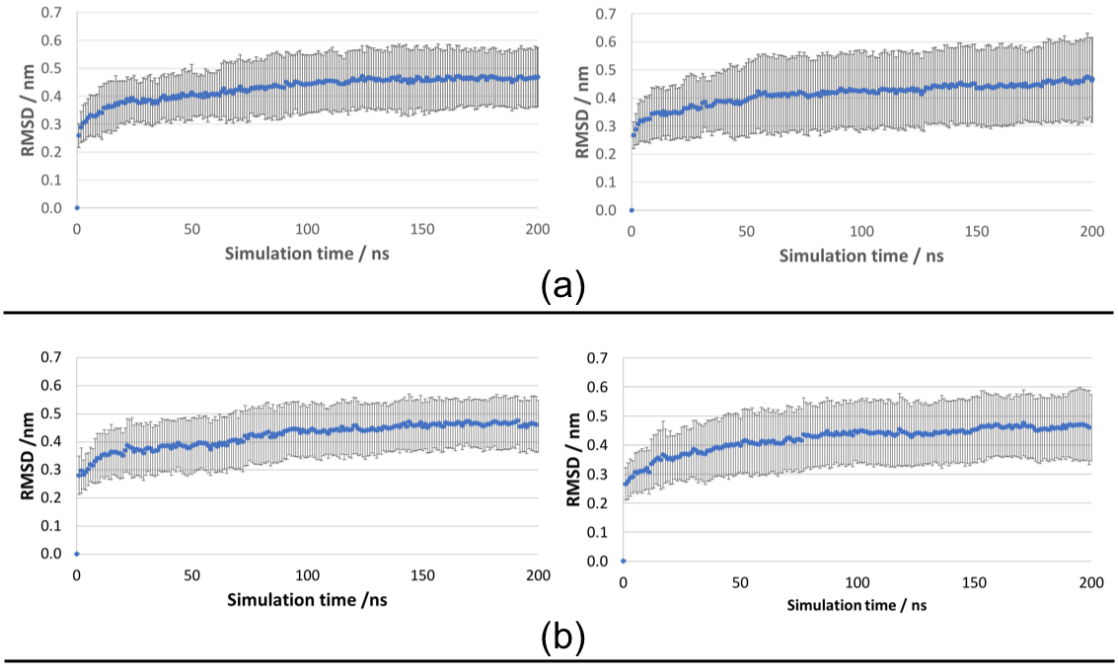


Figure 4.15: The average RMSD value for the backbone atoms of the ECDs as a function of simulation time. Figure (a) shows the RMSD for the top and bottom layer of ECDs respectively for the system with the weak position restraints (grid200) and (b) shows the RMSD value for the top and bottom layer of ECDs respectively for the system with the strong position restraints (grid500). The relative standard deviation is also displayed (2σ).

Next, we examined which region of the ECDs showed the largest flexibility by calculating the RMSF values for the backbone atoms of all residues with a view to understand its functional regions. This property is important because flexibility can accommodate change, meaning that the areas characterised as more flexible might be important for the trans- interaction that forms the pore structure. Figure 4.16 shows the average RMSF values of the two layers of both systems (grid200 & grid500). The graphs seem very similar and when superimposed (see Figure 4.17) it is clear that the unstructured regions not part of any secondary structure are the most flexible areas, as expected. Figure 4.17 presents the superimposed RMSF plots of all systems. If we consider a threshold value of 0.3 nm, which corresponds to approximately 236 for the B-factor (see Equation 19 where B is the B-factor and RMSF is the RMSF value), we find that with respect to the first extracellular loop (residues 29-81) the areas with large

RMSF values are residues 37-41, which connect β1 and β2 strand, residues 58-60 which connect β3 and β4 strand and the unstructured area before the small helix (residues 68-72). Specifically, high RMSF values display glycine 37-valine 41 for all systems and threonine 42 for the bottom ECDs of the system with the large position restraints (grid500). Serine 58- glycine 60 and glutamine 61 show large fluctuations (glutamine 61 does not display an RMSF value higher than 0.3 nm for the top ECDs of the grid500 system). Some backbone atoms of aspartic acid 68 and serine 69, leucine 71, asparagine 72 and leucine 70 show significant displacement and finally, arginine 81 for the top ECDs of the grid200 system, display significant fluctuations thus, their spatial extend of random motion is large. This means that all the above-mentioned residues display larger fluctuations compared to the rest and show significant deviation of their position with respect to a reference position over time. This measure highlights the variable regions of the hairpin loops which connect the β strands as the more flexible areas of the ECDs, and this was expected, since they do not participate in an ordered secondary structure conformation. Moreover, the area close to the small helix which is part of ECL1 shows high RMSF values and this area is important since it is suggested to participate in cis-interactions based on the crystal structure of mouse claudin-15 (the linear interface) (Suzuki et al., 2014). Additionally, considering the second loop (residues 137-163) only methionine 152 depicts a high RMSF value (except for the top ECDs of the grid500 system). Methionine 152 is a residue located in the V2 variable region of the ECDs and has been found important for the trans- interaction (binding) between claudin-1 and a claudin-1 specific binder (a monoclonal antibody) (Nakajima et al., 2015). Note that in an atomistic study that examined the stability of the Suzuki model, and thus the structure was initially fixed, it was concluded that the RMSF values for the backbone atoms were close to 0.24 nm and the generated figure resembles the one reported here (Figure 4.17) highlighting the hairpin loops as very flexible regions (Samanta et al., 2018).

.

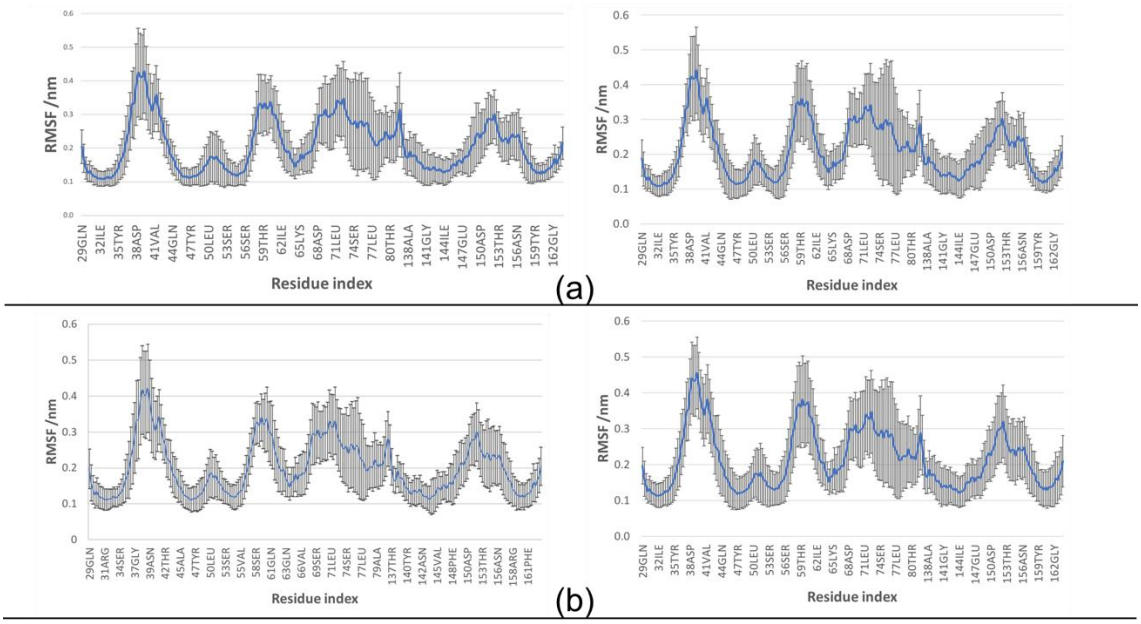


Figure 4.16: The average RMSF values for the backbone atoms of the ECDs for the large grid systems with respect to residue index. Average RMSF values for (a) the top and bottom layer of the ECDs respectively for the grid200 system and (b) the top and bottom layer of the ECDs for the grid500 system. The standard deviation is also displayed (2σ).

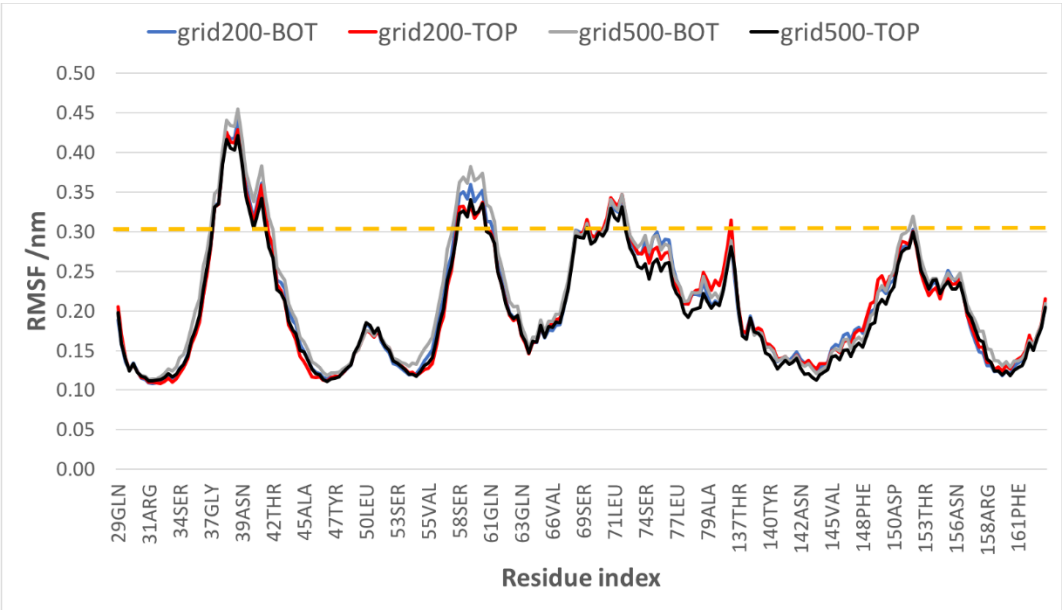


Figure 4.17: The average RMSF value for both grid systems (grid200 & grid500) superimposed in the same graph. The 0.3 nm threshold limit is also displayed on the graph as a dashed line.

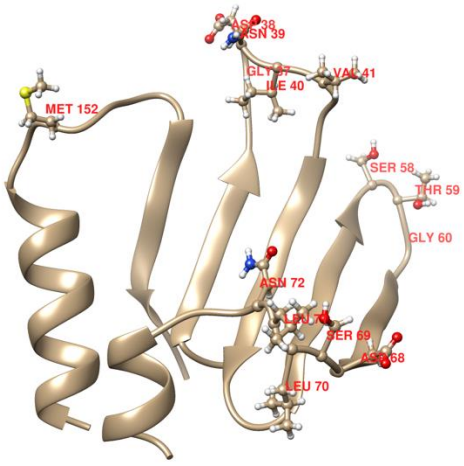


Figure 4.18: The important areas of the ECDs based on their RMSF values are presented in a different representation (balls & sticks) compared to the rest of ECD which is shown in cartoon representation. The ECD was rendered with UCSF Chimera (Pettersen et al., 2004).

These results might be expected but it is remarkable how the areas with residues with hydrophobic side chains due to their nature (solvated in water) can facilitate the so called 'head to head' or trans- interaction. This could be due to both their location (positioned at the top of the ECDs) and their properties (hydrophobicity), a statement which is supported by both the findings of this study as well as by others (Alberini G, 2017, Suzuki et al., 2017, Suzuki et al., 2015, Irudayanathan et al., 2017).

Figure 4.18 offers a visual representation of the flexible areas of the ECD and enhances our understanding of the location of the key residues. These areas are considered crucial for the study of both the cis- and trans- interaction between TJ proteins.

4.3.1.5 The key areas of the ECDs that regulate the trans-interaction are the variable regions

Before we move on to the systematic study of the rotated ECDs, we aim to identify the important areas that were found to interact in the grid systems and establish that these areas where at the top of the ECD particle. We have divided the domain particles into different regions and calculated the number of contacts between these specific regions and protein atoms from the opposing layer. The analysis was done for plenty different areas composing structured and unstructured regions of the ECDs and it highlighted the 'key players' that determine the trans-interaction of claudin-1.

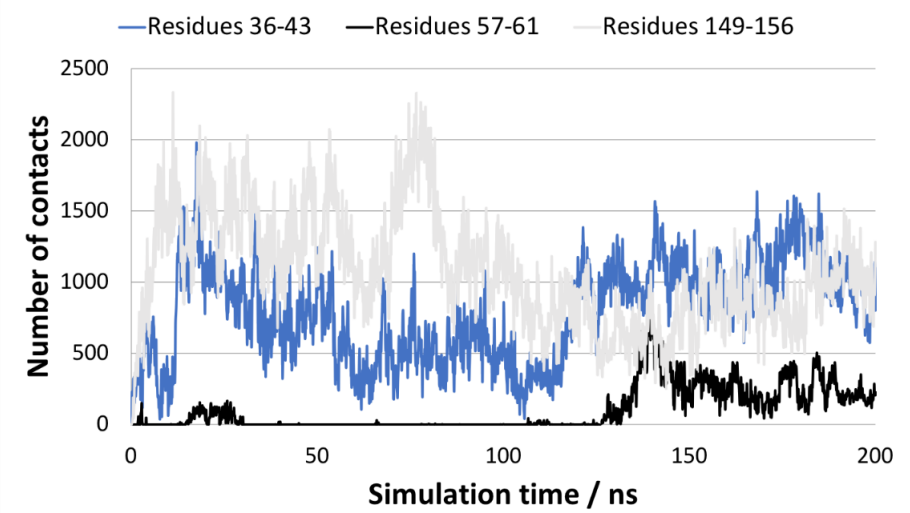


Figure 4.19: The number of contacts between the specified region of the bottom ECD (see label) and protein atoms from the opposing top layer for the grid200 system. The grey line represents the V2 region (residues 149-156) while the blue line shows the respective number of contacts formed by the V1 region (residues 36-43).

Based on a cut off distance (0.5 nm) the number of contacts between the two opposing layers (bottom and top ECDs) was defined. It has been suggested in the literature, that the unstructured regions of the loops are important for the trans- interaction (Irudayanathan et al., 2017, Alberini et al., 2017, Rossa et al., 2014). Here, we confirm that these regions of the bottom ECDs (residues 36-43, residues 57-61 and residues 149-156) form the largest number of contacts with protein atoms from the top layer in both examined systems, characterising them as ‘key’ regions for the trans- interaction.

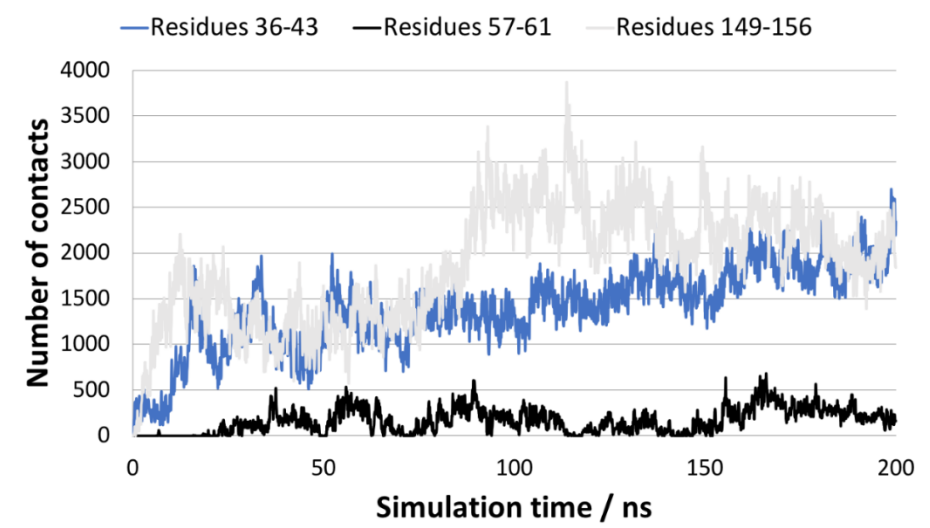


Figure 4.20: The number of contacts between the specified region of the bottom ECD (see labels) and protein atoms from the opposing top layer for the grid500 system. The grey line represents the V2 region (residues 149-156) while the blue line shows the respective number of contacts formed by the V1 region (residues 36-43).

Specifically, the V1 region connecting β1 with β2 strand (residues 36-43) and V2 region of the second loop that connects β5 strand with the third transmembrane region, exhibit the largest number of contacts (see Figure 4.19 and Figure 4.20). During the trajectory, the number of contacts vary as seen in the graphs, but these areas remain the most important ones for the trans- interaction between the ECDs. V1 and V2 seem to compete during the simulation of which area forms more contacts compared to the other. Figure 4.21 shows the ECD particle with highlighted the 'key' trans- interacting regions, and it is clear that both V1 and V2 regions are located higher compared to the other residues, and especially the third variable area (residues 57-61 which have polar side chains); thus, it was expected in some extend that these areas would be important for the claudin-claudin interactions between neighbouring cells. It is important to note that these areas (V1 and V2) also displayed high RMSF values.

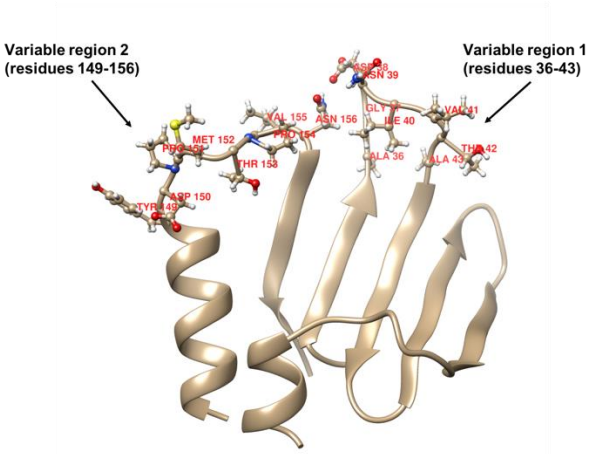


Figure 4.21: The ECD particle in cartoon representation, coloured in gold. The 'key' trans-interacting areas of the ECD are highlighted, by displaying them like 'ball & sticks' (atom colours). The rendering was done with UCSF Chimera (Pettersen et al., 2004).

To summarise, with regards to the self-assembly simulations of the large grids we notice that the ECDs initially placed on two grids facing each other, quickly interact on the same plane forming dimers and occasionally trimers and some also formed trans-contacts with the opposing grid. However, the applied position restraints on the terminal atoms were strong, preventing the ECDs from penetrating deeper into the opposing layer. Considering this result, we decided to move on and examine the trans-interaction between claudin-1 more systematically.

4.3.2 The rotation study

The six new systems comprising the rotation study were named after the rotation angle phi (φ), before the second ECD was turned over and placed on top of the other ECD (phi0, phi60, phi120, phi180, phi240 and phi300). The idea was to encourage the system to sample more conformations by changing systematically their initial arrangement (see Figure 4.5).

Here, the ECDs seem to favour some initial relative arrangements compared to others, meaning they approach each other quickly and remain in a similar configuration for the

remaining trajectory. Otherwise, the ECDs approach each other but break their initial interaction, slightly reorder themselves and find another interface which appears to be more favourable compared to the first initial interaction site. Note that, the ECDs were restrained on a line and thus, their ability to rearrange before interacting again was limited.

4.3.2.1 The ECDs display a dynamic behaviour and penetrate deeper into the opposing layer

We characterised the structural stability of the ECDs (RMSD value) and how much they 'open up' during the trajectory (radius of gyration). We also examined the distance between the top ECD from the bottom ECD (penetration depth), the minimum distance between the particles, the number of contacts between them and the number of formed salt bridges. Most importantly, we calculated a characteristic torsion angle to see if the loops significantly change their configuration upon binding or if they remain the same (see Table 4.1). Additionally, since the end frame was considered the most favourable conformation of the system, we have characterised the end structure of all rotated systems in detail. We do that by displaying the initial and final snapshot of all examined systems next to each other (see Figure 4.27 and Figure 4.28) and with the help of UCSF chimera and its tool Find Clashes/Contacts, we report on the residues that form the largest number of contacts with the opposing ECD (Pettersen et al., 2004). The number of contacts/overlaps was calculated using Equation 20:

$$overlap = r_{VDWi} + r_{DVWj} - d_{ij} - allowance_{ij} \quad (20)$$

where r_{VDWi} is the vdW radius of atom i**,** d_{ij} the distance between atoms i and j, and an allowance distance which for this study was set to 0.1 nm.

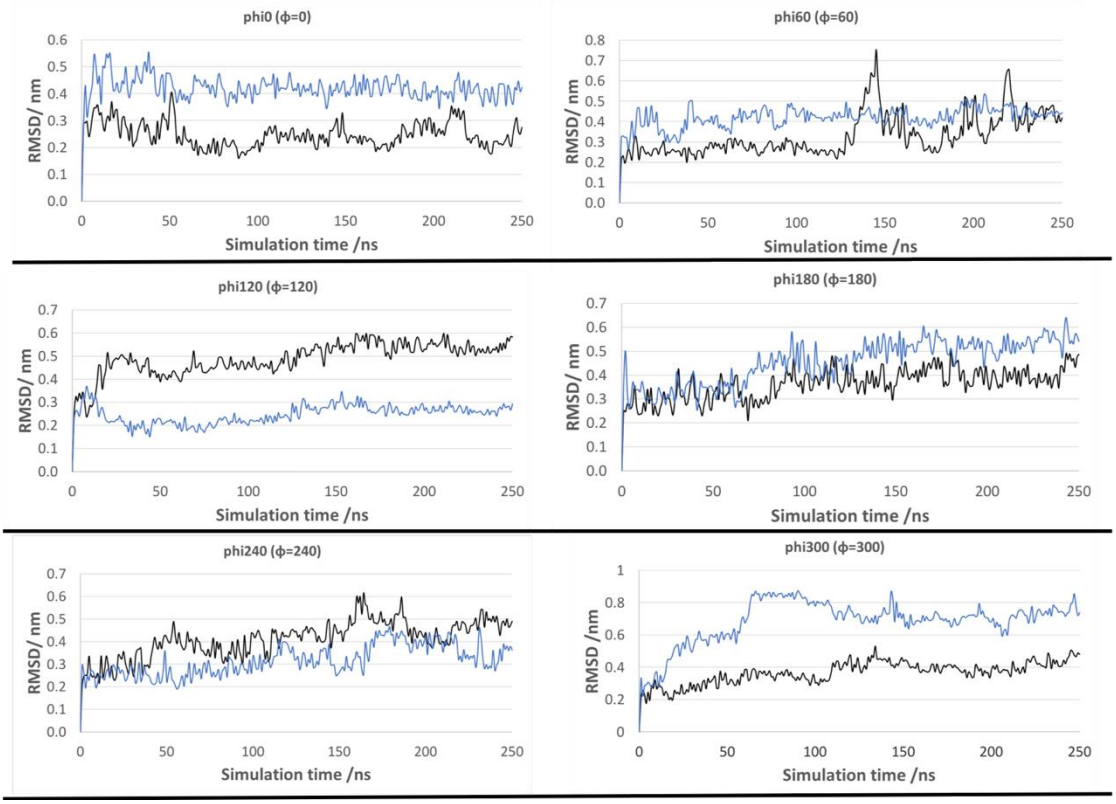


Figure 4.22: RMSD value of both ECDs from the rotation study with respect to simulation time. Each graph is labelled according to the specific system (phi0 to phi300) and the analysis was focused on the backbone atoms of the bottom (black line) and top (blue line) ECD atoms.

Here, the loops displayed some deviation from the initial structure with RMSD values for the backbone atoms slightly over 0.6 nm (see Figure 4.22) but the penetration depth increased significantly with values in the range of 1-4 nm (see Figure 4.25). The radius of gyration of the ECDs, which is indicative of how compact the structure remains is presented in Figure 4.23 and we can conclude that the ECDs display varying trends, in some cases reducing their value while in others slightly increasing it. It is also noteworthy, that the top and bottom ECD particles are not consistent, meaning that in some cases the top ECD has a smaller radius of gyration compared to the bottom one, while in others it is the other way around. However, the forces of the applied position restraints were the same on the two particles and we have utilised PBC, so, the top and bottom should not differentiate amongst each other.

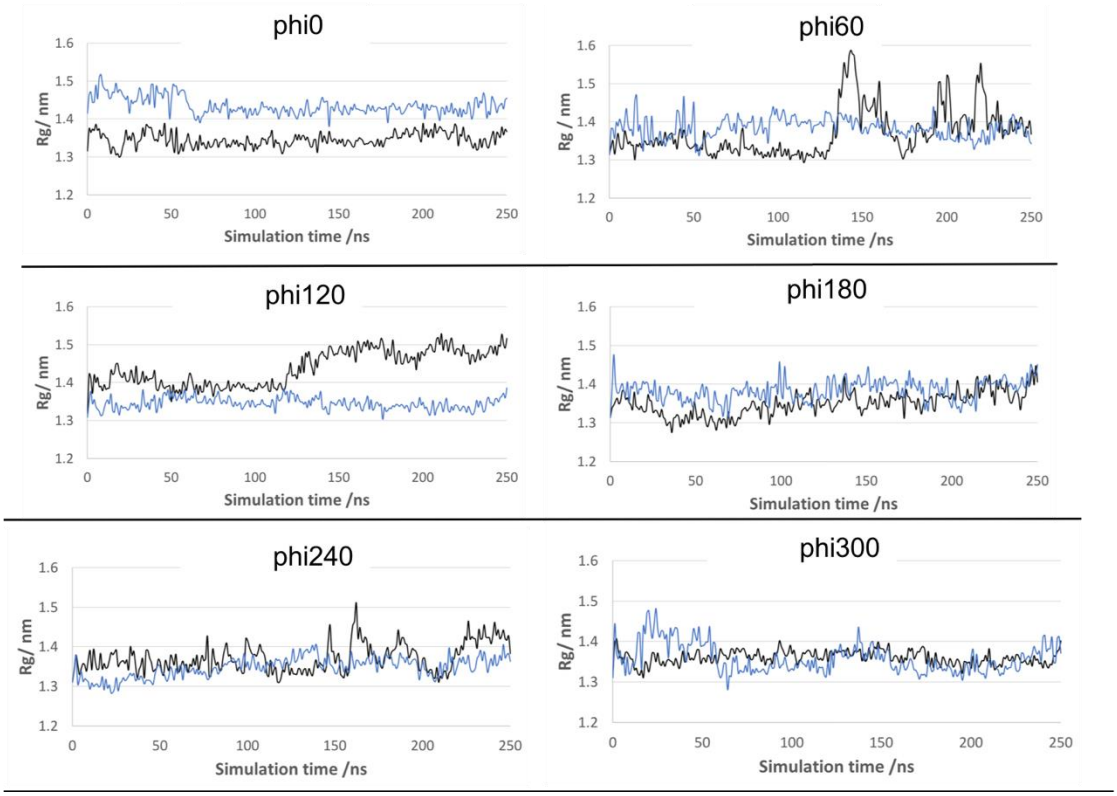


Figure 4.23: Radius of gyration as a measure of the ECDs compactness for the systems comprising the rotation study. The black line corresponds to the backbone atoms of the bottom ECD, while the blue line is for the backbone atoms of the top ECD.

An important factor was how much the ECDs would retain the initial arrangement, so we examined the torsion angle omega (ω) formed between the vector connecting the terminal residues threonine 137 and glutamine 29 having as an axis the vector connecting glutamine 29-glutamine 29. This torsion angle between the two vectors, is different than the initial angle (phi) used to rotate the ECD on the x-y plane before placing the particles opposite each other. The analysis shows that the initial angles do not change upon binding and they fluctuate within about less than 22 degrees from the average angle (see Table 4.1) meaning that the terminal atoms of the ECDs do not change significantly their relative spatial arrangement when interacting. We acknowledge however, that the relative position restraints might play a role in this.

Table 4.1: The torsion angle ω formed between the vectors connecting residues threonine 137-glutamine 29 having as an axis the vector connecting the two glutamines (GLN29-GLN29). The standard deviation is also displayed.

Initial torsion angle (ω)	**Average torsion angle (ω)**	**Standard deviation (±σ)**
-17	-28	±11
41	40	±8
99	100	±5
161	163	±6
-139	-142	±7
-80	-72	±7

Using a cut-off distance (0.6 nm) the number of contacts formed between the opposing ECDs (Figure 4.24), the minimum distance between them (Figure 4.26) and their penetration depth (Figure 4.25) was calculated during the trajectory. All these quantities are used to characterise any given interface and are good indicators of the stability of the interacting pairs.

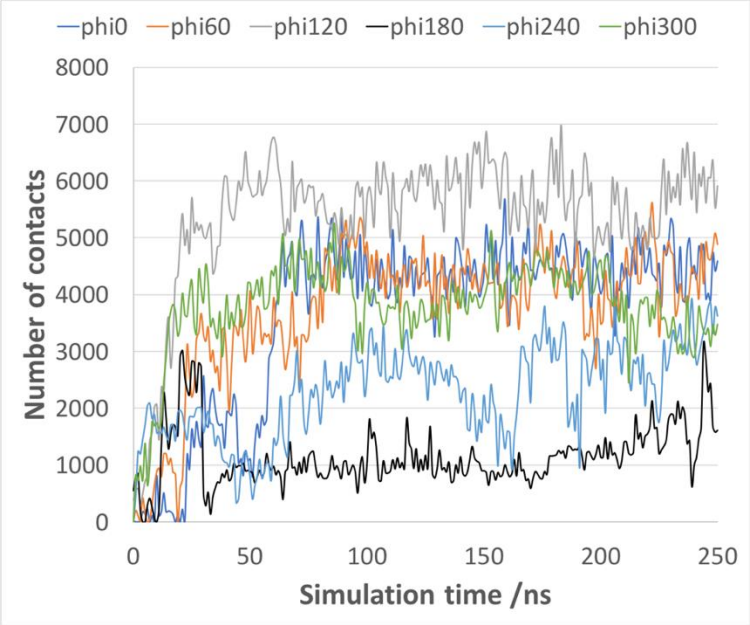


Figure 4.24: Number of contacts formed between the ECD particles during the 250 ns trajectory for the phi0-phi300 systems. The distance criterion used was 0.6 nm, thus, atoms having a smaller distance were automatically considered that they formed a contact.

With regards to the number of contacts formed between the ECDs during the trajectory, we note that the system with the largest number of contacts formed is phi120, while the one with the smallest number is phi180, which also displayed varying interfaces during the 250 ns trajectory (see next section). Phi240 displays the largest fluctuations in the number of formed contacts after 100 ns when most of the systems have equilibrated. As previously mentioned, the end frame is considered a stable configuration of the system. Hence, we have also calculated with UCSF Chimera and its tool Find Clashes/Contacts the number of contacts formed between the ECDs at the last frame (Table 4.2). Note that the way a contact is defined in the Find Clashes/Contacts tool in UCSF Chimera (Pettersen et al., 2004) is different from the distance criterion that GROMACS uses to define the equivalent number of contacts between the ECDs (Abraham et al., 2015). However, from Figure 4.24 and Table 4.2 we can infer that the phi120 system appears to have established a strong interface, as well as the phi0 and phi60, where atoms from one ECD are interacting with a large number of atoms from the opposite ECD. On the other hand, phi180 has the smallest number of contacts in both analysis.

Table 4.2: Number of contacts formed between claudin ECDs in the rotation study. The contacts were determined at the end frame with UCSF chimera and its structural analysis tool Find Contacts (Pettersen et al., 2004).

Examined System	**No of contacts**
phi0	530
phi60	473
phi120	600
phi180	123
phi240	329
phi300	385

The minimum distance between the trans-interacting ECDs was defined as the minimum distance between any atom of the bottom ECD with any atom of the top ECD. Figure 4.26 demonstrates that all systems (phi0-phi300) have a similar small minimum distance (approximately 0.2 nm), which is indicative of how close the particles were. However, the penetration depth -vertical distance (defined as the difference in the coordinates z2-z1) between the atoms of the neighbouring ECDs, is more informative and distinguishes the rotated systems better (see Figure 4.25). Phi0 depicts the largest average penetration depth (3.5 nm) compared to the other systems, while phi180 the smallest one (0.5 nm). The other systems, namely phi60, phi120, phi240 and phi300 show an average penetration depth of 2.3 nm, 1.4 nm, 1.3 nm and 1.1 nm respectively averaged over the last 200 ns of the trajectory.

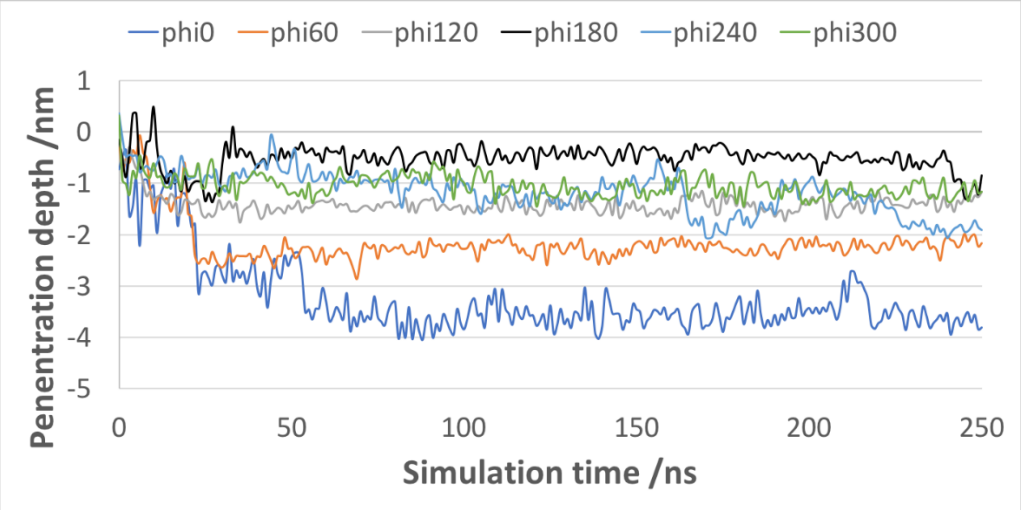


Figure 4.25: The penetration depth for all systems comprising the rotation study with respect to simulation time. The depth was measured as the difference in the Z coordinate of the lowest (Zmin) atom coordinate of ECD2 minus the highest (Zmax) atom coordinate of ECD1.

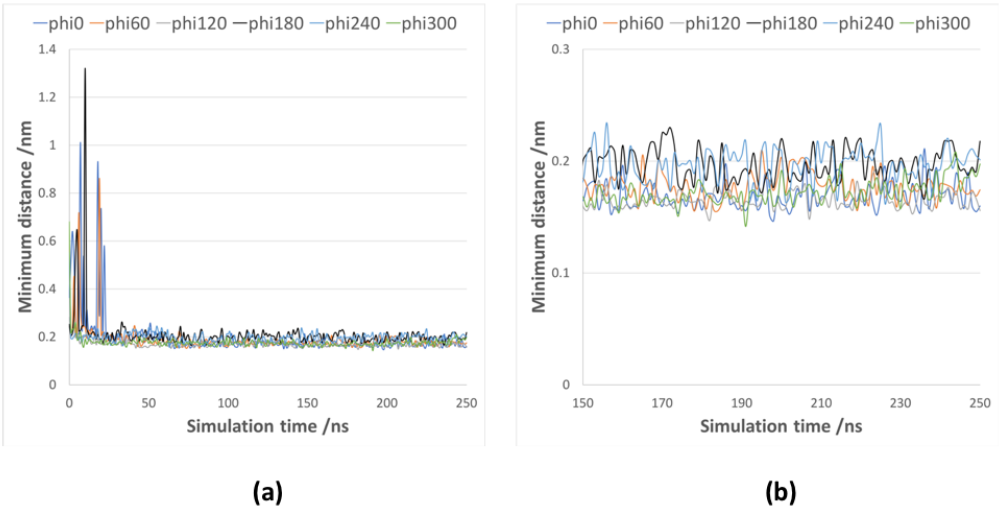


Figure 4.26: (a) Minimum distance between the ECDs in the phi0-phi300 systems with respect to simulation time. (b) The graph is focused on the last 100 ns of the trajectory showing the minimum distance between the ECDs as a function of simulation time. The graph labels show the name of the examined system (phi0-pi300).

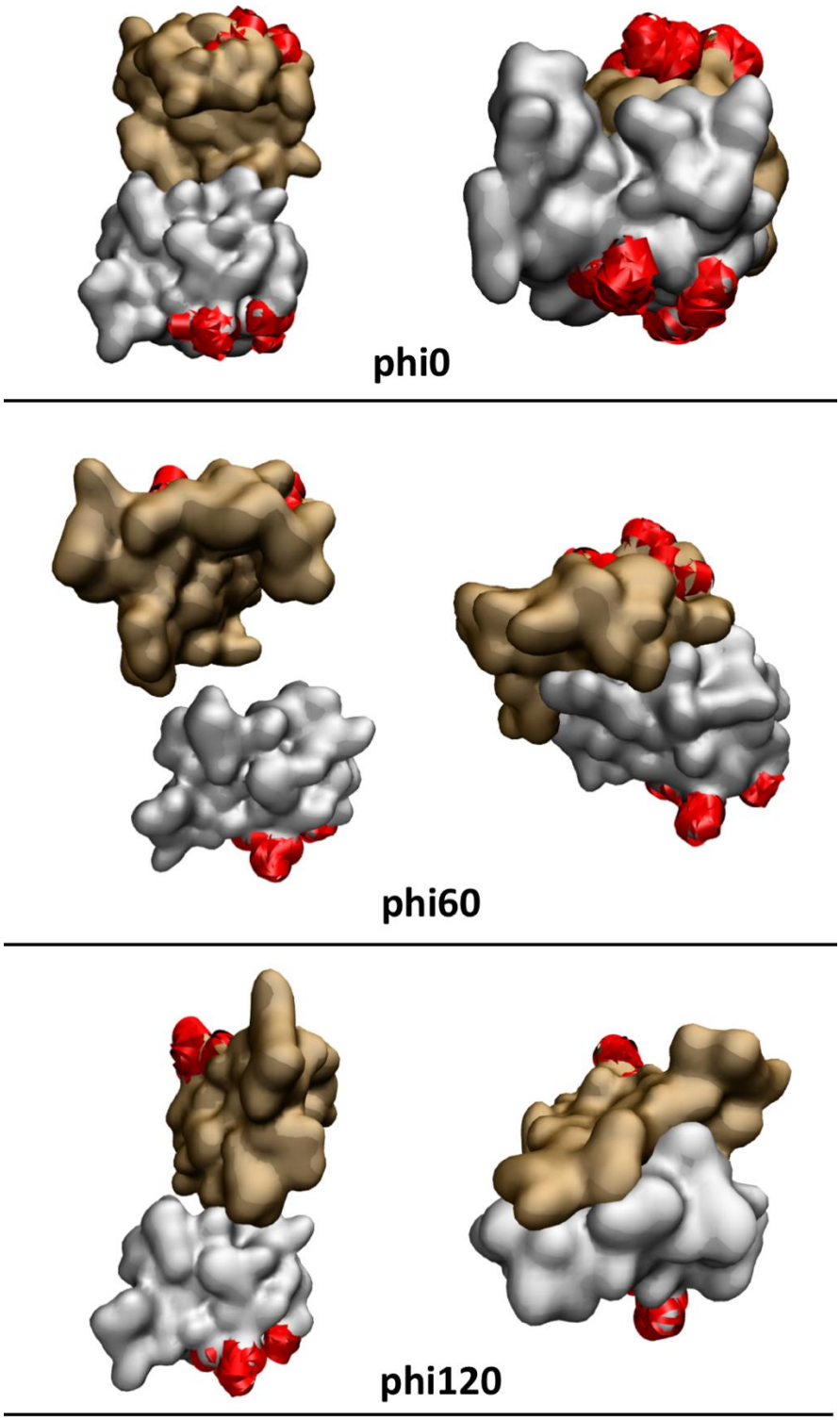


Figure 4.27: The start and end frame next to each other of the rotated systems phi0, phi60 and phi120 respectively. The ECDs are in surface representation coloured silver and gold and the atoms of the terminal residues (GLN29, ARG81, THR137 and GLN163) are shown as red spheres (of 0.3 nm radius).

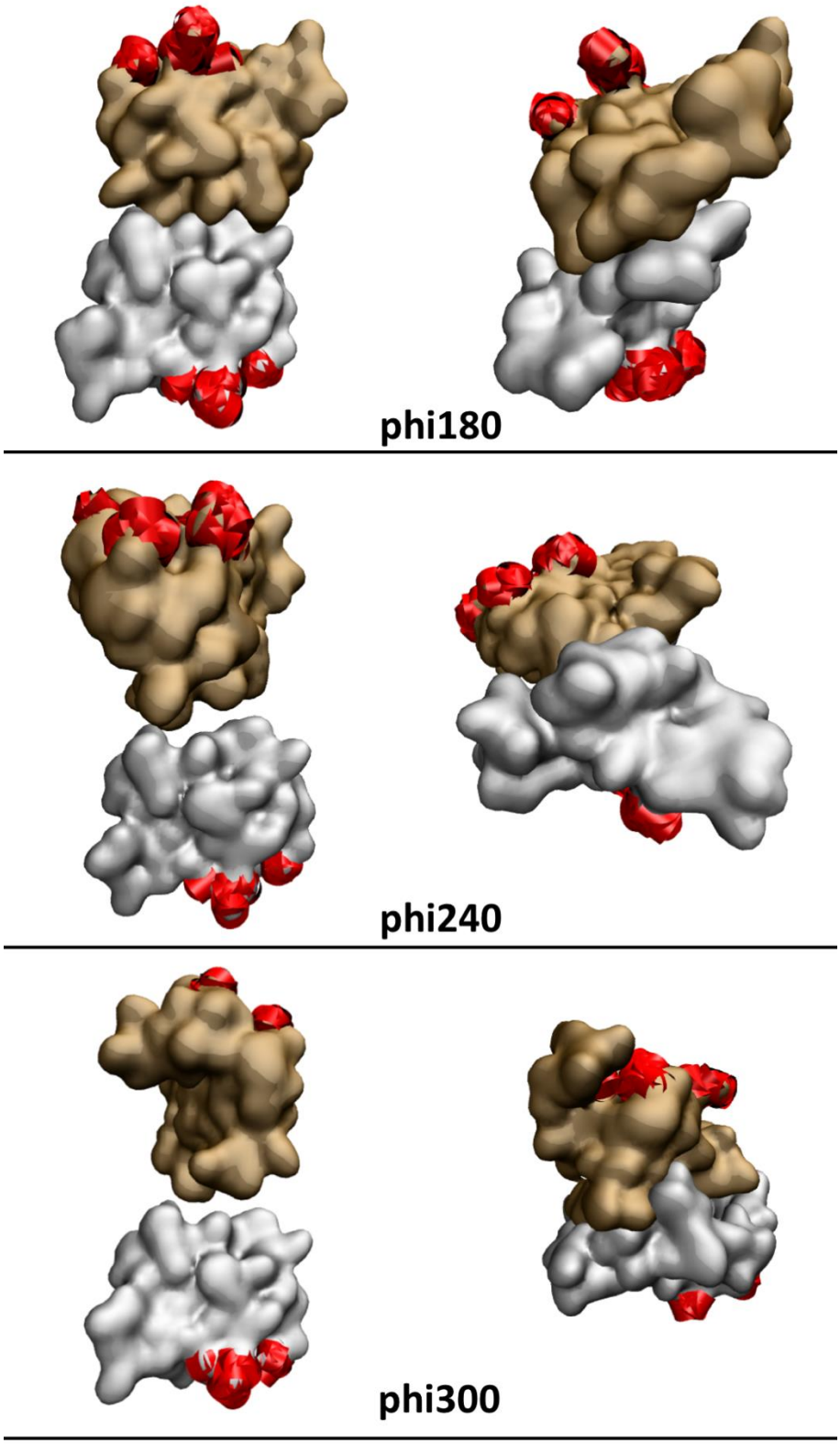


Figure 4.28: The start and end frame next to each other, of the rotated systems phi180, phi240 and phi300 respectively. The ECDs are in surface representation coloured silver and gold and the atoms of the terminal residues (GLN29, ARG81, THR137 and GLN163) are shown as red spheres (of 0.3 nm radius).

4.3.2.2 The observed trans- interfaces between the ECDs of the phi0 to phi300 systems

After viewing the trajectories of all systems, we observed that in some cases pore-like structures were formed (see Figure 4.29) and the ECDs did not change significantly their organisation for the remaining of the trajectory. While in some other the ECDs kept on 'sliding' towards each other increasing the penetration depth. It is important to note, that in this study we observed similar trans-interfaces as in the big grid systems, where the ECDs interacted through the top of the particle (hairpin loops). However here they increased the penetration depth forming contacts with areas of the ECD that are closer to the β-sheet domain and the small helix of the first loop.

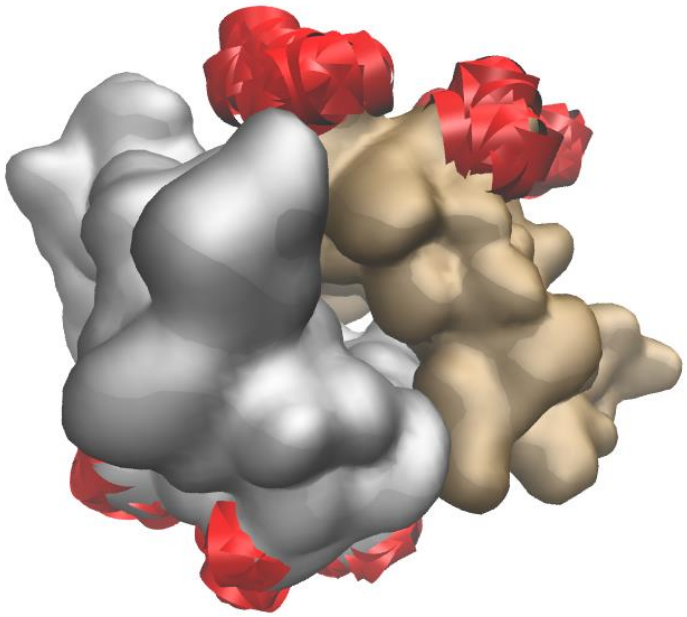


Figure 4.29: The phi0 system at the end frame. Each ECD is in surface representation and coloured differently (gold and silver). The end frame structure shows that a small pore was formed between the ECDs (rendered with VMD (Humphrey et al., 1996)).

Specifically, in the rotate phi0 (φ=0) system the ECDs exhibit an RMSD value of less than 0.5 nm and they quickly interact through the variable regions (V1 and V2 hairpin loops) of one ECD and via the area close to the small helix of the opposing ECD (see

Figure 4.27). The penetration depth is large, approximately 3.5 nm (averaged over the final 200 ns of the trajectory), and the loops remain roughly in this configuration for the rest of the trajectory (see Figure 4.25). UCSF chimera gave a total of 530 contacts between the atoms of the ECDs with leucine 71, serine 75, methionine 152 and methionine 46 forming the largest number of contacts with the opposing ECD (see Table 4.2). This trajectory resulted in a pore-like structure (see Figure 4.29 and Figure 4.30) where the β4 strands are not participating in the trans-interface so they are free to interact with other TJ proteins on the same plane (cis-interaction); note that this is consistent with the Suzuki model that proposes a cis-interface formed between the β4 strands. It is also interesting to analyse the formation of salt bridges, as this is one of the strongest non-covalent interaction that we can observe. The salt bridges formed are presented in the Appendix A2 where we monitored the distance between the oxygen of the acidic residue and the nitrogen of the basic residue. If the aforementioned distance was less than 0.32 nm, it was considered that the two corresponding residues have formed a salt bridge (VMD salt bridge analysis tool (Humphrey et al., 1996)). Phi0 and phi180 demonstrated the smallest number of salt bridges formed compared to the other systems (8 salt bridges were formed in both systems).

The starting arrangement in the rotate phi60 (φ=60) system does not seem to favour the trans-interaction between the ECDs. The top ECD even after 250 ns does not seem to have 'settled' based on the RMSD value. The average penetration depth is approximately 2.3 nm (averaged after the first 50 ns) and the number of contacts formed between the particles fluctuates a lot (see Figure 4.25 and Figure 4.24). Again the β4 strands are left free to interact with other proteins on the same plane but the orientation of the ECDs is slightly different compared to the phi0 system. Here two different interfaces were formed; specifically, the opposing ECDs interact via the tip of the second loop and the area close to the small helix from the other ECD, and via the V2 area of one ECD and the β1 and β2 strands (area under the V1 hairpin loop) of the other ECD. The area close to the small helix of the bottom ECD has unfolded and this might be because it does not interact with the opposing ECD. The number of contacts at the end frame was high (473 contacts calculated with the Find Clashes/Contacts tool in the structure analysis of chimera (Pettersen et al., 2004)) and the key residues were methionine 152, valine 41, proline 154 and tyrosine 33. There were also 13 salt bridges formed during the trajectory.

With regards to the trajectory of the phi120 (φ=120) system, we notice that in the end frame the ECDs display a different organisation compared to the phi0 and phi60 systems (see Figure 4.27). The position of the β4 strand was different; instead of facing the opposing β4 strand as in phi0 system, it was opposite the long transmembrane helix (see Figure 4.30). The number of contacts at the end frame calculated with chimera was 600, which is the largest number of contacts formed compared to the other systems (see Table 4.2), and the important residues based on the number of contacts they formed with atoms from the opposing ECD were methionine 152, tyrosine 35, asparagine 72, lysine 65, proline 154 and valine 55. The number of salt bridges formed during the trajectory was 11.

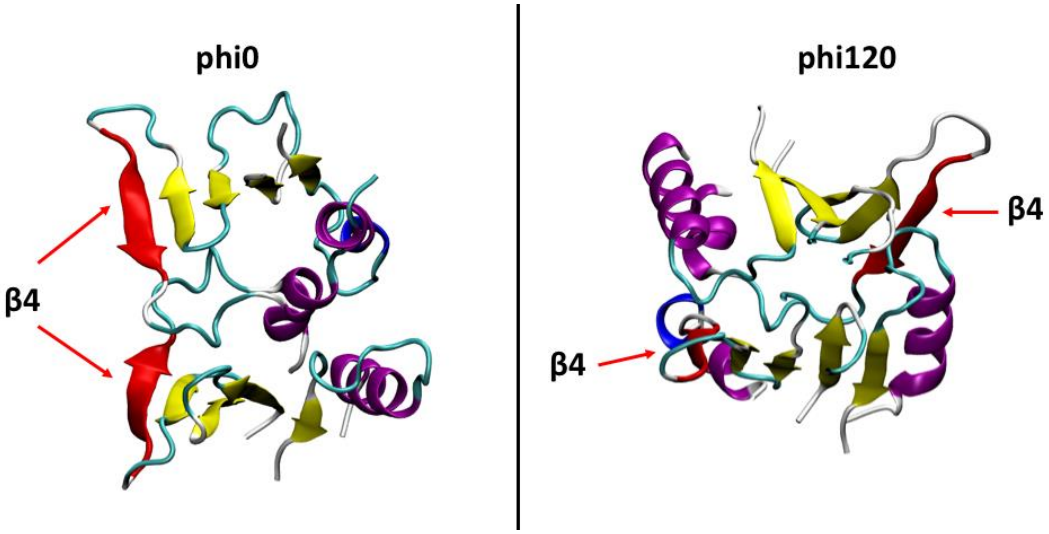


Figure 4.30: The different organisation of the ECDs of the phi0 and phi120 systems (top view-end frame). The pictures show the different relative arrangement of the β4 strands in the systems. In phi0 the β4 strands are opposite each other and are free to interact with other protein atoms on the same plane, while in the phi120 the β4 strands are opposite the long helix of the third transmembrane.

The phi180 system displays an interesting and very dynamic behaviour because the formed trans-interface between the ECDs changes throughout the trajectory. Initially, the ECDs interact via the tip of the second loop but this is not a strong interaction since the ECDs break it and slightly rearrange themselves before interacting again. The remarkable thing is that after 40 ns the ECDs again approach each other through the

same area (top of the second loop- V2 region). Figure 4.31 shows the distance between methionines 152 from the two particles with respect to simulation time, because methionine 152 is strategically located at the top of the ECD and it has been characterised as important for the trans-interaction of claudin-1 in experiments (Nakajima et al., 2015). From the graph we can conclude that the interaction between the V2 region is strong because the distance between the methionines has an average value of 0.8 nm but towards the end of the trajectory the ECDs move away from this area. It seems that after approximately 200 ns the ECL2 of the top ECD 'slides' towards the β2 strand of the bottom ECD. Note that this system (φ=180) demonstrates the smallest number of contacts (see Table 4.2 and Figure 4.24) compared to the others, the 'key' residues being proline 151, methionine 152 and isoleucine 40. The system also has the smallest number of formed salt bridges alongside the phi0 system.

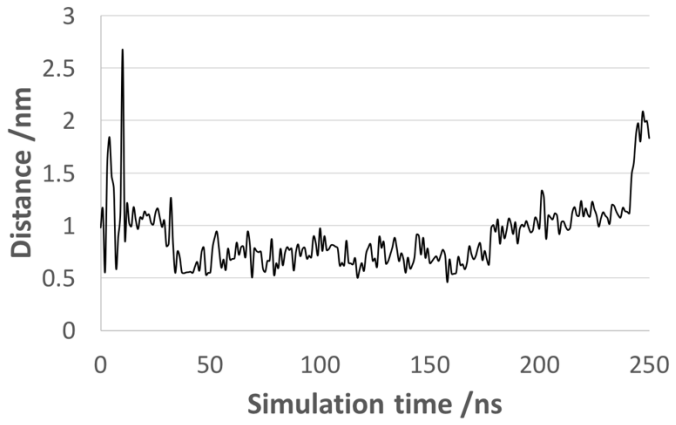


Figure 4.31: The graph displays the distance between the residues methionine 152 of the opposing ECDs of the phi180 system. It is clear that the interaction is strong, however, towards the end of the trajectory the ECDs rearrange their relative position.

Considering the phi240 system, their initial arrangement was biasing the ECDs to interact through their second loop since it was the part of them that was placed closer (Figure 4.5). Indeed this area was the first interaction site, but the ECDs did not stay in this conformation but rather penetrated deeper and in the end frame the top ECD was closer to the area above the β1 and β2 strand. The Find Contacts tool from Chimera pointed out that methionine 152, tyrosine 33, leucine 64 and phenylalanine 67 were the residues that form the largest number of contacts with the opposing ECD. In this system,

the RMSD value for the bottom ECD was high and we notice that the helical part of the ECD has partially unfolded. There were also 14 salt bridges formed during the simulation, which is the largest number of salt bridges formed amongst all other systems comprising the rotation study.

Lastly, the phi300 system displayed a similar behaviour with the phi240 system, regarding the preferred interaction site but the top ECD displayed a large RMSD value (see Figure 4.22). The top ECD becomes slightly more 'compact' as seen in the reduction of its radius of gyration (see Figure 4.23) and its edges (long helix and β4 strand) appeared closer, compared to its initial conformation. The important residues were valine 155, proline 151, phenylalanine 148 and methionine 152, they are all located at the second smaller loop of claudin-1, and they formed the largest number of contacts with the opposite ECD. The average penetration depth was small (approximately 1.1 nm) because the top ECD interacted with the middle part of the bottom ECD rather than its ends, so the movement of the top ECD was restricted by the bottom ECD (see Figure 4.25). This systems formed 13 salt bridges during the trajectory.

From the detailed description of all systems comprising the rotation study, we notice that phi0 stands out as a stable configuration. The ECDs demonstrated a large number of contacts between them, small RMSD and radius of gyration values, a big penetration depth and the system generally seemed to be 'settled' in this interaction interface from the onset of the simulation till the end of the trajectory. From Figure 4.32 we notice that the trans-interaction between the ECDs is compatible with the potential interfaces that could be formed while two claudin proteins are embedded in the membrane, meaning that the tilt angles of the proteins are not high. In the picture both ECDs are coloured differently and the respective protein is superimposed on each ECD to give a clearer picture of how the end structure of the trans-interacting ECDs would be if the rest of the protein body was present.

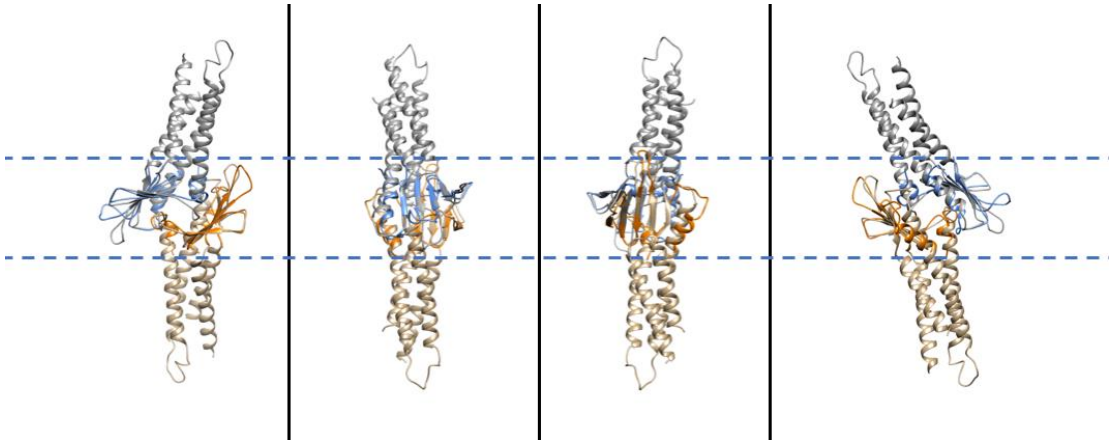


Figure 4.32: The whole claudin-1 proteins superimposed on the ECDs from the phi0 system (end frame). The blue dashed lines represent the hypothetical boundaries of the opposing cell membranes. Each element (ECDs and proteins) are coloured differently in cartoon representation. Specifically, the bottom ECD is orange and the bottom protein is gold, while the top ECD is light blue and the top claudin is grey. Rendered with VMD from different side views (Humphrey et al., 1996).

We noted that the end structure of the phi0 system seems to have a small pore (seen in Figure 4.29 and Figure 4.32). We have used the MOLEonline web interface that is used for characterisation of channels, tunnels and pores in macromolecular structures (Berka et al., 2012). For the specific system it calculated various tunnels as seen in Figure 4.33 but the bottleneck of all was smaller than 0.12 nm and none seemed to be through the whole pore structure. This clearly displays the barrier properties of claudin-1 in the extracellular space.

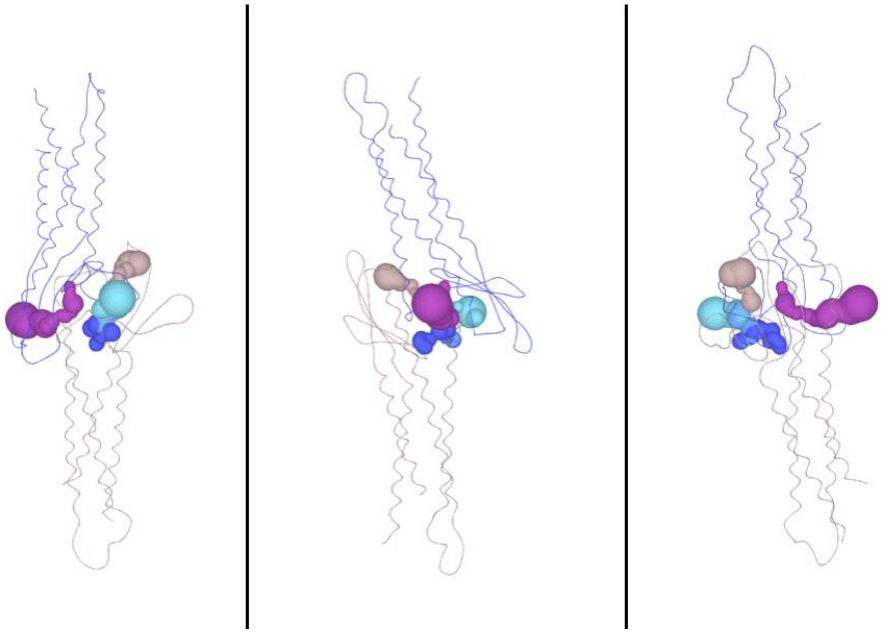


Figure 4.33: The figure displays some of the tunnels in different colours, calculated with the MOLEonline platform for the phi0 system. The two proteins are in cartoon representation.

It should be emphasised that there are limitations in this second study. The top ECD was systematically rotated with an angle (phi, φ) spanning 0 to 300 degrees but in order to avoid clashes with the bottom ECD we had to translate the top ECD. Usually the top of the second loop of claudin-1 would clash with the respective second loop of the bottom ECD. The possibilities to translate the top ECD are numerous because it can be translated both in the x- and y- direction with varying distances. Therefore, the number of systems here may not be sufficient to observe all the possible trans-interfaces between claudin-1. Additionally, the ECDs were restrained on a line (z-axis) so, they could not move far away from this line. As a consequence no significant change in the orientation between the ECDs was observed prior to interaction. Nevertheless, the study emphasised that the trans-interaction of claudin-1 results in very compact structures that show a barrier property filling the extracellular space between cells.

4.3.3 Binding energies of the trans–dimers isolated from the self-assembly simulations and the rotation study

A study of single molecule force spectroscopy demonstrated that claudin-1 molecules form weak and short-lived interactions, thus, supporting the model of TJs dynamic nature, where the strands break and reseal the intercellular gap, rather than being a static complex (Lim et al., 2008b). In the case of adherens junctions, which keep the cells attached, the family of cadherins is important for stabilising the connection between adjacent cells. The bond strength of E-cadherins, the fundamental proteins found at the adherens junction, was found to be 39 and 51 pN at a rate of $0.01s^{-1}$ while claudins had a bond strength of 21 and 48 pN at a rate of 1.35 s^{-1} (Lim et al., 2008b). It was also demonstrated in the aforementioned study, that the homophilic claudin-1/claudin-1 bond shows a higher dissociation rate compared to E-cadherins. However, Panorchan et al., showed that N-cadherin and E-cadherin molecules show differences in their binding mechanisms, both quantitative and qualitative (Panorchan et al., 2006). Specifically, they found that E-cadherin-E-cadherin bond can withstand mean forces up to 73 pN (for a loading rate of 1,000 pN/second) and 157 pN (for a loading rate of 10,000 pN/sec). The N-cadherin-N-cadherin bond could withstand forces up to 30 and 40 pN for the same loading rates. The values are quite different than the previous reported ones. Note that the basic principle of single molecule force spectroscopy uses a molecular force probe to probe the cell-cell interactions, where a cell is deposited on a cantilever and the cantilever is positioned over another target cell. The cantilever moves and the breaking of the bonds causes deflections of the cantilever which are translated into time-dependent forces and recorded as a function of the distance between cells (Panorchan et al., 2006). The force is the height of the peak in the force-distance curves while the loading rate is related to the velocity of the cantilever.

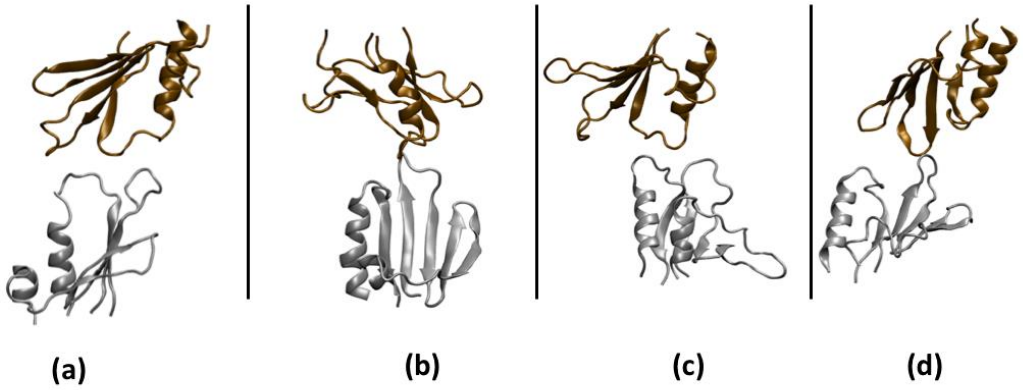


Figure 4.34: The isolated trans-interacting dimers from the big grid systems. The ECDs are in cartoon representation and coloured differently. The rendering was done with VMD (Humphrey et al., 1996).

To estimate the relative binding energies of the formed trans-dimers we used the MM/PBSA method (Kumari et al., 2014). The method was employed for both the dimers isolated from the self-assembly simulations (grid systems) and for the ones from the rotation study. Four indicative dimers were chosen from the grid systems (see Figure 4.34) that interacted via the top of the ECDs and specifically via the V1 and V2 areas. Only two dimers showed marginally favourable binding energy (-2 KJ mol^{-1} and -17 KJ mol^{-1}), specifically the second and fourth dimer (Figure 4.34 (b) and (d)) when the single frame approach was used. This slightly favourable (small) binding energy is in agreement with the study of single molecule force spectroscopy mentioned previously that revealed relatively weak interactions between claudins. When we examined the binding for the last frames of the trajectory, we mostly obtained unfavourable binding energies.

The binding energies demonstrated by the dimers isolated from the rotation study were larger (see Table 4.3). The energies were mostly in the range of -45 to -52 KJ mol^{-1} but note that the binding energies of two systems were rather high (approximately -230 KJ mol^{-1}). Specifically, the last 10 frames of the trajectory were isolated and were used to predict the binding energy between the ECD particles (analysis over 1 ns). Based on their binding free energy the ranking was phi0> phi240> phi180> phi120> phi60> phi300. Both phi0 and phi240 show a high binding energy almost four times larger than the other systems.

Table 4.3: The binding energies of isolated dimers from the phi0-phi300 systems examined with the MM/PBSA method. The analysis was done over the last 10 frames of the trajectory (1ns).

System	**Binding energy/ KJmol^{-1}**
phi0	-237 $\pm$ 28
phi60	-47 $\pm$ 57
phi120	-48 $\pm$ 64
phi180	-52 $\pm$ 52
phi240	-231 $\pm$ 44
phi300	-45 $\pm$ 42

To put these results in perspective, the trans-interaction of claudin-5 has been reported to be -8.3 kcal mol^{-1} (-34.73 KJ mol^{-1}) using the interactive webserver PISA while the E-cadherins showed a binding energy of -60.3 kcal mol^{-1} (-252.3 KJ mol^{-1}) (Irudayanathan et al., 2017, Schymkowitz et al., 2005, Krissinel and Henrick, 2007). Furthermore, in Chapter 3 we calculated the binding energy for the cis-dimers to be in the range of -58 to -21 KJ mol^{-1} with the accurate PMF method and -166 to -20 with the MM/PBSA method.

The MM/PBSA results demonstrate that the most stable dimers observed are the phi0 and phi240 systems that show a similar binding energy which is significantly higher than the one of the other systems. The other systems show similar values, ranging from -52 to -45 KJ mol^{-1}. Note that phi240 had the highest number of formed salt bridges. The method suggests that the binding energies for the trans-interaction, as calculated for the phi0-phi300 systems, are higher compared to the ones calculated for the cis-dimers when phi0 and phi240 are considered, a result that was not expected. But when the penetration depth is small (as observed in the self-assembly grid systems) the binding energy is small, smaller than the respective one for the cis-dimers. However,

due to the limitations of the MM/PBSA method we report these binding energies with caution.

To conclude, when the penetration depth is small (as in the dimers isolated from the self-assembly simulations) the interacting ECDs display a small binding energy, while in the case of the rotation study, the binding energy is high. Specifically, for phi0 and phi240 considerably higher than the binding energy calculated for the cis-dimers in Chapter 3.

4.4 General Comments and Conclusions

The trans-interaction between claudins regulates the barrier and channel properties of the TJ barrier/pore. In this chapter, atomistic MD simulations were carried out in an implied bilayer to characterise the formed trans-interfaces between the ECD particles of claudin-1. The ECD particles, which are the main regulators of the trans-interaction, were free to interact and able to form a variety of interfaces. The novelty of the study was that the (trans-)interaction was examined with atomistic resolution and without dictating the ECDs initial arrangement based on a proposed model.

The prevailed Suzuki model shows that a single pore is formed when 4 claudin monomers interact (2 from each cell) and a double pore is formed between 8 interacting monomers (4 from each cell) (Alberini G, 2017, Suzuki et al., 2015). The double pore involves both (cis-) interfaces, namely the ‘face to face’ and ‘linear’ interface (Suzuki et al., 2015). Other studies have suggested alternative interfaces with a slightly different arrangement between claudins (e.g. different relative angle between the monomers) (Zhao et al., 2018, Irudayanathan et al., 2018, Piontek et al., 2017). Here, the exact single or double pore structure proposed by the Suzuki group was not observed, although some similar interfaces between claudin dimers did occur (Figure 4.32). Note that in the rotation study only two ECDs were examined. Moreover, the Suzuki model refers to mouse claudin-15, which forms pores while claudin-1 is a predominantly barrier forming claudin. The sequence similarity between them is approximately 33% measured with Crystal Omega an on-line tool for sequence alignment (Sievers et al., 2011). The barrier structure was dense (see Figure 4.35) but the system needed more time to fully develop a compact barrier as seen in CG models of the ECDs, and perhaps form more stable interactions that might resemble the Suzuki model.

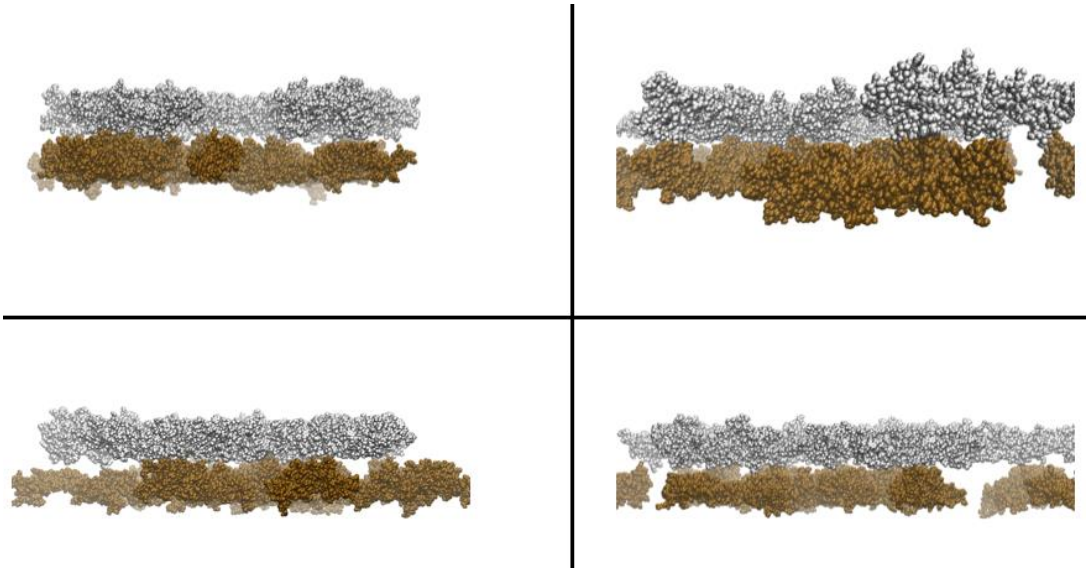


Figure 4.35: The barrier formed by claudin-1 ECDs. The top and bottom layers are coloured differently and the protein atoms are shown as vdW spheres. Rendering was done from different perspectives.

With regards to the results from the grid study, we noticed that the ECDs quickly interacted on the same plane forming dimers and occasionally trimers and some also formed trans- contacts with the atoms from the opposing grid of proteins. The calculated RMSD and RMSF average values, were consistent with the ones previously reported in literature and Chapter 3 of the thesis. The morphology of the strand was also in agreement with other computational modelling studies and the previous chapter. The cross-linked network of particles was reproduced, and some areas were more dense compared to others. Moreover, the angle distribution graphs demonstrated that there is a plethora of formed cis-interfaces, failing to highlight a single favourable interaction site. However, the applied position restraints on the terminal atoms were strong, preventing the ECDs from penetrating deeper into the opposing layer and the isolated dimers were characterised with marginally favourable binding energies. The important trans-interfaces were the variable areas V1 and V2, which are also the ones that are located higher compared to the rest protein atoms of the ECD.

The grid systems and the systematic rotation study have both highlighted the variable loop regions (V1 and V2 areas) as the ‘key players’ of the trans-interaction of claudin-1. Methionine 152 from the V2 region seems to be a ‘key’ residue for the trans-

interaction. This finding is also supported by experimental studies of claudin-1 where it interacts with a binder, specifically a monoclonal antibody (Nakajima et al., 2015). We further identified residues of the area close to the small helix (71-75) as well as the aromatic amino acids tyrosines 33 and 35 as 'key' residues stabilising the trans-interface in the rotation study. In another study that used a combination of experiments and in computational modelling methods, tyrosine 33 (alongside the critical amino acids 63-66) was found to be important because it interacts with lysine 148 of the tetraspanin CD81 forming a suggested complex in cells (Davis et al., 2012). Note that this complex of claudin-1/CD81 is suggested to be outside the TJs, located at non junctional pools of claudin-1 and it is an entry point in the host cells for the Hepatitis C virus.

The rotation study concluded that phi0 and phi120 are stable configurations showing a large number of contacts and that the ECDs largely maintain their arrangement (small RMSD values). In contrast to some other orientations (e.g. phi180) where the ECDs display a dynamic behaviour and seem to be unsettled. The aforementioned two systems were highlighted as more favourable configurations compared to the others where the ECDs displayed a large deviation from the initial structure or they constantly tried to rearrange themselves throughout the trajectory. Those two interfaces although different as seen in Figure 4.30 show that the variable regions and the area close to the small helix, are important areas for the trans-interaction of claudin-1. It should be mentioned that Veshnyakova at el., found that the substitution of glutamic acid 48 and serine 53 with other amino acids (E48K and S53E respectively) increased the permeability of ions through the pore (Veshnyakova et al., 2012). It was suggested though that these mutations likely affect the electrostatic propensity and do not change the structure of claudin-1 (Veshnyakova et al., 2012, Irudayanathan et al., 2017). Both amino acids are not key for the trans interfaces observed here and are located in the centre of the formed structure in phi0 and phi120 systems (see Figure 4.30 and Figure 4.32). Specifically, in the symmetrical end structure of the phi0 system they are located at the same side of the pore and possibly towards the entrance of the pore.

In the rotation study, the ECDs could penetrate further than in the grid systems. The identified dimers of the grid study could likely be intermediates with small lifetimes, while the more compact structures-dimers identified in the rotation study could be more favourable and stable interactions (the binding energies are in agreement with such

hypothesis). The binding energies (except the ones for phi0 and phi240) were similar to the ones reported in Chapter 3 that considered the cis-interaction. The results suggest that claudin-1 forms tight barriers consistent with its predominantly barrier function (Günzel and Alan, 2013, Inai et al., 1999).

The importance of the results is demonstrated by the fact that although some claudins show high sequence identity they do not form the same interactions and thus, each claudin should be examined separately. Each individual claudin however, is thought to favour some interfaces compared to others and the results obtained here are specific for claudin-1. The distinguished features of phi0, phi180 and phi240 the energetically most stable interacting pairs observed, as well as the phi120 which was also characterised important, were characterised in detail. The trans-interfaces identified here are likely to play crucial roles in TJ assembly.

Currently, the architecture of TJ pore cannot be elucidated due to limitations in isolating a TJ strand and visualising it. The molecular organisation of the TJ proteins is expected to be complex, especially if ones considers the fact that there are many different claudins expressed in a specific tissue as well as other TJ proteins. It is now clear, that claudin self-assembly involves a plethora of formed interfaces, which are required for strand formation and the inherent flexibility that these dynamic structures show (Zhao et al., 2018). However, the trans-interaction between claudin-1 is dictated from the specific regions highlighted here. The trans-interaction between different claudins (homotypic or heterotypic interactions) could regulate the pore structure and perhaps the suggested ‘opening’ and ‘closing’ events of the pore. The key residues mentioned here are found to facilitate the homotypic binding between claudin-1 and could serve as good indicators of the heterotypic binding of claudin-1 with other proteins from the claudin family.

5 Self-Assembly of Claudin-1 in Lipid Bilayers

5.1 Introduction

In this chapter, we investigate the claudin-claudin interactions whilst the claudins are embedded in lipid bilayers, and how the different bilayer lipid compositions affect these interactions. These are relatively large-scale simulations and have been carried out using coarse-grained (CG) molecular dynamics. These simulations have been supplemented with atomistic simulations of the pure lipid bilayers and a single claudin particle embedded in each of the bilayers. The focus is on the lipid membranes of the skin epidermis, in particular the *stratum granulosum* (SG) where the tight junctions (TJs) exist. In Chapters 3 and 4 of the thesis, we addressed the question of how the TJ proteins self-assemble in an *implied* bilayer, focusing on the proteins' extracellular region only. The advantage of investigating the behaviour of only the extracellular domain particles (ECDs) was that we could examine the self-assembly of a relatively large system in *atomistic* detail. However, there might be cases where the transmembrane region of the claudin stabilises the already formed interaction sites between the claudins, and as a result they might slightly change their orientation. For instance, the transmembrane region has been found important for the self-assembly of claudin-5 with a note that the formed interfaces are largely dominated by their ECDs when the claudins are embedded in complex bilayer systems (Irudayanathan et al., 2015). Earlier it was suggested that non-conserved residues from the second extracellular loop (ECL2) and the third transmembrane region of claudins-3 and -5 are important for their folding and assembly into the TJs (Rossa et al., 2014). As noted in previous chapters, the ECDs (heads of the 'Olympic torches') are indeed larger in diameter and mostly overhung the transmembrane region of the protein. Therefore, they should dictate the formed interfaces as the two claudins approach each other. This

chapter investigates whether the interfaces formed in the previous studies are observed here too, and whether some are more popular than others. Furthermore, we investigate how the different lipids of the bilayer affect the spatial position of the claudins with regards to their tilt angle and their hydrophobic depth.

The study systematically investigates systems of increasing complexity, starting with a pure phospholipid bilayer and then including other relevant lipids with a view to mimicking the membrane micro-environment of epithelial cells found in the skin epidermis. Lipids are very diverse in their structure and function, so studying a very complex multi-component bilayer can be challenging. It is difficult for example, to examine local deformations in such inhomogeneous systems, induced by differences in the length of the acyl chains of the different lipids. Nevertheless, we hope that by strategically choosing the lipids (or a representative of a group of lipids) that predominate in the cell membrane of interest, we can develop useful insights. Resolving protein-protein interactions whilst embedded in complex lipid bilayers is challenging and it stretches the computational resources that we have available.

5.1.1 Skin strata and their lipid composition

Biological membranes are complex systems, comprising a variety of lipids (and other components such as proteins and carbohydrates) that might exist in various phases and could affect the self-assembly of any transmembrane protein complex. The diversity in lipid composition and the ability of computer simulations to address such systems is an area of interest due to the offered molecular resolution of the simulations comparable to experimental methods such as X-ray diffraction and therefore, simulations can be used to interpret experiments (Bennett and Tieleman, 2013).

TJs along with the *stratum corneum* (SC), the outermost layer of our skin, compose the two major mechanical barriers of our epidermis although for a long time the epidermal barrier has been attributed only to the lipids of the SC (Kirschner et al., 2010, Yoshida et al., 2013, Brandner, 2009). These mechanical shields alongside the skin microbiome and the chemical and immunological barrier, act together to effectively protect our body from external threats and loss of internal fluids (Bäsler et al., 2016). However, in the

case where the TJs are of primary interest (e.g. when the SC is breached or when examining the barrier function of hair follicles), TJs alone constitute a major 'gate' keeper in order to protect the underlying tissue (Zorn-Kruppa et al., 2018). As previously stated, TJs are localised in the SG, right below the SC, which is the layer of non-viable cells called corneocytes. These enucleated cells are embedded in a complex lipid matrix ('brick and mortar' model) and are of great importance since they compose the first line of defence of our organism. The lipid composition and structure of the SC have been extensively studied and differ to the one viable cell bilayers exhibit. The dense lamellar structured lipid matrix is an approximate of 1:1:1 molar ratio of ceramides: free fatty acids: cholesterol making evident the lack of phospholipids. MD simulations have been employed to better understand the behaviour of this complex mixture of lipids with a view to understand how these lipids are organised (Lundborg et al., 2018).

Thus, focusing on the skin epidermis, although there are many computational studies examining the molecular organisation of the epidermal lipids, most of these studies focus on the SC (Lundborg et al., 2018, Notman and Anwar, 2013, Iwai et al., 2012, Gupta et al., 2016, MacDermaid et al., 2015). In contrast, the TJs are found in the SG for which limited information exists regarding its composition and structure.

Other model membranes have also been studied with computer simulations but model bilayers with a composition close to the lipid composition of SG has never been examined to the best of our knowledge. The composition between the two layers is significantly different, because during keratinisation (a process that epithelial cells undergo in our epidermis) extensive changes happen in the morphology and histochemistry of cells as well as in their lipid composition (Feingold, 2007).

The lamellar bodies of the epidermis play a crucial role in the lipid changes that happen in the skin and the epidermal permeability barrier (Raymond et al., 2008b). The lamellar bodies are specialised secretory organelles that contain phospholipids, glucosylceramides, sphingomyelin and cholesterol (Feingold and Elias, 2014, Feingold, 2007). Lamellar bodies secrete their contents in the upper keratinocyte layers (Raymond et al., 2008a). The lipids of the lamellar bodies are the precursors of the SC extracellular lipids. The phospholipids are broken down by phospholipases to free fatty acids and glycerol, and glycoceramides and sphingomyelin are broken down to ceramides.

Table 5.1: Lipid composition of isolated lipid fractions from neonatal mouse stratum granulosum and stratum corneum (Elias et al., 1979). The values are average weight percent of total lipid recovered from thin-layer chromatography plates and standard deviations from three experiments.

	Stratum Granulosum / %w/w	**Stratum Corneum / % w/w**
Phospholipids (subtotal)	5.5%	5.2%
Phospatidylethanolamine	1.77 ± 0.15	1.96 ± 0.84
Phospatidylcholine	2.27 ± 0.23	1.67 ± 0.74
Sphingomyelin	1.47 ± 0.30	1.63 ± 0.42
Neutral lipids	25.3%	48.4%
Sterol esters	3.43 ± 1.27	6.63 ± 1.32
Triglycerides	2.57 ± 0.25	3.87 ± 0.90
Free fatty acids	2.97 ± 1.14	6.87 ± 1.98
Free sterols	3.87 ± 0.15	9.07 ±5.32
Glycosphingolipids	7.03 ± 1.06	16.8 ± 5.95

Table 5.2: Variations in lipid composition during human epidermal differentiation and cornification (weight ± standard error of mean)(Lampe et al., 1983).

	Stratum Granulosum / % w/w	**Stratum Corneum / % w/w**
Polar lipids	25.3 ± 2.6	4.9 ±1.6
Cholesterol sulfate	5.5 ± 1.3	1.5 ± 0.2
Neutral lipids	56.5 ± 2.8	77.7 ± 5.6
Sphingolipids	11.7 ± 2.7	18.1 ± 2.8

With regards to the lipid composition of SG, there are two rather outdated experimental studies (see Table 5.1 and Table 5.2) that form the basis of our study. But we incorporate new knowledge considering the composition of other skin layers (e.g. SC), bearing in mind that the model membranes that we wish to study need to be relatively simple. The lipid environment could play an important role in the aggregation of the transmembrane claudins by changing their relative position in the bilayer. The self-assembly of proteins is driven by fundamental protein interactions such as hydrophobic, electrostatic, van der Waals and hydrogen bonding, as well as, thermodynamic and kinetic factors (McManus et al., 2016).

In this chapter, MD simulations were employed to examine the way claudins self-assemble into TJ strands whilst embedded in lipid bilayers using CG models. It is hypothesised that different lipid environments will affect the resulting strand network of claudins due to the varying bilayer thickness as well as the ability of different lipids to interact with specific residues of the claudin. Specifically, we examined claudin-1 that is a major constituent of TJs and is required for the normal barrier function of our skin (Furuse et al., 2002). The focus was on the lipid membranes of the SG. However, we acknowledge two facts, that the lipid composition of SG is largely unknown, and that claudins occur in many different lipid membranes, so, there is a possibility that

claudins are not matched with a particular lipid environment. Thus, it is not crucial to match claudin-1 with a model lipid bilayer of specific composition. But rather the key components added should be prevalent in biological membranes. The self-assembly was also examined with metadynamics. Furthermore, we examined the effect of claudin concentration in the formed interfaces and strand's ultrastructure.

5.2 Methodology

We carried out five sets of simulations:

i. Atomistic simulations of pure lipid bilayers that were considered to be representative of SG lipids and which served as controls.
ii. Atomistic simulations of a single claudin-1 particle embedded in the selected lipid bilayers.
iii. Self -assembly simulations of claudin-1 particles in each of the selected lipid bilayers using CG models.
iv. Self-assembly simulations of claudin-1 particles in 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) bilayers as a function of three concentrations of claudin-1 using CG models to look at supersaturation effects.
v. Metadynamics simulations to resolve the self-assembly of claudin-1 using as a collective variable the separation distance between the centre of mass of the claudins.

5.2.1 The five distinct sets of simulations examined

The first set of simulations comprised a pure lipid bilayer containing only 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) which we label as System_1, then we added an additional component to the bilayer, specifically cholesterol (System_2) and glycolipid (System_3) (see Table 5.3 and Figure 5.2). Cholesterol is abundant in cell membranes and is one of the most studied molecules as a part of model membranes because it is associated with the lipid raft domain hypothesis (Berkowitz, 2009, Bennett and Tieleman, 2013). Cholesterol sulfate was found in peak levels in SG (Lampe et al., 1983). Glycolipids play an important role in various cellular functions and are often found on the surface of cell membranes (Lee et al., 2018). Glycolipids consist of a lipid

and a carbohydrate and can be subdivided into several classes. In this study, we added a glycosphingolipid where its lipid part was a ceramide, specifically, ceramide181 that has two acyl chains with 18 carbon atoms and a double bond each (similarly to DOPC) and the carbohydrate was glucose. The experimental paper that describes the lipid composition of skin epidermis revealed that in all epidermis layers examined, the fatty acid composition ranged from C12 to C24 with C16-18 representing the major species; and specifically oleic acid (C18:1) was found in large quantities (Lampe et al., 1983). Figure 5.1 shows the structure of the aforementioned lipid molecules generated with ChemDraw (Li et al., 2004).

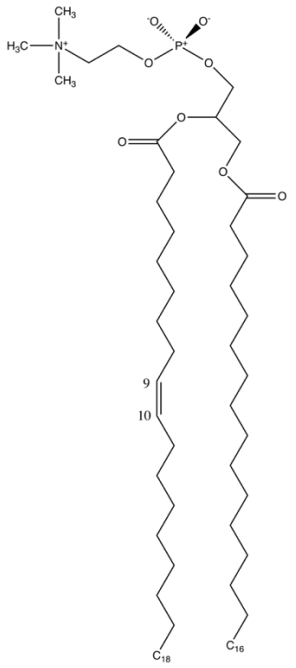
Table 5.3: The lipid composition of the model bilayer systems studied alone and with a claudin particle embedded. System_1 had only DOPC molecules (pure lipid bilayer), while systems_2 and_3 had DOPC molecules plus cholesterol and glycolipids (GLPA) respectively. The number of water molecules (TIP3P) is also shown. The number of lipids in each leaflet is reported in the brackets.

	System_1	**System_2**	**System_3**
DOPC	128 (64+64)	108 (54 + 54)	116 (58 + 58)
CHOLESTEROL		20 (10 + 10)	
GLPA			12 (6 + 6)
TIP3P	5016	4714	6578

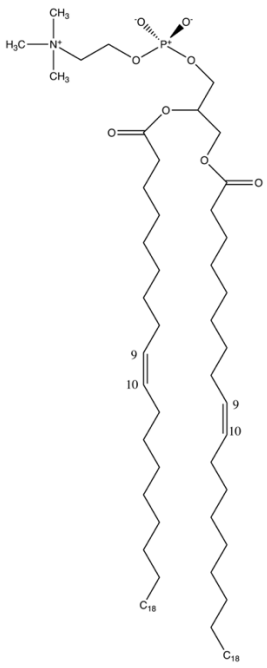
With regards to their concentration, the number of cholesterol molecules was based on a 15% mol concentration (128 * 0.15 = 19.2 molecules) while the glycolipids had a 10% mol concentration (128 * 0.1 = 12). The solubility of cholesterol in DOPC multilayers is approximately 40% in mol fraction (Hung et al., 2007)) but the percentage of both molecules vary in different cell membranes. The approximate lipid composition was calculated based on a similar concentration to the one found in the endoplasmic reticulum (Alberts et al., 2015) and observations from the two experimental papers

(Elias et al., 1979, Lampe et al., 1983). The lipid composition of isolated fractions of SC and SG are presented in Table 5.1 and Table 5.2, copied from the two aforementioned experimental studies, where we can see, for example, the weight percentage of glycosphingolipids is approximately 10 % w/w. We chose to add these lipids and studied their lipid-lipid interactions and protein-lipid interactions as they serve as representative lipid types in the SG, though we appreciate the complexity of SG lipids is substantial. The expectation was that these representatives lipid types present in a moderate concentration would enhance our understanding of the aforementioned interactions that lead to the self-assembly of claudin-1.

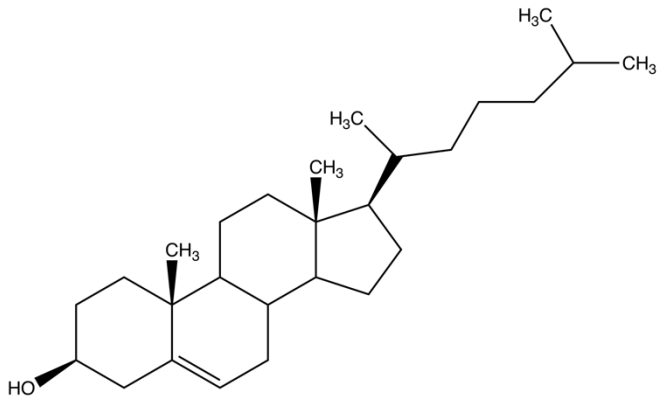
The next set of simulations composed of a single claudin particle embedded in the model bilayers and we mostly focused on the behaviour of the claudin-1 (see Figure 5.3). The same model of claudin-1 with the previous studies was used. Subsequently, the self-assembly of claudins in lipid bilayers was investigated with CG models. The CG systems comprised 8x8 grids of separated claudins (centre of mass separation distance of approximately 7 nm along the x- and y-plane) embedded in three different lipid bilayers of composition similar to the atomistic systems (see Table 5.4 and Figure 5.17). CG_1 had DOPC lipids, CG_2 also had cholesterol molecules added, while in CG_3 we examined the general model of a glycosphingolipid (monosialotetrahexosylganglioside lipid (DPG1)) corresponding to the atomistic (18:1/18:0). The systems were simulated in the *NPT* ensemble for a total simulation period of 8.5 microseconds (CG_1 and CG_2) and 10 microseconds (CG_3) using a time step equal to 15 femtoseconds.



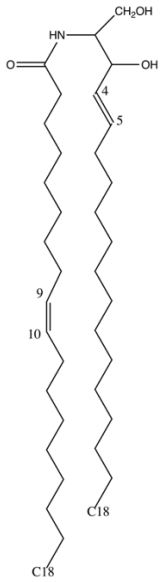
(a) POPC



(b) DOPC



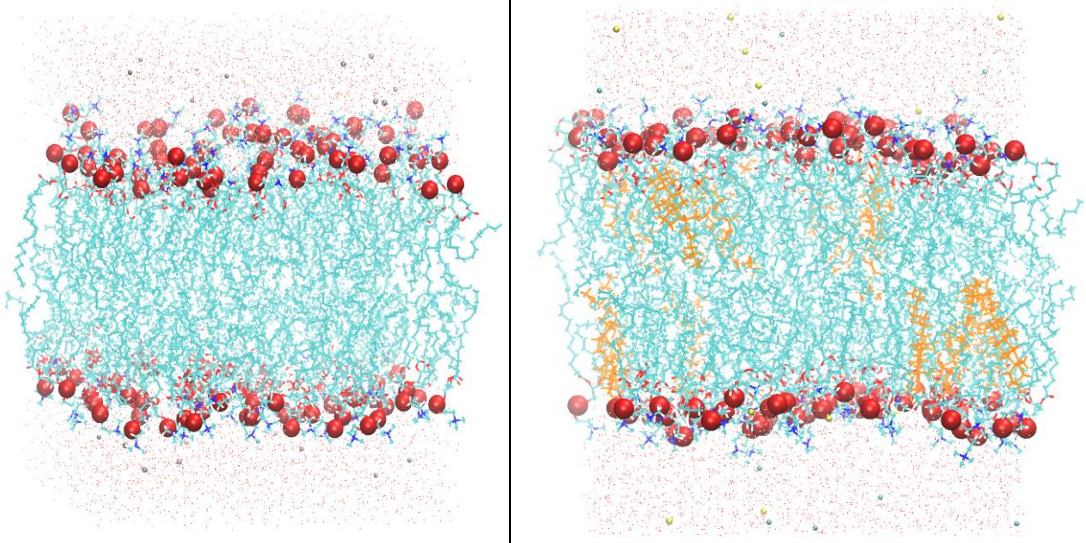
(c) CHOLESTEROL



(d) CER181

Figure 5.1: The structure of the major lipid types examined in this study. (a) 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine, POPC (16:0-18:1 PC), (b) 1,2-dioleoyl-sn-glycero-3-phosphocholine, DOPC (18:1 (Δ9-Cis) PC), (c) cholesterol and (d) ceramide 181, N-stearoyl-D-erythro-sphingosine (d18:1/18:1).

All systems were solvated in water with added ions at a physiological concentration of 0.15M NaCl. Some major guidelines that we complied with when we set up the systems, was that DOPC was reported to have a ratio of 32.8 water/lipids (Kučerka et al., 2006) and that it is advised that the phospholipid and glycolipid to have roughly the same chain length for homogenicity. Additionally, the lipids should be evenly distributed in both bilayer leaflets (see Table 5.3).



(a)	(b)

Figure 5.2: (a) The DOPC pure bilayer system and (b) the DOPC and cholesterol bilayer system rendered with VMD. The DOPC molecules are shown in line representation coloured opaque, while the cholesterol molecules are also in lines but coloured orange. Water molecules are shown as red points and ions as yellow CPK spheres. The phosphorus atoms of the DOPC headgroups are shown as red vdW spheres.

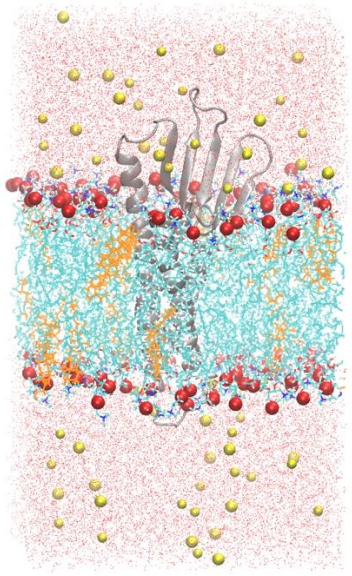


Figure 5.3: Claudin-1 embedded in a complex bilayer composed of DOPC and cholesterol molecules. Water molecules are red points, ions are represented as yellow CPK spheres, DOPC molecules are opaque lines while cholesterol is coloured orange. Claudin-1 is in cartoon representation and coloured silver. The image was rendered with VMD (Humphrey et al., 1996).

Table 5.4: The composition of the three examined CG models. Each system was initially set up as a grid of 8x8 separated proteins.

	CG_1	**CG_2**	**CG_3**
PROTEINS	64	64	64
DOPC	8448	7168	7680
CHOLESTEROL		1280	
DPG1			768

Table 5.5: The composition of the systems that had claudin-1 embedded in POPC lipid bilayers to examine supersaturation effects.

	POPC_1	**POPC_2**	**POPC_3**
Claudin-1	36	36	36
POPC	6984	2268	1584

We also studied another set of simulations which investigated the self-assembly of claudin-1 embedded in POPC lipid bilayers as a function of claudin-1 concentration, to investigate whether supersaturation can affect the emerging morphology of the claudin strands. The parameters for the simulations were the same as in the previous CG systems. The grid systems comprised 6x6 separated claudins with a varying centre of mass separation distance. The number of lipids per system is presented in Table 5.5. It was not possible to set up a more dense box because it would have resulted in inadequate lipids surrounding the protein. The water molecules were in the range of approximately 36000-130000. The POPC_2 and POPC_3 systems were simulated in the *NPT* ensemble for 5 microseconds, while POPC_1 was extended to 10 microseconds.

The final set of simulations employed, aimed to elucidate the aggregation of claudins with well-tempered metadynamics. In these studies, we employed GROMACS patched with the plug-in PLUMED (Bonomi et al., 2009) used to run metadynamics simulations (further on metadynamics on the next chapter). The systems were four in total each composed of two claudins embedded in pure POPC lipid bilayers (348 lipids) where we looked at different interfaces (see Figure 5.4). The position of one protein was kept fixed with a force equal to 1000 KJ mol^{-1} nm^{-2}. Each system was simulated in the *NPT* ensemble for a total simulation period of 5.5 microseconds using a time step equal to 15 femtoseconds. The distance between the centre of mass of the proteins (in the x-y plane) was originally 7.5 nm and a bias potential was added to the distance to enhance sampling. The chosen collective variable i.e. the Pythagorean distance between the centre of mass of each claudin, was chosen to describe their aggregation. The width of the Gaussian potential was 0.2 (sigma, σ=0.2), the Gaussian height was set to 1.2 KJ

mol^{-1}, the bias factor was 10 and the stride of Gaussian deposition was 200 (more details on these values in Chapters 2 and 6).

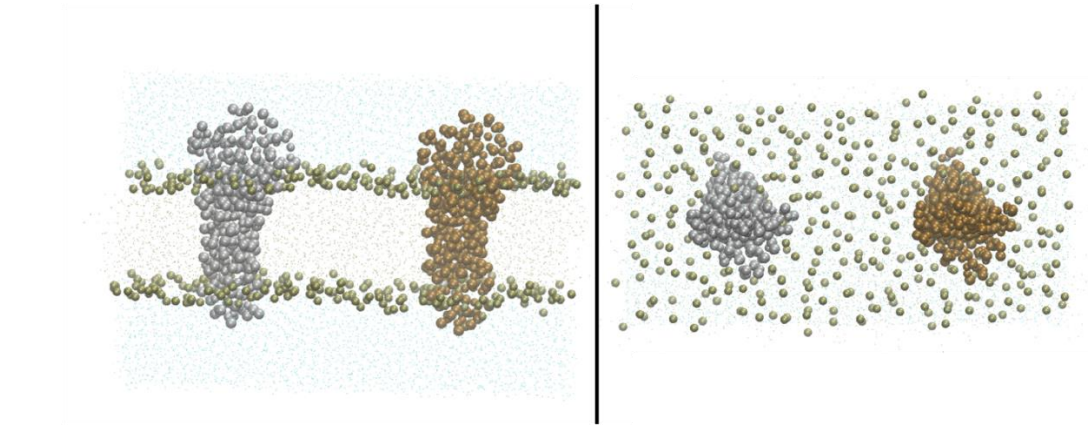


Figure 5.4: Two proteins embedded in a POPC lipid bilayer from a side and top view (CG models). Ions and water molecules are blue dots, the phosphate head groups are vdW spheres and the claudins are shown as silver and brown vdW spheres.

5.2.2 Technical details of the molecular dynamics simulations

The parameters for the proteins, lipids and ions (atomistic simulations) were based on the CHARMM36 force field (Best et al., 2012) while for the CG models we used the Martini force field (Marrink et al., 2007, Monticelli et al., 2008) coupled with the ELNEDIN network to keep the tertiary structure of the protein fixed (Periole et al., 2009). Note that with regards to lipid parameters, the CHARMM force field is the only force field reported in a study to produce the chain inequivalence (*sn-1* being the leading chain) compared to the other widely used force field for lipids (i.e. Berger lipid parameters) (Berger et al., 1997, Pezeshkian et al., 2018, Lee et al., 2014, Klauda et al., 2010). The water model used was TIP3P matching the CHARMM force field parameters. All simulations were run with the MD code GROMACS (Abraham et al., 2015, Bjelkmar et al., 2010). The membrane systems were generated with the membrane set up tool CHARMM-GUI (Lee et al., 2018). During the atomistic simulations, the temperature was set to 310K controlled through the Nose-Hoover thermostat and the pressure was 1 atm controlled through the Parrinello-Rahman barostat, while for the CG systems with regards to temperature coupling, we employed velocity rescaling (v-rescale) with a stochastic term to ensure that a proper ensemble

was generated. The coupling groups were protein atoms (where applicable), lipids, and solvent molecules along with ions. Note that the pressure coupling was semi-isotropic (the pressure coupling is isotropic in the x- and y- direction (the plane of the bilayer) but different in the z-direction). For the atomistic simulations, the LINCS algorithm was used to constrain the H-bonds and a 1.2 nm cut-off distance was set for the van der Waals (vdW) and Coulombic interactions. The particle-mesh Ewald summation method was used to calculate the electrostatics using cubic interpolation (ewald r-tol set to 10^{-5}). While for the CG runs, the equivalent cut-off was set to 1.1 nm for the coulombic interactions which employed a reaction-field and 1.1 nm for the vdW interactions (cut-off scheme with potential-shift-Verlet). All systems were energy minimised and equilibrated prior to running the MD simulations. The systems with atomistic resolution were simulated in the *NPT* ensemble for a total simulation period of 100 ns (time step= 2 femtoseconds), whilst the CG simulations were run for the time specified previously using a time step equal to 15 femtoseconds for stability.

5.3 Results and Discussion

5.3.1 Properties of pure lipid bilayers

The properties of the pure bilayers examined were in good agreement with the reported experimental and computational values in the literature, providing evidence that the force field successfully reproduces the properties of the examined bilayers. The simulations can capture the dynamic configurations of lipid molecules at a level of atomic detail (Pezeshkian et al., 2018). We examined a variety of biophysical data such as area per lipid, densities of some representative groups, bilayer thickness, headgroup orientations and acyl chain order parameters. These properties can be validated with data from e.g. spin-label electron paramagnetic resonance, NMR and Forster resonance energy transfer studies (Kuprusevicius et al., 2011, Rauscher et al., 2015, Best et al., 2015). The three model bilayers of the first set of simulations, had different lipid composition but the total number of lipids was kept constant (see Table 5.3). For each of the model bilayers, we analysed the average area per lipid, the bilayer thickness and the diffusion coefficient of their components presented in Table 5.6. The area per lipid

is an approximation however, because we assumed that each of the lipid types (DOPC, cholesterol, and glycolipids) occupied the same area.

The area per lipid was calculated after we extracted the box size from every frame by multiplying the x- and y-dimensions and dividing by the total number of lipid molecules per leaflet. The average area per lipid for the pure DOPC bilayer was 0.68 nm^2 (see Figure 5.5) which was significantly reduced when we inserted the cholesterol molecules. The calculated area per lipid was in good agreement with previous studies. For example, in a study that examined properties of fully hydrated fluid phases of DOPC bilayers, the area per lipid was found experimentally to be 0.67 nm^2 and the hydrophobic thickness of the bilayer equal to 3.67 nm (head-to-head spacing) with x-ray and neutron scattering data at 30^0C (Kučerka et al., 2008). Petrache et al., investigated structural properties of charged phosphatidylserine bilayers with X-ray diffraction and NMR spectroscopy, and reported a value equal to 0.73 nm^2 for the average area of DOPC (Petrache et al., 2004). Pan et al., examined the temperature dependence of the structure, bending rigidity and bilayer interactions of DOPC bilayers with X-ray diffuse scattering and reported the area/molecule equal to 0.76 nm^2 at 45^0C and 0.72 nm^2 at 30 ^{0}C (Pan et al., 2008). We notice the good agreement between the experimental and our computational results.

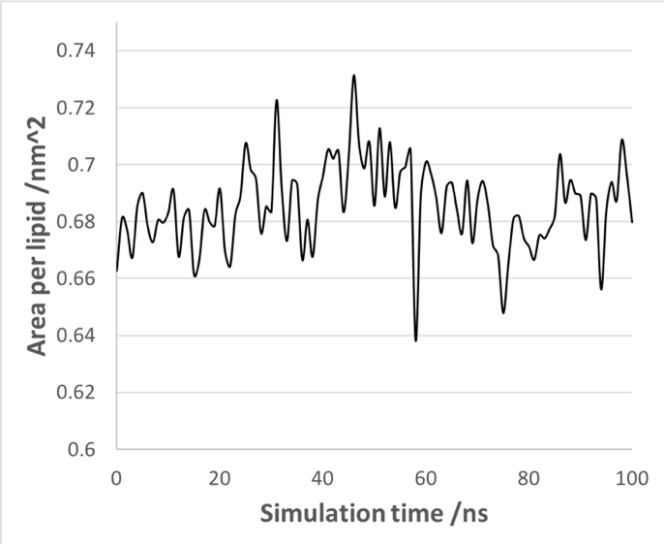


Figure 5.5: The area per lipid of the DOPC bilayer with respect to simulation time (System_1).

The area per lipid in the presence of cholesterol molecules was decreased as presented in Table 5.6. The table also shows that the bilayer thickness was increased, namely from 3.79 to 3.87 nm. Cholesterol plays a major role in regulating the fluidity and the mechanical properties of bilayers and it is known to cause a condensing effect in DOPC bilayers because the area per lipid is decreased in its presence. Cholesterol also increases the phosphate to phosphate distance in the lipid bilayer (thickening effect). This is probably because the insertion of cholesterol makes the aliphatic acyl chains more perpendicular to the bilayer normal (Hung et al., 2007). The calculation of area per lipid after the insertion of cholesterol is not straightforward because in the bilayer there are now two types of lipids. Some methods assume that both DOPC and cholesterol molecules have a cylindrical shape and the same height, but this can result in contradicting measurements between the individual area per lipid calculations and the overall result at high cholesterol mole fractions (Alwarawrah et al., 2010). Both the molecular packing and the tilt angle of cholesterol are important factors that need to be considered when calculating the properties of bilayers. In this study, we also assumed that the two lipids have roughly the same shape (cylindrical) and height.

The thickness of the pure bilayers presented in Table 5.6 was calculated as the distance between the phosphorus atoms of the lipid headgroups from the density profiles (see Figure 5.6). Regarding the pure DOPC bilayer, the average thickness was 3.79 nm, which is slightly larger than the experimental value observed by Kucerka et al., namely 3.67 nm. The value is also close to the ones reported in other computational modelling studies. For example, in a study that examined the cholesterol condensing effect in DOPC lipid bilayers with atomistic MD simulations (where different lipid compositions were examined) the thickness ranged between 3.94-4.67 nm (Alwarawrah et al., 2010) and Skjevik et al., reported the value of bilayer thickness equal to 3.79 ± 0.04 nm, measured as the distance between the phosphate peaks in the time averaged electron density profiles (Skjevik et al., 2016). Skjevik et al., reported experimental values for the thickness from various studies in the range of 3.5-3.7 nm and an area per lipid of 0.68 nm^2 for the DOPC simulations using the CHARMM C36 lipid force field parameters (Skjevik et al., 2016). Table 5.6 demonstrates that the bilayer thickness

increased when cholesterol was added in contrast to the addition of glycolipids (System_3).

Table 5.6: Average area per lipid, thickness and diffusion coefficient (lateral diffusion) of the model bilayers (System_1, System_2 and System_3).

	Area per lipid/ nm²	**Thickness /nm**	**Diffusion Coefficient (lateral diffusion plane x-y) (MSD/cm² s⁻¹)**
DOPC (System_1)	0.69 ±0.01	3.79	0.011 (± 0.0010) x 10^{-5} (P atoms)
DOPC+Cholesterol (System_2)	0.60 ±0.01	3.87	0.014 (± 0.0050) x 10^{-5} (P atoms) 0.019 (± 0.0062) x 10^{-5} (Cholesterol)
DOPC+GLPA (System_3)	0.67±0.01	3.69	0.013 (± 0.0000) x 10^{-5} (P atoms) 0.013 (± 0.0038) x 10^{-5} (BGLC) 0.015 (± 0.0063) x 10^{-5} (CER)

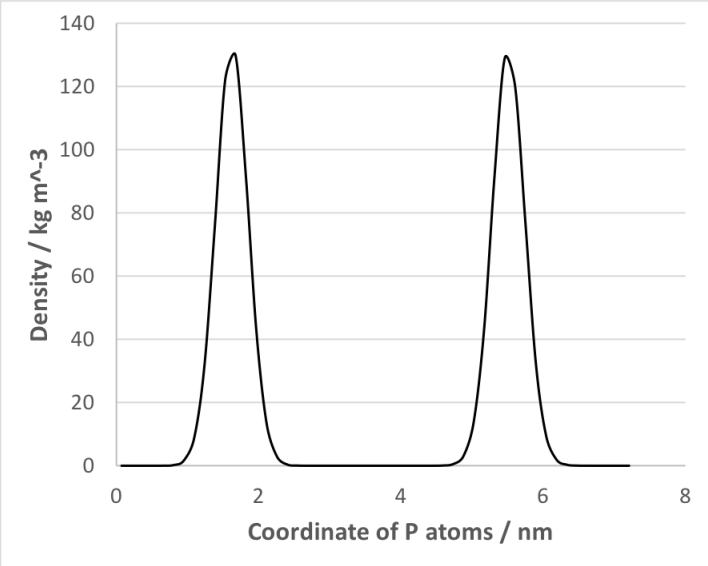


Figure 5.6: The density of the phosphorus atoms across the bilayer (z coordinate) for System_1. The density peak positions were used to calculate the thickness of the bilayer as the distance between the phosphorus atoms of each leaflet (peaks). The phosphorus atoms are considered representative of the bilayer margins.

From the MD simulation, we can also calculate lateral diffusion coefficients (within a plane) using the Einstein relation:

$$\lim_{t\to\infty}\langle\|r_i(t) - r_i(0)\|^2\rangle_{\,i\epsilon A} = 6\,D_A t \qquad (21)$$

where D_A is the self-diffusion coefficient of particle i of type A and $r_i(t)$ and $r_i(0)$ are the positions of atom i at time t and 0 respectively (Allen and Tildesley, 1989). The coefficients are presented in Table 5.6 for all model bilayers using the phosphate atom (representative of DOPC), the cholesterol molecules and the ceramide (lipid -residue) and glucose (sugar- residue) of the glycolipids. The data were fitted from 10 ns to 90 ns and the mean square displacement was calculated with the GROMACS tool, gmx msd (Abraham et al., 2015). The data suggest that cholesterol is more mobile compared to the other lipid types and that DOPC slightly increased its diffusion coefficient after

another component was added in the model bilayer (Systems_2 and_3). Heterogeneous systems tend to be less structured and hence show greater diffusion.

Lipids in the bilayer are highly dynamic, rotating around their axis, wobbling, diffusing in the same plane and showing transverse motion across the bilayer (flip-flop) but these motions happen in different time scales (Vermeer et al., 2007). Most of these movements influence the order parameters of the lipid acyl chains. The deuterium order parameters (S_{CD}) can be calculated from MD and subsequently compared to experimental data from NMR to calibrate the simulations. The S_{CD} of the two acyl chains (*sn1* and *sn2*) of the DOPC lipid molecules are presented in Figure 5.7 and Figure 5.8 for all model bilayers. For this analysis, we created an index group that contained only the carbon atoms along the lipid acyl chains and employed the gmx_order GROMACS command to compute the order parameters per atom. The deuterium order parameters S_{CD} were calculated using Equation 22:

$$S_{CD} = \frac{1}{2}(3\langle cos^2\theta\rangle - 1) \qquad (22)$$

where the angle θ is calculated between C-H and the normal to the bilayer along the z-axis. With regards to both *Sn1* and *Sn2*, the pure DOPC bilayer shows the lowest values while when cholesterol is inserted the values increased. The insertion of the glycolipids also affected the values but not significantly. The order parameter quantifies the degree of order of the aliphatic acyl chains that form the hydrophobic core region of the bilayer. When the deuterium order parameters have low values, this implies more disorder, therefore the insertion of both cholesterol and glycolipids induced greater order in both acyl chains of the DOPC lipids. Cholesterol has been found to increase the order parameters in lipid bilayers but the amount of additional ordering differs on the type of lipids (Vermeer et al., 2007). The pure DOPC values are in good agreement with previous reported ones by Venable et al., who examined the mechanical properties of lipid bilayers with MD simulations (Venable et al., 2015) and Skjevik et al., who examined the self-assembly of lipid bilayers (Skjevik et al., 2016). Note that it was not

possible to calculate the S_{CD} for the terminal atoms because they do not have neighbouring atoms to define the local molecular axis.

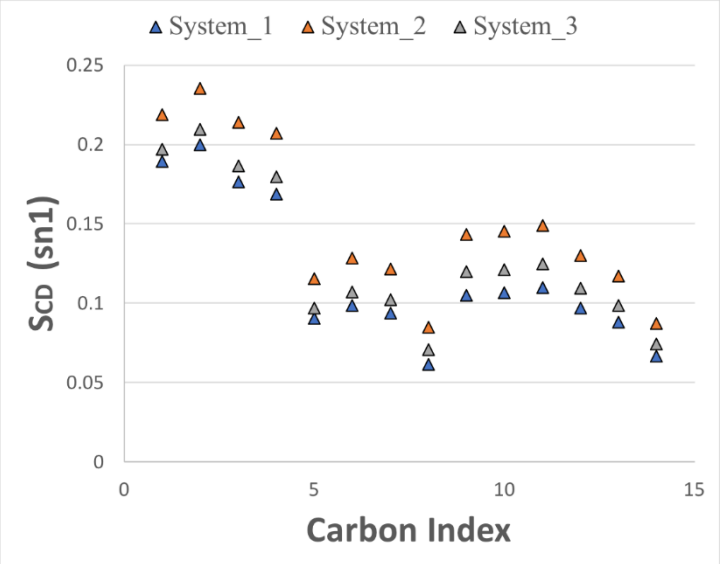


Figure 5.7: The deuterium order parameters (SCD) of the first acyl chain sn1 vs carbon atom for the three examined lipid bilayer systems (see graph label on top).

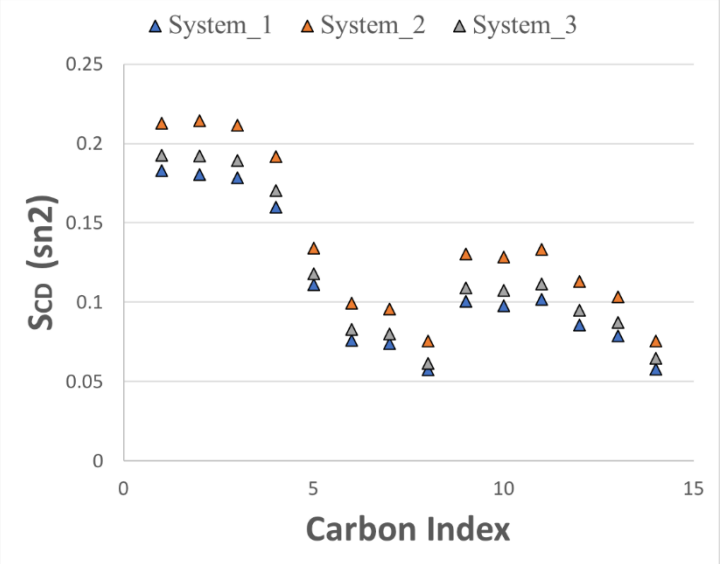
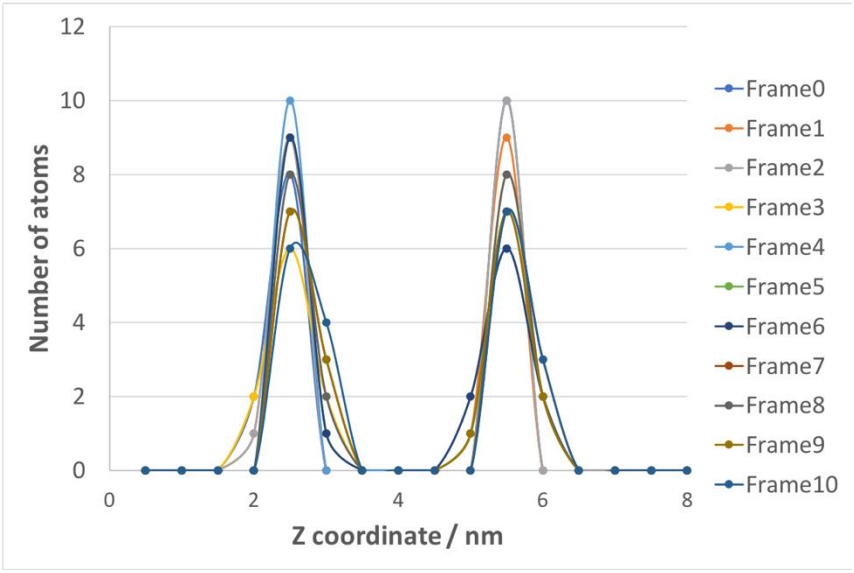


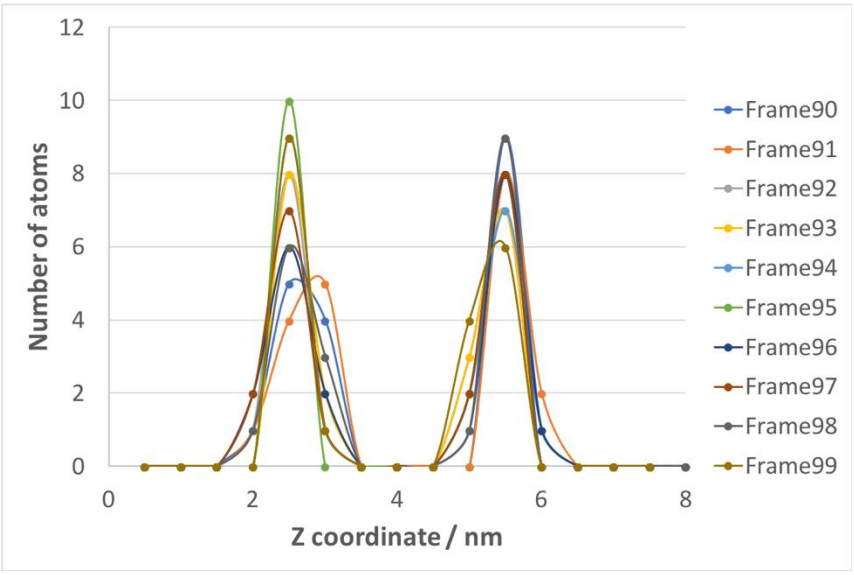
Figure 5.8: The deuterium order parameter (SCD) of the second acyl chain sn2 vs carbon atom for the three examined lipid bilayer systems (see graph label on top).

It is also interesting to examine the relative position of the added components in the model bilayers e.g. the cholesterol molecules. We developed a python script that calculates the number of lipids (or more specifically the total number of reference atoms) within a range of coordinates (with regards to the z-dimension). The script investigates where the cholesterol molecules reside and how much their position changes in time with regards to the bilayer normal (see Figure 5.9). The script 'slices' the box every 0.5 nm and counts how many reference atoms are present in this z-range (number density). The reference atom for the DOPC molecule was the phosphorus (P) atom while for the cholesterol was the carbon atom connected to the hydroxyl group (atom type C3) and the carbon atom at the hydrocarbon tail (atom type C25) which is opposite the C3 (see Figure 5.10).

The relative position of cholesterol molecules (focusing on the C3 atom) at the beginning and towards the end of the trajectory is presented in Figure 5.9. The molecules resided roughly in the same region of each bilayer leaflet (with regards to the z-dimension) within an approximately 0.1 nm variation in their z-coordinate. Thus, the cholesterol did not change significant its position from the original bilayer structure. The average values of the z-dimension of the molecules were 2.31 nm and 5.13 nm (C3atom) in the lower and upper leaflet respectively. Figure 5.10 demonstrates that the alkyl chain of the cholesterol molecules prefer to reside in the centre of the lipid bilayer (since it has a hydrophobic character) further away from the DOPC headgroups while the carbon connected to the hydroxyl group is closer to the phosphate atoms, as expected.



(a)



(b)

Figure 5.9: The location of the cholesterol molecules in the simulation box for the first frames (a) and the end frames of the trajectory (System_2). Frames 0-10 correspond to the first ns and frames 90-99 to the last ns of the trajectory (sampled every 100 ps).

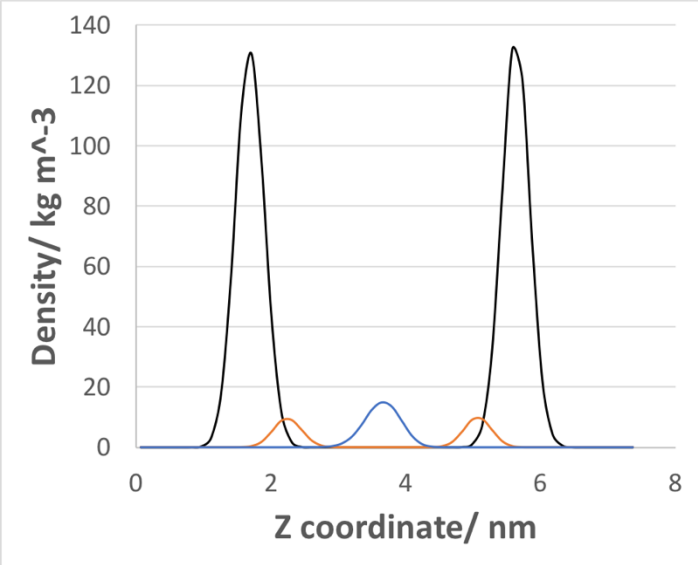


Figure 5.10: The partial density of the P atoms (DOPC lipids) with respect to their Z coordinate (black line). The graph also shows the densities of the carbon atoms of cholesterol, type C3 (carbon atom connected to the hydroxyl group; orange line) and atom type C25 (carbon atom at the hydrocarbon tail; blue line). The two carbon atoms are further away from each other in the cholesterol molecule.

To conclude, the properties calculated from the simulations were all in agreement with the experimental observations and previous simulations therefore, giving us confidence in the formed bilayer structures.

5.3.2 Structural stability of claudin-1 embedded in lipid bilayers

In this section we focus on characterising the overall structural stability of the embedded claudin-1, its tilt angle relative to the bilayer and its hydrophobic depth into the bilayer. In Chapter 3, we investigated how much claudin's ECD domain differs when it is alone and solvated in a simulation box from when the whole protein is embedded in a bilayer. There, we presented in detail the structural changes of claudin-1 when embedded in the DOPC bilayer. Specifically, the 'opening up' and vertical angles of each individual β strand were characterised, as well as, the radius of gyration and the tilt angle of the long helix (see Figure 3.8 and Figure 3.10).

For each of the lipid bilayers examined, claudin-1 did not show significant deviation from its initial structure with RMSD values being less than 0.4 nm (for the backbone atoms) of all systems. The hydrophobic depth, measured as the difference between the z coordinate of the centre of mass of the protein minus the centre of mass of the DOPC lipids, is presented in Figure 5.11. The average depth is similar in all three bilayers, namely 0.91 ± 0.07 nm for System_1 (black line), 0.95 ± 0.08 nm for System_2 (orange line), and 0.92 ± 0.06 nm for System_3 (blue line), averaged over the last 50 ns where the systems seem to have converged to similar average values. System_2 shows a slightly larger hydrophobic depth compared to the other two.

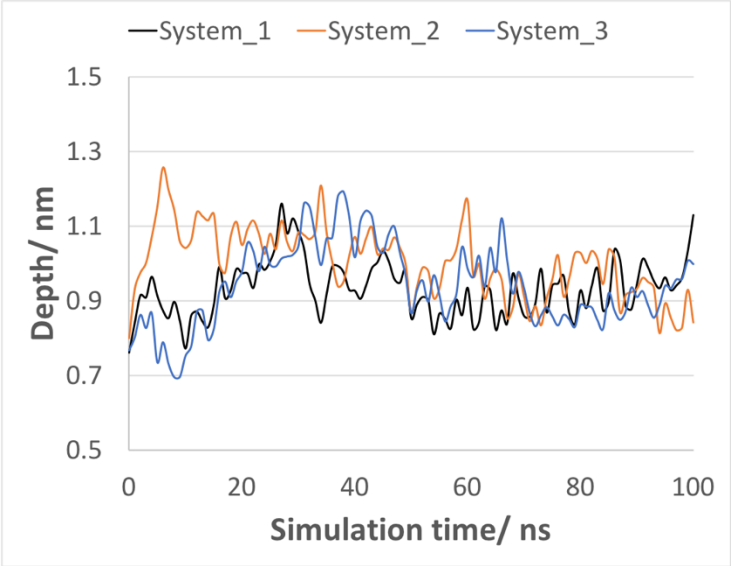


Figure 5.11: Hydrophobic depth of claudin-1 as a function of simulation time for the second set of simulations with the embedded claudin-1 (see label of the graph).

The position of claudin-1 in the bilayer can be calculated through an analysis tool called GridMAT-MD (Allen et al., 2009). GridMAT-MD measures the apparent thickness and area per lipid of a bilayer. This tool can account for more than one lipid types and the presence of an embedded protein. We employed GridMAT-MD to calculate the thickness of the bilayers whilst the claudin was embedded. The thickness is monitored through the different colours in the plot and the colour scale is presented at the right-hand side of the graph. For the DOPC pure bilayer the thickness is presented in Figure 5.12 and for Systems_2 and _3 the respective graphs are in the Appendix A2 of the

thesis. The area of the graph that displays high thickness is depictive of the claudin's position. The bilayer thickness analysis was carried out for the end frame structure of the trajectory.

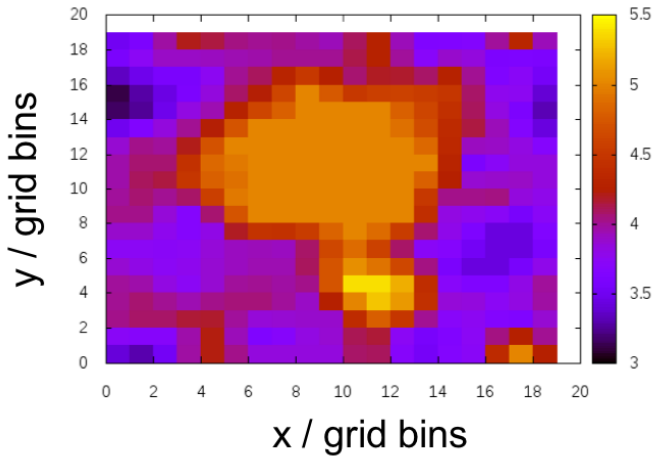


Figure 5.12:The bilayer thickness of the DOPC bilayer (System_1) with a claudin-1 embedded. The thickness was calculated with the tool GridMAT-MD and the graphics generated with Gnuplot using a 20x20 grid distribution (Williams et al., 2008).

The relative position of each protein's residue with regards to the surface of the bilayer is also important because it defines the extracellular part of the protein. The ECD forms the barrier/pore at the TJs and its charged residues exposed to the extracellular space can define the charge specificity of the pore. The densities of residues that are expected to be close to the phosphate atoms are presented in Figure 5.13 where we can also observe the density of the latter. From the examined residues 27, 28, 29 and 30 we concluded that residues 28 and 29 show similar Z coordinate to the phosphate atoms, thus, are closer to the bilayer surface (System_1). Additionally, Figure 5.13 focuses on the densities of residues 28 and 29 for the other two model bilayers and clearly shows that in all three systems these residues are consistently close to the margin of the lipid bilayer. This is in agreement with the UniProtKB database which has set residue 29 at the beginning of the ECD (Consortium, 2018). Also, since the depth of the protein presented in Figure 5.11 does not change significantly in all three bilayers, roughly the same residues should remain close to the bilayer surface.

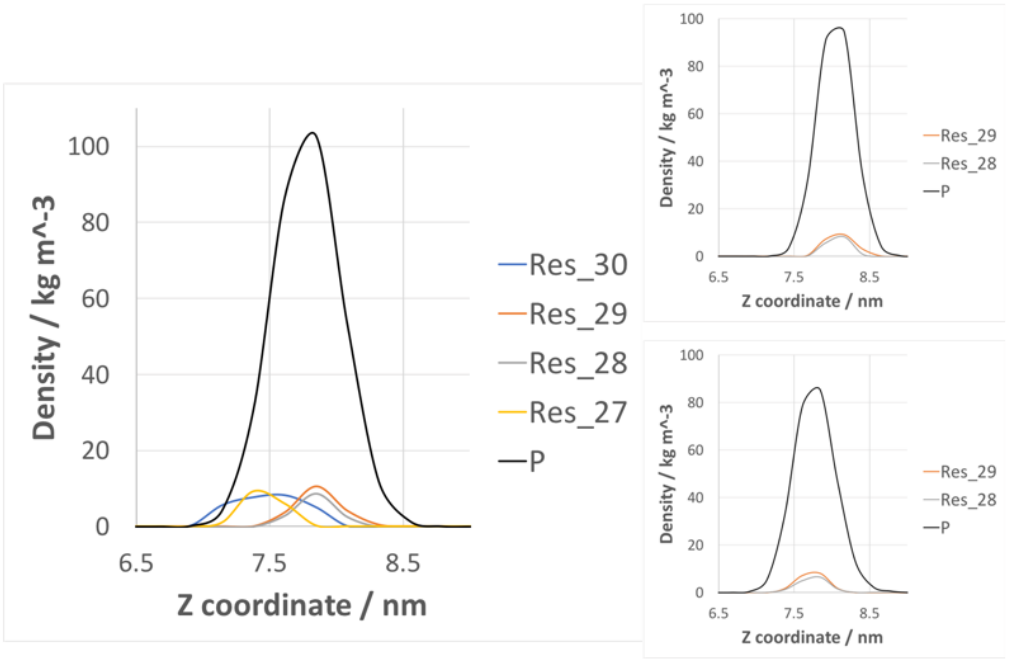


Figure 5.13: The densities of various residues that are near the phosphate atoms of the DOPC lipids (top leaflet) in all three bilayer systems. The top right graph is for System_2, the bottom right one is for System_3, while the bigger graph is for the pure DOPC bilayer (System_1).

The protein's 'tilt' angle defined as the angle between the bilayer normal and the vector connecting methionine 84 and methionine 102 is presented in Figure 5.14. The methionine residues were chosen because they reside roughly at the centre of the transmembrane helical bundle as seen from the top and bottom of the bilayer, respectively. The average values for the last 50 ns were 26.1^0, 16.1^0 and 14.7^0 degrees for System_1,_2 and_3 respectively, highlighting the largest value in the case of the pure DOPC bilayer. The values fluctuated mostly within ± 5 degrees. A large tilt angle would imply that the interaction between claudins is possibly driven by their ECD, because the other end of the protein (intracellular tails) has already formed interactions with linkers that link claudins to the actin cytoskeleton. To summarise, claudin-1 embedded in the pure DOPC bilayer showed the smallest hydrophobic depth and the largest tilt angle implying that the ECD particle could play a more prominent role in the

aggregation of claudin-1 in this bilayer. Note however, that the tilt angle was not dramatically increased to indicate that the DOPC bilayer is not representative of true biology. In Chapter 6, we report that claudin-1 has approximately 3.2 ± 0.1 nm hydrophobic thickness and an optimal tilt angle of 20 ± 3^0 degrees (see Table 6.1) calculated with the orientation of proteins in membranes (OPM) database (Lomize et al., 2011). In a recent in silico study, claudin-5 was reported to have a similar optimal hydrophobic thickness namely 3.1 ± 0.2 nm, making bilayers of varying thickness in the range of 2.7-3.8 nm appropriate to study the effect of hydrophobic mismatch (Irudayanathan et al., 2015). The hydrophobic mismatch is defined as the difference in the length of the intramembrane helical structure with the membrane thickness.

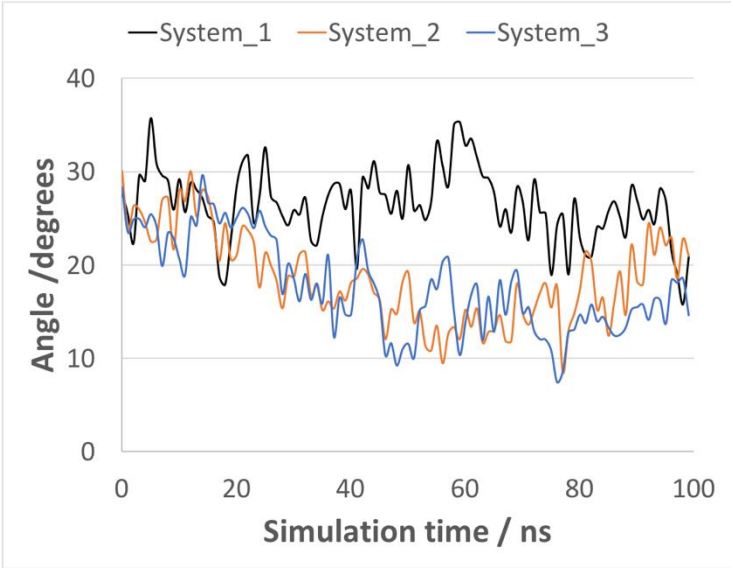


Figure 5.14: The tilt angle of the helical bundle in the three different model bilayers with respect to simulation time.

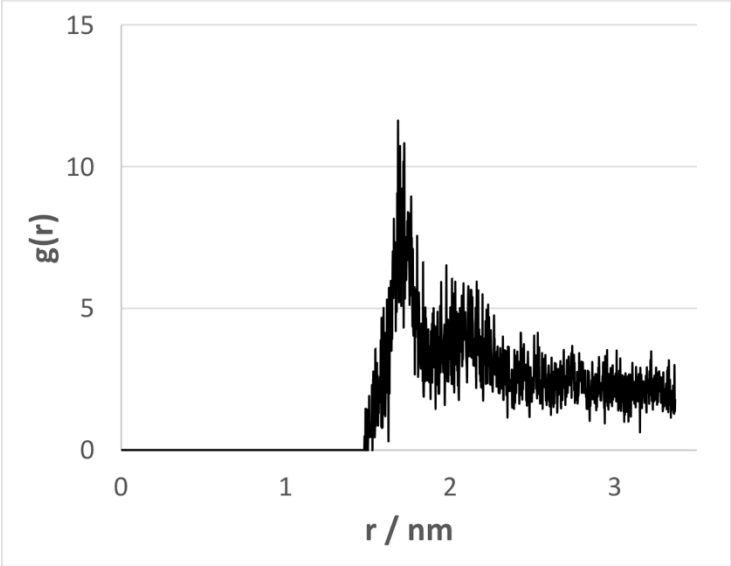


Figure 5.15: Cholesterol's radial distribution function g(r) measured from the transmembrane region of claudin-1.

Considering System_2, it is interesting to see how close the cholesterol molecules approach the helical bundle of the protein. We created two groups, namely all the residues of the transmembrane region of the protein (residues: 8-28, 82-102, 116-136, 164-184) and the C3 carbon atoms connected to the hydroxyl group of the cholesterol molecules. The radial distribution function (g(r)) of the C3 carbon atom of cholesterol was calculated from the centre of mass of the transmembrane region of the protein which was the reference group. The density of cholesterol molecules as a function of distance from the transmembrane region is presented in Figure 5.15. The closest distance on average was 1.68 nm and there were peaks at 1.69 nm, 1.71 nm and 1.72 nm. This means that the cholesterol molecules were quite close to the transmembrane region of the protein considering that the diameter of the transmembrane region is approximately 2 nm. However, when we similarly examined the radial distribution function of the third bilayer (System_3) we found that the glycolipids resided even closer to the protein. Figure 5.16 shows that there were many peaks but the closest one was at 1.53 nm a distance smaller than the one calculated for the cholesterol molecules. In the next section, where we examined the self-assembly of claudin-1 we demonstrate how the glycolipids appeared to have surrounded the proteins.

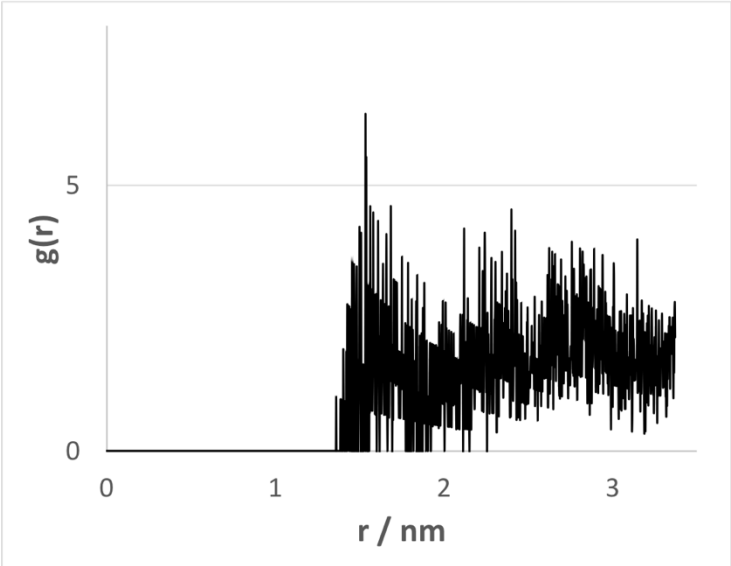


Figure 5.16: The radial distribution function (g(r)) of the glycolipid from the transmembrane region of claudin-1 (reference group).

5.3.3 Self-assembly of claudin-1 in lipid bilayers of different composition

In this section, we examined the self-assembly of claudin-1 in the three different model bilayers namely CG_1, CG_2 and CG_3 (see Table 5.4). The initial set up consisted of 8x8 grids of claudin-1 molecules embedded in the lipid bilayers with a centre of mass separation distance of approximately 7 nm (see Figure 5.17). The focus was to determine whether the proteins interacted primarily through their ECD particle or transmembrane region and whether the different lipid composition affected the strand morphology.

The claudins initially placed on a grid, quickly interacted and within 2 μs they formed less than 20 aggregates except for the CG_3 system where the aggregation was significantly slower (see Figure 5.24). Although the simulation time of the latter system was extended, it still did not reach the same number of formed aggregates. The claudins formed mostly dimeric interfaces and occasionally some trimeric ones were observed possibly facilitating the branching of the strand. The end structures of CG_1-3 are presented in Figure 5.18, Figure 5.19 and Figure 5.20, respectively. Claudins in most cases formed single row aggregates similarly to the atomistic models of the ECDs

observed in Chapters 3 and 4, while in the DOPC bilayer we also observed ring-like aggregates. The last frame of CG_2 is presented in Figure 5.19, where a double-strand formation is observed indicated by the orange rectangle. Overall, the morphology of the claudin networks was similar, showing a largely 'linear' arrangement with some cross-link, except from CG_3 which seems to need more time.

Figure 5.17: The initial set up of the grid systems (CG_1, CG_2 and CG_3). The claudin-1 particles are coloured yellow (in vdW representation) and the DOPC lipids are coloured ochre in lines representation.

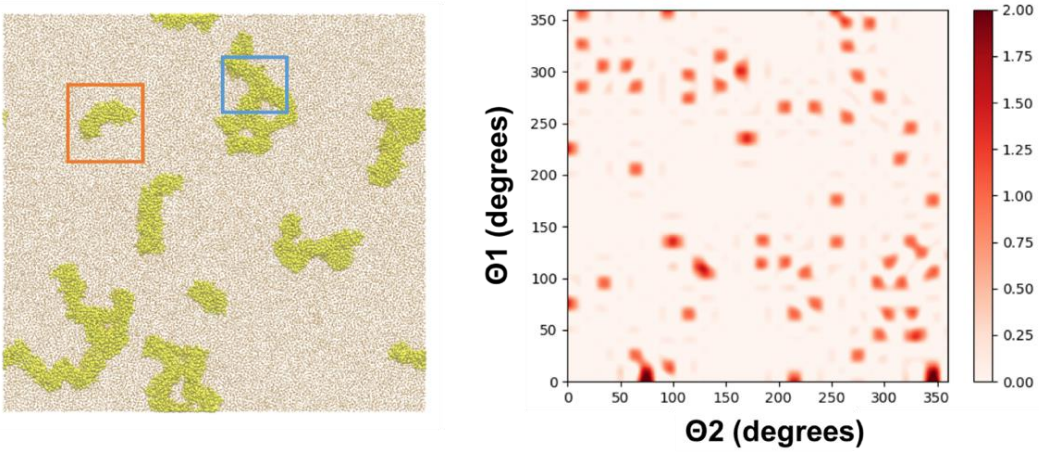


Figure 5.18: The end frame structure of CG_1 (at approximately 8.5 μs) and the θ1 and θ2 angle distribution (frequency) of the formed interfaces. Claudin-1 is in yellow, DOPC molecules coloured ochre while water molecules and ions are not displayed. The rectangles indicate the isolated aggregates that were back-mapped to atomistic models.

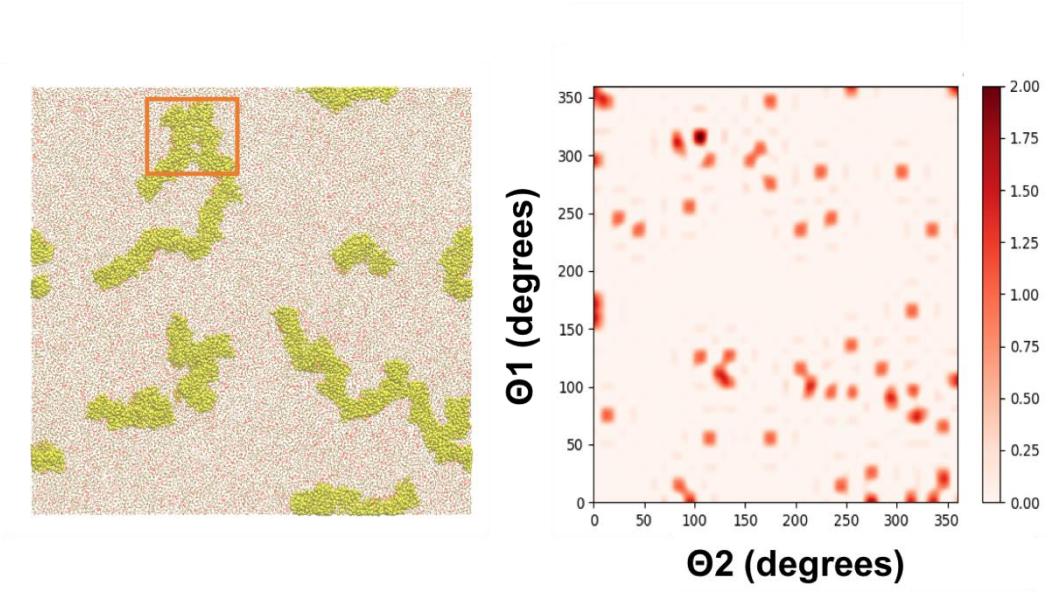


Figure 5.19: The end frame structure of CG_2 (at approximately 8.5 μs) and the respective θ1 and θ2 angle distribution (frequency) of the formed interfaces. Claudin-1 is in yellow, DOPC molecules coloured ochre, cholesterols are shown as red lines while water molecules and ions are not displayed. The rectangle indicates the tetramer which was isolated from the last frame and converted into atomistic models.

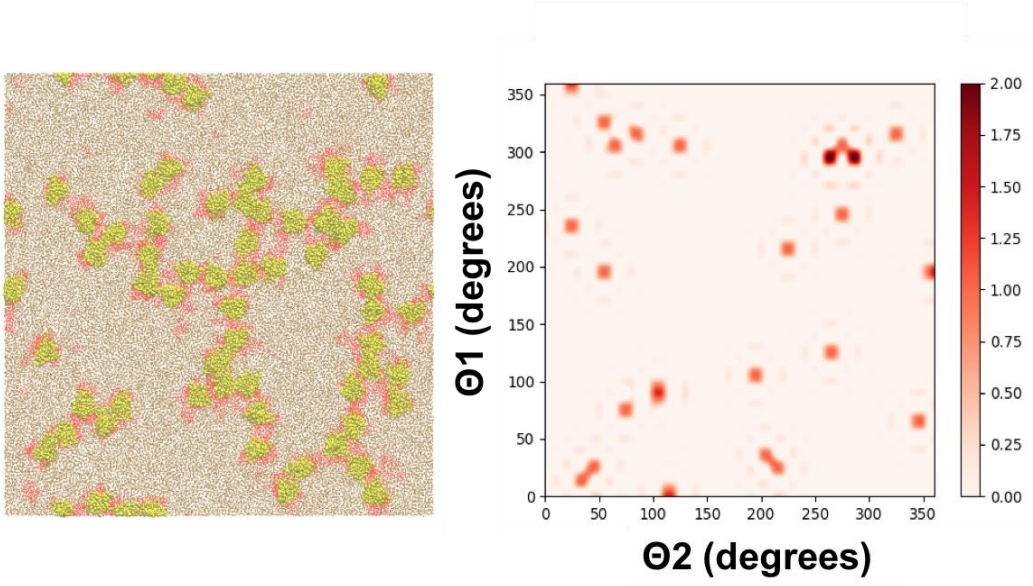


Figure 5.20: The end frame structure of CG_3 (at approximately 10 μs) and the respective θ1 and θ2 angle distribution (frequency) of claudin-1 next to it. Claudin-1 is in yellow, DOPC molecules coloured ochre, glycolipids are shown as red lines while water molecules and ions are not displayed.

The morphologies of the aggregates are more clearly presented in Figure 5.21 where the claudins are presented in Bendix representation. When visualising CG models (e.g. a protein), there are no bonds (lines) to clearly depict the connectivity between the CG beads. Therefore, Bendix was used for better representation of the proteins providing a more clear view of the formed interfaces (Dahl et al., 2012). The morphologies of the strands can be characterised as previously a cross-linked network of claudins in a linear arrangement and it is remarkable how the morphology of the strand is consistently observed in all the systems we have examined (in addition to different pure phospholipids, or more complex bilayers that we do not specifically mention here). Also, the morphology is in agreement with the ECD studies mentioned previously (Chapters 3 and 4) and other published computational studies which examine the aggregation of claudins in lipid bilayers (Irudayanathan et al., 2015, Irudayanathan et al., 2017, Irudayanathan et al., 2018).

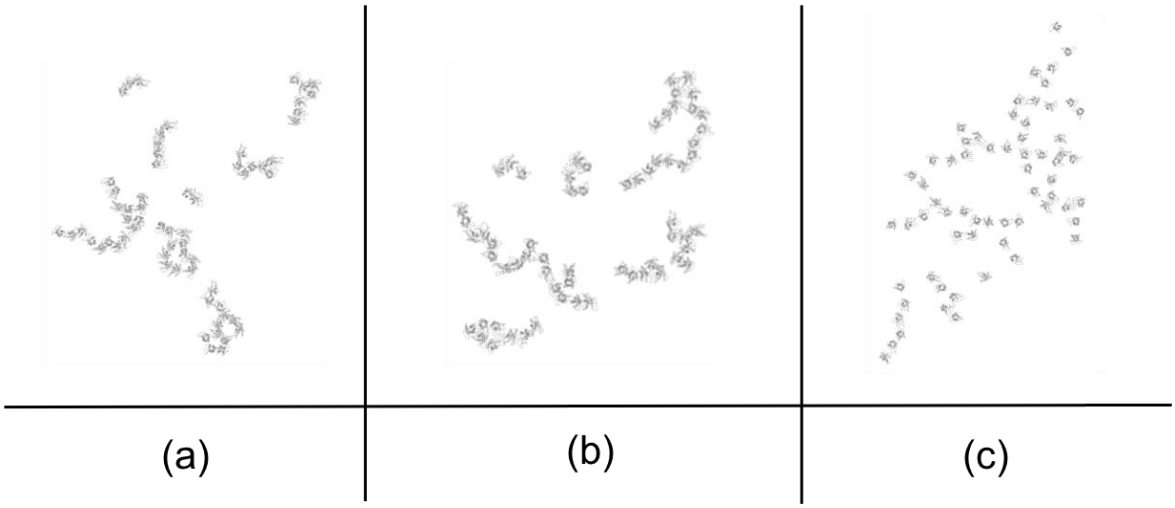


Figure 5.21: The end frame of (a) CG_1, (b) CG_2 and (c) CG_3, where the proteins are in Bendix representation and lipids, water molecules and ions are not displayed (Dahl et al., 2012).

It is evident that in CG_3 the glycolipids were very close to the protein kind of surrounding it, which can also be seen more clearly in Figure 5.23 where the proteins are not displayed and in Figure 5.22 where the radial distribution value is presented.

This probably prevented the proteins from approaching each other closely and made their diffusion in the bilayer slower. The diffusion coefficient of claudins in CG_1-3 was calculated equal to 0.007 x 10^{-5} , 0.005 x 10^{-5} and 0.003 x 10^{-5} cm^2 s^{-1} respectively, demonstrating that the diffusion of claudins in CG_3 was indeed slower. It seems that the glycolipids occur at the interfaces too, and hence must modulate the interaction at the interface. The number of aggregates formed is presented in Figure 5.24 and it is clear that CG_1 and CG_2 behaved similarly in terms of formed aggregates, but CG_3 had formed the largest number of dispersed aggregates and needs more time to obtain a single aggregate. The aggregates were determined based on a cut-off distance, namely 0.7 nm.

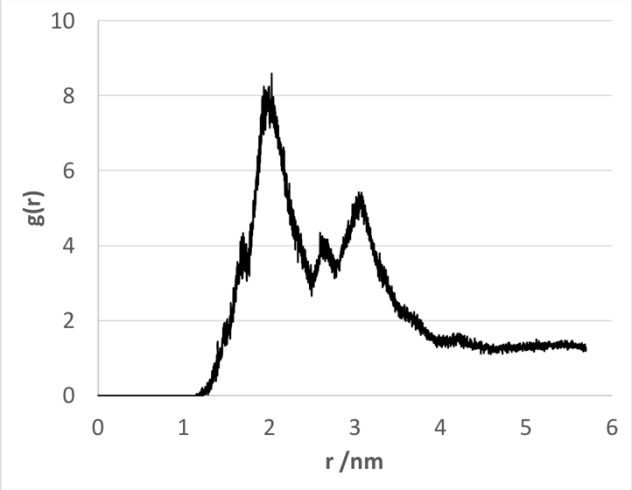


Figure 5.22: The radial distribution function (g(r)) of the glycolipids from the transmembrane region of the proteins (CG_3) calculated over the last 3 μs of the trajectory.

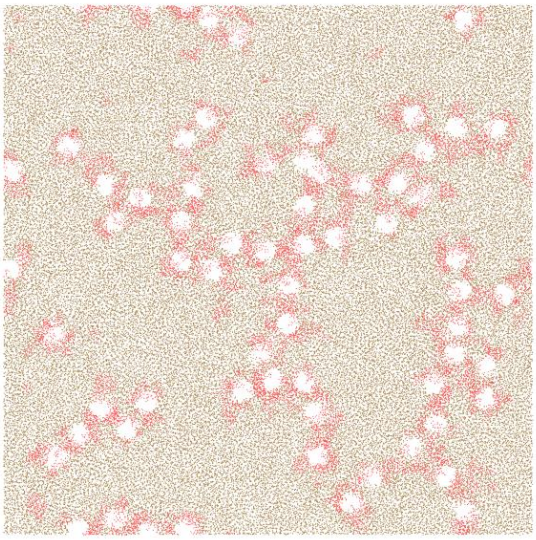


Figure 5.23: The figure displays only the lipids of the CG_3 system (final frame). It is clear that the glycolipids (red lines) are not evenly dispersed in the bilayer but rather reside close to claudin-1.

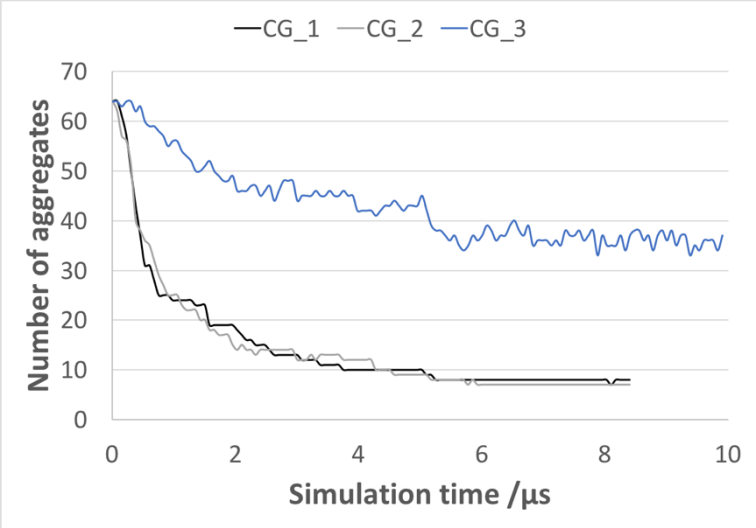


Figure 5.24: Number of aggregates formed in the CG_1 (black line), CG_2 (grey line) and CG_3 (blue line) systems with respect to simulation time.

The relative orientation angles of the claudin particles, the θ1-θ2 distributions (for definition of the angles see Chapter 3), are presented next to the end frames of the CG

systems. The distributions for CG_1 and CG_2 are quite similar and in agreement with the ones observed in Chapter 3 of the thesis because they show a plethora of formed interfaces. There is considerably variety in terms of relative orientation of the particles that characterise the interacting interfaces. CG_3 showed less formed interfaces since there were more dispersed clusters. In all cases the middle region of the graph remains an unpopular area while the corners are more favourable regions (as in the ECD study). The python script (see Appendix A1) uses a minimum distance of 3.5 nm (below which the proteins were thought to interact) that is assumed as the largest distance between the centre of mass of the proteins.

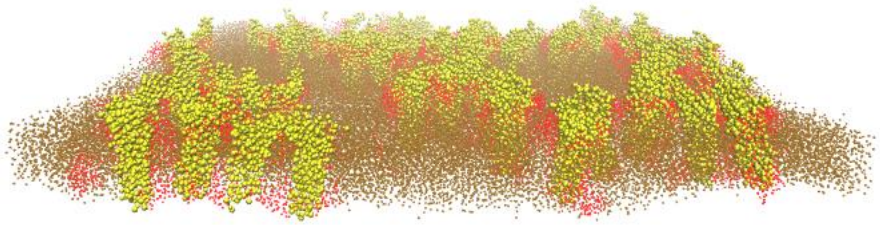


Figure 5.25: The CG_3 system from a side view where the lipid bilayer's curvature can be clearly seen. Proteins are coloured yellow, DOPC molecules are in ochre and glycolipids are coloured red. Water molecules and ions are not displayed.

In addition to the Bendix representation, we have used a method called backmapping or reverse coarse-graining (Wassenaar et al., 2014). The method converts the CG model into an atomistic one and thus, reintroduces atomistic detail. We have isolated some characteristic aggregates from CG_1 and CG_2, all of which were indicated with a rectangle at the end frame pictures. Then, we employed the martini all-atom converter from the CHARMM-GUI platform and obtained the atomistic models (see Figure 5.26, Figure 5.27 and Figure 5.28). The isolated pairs show some favourable areas mostly in the ECD particle, through which the proteins interact.

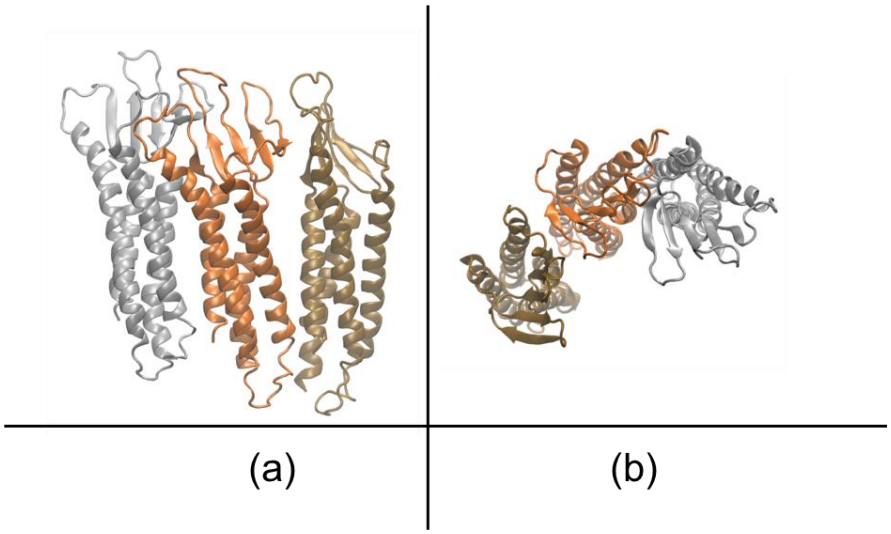


Figure 5.26: The isolated aggregate from the CG_1 system (orange rectangle). Each claudin is coloured differently in cartoon representation and we can see the proteins from a side (a) and top (b) view.

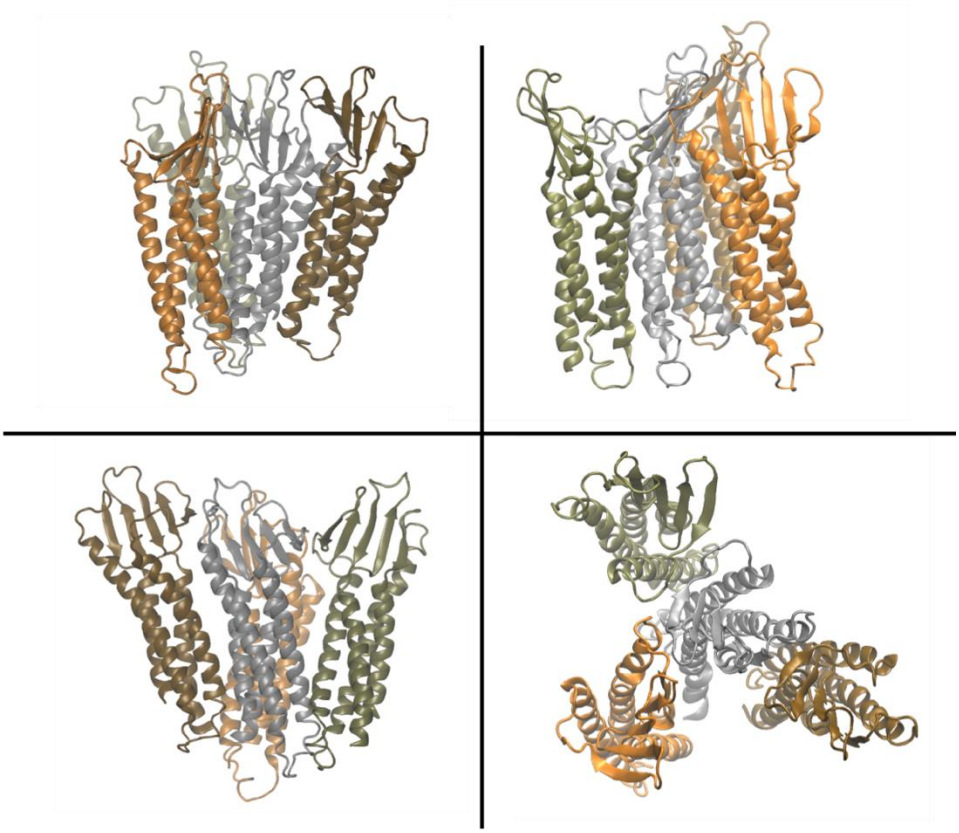


Figure 5.27: Different perspectives of an isolated aggregate from the CG_1 system (blue rectangle). Each claudin is coloured differently in cartoon representation. Images were prepared with VMD.

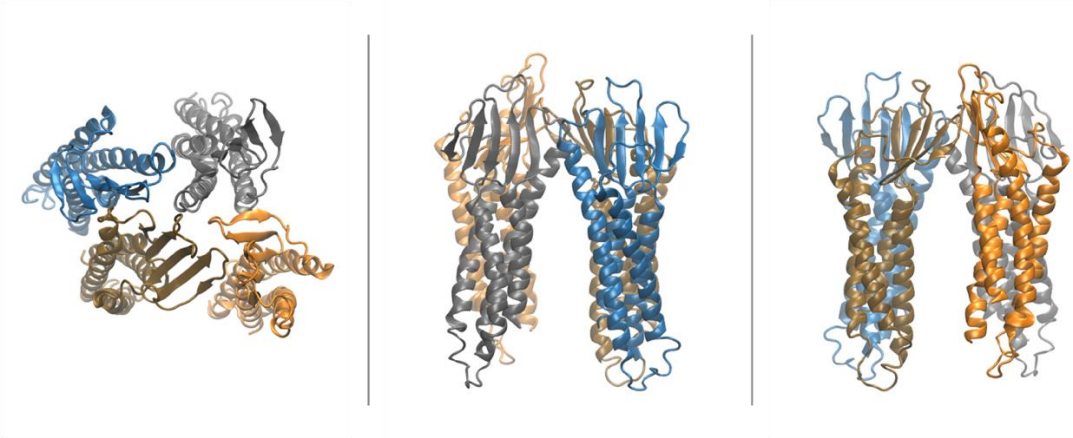


Figure 5.28: The isolated tetramer (orange rectangle) from CG_2 system from different views. Each claudin is coloured differently and in cartoon representation.

The isolated aggregates demonstrate that a variety of dimeric interfaces were formed and that the monomers were also able to interact via the back of the β-sheet domain because of the bilayers curvature (see Figure 5.25). The double row aggregate presented in Figure 5.28 is different than the one proposed by Suzuki et al. (Suzuki et al., 2015). However, the blue and grey claudins from Figure 5.28 resemble the arrangement of another dimeric interface proposed for claudin-5 by Irudayanathan et al., (Irudayanathan et al., 2017). The proposed latter dimer was isolated from an MD simulation and trans-interacting dimers were produced with docking, suggesting another putative pore model which was supported by biochemical results.

From the isolated aggregates it appears that the claudins interacted primarily through their ECD particles. The total number of contacts formed between the ECD particle and the transmembrane region with other protein atoms, are presented in Figure 5.29 and Figure 5.30 and also tabulated in Table 5.7 for all examined systems. Table 5.7 presents the average number of contacts per domain and clearly shows that the ECD particle forms on average more contacts compared to the transmembrane region (averaged after 5 μs). The ECD particle had 181 CG beads and the transmembrane region 157 in total. CG_1 and CG_2 show similar average number of formed contacts with regards to their ECD particle while CG_2 depicts less contacts formed via the transmembrane region than CG_1. CG_2 had cholesterol molecules in the lipid bilayer, which probably

reduced the number of contacts formed between claudins via the transmembrane region, thus making the ECD particle the dominant interface. It is notable that CG_3 depicts a very low number of contacts, with claudins in this system almost not forming any transmembrane contact. Generally, the ECD particle formed a larger number of contacts with other protein atoms, characterising it as the 'key' player of the aggregation as hypothesised in our previous study (Chapter 3). However, note that the systems probably need more time to form the final energetically favourable structure. Due to their high binding energy (calculated in the previous chapters), it is suggested that the systems were not able to reach the overall lowest free energy configuration, highlighting again the problem of kinetic trapping of the proteins. Moreover, claudins in CG_3 clearly need more time to aggregate.

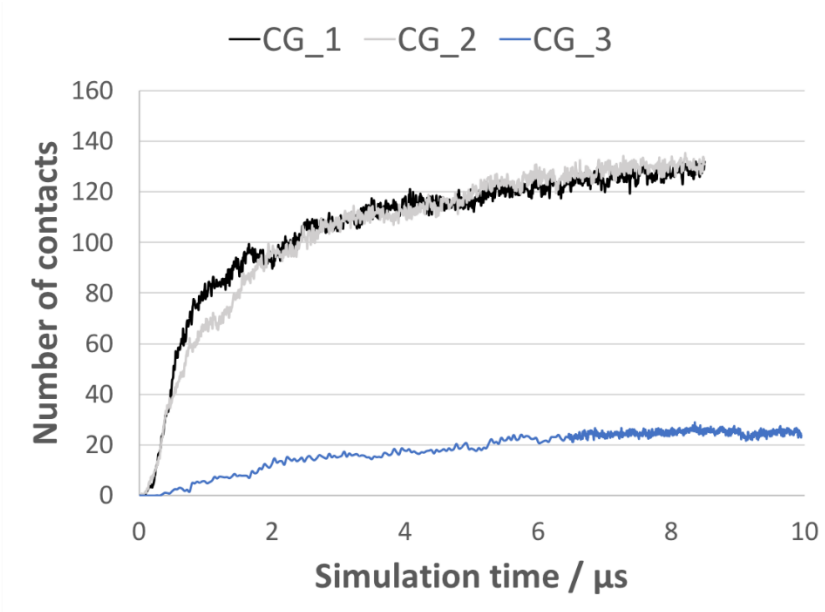


Figure 5.29: Average number of contacts between an ECD particle and other claudin particles during the trajectory of the three examined self-assembly systems (see graph label). The number of contacts is per ECD particle (averaged over a total of 64 particles).

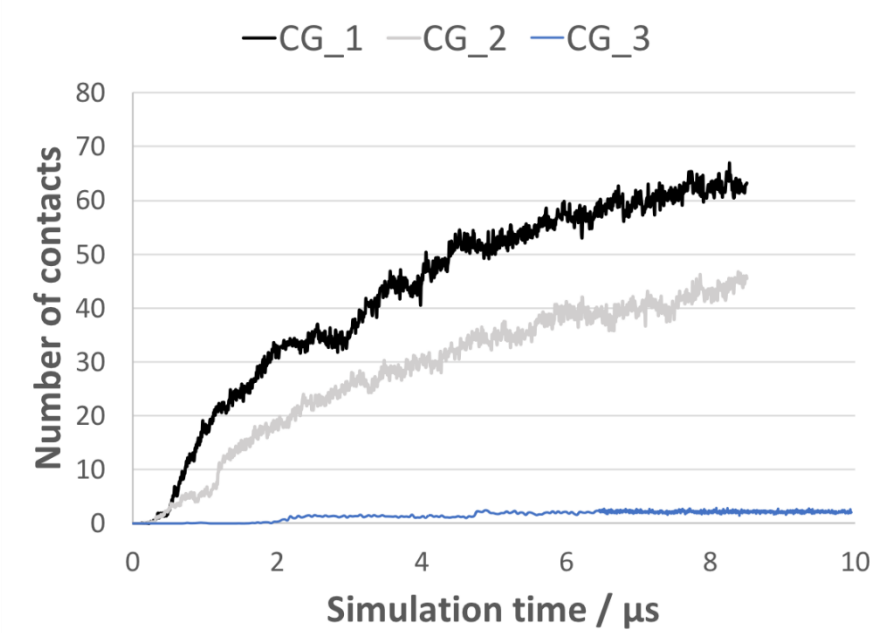


Figure 5.30: Average number of contacts between the transmembrane region and other claudin particles during the trajectory of the three examined self-assembly systems (see graph label). The number of contacts is per transmembrane domain (averaged over 64 transmembrane domains).

Table 5.7: The average number of contacts formed between the ECD particle and the transmembrane domain and other claudin particles in the CG_1, CG_2 and CG_3 systems (averaged after 5 μs).

	ECD particle	**Transmembrane domain**
CG_1	124 ± 4	59 ± 3
CG_2	127 ± 3	40 ± 3
CG_3	25 ± 1	2 ± 0

5.3.4 The effect of different protein concentration on strand morphology

Any aggregation process requires that the system is supersaturated with the biomolecule of interest in terms of concentration. The concentration determines the driving force (the chemical potential) for aggregation. In view of this we investigated the self-assembly of claudins in POPC bilayers as a function of claudin concentration. We examined the architecture of the claudin based TJ strands, and whether at low concentration the aggregates are different than the ones observed previously in the model bilayers. The expectation was that at the lowest concentration, although it takes longer for the molecules to aggregate, the final structure would have a lower free energy as the claudin particles have greater time to re-arrange and order themselves. Therefore, the architecture of the network could be altered.

Recently, Krystofiak et al., revealed with carbon replicas and EM that the TJs intramembranous fibrils have a double-stranded morphology (Krystofiak et al., 2019). The method can be used to visualise membrane-embedded protein complexes and reveal structural features. The authors suggested that this 'provides direct evidence for the double-stranded structure of the tight junctions intramembrane fibrils in their native locations' (Krystofiak et al., 2019). Thus, we also investigated whether this double-row arrangement of claudins would be observed if the systems had more time to aggregate (low concentration).

We examined three systems, namely POPC_1, POPC_2 and POPC_3 with varying protein concentration (see Table 5.5). The end frame structures of the two latter systems are presented in Figure 5.31 and we can see that for POPC_2 the claudins adopt a linear structure while in the POPC_3 system, the structure showed more branching points. Figure 5.32 demonstrates that POPC_1 needed more time to form a single aggregate, but the end structure again was very similar to the ones we have examined previously in the CG systems albeit having more 'closed'/ ring-like structures. These ring-like aggregates were also observed previously in the CG_1 system. At supersaturation (high concentration), the chemical potential is high, the driving force is high, and molecules are forced to reduce their interaction with the solvent, and aggregate. They do not have sufficient time to equilibrate and lead to non-equilibrium structures. The structure is governed by aggregation kinetics and molecules become kinetically locked. The

POPC_1 system had the lowest protein concentration, so the proteins' diffusion was slow and that prevented the aggregation of the proteins within 5 microseconds (as in POPC_2 and_3). This system was further extended to 10 microseconds.

With regards to the system with the highest concentration (POPC_3), again it was demonstrated that the ECD particle was the proteins 'key' region showing a larger number of contacts compared to the transmembrane region. Specifically, the average number of contacts between the ECD and other protein atoms was 138 ± 5 while the corresponding number for transmembrane contacts was 69 ± 5 (averaged over the last 4 μs of the trajectory).

There were many cis dimeric interfaces observed in the study as in the previous chapters. A single specific interface would result in either dimers or a linear strand. The higher order ring-like assemblies observed in Figure 5.18 and Figure 5.32 could be the 10 nm particles observed with EM, but it is difficult to assess whether these assemblies correlate with the particle morphologies observed with the EM images.

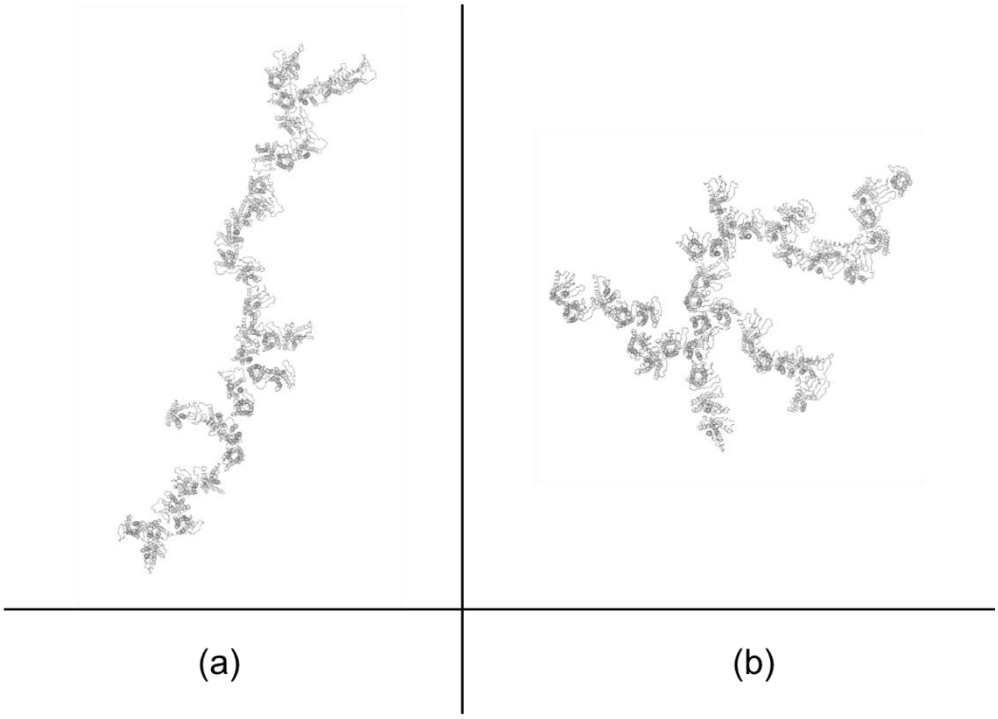


Figure 5.31: The end frames of the two more dense grid systems (a) POPC_2 and (b) POPC_3. The images show only the proteins in Bendix representation (rendered with VMD).

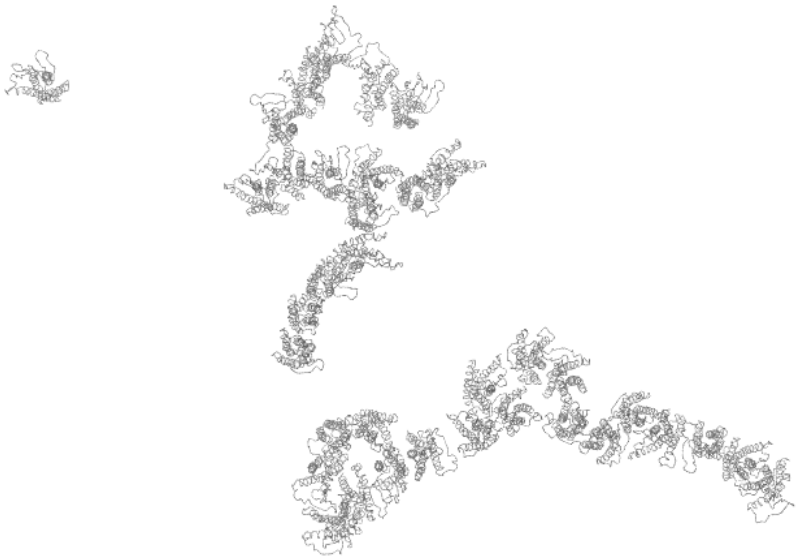


Figure 5.32: The end frame of POPC_1 (at 10 microseconds) where only claudins are shown in Bendix representation (prepared with VMD).

The study concluded that even in the very dense systems where claudins were initially placed very close to each other, the network of proteins demonstrated a single row strand architecture (see Figure 5.31) although some more dense areas were observed where trimeric interfaces formed the cross-linking points. Most importantly, even in the lowest concentration where claudins had more time to rotate and re-arrange themselves the molecular organisation of the claudin network was similar.

5.3.5 Characterisation of cis-interaction between claudins while embedded in lipid bilayers using metadynamics simulations

In the previous chapters as well as in this chapter we demonstrated that claudin-1 did self-assemble but in multiple morphologies sometimes with or without symmetry. Possibly, these aggregates were kinetically trapped, unable to rearrange and develop to equilibrium structures, because the free energy landscape (FES) of large proteins that have many degrees of freedom are complex. In Chapter 2, we discussed how MD simulations coupled with metadynamics can systematically explore more phase space and hence, lead to the most favourable binding configuration of claudins. However, in

order to assess the stability of each aggregate, we need to determine the free energies of all the possible configurations. Such a study for relatively large systems can be difficult and time-consuming before one reaches the desired result, even with metadynamics simulations. Here, we report on the efforts made to determine the most favourable dimers of claudin-1 molecules using metadynamics looking at different interfaces.

It is evident that the claudin-1 molecule is not the same (in size and shape) in all directions. For example, the area behind the β-sheet in the head of the 'Olympic torch' kind of prevents the molecules from interacting closely (through their transmembrane region), unless the β-sheet changes its orientation significantly or the tilt angle of the protein is large as observed previously in the CG models due to bilayers curvature (see Figure 5.25). In view of this, we have set up various systems composed of two claudins embedded in a POPC lipid bilayer starting from a different arrangement. The position of one protein was fixed during the simulation.

During the metadynamics simulations, the bias potential acted on the Pythagorean distance (x-y plane) between the centre of mass of the proteins, as this is definitely an important degree of freedom when we examine their aggregation. During the simulations we also monitored other collective variables (CVs) i.e. specific angles formed between residues. In order to elucidate the cis-cis interaction, the set of CVs should include all the relevant degrees of freedom. Hence, to describe the aggregation we possibly need three CVs i.e. the distance between the monomers and two angles to describe their relative position (including the difference of the centre of mass of the particles in the z-dimension).

After the simulations finished, we extracted the free energy profile of all systems (PMF calculation). The change in the free energy as a function of the distance between the centre of mass of the claudins is presented in Figure 5.33. We conclude from the graph that system X2 had the lowest free energy at 2.15 nm, namely -348.73 KJ mol^{-1}. The binding energies determined here are similar to the ones reported for claudin-5 by Irudayanathan et al., (Irudayanathan et al., 2015). The extracted lowest free energy dimer (from X2) is presented in Figure 5.34 and Figure 5.35 and it is seems that this configuration is similar to the linear arrangement of claudins. The Y1 system also displayed a low free energy at the same distance and a similar configuration as X2. The lowest free energy dimer X2 has been observed in our studies often, and it is similar in

gross terms to the Dimer D arrangement of the ECDs identified in Chapter 3 (see Figure 3.19).

The CG model from the metadynamics simulation was reversed to an atomistic one with the all-atom converter from CHARMM-GUI and we determined the key amino acids using the tool FindContacts/overlaps of UCSF Chimera (Pettersen et al., 2004). The main interaction site with regards to the ECD was the area behind the β2 and β3 strand and the ECL2 of the other monomer (Figure 5.35). The 'key' amino acids were GLN 29, TYR 47, VAL 135, GLN 146, TYR 149 and MET 152 from the ECD and ILE 12 and LEU residues from the transmembrane region. In the lipid bilayer, it is interesting that leucine side chains from one alpha helix (first transmembrane helix) interdigitate with those from the alpha helix of the third transmembrane facilitating dimerisation. Hydrophobic side chains (LEU 83, 127, 128 &131 and LEU 9, 13 & 16) extend from the helices and form a tight packing stabilising the interface. The leucine zipper has been suggested as a stabilising interaction for claudin-5 where LEU 83, 90, 124, 131 from the second and third transmembrane domain formed a symmetric hydrophobic interface in two observed dimers (Irudayanathan et al., 2015). The authors suggested that since there is a high evolutionary conservation of these residues among classic claudins it can be inferred that these leucine residues are important for dimerisation, stability and TJ assembly. Dimer X2 did not demonstrate an interaction between LEU 90 or 134 so maybe slight variations in the relative angles between the monomers can involve all the aforementioned leucine residues.

It is pertinent to note, that this is not a final result, since we should run the simulations for longer and perhaps need to bias more parameters; as mentioned previously more than one degrees of freedom are needed to fully characterise the aggregation of such large proteins.

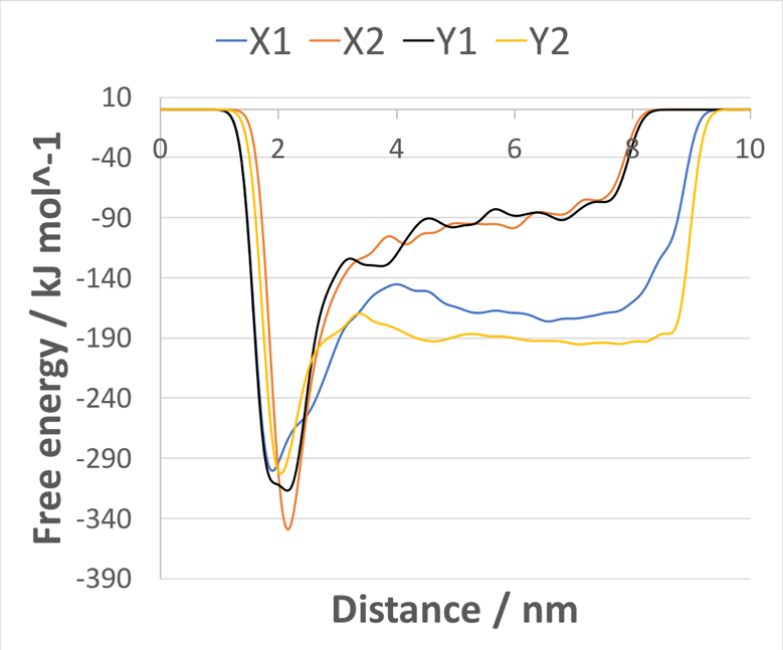


Figure 5.33: The free energy of the X1, X2, Y1 and Y2 systems with respect to the distance between the centre of mass of the claudin monomers.

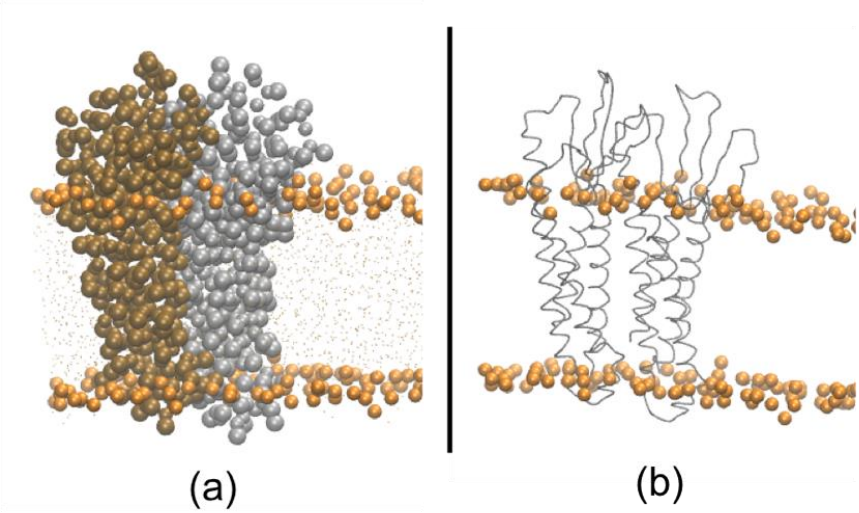


Figure 5.34: The X2 system at the time step with the lowest free energy. (a) The proteins are coloured differently and displayed as vdW spheres while the phosphate atoms of the lipids are orange spheres. In (b) the proteins are shown in a Bendix representation. Water molecules and ions are not displayed (rendered with VMD (Humphrey et al., 1996)).

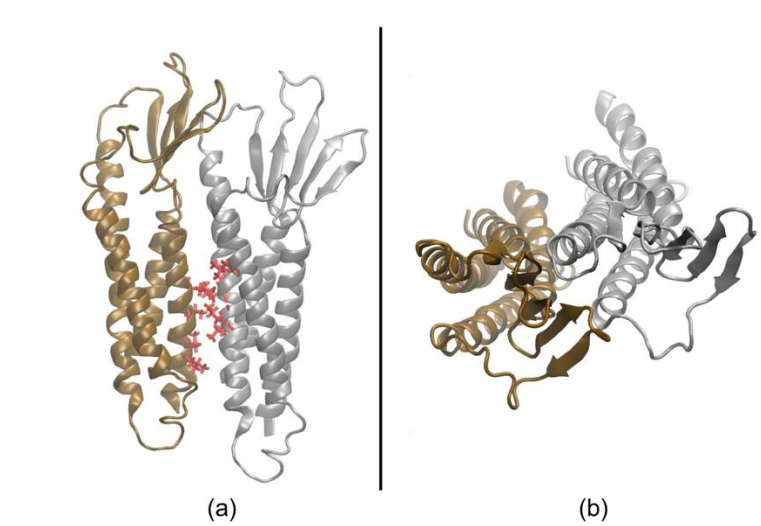


Figure 5.35: The dimer from the X2 system at the lowest free energy frame, in cartoon representation from a (a) side and (b) top view. In the side view image the key leucine residues are represented as red bonds.

In the above mentioned systems, we did not observe many binding and unbinding events between claudin-1 monomers. A large number of binding and unbinding events would give us confidence in the correctness of the reconstructed free energy landscape. Hence, the next step was to add an additional external bias potential in the form of Gaussians in addition to the applied metadynamics bias, to encourage the system to break their initial interaction. For this study, we used the same system X2 as previously, but we removed the position restraints of the centre protein. We also employed the function MATHEVAL from PLUMED, to add a constant function, centred at a specific centre of mass separation distance with a specific height (both variables, namely the centre and the height of the function were altered in our efforts). As a result, we observed more binding events (3-4 events) in some simulations but still these were not enough. The problem was that the proteins seemed to be struggling to get back together.

To conclude, our efforts to employ metadynamics in order to elucidate with confidence the aggregation of claudins, did not result in definite conclusions. We need to observe more binding and unbinding events between the two proteins to give us the reassurance that this is the most accurate description of their FES and thus, future work is suggested.

5.4 General Comments and Conclusions

In this chapter, we employed CG models to study the membrane driven interactions between claudin-1 monomers. This enabled us to access longer time scales and examine complex systems. i.e. proteins embedded in complex lipid bilayers, but the loss of atomistic detail was an important compromise.

Initially, we demonstrated the good agreement between the commonly examined properties of the model lipid bilayers (e.g. area per lipid and bilayer thickness) with experiments and computational modelling studies. The condensing and thickening effect when cholesterol is added in a phospholipid bilayer was also confirmed. The deuterium order parameters of the DOPC acyl chains demonstrated the increased order when cholesterol and glycolipids were inserted in the pure DOPC lipid bilayer. The glycolipids resided close to the protein rather than being evenly distributed in the bilayer, and cholesterol 'hid' its hydrophobic tail in the core of the lipid bilayer. The insertion of the protein in the model bilayers did not cause significant changes in its structure (low RMSD values) and the tilt angle and hydrophobic depth did not fluctuate significantly, suggesting that the lipid environments studied here, do not cause large changes in the structure of claudin-1. Claudin-1 in the pure DOPC bilayer demonstrated the largest tilt angle and smaller hydrophobic depth, suggesting that the ECD particle would play an important role in the self-assembly simulations.

The grids of claudins embedded in the model bilayers revealed that claudins quickly interact and show the same overall strand architecture as the studies in Chapters 3 and 4, where the hypothesis was that the ECD particle is the dominant -key player of claudins' self-assembly. Certainly, the invaluable atomistic detail observed in the previous chapters was lost in the CG studies. However, the linear cross-linked network morphology of the TJ strand was remarkably consistently observed in all examined systems and the combined distribution of the relative angles ($\theta 1$ and $\theta 2$) highlighting the plethora of formed interfaces, was also very similar. The analysis demonstrated that the head of the 'Olympic torch' forms more contacts with other protein atoms compared to the transmembrane region which was an important observation that provided support to our initial hypothesis. Occasionally, in the implied bilayer simulations we had observed a height difference between the terminal atoms of the ECDs, which now is also proved acceptable since the same phenomena was observed in the protein-

membrane simulations of this chapter and was created and justified by the membrane's curvature (as expected). The diffusion of the proteins in the bilayer that had glycolipids was slower compared to the other systems, preventing the proteins from forming tightly packed dimer configurations. The glycolipids were also positioned at the interface perhaps stabilising the transmembrane-transmembrane contacts.

Another factor we considered, was the effect of varying protein concentration on the claudin-based strand morphology and formed cis-dimeric interfaces. Here, the system with the lowest concentration should reveal the energetically most favourable structure as the claudins had the opportunity to shuffle and re-arrange prior to binding. Again, the architecture observed was a linear cross-linked strand network with some trimeric interfaces that justify the cross-linking organisation. Ring-like higher order assemblies were observed that could be related to the 10 nm particle observed with freeze-fracture EM.

Unfortunately, with regards to the metadynamics simulations, although our initial results were promising, future work is needed to fully clarify the FES of claudin-1. The leucine interactions observed in the lipid bilayer have been previously suggested as a stabilising interaction for classic claudins (Irudayanathan et al., 2015). Future studies should focus on biasing more degrees of freedom and observing more binding and unbinding events, to give the confidence that the reconstructed FES is correct. Metadynamics is a valuable method that offers great knowledge of complex systems demonstrated in literature as well as in the next chapter of the thesis where we examined single point mutations in claudin-1.

The importance of an efficient epidermal barrier cannot be overemphasised and the TJs that exist in our epidermis play a crucial role in the establishment of the skin barrier. The representative skin lipids examined in this chapter show that the morphology of the TJ strand can be affected by the type of lipids expressed in the membrane. For example, the glycolipids were highly localised to the claudin particles and this may play an important role in modulating the development and the final structure of the claudin strands. Thus, the focus on the biology of the skin can demonstrate how a condensed robust structure can block molecules from going through the paracellular route.

6 Effect of Single Point Mutations on Claudin-1 Structure and Assembly

6.1 Introduction

The modelling studies from the previous chapters, alongside experimental studies from the literature, have highlighted the importance of specific amino acids located mostly in the extracellular domain (ECD) particle of claudin-1. These mutants were found to either affect the aggregation of claudins or alter the paracellular barrier properties of the TJs. Therefore, in this chapter we focus on examining whether the identified single point mutations change the secondary and tertiary structure of claudin-1, providing hints about how these changes could affect the protein-protein interactions at the TJs. Protein–protein interactions are essential for cellular function and processes that occur in multicellular organisms. So, it is essential to ascertain the true structure and stability of the mutated protein since they can affect the aforementioned interactions. Furthermore, we investigate if the mutations cause steric effects (e.g. steric clashes) or affect the electrostatic interactions with residues of their microenvironment (e.g. form salt bridges or repel each other) when a general disruption of the structure is not observed. Note that the exact spatial position of each amino acid is also important, because it can either be in the centre or at the ‘entrance’ of the pore structure, and consequently, affect the charge selectivity of the TJ pore (functional site) without affecting the tertiary structure of the protein.

To better understand the effect of mutations, one needs to characterise local structural changes or the overall stability of the structure of the mutated protein and identify possible effects that the mutation triggers. For example, in an experimental study on the protein lysozyme of phage T4, it was found that when mutating six ‘cavity creating’

amino acids, the structural and thermal stability decreased (Eriksson et al., 1992). The substitutions were 'cavity creating' because the relatively bulkier leucine residues were substituted by alanines. Furthermore, the authors found a relation between the conformational change and the hydrophobic effect, i.e. 'the energy of stabilisation provided by the transfer of hydrocarbon surfaces from solvent to the interior of proteins' (Eriksson et al., 1992). The T4 phage lysozyme is considered a good model for studying the structure and stability of proteins as the crystal structures of many reported mutants have been resolved (Baase et al., 2010). It was also found that the extent of structural change varied between the mutants. In the case of four of the mutants, the structure relaxed and the changes were characterised as slight readjustments rather than reorganisation (Eriksson et al., 1992).

With regards to the TJ proteins, there are reports of mutants that alter the charge selectivity of the TJ pore (Colegio et al., 2003, Colegio et al., 2002, Krug et al., 2012b, Van Itallie et al., 2006, Piontek et al., 2017) and cases where the strand formation was prevented altogether (Zhao et al., 2018, Piontek et al., 2008, Piontek et al., 2017). Since TJs regulate the paracellular route in epithelial tissues, the mutations of TJ proteins have also been implicated with diseases. For example, mutations in claudin-16 and claudin-14 genes cause hereditary hypomagnesemia and hereditary deafness respectively (Sawada, 2013). Mutations of claudin-1 gene causes neonatal sclerosing cholangitis with ichthyosis, which is a liver disease where patients present epidermal scaling and jaundice amongst other symptoms (Hadj-Rabia et al., 2004).

Considering claudin-1, an indispensable protein for the epidermal barrier, experiments have demonstrated that mutating specific residues in the ECD affected both the TJ strand formation and its barrier properties; since the mutant induced a pore into the structure, which otherwise would not exist. The specific claudin-1 mutants examined were E48K, where the glutamic acid at position 48 is substituted by lysine, S53E, where serine at position 53 is substituted by glutamic acid, K65D, where lysine is substituted by aspartic acid at position 65, and D68S, where aspartic acid at position 68 is substituted by serine (see Figure 6.1).

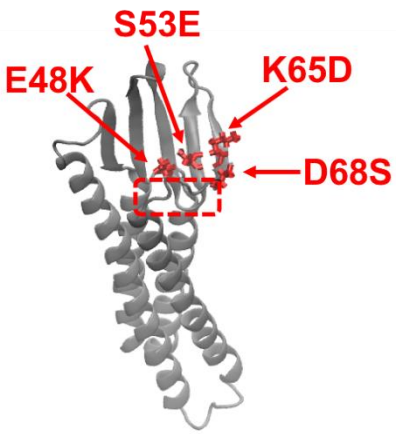


Figure 6.1: Claudin-1 in cartoon representation. The important amino acids of the first extracellular loop that have been targeted by mutation studies are labelled and coloured red (bond representation). Additionally, the unstructured area of the first loop located after the β4 strand is shown in a red dashed frame.

The above mentioned mutations were based on substituting the charged and polar residues of the barrier forming claudin-1 with those of the pore forming claudin-2, at the corresponding positions (Piontek et al., 2017). Note that some mutants had no major effect on the ultrastructure of TJs (morphology of TJ strands), while others failed to display TJ strand-like structures. The study of Piontek et al., demonstrated that substituting position 65 of claudin-1 (K65D) blocks trans-interaction and impedes the formation of TJ-like strands, unlike the other examined mutants (D68S, E48K and S53E) (Piontek et al., 2017). The charge at position 65 is conserved amongst almost all barrier forming claudins. Barrier forming claudins also have an aspartic acid at a nearby position (position 68 in claudin-1) suggesting that an intermolecular electrostatic interaction could define the barrier versus channel forming claudins. Replacing the positively charged lysine with a negatively charged aspartic acid abolished the formation of the salt bridge that served as a stable interaction between the β4-strand and the unstructured area of the first extracellular loop (see red dashed frame in Figure 6.1). Lysine at position 65 was also found important for the cis-interaction in Chapter 3 of this thesis.

In another experimental study, Veshnyakova et al., demonstrated that the claudin-2 mimicking substitution at position 53 (S53E) disturbed the barrier function of claudin-

1 since it induced a paracellular pore within Madin-Darby Canine kidney cells (Veshnyakova et al., 2012). The study demonstrated that both examined mutants, namely E48K and S53E, reduced the trans-epithelial resistance and increased the permeability of ions (e.g. Na^+ and Cl^-). Moreover, it was suggested that S53E could induce charge unselective pores. S53E and potentially D68S were believed to be involved in sealing of the paracellular space (Veshnyakova et al., 2012). The other examined mutant K65D, showed no significant change of the TJ pore permeability although it did appear to reduce the transepithelial resistance, but the effect was considered marginal (Veshnyakova et al., 2012). This result partially contradicts the finding of Piontek et al., where the position 65 was characterised as being critical for TJ strand formation.

In this chapter, we investigate using enhanced sampling techniques the effect of the above mentioned single point mutations, namely E48K, S53E, K65D and D68S, in the structure of a single claudin and discuss the potential impact on the organisation of the claudin-based strands upon aggregation. The above mutants were reported in experimental studies as having either a crucial or redundant role in the morphology of the TJ strand or in the properties of the TJ barrier. By matching the regions of the mutants that changed (compared to the corresponding ones of the wild-type claudin-1) with the experimental findings, we can strengthen the conclusions drawn from the previous chapters regarding both the critical regions for aggregation and barrier formation.

6.1.1 Enhanced sampling techniques

A limitation of MD simulations is that a system can become trapped in local free energy minima, being unable to cross high energy barriers over the timescale of standard MD (kinetic locking). Moreover, some phenomena such as the aggregation of proteins, can involve complex energy landscapes with aggregation occurring over a large range of time scales that are difficult to access with brute force MD simulations. A way to overcome these possible hinders is to use enhanced sampling techniques such as metadynamics, umbrella sampling and replica exchange (Laio and Gervasio, 2008, Barducci et al., 2011, Bernardi et al., 2015). These techniques encourage the system to

reach the most favourable configurations corresponding to the lowest free energy by overcoming the relevant energy barriers, as discussed in detail in Chapter 2.

Briefly, in the replica exchange technique, the temperature of the system is modified in order to facilitate better sampling. Many replicas of a system are simulated in parallel (at the same time) within a range of different temperatures with the possibility of exchanging the conformation between neighbouring replicas. Hence, the system can overcome the energy barriers, because of the additional kinetic energy and access regions of conformational space that would be rarely sampled at the standard temperature of interest. (Earl and Deem, 2005, Qi et al., 2018). This method will ultimately lead to more favourable conformations of the molecule and characterising the final structure and other stable ones (found at local energy minima) can shed light into specific areas of the molecule that are important for its function. For instance, during the folding of a protein a metastable state, which can be a partially folded state, can be an important functional site for the correct folding of the protein. Or in the case of a sealing claudin, the final overall stable structure could characterise the properties of the paracellular TJ barrier. Some of the recent applications of replica exchange include studies examining protein folding, aggregation and receptor-ligand binding (Sugita et al., 2012, Kokubo et al., 2013, Qi et al., 2018). The method however has high demands in terms of computational resources as many replicas are needed to span a sufficient temperature range and/or the simulations need to be run for a long time to ensure sufficient sampling (convergence). The Hamiltonian replica exchange method could also be used; in that case the different replicas are simulated at constant temperature, but the Hamiltonian of each of the replicas is varied (Meli and Colombo, 2013, Affentranger et al., 2006). In the latter method, the force field parameters are modified (usually the inter-particle non-bonded parameters) in order to overcome the limitations of temperature based replica exchange. The advantage over temperature-based replica exchange is that Hamiltonian exchange by design affects only the selected molecules of the system (rather than the whole system), thus limiting ΔU between the replicas (which means greater overlap of the energy distributions) and leading to a significant reduction in the number of replicas required.

Metadynamics on the other hand, adds a history dependent potential, in the form of Gaussian functions (Gaussians), that prevents the system from revisiting the same areas

of phase space and thus, facilitates better sampling (Barducci et al., 2011, Laio and Parrinello, 2002). This method has found applications in protein folding, phase transitions and conformational changes amongst others (Bernardi et al., 2015, Bussi et al., 2006). Recently, metadynamics was used to break down the free energy map into its entropic and enthalpic contribution for a small system and therefore, provide information of the role they play into ensuring thermodynamic stability of any given system (Gimondi et al., 2018).

These methods can help ascertain the structural differences between the wild-type protein and the mutants, and hence identify the important regions for the cis- and trans-interaction between claudins. It would also be interesting to compare the results obtained here, with X-ray crystallographic data or NMR spectroscopy data. However, there is currently lack of such structural data.

To summarise, here we use replica exchange and metadynamics simulations to examine the similarities and differences between the structures of the wild-type claudin-1 and the selected mutants. In experiments, the mutants were found to alter the characteristics of the TJs or were found not to significantly affect its ultrastructure (the overall morphology of the strand). Here, we have identified the favourable structures of the proteins using both enhanced sampling methods and compared the end structures found, reporting similarities and differences. There were differences between the structures (e.g. in the specific values of the backbone dihedral angles phi (φ) and psi (ψ) of the mutated residues), but there were also important similarities-trends between them. The systems studied consisted of a single protein embedded in a pure lipid bilayer, with a view to identifying the protein's lowest free energy structure.

6.2 Methodology

6.2.1 Homology model of the protein and the topology of the mutants

A homology model of claudin-1 was generated with I-TASSER, an online platform for protein structure and function prediction (Yang et al., 2015). The model was based, amongst other fragments, on the known crystal structures of mouse claudin-15 (PDB:4P79), mouse claudin-19 (PDB:3X29), and human claudin-4 (PDB:5B2G)

(Suzuki et al., 2014, Saitoh et al., 2015, Shinoda et al., 2016) (see Appendix A2). The unstructured intracellular tail of the model was truncated, because it does not participate in claudins' interactions whilst being embedded in the cell membrane.

The homology model was then mutated at the relevant position with the mutagenesis wizard in the PYMOL molecular graphics system (DeLano, 2002) and the most frequent rotamer was chosen. Particular care was taken to ensure that when the new atoms are inserted, they do not cause any steric clashes or overlaps with the surrounding atoms. The claudin-1 and mutants (E48K, S53E, K65D and D68S) were then embedded in an 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer and immersed in water and ions at a physiological concentration of 0.15M NaCl (see Figure 6.2). This resulted in approximately 53000 atoms in the simulation box for each system. The lipid-bilayer patch consisted of approximately 134 lipids (upper leaflet ≅ 65 and lower leaflet ≅ 69 lipids) while water molecules ranged between 10689-10873. The orientation of each protein (the tilt angle) prior to its positioning in the bilayer, was calculated from the orientation of proteins in membranes database OPM (Lomize et al., 2011), which provides the orientation of proteins with respect to the lipid bilayer. The initial hydrophobic thickness of each protein ranged from 3.14 to 3.18 nm whilst the tilt angle varied from 19 to 24 degrees (see Table 6.1). Hydrophobic thickness is the width of the hydrophobic transmembrane segment of an integral membrane protein and it is an important parameter since it defines the position of the protein in a lipid bilayer.

Table 6.1: The initial tilt angle and hydrophobic thickness of human claudin-1 and mutants as calculated with the orientation of proteins in membranes database (Lomize et al., 2011).

Protein	**Depth/ hydrophobic thickness (nm)**	**Tilt angle / degrees**
Claudin-1	3.18 ± 0.1	20 ± 3
E48K	3.18 ± 0.1	20 ± 3
S53E	3.14 ± 0.2	22 ± 3
K65D	3.18 ± 0.1	19 ± 4
D68S	3.18 ± 0.1	24 ± 3

6.2.2 Technical details of the MD simulations

The all atom protein/membrane systems were built with the web platform CHARMM-GUI (Wu et al., 2014) used for embedding proteins into pure or complex lipid bilayers. The force field utilised was CHARMM36 (Lee et al., 2014) with explicit TIP3P water model and all simulations were carried out using the GROMACS platform (Abraham et al., 2015). For the metadynamics simulations, we employed GROMACS patched with PLUMED, an open source library (package version 2) (Bonomi et al., 2009). The bonds involving hydrogen atoms were constrained using the LINCS algorithm. The particle mesh Ewald method was used to calculate the long-range electrostatic interactions with a real-space cut-off at 1.2 nm, cubic interpolation, Fourier spacing equal to 0.16, and a precision (parameter -rtoler in GROMACS) of 10^{-5}. The same cut-off distance was used for the van der Waal's interactions with a switch function at 1.0 nm. All systems were energy minimised using the steepest descent algorithm and subsequently equilibrated at constant temperature and pressure (*NPT* ensemble) for 200 ps. During the preparation, we used the Berendsen thermostat and barostat to allow smooth scaling of the box. After the preparation stage, we employed the Parinello-

Rahman barostat and Nose-Hoover thermostat for data collection (the time constants for temperature and pressure coupling were tau_t=1 ps and tau_p= 5 ps, respectively).

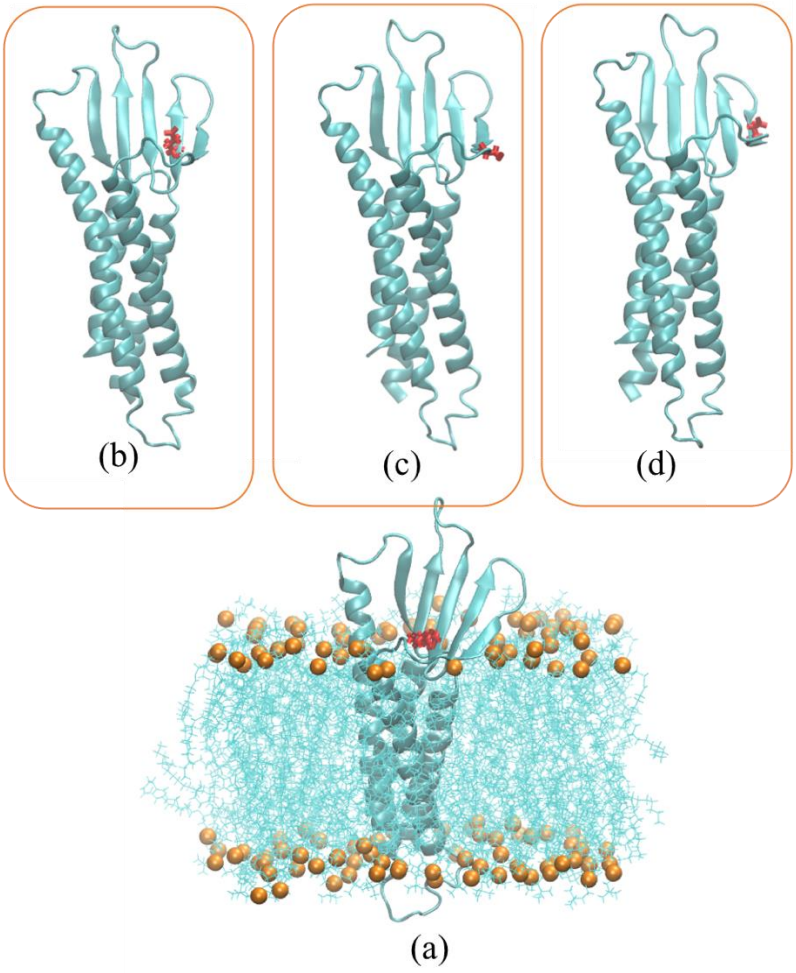


Figure 6.2: (a) The E48K mutant in cartoon representation embedded in a POPC lipid bilayer. The mutated residue is coloured red and shown in bond representation while the lipids are shown as opaque lines and their head groups as orange vdW spheres. Water and ions are not shown for clarity. The (b) S53E (c) K65D and (d) D68S mutant in cartoon representation with the corresponding mutated residues coloured red. Images were rendered with VMD (Humphrey et al., 1996).

6.2.3 Replica exchange simulations

There were sixteen replicas in total, spanning a range of temperatures from 310.00 to 335.67 K (specifically, 310.00 K, 311.66 K, 313.33 K, 315.00 K, 316.68 K, 318.37 K, 320.07 K, 321.78 K, 323.49 K, 325.20 K, 326.93 K, 328.66 K, 330.40 K, 332.15 K, 333.91 K and 335.67 K) at the constant pressure of 1 bar. The systems were simulated for a period of 100 ns using a 2 femtosecond timestep. The decision of how many replicas and what range of temperature they should span is critical and not trivial because sampling of all the relevant conformations is important. The higher temperatures should be such that the system is not trapped in any local energy minima and the number of replicas such that the exchange probability ensures that swapping is sufficient. We have considered all the above while bearing in mind that the system's size is large, the energy landscape of proteins can be complex and rough, that the lowest temperature should be kept at the desired temperature of 310 K for analysis and we have also considered the computational cost. The target is to achieve the best possible sampling with minimum computational cost.

The method is considered successful when the average exchange probability between the replicas is close to the empirical value of 20% (Earl and Deem, 2005). Kone at al., considered an acceptance probability of 23% as optimal (Kone and Kofke, 2005). The most favourable structure from the replica exchange simulation is considered to be the structure after convergence from the replica at the target temperature (310K) and the latter part of the trajectory can be used to calculate the desired properties (e.g. thermodynamic and structural data).

6.2.4 Metadynamics simulations

Metadynamics can act on various degrees of freedom or collective variables (CVs) that best describe the system. The aim is to extract the free energy landscape with regards to the specific CVs examined, as discussed in detail in Chapter 2. We aim to get the lowest free energy structure of the mutants, thus, we employ metadynamics on two collective variables, i.e. the dihedral backbone angles phi (φ) and psi (ψ) of the mutated residue. The four atoms making up φ are the carbonyl carbon from the previous residue, the connecting α-carbon, an amide nitrogen and the carbonyl carbon of the specific

residue, while the atoms that make up the ψ angle are the amide nitrogen, a carbonyl carbon, an α-carbon and a second nitrogen, to describe the rotation about the N-C_A peptide bond and the rotation about the C_A-C peptide bond respectively (see Figure 6.3). By biasing the system to sample the whole range of φ and ψ angles, we can identify the energetically most favourable conformation of the protein examined. However, we acknowledge that there are numerous more combinations of CVs that could be examined (e.g. considering other angles of the mutated residue as the ones of the side chain (chi)). But we believe that the two angles φ and ψ are structurally the significant ones for the examined systems; the specific angles are also plotted in the Ramachandran plot which is important for studying protein conformation and shows their permitted combinations.

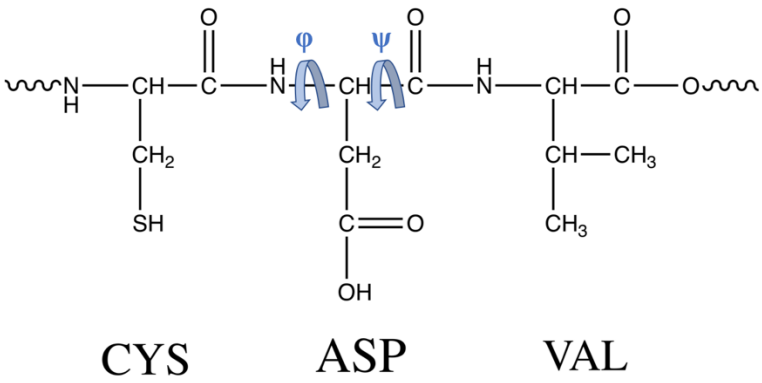


Figure 6.3: An illustration of the peptide dihedral angles φ and ψ in a polypeptide chain. The sequence chosen cysteine 64 -aspartic acid 65 – valine 66, is part of the amino acid sequence of the mutant K65D (image generated with ChemDraw (Li et al., 2004)).

The studied systems were the same as those investigated with replica exchange (one protein embedded in a POPC bilayer). The systems were simulated for 200-300 ns depending on when metadynamics had converged for each examined system. In order to assess the convergence, we monitored the CVs as a function of simulation time and the Gaussian height. It is important for the CVs to visit all the possible values multiple times (i.e. to have many re-crossings) and the height of the bias potential to progressively decrease (tends to zero). Both factors imply that the system has sampled enough conformational substates and it has converged. However, the best criterion is to

show that the estimate of the free energy with respect to simulation time converges towards the end of the simulation (apart from a constant offset). That is why we also present the reconstructed profiles of the mutants in a single graph that shows how similar the profiles are towards the end of the simulation.

In order to employ metadynamics we used the plumed.dat file, specifying well-tempered metadynamics on the two CVs, the dihedral angles φ and ψ (see Appendix A1). The well-tempered recipe means that the height of the Gaussians can change - adapt during the production run. A Gaussian was deposited every 500 steps, with an initial height of 1.2 KJ mol^{-1} and width (sigma value, σ) equal to 0.35 degrees. The bias factor γ is defined as the ratio between the temperature of the CVs ($\Delta T + T$) and the temperature of the system (T):

$$\gamma = \frac{\Delta T + T}{T} \tag{23}$$

and was set to 10. This is important to ensure that all the relevant free-energy barriers are crossed within the time-scale of the simulations. Essentially, the bias factor defines how quickly the Gaussian height decreases: a smaller bias factor results in a faster decrease of the Gaussian height. For example, if ΔT is set to 0 then the run will be unbiased. On the other hand, if ΔT is infinite, there will be no scaling of the height of the Gaussian hills (the bias potential function), therefore, the employed metadynamics run is a non- well-tempered. The temperature and pressure of the system was kept constant at 310K and 1bar respectively, with a Nose-Hoover thermostat and a Parrinello-Rahman barostat (the time constants for temperature and pressure coupling were tau_t=1 ps and tau_p= 5 ps). For analysis, all the values of the CVs and the metadynamics bias potential were stored in a COLVAR file and the Gaussians were written in a HILLS file as a grid, with a grid range set to [-π, π].

6.3 Results and Discussion

Prior to analysis of the data it is crucial to ensure that the methods have converged. Hence, with regards to the replica exchange simulations, we report on the exchange probabilities between the replicas, and assess the overlap between the generated potential energy histograms. While for the metadynamics simulations, we show how the height of the bias potential (Gaussians height) constantly decreased. Also, by monitoring the values of the CVs with respect to simulation time, we demonstrate how the system has explored all the possible values many times (sufficient re-crossings). Lastly, we show in a single graph that the reconstructed free energy profiles are similar towards the end of the simulation. All these factors demonstrate that the methods were successful and the simulations have converged.

For analysis, we focused on the similarities and differences between the identified low free energy structures of claudin-1 and its mutants with both methods. There was a focus on the structure of the first extracellular loop (ECL1) because the mutated residues are all located on this first loop and mostly on its second half (see Figure 6.1). If replica exchange is run for sufficient time, the end of the trajectory will provide the most favourable conformations, while metadynamics simulations generates the free energy surface (FES) of the system with regards to the CVs examined. Then the lowest free energy structure can be extracted from the trajectory based on the specified timestamp.

6.3.1 Potential energy distributions and exchange probability in the replica exchange simulations

The simulations of the wild-type claudin-1 and mutants yielded a satisfying exchange probability between the replicas. The probability is presented in Table 6.2 and it was in the range of 0.24-0.32, well within the acceptable limits. We have also extracted the potential energy of each replica and generated their histograms using 20 bins. These distributions of the potential energy of claudin-1 are presented in Figure 6.4 and of the mutants in Figure 6.5, where each replica is coloured differently and is slightly transparent to demonstrate the overlap between the potential energy distributions. Note that sufficient overlap is important for an exchange to happen. It is clear that in all cases

the distributions show substantial overlap, which also means that a slightly bigger range of temperatures with the same number of replicas could be achieved.

Table 6.2: The average exchange probabilities between the replicas of claudin-1 and mutants (E48K, S53E, K65D and D68S).

Protein	**Average exchange probability between replicas**														
	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15
Claudin-1	0.26	0.29	0.3	0.31	0.31	0.31	0.31	0.31	0.32	0.31	0.31	0.31	0.3	0.3	0.27
E48K	0.24	0.26	0.27	0.27	0.27	0.26	0.26	0.27	0.26	0.27	0.27	0.27	0.27	0.27	0.25
S53E	0.25	0.26	0.26	0.27	0.27	0.27	0.26	0.27	0.27	0.27	0.26	0.27	0.27	0.27	0.26
K65D	0.25	0.26	0.27	0.27	0.26	0.26	0.26	0.27	0.27	0.26	0.27	0.27	0.27	0.26	0.25
D68S	0.26	0.29	0.3	0.3	0.31	0.3	0.3	0.3	0.31	0.3	0.31	0.31	0.3	0.29	0.26

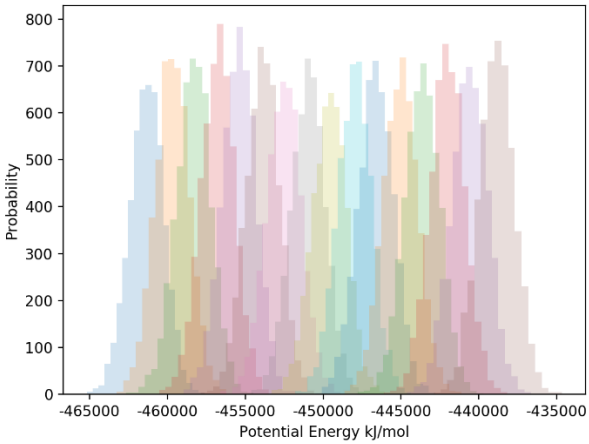


Figure 6.4: Potential energy distributions of the wild-type claudin-1 replicas. The distributions are coloured differently and are transparent to demonstrate the overlap between them.

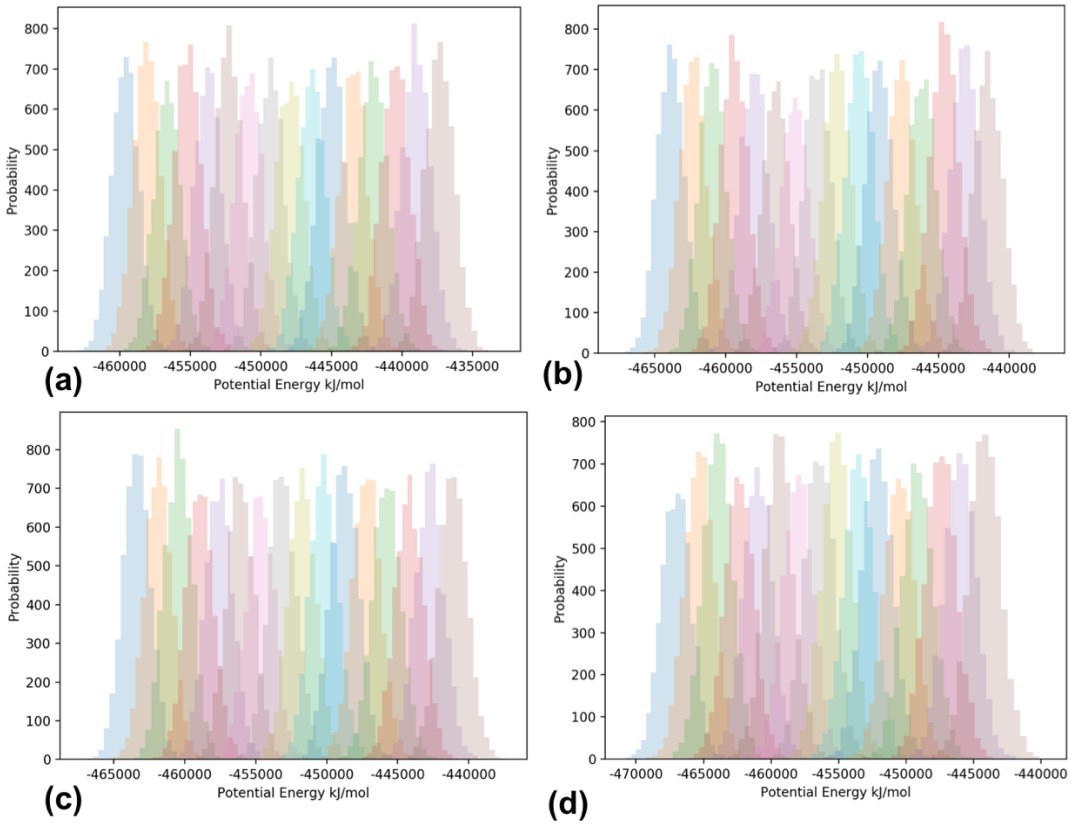


Figure 6.5: Potential energy distributions of the replicas of (a) E48K, (b) S53E, (c) K65D and (d) D68S mutants. The number of bins used to generate the distributions was 20 and each replica is shown in a different colour.

6.3.2 Structural characterisation of claudin-1 and mutants

6.3.2.1 E48K and D68S change their initial structure, the mutation affecting the first and second loop respectively

The structural differences between the wild-type claudin-1 and mutants are important because they provide evidence of the critical areas responsible for the aggregation of claudin-1 and the functional sites of the TJ barrier structure. For instance, one can match the specific areas that changed upon mutation with experimental observations e.g. the 'opening up' of the barrier or the electrostatic potential within the pore. Both the initial models of the wild-type protein and the mutants are superimposed on each other in Figure 6.6 with UCFS Chimera (Pettersen et al., 2004). The figure shows that the models of the mutants were very similar (in terms of secondary structure) with the model of the wild-type claudin-1. Therefore, we can compare the results between the wild-type claudin with the ones obtained for the mutants.

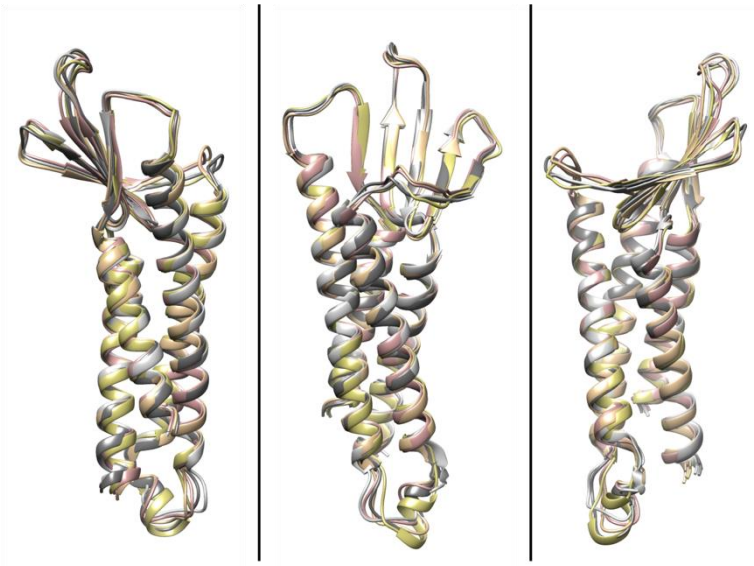


Figure 6.6: Initial models of the four mutant structures superimposed on the initial model of wild-type claudin-1. The proteins are in cartoon representation and coloured differently.

For analysis we omitted the first 10 ns of the trajectory of the lowest temperature replicas (both wild-type and mutants). The average RMSD values for the backbone

atoms are presented in Table 6.3. The analysis is for the whole claudin and also broken down to the first and second loop (ECL1: residues 29-81, ECL2: residues 137-163). We notice that all claudins exhibited similar values. With regards to the backbone atoms, all mutants showed a larger RMSD value compared to wild-type claudin-1. E48K and S53E showed the highest RMSD values, implying a larger deviation from their starting structure. Focusing the analysis on the first loop, all mutants except S53E showed a larger RMSD value (compared to the corresponding value of claudin-1) while considering ECL2 only D68S displayed a larger value compared to the wild-type. This means that the ECL2 was not affected so much by the mutations, because as previously mentioned the substitutions were in the first loop. The substitution of the charged residues at positions 48 and 65 which altered the original charge of the side chains of the residues at these positions has affected the first loop significantly. The charge at position 65 is important for all barrier forming claudins and K65D has now an oppositely charged residue in this position. So, instead of a favourable -stabilising electrostatic interaction between LYS65 and ASP68 we now observe a repulsion between the residues at these positions. Additionally, S53E showed a high overall deviation from its initial structure (and thus also from the wild-type) since the substitution involved the insertion of a residue with charged side chain in the position of a polar one, next to an oppositely charged residue, namely GLU48.

6.3.2.2 Replica exchange simulations suggest that all mutants slightly 'open up' their first extracellular loop

The radius of gyration (R_g) is an indication of how compact the structure remains characterising the tertiary structure of the protein. It is defined as the root mean square distance from each atom of the protein to the centroid. The R_g of the wild-type claudin-1 and mutants is presented in Table 6.4. Here, when we focused on ECL1 we noticed that all mutants have increased their radius of gyration, meaning they increased their extent or showed a looser 'packing' with respect to their initial conformation. Thus, the structure of the ECD seems to have extended its volume slightly, which is likely to affect the pore structure formed between claudins, since the extracellular loops form the pore. However, the underlying differences are within the sigma values and therefore are not considered significant.

Table 6.3: The average RMSD values of claudin-1 and mutants. The analysis is broken down to protein backbone atoms and the backbone atoms of the first (ECL1) and second (ECL2) extracellular loops. The standard deviation from the average value is also displayed (± σ).

Protein	**RMSD_BB /nm**	**RMSD_ECL1/ nm**	**RMSD_ECL2/ nm**
Claudin-1	0.25 ± 0.04	0.28 ± 0.06	0.15 ± 0.04
E48K	0.28 ± 0.04	0.30 ± 0.06	0.13 ± 0.04
S53E	0.28 ± 0.04	0.27 ± 0.06	0.15 ± 0.05
K65D	0.27 ± 0.04	0.29 ± 0.06	0.14 ±0.04
D68S	0.27 ± 0.05	0.28 ± 0.05	0.17 ± 0.07

Table 6.4: The radius of gyration R_g of claudin-1 and mutants (averaged after the first 10 ns). The analysis is broken down to protein backbone atoms and the backbone atoms of the first and second extracellular loops. The standard deviation from the average value is also displayed (± σ).

Protein	***Rg*_BB /nm**	***Rg*_ECL1/ nm**	***Rg*_ECL2/ nm**
Claudin-1	2.05 ± 0.02	1.26 ± 0.05	0.94 ± 0.03
E48K	2.04 ± 0.03	1.31 ± 0.04	0.94 ± 0.02
S53E	2.04 ± 0.02	1.28 ± 0.04	0.95 ± 0.03
K65D	2.05 ± 0.02	1.28 ± 0.03	0.94 ± 0.02
D68S	2.06 ± 0.02	1.29 ± 0.04	0.96 ± 0.05

The breakdown of the number of residues adopting a β-sheet and an α-helix conformation as well as the average number of hydrogen bonds with respect to simulation time are presented in Table 6.5, as indicators of structural stability. The number of average hydrogen bonds is high in the range of 139-141 and the fluctuations are relative small (see Figure 6.7). A general observation from Table 6.5 is that with regards to both hydrogen bonds and the number of residues adopting a regular secondary structure element, the differences in the values between the proteins are relatively small and the fluctuations are in the range of these differences, so they are considered subtle.

Table 6.5: The table shows the average number of hydrogen bonds formed between protein atoms during the trajectory and the average number of residues adopting a β-sheet or an α-helix conformation for claudin-1 and mutants (after convergence). The number in parenthesis in the fourth column, is the equivalent number of residues when we consider only the ECD. The standard deviation from the average value is also displayed (± σ).

Protein	**No Hydrogen bonds**	**Residues (that adopt β-sheet conformation)**	**Residues (that adopt an α-helix conformation)**
Claudin-1	140 ± 6	20 ± 4	105 ± 4 (17 ± 1)
E48K	140 ± 5	22 ± 2	104 ± 2 (17 ± 1)
S53E	141 ± 6	21 ± 3	104 ± 3 (16 ± 2)
K65D	140 ± 6	20 ± 2	106 ± 3 (17 ± 1)
D68S	139 ± 6	21 ± 3	105 ± 4 (17 ± 2)

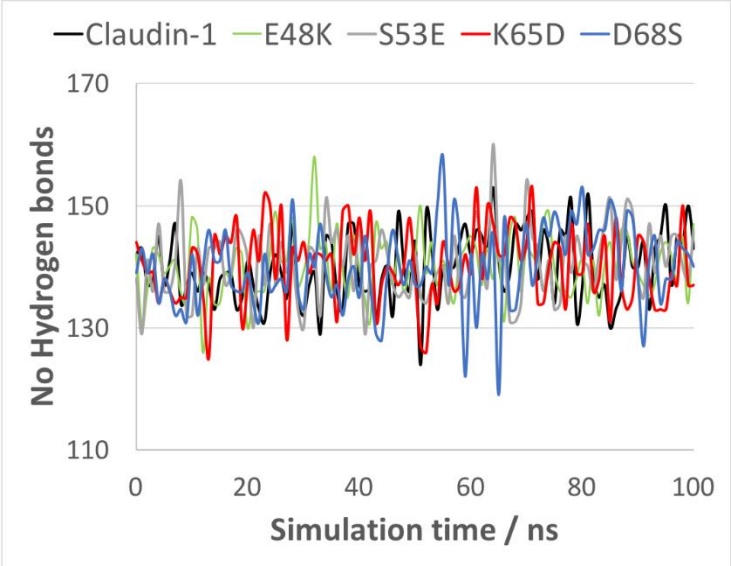


Figure 6.7: The number of hydrogen bonds formed during the trajectory between protein atoms for the wild-type and the mutants (see graph label on top).

The number of residues that adopt a β-sheet or an α-helix conformation was not significantly different amongst the examined claudins. Noticeably, all mutants slightly increased the average number of residues that form β strands. On the contrary the mutants depicted a smaller value of residues adopting an α-helix conformation compared to the wild-type claudin (except K65D in both cases). This means that the β-sheet content is slightly increased upon mutating the polar or charged residues of the ECD of claudin-1, implying that the ECL1 remains slightly more structured (folded). However, when observing the end frame structures in Figure 6.11 and Figure 6.14, it is evident that E48K and D68S mostly increased the number of residues part of the β5 strand which was reduced in the wild-type claudin. Regarding the helical content, we noticed that the mutants slightly reduced their helical conformation that could potentially affect the folding of ECL2. To examine this, we focused the analysis on the ECD particle (from Table 6.5 see parenthesis) and concluded that this is indeed an area that was affected more (the long helix of ECL2) compared to the tightly packed left-handed four helices of the transmembrane region. Note that this longer third transmembrane region is important for the aggregation of claudins and its tilt angle can affect the relative population of the different dimeric interfaces observed (Nakamura et al., 2019).

6.3.2.3 A 'key' difference is in the β sheet orientation of the K65D mutant

The deviation from the starting structure was also characterised in terms of the orientation of the specific secondary structure elements (β1: residues 31-35, β2: residues 44-48, β3: residues 52-56, β4: residues 62-66 and β5: residues 157-161). Specifically, we calculated the 'opening up' angle presented in Table 6.6 and the tilt angle presented in Table 6.7 for each individual β strand. The definition of the angles was given in Chapter 3. Briefly, the tilt of a particular β strand is the angle between the vector that defines its axial direction and the vertical axis, and the opening angle is the one between the long helix and the individual β strand of the ECD particle.

Table 6.6: The average 'opening' angles of the individual β-strands with the respective standard deviation (±σ). The underlined values indicate the strand closer to where each mutation occurred.

Protein	**Opening angle / degrees**				
	β1	β2	β3	β4	β5
Claudin-1	33 ± 6	47 ± 9	50 ± 10	72 ± 17	30 ± 10
E48K	28 ± 5	<u>47 ± 9</u>	50 ± 8	68 ± 9	26 ± 8
S53E	29 ± 6	46 ± 9	<u>51 ± 11</u>	70 ± 12	32 ± 9
K65D	29 ± 6	47 ± 9	55 ± 8	<u>76 ± 10</u>	28 ± 7
D68S	32 ± 7	47 ± 11	53 ± 13	73 ± 18	29 ± 8

The average 'opening' angles show some deviation from the respective average angles of claudin-1 mostly within 5 degrees difference. The 'opening' angles of the β1 strand were reduced for all mutants compared to claudin-1. S53E demonstrated a bigger angle

for the β5 strand compared to the β1 strand which implies that the β5 strand of ECL2 'opens up' (near the edge of the β-sheet domain). S53E has previously been found to disturb the barrier properties of claudin-1 and induced charge-unselective pores (Veshnyakova et al., 2012). This area of the β-sheet domain was also found to be important for the stability of the trans-dimers in Chapter 4 of the thesis. Another important finding from this analysis was that K65D displayed the largest difference between the β2 and β3 strands. The end frame structure of the mutant presented in Figure 6.13 clearly shows how the distance between these strands was increased resulting in a different orientation of the edge of the extended β-sheet domain which is regarded as the 'key' interface in the face-to-face Suzuki model.

Table 6.7: The average tilt angle of a particular secondary structure element and the vertical axis. The standard deviation from the average value is also displayed (±σ) and the underlined values depict the strand closer to where the mutation occurred.

Protein	**Tilt angle-Angle relative to the vertical axis / degrees**				
	β1	β2	β3	β4	β5
Claudin-1	16 ± 8	28 ± 10	37 ± 13	63 ± 17	18 ± 9
E48K	16 ± 8	<u>29 ± 10</u>	36 ± 12	57 ± 10	20 ± 8
S53E	17 ± 9	33 ± 13	<u>44 ± 11</u>	65 ± 11	18 ± 10
K65D	17 ± 8	33 ± 12	46 ± 11	<u>68 ± 12</u>	19 ± 8
D68S	15 ± 8	30 ± 11	43 ± 12	64 ± 14	16 ± 9

When considering the tilt angles of the β strands relative to the vertical axis, we noticed that the mutants mostly increased these values except for E48K (β3 and β4) and D68S (β1 and β5). Therefore, the mutants also showed the tendency to become slightly more horizontal, as the sole ECD of claudin-1 in Chapter 3. It is worth mentioning that the

corresponding tilt angle of the long helix (ECL2) was calculated as 32^0±7, 33^0±5, 28^0±8, 33^0±7 and 28^0±8 degrees for claudin-1, E48K, S53E, K65D and D68S, respectively (averaged after the first 10 ns of the trajectory for the lowest temperature replicas).

To conclude, although the relative orientation analysis did not reveal big differences between the individual mutants and their original model (which was similar to the model of the wild-type claudin-1) it highlighted two key areas, the edge of the β-sheet domain (β3 and β4) and the area between β1 and β5 strands as the areas that changed in K65D and S53E, respectively. These regions were found critical in this thesis for the cis- and trans- interaction of claudin-1, respectively. The tilt angle and opening angle of the individual β-strands of claudin-1 are presented in Figure 6.8 and Figure 6.9, respectively. They both demonstrate that the β4 strand located at the edge of the ECD (the head of the 'Olympic torch') fluctuated more compared to the other β strands.

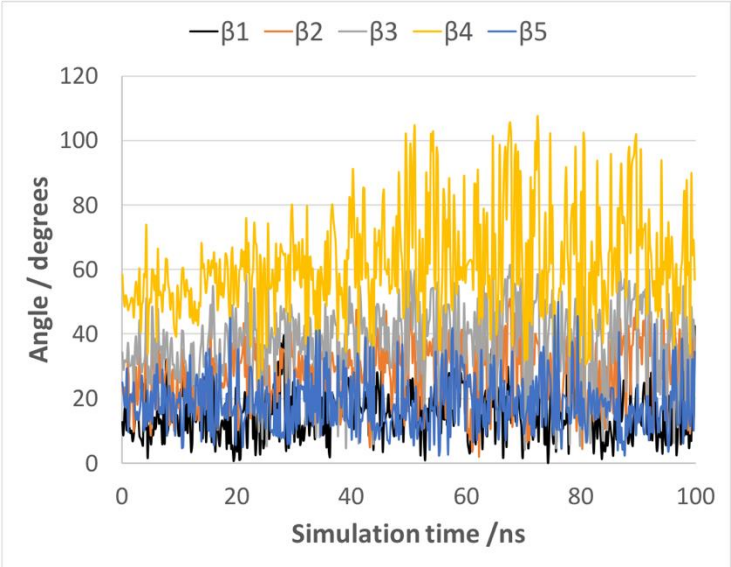


Figure 6.8: The tilt angle of the β strands of claudin-1 with respect to simulation time (of the lowest temperature replica).

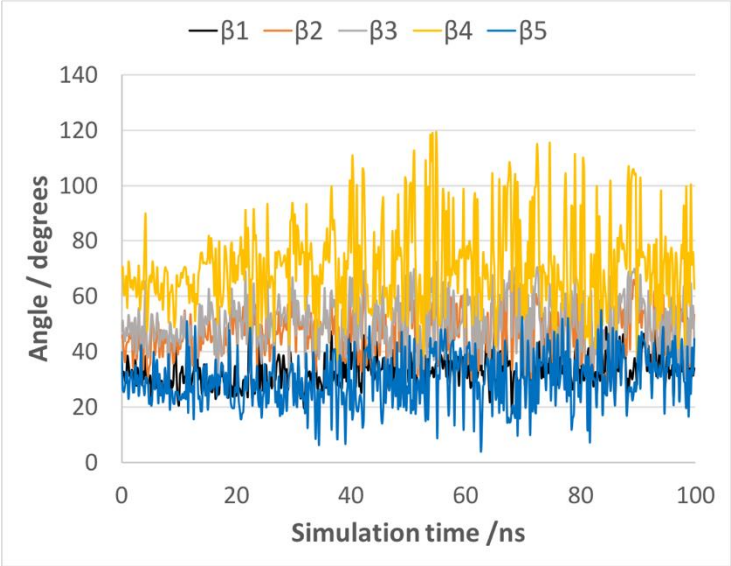


Figure 6.9: The opening angle of the β strands of claudin-1 with respect to simulation time (of the lowest temperature replica).

6.3.3 Characterisation of the end frame structures of the wild-type protein and mutants

In this section, we characterise in detail the end frame structures of the lowest temperature replicas of both the wild-type protein and the mutants; since they are considered the most favourable conformation of the molecules. To illustrate more clearly the differences between them, we superimposed their structures with UCSF Chimera (Pettersen et al., 2004). Firstly, the wild-type claudin-1 was examined, and then we used the end structure of the wild-type claudin and superimposed the final structures of the mutants on it for comparison. This helped us identify any differences between the two end structures. Subsequently, we tried to rationalise the experimental findings mentioned in the introduction with specific deviation in the final structures.

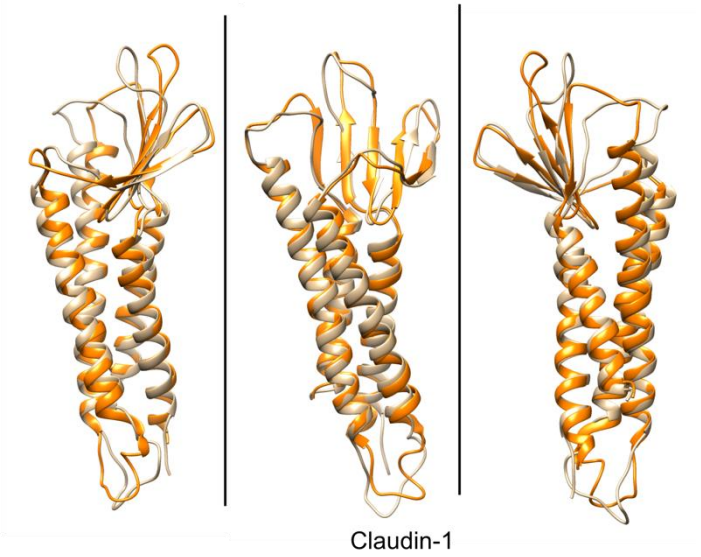


Figure 6.10: The final frame structure of Claudin-1 (gold) superimposed on the original model (orange) in cartoon representation from different viewing points.

The end frame structure (gold) superimposed on the starting equilibrated structure of the wild-type claudin-1 (orange) from the lowest temperature trajectory is presented in Figure 6.10. The value of the RMSD suggested that the differences between them were not significant. The figure shows that the biggest difference was concentrated on the ECL2 and particularly in the β5 strand which lost its β folding. There is also a small difference in the orientation of the β3 strand and in the orientation of the short loop that connects β1 and β2 strand (V1 region). However, the short loops are generally expected to fluctuate a lot since they do not form a regular secondary structure. Also, the number of residues part of β5-strand did fluctuate during the trajectory but a significant trend was not observed. The alpha helical transmembrane structure remained mostly the same with a small difference in the tilt angle at the edge of the third transmembrane helix.

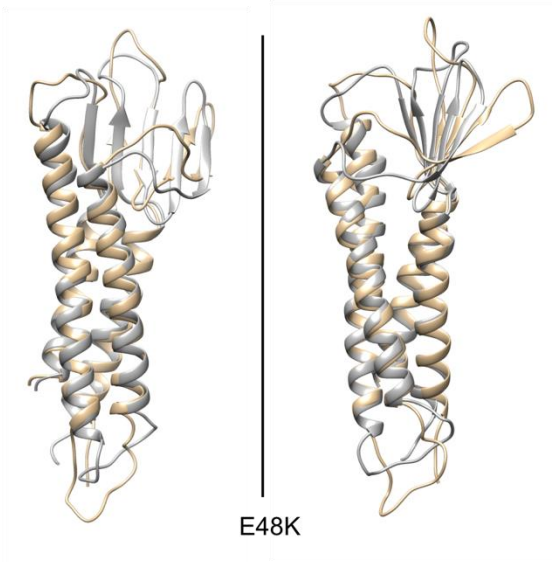


Figure 6.11: The final frame structure of E48K (silver) superimposed on the final frame structure of the wild-type claudin-1(gold).

The E48K has changed its structure as observed in Figure 6.11 in the sense that the β-sheet content was increased and the β strands did not become more horizontal but rather stayed close to the β5 strand from the second loop. The β3 and β4 strands also displayed different orientation compared to claudin-1. Another area that seems to be affected was the one between the β4 strand and the second transmembrane helix composed of mostly coils. This area seems to be lower than the corresponding area in the wild-type protein. Furthermore, we noticed that there were some amino acids from the third transmembrane long helix (GLU147, PHE148 and TYR 149) that became turns and lost their α-helical conformation while at the same time it looks like a small bend (kink) was created in the helix. Generally, it appears like this mutant did not change its structure a lot, because the β-sheet domain seems more ordered and compact. But the RMSD value was rather high and the R_g value of the ECL1 was increased maybe due to the opening up in the unstructured area mentioned above. It should be mentioned that the superimposition was between the end frame of the wild-type and the mutant, but the RMSD and R_g measured the deviation from the mutants initial model.

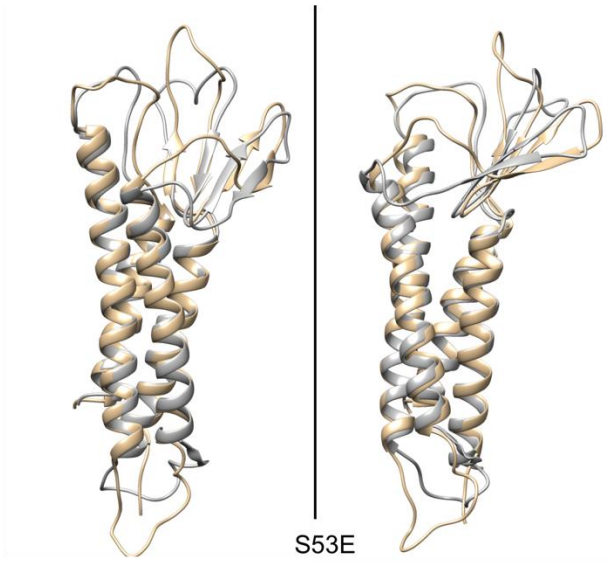


Figure 6.12: The final frame structure of the S53 mutant (silver) superimposed on the final frame structure of the wild type claudin-1 (gold).

With regards to the S53E, it appears that this mutant has changed its structure significantly. The end frame structure presented in Figure 6.12 demonstrates how the β strands of the ECD change their conformation. The β3 and β4 remain close, but the β2 strand moves away from the β1 that stayed closer to the β5 strand. This resulted in different opening angles between the β strands. The β1 and β5 strands lost some residues from their extended conformation thus, their lengths were reduced. The mutant also demonstrated that the unstructured area shown in the red dashed frame in Figure 6.1 was lower than the corresponding one of claudin-1 (similarly to E48K). The specific mutant was found to increase the permeation of ions through the TJ in experiments and the result obtained here, confirms the critical role of the central region of the β-sheet domain that is expected to influence strongly the barrier structure of the TJs. The β1-strand was also found important for the trans-interaction in Chapter 4 of the thesis (tyrosine residues 33 and 35).

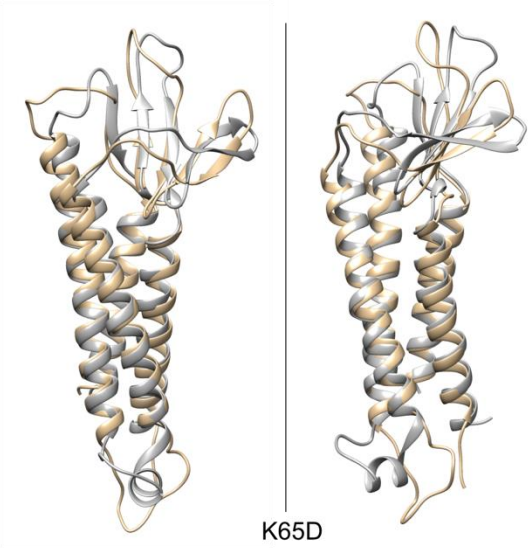


Figure 6.13: The final frame structure of K56D (silver) superimposed on the final frame structure of the wild-type claudin-1(gold).

The final frame structure of K65D (silver) superimposed on the final frame structure of claudin-1 (gold) is presented in Figure 6.13. The figure demonstrates that the mutant shows significant changes in the structure of the extended β-sheet domain. Specifically, the β3 and β4 strand became more horizontal and moved away from β2 strand. It seems like the domain 'split' in the middle of the extended conformation resulting in differing orientations of the β-strands. Additionally, the unstructured area at the front of the β-sheet (dashed frame in Figure 6.1) moved further away (higher) from the bilayer's surface. Both sides of the ECD particle, namely the unstructured area and the edge of the domain (towards the β4 strand) are critical for the linear and face-to-face interface based on the Suzuki model (Suzuki et al., 2015). Also, the back side of the β-sheet participates in the former interface. This mutant impeded the formation of TJ-like strands in experiment (Piontek et al., 2017). It was suggested that the substitution directly or indirectly impedes the trans-interaction of claudin-1 (measured as the enrichment of constructs at contacts between claudin expressing cells). The unstructured region at the front of the β-sheet was also found to participate in the trans-interaction of claudin1- in Chapter 4. Thus, it can be argued that significant structural changes are observed in critical regions of the K65D mutant with replica exchange simulations.

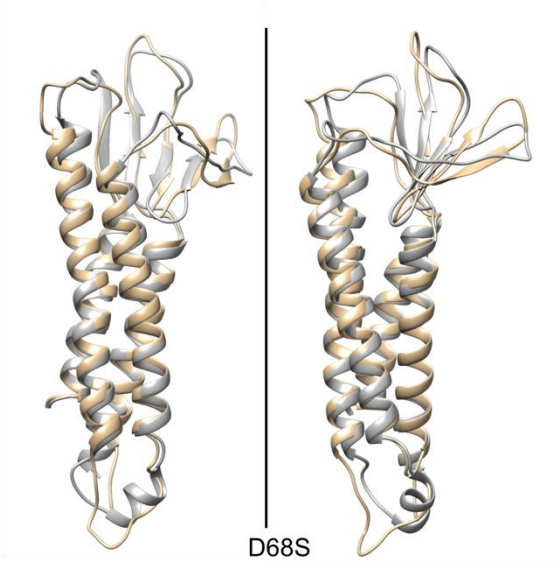


Figure 6.14: The final frame structure of D68S (silver) superimposed on the final frame structure of the wild-type claudin-1(gold).

The superimposed final structure of D68S (silver) on the wild-type claudin-1 (gold) presented in Figure 6.14 demonstrates that D68S also displayed a difference in the relative orientation of β3 and β4 strand like K65D did but not so intense. The β4 strand was more horizontal while its length was reduced (only residues 63-64 are part of the extended β-sheet conformation). On the contrary, the β5 strand displayed an increased length compared to claudin-1.

To summarise, the area above the β-sheet is important for the properties of the TJ barrier (see Chapter 4). A compact β-sheet domain helps strengthen the barrier structure, while an increased distance between the strands (as observed in S53E and K65D) would probably result in the formation of a small pore structure. The 'loosening' of the compactness of the β-sheet would definitely affect the trans-interaction that possibly starts through the variable regions V1 and V2 but slides further in the centre of the β-sheet for increased stability and flexibility (see Chapter 4). Additionally, the β4 strand alongside the unstructured area after it (red dashed frame in Figure 6.1), affect the cis-interaction of claudin-1 (see Chapter 3) and both areas changed their orientation in K65D, a mutant that abolished the formation of TJ strands. The area at the back of the

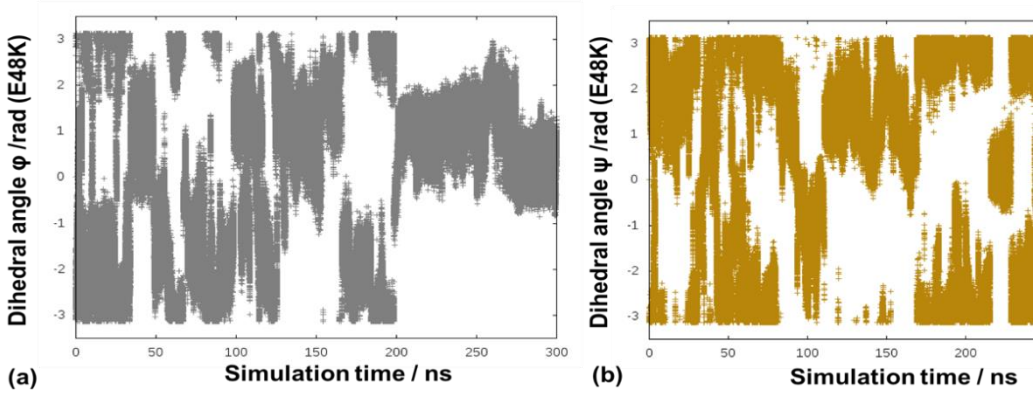
β-sheet domain (mainly behind β1, β2 and β5) are part of the formed linear interface, so when the distance between the strands is increased it could potentially affect the formed cis-interfaces. The kink of the long helix of ECL2 observed in the other cases possibly affects the relative population of dimeric interfaces but does not abolish the cis- or trans-interaction between claudins.

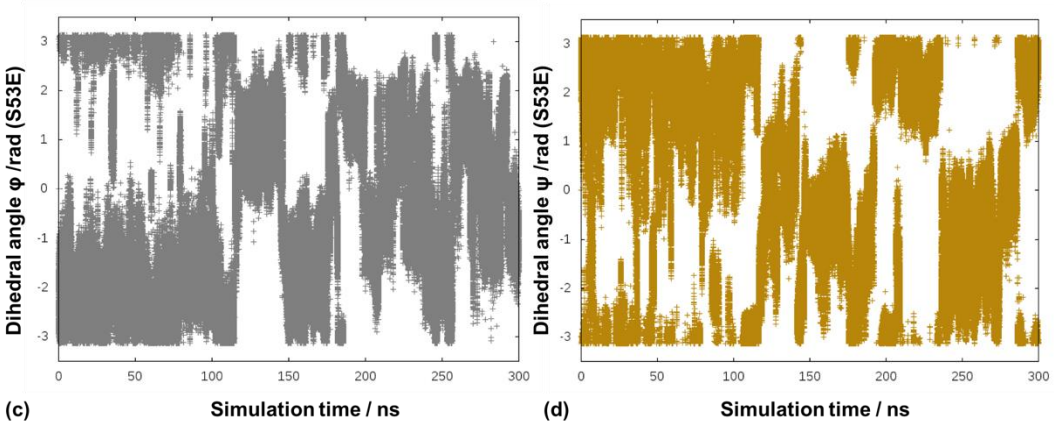
To conclude, although initially the differences between the wild-type protein and mutants were not significant in terms of RMSD and R_g values and relative orientation of the β-strands, it is clear that the differences observed in the superimposed final structures of S53E and K65D could explain the experimental observations of Piontek et al., and Veshnyakova et al., (Piontek et al., 2017, Veshnyakova et al., 2012). However, replica exchange simulations did not span a big range of temperatures and the energy barriers that the proteins overcame, are not expected to be very high. On the contrary, metadynamics simulations offer the benefit of exploring the free energy surface more effectively and the computational cost is significantly lower. Hence, in the next session we discuss the results obtained from the latter method and present a comparison between the lowest free energy structures obtained with the two enhanced sampling techniques.

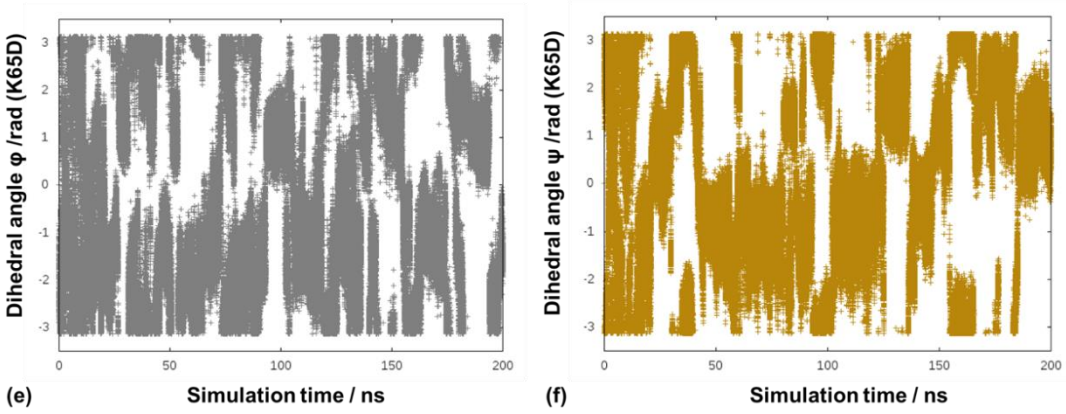
6.3.4 Metadynamics simulations to estimate the free energy surface of the mutants

6.3.4.1 Assessing the convergence of metadynamics simulations

Prior to analysis, it is important to ensure that the claudins during the metadynamics simulations were able to visit all the relevant conformations; meaning they were able to diffuse freely over the whole range of values that the CVs can take. To ensure this, the two dihedral angles φ and ψ were monitored during the simulation. The φ and ψ angles as a function of simulation time are presented in Figure 6.15 and it is clear that there were many re-crossings through the CV space for all examined mutants. The primary reason to report the angles in radians is because by default the metadynamics code uses radians and also for consistency with the other analysis tools used later in the chapter (Metadyn View).


Dihedral angle φ /rad (E48K)
Simulation time / ns
(a)
Dihedral angle ψ /rad (E48K)
Simulation time / ns
(b)


Dihedral angle φ /rad (S53E)
Simulation time / ns
(c)
Dihedral angle ψ /rad (S53E)
Simulation time / ns
(d)


Dihedral angle φ /rad (K65D)
Simulation time / ns
(e)
Dihedral angle ψ /rad (K65D)
Simulation time / ns
(f)

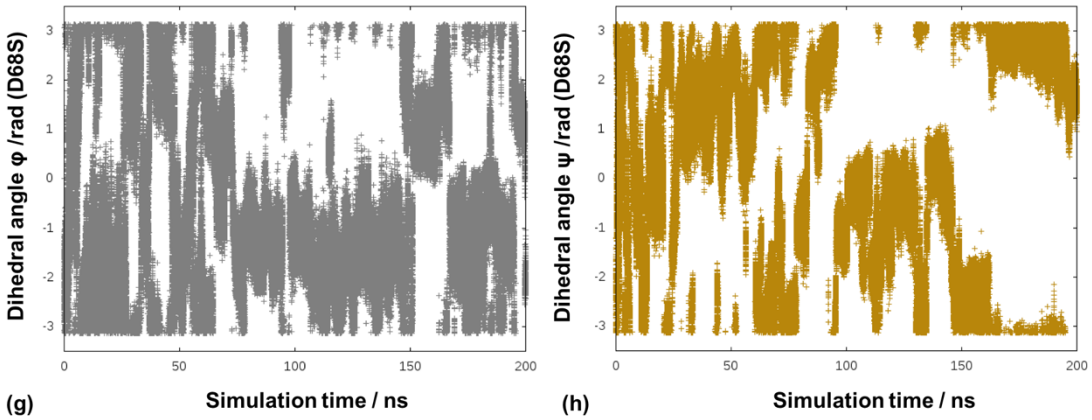
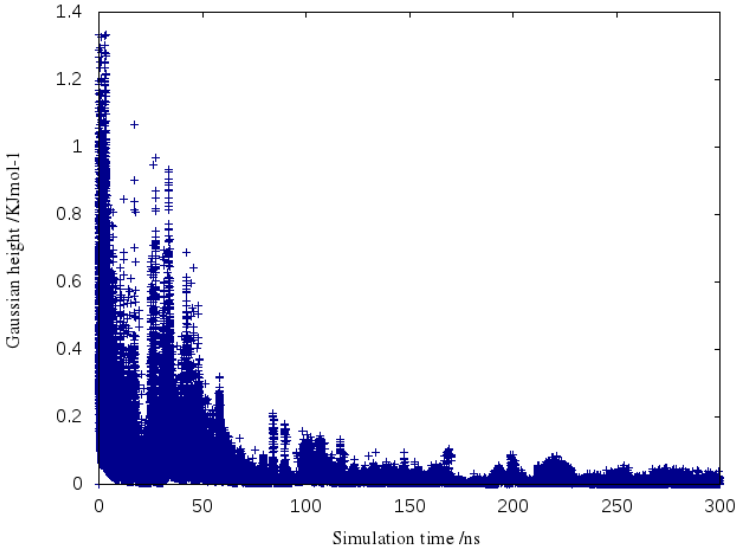


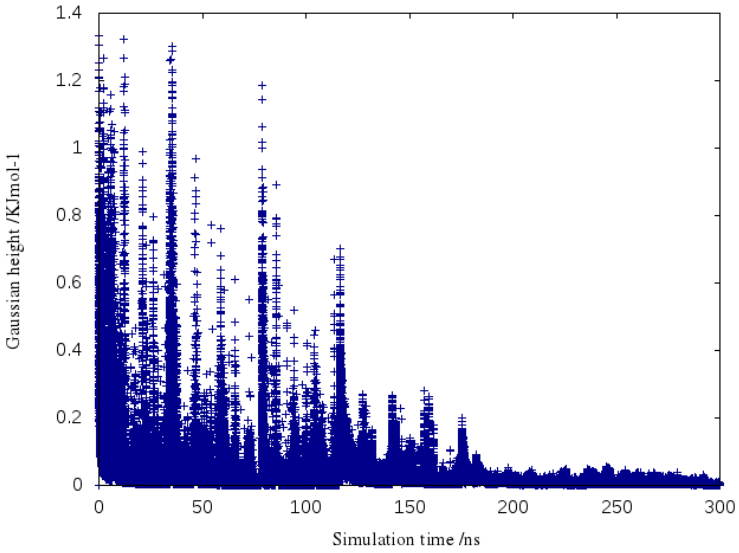
Figure 6.15: Time evolution of the CVs φ and ψ. The φ ((a),(c),(e) and (g)) and ψ ((b),(d),(f) and (h)) dihedral angles of the mutated residues of E48K, S53E K65D, and D68S respectively. Gnuplot was used to prepare the graphs (Williams et al., 2008).

The re-crossings seen in Figure 6.15 are an indication that the metadynamics simulations have converged and so the estimated bias potential is close to yielding a flat potential energy surface. We notice that for E48K (Figure 6.15 (a) and (b)) after 200 ns the system seems to be confined in a specific region of φ and ψ values. However, during the simulation the angles changed a lot and there were enough re-crossings. Therefore, we can assume that the bias potential has converged smoothly to the estimate of the free energy for this system too.

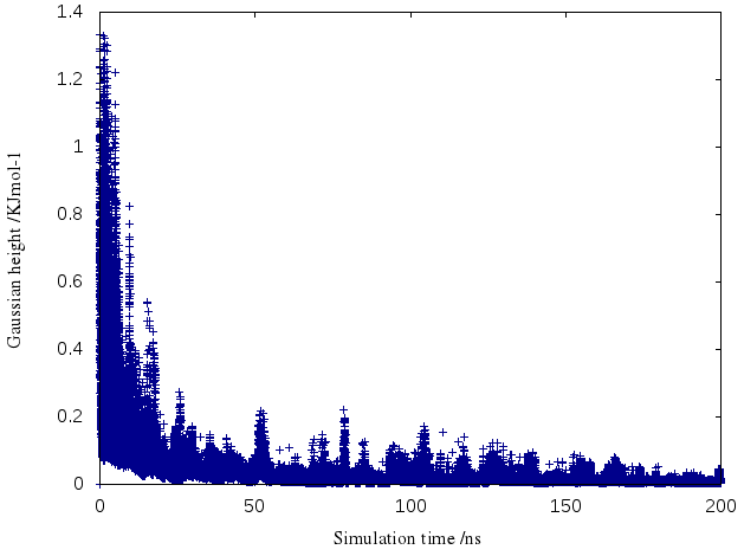
It is also important to monitor the height of the added Gaussians to determine whether the run has converged according to the well-tempered recipe. If the height decreases and tends to become zero, this means that the added potential is also zero and the free energy surface has flattened. Hence, it is a strong indication that one has reconstructed the free energy landscape with success. The height of the added Gaussians with respect to simulation time for all examined systems is presented in Figure 6.16. We notice that the height at the beginning of the simulations was slightly higher than the initial height (1.2 KJ mol^{-1}) set in the plumed.dat file, and this is because the height was re-scaled using the bias factor. The graphs clearly show the constant decay of the Gaussian height during the simulations.


Gaussian height /KJmol-1
Simulation time /ns
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
50
100
150
200
250
300

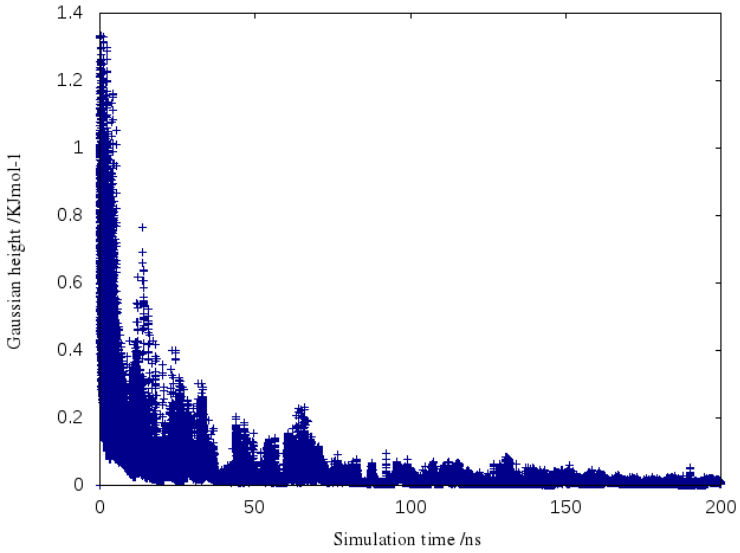
(a)


Gaussian height /KJmol-1
Simulation time /ns
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
50
100
150
200
250
300

(b)



(c)



(d)

Figure 6.16: The height of the added Gaussians with respect to simulation time for the (a) E48K (b) S53E (c) K65D and (d) D68S mutants (Gnuplot was used for plotting the data (Williams et al., 2008)).

Figure 6.16(a) shows that towards 200 ns the E48K mutant exhibited a small increase in the Gaussian's height and Figure 6.16(b) shows that the Gaussian height increased in

many instants during the trajectory (S53E mutant). Therefore, we have extended both simulations for another 100 ns, resulting in a total duration of 200 ns for K65D and D68S, and 300 ns for E48K and S53E.

Finally, metadynamics can generate the free energy surface (FES) of the mutants as a function of the CVs. The FES provides invaluable information about their conformational stability, since it shows for a specific combination of CVs the estimated free energy and thus, the likelihood of the particular conformation. One can calculate a one-dimensional free energy profile from the two-dimensional metadynamics simulations, which can be used (see below) to ascertain convergence.

Figure 6.17 gives the estimate of the free energy of the D68S mutant as a function of the dihedral angle φ (grey), and of ψ (gold). The corresponding graphs for the other mutants can be found in the thesis Appendix A2.

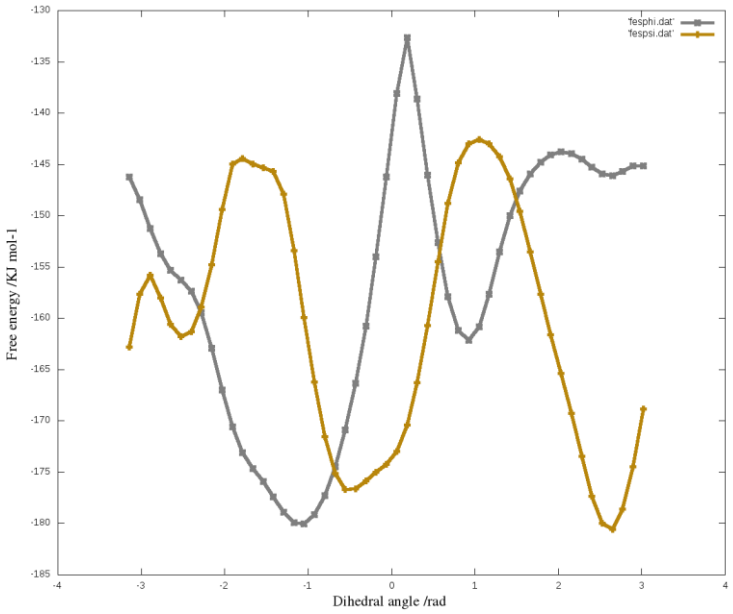


Figure 6.17: The change in the free energy with respect to φ (grey) and ψ (gold) dihedral angles of the D68S mutant. The FES was constructed as a function of either φ or ψ and the relevant energy barriers are shown. The graph was produced with Gnuplot (Williams et al., 2008).

To test for convergence we plot the 1-dimensional free energy surfaces repeatedly as a function of simulation time in Figure 6.18. Towards the end the reconstructed profiles should be similar and only apart by a small constant offset. This is another strong indication that the metadynamics simulations have converged. In order to do that, the sum_hills tool available in PLUMED was utilised and its stride option enabled to give an estimate of the free energy profiles every 500 Gaussians kernels deposited (e.g. run the command: plumed sum_hills –hills HILLS –idw phi/psi –kt 2.5 –mintozero —stride 500). With the option -idw one can define the desired variable (e.g. φ) and integrate out the other variable (i.e. ψ).The global minimum was set to zero in all profiles and the resulting plot for the D68S mutant is presented in Figure 6.18.The reconstructed profiles for the other mutations are in the Appendix A2 of the thesis.

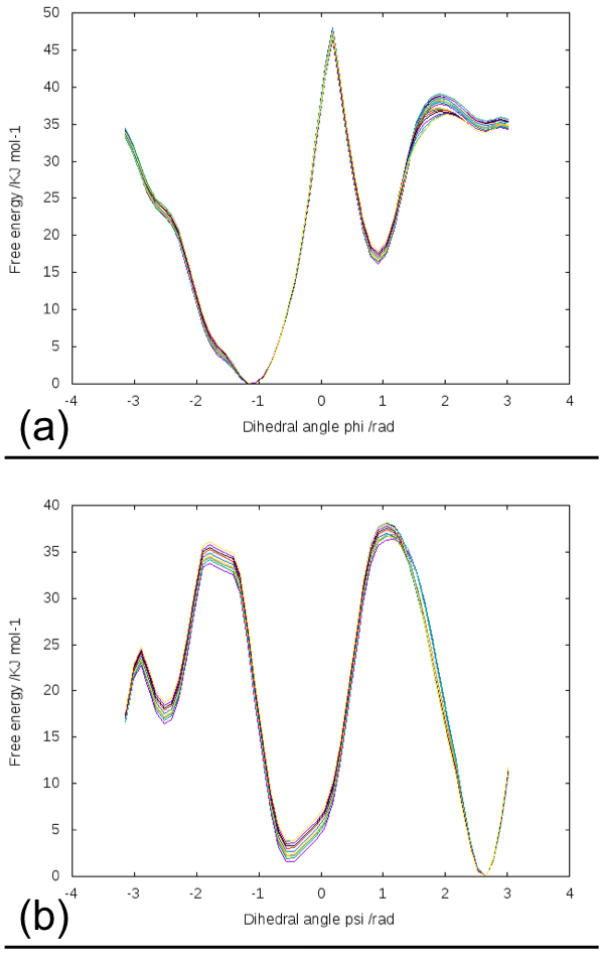


Figure 6.18: The reconstructed free energy profile (every 500 gaussians) for the D68S mutant with respect to the dihedral angles (a) φ and (b) ψ. The graphs show the last 20 estimates of the FES coloured differently (from fes_380.dat to fes_400.dat).

6.3.4.2 Metadynamics simulations highlight the important local regions of the wild-type protein affected by the mutations

Metadyn View is a user-friendly web-based viewer to visualise and analyse the FES calculated by metadynamics methods (Hošek and Spiwok, 2016). It reads the HILLS file produced during the simulation which contains a list of the Gaussian kernels deposited (Hošek and Spiwok, 2016). We used this tool to produce the FES of all examined mutants. The generated graphs depict the relationship between the dihedral angles (φ and ψ) and the change in free energy as a colour plot. The colours used for the energy are shown as a colour bar on the right-hand side of the plots, the x-axis is the value of the angle φ, and the y-axis is the dihedral angle ψ.

The energy surface of the E48K mutant is presented in Figure 6.19 and we notice that there is a single lowest free energy basin coloured dark blue (bias ≅ -187 KJ mol^{-1}). The difference in the free energy between the two extreme points in this specific basin is low (approximately 0.37 KJ mol^{-1}) and between these points and points from other low energy basins more than 10 KJ mol^{-1}. The values of the CVs in the two extreme points in the lowest energy basin, were approximately φ≅0.55 rad and ψ≅1.04 rad with energy equal to -187.03 KJ mol^{-1} depicted by a red star in the plot, and φ≅1.25 rad and ψ ≅ 0.31 rad with free energy equal to -186.66 KJ mol^{-1} depicted by a yellow star. We notice that although the energies were very close the values of the angles were significantly different. Given that the energies are close, we could consider the two minimum as essentially a part of the same minimum basin. The red circle in Figure 6.19 corresponds to the combination of φ and ψ values, of the structure extracted from the last frame of the replica exchange simulation. Notice that this structure is also found in an area with low free energy but not in the basin with the lowest free energy. This indicates as previously suggested that replica exchange did not manage to overcome all the relevant high-energy barriers, perhaps because of the small temperature range used.

One very useful functionality of the Metadyn View is that it provides the user with the values of the closest HILLS file to a specific point (time records). By clicking on the point of interest on the plot, Metadyn View will display the hills which are closer to the selected point in the CV space and thus, we can extract the relevant frames from the

trajectory. We can also double-check this information from the files created during the simulation.

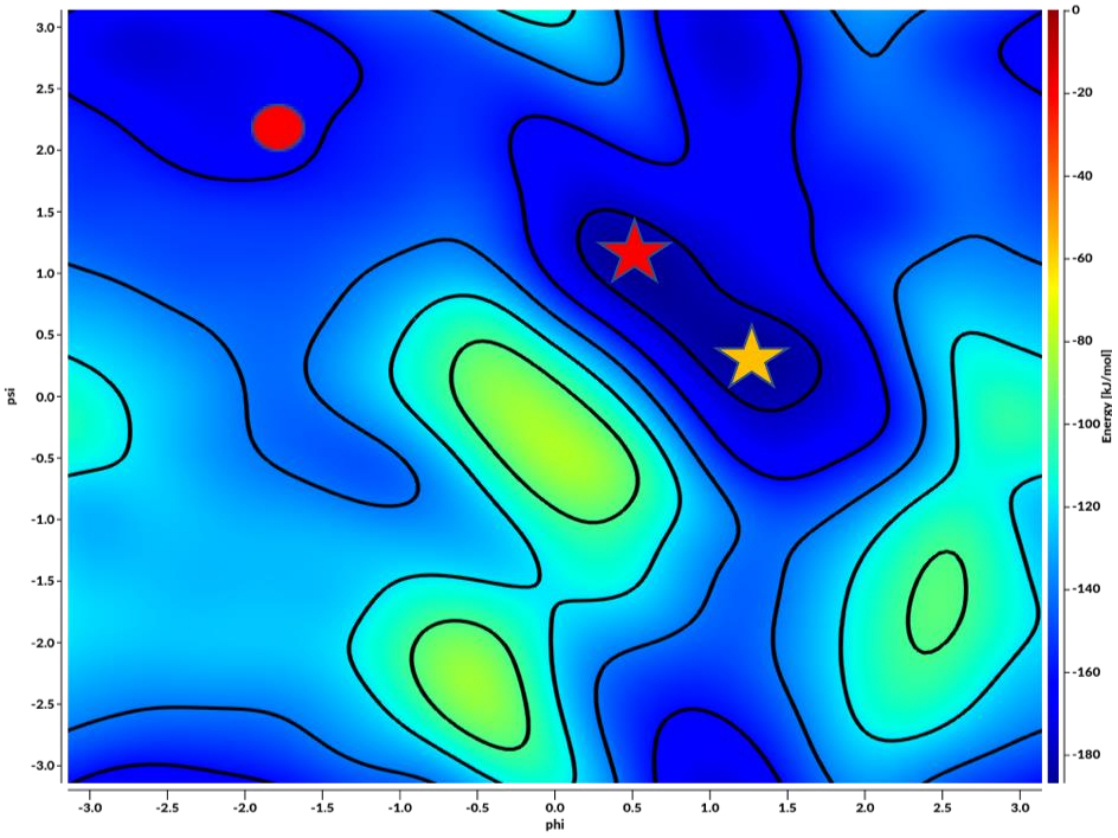


Figure 6.19: The free energy surface of E48K mutant visualised with Metadyn View. The gradient of the free energy is shown in the bar on the right-hand side of the plot and the angles φ and ψ on the x- and y-axis respectively. There are two extreme energy points in the lowest energy basin depicted by a red and yellow star. Also, a red circle depicts the φ and ψ angles of the structure extracted from the replica exchange simulation.

We have extracted the two relevant frames (suggested time stamps) of the E48K mutant which are presented in Figure 6.20. We conclude that there was significant deviation in the structure of the head of the 'Olympic torch' from its starting conformation and specifically in the β-sheet domain. The position 48 is strategically placed in the middle of the extended β-sheet and its mutation seems to affect significantly the connection between the neighbouring β1 and β3 strands and affect the coherence and coordinated movement of the β-sheet in total. The β2 strand where the mutation occurred, has

significantly reduced the number of residues that adopt a β strand conformation. It is important to note that there were also deviations in the intracellular loop region of the mutants but here we focus mainly on the extracellular region. This result is not in agreement with the result from the replica exchange simulations, where the mutant did not display significant deviation in the extended β-sheet structure.

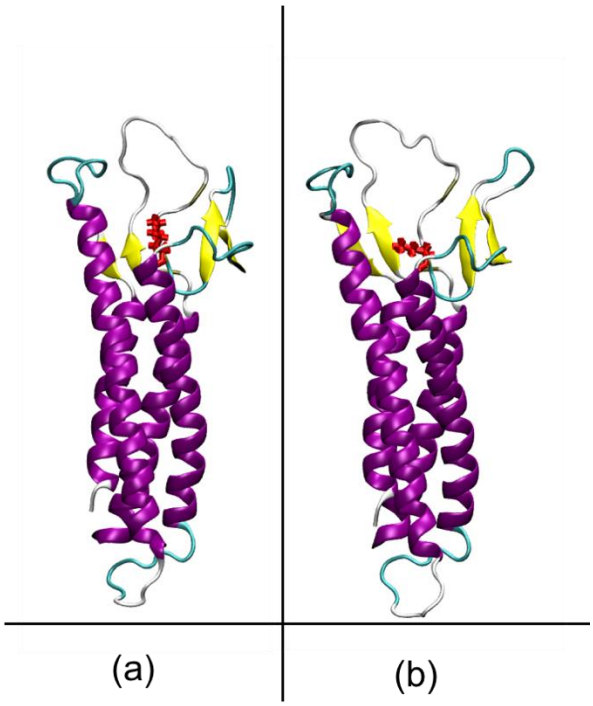


Figure 6.20: The two lowest free energy structures of E48K according to Metadyn View (timestamp (a) t=249800 ps and (b) t= 276500 ps) where we can see that the chi angles of the side chain are also different. The claudin is shown in cartoon representation and coloured according to the secondary structure elements while the mutated residue is coloured red and shown in bond representation. Rendered with VMD (Humphrey et al., 1996).

For synopsis, Table 6.8 shows the φ and ψ angles of the lowest free energy structures for the E48K mutant calculated with the replica exchange simulations and the metadynamics ones. The two methods calculated different values for the dihedral angles but as we can see in Figure 6.19 they are both in low free energy regions.

Table 6.8: The φ and ψ values of the E48K mutant from both the replica exchange and metadynamics simulations. The angles are reported in radians and in the parenthesis in degrees.

E48K	**φ /rad / (degrees)**	**ψ /rad / (degrees)**
Replica exchange	-1.8 (-103^0)	2.3 (132^0)
Metadynamics	0.6 (34^0)	1.0 (57^0)

With regards to the next examined mutant, namely the S53E, we follow the same protocol and calculate the values for the φ and ψ dihedral angles of the lowest free energy structure and compare them to the ones calculated with the replica exchange simulations. Figure 6.21 illustrates the FES of the mutant visualised with Metadyn View and there is again a single deep lowest free energy basin with two extreme points (energy = -185.91 KJ mol^{-1} where φ= -2.28 and ψ= 2.35 rad (red star) and energy= -185.73 KJ mol^{-1} where φ= -2.49 and ψ= 2.61 rad (yellow star)). The values of the dihedral angles are not very different as the energy basin is not very wide.

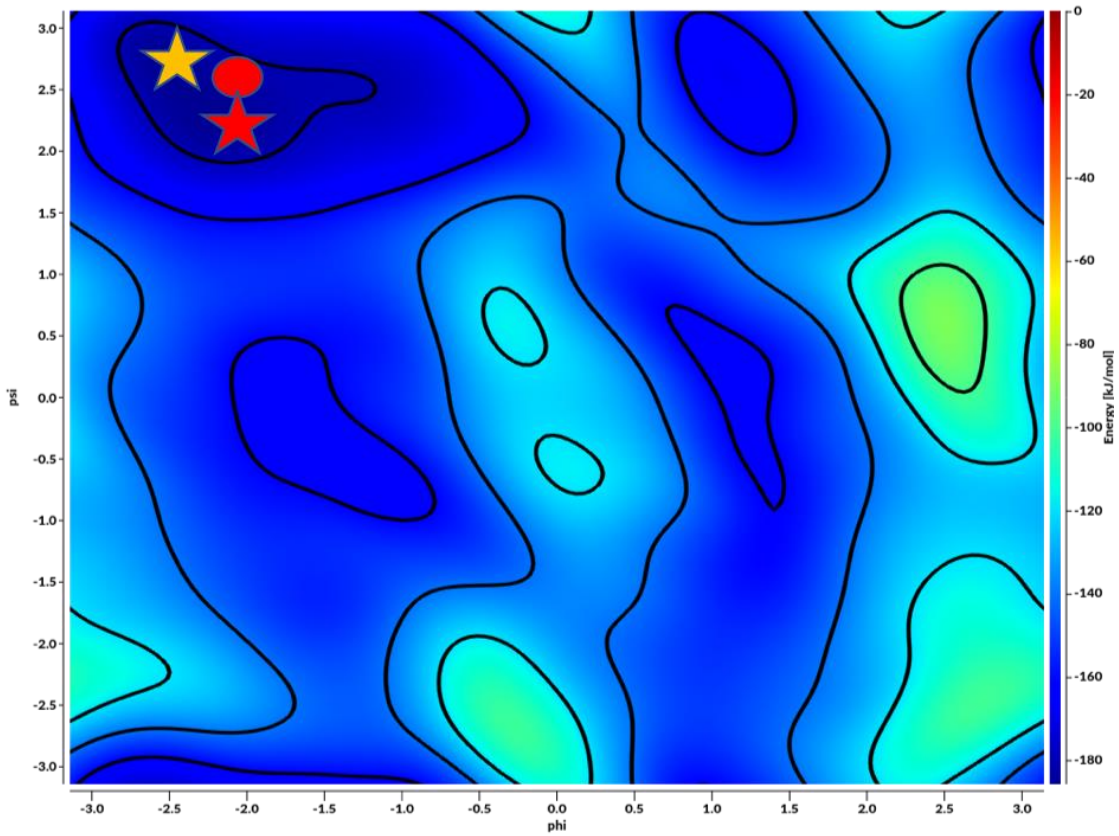


Figure 6.21: The free energy surface of S53E mutant visualised with Metadyn View. The dihedral angles φ and ψ are shown in the x- and y- axis. The bar on the right-hand side of the plot shows the gradient of the free energy. There are two extreme energy points in the lowest energy basin shown as red and yellow stars and a red circle depicts the lowest energy structure calculated with replica exchange.

The closest hills were at 14361 ps, 105844 ps and 42092 ps and the related structures are presented in Figure 6.22. Noticeably, the structure at the second timestamp ((b) in Figure 6.22) was different from the other two. It displayed an increased distance between β5/β1 strands which is in agreement with the replica exchange result. The (b) and (c) structures also showed an increased tilt angle in the second transmembrane helix while (a) and (c) showed that there is a 'kink' in the long helix of ECL2 which is considered important for the claudin-claudin cis-interaction (Nakamura et al., 2019). It has been reported in the literature that the presence or absence of the helix bending, can change the positions of residues related to claudin interactions that could also result in

differences in the morphology and adhesiveness of the TJ strand (Nakamura et al., 2019).

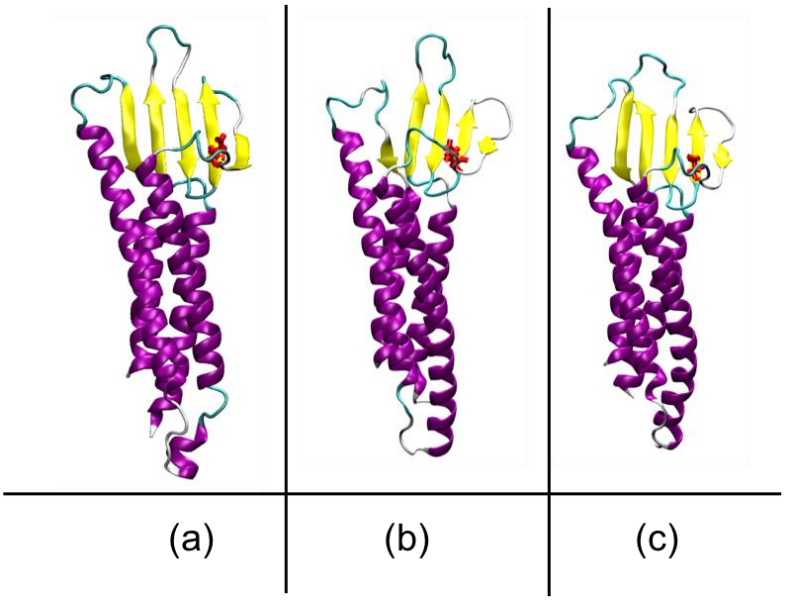


Figure 6.22: The lowest free energy structures of S53 according to Metadyn View and the closest hills file. The time stamps for the extracted frames were at (a) t=14361 ps, (b) t=105844 ps and (c) t=42092 ps. The protein was rendered with VMD and shown in cartoon representation while the mutated residue is coloured red and shown in bond representation (Humphrey et al., 1996).

Similarly, Table 6.9 summarises the φ and ψ angles for S53E calculated with both methods as described previously, and we note that there is not a significant difference between them (roughly less than 15^0 degrees). Hence, with regards to these angles the methods are in agreement and also in the fact that the structure opens up close to the centre of the β-sheet ((b) in Figure 6.22). However, the large deviation in the extended β-sheet conformation seen in replica exchange was not observed here.

Table 6.9: The φ and ψ values of the S53E mutant calculated with both replica exchange and metadynamics simulations. The angles are shown in radians and in the parenthesis in degrees.

S53E	**φ / rad/ (degrees)**	**ψ / rad/ (degrees)**
Replica exchange	-2.2 (-125^0)	2.5 (146^0)
Metadynamics	-2.3 (-132^0)	2.3 (132^0)

The FES of the K65D is presented in Figure 6.23 and it is clear that there is a deep well of low free energy (bias ≅ -169.85 KJ mol^{-1}) at φ= -1.36 rad and ψ= -0.61 (see red star). The Metadyn View calculated the closest hills at 1711 ps, 65997 ps and 93157 ps. The relevant structures are presented in Figure 6.24 where we notice that the β3 and β4 strand became more horizontal and it seems that the hairpin loop area was 'pulled back'. If we assume that this protein region is located at the mouth of the barrier, this could mean that the radius of the pore at the entrance is significantly increased; while if this area is in the middle/heart of the pore, this could also affect the permeability of small molecules and ions through the TJ pore (different type of pore models are discussed in Chapter 4 of the thesis). Most importantly however, because the mutant abolished the TJ-like strand formation, this means that this area is critical for the aggregation of claudin-1. When we compare the results presented here with the ones from the replica exchange simulations we find that in both cases the β3 and β4 strand becomes more horizontal and the unstructured area of the first loop tends to move higher than the surface of the lipid bilayer, thus they are in agreement. The fact that this area is higher while the small loop connecting β3 and β4 is pulled back, suggests that the favourable cis-interaction seen in the linear interface (Chapter 3) might be blocked, explaining why these mutants failed to display TJ like-strands in the experimental study. Note that the middle area of Figure 6.23 is an area associated with high energy barriers (values of φ in the range of -0.5 : 0.5 rad) hence, these values are considered unfavourable for the dihedral angle φ.

Table 6.10 presents the values of the two CVs as calculated by both replica exchange and metadynamics simulations and we note that the values are different, unlike the S53E

mutant. However, we acknowledge that in such big systems maybe replica exchange needs more time to converge or maybe more replicas to span a bigger region of temperature. From Figure 6.19, Figure 6.21 and Figure 6.23 it is clear that the dihedral angles calculated with replica exchange are in low free energy regions too, but not in the lowest basin. There might be other favourable structures found in local free energy minima that depict different φ and ψ values but these structures could also be important states that could affect the function of the protein.

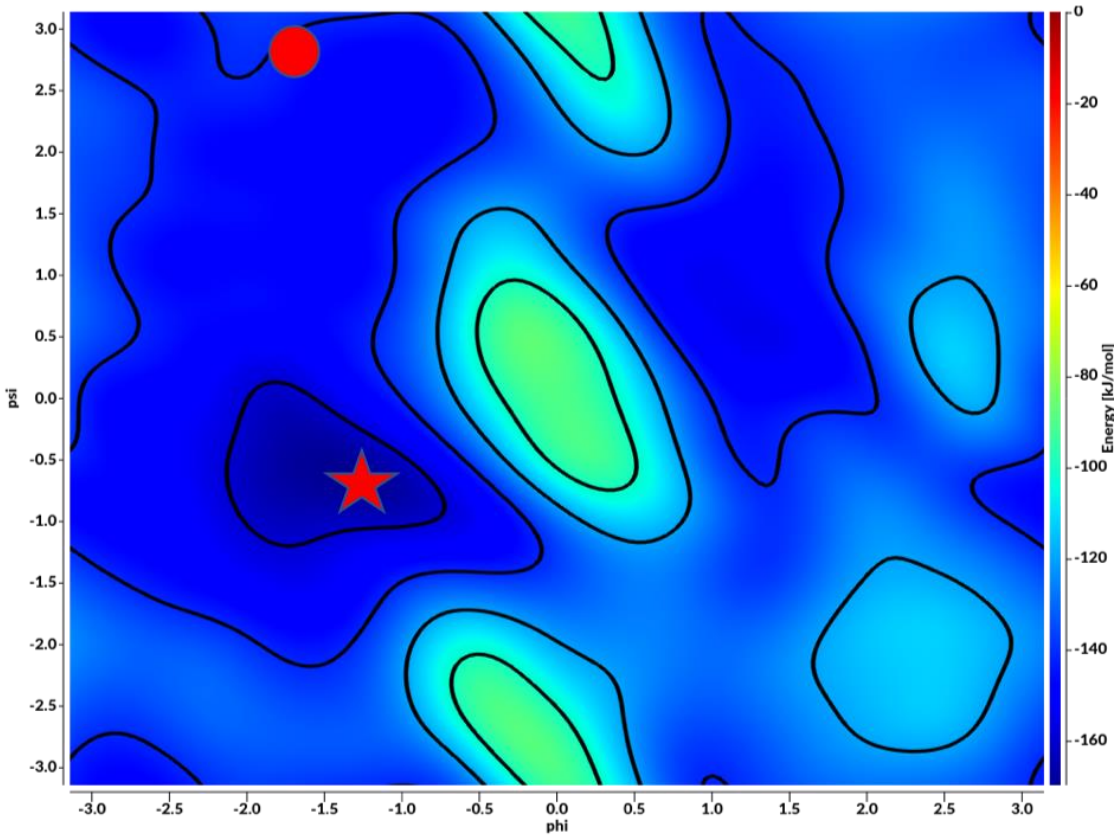


Figure 6.23: The free energy surface of K65D mutant with respect to φ and ψ angles which are shown in the x- and y- axis respectively. The colour bar is on the right-hand side of the graph and it shows that areas of dark blue colour are the ones with the lowest free energy. The star depicts the lowest free energy point of the graph. The red circle shows the φ and ψ combination of the lowest free energy structure as calculated with replica exchange simulations.

Table 6.10: The φ and ψ values of the K65D lowest free energy structure calculated with both the replica exchange and metadynamics simulations. The angles are shown in radians and in the parenthesis in degrees.

K65E	**φ / rad / (degrees)**	**ψ / rad / (degrees)**
Replica exchange	-1.9 (-110^0)	2.8 (160^0)
Metadynamics	-1.29 (-74^0)	-0.68 (-39^0)

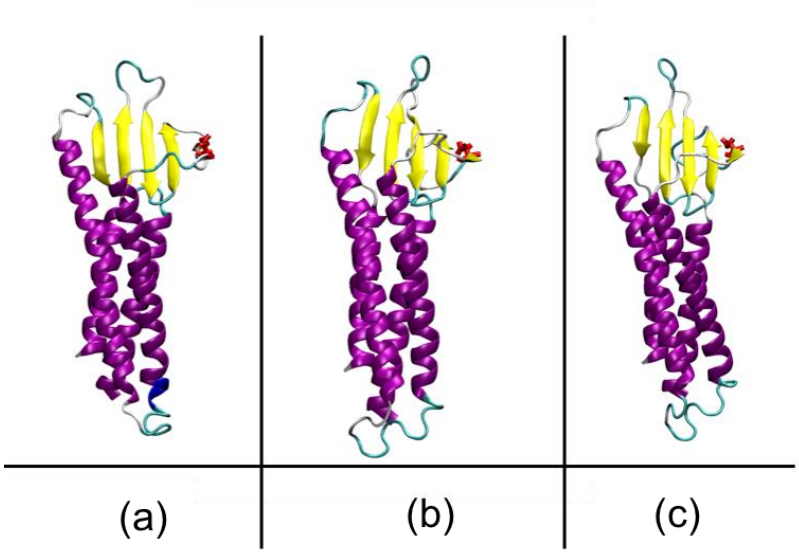


Figure 6.24: The lowest energy structures of the K65D mutant according to Metadyn View and the closest hills file. The time stamps for the extracted frames were at (a) t=1711 ps, (b) t=65997 ps and (c) t=93157 ps. The protein was rendered with VMD and shown in cartoon representation while the mutated residue is coloured red and shown in bond representation.

Lastly, the D68S mutant had two distinct low energy states presented in Figure 6.25 with the associated energies having a difference of less than 4 KJ mol^{-1}. Specifically, the point depicted by the red star had a bias potential equal to -177.05 KJ mol^{-1} at φ=-1.06 rad and ψ=2.64 rad with the closest hills at 88579 ps, 178455 ps and 188303 ps, while the point depicted by the yellow star had a bias potential equal to -174.06 KJ mol-

[1] at φ=-1.15 rad and ψ= -0.529 rad with the closest hills at 16581 ps, 103124 ps and 121353 ps. For simplicity, we name the states depicted by the red star as Point A and the ones from the yellow star as Point B (see Figure 6.25). We further isolated the protein structures that correspond to the close hills files (time stamps) for both points and present them in Figure 6.26 and Figure 6.27 respectively.

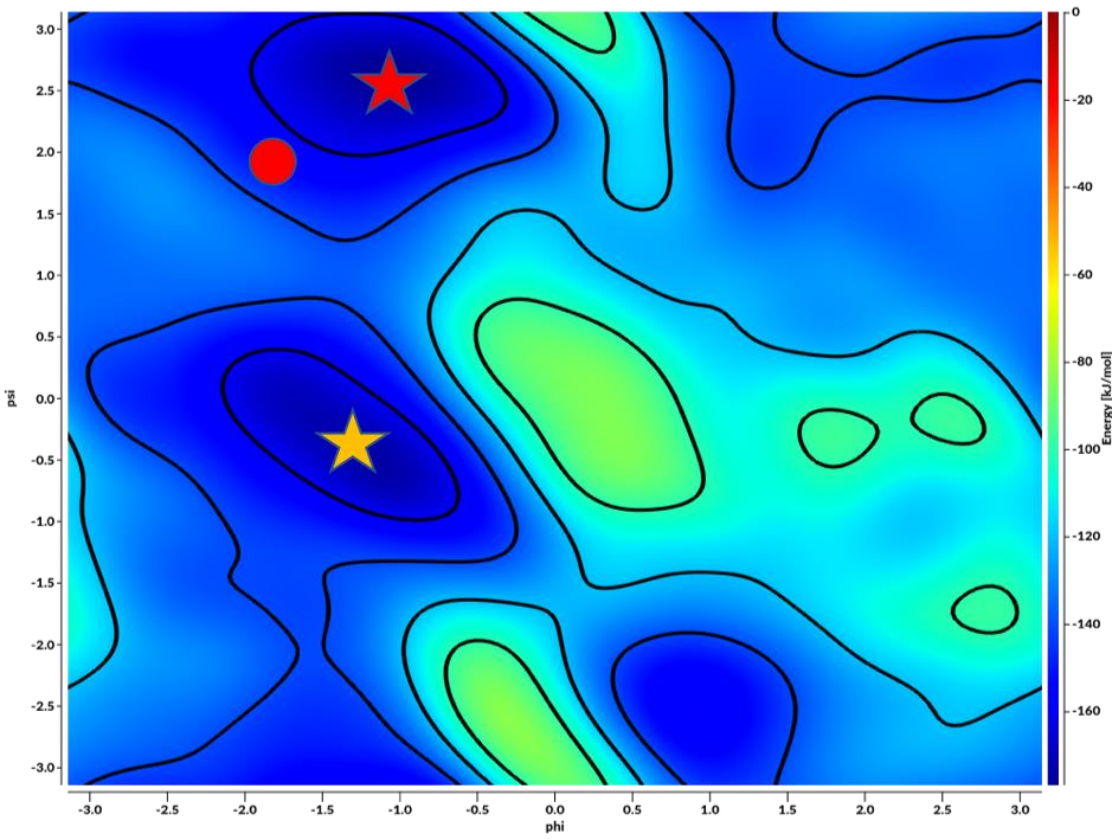


Figure 6.25: The free energy surface of D68S mutant with respect to φ and ψ angles (shown in the x- and y- axis respectively). The colour bar is on the right-hand side of the graph and it shows that areas of dark blue colour are the ones with the lowest free energy. The stars depict the lowest free energy points of the graph with the red showing the global lowest free energy conformation (Point A). The red circle depicts the angles of the structure calculated with replica exchange.

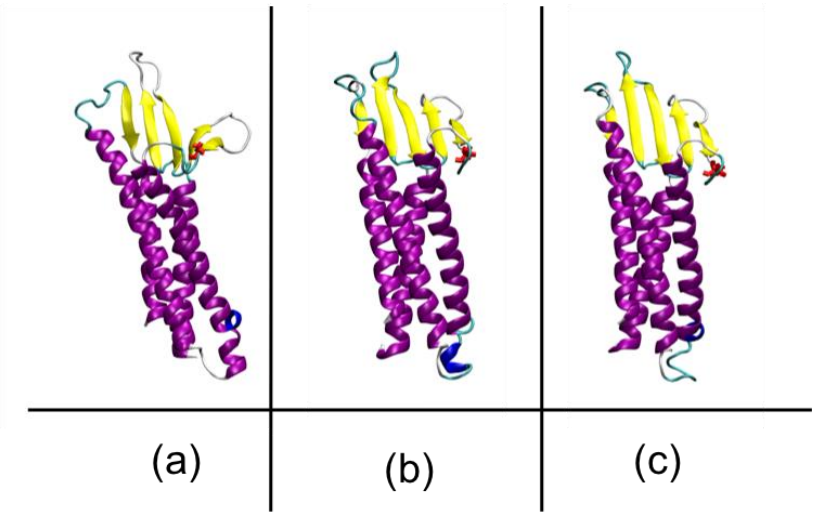


Figure 6.26: The lowest free energy structures of the D68S mutant according to Metadyn View and the closest hills file for point A. The time stamps for the extracted frames were at (a) t=88579 ps, (b) t=178455 ps and (c) t=188303 ps. The protein was rendered with VMD and shown in cartoon representation while the mutated residue is coloured red and shown in bond representation.

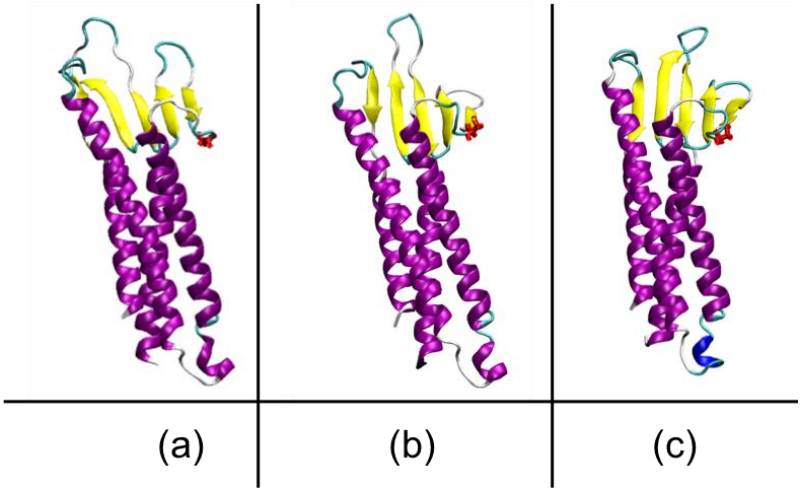


Figure 6.27: The lowest energy structures of the D68S mutant according to Metadyn View and the closest hills file for point B. The time stamps for the extracted frames were at (a) t=16581 ps, (b) t=103124 ps and (c) t=121353 ps, respectively. The protein was rendered with VMD and shown in cartoon representation while the mutated residue is coloured red and shown in bond representation.

The isolated structures show various differences that are mostly located in the orientation of the β3 and β4 strand and also in the distance between β1 and β5 strand (in the case of the structures isolated from point B) as well as in the formed small kink

of the helix at the ECL2. It is also evident that the small unstructured loops connecting the β-strands also adopt different conformations. The unstructured area of the first loop appears higher than its original position relevant to the surface of the lipid bilayer in all structures which also agrees with the result from the replica exchange simulations. Note that both K65D and D68S display this behaviour. Also, the orientation of the β4 strand is different. Finally, the calculated φ and ψ angles from replica exchange and metadynamics are different as presented in Table 6.11, but appear to be in the same low free energy region (see Figure 6.25).

Table 6.11: The φ and ψ values of the D68S lowest free energy structures calculated with both replica exchange and metadynamics. The angles are shown in radians and in the parenthesis in degrees.

D68S	**φ / rad / (degrees)**	**ψ / rad / (degrees)**
Replica exchange	-1.7 (-96^0)	1.9 (107^0)
Metadynamics (point A)	-1.1 (-63^0)	2.6 (149^0)
Metadynamics (point B)	-1.2 (-69^0)	-0.5 (-29^0)

To conclude, the construction of the FES for all examined mutants offers invaluable information with regards to the native state of the protein because it shows the lowest free energy structures, hence, the ones with the higher probability to occur. Also, we can isolate other 'metastable' states that could provide additional information about the function of the protein. Clearly the metadynamics offer an advantage compared to the more time-consuming replica exchange simulations. Replica exchange and metadynamics contrary to our expectations did not always agree on the estimated φ and ψ angles although many similarities with respect to the favourable structures were found and the former one also depicted structures from (local) low free energy basins.

6.4 General Comments and Conclusions

In this chapter, we tried to reveal and characterise the native state of claudin-1 and its mutants that were found in experimental studies to be either crucial or redundant for the protein's function. We did this by employing two different enhanced sampling techniques, namely replica exchange and metadynamics simulations. Although there are limitations in both methods, they surely offer the clear advantage of sampling more conformations compared to employing brute force MD simulations of the same duration. The native state of a protein is believed to be at the global free energy minimum of the FES but note that other selected metastable states could play a role in the biological function of the protein.

Replica exchange simulations revealed that the examined mutants displayed a small deviation from their starting structure during the lowest temperature replica trajectory, and when we focused on ECL1, we found that the radius of gyration was increased which means that this area slightly 'opened up' and occupied more space. The relative orientation of the individual β-strands was also characterised in terms of both tilt and 'opening' angles. Although the differences were minor, they revealed some key changes in the orientation of β5/β1 strands (S53E) and β3/β4 strands (K65D) providing hints about the different effects of these mutations. Additionally, the superimposed low free energy structures of the mutants projected on the wild-type claudin-1, clearly depicted the local structural differences induced in the mutants that could be matched with the experimental findings.

Contrary to our expectations there were differences in the predicted structures between the two methods, especially with regards to the φ and ψ dihedral angles. Nevertheless, we did find some similarities between them and as emphasised previously both methods depicted structures from low free energy basins. For example, metadynamics showed a large change in the extended β-sheet conformation of the E48K mutant, unlike the replica exchange method. The latter method although calculated a relatively large RMSD value for the replica at the lowest temperature, there were no obvious changes in the β-sheet domain (in terms of loss of regular secondary structures elements). Metadynamics demonstrated that E48K displayed an increased distance in the centre of the β-sheet (β2-β3 strands) and it has been suggested that this area is near the centre of the pore structure. Indeed the difference in the structure observed here, could potentially

rationalise the increased permeability of ions through the pore seen in the experimental study. Considering the S53E mutant, replica exchange suggested a large deviation in the structure and specifically, the structure 'opened up' at the edge of the β-sheet domain. This finding provides evidence of the importance of this site for ion permeation, since the mutant was found in experiment to induce charge unselective pores in the TJs. Metadynamics did demonstrate a large distance between the strands in one extracted frame structure, but overall we did not observe the large deviation in the conformation of the ECD particle seen with replica exchange.

The K65D mutant changed significantly the position of almost half of the β-sheet domain (β3 and β4 became more horizontal) in both methods. Also, the unstructured area between the β4 strand and the second transmembrane helix, which is considered to be important for the claudin-1 interactions, was also affected. This result is important because the claudin-2 mimicking substitution K65D of claudin-1 possibly blocks the trans-interaction and impedes the formation of the TJ strands (Piontek et al., 2017). The K65D demonstrated large changes in the structure of the ECD particle -the head of the 'Olympic torch'- that have been connected with the two putative interfaces of the so-called Suzuki model. Moreover, both enhanced sampling methods agreed in the aforementioned result. This provides some confidence that this region of the protein is critical for the claudin-1 interactions. Similarly, the mutation at position 68 affected the same area of the ECD particle as K65D. The experimental studies also suggested that the D68S could potentially affect the sealing of the paracellular pore a property that might be related to the difference in orientation of the unstructured area of the fist loop, as indicated by both methods. This area is a key region for the linear interface found in Chapter 3 (Dimer D) and is believed to be involved in the trans-interaction too (Chapter 4 and Suzuki et al., model).

It should be mentioned that Veshnyakova et al., suggested the E48K and S53E substitutions likely affect only the electrostatic property and not the structure of the mutants (Veshnyakova et al., 2012) and that Irudayanathan et al., verified that the secondary and tertiary structure of claudin-1 were unaffected (Irudayanathan et al., 2017). However, how exactly they did this is not mentioned in the study.

Here, the substitution of the amino acids with ones of opposite charge, clearly affected the structure at the cases of position 65 and 68 and is expected to alter the electrostatic

microenvironment of the pore in the case of the E48K mutant (the charged side chains are in the middle of the ECD particle). Also, substituting the polar serine with glutamic acid (S53E) affected the coherence of the β-sheet since this position is roughly in the centre of the β-sheet and 'opened up' the structure possibly inducing a pore in the barrier.

The novel knowledge offered in this chapter needs further validation with experimental studies, preferably validation with crystal structures of the mutants. But it is important that the methods agreed in specific results that could potentially rationalise the properties of claudin-1 and its mutants as indicated by experiments. There are limitations when employing enhanced sampling techniques, especially when assessing the convergence and success of the simulations since there are many quantitative and qualitative criteria. But here, we monitored and reported on the important factors that could characterise the convergence in both methods.

Future studies could focus on clustering the structures from the low free energy basins that have similar φ and ψ values, and examine their similarities and differences considering the overall structure. This would lead to clusters of structures with characteristic folding and their corresponding occurrence during the simulations, rather than characterising a single lowest free energy conformation.

The ability of 'key' residues from the first loop to either seal the TJ pore or form channels has important biomedical significance, since the TJs regulate the permeation of molecules through the intercellular space establishing the physiological function of our organism.

7 Concluding Remarks and Future Work

7.1 Thesis Conclusions

The molecular architecture of the TJs is largely unknown and current experimental techniques cannot elucidate alone the protein-protein interactions that lead to the formation of the TJ barrier/channel structure. Claudins, the major proteins in the TJs, are expressed in various epithelial and endothelial tissues and show structural variation which determines their function (Günzel and Alan, 2013, Daugherty et al., 2007). Claudin assembly results in the formation of polymeric strands or fibrils surrounding the cell, which regulate the molecular transport through the paracellular route. The strands can be a product of homo- and heteromeric interaction between claudins, the mosaic of which determine the characteristics of the TJ strand (Zihni et al., 2016). Additionally, other integral proteins such as occludin also incorporate into the strands. The strands determine the nature of the paracellular pathway which may function either as a barrier or as channels that are selective for cations, anions, or water, the charge selectivity being influenced by specific amino acids with charged side chains located at the ECD particle of claudins (Colegio et al., 2002, Colegio et al., 2003, Van Itallie and Anderson, 2004).

The research questions addressed in this thesis were: How do claudins form contiguous strands? Do claudins have preferred cis-cis interfaces? What are the primary determinants of their aggregation and function? How is the barrier between the cells formed? How does the variability in the lipid composition of the membrane affect the aggregation of claudins (with a view to mimicking the skin lipids)? How do specific single point mutations affect the secondary and tertiary structure of claudins? The mutants were found in experiments to affect the structure and function of TJs. We focused on the behaviour of claudin-1, a predominantly barrier forming claudin that is required for the normal barrier function of epidermis.

Towards addressing these objectives, four distinct studies were carried out using molecular simulation as it offers the atomistic resolution needed to access the spatio-temporal domain of large bio-molecular complexes.

The first study (Chapter 3) investigated the side-by-side (cis-)interaction between claudins. To maintain the near-atomic resolution needed for this study, we employed the concept of an *implied* bilayer and probed the interaction of claudin-1 monomers within a membrane (cis-) and in the next study (Chapter4) between (trans-) neighbouring membranes. By simulating only the ECD particle, which is larger in diameter and laterally overhangs much of the transmembrane domain, the length and timescale issues of atomistic-resolution MD simulations were overcome in a creative way. The self-assembly simulations comprised a large population of claudin's extracellular domains solvated at a physiological salt concentration, which challenges computational resources. The domain particles were free to interact on the plane of the implied bilayer (x-y plane). These studies demonstrate that the question of claudin self-assembly can be successfully addressed at atomistic resolution by using the implied membrane approach. Remarkably, the emergent branched, network strand structure, resembled electron micrographs of TJs and revealed a plethora of formed cis-cis interactions. The identified dimers were characterised by differing relative orientation angles between the claudin protomers with some dominant cis-interfaces. The plethora of formed orientation angles explains the branching of the strands that do not show any unique branching angle. Note that the statistics were poor as we have averaged over only the last configuration of the trajectory. A main conclusion was that the large number of cis-interfaces determined, support the cross-linked morphology of the strand network. The TJ strands have diverse curvatures and different branching angles. It is also noteworthy that the cis-interaction occurred in the absence of trans-interacting partners from neighbouring cells. Networks of similar morphology have been reported earlier by another group that studied the organisation of TJs at the blood-brain barrier (BBB) and the self-assembly of classic claudins using molecular simulation (Irudayanathan et al., 2015, Irudayanathan et al., 2017, Irudayanathan et al., 2018).

The self-assembly simulations reveal that there is little scope for optimisation or dynamic re-arrangement of the emergent strands because the particles become kinetically locked. This suggests strong binding interactions. To confirm this, we

carried out potential of mean force calculations for selected dimer interactions using umbrella sampling. The free energies of binding indicate strong interactions suggesting that the biological TJ structures are out of equilibrium, being kinetically locked. The calculated energies were estimated to be in the range 8-22 k_BT at 310 K. The significance of the binding free energies being at least an order of magnitude greater than k_BT (the thermal energy) is that once the claudins come together they are unlikely to unbind again. The implication is that the formed claudin strands are not in a thermodynamic equilibrium state. It appears that nature favours strong interactions between the component proteins in TJs, probably because TJs based on a claudin-claudin binding free energy of a few k_BT would likely fail to perform their fundamental function, which is to form a seal between cells. Regardless if the claudins form channels, their primary role is to seal the gap and then regulate the diffusion of substances between cells, therefore all claudins are barrier forming.

The relative stability of the cis-interacting dimers revealed that some interfaces are stronger than others, specifically the Dimer D (similar to the linear interface) and the face-to-face interface from the putative Suzuki et al., model (Suzuki et al., 2015). This provides support in the above mentioned model for the claudin-based paracellular channels. Further, due to the atomistic resolution, we were able to identify residues acting as 'hot spots' that were consistent with reported experimental mutation studies. The cis-cis interactions were mostly regulated by the unstructured area of the first extracellular loop, the β4 strand and the back of the β-sheet domain. Figure 7.1 shows the above mentioned regions of claudin-1 that dominated the cis-cis interfaces. The identified critical residues based on the formed number of contact with others, were phenylalanines (67, 148 & 161), valines (145 & 155), leucines (70 & 73), tyrosines (35, 149 & 159), lysine (65), arginine (158) and aspartic acid (150). The atomistic resolution enabled us to characterise in detail the particular residues that interacted strongly, and hence, played a key role in stabilising the dimers. Additionally, the MM/PBSA method revealed the residues characterised by favourable contributions to the overall binding energy. Important residues include glutamine (29), arginine (31), lysine (65), threonine (137) and arginines (143 & 158). The fact that the back of the β-sheet domain is a broad region giving rise to redundant interaction sites (wide cis-interaction pocket) might offer the claudin-based strands the flexibility they need to overcome mechanical stress and stress resulting from osmotic pressure imbalances.

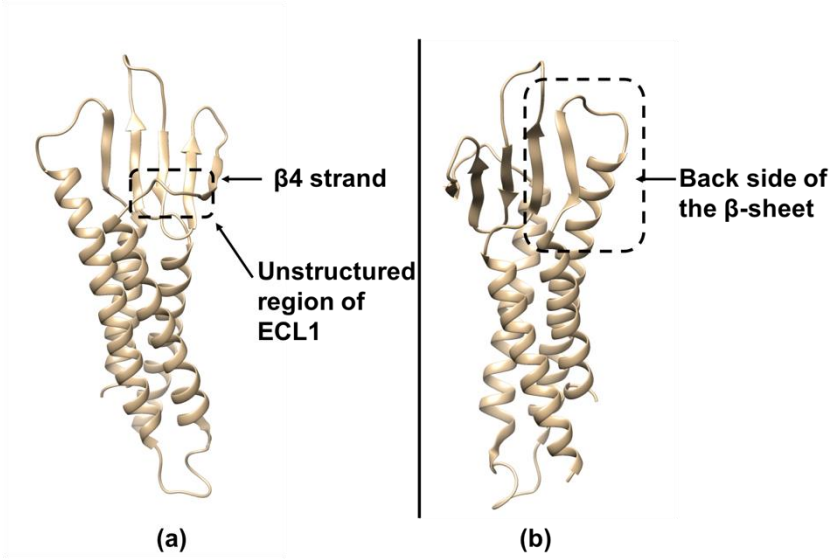


Figure 7.1: The 'key' regions of claudin-1 that regulate the side-by-side interactions.

The so-called Suzuki model, which offers a plausible description of the pore structure, has attracted attention and in silico modelling studies have assessed its accuracy (Alberini G, 2017, Samanta et al., 2018, Alberini et al., 2018, Zhao et al., 2018). The limitation of these studies however was that they examined the accuracy of the model on the basis of its stability in a MD simulation over a relatively short timescale (260 ns for the single pore structure and 35 ns for the double pore structure). In Chapter 4, we examined instead the trans-interaction by way of self-assembly using two bilayers, i.e. without biasing the starting configuration. We again employed the implied bilayer concept in a self-assembly study and in a related study involving systematic relative rotation of two claudin ECDs interacting along the trans-axis to characterise the trans-interaction. With hindsight, the results of the former study are somewhat limited as the restraint forces that keep the ECD particles on their implied membrane were rather strong, limiting the penetration depth of the ECD particles from one bilayer to the other. Having said this, the claudin particles did interact both in and between the implied bilayer planes, though the maximum penetration depth was not quite optimal. The specific regions of the protein that dominated the observed trans-interaction were the small unstructured loops connecting the β1/β2 strands (V1 region) and the β5-strand with the long helix of the second extracellular loop (V2 region). The V2 region is

suggested to affect the trans-interactions of claudin-3 and -5 (Rossa et al., 2014) and both variable regions were characterised important for the head-to-head interaction in an in silico study (Irudayanathan et al., 2017). Sequence alignment amongst members of the claudin family shows that these regions have mostly non-conserved residues. In the systematic rotation study, we found that the penetration depth increased and that some trans-interfaces were preferred. The preferred trans-dimers exhibited high binding energies, a large number of contacts, and they did not change their configuration significantly throughout the trajectory indicating stability. This study also demonstrated that some interacting monomers initially formed contacts between the variable regions but further slightly go deeper/'slide' down towards the adjacent monomer. Perhaps the 'head-to-head' interaction mechanism involves the formation of other (intermediate) binding modes to offer tolerance to the protein interaction networks (like a 'back up' mechanism). The studies reveal the critical role of methionine 152 in sealing the paracellular space and forming an efficient barrier at the TJs further supported by experiment (Nakajima et al., 2015). We also identified residues of the area close to the small helix (71-75) as well as the aromatic amino acids tyrosines 33 and 35 (from β1 strand) as 'key' residues stabilising the trans-interface. The overall picture with regards to the trans-interaction was more clear, identifying the variable regions of the ECDs as the 'hot spots' as well as the area in their proximity. The resulted structure was devoid of channels, thus the passage of small ions and molecules would be blocked, confirming the barrier properties of claudin-1.

Claudins are tetraspan membrane proteins, so their transmembrane domain might stabilise their aggregation. Therefore, we examined the membrane driven claudin interactions in Chapter 5 and the primary goal was to ascertain if the same interfaces would be observed in the protein-membrane systems. This type of assembly simulations are computationally expensive because they need to be performed over a sufficient time to obtain reliable data. The transmembrane helical bundle provides a rigid structure that stabilises the protein in the bilayer and it has mostly non-polar residues; only a single residue has a charged side chain, namely arginine 117. The hydrophobic mismatch, which is the difference between the hydrophobic depth of the protein and the thickness of the bilayer is expected to play an important role in the folding and stability of claudins. We did not expect the cis-cis interfaces to be driven by non-polar interactions in an already hydrophobic environment. However, we did observe that the non-polar

leucine residues located in the transmembrane helices, stabilised further a specific dimer which was similar to the linear Suzuki interface. It is pertinent to note that a leucine zipper has been suggested as a stabilising interaction in another computational study of the BBB claudins -3 and -5 (Irudayanathan et al., 2015). However, the metadynamics studies that revealed the former interaction site in Chapter 5 needs further investigation since the newly formed non-optimal dimers have long life-times and were not able to unbind, shuffle and re-arrange during the simulations.

The effect of different lipid compositions on the relative orientation of the protein was also investigated in Chapter 5. The lipids were chosen to mimic the composition of the specific layer of skin where TJs exist. However, considering that TJs are found in other tissues, our target was to choose representative lipids of key lipid groups relevant to the lipid membrane in the stratum granulosum. The systems comprised pure phospholipid bilayers and binary mixtures of lipids (phospholipids with cholesterol and glycolipids). The condensing and thickening effect upon inserting cholesterol in a phospholipid bilayer was confirmed, as was the increased order of the acyl chains of the phospholipids in the binary model bilayers. The relative position of the glycolipids and cholesterol was also investigated demonstrating that the examined model bilayers did not cause dramatic changes in the folding or position of the protein with regards to its tilt angle and hydrophobic depth. Claudin-1 in the phospholipid bilayer demonstrated the largest tilt angle and smaller hydrophobic depth, making the ECD particle (the head of the ‘Olympic torches’) the primary domain for aggregation.

The self-assembly simulations of the claudins embedded in the different bilayers reveal that the claudins initially form contacts through the ECDs, but subsequently interact through the helix bundle resulting in greater packing. Except in the case of the model bilayer that had glycolipids, which slowed down the diffusion of claudins and prevented them from forming tightly packed dimer configurations. The glycolipids were positioned at the interfaces between claudins. The claudins formed a linear cross-linked strand network with some trimeric interfaces that justify the cross-linking organisation. Ring-like structures were also observed. TJs appear as strands of particles (approximately 10nm in diameter) with freeze-fracture EM. The ring-like structures observed in the studies could be relevant to the particles observed experimentally. The morphology of the claudin-based strand was similar to the ones observed in the previous

studies using atomistic resolution (Chapters 3 and 4). Interestingly, even in the systems that examined the effect of varied protein concentration, we observed the same strand morphology. In the latter study, we expected to ascertain the energetically most favourable structure as the low protein concentration should give the proteins enough time to rearrange and equilibrate prior to binding. The almost identical architecture of the claudin-based strand was consistently observed in all systems studied.

The nanoscale assemblies have provided insight into claudin-claudin interactions and improved our understanding of the factors that define the strand's architecture and shape. However, the time scales used are too short to capture the dynamics of formation of such big biological complexes with MD; even when using simplified representations, i.e. coarse-grained models. Nevertheless, the finely tuned interactions amongst claudin-1 are now better characterised and the new models can be used to test new hypothesis and describe functional sites.

Lastly, the effects of specific mutations on the structure and function of claudin-1 were examined in Chapter 6. The mutants were found in experiments to abolish the formation of the strands or induce pores in the barrier structure. Some of the examined mutants were considered redundant for the function of TJs. The study attempted to identify the lowest free energy structures for each of the mutants using advanced sampling methods (replica exchange and metadynamics) and compared these structures with the wild-type claudin-1. The replica exchange simulations perhaps were not fully converged, they probably needed to span a larger temperature range to overcome all the relevant energy barriers. Metadynamics is a computationally effective method that efficiently accelerates sampling. We reconstructed with metadynamics the free energy surfaces of all mutants to identify the low free energy structures. Remarkably, these structures were able to rationalise and explain the experimental observations. E48K depicted a big deviation in the structure of the β-sheet domain while S53E displayed an 'opening up' close to the centre of the β-sheet domain. Both mutants were found experimentally to induce pores in the TJs. K65D demonstrated structural changes in 'key' regions of the ECD involved in both the cis-and trans-interaction of claudin-1. This mutant abolished the formation of the TJ strand in experiment. Lastly, D68S revealed similar behaviour to K65D and it has been suggested that this mutation could potentially affect the sealing of the paracellular pore. The changes between the favourable structure of the mutants

and the wild-type claudin provided hints about the functional sites of the protein (after being matched to the experimental findings). The functional sites were in agreement with the findings from the previous chapters of the thesis. It should be emphasised that the differences were mostly on local regions and specific orientations of the secondary structure elements rather than a global deviation of the structure. Obtaining the free energy landscape of a system is not something trivial, but it compensates the crucial information gained especially when studying systems that undergo complex conformational changes.

The thesis findings have contributed in the elucidation of claudins aggregation and function, by highlighting the key players for the self-assembly and barrier structure. The results obtained here can lead to new questions - new hypothesis that will promote research in the field.

7.2 Limitations of the Studies, Outstanding Questions and Future Work

Computational modelling studies suffer from time and length scale limitations that constrain the range of problems that can be addressed, e.g. to obtain the detailed mechanisms that govern the aggregation of large membrane proteins. The energy barriers related with such processes could be too high to cross and as a consequence the system remains trapped in local regions of the free energy landscape. Moreover, these biological processes need longer time scales. MD trajectories at best sample up to about the μs timescale. Consequently, many complex bio-molecular processes can be computationally unfeasible to study.

The self-assembly simulations studies examined in the thesis have some limitations. The cis-interaction was investigated with the hypothesis that the dominant interface for aggregation was the ECD (the head of the ‘Olympic torch’) being larger than the transmembrane body. The time scales typically used in MD simulations, resulted in the particles becoming kinetically ‘trapped’ due to their strong binding energy. The heads of the ‘Olympic torches’ are surely the decisive factors that mediate the trans-interaction (between adjacent cells) since they are located in the extracellular region. However, it

might be a prerequisite for claudins to be properly folded and oligomerised (e.g. form stable dimers) before further transport along the secretory pathway. In that case, stable oligomers would then interact via their ECDs with the neighbouring cells to seal the intercellular space. Meaning that the cis-interaction might guide the trans-interaction. It is likely that claudin-claudin intermediates are formed in the late secretory pathway (Koval, 2013). We were reluctant to arbitrary choose a cis-dimer and subsequently with docking (or another similar method) to try and generate the barrier structure. Thus, we preferred to maintain the freedom of the 'Olympic torches' and explore their orientation space in Chapter 4.

The rotation study which examined through the systematic rotation of two ECDs the trans-interaction of claudins, was set up in order to explore systematically the orientation sampling. Although this is an advantage of the study, it can also be considered its limitation since it explored that by using monomers and not pre-formed cis-dimers (as previously mentioned). Another limitation of the rotation study was that there were numerous ways to translate the top particle compared to the bottom one, along the x- and y- direction. Hence, other initial configurations could be perceived.

In future studies, it is strongly recommended to use enhanced sampling methods to overcome the limits of MD simulations. But as pointed out in Chapter 5 this can be time consuming and our efforts to employ metadynamics to ascertain the cis-cis interfaces although fruitful, did not give us the confidence that the free energy profiles obtained were correct. Finding the proper collective variables is not a trivial process.

With regards to the mutation studies in Chapter 6, there were cases where metadynamics pointed out more than one low free energy structures. The isolated structures in some cases had structural differences, hence in future work clustering structures isolated from the lowest free energy basin is advised. Additionally, resolving the crystal structures of the mutants could shed light in the structural changes that happen upon mutation.

Interactions between membrane proteins are fundamental for biology and they offer a possible intervention route for the pharmaceutical industry. Defining the exact molecular organisation of the TJ strand, will enhance our understanding of how the barrier is created, regulated and altered in pathology. The development of new therapeutic strategies that alter by design the barrier properties can promote the well-

being of patients that suffer from TJs related diseases. Also, the study of small molecules that could potentially act as TJ modulators and reversibly open the barrier is an area of increased interest. In silico methods can provide invaluable information in such studies.

TJs play a central role in homeostasis since they regulate the passage of molecules through the cells, which is related to inflammation, oedema and blood-borne metastasis (Sawada, 2013). Moreover, their fence function maintains cell polarity and therefore is involved in cancer cell biology. Their involvement in cancer can be investigated considering two aspects, first the changes that they undergo as a cell-apparatus and as individual components, e.g. claudins. The involvement of TJs in cancer is an area of increased interest nowadays and future work should continue to focus on this.

Simplifying a model in order to form specific hypothesis and try to tackle the problem is a common method. In this thesis, we addressed fundamental questions regarding the TJs. However, there are more that need to be addressed, for example: Which is the structural mechanism that enables charge and size selective diffusion through the junctions? Do specific areas affect the type of pores formed, for example by changing the orientation of the ECD particle, which is the regulator of the trans-interaction? Is there a motif for the heterotypic binding between different claudins? Future studies in the field could examine the systematic mutation of important residues of the identified critical regions for the assembly of claudin-1 and study their effect on the integrity and functionality of TJs. Additionally, the efforts to resolve more crystal structures of either sole claudins or claudins in complex with the Clostridium perfringens enterotoxin should continue, because they provide possible mechanisms for the assembly and disassembly of claudin interactions respectively. Small variations in the resolved crystal structures might influence the functionality of the complex and establish the decisive determinants that enable their vital functions.

Many of these molecular-level results need to be verified by future experiments as they have physiological and biomedical significance. For example, the stability of the dimers could be assessed and compared to other protein oligomers from other cell-cell junctions to establish the strength of interactions. Some atomic force spectroscopy data exist, but more work is needed to understand for example how TJs respond to mechanical forces. The relationship between different cell-cell junctions should also be

investigated as there might be similarities between them. For example, do claudins form trans-intermediate distinct conformations as cadherins do? Do claudins form oligomers before they interact with the adjacent membrane as connexins do? How does external stimuli affect these cell-cell structures? These are subjects of on-going studies that will further enhance our understanding of this remarkable and complex structure.

8 References

ABASCAL, F. & ZARDOYA, R. 2013. Evolutionary analyses of gap junction protein families. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1828**,** 4-14.

ABRAHAM, M. J., MURTOLA, T., SCHULZ, R., PÁLL, S., SMITH, J. C., HESS, B. & LINDAHL, E. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX,* 1**,** 19-25.

AFFENTRANGER, R., TAVERNELLI, I. & DI IORIO, E. E. 2006. A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. *Journal of Chemical Theory and Computation,* 2**,** 217-228.

ALAN, S. 2010. Structure–Function Studies of the Claudin Pore. *Current Topics in Membranes,* 65**,** 79-95.

ALBERINI, G., BENFENATI, F. & MARAGLIANO, L. 2017. A refined model of claudin-15 tight junction paracellular architecture by molecular dynamics simulations. *PloS one,* 12**,** e0184190.

ALBERINI, G., BENFENATI, F. & MARAGLIANO, L. 2018. Molecular Dynamics Simulations of Ion Selectivity in a Claudin-15 Paracellular Channel. *The Journal of Physical Chemistry B,* 122**,** 10783-10792.

ALBERINI G, B. F., MARAGLIANO L 2017. A refined model of claudin-15 tight junction paracellular architecture by molecular dynamics simulations. *PLoS ONE* 12.

ALBERTS, B., JOHNSON, A., LEWIS, J., MORGAN, D., RAFF, M., ROBERTS, K. & WALTER, P. 2015. *Molecular biology of the cell* Garland Science Taylor & Francis Group.

ALLEN, M. P. 2004. Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins,* 23**,** 1-28.

ALLEN, M. P. & TILDESLEY, D. J. 1989. *Computer simulation of liquids*, Oxford university press.

ALLEN, W. J., LEMKUL, J. A. & BEVAN, D. R. 2009. GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics. *Journal of computational chemistry,* 30**,** 1952-1958.

ALWARAWRAH, M., DAI, J. & HUANG, J. 2010. A molecular view of the cholesterol condensing effect in DOPC lipid bilayers. *The journal of physical chemistry B,* 114**,** 7516-7523.

ANDERSON, J. M. 2001. Molecular structure of tight junctions and their role in epithelial transport. *Physiology,* 16**,** 126-130.

ANDERSON, J. M. & VAN ITALLIE, C. M. 2009. Physiology and function of the tight junction. *Cold Spring Harbor perspectives in biology,* 1**,** a002584.

ANGELOW, S. & ALAN, S. 2009a. Cysteine Mutagenesis to Study the Structure of Claudin-2 Paracellular Pores. *Annals of the New York Academy of Sciences,* 1165**,** 143-147.

ANGELOW, S. & ALAN, S. 2009b. Structure-function studies of claudin extracellular domains by cysteine-scanning mutagenesis. *Journal of Biological Chemistry,* 284**,** 29205-29217.

BAASE, W. A., LIU, L., TRONRUD, D. E. & MATTHEWS, B. W. 2010. Lessons from the lysozyme of phage T4. *Protein Science,* 19**,** 631-641.

BARDUCCI, A., BONOMI, M. & PARRINELLO, M. 2011. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science,* 1**,** 826-843.

BÄSLER, K., BERGMANN, S., HEISIG, M., NAEGEL, A., ZORN-KRUPPA, M. & BRANDNER, J. M. 2016. The role of tight junctions in skin barrier function and dermal absorption. *Journal of Controlled Release*.

BENNETT, W. D. & TIELEMAN, D. P. 2013. Computer simulations of lipid membrane domains. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1828**,** 1765-1776.

BENSOUILAH 2012. Skin structure and function

BERENDSEN, H. J., POSTMA, J. V., VAN GUNSTEREN, W. F., DINOLA, A. & HAAK, J. 1984. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics,* 81**,** 3684-3690.

BERGER, O., EDHOLM, O. & JÄHNIG, F. 1997. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical journal,* 72**,** 2002-2013.

BERKA, K., HANÁK, O., SEHNAL, D., BANÁŠ, P., NAVRATILOVA, V., JAISWAL, D., IONESCU, C.-M., SVOBODOVÁ VAŘEKOVÁ, R., KOČA, J. & OTYEPKA, M. 2012. MOLE online 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic acids research,* 40**,** W222-W227.

BERKOWITZ, M. L. 2009. Detailed molecular dynamics simulations of model biological membranes containing cholesterol. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1788**,** 86-96.

BERNARDI, R. C., MELO, M. C. & SCHULTEN, K. 2015. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA)-General Subjects,* 1850**,** 872-877.

BEST, R. B., HOFMANN, H., NETTELS, D. & SCHULER, B. 2015. Quantitative interpretation of FRET experiments via molecular simulation: force field and validation. *Biophysical journal,* 108**,** 2721-2731.

BEST, R. B., ZHU, X., SHIM, J., LOPES, P. E., MITTAL, J., FEIG, M. & MACKERELL JR, A. D. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. *Journal of chemical theory and computation,* 8**,** 3257-3273.

BJELKMAR, P., LARSSON, P., CUENDET, M. A., HESS, B. & LINDAHL, E. 2010. Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *Journal of Chemical Theory and Computation,* 6**,** 459-466.

BONOMI, M., BRANDUARDI, D., BUSSI, G., CAMILLONI, C., PROVASI, D., RAITERI, P., DONADIO, D., MARINELLI, F., PIETRUCCI, F. & BROGLIA, R. A. 2009. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications,* 180**,** 1961-1972.

BRANDNER, J., ZORN-KRUPPA, M., YOSHIDA, T., MOLL, I., BECK, L. & DE BENEDETTO, A. 2015. Epidermal tight junctions in health and disease. *Tissue barriers,* 3**,** e974451.

BRANDNER, J. M. 2009. Tight junctions and tight junction proteins in mammalian epidermis. *European Journal of Pharmaceutics and Biopharmaceutics,* 72**,** 289-294.

BRANDNER, J. M., KIEF, S., GRUND, C., RENDL, M., HOUDEK, P., KUHN, C., TSCHACHLER, E., FRANKE, W. W. & MOLL, I. 2002. Organization and formation of the tight junction system in human epidermis and cultured keratinocytes. *European journal of cell biology,* 81**,** 253-263.

BUSSI, G., GERVASIO, F. L., LAIO, A. & PARRINELLO, M. 2006. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *Journal of the American Chemical Society,* 128**,** 13435-13441.

CAPALDO, C. T. & NUSRAT, A. Claudin switching: Physiological plasticity of the Tight Junction. Seminars in cell & developmental biology, 2015. Elsevier, 22-29.

CEREIJIDO, M. & ANDERSON, J. M. 2001. *Tight junctions*, Crc Press.

CHIBA, H., OSANAI, M., MURATA, M., KOJIMA, T. & SAWADA, N. 2008. Transmembrane proteins of tight junctions. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1778**,** 588-600.

CLAUDE, P. 1978. Morphological factors influencing transepithelial permeability: A model for the resistance of theZonula Occludens. *The Journal of membrane biology,* 39**,** 219-232.

COLEGIO, O. R., VAN ITALLIE, C., RAHNER, C. & ANDERSON, J. M. 2003. Claudin extracellular domains determine paracellular charge selectivity and resistance but not tight junction fibril architecture. *American Journal of Physiology-Cell Physiology*.

COLEGIO, O. R., VAN ITALLIE, C. M., MCCREA, H. J., RAHNER, C. & ANDERSON, J. M. 2002. Claudins create charge-selective channels in the paracellular pathway between epithelial cells. *American Journal of Physiology-Cell Physiology,* 283**,** C142-C147.

CONSORTIUM, U. 2018. UniProt: a worldwide hub of protein knowledge. *Nucleic acids research,* 47**,** D506-D515.

CORDING, J., ARSLAN, B., STAAT, C., DITHMER, S., KRUG, S. M., KRÜGER, A., BERNDT, P., GÜNTHER, R., WINKLER, L. & BLASIG, I. E. 2017. Trictide, a tricellulin-derived peptide to overcome cellular barriers. *Annals of the New York Academy of Sciences,* 1405**,** 89-101.

CRAMER, C. J. 2013. *Essentials of computational chemistry: theories and models*, John Wiley & Sons.

DA SILVA, P. P. & KACHAR, B. 1982. On tight-junction structure. *Cell,* 28**,** 441-450.

DAHL, A. C. E., CHAVENT, M. & SANSOM, M. S. 2012. Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics,* 28**,** 2193-2194.

DAUGHERTY, B. L., WARD, C., SMITH, T., RITZENTHALER, J. D. & KOVAL, M. 2007. Regulation of heterotypic claudin compatibility. *Journal of Biological Chemistry,* 282**,** 30005-30013.

DAVIS, C., HARRIS, H. J., HU, K., DRUMMER, H. E., MCKEATING, J. A., MULLINS, J. G. & BALFE, P. 2012. In silico directed mutagenesis identifies the CD 81/claudin-1 hepatitis C virus receptor interface. *Cellular microbiology,* 14**,** 1892-1903.

DELANO, W. L. 2002. PyMOL.

DELI, M. A. 2009. Potential use of tight junction modulators to reversibly open membranous barriers and improve drug delivery. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1788**,** 892-910.

EARL, D. J. & DEEM, M. W. 2005. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics,* 7**,** 3910-3916.

ELIAS, P. M., BROWN, B. E., FRITSCH, P., GOERKE, J., GRAY, G. M. & WHITE, R. J. 1979. Localization and composition of lipids in neonatal mouse stratum granulosum and stratum corneum. *Journal of Investigative Dermatology,* 73**,** 339-348.

ERIKSSON, A. E., BAASE, W. A., ZHANG, X.-J., HEINZ, D. W., BLABER, M., BALDWIN, E. P. & MATTHEWS, B. W. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science,* 255**,** 178-183.

FARQUHAR, M. G. & PALADE, G. E. 1963. Junctional complexes in various epithelia. *The Journal of cell biology,* 17**,** 375-412.

FEINGOLD, K. R. 2007. Thematic review series: skin lipids. The role of epidermal lipids in cutaneous permeability barrier homeostasis. *Journal of lipid research,* 48**,** 2531-2546.

FEINGOLD, K. R. & ELIAS, P. M. 2014. Role of lipids in the formation and maintenance of the cutaneous permeability barrier. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids,* 1841**,** 280-294.

FOLDVARI, M. 2000. Non-invasive administration of drugs through the skin: challenges in delivery system design. *Pharmaceutical science & technology today,* 3**,** 417-425.

FÖRSTER, C. 2008. Tight junctions and the modulation of barrier function in disease. *Histochemistry and cell biology,* 130**,** 55-70.

FUKASAWA, M., NAGASE, S., SHIRASAGO, Y., IIDA, M., YAMASHITA, M., ENDO, K., YAGI, K., SUZUKI, T., WAKITA, T. & HANADA, K. 2015. Monoclonal antibodies against extracellular domains of claudin-1 block hepatitis C virus infection in a mouse model. *Journal of virology,* 89**,** 4866-4879.

FURUSE, M. 2010a. Introduction: claudins, tight junctions, and the paracellular barrier. *Current Topics in Membranes,* 65**,** 1-19.

FURUSE, M. 2010b. Molecular basis of the core structure of tight junctions. *Cold Spring Harbor perspectives in biology,* 2**,** a002907.

FURUSE, M., HATA, M., FURUSE, K., YOSHIDA, Y., HARATAKE, A., SUGITANI, Y., NODA, T., KUBO, A. & TSUKITA, S. 2002. Claudin-based tight junctions are crucial for the mammalian epidermal barrier a lesson from claudin-1–deficient mice. *The Journal of cell biology,* 156**,** 1099-1111.

FURUSE, M., HIRASE, T., ITOH, M., NAGAFUCHI, A., YONEMURA, S. & TSUKITA, S. 1993. Occludin: a novel integral membrane protein localizing at tight junctions. *The Journal of cell biology,* 123**,** 1777-1788.

FURUSE, M., SASAKI, H., FUJIMOTO, K. & TSUKITA, S. 1998. A single gene product, claudin-1 or-2, reconstitutes tight junction strands and recruits occludin in fibroblasts. *The Journal of cell biology,* 143**,** 391-401.

FURUSE, M., SASAKI, H. & TSUKITA, S. 1999. Manner of interaction of heterogeneous claudin species within and between tight junction strands. *The Journal of cell biology,* 147**,** 891-903.

GARROD, D. & CHIDGEY, M. 2008. Desmosome structure, composition and function. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1778**,** 572-587.

GENHEDEN, S. & RYDE, U. 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery,* 10**,** 449-461.

GIEPMANS, B. N. & VAN IJZENDOORN, S. C. 2009. Epithelial cell–cell junctions and plasma membrane domains. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1788**,** 820-831.

GIMONDI, I., TRIBELLO, G. A. & SALVALAGLIO, M. 2018. Building maps in collective variable space. *The Journal of chemical physics,* 149**,** 104104.

GONG, Y., RENIGUNTA, V., ZHOU, Y., SUNQ, A., WANG, J., YANG, J., RENIGUNTA, A., BAKER, L. A. & HOU, J. 2015. Biochemical and biophysical analyses of tight junction permeability made of claudin-16 and claudin-19 dimerization. *Molecular biology of the cell,* 26**,** 4333-4346.

GONSCHIOR, H., HAUCKE, V. & LEHMANN, M. 2020. Super-Resolution Imaging of Tight and Adherens Junctions: Challenges and Open Questions. *International Journal of Molecular Sciences,* 21**,** 744.

GONZALEZ-MARISCAL, L., NAVA, P. & HERNANDEZ, S. 2005. Critical role of tight junctions in drug delivery across epithelial and endothelial cell layers. *The Journal of membrane biology,* 207**,** 55-68.

GOODENOUGH, D. A. & PAUL, D. L. 2009. Gap junctions. *Cold Spring Harbor perspectives in biology,* 1**,** a002576.

GOWERS, R. J., LINKE, M., BARNOUD, J., REDDY, T. J., MELO, M. N., SEYLER, S. L., DOTSON, D. L., DOMANSKI, J., BUCHOUX, S. & KENNEY, I. M. MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. Proceedings of the 15th Python in Science Conference, 2016. 98-105.

GREBENKÄMPER, K. & GALLA, H.-J. 1994. Translational diffusion measurements of a fluorescent phospholipid between MDCK-I cells support the lipid model of the tight junctions. *Chemistry and physics of lipids,* 71**,** 133-143.

GÜNZEL, D. 2017. Claudins: vital partners in transcellular and paracellular transport coupling. *Pflügers Archiv-European Journal of Physiology,* 469**,** 35-44.

GÜNZEL, D. & ALAN, S. 2013. Claudins and the modulation of tight junction permeability. *Physiological reviews,* 93**,** 525-569.

GUPTA, R., SRIDHAR, D. & RAI, B. 2016. Molecular dynamics simulation study of permeation of molecules through skin lipid bilayer. *The Journal of Physical Chemistry B,* 120**,** 8987-8996.

HADJ-RABIA, S., BAALA, L., VABRES, P., HAMEL-TEILLAC, D., JACQUEMIN, E., FABRE, M., LYONNET, S., DE PROST, Y., MUNNICH, A. & HADCHOUEL, M. 2004. Claudin-1 gene mutations in neonatal sclerosing cholangitis associated with ichthyosis: a tight junction disease. *Gastroenterology,* 127**,** 1386-1390.

HARTSOCK, A. & NELSON, W. J. 2008. Adherens and tight junctions: structure, function and connections to the actin cytoskeleton. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1778**,** 660-669.

HAS, C. & SITARU, C. 2013. *Molecular dermatology*, Humana Press.

HASHIMOTO, K. 1971. Intercellular spaces of the human epidermis as demonstrated with lanthanum. *Journal of Investigative Dermatology,* 57**,** 17-31.

HASHIMOTO, Y., YAGI, K. & KONDOH, M. 2016. Current progress in a second-generation claudin binder, anti-claudin antibody, for clinical applications. *Drug discovery today,* 21**,** 1711-1718.

HEINEMANN, U. & SCHUETZ, A. 2019. Structural features of tight-junction proteins. *International Journal of Molecular Sciences,* 20**,** 6020.

HISTOLOGY, L. 2011. Chapter 2: Epithelial Tissue. Available from: https://learnhistology.wordpress.com/2011/08/12/chapter-2-epithelial-tissue/.

HOŠEK, P. & SPIWOK, V. 2016. Metadyn View: Fast web-based viewer of free energy surfaces calculated by metadynamics. *Computer Physics Communications,* 198**,** 222-229.

HOSPITAL, A., GOÑI, J. R., OROZCO, M. & GELPÍ, J. L. 2015. Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry: AABC,* 8**,** 37.

HOU, J., RENIGUNTA, A., YANG, J. & WALDEGGER, S. 2010. Claudin-4 forms paracellular chloride channel in the kidney and requires claudin-8 for tight junction localization. *Proceedings of the National Academy of Sciences,* 107**,** 18010-18015.

HUB, J. S., DE GROOT, B. L. & VAN DER SPOEL, D. 2010. g_wham— A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of chemical theory and computation,* 6**,** 3713-3720.

HUG, S. 2013. Classical molecular dynamics in a nutshell. *Biomolecular Simulations: Methods and Protocols***,** 127-152.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: visual molecular dynamics. *Journal of molecular graphics,* 14**,** 33-38.

HUNG, W.-C., LEE, M.-T., CHEN, F.-Y. & HUANG, H. W. 2007. The condensing effect of cholesterol in lipid bilayers. *Biophysical journal,* 92**,** 3960-3967.

IKENOUCHI, J., FURUSE, M., FURUSE, K., SASAKI, H., TSUKITA, S. & TSUKITA, S. 2005. Tricellulin constitutes a novel barrier at tricellular contacts of epithelial cells. *The Journal of cell biology,* 171**,** 939-945.

INAI, T., KOBAYASHI, J. & SHIBATA, Y. 1999. Claudin-1 contributes to the epithelial barrier function in MDCK cells. *European journal of cell biology,* 78**,** 849-855.

IRUDAYANATHAN, F. J., TRASATTI, J. P., KARANDE, P. & NANGIA, S. 2015. Molecular Architecture of the Blood Brain Barrier Tight Junction Proteins–A Synergistic Computational and In Vitro Approach. *The Journal of Physical Chemistry B,* 120**,** 77-88.

IRUDAYANATHAN, F. J., WANG, N., WANG, X. & NANGIA, S. 2017. Architecture of the paracellular channels formed by claudins of the blood–brain barrier tight junctions. *Annals of the New York Academy of Sciences*.

IRUDAYANATHAN, F. J., WANG, X., WANG, N., WILLSEY, S. R., SEDDON, I. A. & NANGIA, S. 2018. Self-Assembly Simulations of Classic Claudins–insights into the Pore Structure, Selectivity and Higher Order Complexes. *The Journal of Physical Chemistry B*.

IWAI, I., HAN, H., DEN HOLLANDER, L., SVENSSON, S., ÖFVERSTEDT, L.-G., ANWAR, J., BREWER, J., BLOKSGAARD, M., LALOEUF, A. & NOSEK, D. 2012. The human skin barrier is organized as stacked bilayers of fully extended ceramides with cholesterol molecules associated with the ceramide sphingoid moiety. *Journal of Investigative Dermatology,* 132**,** 2215-2225.

JENSEN, F. 2007. *Introduction to computational chemistry*, John Wiley & Sons.

KABSCH, W. & SANDER, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* 22**,** 2577-2637.

KAN, F. 1993. Cytochemical evidence for the presence of phospholipids in epithelial tight junction strands. *Journal of Histochemistry & Cytochemistry,* 41**,** 649-656.

KIRSCHNER, N., HOUDEK, P., FROMM, M., MOLL, I. & BRANDNER, J. M. 2010. Tight junctions form a barrier in human epidermis. *European journal of cell biology,* 89**,** 839-842.

KLAUDA, J. B., VENABLE, R. M., FREITES, J. A., O'CONNOR, J. W., TOBIAS, D. J., MONDRAGON-RAMIREZ, C., VOROBYOV, I., MACKERELL JR, A. D. & PASTOR, R. W. 2010. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *The journal of physical chemistry B,* 114**,** 7830-7843.

KOKUBO, H., TANAKA, T. & OKAMOTO, Y. 2013. Two-dimensional replica-exchange method for predicting protein–ligand binding structures. *Journal of Computational Chemistry,* 34**,** 2601-2614.

KONE, A. & KOFKE, D. A. 2005. Selection of temperature intervals for parallel-tempering simulations. *The Journal of chemical physics,* 122**,** 206101.

KOVAL, M. 2013. Differential pathways of claudin oligomerization and integration into tight junctions. *Tissue barriers,* 1**,** e24518.

KRAUSE, G., PROTZE, J. & PIONTEK, J. Assembly and function of claudins: Structure–function relationships based on homology models and crystal structures. Seminars in cell & developmental biology, 2015. Elsevier, 3-12.

KRAUSE, G., WINKLER, L., MUELLER, S. L., HASELOFF, R. F., PIONTEK, J. & BLASIG, I. E. 2008. Structure and function of claudins. *Biochimica et Biophysica Acta (BBA)-Biomembranes,* 1778**,** 631-645.

KRAUSE, G., WINKLER, L., PIEHL, C., BLASIG, I., PIONTEK, J. & MÜLLER, S. L. 2009. Structure and function of extracellular claudin domains. *Annals of the New York Academy of Sciences,* 1165**,** 34-43.

KRAUSE, M. R. & REGEN, S. L. 2014. The structural role of cholesterol in cell membranes: from condensed bilayers to lipid rafts. *Accounts of chemical research,* 47**,** 3512-3521.

KRISSINEL, E. & HENRICK, K. 2007. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology,* 372**,** 774-797.

KRUG, S. M., GÜNZEL, D., CONRAD, M. P., LEE, I. F. M., AMASHEH, S., FROMM, M. & YU, A. S. 2012a. Charge-selective claudin channels. *Annals of the New York Academy of Sciences,* 1257**,** 20-28.

KRUG, S. M., GÜNZEL, D., CONRAD, M. P., ROSENTHAL, R., FROMM, A., AMASHEH, S., SCHULZKE, J. D. & FROMM, M. 2012b. Claudin-17 forms tight junction channels with distinct anion selectivity. *Cellular and Molecular Life Sciences,* 69**,** 2765-2778.

KRUG, S. M., SCHULZKE, J. D. & FROMM, M. Tight junction, selective permeability, and related diseases. Seminars in cell & developmental biology, 2014. Elsevier, 166-176.

KRYSTOFIAK, E. S., HEYMANN, J. B. & KACHAR, B. 2019. Carbon replicas reveal double stranded structure of tight junctions in phase-contrast electron microscopy. *Communications biology,* 2**,** 98.

KUČERKA, N., NAGLE, J. F., SACHS, J. N., FELLER, S. E., PENCER, J., JACKSON, A. & KATSARAS, J. 2008. Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. *Biophysical journal,* 95**,** 2356-2367.

KUČERKA, N., TRISTRAM-NAGLE, S. & NAGLE, J. F. 2006. Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *The Journal of membrane biology,* 208**,** 193-202.

KUMARI, R., KUMAR, R., CONSORTIUM, O. S. D. D. & LYNN, A. 2014. g_mmpbsa— A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of chemical information and modeling,* 54**,** 1951-1962.

KUPRUSEVICIUS, E., WHITE, G. & OGANESYAN, V. S. 2011. Prediction of nitroxide spin label EPR spectra from MD trajectories: application to myoglobin. *Faraday discussions,* 148**,** 283-298.

LAIO, A. & GERVASIO, F. L. 2008. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics,* 71**,** 126601.

LAIO, A. & PARRINELLO, M. 2002. Escaping free-energy minima. *Proceedings of the National Academy of Sciences,* 99**,** 12562-12566.

LAMPE, M. A., WILLIAMS, M. L. & ELIAS, P. M. 1983. Human epidermal lipids: characterization and modulations during differentiation. *Journal of Lipid Research,* 24**,** 131-140.

LEACH, A. R. 2001. *Molecular modelling: principles and applications*, Pearson education.

LEE, D. B., JAMGOTCHIAN, N., ALLEN, S. G., ABELES, M. B. & WARD, H. J. 2008. A lipid-protein hybrid model for tight junction. *American Journal of Physiology-Renal Physiology,* 295**,** F1601-F1612.

LEE, J., PATEL, D. S., STÅHLE, J., PARK, S.-J., KERN, N. R., KIM, S., LEE, J., CHENG, X., VALVANO, M. A. & HOLST, O. 2018. CHARMM-GUI membrane builder for complex biological membrane simulations with glycolipids and lipoglycans. *Journal of chemical theory and computation,* 15**,** 775-786.

LEE, S., TRAN, A., ALLSOPP, M., LIM, J. B., HÉNIN, J. R. M. & KLAUDA, J. B. 2014. CHARMM36 united atom chain model for lipids and surfactants. *The journal of physical chemistry B,* 118**,** 547-556.

LI, J., ANGELOW, S., LINGE, A., ZHUO, M. & YU, A. S. 2013. Claudin-2 pore function requires an intramolecular disulfide bond between two conserved extracellular cysteines. *American Journal of Physiology-Cell Physiology,* 305**,** C190-C196.

LI, Z., WAN, H., SHI, Y. & OUYANG, P. 2004. Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/Draw, and ChemSketch. *Journal of Chemical Information and Computer Sciences,* 44**,** 1886-1890.

LIANG, G. H. & WEBER, C. R. 2014. Molecular aspects of tight junction barrier function. *Current opinion in pharmacology,* 19**,** 84-89.

LIM, T. S., VEDULA, S. R. K., HUNZIKER, W. & LIM, C. T. 2008a. Kinetics of adhesion mediated by extracellular loops of claudin-2 as revealed by single-molecule force spectroscopy. *Journal of molecular biology,* 381**,** 681-691.

LIM, T. S., VEDULA, S. R. K., KAUSALYA, P. J., HUNZIKER, W. & LIM, C. T. 2008b. Single-molecular-level study of claudin-1-mediated adhesion. *Langmuir,* 24**,** 490-495.

LOMIZE, M. A., POGOZHEVA, I. D., JOO, H., MOSBERG, H. I. & LOMIZE, A. L. 2011. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic acids research,* 40**,** D370-D376.

LUISSINT, A.-C., ARTUS, C., GLACIAL, F., GANESHAMOORTHY, K. & COURAUD, P.-O. 2012. Tight junctions at the blood brain barrier: physiological architecture and disease-associated dysregulation. *Fluids and Barriers of the CNS,* 9**,** 1.

LUNDBORG, M., NARANGIFARD, A., WENNBERG, C. L., LINDAHL, E., DANEHOLT, B. & NORLÉN, L. 2018. Human skin barrier structure and function analyzed by cryo-EM and molecular dynamics simulation. *Journal of structural biology,* 203**,** 149-161.

MACDERMAID, C. M., DEVANE, R. H., KLEIN, M. L. & FIORIN, G. 2015. Dehydration of Multilamellar Fatty Acid Membranes: Towards a Computational Model of the Stratum Corneum. *Biophysical Journal,* 108**,** 243a.

MADDEN, M. E. & SARRAS, M. P. 1985. Development of an apical plasma membrane domain and tight junctions during histogenesis of the mammalian pancreas. *Developmental biology,* 112**,** 427-442.

MANIBOG, K., LI, H., RAKSHIT, S. & SIVASANKAR, S. 2014. Resolving the molecular mechanism of cadherin catch bond formation. *Nature communications,* 5**,** 3941.

MARKOV, A. G., ASCHENBACH, J. R. & AMASHEH, S. 2015. Claudin clusters as determinants of epithelial barrier function. *IUBMB life,* 67**,** 29-35.

MARKS, J., MILLER, J. J. & LOOKINGBILL, D. 2006. Principles of dermatology. *Lookingbill and Marks‟ principles of dermatology (4 th ed.). Philadelphia, PA: Saunders Elsevier***,** 3-50.

MARRINK, S. J., DE VRIES, A. H. & MARK, A. E. 2004. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B,* 108**,** 750-760.

MARRINK, S. J., RISSELADA, H. J., YEFIMOV, S., TIELEMAN, D. P. & DE VRIES, A. H. 2007. The MARTINI force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B,* 111**,** 7812-7824.

MCMANUS, J. J., CHARBONNEAU, P., ZACCARELLI, E. & ASHERIE, N. 2016. The physics of protein self-assembly. *Current opinion in colloid & interface science,* 22**,** 73-79.

MEERTENS, L., BERTAUX, C., CUKIERMAN, L., CORMIER, E., LAVILLETTE, D., COSSET, F.-L. & DRAGIC, T. 2008. The tight junction proteins claudin-1,-6, and-9 are entry cofactors for hepatitis C virus. *Journal of virology,* 82**,** 3555-3560.

MELI, M. & COLOMBO, G. 2013. A Hamiltonian replica exchange molecular dynamics (MD) method for the study of folding, based on the analysis of the stabilization determinants of proteins. *International journal of molecular sciences,* 14**,** 12157-12169.

MENON, G. K. 2002. New insights into skin structure: scratching the surface. *Advanced drug delivery reviews,* 54**,** S3-S17.

MICHAUD-AGRAWAL, N., DENNING, E. J., WOOLF, T. B. & BECKSTEIN, O. 2011. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry,* 32**,** 2319-2327.

MILATZ, S., PIONTEK, J., HEMPEL, C., MEOLI, L., GROHE, C., FROMM, A., LEE, I. F. M., EL-ATHMAN, R. & GÜNZEL, D. 2017. Tight junction strand formation by claudin-

10 isoforms and claudin-10a/-10b chimeras. *Annals of the New York Academy of Sciences,* 1405**,** 102-115.

MILATZ, S., PIONTEK, J., SCHULZKE, J.-D., BLASIG, I. E., FROMM, M. & GÜNZEL, D. 2015. Probing the cis-arrangement of prototype tight junction proteins claudin-1 and claudin-3. *Biochemical Journal,* 468**,** 449-458.

MINETA, K., YAMAMOTO, Y., YAMAZAKI, Y., TANAKA, H., TADA, Y., SAITO, K., TAMURA, A., IGARASHI, M., ENDO, T. & TAKEUCHI, K. 2011. Predicted expansion of the claudin multigene family. *FEBS letters,* 585**,** 606-612.

MITIC, L. L., UNGER, V. M. & ANDERSON, J. M. 2003. Expression, solubilization, and biochemical characterization of the tight junction transmembrane protein claudin-4. *Protein Science,* 12**,** 218-227.

MONTICELLI, L., KANDASAMY, S. K., PERIOLE, X., LARSON, R. G., TIELEMAN, D. P. & MARRINK, S.-J. 2008. The MARTINI coarse-grained force field: extension to proteins. *Journal of chemical theory and computation,* 4**,** 819-834.

MORITA, K. & MIYACHI, Y. 2003. Tight junctions in the skin. *Journal of dermatological science,* 31**,** 81-89.

MRSNY, R. J., BROWN, G. T., GERNER-SMIDT, K., BURET, A. G., MEDDINGS, J. B., QUAN, C., KOVAL, M. & NUSRAT, A. 2008. A key claudin extracellular loop domain is critical for epithelial barrier integrity. *The American journal of pathology,* 172**,** 905-915.

NAKAJIMA, M., NAGASE, S., IIDA, M., TAKEDA, S., YAMASHITA, M., WATARI, A., SHIRASAGO, Y., FUKASAWA, M., TAKEDA, H. & SAWASAKI, T. 2015. Claudin-1 binder enhances epidermal permeability in a human keratinocyte model. *Journal of Pharmacology and Experimental Therapeutics,* 354**,** 440-447.

NAKAMURA, S., IRIE, K., TANAKA, H., NISHIKAWA, K., SUZUKI, H., SAITOH, Y., TAMURA, A., TSUKITA, S. & FUJIYOSHI, Y. 2019. Morphologic determinant of tight junctions revealed by claudin-3 structures. *Nature communications,* 10**,** 816.

NOTMAN, R. & ANWAR, J. 2013. Breaching the skin barrier—Insights from molecular simulation of model membranes. *Advanced drug delivery reviews,* 65**,** 237-250.

PAN, J., TRISTRAM-NAGLE, S., KUČERKA, N. & NAGLE, J. F. 2008. Temperature dependence of structure, bending rigidity, and bilayer interactions of dioleoylphosphatidylcholine bilayers. *Biophysical journal,* 94**,** 117-124.

PANORCHAN, P., THOMPSON, M. S., DAVIS, K. J., TSENG, Y., KONSTANTOPOULOS, K. & WIRTZ, D. 2006. Single-molecule analysis of cadherin-mediated cell-cell adhesion. *Journal of cell science,* 119**,** 66-74.

PAQUET, E. & VIKTOR, H. L. 2015. Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review. *BioMed research international,* 2015.

PERIOLE, X., CAVALLI, M., MARRINK, S.-J. & CERUSO, M. A. 2009. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *Journal of chemical theory and computation,* 5**,** 2531-2543.

PETRACHE, H. I., TRISTRAM-NAGLE, S., GAWRISCH, K., HARRIES, D., PARSEGIAN, V. A. & NAGLE, J. F. 2004. Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. *Biophysical journal,* 86**,** 1574-1586.

PETTERSEN, E. F., GODDARD, T. D., HUANG, C. C., COUCH, G. S., GREENBLATT, D. M., MENG, E. C. & FERRIN, T. E. 2004. UCSF Chimera—a visualization system for

exploratory research and analysis. *Journal of computational chemistry,* 25**,** 1605-1612.

PEZESHKIAN, W., KHANDELIA, H. & MARSH, D. 2018. Lipid configurations from molecular dynamics simulations. *Biophysical journal,* 114**,** 1895-1907.

PIEHL, C., PIONTEK, J., CORDING, J., WOLBURG, H. & BLASIG, I. E. 2010. Participation of the second extracellular loop of claudin-5 in paracellular tightening against ions, small and large molecules. *Cellular and Molecular Life Sciences,* 67**,** 2131-2140.

PIONTEK, A., ROSSA, J., PROTZE, J., WOLBURG, H., HEMPEL, C., GÜNZEL, D., KRAUSE, G. & PIONTEK, J. 2017. Polar and charged extracellular residues conserved among barrier-forming claudins contribute to tight junction strand formation. *Annals of the New York Academy of Sciences*.

PIONTEK, J. R., WINKLER, L., WOLBURG, H., MULLER, S. L., ZULEGER, N., PIEHL, C., WIESNER, B., KRAUSE, G. & BLASIG, I. E. 2008. Formation of tight junction: determinants of homophilic interaction between classic claudins. *The FASEB Journal,* 22**,** 146-158.

PONDER, J. W. & CASE, D. A. 2003. Force fields for protein simulations. *Advances in protein chemistry,* 66**,** 27-85.

QI, R., WEI, G., MA, B. & NUSSINOV, R. 2018. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example. *Peptide Self-Assembly.* Springer.

QUINN, P. J. 2010. A lipid matrix model of membrane raft structure. *Progress in lipid research,* 49**,** 390-406.

RAJAGOPAL, N., IRUDAYANATHAN, F. J. & NANGIA, S. 2019. Palmitoylation of Claudin-5 Proteins Influences Their Lipid Domain Affinity and Tight Junction Assembly at the Blood–Brain Barrier Interface. *The Journal of Physical Chemistry B,* 123**,** 983-993.

RAUSCHER, S., GAPSYS, V., GAJDA, M. J., ZWECKSTETTER, M., DE GROOT, B. L. & GRUBMÜLLER, H. 2015. Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. *Journal of chemical theory and computation,* 11**,** 5513-5524.

RAYMOND, A.-A., DE PEREDO, A. G., STELLA, A., ISHIDA-YAMAMOTO, A., BOUYSSIE, D., SERRE, G., MONSARRAT, B. & SIMON, M. 2008a. Lamellar Bodies of Human Epidermis Proteomics Characterization by High Throughput Mass Spectrometry and Possible Involvement of CLIP-170 in their Trafficking/Secretion. *Molecular & Cellular Proteomics,* 7**,** 2151-2175.

RAYMOND, A.-A., DE PEREDO, A. G., STELLA, A., ISHIDA-YAMAMOTO, A., BOUYSSIE, D., SERRE, G., MONSARRAT, B. & SIMON, M. 2008b. Lamellar bodies of human epidermis: proteomics characterization by high throughput mass spectrometry and possible involvement of CLIP-170 in their trafficking/secretion. *Molecular & Cellular Proteomics,* 7**,** 2151-2175.

ROSE, P. W., BERAN, B., BI, C., BLUHM, W. F., DIMITROPOULOS, D., GOODSELL, D. S., PRLIĆ, A., QUESADA, M., QUINN, G. B. & WESTBROOK, J. D. 2010. The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic acids research,* 39**,** D392-D401.

ROSE, P. W., PRLIĆ, A., ALTUNKAYA, A., BI, C., BRADLEY, A. R., CHRISTIE, C. H., COSTANZO, L. D., DUARTE, J. M., DUTTA, S. & FENG, Z. 2016. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic acids research*, gkw1000.

ROSENTHAL, R., MILATZ, S., KRUG, S. M., OELRICH, B., SCHULZKE, J.-D., AMASHEH, S., GÜNZEL, D. & FROMM, M. 2010. Claudin-2, a component of the tight junction, forms a paracellular water channel. *J Cell Sci,* 123**,** 1913-1921.

ROSSA, J., PLOEGER, C., VORREITER, F., SALEH, T., PROTZE, J., GÜNZEL, D., WOLBURG, H., KRAUSE, G. & PIONTEK, J. 2014. Claudin-3 and claudin-5 protein folding and assembly into the tight junction are controlled by non-conserved residues in the transmembrane 3 (TM3) and extracellular loop 2 (ECL2) segments. *Journal of Biological Chemistry,* 289**,** 7641-7653.

ROUX, B. 1995. The calculation of the potential of mean force using computer simulations. *Computer physics communications,* 91**,** 275-282.

ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols,* 5**,** 725-738.

SAITOH, Y., SUZUKI, H., TANI, K., NISHIKAWA, K., IRIE, K., OGURA, Y., TAMURA, A., TSUKITA, S. & FUJIYOSHI, Y. 2015. Structural insight into tight junction disassembly by Clostridium perfringens enterotoxin. *Science,* 347**,** 775-778.

SAMANTA, P., WANG, Y., FULADI, S., ZOU, J., LI, Y., SHEN, L., WEBER, C. & KHALILI-ARAGHI, F. 2018. Molecular determination of claudin-15 organization and channel selectivity. *The Journal of general physiology*, jgp. 201711868.

SAWADA, N. 2013. Tight junction-related human diseases. *Pathology international,* 63**,** 1-12.

SCHYMKOWITZ, J., BORG, J., STRICHER, F., NYS, R., ROUSSEAU, F. & SERRANO, L. 2005. The FoldX web server: an online force field. *Nucleic acids research,* 33**,** W382-W388.

SHAHZAD, Y., LOUW, R., GERBER, M. & DU PLESSIS, J. 2015. Breaching the skin barrier through temperature modulations. *Journal of Controlled Release,* 202**,** 1-13.

SHEA, J.-E. & LEVINE, Z. A. 2016. Studying the early stages of protein aggregation using replica exchange molecular dynamics simulations. *Protein Amyloid Aggregation.* Springer.

SHEN, L., WEBER, C. R., RALEIGH, D. R., YU, D. & TURNER, J. R. 2011. Tight junction pore and leak pathways: a dynamic duo. *Annual review of physiology,* 73**,** 283-309.

SHEN, L., WEBER, C. R. & TURNER, J. R. 2008. The tight junction protein complex undergoes rapid and continuous molecular remodeling at steady state. *The Journal of cell biology,* 181**,** 683-695.

SHINODA, T., SHINYA, N., ITO, K., OHSAWA, N., TERADA, T., HIRATA, K., KAWANO, Y., YAMAMOTO, M., KIMURA-SOMEYA, T. & YOKOYAMA, S. 2016. Structural basis for disruption of claudin assembly in tight junctions by an enterotoxin. *Scientific Reports,* 6.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M. & SÖDING, J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology,* 7.

SKJEVIK, Å. A., MADEJ, B. D., DICKSON, C. J., LIN, C., TEIGEN, K., WALKER, R. C. & GOULD, I. R. 2016. Simulation of lipid bilayer self-assembly using all-atom lipid force fields. *Physical Chemistry Chemical Physics,* 18**,** 10573-10584.

STAEHELIN, L. A. 1974. Structure and function of intercellular junctions. *International review of cytology,* 39**,** 191-283.

STEVENSON, B. R., ANDERSON, J. M., GOODENOUGH, D. A. & MOOSEKER, M. S. 1988. Tight junction structure and ZO-1 content are identical in two strains of Madin-Darby canine kidney cells which differ in transepithelial resistance. *The Journal of cell biology,* 107**,** 2401-2408.

STEVENSON, B. R., SILICIANO, J. D., MOOSEKER, M. S. & GOODENOUGH, D. A. 1986. Identification of ZO-1: a high molecular weight polypeptide associated with the tight junction (zonula occludens) in a variety of epithelia. *The Journal of cell biology,* 103**,** 755-766.

SUGITA, Y., MIYASHITA, N., LI, P.-C., YODA, T. & OKAMOTO, Y. 2012. Recent applications of replica-exchange molecular dynamics simulations of biomolecules. *Current Physical Chemistry,* 2**,** 401-412.

SUZUKI, H., ITO, Y., YAMAZAKI, Y., MINETA, K., UJI, M., ABE, K., TANI, K., FUJIYOSHI, Y. & TSUKITA, S. 2013. The four-transmembrane protein IP39 of Euglena forms strands by a trimeric unit repeat. *Nature communications,* 4**,** 1766.

SUZUKI, H., NISHIZAWA, T., TANI, K., YAMAZAKI, Y., TAMURA, A., ISHITANI, R., DOHMAE, N., TSUKITA, S., NUREKI, O. & FUJIYOSHI, Y. 2014. Crystal structure of a claudin provides insight into the architecture of tight junctions. *Science,* 344**,** 304-307.

SUZUKI, H., TANI, K. & FUJIYOSHI, Y. 2017. Crystal structures of claudins: insights into their intermolecular interactions. *Annals of the New York Academy of Sciences,* 1397**,** 25-34.

SUZUKI, H., TANI, K., TAMURA, A., TSUKITA, S. & FUJIYOSHI, Y. 2015. Model for the architecture of claudin-based paracellular ion channels through tight junctions. *Journal of molecular biology,* 427**,** 291-297.

TAKAHASHI, A., SAITO, Y., KONDOH, M., MATSUSHITA, K., KRUG, S. M., SUZUKI, H., TSUJINO, H., LI, X., AOYAMA, H. & MATSUHISA, K. 2012. Creation and biochemical analysis of a broad-specific claudin binder. *Biomaterials,* 33**,** 3464-3474.

TAKEICHI, M. 2006. Shoichiro Tsukita: a life exploring the molecular architecture of the tight junction. *The Journal of cell biology,* 172**,** 321-323.

TAMURA, A. & TSUKITA, S. Paracellular barrier and channel functions of TJ claudins in organizing biological systems: advances in the field of barriology revealed in knockout mice. Seminars in cell & developmental biology, 2014. Elsevier, 177-185.

TRIBELLO, G. A., BONOMI, M., BRANDUARDI, D., CAMILLONI, C. & BUSSI, G. 2014. PLUMED 2: New feathers for an old bird. *Computer Physics Communications,* 185**,** 604-613.

TSCHEIK, C., BLASIG, I. E. & WINKLER, L. 2013. Trends in drug delivery through tissue barriers containing tight junctions. *Tissue barriers,* 1**,** e24565.

TSUKITA, S. & FURUSE, M. 2000. Pores in the wall claudins constitute tight junction strands containing aqueous pores. *The Journal of cell biology,* 149**,** 13-16.

TSUKITA, S., FURUSE, M. & ITOH, M. 2001. Multifunctional strands in tight junctions. *Nature reviews Molecular cell biology,* 2**,** 285-293.

TURIN, L., BEHE, P., PLONSKY, I. & DUNINA-BARKOVSKAYA, A. 1991. Hydrophobic ion transfer between membranes of adjacent hepatocytes: a possible probe of tight junction structure. *Proceedings of the National Academy of Sciences,* 88**,** 9365-9369.

TURKSEN, K. & TROY, T.-C. 2010. Claudin is Skin Deep. *Current Topics in Membranes,* 65**,** 255-272.

TURNER, J. R., BUSCHMANN, M. M., ROMERO-CALVO, I., SAILER, A. & SHEN, L. The role of molecular remodeling in differential regulation of tight junction permeability. Seminars in cell & developmental biology, 2014. Elsevier, 204-212.

VAN ITALLIE, C. M. & ANDERSON, J. M. 2004. The role of claudins in determining paracellular charge selectivity. *Proceedings of the American Thoracic Society,* 1**,** 38-41.

VAN ITALLIE, C. M. & ANDERSON, J. M. Architecture of tight junctions and principles of molecular composition. Seminars in cell & developmental biology, 2014. Elsevier, 157-165.

VAN ITALLIE, C. M., HOLMES, J., BRIDGES, A., GOOKIN, J. L., COCCARO, M. R., PROCTOR, W., COLEGIO, O. R. & ANDERSON, J. M. 2008. The density of small tight junction pores varies among cell types and is increased by expression of claudin-2. *Journal of cell science,* 121**,** 298-305.

VAN ITALLIE, C. M., ROGAN, S., YU, A., VIDAL, L. S., HOLMES, J. & ANDERSON, J. M. 2006. Two splice variants of claudin-10 in the kidney create paracellular pores with different ion selectivities. *American Journal of Physiology-Renal Physiology,* 291**,** F1288-F1299.

VECCHIO, A. J. & STROUD, R. M. 2019. Claudin-9 structures reveal mechanism for toxin-induced gut barrier breakdown. *Proceedings of the National Academy of Sciences,* 116**,** 17817-17824.

VENABLE, R. M., BROWN, F. L. & PASTOR, R. W. 2015. Mechanical properties of lipid bilayers from molecular dynamics simulation. *Chemistry and physics of lipids,* 192**,** 60-74.

VERMEER, L. S., DE GROOT, B. L., RÉAT, V., MILON, A. & CZAPLICKI, J. 2007. Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with 2 H NMR experiments. *European Biophysics Journal,* 36**,** 919-931.

VESHNYAKOVA, A., KRUG, S. M., MUELLER, S. L., PIONTEK, J., PROTZE, J., FROMM, M. & KRAUSE, G. 2012. Determinants contributing to claudin ion channel formation. *Annals of the New York Academy of Sciences,* 1257**,** 45-53.

WASSENAAR, T. A., PLUHACKOVA, K., BÖCKMANN, R. A., MARRINK, S. J. & TIELEMAN, D. P. 2014. Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of chemical theory and computation,* 10**,** 676-690.

WEBER, C. R. 2012. Dynamic properties of the tight junction barrier. *Annals of the New York academy of sciences,* 1257**,** 77.

WERTZ, P. & NORLÉN, L. 2004. " Confidence Intervals" for the" True" Lipid Composition of the Human Skin Barrier? *BASIC AND CLINICAL DERMATOLOGY,* 26**,** 85-106.

WILLIAMS, T., KELLEY, C., LANG, R., KOTZ, D. & CAMPBELL, J. 2008. 1 gnuplot.

WOODCOCK, L.-V. 1971. Isothermal molecular dynamics calculations for liquid salts. *Chemical Physics Letters,* 10**,** 257-261.

WU, E. L., CHENG, X., JO, S., RUI, H., SONG, K. C., DÁVILA-CONTRERAS, E. M., QI, Y., LEE, J., MONJE-GALVAN, V. & VENABLE, R. M. 2014. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *Journal of computational chemistry,* 35**,** 1997-2004.

YANG, J., YAN, R., ROY, A., XU, D., POISSON, J. & ZHANG, Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nature methods,* 12**,** 7-8.

YOSHIDA, K., YOKOUCHI, M., NAGAO, K., ISHII, K., AMAGAI, M. & KUBO, A. 2013. Functional tight junction barrier localizes in the second layer of the stratum granulosum of human epidermis. *Journal of dermatological science,* 71**,** 89-99.

ZHAO, J., KRYSTOFIAK, E. S., BALLESTEROS, A., CUI, R., VAN ITALLIE, C. M., ANDERSON, J. M., FENOLLAR-FERRER, C. & KACHAR, B. 2018. Multiple claudin–claudin cis interfaces are required for tight junction strand formation and inherent flexibility. *Communications Biology,* 1**,** 50.

ZIHNI, C., MILLS, C., MATTER, K. & BALDA, M. S. 2016. Tight junctions: from simple barriers to multifunctional molecular gates. *Nature Reviews Molecular Cell Biology*.

ZORN-KRUPPA, M., VIDAL-Y-SY, S., HOUDEK, P., WLADYKOWSKI, E., GRZYBOWSKI, S., GRUBER, R., GORZELANNY, C., HARCUP, J., SCHNEIDER, S. W. & MAJUMDAR, A. 2018. Tight Junction barriers in human hair follicles–role of claudin-1. *Scientific reports,* 8**,** 1-16.

9 Appendices

9.1 Appendix A1

A1.1 The python script used to calculate the respective rotation angles θ1 and θ2 (about the z-axis) of each of the two domain particles involved in any given dimer interaction:

```
import numpy as np
import MDAnalysis as mda
import matplotlib.pyplot as plt
import math
u = mda.Universe("file.gro", "file.xtc")
u.trajectory[-1]
uni = u.select_atoms('protein')
CoM = []
bin1 = []
bin2 = []
c = [0,0,1]
step = insert here the number of atoms in a protein
numOfProteins = insert here the total number of proteins

for i in range(0, numOfProteins):
        CoM.append( uni.atoms[step*i:step*i+step-1].center_of_mass() )
A = u.select_atoms('resname GLN and resnum 29 and name CA')
B = u.select_atoms('resname THR and resnum 137 and name CA')
dbin = 10
sizeOfArray = int(math.floor(360 / dbin))
distribution = np.zeros((sizeOfArray, sizeOfArray), int)

for i in range(0, numOfProteins-1):
        for j in range(i+1, numOfProteins):
              CoM[i][2] = 0
              CoM[j][2] = 0
              A[i].position = [A[i].position[0], A[i].position[1], 0]
              B[i].position = [B[i].position[0], B[i].position[1], 0]

              A[j].position = [A[j].position[0], A[j].position[1], 0]
              B[j].position = [B[j].position[0], B[j].position[1], 0]
              dist = np.linalg.norm(CoM[i] - CoM[j])
              # when in proximity calculate the orientation angles
              if (dist < 35):
                     AB = A[i].position[0:3] - B[i].position[0:3]
                     BC = B[j].position[0:3] - B[i].position[0:3]
                     AB /= np.linalg.norm(AB)
                     BC /= np.linalg.norm(BC)
                     tripleProduct1 = np.dot(AB, np.cross((BC),c))

                     angle1= math.acos(np.dot(AB, BC))
                     if tripleProduct1 < 0:
                            angle1= 2 * math.pi -angle1
                     theta1 = angle1 * 180 / math.pi
                     bin1 = sizeOfArray - 1 - int(math.floor(theta1 / dbin))

                     AB = A[j].position[0:3] - B[j].position[0:3]
```

```
                    BC = B[i].position[0:3] - B[j].position[0:3]
                    AB /= np.linalg.norm(AB)
                    BC /= np.linalg.norm(BC)
                    tripleProduct2 = np.dot(AB, np.cross((BC), c))

                    angle2 = math.acos(np.dot(AB, BC))
                    if tripleProduct2 < 0:
                          angle2 = 2 * math.pi - angle2
                    theta2 = angle2 * 180 / math.pi
                    bin2 = int(math.floor(theta2 / dbin))

                    distribution[int(bin1)][int(bin2)]                          =
distribution[int(bin1)][int(bin2)] + 1
print(distribution)


image = distribution
plt.imshow(image, interpolation='sinc', cmap='Reds', extent=[0,360,0,360])
plt.colorbar()
plt.title('Combined distribution of Theta 1 and Theta 2.')
plt.xlabel('Theta 2')
plt.ylabel('Theta 1')
plt.show()
```

A1.2 The python script that calculates the relative penetration depth of the opposing ECDs in Chapter 4:

```
import numpy as np
import MDAnalysis as mda
import matplotlib.pyplot as plt
import sys
#create universe
u = mda.Universe("file.gro", "file.xtc", use_periodic_selections = True)
step = insert here the number of atoms in the protein
numOfProteins = insert here the number of proteins

for ts in u.trajectory:
        #select only protein atoms
        uni = u.select_atoms('protein')
        # find the centre of mass of all ECDs and store them on an array
        ProteinArrayCoM = []
        for i in range(0, numOfProteins):
             ProteinArrayCoM.append(           uni.atoms[step*i:step*i+step-
1].center_of_mass() )
        # create an array with all the max Z of each loop
        ProteinArrayZMax = []
        for i in range(0, numOfProteins):
             zmax = -sys.maxint-1
             zID = 0
             zPosition = [0,0,0]
             for j in range(step*i, step*i+step-1):
                    pos = uni.atoms[j].position
                    if ( pos[2] > zmax):
                          zmax = pos[2]
                          zID = uni.atoms[j].id
                          zPosition = pos
             ProteinArrayZMax.append( (zID, zPosition[2]) )

        # create an array with all the min Z of each loop
        ProteinArrayZMin = []
        for i in range(0, numOfProteins):
```

```
            zmin = sys.maxint
            zID = 0
            zPosition = [0,0,0]
            for j in range(step*i, step*i+step-1):
                  pos = uni.atoms[j].position
                  if ( pos[2] < zmin):
                        zmin = pos[2]
                        zID = uni.atoms[j].id
                        zPosition = pos
            ProteinArrayZMin.append( (zID, zPosition[2]) )

      #check if dx and dy are less than the cutoff between pairs
      for prot1 in range(0, numOfProteins/2):
            for prot2 in range(numOfProteins/2, numOfProteins):
                        ProteinA_XYZ     =     (ProteinArrayCoM[prot1][0],
ProteinArrayCoM[prot1][1], ProteinArrayCoM[prot1][2] )
                        ProteinB_XYZ     =     (ProteinArrayCoM[prot2][0],
ProteinArrayCoM[prot2][1], ProteinArrayCoM[prot2][2])

                        dx=abs(ProteinB_XYZ[0] - ProteinA_XYZ[0])
                        dy=abs(ProteinB_XYZ[1] - ProteinA_XYZ[1])
                        dz=abs(ProteinB_XYZ[2] - ProteinA_XYZ[2])

                        if (dx<20 and dy<20 and dz<70):
                              depth1=(ProteinArrayZMin[prot2][1]-
ProteinArrayZMax[prot1][1])
                              frame = u.trajectory.frame
                              print frame,depth1
```

A1.3 The plumed.dat file used to activate well-tempered metadynamics on the dihedral angles φ and ψ of the mutants in Chapter 6:

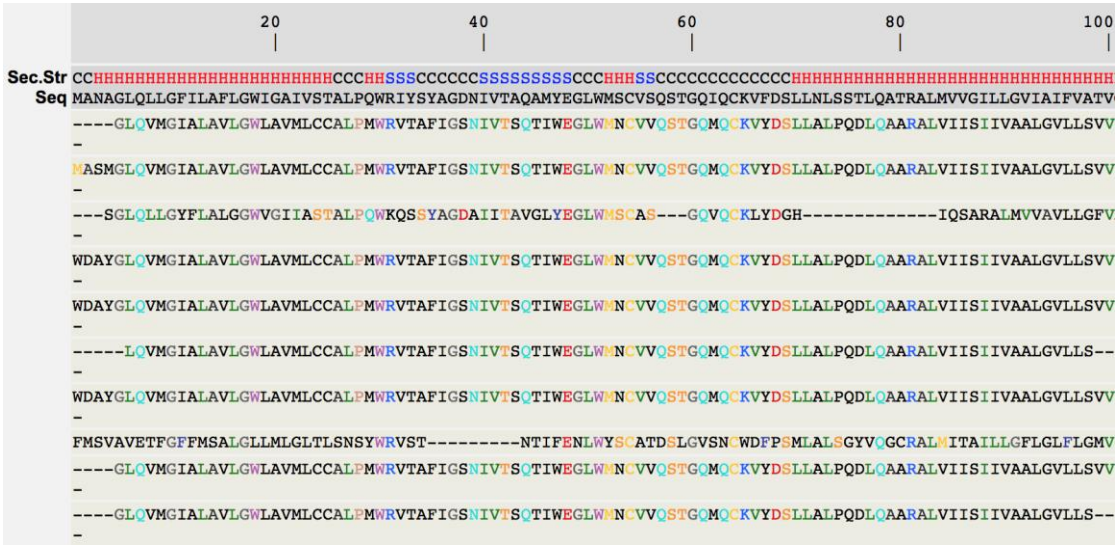
```
# First load information about the molecule.
MOLINFO STRUCTURE=file.pdb
# Group the protein atoms
pro1: GROUP ATOMS=insert protein atoms
#Make the protein whole
WHOLEMOLECULES ENTITY0=pro1
#Monitor chi1 angle
chi1: TORSION ATOMS=@chi1-53
# set up two variables for phi and psi dihedral angles
phi: TORSION ATOMS=@phi-53
psi: TORSION ATOMS=@psi-53
# Activate well-tempered metadynamics on phi and psi
METAD ...
LABEL=metad
ARG=phi,psi
#deposit a Gaussian every 500 steps with initial height equal to 1.2 kJ/mol
PACE=500 HEIGHT=1.2
# set the bias factor and temperature
BIASFACTOR=10
TEMP=310
#Gaussian width (sigma) for both CVs.
SIGMA=0.35,0.35
#Gaussians should be written to a file & also stored on grid
FILE=HILLS GRID_MIN=-pi,-pi  GRID_MAX=pi,pi
...
# Print the CVs and the value of the bias potential on COLVAR file
PRINT ARG=chi1,phi,psi,metad.bias  FILE=COLVAR STRIDE=100
```

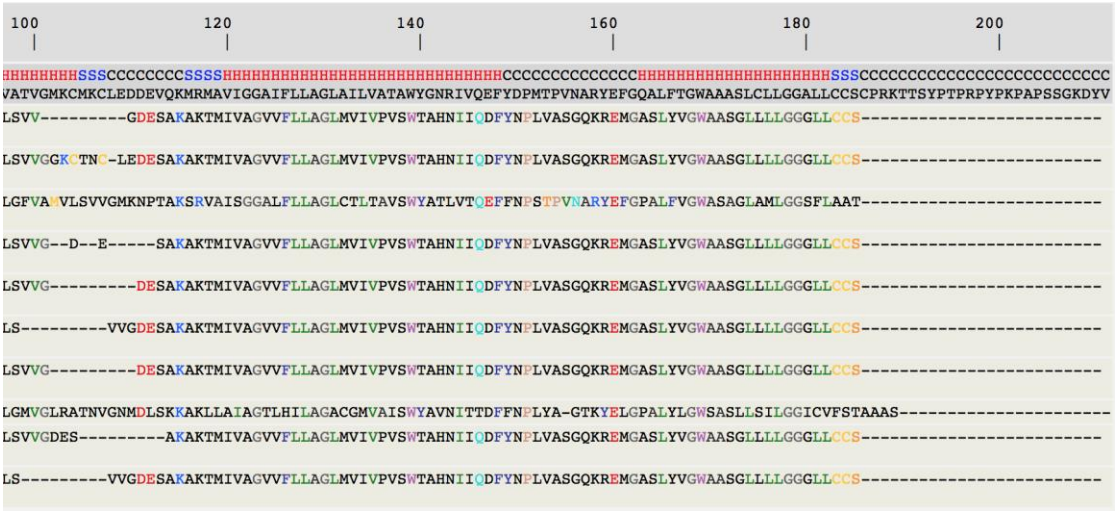
9.2 Appendix A2

A2.1 The ten templates of the highest significance used in the threading alignments to build the model of claudin-1 with I-TASSER. Identity 1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence, while identity 2 is the corresponding percentage between the whole template chains with the query sequence.

Rank	**PDB hit**	**Identity 1**	**Identity 2**
1	5B2G (claudin-4)	0.48	0.4
2	5B2G (claudin-4)	0.48	0.43
3	3X29(claudin-19)	0.54	0.47
4	5B2G (claudin-4)	0.47	0.41
5	5B2G (claudin-4)	0.48	0.41
6	5B2G (claudin-4)	0.47	0.4
7	5B2G (claudin-4)	0.48	0.41
8	4P79 (claudin-15)	0.32	0.29
9	5B2G (claudin-4)	0.48	0.4
10	5B2G (claudin-4)	0.48	0.4

The corresponding sequence alignment (split into two pictures) of the proteins-templates used to generate the model of claudin-1. The highlighted residues correspond to those which are identical to the residue in the query sequence (sequence on top).



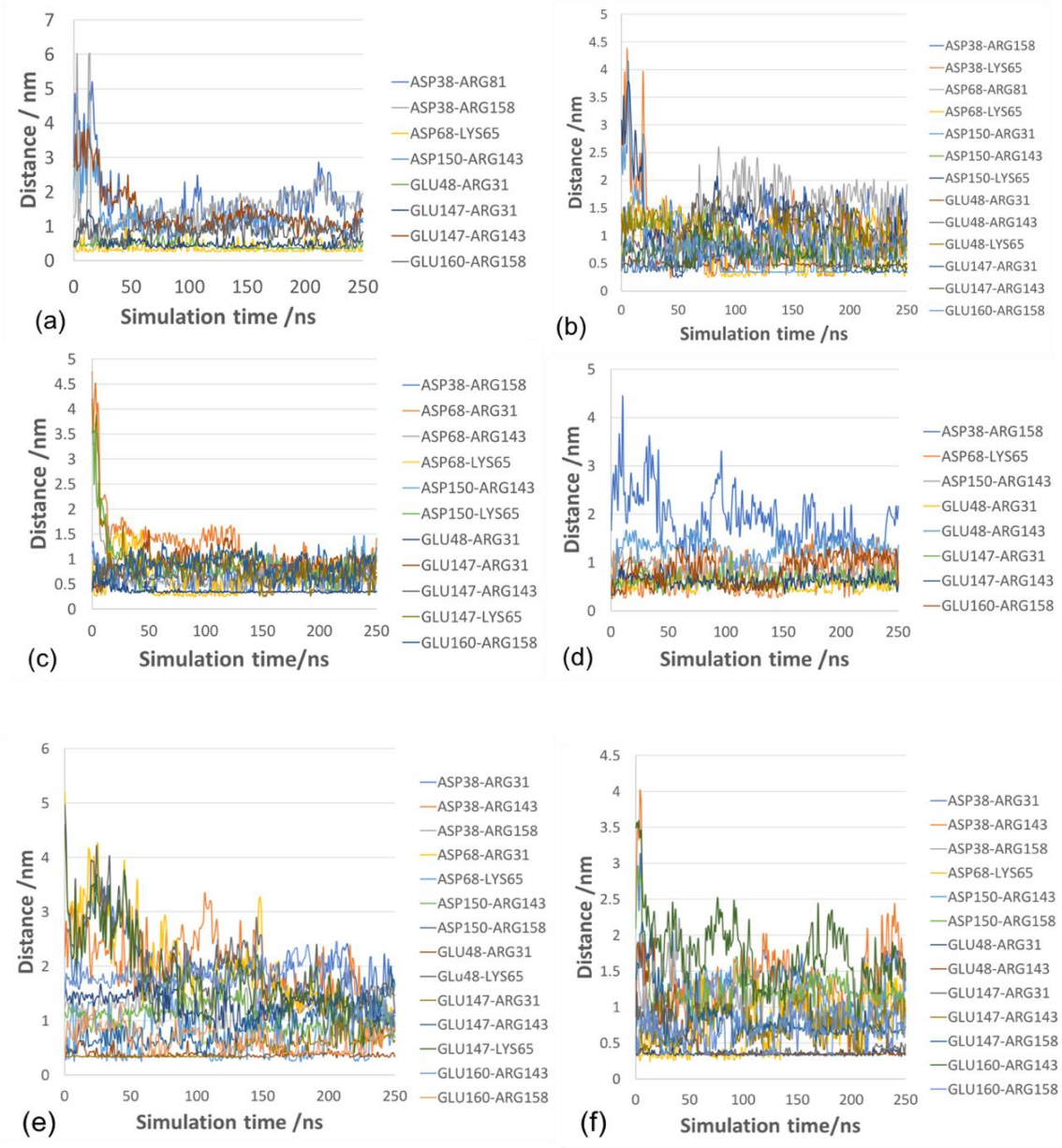


A2.2 Sequence alignment of claudins -1, -2, -4, -15 and -17 generated with Crystal Omega. Red colours are for small and hydrophobic (including aromatic) residues, blue and magenta are acidic and basic residues respectively, while serine, threonine, asparagine, glutamine, tyrosine, histidine, cysteine and glycine are coloured green.

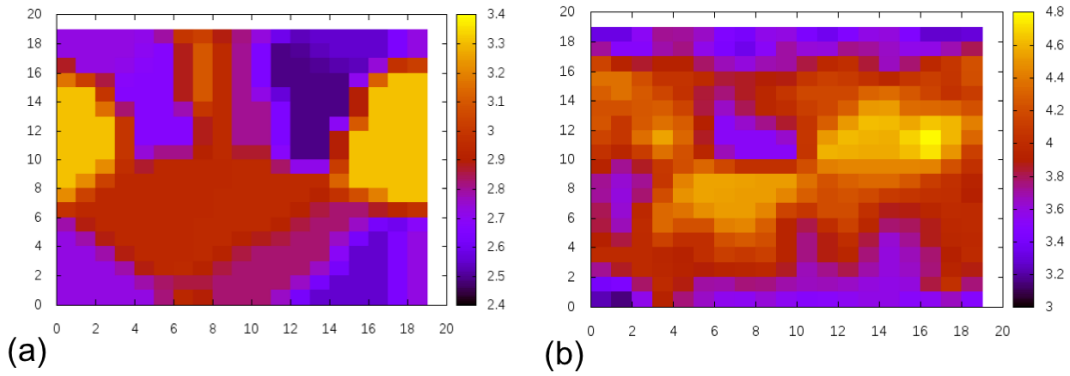
Asterisks (*) indicate positions which have a single, fully conserved residue, while a colon (:) and a period (.) indicate conservation between groups of strongly and weakly similar properties respectively.

```
sp|Q9Z0S5|CLD15_MOUSE      -MSVAVETFGFFMSALGLLMLGLTLSNSYWRVSTVHG-NVITTNTIFENLWYSCATDSLG 58
sp|O88552|CLD2_MOUSE       MASLGVQLVGYILGLLGLLGTSIAMLLPNWRTSSYVGASIVTAVGFSKGLWMECATHSTG 60
sp|Q8BXA6|CLD17_MOUSE      MAFYPLQIAGLVLGFFGLVGTIGTTLLPQWRVSAFIGSNIIIFERIWEGLWMNCIQQAMV 60
sp|O95832|CLD1_HUMAN       MANAGLQLLGFILAFLGWIGAIVSTALPQWRIYSYAGDNIVTAQAMYEGLWMSCVSQSTG 60
sp|O35054|CLD4_MOUSE       MASMGLQVLGISLAVLGWLGIILSCALPMWRVTAFIGSNIVTAQTSWEGLWMNCVVQSTG 60
                                ::  *  :. :* :    :     **  :  * .::      :.** .*  .:

sp|Q9Z0S5|CLD15_MOUSE      VSNCWDFPSMLALSGYVQGCRALMITAILLGFLGLFLGMVGLRCTNVGNMDLSKKAKLLA 118
sp|O88552|CLD2_MOUSE       ITQCDIYSTLLGLPADIQAAQAMMVTSSAMSSLACIISVVGMRCTVFCQDSR-AKDRVAV 119
sp|Q8BXA6|CLD17_MOUSE      TLQCKFYNSILALPPVLEAARALMCVAVALALVALIIGICGMKQLQCTGSSERVKAYLLG 120
sp|O95832|CLD1_HUMAN       QIQCKVFDSLLNLSSTLQATRALMVVGILLGVIAIFVATVGMKCMKCLEDDEVQKMRMAV 120
sp|O35054|CLD4_MOUSE       QMQCKMYDSMLALPQDLQAARALMVISIIVGALGMLLSVVGGKCTNCMEDETVKA-KIMI 119
                             :*  : ::* *   ::. :*:*  .  :. :. ::.  * :       .      :

sp|Q9Z0S5|CLD15_MOUSE      IAGTLHILAGACGMVAISWYAVNITTDFFNPLY-AGTKYELGPALYLGWSASLLSILGGI 177
sp|O88552|CLD2_MOUSE       VGGVFFILGGILGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISALFSLVAGV 179
sp|Q8BXA6|CLD17_MOUSE      TSGVLFILTGIFVLIPVSWTANIIIRDFYDPTVHAGQKRELGGALFLGWATAAVLFIGGG 180
sp|O95832|CLD1_HUMAN       IGGAIFLLAGLAILVATAWYGNRIVQEFYDPMTPVNARYEFGQALFTGWAAASLCLLGGA 180
sp|O35054|CLD4_MOUSE       TAGAVFIVASMLIMVPVSWTAHNVIRDFYNPMVASGQKREMGASLYVGWAASGLLLLGGG 179
                            .*...:: .   ::  :*    :  :*:.*    . : *:* :*: *  :: . ::.*

sp|Q9Z0S5|CLD15_MOUSE      CVFSTCCCSSKEEPATRAGL------PYKPSTVVIPRATSDESDISFGKYGKNAYV     227
sp|O88552|CLD2_MOUSE       ILCFSCSPQGNRTNYYDGYQAQPLATRSSPRSAQQPKAK-----SEFNSYSLTGYV     230
sp|Q8BXA6|CLD17_MOUSE      LLCGYCCCNRKERWHRYPVPAYRVPQKDNQRNVTVPR------------KSSTSYV     224
sp|O95832|CLD1_HUMAN       LLCCSC--PRKTTSYPTPRP-YPKPAPSS----------------------GKDYV     211
sp|O35054|CLD4_MOUSE       LLCCSC--PPRSNDKPYSAK-YSAARSVP----------------------ASNYV     210
                            :   *    .                                        . **
```

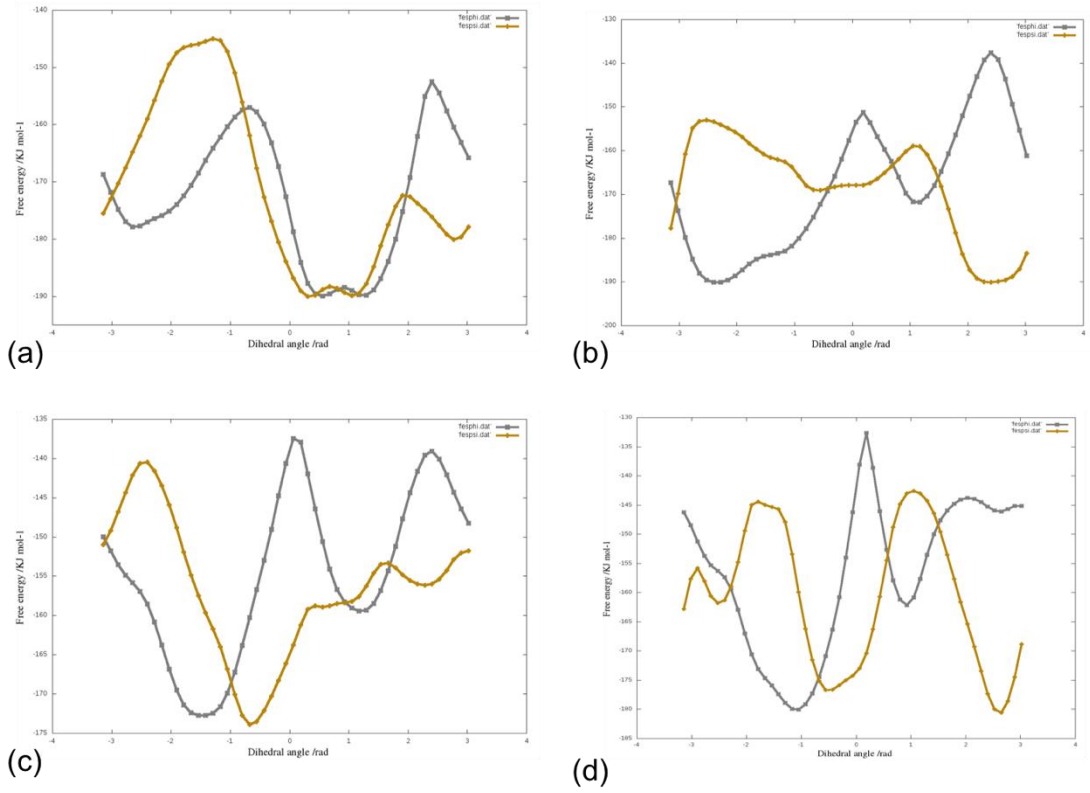
A2.3 The salt bridges formed during the trajectory of (a) phi0, (b) phi60, (c) phi120, (d) phi180, (e) phi240 and (f) 300 systems in Chapter 4. The label on the right-hand side shows the 3-letter code name of the residues with their index number. Each line presents the distance between the oxygen of the acidic residue (aspartic acid and glutamic acid) and the nitrogen of the basic residue (arginine and lysine).



A2.4 The bilayer thickness in a colour plot calculated with GridMAT-MD of (a) System_2 and (b) System_3 with a claudin-1 particle in Chapter 5 (x- and y- axis display the bin size).



A2.5 The estimate of the free energy with respect to φ (grey) and ψ (gold) dihedral angles of the (a) E48K, (b) S53E, (c) K65D and (d) D68S mutants:



A2.6 The reconstructed free energy profiles of the mutants (a),(b) E48K (c),(d) S53E and (e),(f) K65D mutants with respect to the dihedral angles φ and ψ respectively. The graphs present the last 20 estimates of the FES coloured differently.