Molecular Organization of Tight Junction Protein Strands: Molecular Dynamics Simulation of the Self-Assembly of Extracellular Domain Particles of Claudin 1

Fitsiou, E.[1], Noro, M.[2,3], Singleton, S.[2] Warren, P.[2] Anwar, J.[1,*]

[1] Chemical Theory and Computation, Department of Chemistry, Lancaster University, Lancaster, LA1 4YB, UK.

[2] Unilever R \& D Port Sunlight, Quarry Road East, Bebington, Wirral, CH63 3JW UK

[3] Current address: The Hartree Centre, STFC Daresbury Laboratory, Sci-Tech Daresbury, Keckwick, Daresbury, Warrington WA4 4AD UK.

1. Introduction

Tight junctions are cell-cell contact structures found in epithelial and endothelial tissues, located at the contact region between neighbouring cells, towards their apical side. They regulate the permeability of small molecules and ions through the intercellular space (paracellular pathway) by either blocking their passage or allowing some molecules with appropriate charge and size to go through. A functional tight junction barrier is critical to the physiology of the body. Its dysregulation can lead to pathologies such as inflammation, metastasis and edema [1]. For instance, mice that lack a key tight junction protein, claudin 1, die after birth due to excessive water loss across the skin. Tight junctions are also the target of several viruses including the hepatitis C virus and bacteria such as the bacterium *Clostridium perfringens* that produces the enterotoxin responsible for food poisoning. They are also targets of strategies for enhancing drug delivery of large molecules including proteins across the gastrointestinal tract. Further, there are numerous hereditary diseases that are linked with mutations of tight junction proteins, which include hypomagnesemia, deafness, neonatal sclerosing cholangitis with ichthyosis and familiar hypercholanemia.

An important research question is the nature of the molecular architecture of the proteins that make up the tight junctions, a better understanding of which would enhance our knowledge of the basic principles of paracellular permeability. At the molecular level, tight junctions are a multi-protein complex comprising transmembrane proteins, cytoplasmic plaque proteins, signalling proteins and adaptors that connect them to the actin cytoskeleton. The transmembrane proteins include primarily the claudins, tight junction associated MARVEL proteins (TAMPs), and JAMs. The claudins are considered to be the fundamental functional and structural elements of tight junctions as they alone can reconstitute tight junction-like strands in cells that normally lack these structures. Structurally, claudins look like an Olympic torch with a transmembrane region (the body of the torch) comprising a left-handed four-helix bundle and an extracellular part (the head of the torch) composed of two loops that are integrated via β-sheets (Figure 1). It is evident from Figure 1 that the extracellular domain (ECD) (the 'torch' head) is larger in diameter compared to the rest of the body.

Whilst we have structural data on claudins, how the tight junction proteins are organised to make up the tight junction is not wholly resolved. The molecular organisation of the tight junction claudins in their native environment i.e. when embedded in lipid membranes is currently inaccessible by experiment. This question of molecular organisation has been tackled

by molecular 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 tight junctions [2,3] using coarse grained models [4]. The CG approach enabled the relatively-slow molecular re-organisation of the claudins in the membranes to be tracked. CG models reduce the 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. CG models by their very nature involve a loss of atomistic resolution. Further, 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. Here, we employ molecular dynamics (MD) simulation using *atomistic resolution* (in contrast to CG representation) to examine the self-assembly of ECD particles of claudin 1 within a membrane. The ECDs are larger in diameter and laterally overhang much of the transmembrane domain, and hence are considered to be the primary determinants of the claudin assembly. 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-assembly in an *implied* membrane. Thus, the ECD particles are constrained to lie on a plane that serves as the membrane.

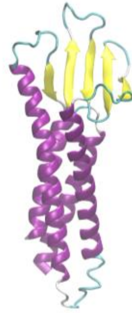

Figure 1. Secondary structure representation of Claudin 1

2. Methodology

We examined the behaviour of a *single*, full claudin particle embedded in a lipid membrane and the claudin 1 extracellular domain (ECD) in an *implied* membrane 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.

3. Results and Discussion

3.1 Claudin 1 embedded in a phospholipid bilayer maintains its structural stability in molecular dynamics simulation – The simulation of the single, full claudin particle embedded in a lipid bilayer revealed that the starting structure remains stable (on the basis of root mean square deviation (RMSD)) throughout the 100 ns trajectory. The secondary structure elements and the tilt angle of the claudin relative to the plane of the membrane are largely maintained. The ECD revealed some opening up of the structure. The conservation of the structure confirmed the quality of the claudin 1 model and the accuracy of the employed CHARMM36 force field.

3.2 Isolated ECD of claudin 1 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 showed that it exhibited greater flexibility and some structural deviation (greater opening up of the structure) relative to the ECD of the full claudin particle embedded in a bilayer.

3.3 Self-assembly of ECDs yields a network of strands – The randomly rotated domain particles, initially located on a grid very quickly come together forming dimers and small clusters, which then aggregate into a morphology that resembles the network of strands (Figure 2) as seen in electron microscopic images. 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 clustering occurring within the first 50 ns. The rapid clustering suggests a high chemical potential (a certain super-saturation) of the domain particles in the system.

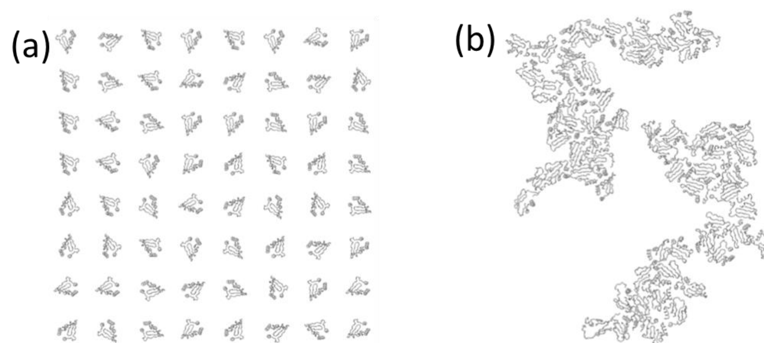

Figure 2. Self-assembly of ECDs. (a) Initial configuration of randomly oriented ECD particles on a grid. (b) Snapshot of the final configuration, 100 ns simulation time. Water molecules and ions are not displayed for clarity and protein atoms are shown in ribbon representation.

3.4 Emergent strands reveal a plethora of interfaces – The ECD particle-particle interfaces within the emergent strands were characterised in terms of respective rotation angles (about the z-axis) of each of the two domain particles involved in any given dimer interaction. The distribution gives the frequency of a specific combination of θ_1 and θ_2 being observed. The rotation angle distribution shows a wide range of values, reflecting the wide variety of observed

interfaces. There are also some particular hotspots, notably four, frequently observed dimers, which have also been observed by others.

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. We confirmed this by estimating the binding free energy between any two interacting ECDs via potential of mean force calculations. The binding free energies are generally large, being within the range of about -21 to -58 kJ mol^{-1}. This equates to 8-22 $k_{\beta}T$ at 310 K, where k_{β} is Boltzmann's constant. The significance of the binding free energies being at least an order of magnitude greater than $k_{\beta}T$ (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? Clearly, tight junctions based on a claudin-claudin binding free energy of a few $k_{\beta}T$ would be readily prone to failure.

3.6 Certain key residues play a significant role in stabilising dimers – The atomistic resolution enables us to characterise in detail the important interactions between the ECD particles. For each of the frequently observed dimer interfaces, there are particular residues that interact strongly and hence play a key role in stabilising the dimers. The simulations confirm the importance of the key residue lysine at position 65, which has been demonstrated by experimentalists to be essential for tight junction strand formation. 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 [5].

4. Concluding remarks

The ECD atomistic simulations reveal rapid aggregation and assembly of the ECDs to yield a network strand structure. The primary driving force for aggregation is the hydrophobic interaction resulting from the numerous hydrophobic residues on the ECD, with leucines 70 and 73 and phenylalanine 161 being key residues. The ECD-ECD interaction free energies ranged from -21 to -57 kJ mol^{-1} indicating strong ECD-ECD interaction. These values in terms of thermal energy translate to about 8-22 $k_{\beta}T$, the implication being that the emergent network strands are likely to be kinetically locked and not at equilibrium. The emergent structure is not linear but rather a branched network, consistent with experimentally observed morphology and previous CG simulations. The cis dimers show considerable variety in term of the relative orientation angles of the claudin particles that characterise the interacting interfaces. The redundancy in interaction interfaces explains 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 atomistic resolution reveals in detail the interacting amino acids that are 'key players', confirming the importance of atomistic resolution in such studies.

[1] Sawada, N (2013) Pathology International, 63,1-12.
[2] Irudayanathan, F.J. et al (2015) J. Phys. Chem. B 120, 77-88.
[3] Irudayanathan, F.J. et al (2017) Annals New York Academy of Sciences 1405, 131—146.
[4] Marrink, S.J. et al. J. Phys. Chem. B, 111, 7812—7824.
[5] Veshnyakova, A. et al. (2012) Annals New York Academy of Sciences 1257, 45-53.