8-2-2021

Mechanisms of Connexin-46 and -50 Intercellular Channel Function and Stability by Molecular Dynamics Simulations

Bassam George Haddad
Portland State University

Mechanisms of Connexin-46 and -50 Intercellular Channel Function and Stability by Molecular Dynamics Simulations

by

Bassam George Haddad

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Chemistry

Dissertation Committee:
Steve Reichow, Chair
Daniel Zuckerman, Co-Chair
Dirk Iwata-Reuyl
Albert Benight
Drake Mitchell

Portland State University
2021

Abstract

Gap junctions make up a class of intercellular channels that characteristically connect the cytoplasm of directly apposed cells through large assemblies, or plaques, constituted by a multitude of intercellular channels. Gap junction mediated intercellular communication is critical for a variety of physiological functions, from coordinating electrical impulses in the heart and brain to maintaining homeostasis in most tissues. There are 21 isoforms of connexins, the constituent subunit of the gap junction, expressed in a tissue dependent manner. Gap junctions formed from different isoforms exhibit distinct biophysical properties, such as gating kinetics and sensitivity, as well as unique permeability and selectivity to solutes suited to the needs of the cell. The molecular mechanism underlying the physiological roles of the gap junctions are poorly understood, due in large part to a lack of high-resolution structures.

The Reichow lab has recently elucidated the structure of two closely related gap junction channels, composed of isoforms expressed in the eye lens (Cx46 and Cx50) using single particle electron cryo-microscopy (CryoEM), providing a means for comparative structural analyses. However, these static structural models provide limited information on the dynamic behaviors that give rise to their complex functions. For my dissertation research, I have utilized molecular dynamics (MD) simulations – a computational microscope that is unrestricted to the limits of electron microscopy – to investigate the dynamic and thermodynamic mechanisms underlying gap junction permeability and selectivity. Using comparative equilibrium MD, I confirmed that Cx46 and Cx50 adopt a stable open state, distinct from the previously determined structure of Cx26; thus,

underscoring the need for high-resolution structures of multiple isoforms. From these simulations, I identified the n-terminal helix (NTH) as the site of isoform-specific behaviors of substrate selectivity. Furthermore, I characterized multiple putative ion binding sites, which were hidden from view in the CryoEM reconstruction due to their transient nature.

A technological advance in membrane protein CryoEM studies, known as a lipid nanodiscs, allowed for my colleague to elucidate a sub-2 Å structure of Cx46 and Cx50 – resolving ordered water molecules and concentric annular layers of lipids within the extracellular leaflet. Equilibrium MD simulations of Cx46 and Cx50 corroborated more than 80% of the solvent densities and recapitulated the extracellular leaflet stabilization – indicating a novel protein-induced gel-phase lipid transition. Moreover, the simulations explained the absence of resolved lipid head-groups in the ensemble reconstructions.

In collaboration with gap junction electrophysiologists, I utilized *in-silico* mutagenesis to swap residues within the NTH between Cx46 and Cx50 to investigate the molecular determinants underlying isoform specific conductance and gating properties of these isoforms. These simulations predicted a swap in junctional conductance, matching closely to experiment, and explained a loss of function in the NTH-swapped Cx46. Additional analysis using computational electrophysiology characterized conductance through gap junctions composed of Cx46, Cx50, and Cx46/50; revealing the permeation mechanism and thermodynamics underlying channel rectification and cation selectivity. In this dissertation, I demonstrate the power of molecular dynamics simulation as a tool which corroborates and extends our knowledge and understanding of gap junction structure and function gained through experimental CryoEM and electrophysiology.

To my parents Traci & George Haddad
Whose unending love, support, and patience made this dissertation possible
and
To my partner Léa Zola
For pushing me to reach high for my dreams

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Reference Sections

Chapter 1

Introduction to Gap Junction Structure and Physiology

1.1 Gap Junctions and Intercellular Communication

Gap junctions are a class of membrane bound proteins, that provide the unique function of directly connecting the cytoplasm of neighboring cells, which facilitate intercellular communication through diffusion of ions and other small molecules[1,2]. Physiologically, gap junctions are responsible for the electrical coupling in heart and brain cells[1] and maintaining homeostasis in most tissues[2]. Aberrations in gap junction expression, folding and/or regulation lead to a myriad of diseases throughout the body – including, deafness[1,3], blindness[4–6], cardiac arrhythmia[1,7], neurodegeneration[8,9], and cancer[3,10–12].

Gap junction intercellular channels have characteristically large pores, providing a pathway for intercellular permeation to a variety of chemical information. While once believed to be large passive channels which are unregulated[13], it is now understood that gap junctions are quite dynamic, gating in response to an array of cellular signals: such as, voltage[14,15], pH[16], and Ca^{2+} [17]. Furthermore, gap junctions are subject to post-translational modifications, e.g., phosphorylation[18], acetylation[19,20] and age-related truncation[21], which alter their biophysical properties (i.e. gating, conductivity, etc.)[21]. Though gap junction function has been extensively studied, the underlying structural mechanisms responsible for these dynamic functions are largely unknown.

In the past decade, high-resolution gap junction structures have provided a foundation upon which assessments of gap junction function can be made. Experimental

and computational studies have made great strides in understanding the mechanism of gap junction gating, and permeability and selectivity (perm-selectivity)[22–25]. Still, there are gaps in knowledge that need to be addressed, such as those that are the focus of this thesis: 1) What are the molecular mechanisms underlying ion permeation and isoform specific selectivity? And 2) How is gap junction structure and function influenced by the lipid environment? These questions aim to extend our understanding of gap junction physiology, and to aide in the development of potential therapeutics for human health and disease.

1.2 Physiological Roles of Gap Junctions

There are 21 connexin isoforms, which are expressed widely throughout the body in a tissue specific manner[1]. Connexins are constituent subunits that oligomerize to form a full dodecameric gap junction intercellular channel (Fig. 2C), each unique in their physiological roles[26]. Gap junctions electrically coupling cardiac cells and neurons and are responsible for synchronizing entire tissues, whether it's the beating of the heart[2], or an electrical synapse in the brain[16,27,28]. Additionally, gap junctions are critical for maintaining lens transparency[29] and homeostasis in most tissues[2,30].

In the heart, gap junctions facilitate the direct electrical coupling of cardiomyocytes', responsible for the synchronization of the beating heart[26]. The human heart beats approximately two billion times in the average human life span with extraordinary regularity[31]. Irregularities in the heartbeat[32] (e.g., arrhythmia) are a sign of illness – it is imperative that consistent and highly coordinated electrical impulses are directly coupled

between cardiomyocytes. This synchronization is achieved by the gap junctions, formed at the boundary of two apposing cells, which narrows to an intercellular distance of approximately 3 nm.

In the brain, gap junctions form the fast-acting synapses between neurons, known as electrical synapses. Gap junctions reduce the intercellular space between the pre- and post-synaptic cells to ~ 3 nm, as compared to the ~ 30 nm distance in the synaptic cleft[27,28] (Fig. 1A). The direct coupling eliminates the need for the diffusion of neurotransmitters across the large post-synaptic cleft, which can take anywhere up to a few milliseconds to transmit[33]. Electrical synapses conduct signals in the nanosecond to microsecond regime, orders of magnitude faster than its chemical counterpart[34,35]. What electrical synapses gain in speed and efficiency, they lose in modularity and strength of signal propagation, thus, they are relegated to more primal brain functions, such as the fight-or-flight response[35].

Gap junctions are not only important in electrically coupled cells, but also play important physiological roles throughout the body, including: general maintenance of homeostasis[2] and mediating signal transduction[26]. An example of this is in the eye lens, which is composed of living cells that require nutrients such as glucose, O_2, and antioxidants[36], the delivery of which is a role that blood plays throughout the rest of the body. However, the lens must remain transparent, and a vascular system would interfere with the transmission of light. Thus, this avascular organ utilizes a specialized “micro-circulatory system” to deliver nutrients and remove waste products[29]. This system uses the Na+/K+-ATPase, together with a range of transporters and channels, to bring

nutrients from the peripheral epithelial cells to the central core, while trafficking of ions, water and waste products from the core back outward is directed by the gap junctions[29] (Fig. 1A). All gap junction can be gated by voltage-potentials, under electrical stimulation[1,37]. However, unlike the gap junctions involved in electrical synapses and cardiac coupling, the lens gap junctions do not encounter action potentials in their physiological environment. Instead lens gap junctions utilize a Na+ gradient, established by Na+/K+-ATPases in the epithelium (i.e., lens-surface), to drive an outward radial flux of ions and water[29] (Fig. 1B).

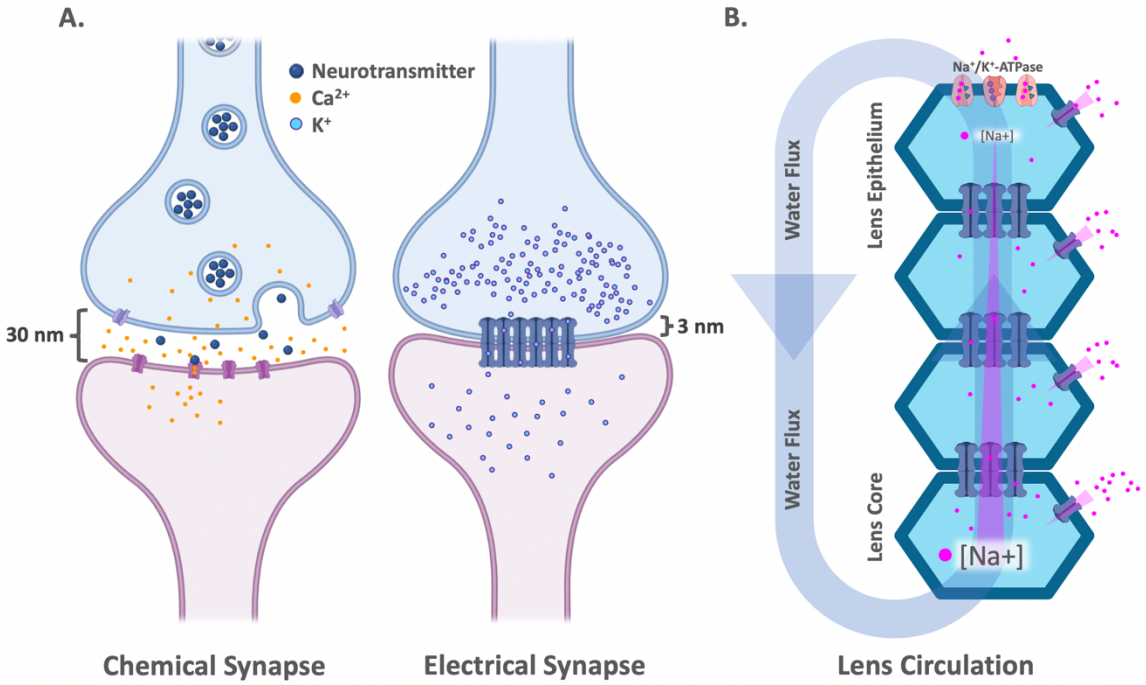


Figure 1. Gap junction physiology. *A) comparison of a chemical synapse and a gap junction mediated electrical synapse. The chemical synapse requires a Ca^{2+} release of neurotransmitters into the 30nm gap known as the synaptic cleft, where they bind to neurotransmitter receptors activating Ca^{2+} channels on the post-synaptic neuron. The electrical synapse allows for direct diffusion of ions along their electro-chemical gradient, passing through the 3 nm gap junction. B) Gap junction mediated, water microcirculatory system in the eye-lens. Na^+/K^+-pumps remove intracellular Na^+ at the epithelium, and Na^+ leak channels allow for Na^+ influx which creates a $[Na^+]$ gradient, driving water flux through the gap junctions, carrying waste products out of the core of the lens. Created with BioRender.com.*

As described here, the gap junctions play fundamental physiological roles throughout the body, therefore aberrations in expression, trafficking, and/or function of gap junction communication may result in a myriad of human diseases[3,8,38]. In the heart, aberrant gap junctions produce cardiac arrhythmias, In the brain, impaired gap junctions result in neurodegenerative diseases. Abnormalities in the expression of gap junctions in the lens and ears result in congenital cataracts (i.e., blindness), and deafness respectively[4,5,39–41]; and more broadly, gap junctions are implicated in many cancers[3,10,42]. Therefore, a complete understanding of the structure, and dynamics which underlie the physiology roles of gap junctions is essential – ultimately working towards the development of new therapies as well as the engineering of tissue-specific drug delivery systems[43].

1.3 Structure and Function of Gap Junction Intercellular Channels

The connexin nomenclature utilizes the molecular weight (in kDa) of the subunit, e.g., connexin-46 (Cx46) has an apparent mass of 46 kDa. Connexins have characteristic biophysical properties: gating sensitivity, conductivity, and perm-selectivity that vary widely between isoforms[44,45]. Multiple isoforms are typically expressed in a single tissue, allowing for modularity in function that can be tuned to the physiological needs of the cell.

Structurally, six connexins come together in the ER or Golgi – depending on certain genetic factors[46] – to form a hemichannel. Two hemichannels from apposing cells then dock together to form a dodecameric intercellular channel[2]. Gap junctions can form via

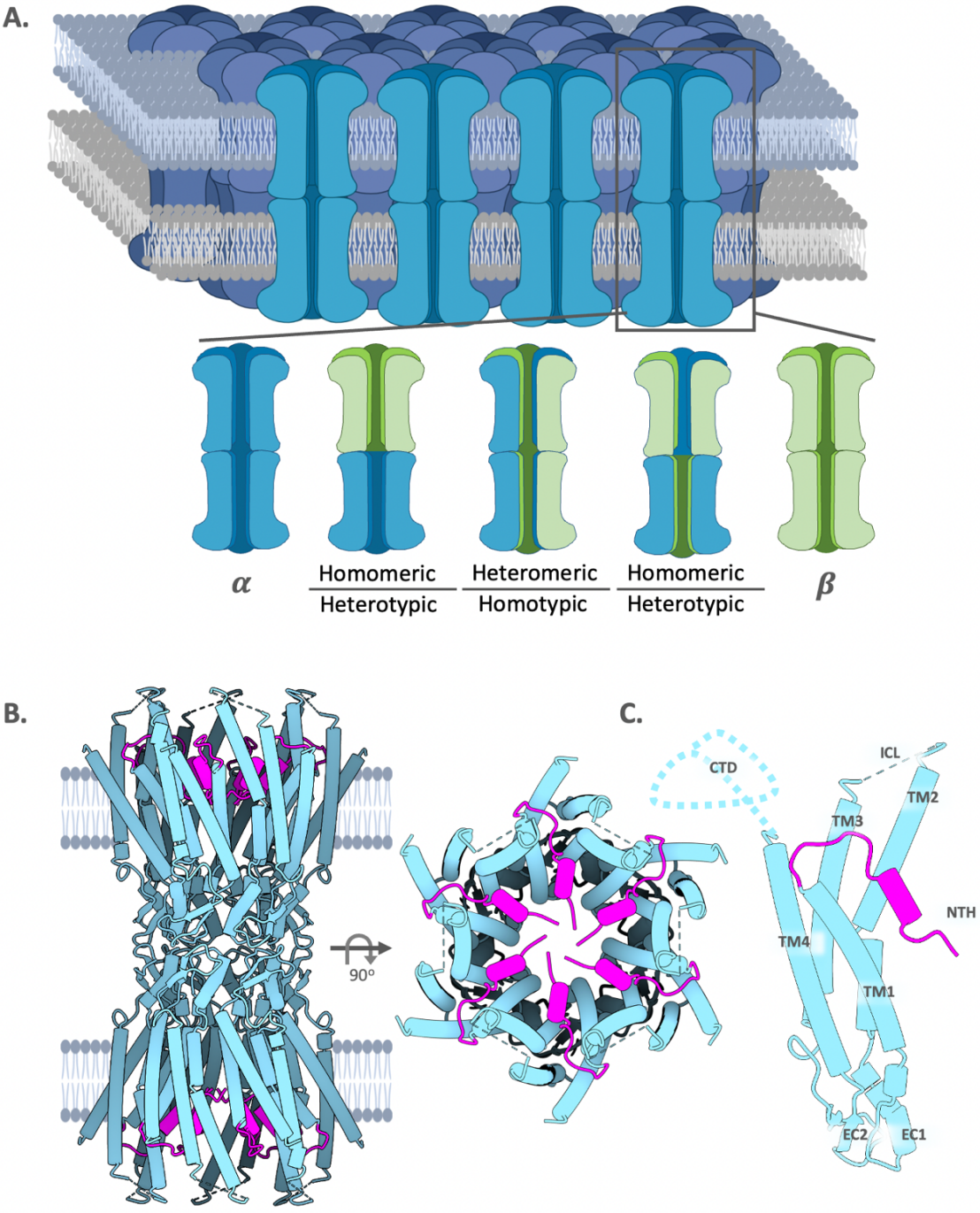


Figure 2. Gap junction structure. *A) Gap junction plaque showing hexagonal packing of intercellular channels (top). Different configurations of heterotypic and heteromeric gap junctions composed of two distinct isoforms (bottom). B) Structure of dodecameric Cx26 intercellular channel with NTH highlighted in magenta; a 90º rotation reveals the 1.4 nm pore. C) Structure of Cx26 monomer, revealing the topological structure of the transmembrane helices.*

the docking of two identical hemichannels or from distinct isoforms, known respectively as homotypic and heterotypic gap junctions (Fig. 2A)[47]. For further modularity and complexity, a hemichannel can be composed entirely of a single connexin isoform (i.e., homomeric), or mixture of isoforms (i.e., heteromeric) (Fig. 2A). The biophysical properties of the resulting heteromeric hemichannel are often intermediate to its parent connexins, but sometimes yield unexpected gain/loss of function[48]. For example, the isoforms Cx26 or Cx32 are both permeable to cAMP, however a heteromeric Cx32/26 channels are effectively impermeable to cAMP[49]. There are 196 permutations for two connexin isoforms to combine and form a gap junction intercellular channel[50]. However, a recent study investigating heteromeric channels formed from Cx26 and Cx30 found a 3:3 (Cx26:Cx30) ratio to be most common[51] – still the rules governing heteromeric stoichiometry are not fully understood.

There are indeed rules governing which isoforms possess heteromeric or heterotypic compatibility. Heteromeric compatibility of connexin isoforms have been organized into two primary groups (R & W) and a third 'other' group, which have been empirically determined by experiments involving co-immunoprecipitation, MS/MS and FRET[52]. Though there appears to be no discernable pattern on inter-subunit interactions, genetic markers in the third transmembrane domain (TM3) correlate well to the empirically derived grouping of connexin isoforms that are heteromerically competent. Connexins with a signature arginine in TM3 (R-type), are stabilized as monomers in the ER, and don't oligomerize until they reach the trans golgi network (TGN)[46]. Whereas the isoforms containing a di-tryptophan motif (W-type) oligomerize in the endoplasmic reticulum golgi

intermediate compartment (ERGIC)[46]. Given that these biosynthetic routes just sort the two groups into folding regions, it is still unclear if there is a mechanism by which particular subunit arrangements may be imposed during the assembly of heteromeric hemichannels.

Unlike the bio-synthetic pathway of co-assembly, heterotypic compatibility is shown to be more structural in nature. Heterotypic compatible isoforms have been similarly separated into distinct groups (Group I/II)[46]. The determinants of Group I/II compatibility appear to be localized to the second extracellular domain (EC2). Genetic analysis into the sequence of EC2 found that there are signature motifs for Group-I (K(R)/N/D), and Group-II (V/H/D) connexins[46]. White et al. demonstrated that by swapping the EC2 of Cx50 with the EC2 of Cx46, the new chimeric Cx50 formed a functional heterotypic channel with Cx43[53]. Notably, there are exceptions to the Group I/II rules. For example, Cx46 and Cx50 are Group-I and compatible, and Cx43 is Group-II yet forms heterotypic junctions with Cx46 but not Cx50[53]. Although it is understood that EC2 confers docking specificity, the specific residues (or contacts) that give rise to heterotypic specificity are still not fully understood.

Following decades of low-resolution structural studies of the gap junction by electron microscopy and AFM, the first atomic structure was resolved by X-ray crystallography in 2009 by Maeda et al.[54]. The Cx26 gap junction channel provided the first atomic-detail insight into connexins (PDB-ID: 2ZW3), resolving many outstanding questions (e.g., which residues are pore-lining etc.) (Fig. 2B). The 3.5 Å[55] crystal structure reveals an n-terminal helix (NTH), four transmembrane helices (TM1-4), two extracellular

domains (EC1-2), as well as an intracellular loop (ICL) and c-terminal domain (CTD) that are disordered and not observed in the crystal structure. Topologically, the NTH, ICL, and CTD are on the intracellular side of the membrane, with the ICL connecting TM2 and TM3 and ending with the CTD extending from TM4 (Fig. 2B). The CTD is the site of largest genetic variation amongst the connexin isoforms and is thought to play a regulatory role in these channels[4]. The extracellular loops EC1 and EC2, which connect the pore-lining TM1 to TM2 domains and membrane facing TM3 to TM4 domains respectively, are characteristically stabilized by three universally conserved disulfide bonds. TM3 and TM4, which have a 60^{o} slant relative to the bilayer-normal, interact directly with the membrane and may form inter-channel contacts in the context of a full gap junction plaque[54].

1.3.1 Role of the N-terminal Helix in Voltage-Gating

The N-terminal helix (NTH) is positioned at the opening of the pore and forms the constriction site at either end of the intercellular channel, making it an ideal candidate for substrate selectivity and voltage sensing. Dual whole-cell and patch-clamp electrophysiological studies of gap junctions have identified the NTH as the voltage sensor, and the putative voltage gate[56,57]. Single channel and whole-cell currents provide information on the magnitude of the permeants, but also the voltage dependence on opening or closing (i.e. gating). There are two known mechanisms of voltage-gating in connexins, a slow voltage-gate and the so-called fast-voltage gate[58] – the structural dynamics underlying these distinct gating mechanisms are heavily disputed[58].

The slow gate is characterized by a series of milliseconds long transitions from the fully open state to a fully closed[58]. Cysteine scanning studies identified the second extracellular loop (EC2) as the domain involved in gross conformation changes[59]. The slow gate has thus appropriately been nick-named the loop gate, typically associated with hemichannel function. The loop gate responds to a variety of chemical stimuli, such as pH, Ca^{2+}, CO_2, and arachidonic acid[60,61], in addition to being sensitive to the transmembrane voltage[23,58]. Given the dual-membrane nature of the gap junction intercellular channels, there are three voltages to consider, the voltage drops across each respective membrane (V_{m1} & V_{m2}), and the potential difference between the two cells (i.e. the transjunctional voltage, V_j)[58].

The fast voltage-gate of an intercellular channel, unlike the loop-gate, responds only to the transjunctional voltage (V_j), and gates on the sub-millisecond timescale[58]. Functional mutagenesis studies of a variety of gap junctions have identified the n-terminal helix as the 'fast' voltage sensing domain[56]. This was demonstrated in an elegant experiment by Verselis et al., where the authors were able to reverse the gating-polarities of the two parent isoforms by swapping the n-termini of two isoforms Cx26 and Cx32[62]. Later it was discovered that the addition of a negative charge to the 2nd residue of Cx32's NTH reverses its gating polarity, highlighting the 2nd residue as being critical for voltage sensing[22]. The same study also indicated the 5th, 8th, and 10th residues as being sensitive to V_j. The NTHs of different isoforms vary in total net charge, including the charge of the 2nd residue, altering the V_j sensitivity and even molecular permeability among isoforms[63].

The molecular mechanism of V_j-gating, while believed to involve the movement of the 2nd n-terminal residues towards the center of the pore[58], is still poorly understood. Homomeric/homotypic gap junctions are open near V_j = 0 mV, and close at large negative and positive voltages due to the symmetry of the channels[64]. V_j-gating is initiated when the intracellular voltage-potential is opposite of the charge at the 2nd (or 3rd) position, making homomeric gap junctions sensitive to a single polarity. Purnick et al. demonstrated that heteromeric hemichannels composed of Cx32 and the Cx32N2E mutant displayed bi-polar voltage sensitivity, *i.e.,* single channel gating at both positive and negative voltages[22]. To understand the effect of heteromeric assembly on gating polarity, Oh et al. showed that the degree of bi-polar sensitivity was dependent on the ratio of the constituent monomers – including hemichannels which have a Cx32 to Cx32N2E ratio of 5:1[65]. This dependence on the isoform ratio demonstrates that only a single NTH is required to initiate voltage gating. These studies underscore the complexity and modularity of gap junctions; however, the molecular underpinnings are still unknown.

1.3.2 Conductance and Selectivity

In addition to voltage-gating, the NTH is expected to be involved in selectivity, as well as conductivity. The pore lining residues define the shape, the charge, and therefore the selectivity of the channel. The gap junction permeation pathway can be divided into three sections: the intracellular vestibule, the pore entrance (or selectivity filter), and the extracellular vestibule. The intracellular vestibule is defined by the intracellular extensions of TM1 and TM2, as well as the ICL. The pore entrance is defined by the NTH,

which is folded into the pore against TM1-2 and forms the constriction site of the intercellular channel. The extracellular vestibule, located at the center of the channel, is formed by the first extracellular domain (EC1).

Unlike highly specific channels such as aquaporins and voltage-gated potassium channels, the pathway of solute permeation in the gap junctions is not so clearly defined (Fig. 3A). In comparison to these canonical channels, the pore of the gap junction is a large (~1.4 nm in diameter) fully solvated cavity, able to accommodate a variety of molecules up to ~1 kDa in size[45] (Fig. 3B).

Unitary conductance of gap junction channels range from as few as 10 pS[66], to 300 pS[67]. For comparison, the high-fidelity KcsA channel (Fig. 3A (bottom)) displays single-

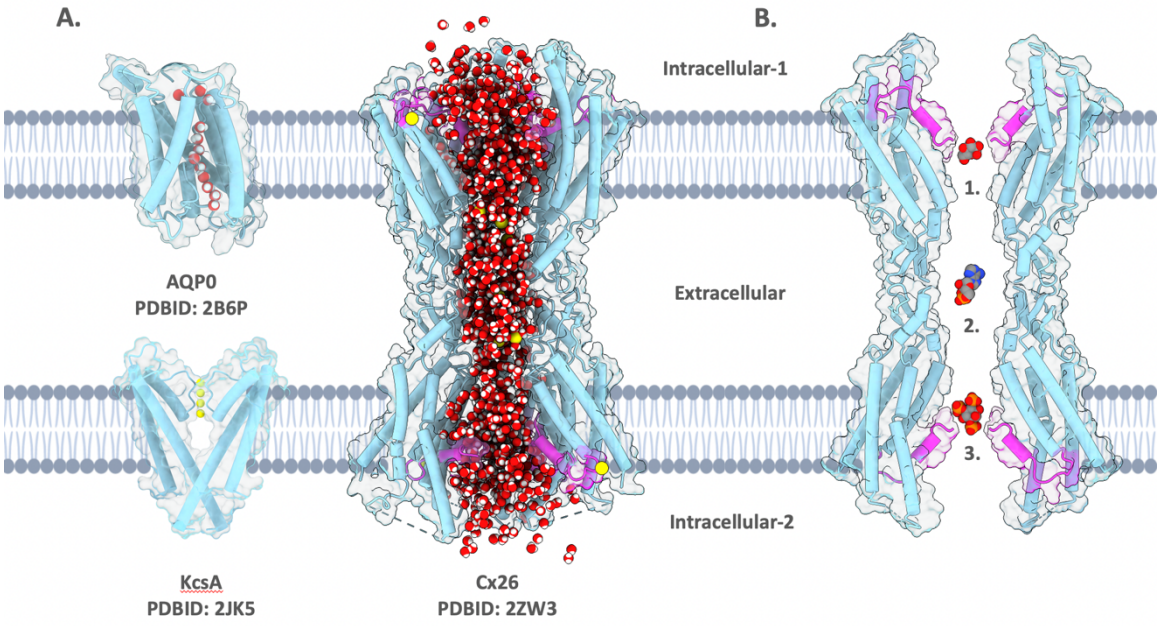


Figure 3: Comparison of "typical" membrane channels to Cx26. *A) Canonical "narrow" pore channels, aquaporin-0 (AQP0, top left) and the bacterial voltage gated K^+ channel (KcsA, bottom left) displaying highly selective pores with single-file substrates. Cx26 gap junction (right) displays a large pore, filled with a bath of solvent. B) Cx26 gap junction with three typical permeants: 1. Glucose, 2. Cyclic AMP (cAMP), and 3. Inositol triphosphate (IP3).*

channel conductance around 96 pS[68]. To gain insight into the determinants of gap junction channel conductance, the Bai group generated a Cx50 homology model from the Cx26 crystal structure. Cx50 forms a high conductance channel (~200–220 pS)[69]. In this study, they mutated the Gly46, a predicted pore-lining residue in EC1, to aspartate, increasing the negative surface-potential within the channel. This led to an increase in unitary conductivity of Cx50 from ~200pS to ~260pS. Based on these results, they concluded that the surface potential, particularly in EC1, is a determinant of unitary conductance[70]. To further test this hypothesis, they generated multiple neutral-to-negative mutations at the NTH and different regions in the pore-vestibule. The general finding was that increased negative potential within the pore results in increase conductance through these cation-selective channels. Remarkably, a mutation deep within the pore-vestibule (V53E) only effected conductance when combined with an n-terminal mutant G8E[71]. This finding tells us that the NTH functions cooperatively with other regions of the channel to define the permeability characteristics.

In the permeation of small ions, gap junction cation-to-anion selectivity ranges from being relatively anionic selective (e.g., Cx31.3 ~13:1)[72] or non-selective, to relatively cationic selectivity (e.g., Cx46: ~10:1)[23]. Ionic selectivity is determined by calculating the reversal-potentials – the voltage at which 0 pA occurs – of hemichannels under asymmetric ionic condition (*e.g.,* $[Na^+]_{in}/[Na^+]_{out}$ = 5/1). The Goldman-Hodgkin-Katz equation can then be applied to determine permeability ratios for any assayed ionic species[73]. This technique has been used to elucidate the “cation-series”, or the order of cation permeabilities ranging from high to low. The cation series of Cx46 hemichannels

largely follows ionic-mobility trends, i.e., $Rb^+ > Cs^+ \approx K^+ > Na^+ > Li^+$ [23], consistent with the idea that general-diffusive forces determine cation selectivity in connexin hemichannels. While cation series have only been experimentally determined in hemichannels due to experimental limitations of the dual whole-cell experiments, relative cation-anion selectivity can be determined using charged dye-transfer experiments[74,75]. While it is often assumed that the gap junction cation series would match hemichannels, structural rearrangements of the pore-lining residues upon hemichannel docking and their effect on selectivity are currently unknown.

Remarkably, despite the large pore size, gap junction channels are able to confer moderate to high selectivity for small molecules (Fig. 3B). The classic approach to studying molecular permeation in gap junctions is to use a dye transfer assay – where a fluorescent dye is injected into one of a pair of gap junction coupled cells and fluorescence intensity from both cells are recorded over discrete periods of time. Under conditions where gap junctional coupling is established, dye molecules can pass through the gap junction into the second cell. There are many different dyes ranging in size, shape, and charge from the negatively charged Lucifer Yellow, to the positively charged ethidium bromide that are used to probe channel selectivity[76]. More recently, fluorescence-based cell-reporter assays have been developed to monitor the coupling efficiency to other physiologically relevant molecules (such as cAMP, ATP and IP_3)[77,78].

The transfer of small molecules through gap junction channels occurs on the milliseconds-minutes timescale, orders of magnitude slower than ion conductance, indicating that the permeation mechanisms are more complex for small molecules

compared to ions. Furthermore, it has been well-established that the selectivity of small molecules does not necessarily reflect permeation/selectivity behaviors toward small ions[79]. The reasons for this are because in the size, shape and anisotropic distribution of charge on molecular messengers alters their hydrodynamics and can affect how the molecule interacts with pore-lining residues. Gap junctions are conduits for large secondary messengers ranging in shape, size and charge (e.g., cAMP, IP3, etc.) in an isoform specific manner[49,80]. Despite close sequence similarity and ion permeation characteristics exhibited between Cx46 and Cx50, the two isoforms are diametrically opposed in their permeability to cAMP; with Cx50 being completely impermeable to cAMP[78]. Intriguingly, shape, size and charge do not necessarily correlate to molecular permeability which indicates that connexins play larger role in perm-selective mechanisms beyond that of being a large, charged tunnel. The discrepancies in the rules underlying perm-selectivity for ions and molecules underscores the need for high-resolution mechanistic investigations into these processes.

1.3.3 Insights into Gap Junction perm-selectivity by computational modeling and molecular dynamics simulation (Before 2018)

Over the last few decades, connexin hemichannels and gap junctions have been heavily studied, and their biophysical properties extensively characterized. Indeed, there is much to be discovered regarding molecular mechanics and dynamics, however, *in-situ* electrophysiology studies lack the spatial resolution required to observe the elusive mechanisms. Since the elucidation of the structure for Cx26 in 2009 by Maede et al. [54,81],

there has been an explosion of computational studies aimed at deciphering the underlying molecular dynamics responsible for the exhibited perm-selectivity of these channels. The Cx26 structure became the archetype for gap junction mechanistic studies and was used for homology modelling of other isoforms[71,82,83]. Such structural studies provide insight beyond that of a static structure and allow observations of the atomic motions underlying ion binding and facilitating molecular permeation.

Despite the fact that the Cx26 crystal structure is that of the full gap junction (Fig. 2A), many of the computational studies were done on the hemichannel, assumed to be well represented by half of the gap junction channel structure. This assumption is based on the observation that hemi-channel conductance is effectively twice that of the full channel, consistent with half of the resistance in a hemichannel with little structural change between these two states[64]. Practically, the computational cost of simulating hemichannels is also favorable.

In 2012, a study lead by Francesco Zonta described the pathway and energetics of potassium (K^+) permeation through Cx26 and Cx30 hemichannels using molecular dynamics simulation[83]. The authors built a Cx30 homology model from Cx26, based on the high sequence identity (77%) between these isoforms. When comparing the equilibrated MD models, the authors found that both structures remained largely the same in the 40ns simulations, as judged by their root mean squared deviations (RMSD) from the starting structures, with the largest structural variations appearing in the NTH. During the simulation, the pore-diameter formed by the NTH increases from < 1 nm in the crystal structure to ~ 1.2 nm in both Cx26 and Cx30. The Cx26 NTH displayed a marked

increase in conformational variability as compared to Cx30, despite the near identical nature of the two NTH sequences, differing only in positions 7 (Gln_{Cx26} & His_{Cx30}) and 9 (Ile_{Cx26} & Phe_{Cx30}). Beyond the NTH, the most critical differences in the two pore profiles are at residues 41 (Lys_{Cx26} & Gln_{Cx30}) and 49 (Ala_{Cx26} & Glu_{Cx30}), significantly increasing the negative potential in the pore vestibule for Cx30 relative to Cx26. Although the pore-vestibule is predominantly negative in both isoforms, the positively charged pore-entrance precluded K^+ ions from permeating either hemichannel for the duration of the simulations. In contrast, both channels experienced Cl^- permeations, despite the fact that Cx26 is known to be cation selective[84]. To better understand the permeation of K^+, the authors pulled the ion through the channel with an externally applied force (i.e., steered-MD; see Chapter 2.3). The work (*i.e.,* energy) profiles of a K^+ being pulled through the Cx30 hemichannel show a single deep energetic well, corresponding to Glu49, whereas Cx26 displayed two shallower energy wells, separated by a peak at Lys41. The estimated *escape-probability,* calculated from the work profiles, correctly predict the relative unitary conductance of Cx30/Cx26 (~1.72_{MD} vs. ~1.45_{Exp}).

Zonta et al. continued their work on the perm-selectivity of Cx26 in the years immediately following their first publication. In their 2013 study on the Cx26 hemichannel, they analyzed the permeation of calcein, a negatively charged fluorophore that is commonly used in dye-transfer studies of gap junctions[12]; demonstrably permeable through Cx26. Molecular permeation occurs on a timescale seven orders of magnitude slower than ionic currents (~10^9 ions/s vs. ~10^2 calcein/s)[85]. To overcome these timescale barriers, they used a common enhanced sampling technique known as

umbrella sampling (chapter 2.3). The calculated energetic barrier for calcein was up to ~27 kcal/mol, a value too high for permeations to occur. To further investigate the effects of charge on the substrate, they parameterized a neutrally charged calcein molecule, which resulted in a permeation barrier that was reduced to ~12 kcal/mol, still insufficient for spontaneous permeation. As a point of comparison, the peak K^+ energy barrier calculated for Cx26 was calculated to be ~1 kcal/mol[83]. Their work led them to the conclusion that the Cx26 crystal structure is not that of the *fully open channel*. The validity of this statement is controversial and raises questions as to the chemical and physical nature of the NTH.

In their following work, Zonta et al. used comparative MD simulations of WT Cx26 and a deafness-causing mutation, M34T[86]. Despite the mutant's pathogenicity, *in-vivo* studies showed it could form hemichannels in the cell membrane that retained 11% of WT activity[87]. The 34th residue of Cx26 is situated in TM1, directly behind Trp3 of the NTH, leading to the hypothesis that M34T disrupts the open state of the NTH. To measure the stability of the open state, they calculated an eccentricity coefficient of the pore entrance, asserting that deviations from an ideally symmetric hexagon is less stable – the mutant was significantly more asymmetric than WT. To assay the destabilizing phenomenon, the authors used an applied force, similar to the ion-pulling experiments, to pull the NTH away from the pore lumen. It took a greater force to remove the WT NTH, indicating the reduced hydrophobic contacts of Thr34 and n-terminal Trp3 as the determining factor of NTH destabilization. This work directly corroborates structural studies of Cx26 which implicate the NTH as the gate in a closed-state, destabilized by the M34A mutation[88,89].

Computational studies by Yun Luo and Andrew Harris extensively characterize the permeation mechanisms of di- and tri-saccharides through Cx26[25]. These studies expand on the calcein work done by Zonta et al. in two key-ways. First, the use of the permeable di-saccharide and impermeable tri-saccharides allows them to directly observe determinants of molecular selectivity. Second, and most importantly, this was the first molecular permeation study to add a neutralizing acetyl-group to the n-terminus – in accordance with mass-spectrometry analyses of native connexin-26[19,20]. As with the calcein study, Luo et al. utilized umbrella sampling to calculate the PMFs for both saccharides, allowing them to observe the energetics of both the permeable and impermeable solutes – difference in peak-barrier was ~2.5 kcal/mol[25]. The peak energy barrier for both saccharides resided within the pore-vestibule near EC1, below the NTH constriction region – a region known as the parahelix. The impermeable tri-saccharide had a similar free energy at the narrowest region of the pore, the NTH.

To better understand the role of the pore-vestibule in the saccharide permeation, Luo et al. analyzed the effects of the permeants on the pore width and found that throughout nearly the entire length of the channel, hemichannels with the permeants present displayed a wider pore. This effect was especially pronounced near the NTH – the variability of the NTH being stabilized through interactions with the saccharides. The most constricting regions of the pore corresponded with increased hydrogen bond interactions, these are however the largest barriers for the permeants. Perhaps more interesting is the effect of Cx26 on the structure of the permeants. The conformational variability of the disaccharide in bulk water and throughout the channel is largely the same – in contrast,

the trisaccharide was forced to sample a compact state otherwise absent in bulk solvent. These data taken together suggest that specific interactions within the channel hinder the trisaccharide from sampling the unfavorable compact conformation required for permeation.

To better understand the contributing factors to the energetics of permeation, the authors performed a force decomposition of the PMF throughout the channel. In addition to interactions with the protein were the interactions with water, either favorable or not, depending on the pore region. Unlike canonical potassium ion channels, like KcsA[90–92], molecular and ionic permeants remain mostly solvated throughout the channel – eliminating the energetic cost of desolvation through a "selectivity filter". Unexpectedly, potassium ions made a minor, albeit significant contribution to the PMF stabilizing the saccharides as they passed through their energetic peaks. This is phenomenon is unlikely to occur in narrower channels and may play a larger role with charged permeants.

This interplay between enthalpic interactions between the permeant and Cx26, and the conformational entropy of the permeant, underscore the complex nature of molecular perm-selectivity in gap junctions[25]. Unlike more specific ion channels the potassium, saccharide and calcein PMFs displayed broad energetic wells and peaks, indicating a lack in specific binding throughout the channel. Despite this, Cx26 still confers a certain degree of selectivity.

1.3.4 Insights into gap junction gating by computational modeling and simulation (Before 2018)

The gap junctions have two voltage-sensitive gates: the fast n-terminal gate, and the slow loop gate. The slow gate is responsible for closing both hemichannels and some full intercellular channels[58]. As discussed in section **1.3.1**, the slow gate is sensitive to transmembrane voltage and extracellular Ca^{2+} and has been localized near the hemichannel docking region. The ability for both hemichannels and gap junctions to participate in slow gating necessitates that the voltage sensor(s) must be luminal facing residues that are not obstructed by the docking interface. Computational and experimental studies have worked to elucidate the residues responsible for gating in hemichannels; however, defining the precise mechanism has remained elusive.

Prior to the elucidation of the Cx26 intercellular channel[54], the subdomains associated with slow voltage gating were identified through mutational studies of Cx32-43EC1, Cx46 and Cx50[93]. Research from the labs of Thaddeus Bargiello and Miduturu Srinivas identified that conformational changes in the parahelix and EC1 are required for loop gating, leading to the hypothesis that voltage induced destabilization of the parahelix-EC1 region results in occlusion of the pore[93].

In 2012 Taekyung Kwon, a computationalist in Ted Bargiello's lab, used MD simulations and *in-silico* mutagenesis to visualize the mechanisms underlying the slow gate initiation. Their simulations identified Van der Waals (VdW) and electrostatic networks which stabilize the parahelix. Both the VdW and electrostatic networks contained residues conserved across many isoforms. There were two critical residues

within the VdW network, V43 and I74 that were chosen for *in-silico* mutagenesis. V43 was identified as a residue within the loop gate, and I74 is a centrally connected residue which forms a stabilizing bridge between V43 and F191[94]. Each of these residues are conserved between Cx26, Cx46, and Cx50 – the latter two isoforms where the subjects of previous experimental work by the same group[95]. A double mutant, Cx26-V43A-I74V, reduced hydrophobic volume, and the overall stability of the conserved VdW network. Importantly, structural fluctuations identified from the VdW network were concurrently correlated to fluctuations in the energetics and dynamics of the electrostatic network. This relationship provides a model of voltage gating; fluctuations to the electric field near charged, pore-lining residues (e.g., E42) would alter the configuration of the electrostatic and VdW networks, destabilizing the parahelix and therefore closing the channel[96]. This hypothetical model still requires experimental structural validation, as the timescale of gating is beyond that of current all-atom MD simulation capabilities.

The next structure of a gap junction came from Brad Bennet in Mark Yeager's laboratory – the Ca^{2+} bound structure of Cx26[97]. Gap junctions and hemichannels are known to be gated by intracellular and extracellular Ca^{2+}, respectively[58]. Remarkably, the structure of the Ca^{2+}-bound Cx26 was nearly identical to that of the previously reported apo-Cx26 structure. The two primary differences being that the NTH is unresolved in both the apo- and Ca^{2+}-bound states reported by the Yeager lab[97], and conformational changes in Ca^{2+}-coordinating residues, namely a near 180^{o} rotation of E47. Interestingly, the three coordinating residues (E42, G45, and E47) were previously identified as critical components of the described electrostatic network implicated in Ca^{2+}-gating of the Cx26

hemichannel[96]. MD simulations of the Ca^{2+}-free and Ca^{2+}-bound Cx26 structures revealed that overall, Ca^{2+} coordination is long-lived (≤ 100ns) and creates an electrostatic barrier for K^+ permeation[97]. Coordination of the bound Ca^{2+} with G45 was highly variable in the simulation, despite Ca^{2+} being stably bound, thus an external restraint was applied to ensure complete coordination throughout the simulation. Energetically, the K^+ PMF of the Ca^{2+}-bound exhibited an increased peak barrier by ~ 1.5 kcal/mol; Cl^- PMFs displayed a ~0.7 kcal/mol reduction in their peak barrier. The authors noted that the PMFs of K^+ and Cl^- permeation were unaffected by the additional G45-Ca^{2+} restraint.

The mechanism of Ca^{2+} binding to Cx26, and its influence on hemichannel gating was further explored by the Contreras laboratory. They performed MD simulations with Ca^{2+} freely diffusing throughout the simulation, albeit initially placed near the extracellular entrance of the hemichannel. Similar to the previous MD work by Bennet et al., it was found that once Ca^{2+} associated with the protein, its interaction was long lived. However, the Ca^{2+} formed a bond with D50 (as compared to E42 and E47 in Bennet et al.)[97]. This and previous work from the Contreras group identified an essential salt-bridge, D50-K61, which leads to enhanced kinetics of voltage-induced hemichannel closure in the absence of Ca^{2+} when disrupted and loss of Ca^{2+} dependence[98].

Taken together, these studies paint a picture of the initial steps of the voltage induced loop gate. Each study identified a conserved electrostatic network that can be affected by voltage – two of which propose a mechanism by which Ca^{2+} binding sensitizes loop gating by destabilizing this network. It is noted in all of the MD studies that the timescale of loop gating is on the order of milliseconds to seconds, and thus acknowledge

the limitations of sampling in their simulations. There are, however, further caveats to consider. First, all but one of the simulations performed are on the hemichannel, despite the lack of an experimental hemichannel structure. Simulations containing Ca^{2+} used nonphysiologically large concentrations of Ca^{2+} – and there were no instances of Ca^{2+} unbinding, an indication of under-sampling. Despite the listed caveats, there is a beautiful consistency in the results from the three independent research groups.

1.3.5 Limitations of the X-ray Model of Cx26.

The NTH is responsible for voltage sensing and fast V_j-gating[56] (Fig. 2B), determining gating polarity, and forms the intercellular opening of the pore – the narrowest region of the channel[22,65,99]. The NTH is a target for modification through post-translational modification (PTM), with each isoform being subject to the PTMs that are specific to their sequence[21,100]. For example, Cx46 and Cx50 have Met1 removed post-translationally, whereas this site is maintained in Cx26[19], as dictated by the sequence specificity of aminopeptidases. Mass spectrometry data have demonstrated that the NT methionine of Cx26 is also, at least partially, acetylated[19]. Such modifications, in addition to isoform differences, alter the charge, shape, and perm-selective properties of gap junction pores[20].

Mass spectrometry of Cx26, expressed in HeLa cells, identified PTMs including acetylation of Met1 as well as seven lysine in the ICL[19]. In addition to these neutralizing modifications by acetylation, γ-carboxyglutamation were identified in the ICL and within the pore that introduce negative charges. None of the modifications were found in the X-

ray structure, likely due to dynamics of the ICL and related sidechains, or low abundance of the modification. Such charge altering modifications heavily influence the perm-selectivity and conductance properties of connexin channels, as discussed in section **1.3.2**.

The effects of these modifications were explored computationally using a combination of all-atom MD and Brownian dynamics (BD) simulations by Taekyung Kwon et al. in the Bargiello lab[20]. The authors first completed the Cx26 model by adding residues missing from the crystal structure, taking Met1 coordinates directly from NMR studies of the NTH. With the addition of Met1 to the crystal structure the pore narrowed below the diameter of a hydrated potassium ion – considered by the authors as evidence that the crystal structure is not that of an open channel. After an initial equilibration, the structure was largely un-altered, with exception to the NTH, which repositioned itself against TM1. An average structure was calculated from the equilibration, with the largest variance occurring with the NTH – the authors then selected a structure from the trajectory most similar to the average for subsequent analysis.

Using grand-canonical Monte-Carlo Brownian dynamics (GCMC/BD)[101], an ion-channel conductivity simulator developed by Benoit Roux, Kwon and Bargiello were able to rapidly simulate current-voltage (I-V) relationships, and K^+/Cl^- permeation energetics of Cx26 with various combinations of PTMs. They found that the equilibrated Cx26, absent of PTMs, had a lower conductance and was completely anion selective. Experimentally, Cx26 is slightly cation selective ($P_{K+}/P_{Cl-} \approx 2.6$)[40]. While a number of PTM combinations yielded I-V curves similar to experiment, it was determined that n-terminal acetylation

was essential to reproduce experimental current and selectivity. Nearly all subsequent computational studies of Cx26, by the Bargiello lab and others, have added an acetylated Met1[20,25].

The GCMC/BD method provides a fast, and robust way to assess the conductivity and selectivity of ion channels, though not without certain caveats. First and foremost, the simulation uses a static protein structure, thus it doesn't include effects of NTH dynamics on perm-selectivity. Furthermore, the simulations are run in an implicit solvent, *i.e.,* only considering water as a continuous dielectric medium. Though all of the approximations in the GCMC/BD are derived from physical-chemical characteristics of the hemichannel, they lack the effects of the ensemble of microstates sampled by the NTH and the rest of the channel.

Originally reported by Maede et al. to be that of the open intercellular channel, multiple computational studies have suggested that the Cx26 model does not represent the fully open state. The ambiguity of the physiological state of the NTH stems from a lack of well-resolved electron density within this region of the protein in the 2009 crystal structure[54]. A computational study done in 2014 even identified an intracellular water pocket situated below the NTH[102], however without confidence in the physiological state of the structure, the conclusions from the study are difficult to interpret. Indeed, of the three atomic models solved at the time, only one (Maeda et al.) had resolved the NTH, highlighting the general instability of the domain on Cx26 under crystallographic conditions.

1.3.6 Interactions with the lipid environment.

Gap junctions, like many membrane-bound proteins, exhibit functional modulation in response to changes in the lipid environment. It is known that gap junctions prefer distinct lipid environments in an isoform dependent manner[103]. *In vitro* experiments with Cx26 and Cx32 channels show different phospholipid classes are indeed capable of modulating channel activity[104]. These studies found optimal cholesterol concentrations at which channel permeability of both Cx26 and Cx32 is maximized (~25 mol%), while higher or lower concentrations of cholesterol lead to decreased hemichannel activity, suggesting this particular lipid may play a role in modulating gap junction activity.

The molecular mechanism(s) by which cell membrane phospholipids and cholesterol affect gap junction channels is almost completely unknown. Prior to 2020, the most detailed mechanistic information on Cx-lipid interactions comes from coarse-grained molecular dynamics (CG-MD) simulations with Cx26[104,105]. Data from these simulations suggest there are coordination sites for phosphatidylcholine (PC) and cholesterol, but this has yet to be corroborated by published experimental studies. It is difficult to study such interactions by X-ray crystallographic methods, which are typically performed on detergent solubilized membrane proteins, albeit highly specific lipids are known to co-crystalize with membrane proteins in some cases. Remarkably little is known regarding the role of the membrane on connexin channel function, especially considering the dynamics and mechanism of gap junction plaque formation.

1.3.7 Conclusion

There are 21 isoforms of connexins, expressed throughout the body, each with distinct biophysical properties suited for their respective cellular environments. There have been a multitude of experimental and computational studies targeted at understanding connexin channel properties, yet a full atomistic picture describing the permeation and selectivity mechanisms of a complete gap junction channel had still been missing. Furthermore, nearly all of the computational modeling in this field have relied on the 2009 Cx26 crystal structure. Indeed, the same structure has been used as a reference to generate homology models of different connexin isoforms. Yet, multiple studies determined that the crystal structure cannot be that of the fully open channel, and instead conclude it to be partially if not fully closed. It is also unclear if Cx26 is a suitable model for all connexins given the broad diversity of channel properties that have been demonstrated by experiment, underscoring the need for additional experimental structures of different connexin isoforms and/or functional states.

References

1. Goodenough DA, Paul DL. Gap junctions. *Cold Spring Harb Perspect Biol*. 2009;1(1). doi:10.1101/cshperspect.a002576
2. Listan Mes¸e G, Richard G, White TW. Gap Junctions: Basic Structure and Function. *J Invest Dermatol*. 2007;127:2516-2524. doi:10.1038/sj.jid.5700770
3. Delmar M, Laird DW, Naus CC, et al. Connexins and Disease. *Cold Spring Harb Perspect Biol*. 2018;10. doi:10.1101/cshperspect.a029348
4. Pal JD, Liu X, Mackay D, et al. Connexin46 mutations linked to congenital cataract show loss of gap junction channel function. *Am J Physiol Cell Physiol*. 2000;279(3):C596-602. doi:10.1152/ajpcell.2000.279.3.C596
5. Berthoud VM, Beyer EC. Oxidative Stress, Lens Gap Junctions, and Cataracts. *Antioxid Redox Signal*. 2009;11(2):339-353. doi:10.1089/ars.2008.2119
6. Dong L, Liu X, Li H, Vertel BM, Ebihara L. Role of the N-terminus in permeability of chicken connexin45.6 gap junctional channels. *Authors J Compil C*. 2006;576(3):787-799. doi:10.1113/jphysiol.2006.113837
7. Van Der Velden HMW, Jongsma HJ. Cardiac gap junctions and connexins: their role in atrial fibrillation and potential as therapeutic targets. *Cardiovasc Res*. 2002;54:270-279. Accessed February 25, 2018. www.elsevier.com
8. García IE, Prado P, Pupo A, et al. Connexinopathies: A structural and functional glimpse. *BMC Cell Biol*. 2016;17(1). doi:10.1186/s12860-016-0092-x
9. Srinivas M, Verselis VK, White TW. Human diseases associated with connexin mutations. *Biochim Biophys Acta - Biomembr*. 2018;1860(1):192-201.

doi:10.1016/j.bbamem.2017.04.024

10. Aasen T, Mesnil M, Naus CC, Lampe PD, Laird DW. Gap junctions and cancer: Communicating for 50 years. *Nat Rev Cancer*. 2016;16(12):775-788. doi:10.1038/nrc.2016.105

11. Hayes MJ, Moss SE. Annexins and disease. *Biochem Biophys Res Commun*. 2004;322(4):1166-1170. doi:10.1016/j.bbrc.2004.07.124

12. chen Q, Boire adrienne, Jin X, et al. Carcinoma-astrocyte gap junctions promote brain metastasis by cGAMP transfer. Published online 2016. doi:10.1038/nature18268

13. Bennett MVL. PHYSIOLOGY OF ELECTROTONIC JUNCTIONS. *Ann N Y Acad Sci*. 1966;137(2):509-539. doi:10.1111/j.1749-6632.1966.tb50178.x

14. Spray DC, Harris AL, Bennett MVL. Equilibrium properties of a voltage-dependent junctional conductance. *J Gen Physiol*. 1981;77(1):77-93. doi:10.1085/jgp.77.1.77

15. Harris AL, Spray DC, Bennett MVL. Kinetic properties of a voltage-dependent junctional conductance. *J Gen Physiol*. 1981;77(1):95-117. doi:10.1085/jgp.77.1.95

16. Spray DC, White RL, Mazet F, Bennett VL. *Regulation of Gap Junctional Conductance*. Vol 17.; 1985.

17. Loewenstein WR. Junctional intercellular communication: the cell-to-cell membrane channel. *Physiol Rev*. 1981;61(4):829-913. doi:10.1152/physrev.1981.61.4.829

18. Lampe PD, Lau AF. Regulation of gap junctions by phosphorylation of connexins.

Arch Biochem Biophys. 2000;384(2):205-215. doi:10.1006/abbi.2000.2131

19. Locke D, Bian S, Li H, Harris AL. Post-translational modifications of connexin26 revealed by mass spectrometry HHS Public Access. *Biochem J*. 424(3):385-398. doi:10.1042/BJ20091140

20. Kwon T, Harris AL, Rossi A, Bargiello TA. Molecular dynamics simulations of the Cx26 hemichannel: Evaluation of structural models with Brownian dynamics. *J Gen Physiol*. 2011;138(5):475-493. doi:10.1085/jgp.201110679

21. Slavi N, Wang Z, Harvey L, Schey KL, Srinivas M. Identification and Functional Assessment of Age-Dependent Truncations to Cx46 and Cx50 in the Human Lens. *Investig Opthalmology Vis Sci*. 2016;57(13):5714. doi:10.1167/iovs.16-19698

22. Purnick PEM, Oh S, Abrams CK, Verselis VK, Bargiello TA. Reversal of the gating polarity of gap junctions by negative charge substitutions in the N-terminus of connexin 32. *Biophys J*. 2000;79(5):2403-2415. doi:10.1016/S0006-3495(00)76485-X

23. Trexler EB, Bennett MVL, Bargiello TA, Verselis VK. Voltage gating and permeation in a gap junction hemichannel. *Proc Natl Acad Sci U S A*. 1996;93(12):5836-5841. doi:10.1073/pnas.93.12.5836

24. Trexler EB, Bukauskas FF, Kronengold J, Bargiello TA, Verselis VK. The First Extracellular Loop Domain Is a Major Determinant of Charge Selectivity in Connexin46 Channels. *Biophys J*. 2000;79(6):3036-3051. doi:10.1016/S0006-3495(00)76539-8

25. Luo Y, Rossi AR, Harris AL. Computational Studies of Molecular Permeation

through Connexin26 Channels. *Biophys J*. 2016;110(3):584-599. doi:10.1016/j.bpj.2015.11.3528

26. Simon AM, Goodenough DA. Diverse functions of vertebrate gap junctions. *Trends Cell Biol*. 1998;8(12):477-483. doi:10.1016/S0962-8924(98)01372-5

27. Harris AL. Electrical coupling and its channels. *J Gen Physiol*. Published online 2018. doi:10.1085/jgp.201812203

28. Bennett MVL. Gap junctions as electrical synapses. *J Neurocytol*. 1997;26(6):349-366. doi:10.1023/A:1018560803261

29. Mathias RT, Kistler J, Donaldson P. The lens circulation. *J Membr Biol*. 2007;216(1):1-16. doi:10.1007/s00232-007-9019-y

30. Cheng C, Nowak RB, Gao J, et al. Lens ion homeostasis relies on the assembly and/or stability of large connexin 46 gap junction plaques on the broad sides of differentiating fiber cells. *Am J Physiol - Cell Physiol*. 2015;308(10):C835-C847. doi:10.1152/ajpcell.00372.2014

31. Weidmann S. The electrical constants of Purkinje fibres. *J Physiol*. 1952;118(3):348-360. doi:10.1113/jphysiol.1952.sp004799

32. Bennett PB, Yazawa K, Makita N, George AL. *Molecular Mechanism for an Inherited Cardiac Arrhythmia*.

33. Sabatini BL, Regehr WG. *Timing of Neurotransmission at Fast Synapses in the Mammalian Brain*.; 1996.

34. Galarreta M, Hestrin S. Electrical synapses between GABA-releasing interneurons. *Nat Rev Neurosci*. 2001;2(6):425-433. doi:10.1038/35077566

35. Connors BW, Long MA. Electrical synapses in the mammalian brain. *Annu Rev Neurosci*. 2004;27:393-418. doi:10.1146/annurev.neuro.26.041002.131128

36. Mathias RT, Rae JL, Baldo GJ. Physiological properties of the normal lens. *Physiol Rev*. 1997;77(1):21-50. doi:10.1152/physrev.1997.77.1.21

37. Bennett M V., Verselis VK. Biophysics of gap junctions. *Semin Cell Biol*. 1992;3(1):29-47. doi:10.1016/S1043-4682(10)80006-6

38. Kelsell DP, Dunlop J, Hodgins MB. Human diseases: Clues to cracking the connexin code? *Trends Cell Biol*. 2001;11(1):2-6. doi:10.1016/S0962-8924(00)01866-3

39. Gao J, Minogue PJ, Beyer EC, Mathias RT, Berthoud VM. Disruption of the lens circulation causes calcium accumulation and precipitates in connexin mutant mice. *Am J Physiol Physiol*. 2018;314(4):C492-C503. doi:10.1152/ajpcell.00277.2017

40. Iossa S, Marciano E, Franzé A. GJB2 Gene Mutations in Syndromic Skin Diseases with Sensorineural Hearing Loss. *Curr Genomics*. 2011;12(7):475-785. doi:10.2174/138920211797904098

41. Kelsell DP, Dunlop J, Stevens HP, et al. Connexin 26 mutations in hereditary non-syndromic sensorineural deafness. *Nature*. 1997;387(6628):80-83. doi:10.1038/387080a0

42. Sinyuk M, Mulkearns-Hubert EE, Reizes O, Lathia J. Cancer Connectors: Connexins, Gap Junctions, and Communication. *Front Oncol*. 2018;8:646. doi:10.3389/fonc.2018.00646

43. Warawdekar UM, Jain V, Patel H, Nanda A, Kamble V. Modifying gap junction

communication in cancer therapy. *Curr Res Transl Med*. 2021;69(1). doi:10.1016/j.retram.2020.09.002

44. Bevans CG, Kordel M, Rhee SK, Harris AL. Isoform composition of connexin channels determines selectivity among second messengers and uncharged molecules. *J Biol Chem*. 1998;273(5):2808-2816. doi:10.1074/jbc.273.5.2808

45. Harris AL. Connexin channel permeability to cytoplasmic molecules. *Prog Biophys Mol Biol*. 2007;94(1-2):120-143. doi:10.1016/j.pbiomolbio.2007.03.011

46. Koval M, Molina SA, Burt JM. Mix and match: Investigating heteromeric and heterotypic gap junction channels in model systems and native tissues. *FEBS Lett*. 2014;588(8):1193-1204. doi:10.1016/j.febslet.2014.02.025

47. Kumar NM, Gilula NB. The gap junction communication channel. *Cell*. 1996;84(3):381-388. doi:10.1016/S0092-8674(00)81282-9

48. Ayad WA, Locke D, Koreen I V, Harris AL. Heteromeric, but not homomeric, connexin channels are selectively permeable to inositol phosphates. *J Biol Chem*. 2006;281(24):16727-16739. doi:10.1074/jbc.M600136200

49. Bevans CG, Kordel M, Rheeţ SK, Harris AL. Isoform Composition of Connexin Channels Determines Selectivity among Second Messengers and Uncharged Molecules*. Accessed February 22, 2018. http://www.jbc.org/content/273/5/2808.full.pdf

50. Cottrell GT, Burt JM. Functional consequences of heterogeneous gap junction channel formation and its influence in health and disease. *Biochim Biophys Acta - Biomembr*. 2005;1711(2):126-141. doi:10.1016/J.BBAMEM.2004.11.013

51. Naulin PA, Lozano B, Fuentes C, et al. Polydisperse molecular architecture of connexin 26/30 heteromeric hemichannels revealed by atomic force microscopy imaging. *J Biol Chem*. 2020;295(49):16499-16509. doi:10.1074/jbc.RA119.012128

52. Yum SW, Zhang J, Valiunas V, et al. Human connexin26 and connexin30 form functional heteromeric and heterotypic channels. *AJP Cell Physiol*. 2007;293(3):C1032-C1048. doi:10.1152/ajpcell.00011.2007

53. White TW, Bruzzone R, Wolfram S, Paul DL, Goodenough DA. Selective interactions among the multiple connexin proteins expressed in the vertebrate lens: The second extracellular domain is a determinant of compatibility between connexins. *J Cell Biol*. 1994;125(4):879-892. doi:10.1083/jcb.125.4.879

54. Maeda S, Nakagawa S, Suga M, et al. Structure of the connexin 26 gap junction channel at 3.5 Å resolution. *Nature*. 2009;458(7238):597-602. doi:10.1038/nature07869

55. Maeda S, Nakagawa S, Suga M, et al. Structure of the connexin 26 gap junction channel at 3.5 Å resolution. *Nature*. 2009;458(7238):597-602. doi:10.1038/nature07869

56. Kronengold J, Srinivas M, Verselis VK. The N-terminal half of the connexin protein contains the core elements of the pore and voltage gates. *J Membr Biol*. 2012;245(8):453-463. doi:10.1007/s00232-012-9457-z

57. Xin L, Bai D. Functional roles of the amino terminal domain in determining biophysical properties of Cx50 gap junction channels. *Front Physiol*. 2013;4 DEC:373. doi:10.3389/fphys.2013.00373

58. Bukauskas FF, Verselis VK. Gap junction channel gating. *Biochim Biophys Acta - Biomembr*. 2004;1662(1-2):42-60. doi:10.1016/j.bbamem.2004.01.008

59. Pfahnl A, Dahl G. *Localization of a Voltage Gate in Connexin46 Gap Junction Hemichannels*.; 1998.

60. Schmilinsky-Fluri G, Valiunas V, Willi M, Weingart R. *Modulation of Cardiac Gap Junctions: The Mode of Action of Arachidonic Acid*. Vol 29.; 1997.

61. Bukauskas FF, Peracchia C. Two distinct gating mechanisms in gap junction channels: CO2-sensitive and voltage-sensitive. *Biophys J*. 1997;72(5):2137-2142. doi:10.1016/S0006-3495(97)78856-8

62. Verselis VK, Ginter CS, Bargiello TA. Opposite voltage gating polarities of two closely related onnexins. *Nature*. 1994;368(6469):348-351. doi:10.1038/368348a0

63. Xin L, Nakagawa S, Tsukihara T, Bai D. Aspartic Acid Residue D3 Critically Determines Cx50 Gap Junction Channel Transjunctional Voltage-Dependent Gating and Unitary Conductance. *Biophys J*. 2012;102(5):1022-1031. doi:10.1016/J.BPJ.2012.02.008

64. Verselis VK, Trexler EB, Bukauskas FF. Connexin hemichannels and cell-cell channels: Comparison of properties. *Brazilian J Med Biol Res*. 2000;33(4):379-389. doi:10.1590/S0100-879X2000000400003

65. Oh S, Abrams CK, Verselis VK, Bargiello TA. Stoichiometry of transjunctional voltage-gating polarity reversal by a negative charge substitution in the amino terminus of a connexin32 chimera. *J Gen Physiol*. 2000;116(1):13-31.

doi:10.1085/jgp.116.1.13

66. Moreno AP, Berthoud VM, Pérez-Palacios G, Pérez-Armendariz EM. Biophysical evidence that connexin-36 forms functional gap junction channels between pancreatic mouse β-cells. *Am J Physiol - Endocrinol Metab*. 2005;288(5 51-5):948-956. doi:10.1152/ajpendo.00216.2004

67. Veenstra RD, Wang H-Z, Beyer EC, Ramanan S V, Brink PR. *Connexin37 Forms High Conductance Gap Junction Channels with Subconductance State Activity and Selective Dye and Ionic Permeabilities*. Vol 66.; 1994.

68. Lemasurier M, Heginbotham L, Miller C. KcsA: It's a potassium channel. *J Gen Physiol*. 2001;118(3):303-313. doi:10.1085/jgp.118.3.303

69. Srinivas M, Costa M, Gao Y, Fort A, Fishman GI, Spray DC. Voltage dependence of macroscopic and unitary currents of gap junction channels formed by mouse connexin50 expressed in rat neuroblastoma cells. *J Physiol*. 1999;517(3):673-689. doi:10.1111/j.1469-7793.1999.0673s.x

70. Tong X, Aoyama H, Tsukihara T, Bai D, Bai D. Charge at the 46th residue of connexin 50 is crucial for the gap-junctional unitary conductance and transjunctional voltage-dependent gating. *Authors J Physiol C*. 2014;592:5187-5202. doi:10.1113/jphysiol.2014.280636

71. Tejada MG, Sudhakar S, Kim NK, Aoyama H, Shilton BH, Bai D. Variants with increased negative electrostatic potential in the Cx50 gap junction pore increased unitary channel conductance and magnesium modulation. *Biochem J*. 2018;475(21):3315-3330. doi:10.1042/BCJ20180523

72. Lee H, Jeong H, Hyun J, et al. Cryo-EM structure of human Cx31 . 3 / GJC3 connexin hemichannel. 2020;(August):1-16.

73. Sterratt DC. Goldman-Hodgkin-Katz Equations. In: *Encyclopedia of Computational Neuroscience*. Springer New York; 2015:1300-1302. doi:10.1007/978-1-4614-6675-8_229

74. Veenstra RD. Size and selectivity of gap junction channels formed from different connexins. *J Bioenerg Biomembr*. 1996;28(4):327-337. doi:10.1007/BF02110109

75. Weber PA, Chang HC, Spaeth KE, Nitsche JM, Nicholson BJ. The permeability of gap junction channels to probes of different size is dependent on connexin composition and permeant-pore affinities. *Biophys J*. 2004;87(2):958-973. doi:10.1529/biophysj.103.036350

76. Nicholson BJ, Weber PA, Cao F, Chang HC, Lampe P, Goldberg G. The molecular basis of selective permeability of connexins is complex and includes both size and charge. *Brazilian J Med Biol Res*. 2000;33(4):369-378. doi:10.1590/S0100-879X2000000400002

77. Valiunas V, Valiuniene L, Brink PR, White TW. Comparison of Second Messenger Permeabilities through Gap Junction Channels Formed by CX43 and CX50. *Biophys J*. 2017;112(3):551a. doi:10.1016/j.bpj.2016.11.2976

78. Valiunas V, Brink PR, White TW. Lens Connexin Channels Have Differential Permeability to the Second Messenger cAMP. Published online 2019.

79. Veenstra RD, Wang HZ, Beblo DA, et al. Selectivity of connexin-specific gap junctions does not correlate with channel conductance. *Circ Res*.

1995;77(6):1156-1165. Accessed April 29, 2019. http://www.ncbi.nlm.nih.gov/pubmed/7586229

80. Kanaporis G, Mese G, Valiuniene L, White TW, Brink PR, Valiunas V. Gap junction channels exhibit connexin-specific permeability to cyclic nucleotides. *J Gen Physiol*. 2008;131(4):293-305. doi:10.1085/jgp.200709934

81. Maeda S, Tsukihara T. Structure of the gap junction channel and its implications for its biological functions. *Cell Mol Life Sci*. 2011;68(7):1115-1129. doi:10.1007/s00018-010-0551-z

82. Nakagawa S, Gong XQ, Maeda S, et al. Asparagine 175 of connexin32 is a critical residue for docking and forming functional heterotypic gap junction channels with connexin26. *J Biol Chem*. 2011;286(22):19672-19681. doi:10.1074/jbc.M110.204958

83. Zonta F, Polles G, Zanotti G, Mammano F. Permeation pathway of homomeric connexin 26 and connexin 30 channels investigated by molecular dynamics. *J Biomol Struct Dyn*. 2012;29(5):985-998. doi:10.1080/073911012010525027

84. Suchyna TM, Nitsche JM, Chilton M, Harris AL, Veenstra RD, Nicholson BJ. Different ionic selectivities for connexins 26 and 32 produce rectifying gap junction channels. *Biophys J*. 1999;77(6):2968-2987. doi:10.1016/S0006-3495(99)77129-8

85. Zonta F, Polles G, Sanasi MF, Bortolozzi M, Mammano F. The 3.5 ångström X-ray structure of the human connexin26 gap junction channel is unlikely that of a fully open channel. *Cell Commun Signal*. 2013;11(1):1-9. doi:10.1186/1478-811X-11-15

86. Zonta F, Buratto D, Cassini C, Bortolozzi M, Mammano F. Molecular dynamics simulations highlight structural and functional alterations in deafness-related M34T mutation of connexin 26. *Front Physiol*. 2014;5 MAR(March):1-9. doi:10.3389/fphys.2014.00085

87. Bicego M, Beltramello M, Melchionda S, et al. Pathogenetic role of the deafness-related M34T mutation of Cx26. doi:10.1093/hmg/ddl184

88. Oshima A, Tani K, Hiroaki Y, Fujiyoshi Y, Sosinsky GE. Three-dimensional structure of a human connexin26 gap junction channel reveals a plug in the vestibule. *Proc Natl Acad Sci*. 2007;104(24):10034-10039. doi:10.1073/pnas.0703704104

89. Oshima A, Tani K, Hiroaki Y, Fujiyoshi Y, Sosinsky GE. Projection structure of a n-terminal deletion mutant of connexin 26 channel with decreased central pore density. *Cell Commun Adhes*. 2008;15(1-2):85-93. doi:10.1080/15419060802013588

90. Gouaux E, MacKinnon R. Principles of selective ion transport in channels and pumps. *Science (80-)*. 2005;310(5753):1461-1465. doi:10.1126/science.1113666

91. Doyle DA, Cabral JM, Pfuetzner RA, et al. The structure of the potassium channel: Molecular basis of K+ conduction and selectivity. *Science (80-)*. 1998;280(5360):69-77. doi:10.1126/science.280.5360.69

92. Bernèche S, Roux B. Energetics of ion conduction through the K+ channel. *Nature*. 2001;414(6859):73-77. doi:10.1038/35102067

93. Tang Q, Dowd TL, Verselis VK, Bargiello TA. Conformational changes in a pore-forming region underlie voltage-dependent “loop gating” of an unapposed

connexin hemichannel. *J Gen Physiol*. 2009;133(6):555-570. doi:10.1085/jgp.200910207

94. Kwon T, Roux B, Jo S, Klauda JB, Harris AL, Bargiello TA. Molecular dynamics simulations of the Cx26 hemichannel: Insights into voltage-dependent loop-gating. *Biophys J*. 2012;102(6):1341-1351. doi:10.1016/j.bpj.2012.02.009

95. Srinivas M, Kronengold J, Bukauskas FF, Bargiello TA, Verselis VK. Correlative studies of gating in Cx46 and Cx50 hemichannels and gap junction channels. *Biophys J*. 2005;88(3):1725-1739. doi:10.1529/biophysj.104.054023

96. Kwon T, Roux B, Jo S, Klauda JB, Harris AL, Bargiello TA. Molecular dynamics simulations of the Cx26 hemichannel: Insights into voltage-dependent loop-gating. *Biophys J*. 2012;102(6):1341-1351. doi:10.1016/j.bpj.2012.02.009

97. Bennett BC, Purdy MD, Baker KA, et al. An electrostatic mechanism for Ca2+-mediated regulation of gap junction channels. *Nat Commun*. 2016;7:8770. doi:10.1038/ncomms9770

98. Lopez W, Ramachandran J, Alsamarah A, Luo Y, Harris AL, Contreras JE. Mechanism of gating by calcium in connexin hemichannels. *Proc Natl Acad Sci U S A*. 2016;113(49):E7986-E7995. doi:10.1073/pnas.1609378113

99. Oh S, Rivkin S, Tang Q, Verselis VK, Bargiello TA. Determinants of gating polarity of a connexin 32 hemichannel. *Biophys J*. 2004;87(2):912-928. doi:10.1529/biophysj.103.038448

100. Wang Z, Schey KL. Phosphorylation and truncation sites of bovine lens connexin 46 and connexin 50. *Exp Eye Res*. 2009;89(6):898-904.

doi:10.1016/j.exer.2009.07.015

101. Lee K Il, Jo S, Rui H, et al. Web interface for brownian dynamics simulation of ion transport and its applications to beta-barrel pores. *J Comput Chem*. 2012;33(3):331-339. doi:10.1002/jcc.21952

102. Araya-Secchi R, Perez-Acle T, Kang SG, et al. Characterization of a novel water pocket inside the human Cx26 hemichannel structure. *Biophys J*. 2014;107(3):599-612. doi:10.1016/j.bpj.2014.05.037

103. Locke D, Liu J, Harris AL. Lipid Rafts Prepared by Different Methods Contain Different Connexin Channels, but Gap Junctions Are Not Lipid Rafts †. *Biochemistry*. 2005;44:13027-13042. doi:10.1021/bi050495a

104. Locke D, Harris AL. Connexin channels and phospholipids: Association and modulation. *BMC Biol*. 2009;7(1):52. doi:10.1186/1741-7007-7-52

105. Hung A, Yarovsky I. Gap Junction Hemichannel Interactions with Zwitterionic Lipid, Anionic Lipid, and Cholesterol: Molecular Simulation Studies. *Biochemistry*. 2011;50:1492-1504. doi:10.1021/bi1004156

Chapter 2

Molecular Dynamics Theory and Methodology

2.1 Historical Overview of Molecular Dynamics Simulation Applied to Membrane Channels

Biology is dynamic. That is to say that every biological process is constantly in flux, both in terms of the generation and consumption of chemical and potential energy, and the physical movements of cells, organelles, biomolecules (e.g., protein, RNA, etc.) and solvents. This is obvious when considering membrane-channels, as they facilitate the movement of solutes across cellular membranes that are driven by electro-chemical potentials. In biochemistry, the central dogma is that structure IS function. With the application of X-ray crystallography to biological molecules introduced in the 1960's came the ability to understand and predict biomolecular functional behavior from near-atomic resolution structures[1]. However, as stated above *biology is dynamic*, and thus there is a limit to the information which can be extracted from *static* crystal structures. While the motto "Structure is Function" is commonly touted by structural biologists to emphasize the empirical nature of molecular machines – here I will discuss an alternative paradigm, "Dynamics is Function".

The first molecular dynamics (MD) simulation of a biological molecule was performed in 1977, describing thermal motions of the trypsin inhibitor, by Andrew McCammon and Martin Karplus[2]. An MD simulation takes the three-dimensional (3D) coordinates of atoms, which have a mass and charge, and integrates Newton's equations of motion to generate a trajectory[3]. Key findings from the study, were that folded proteins have fluid-

like motions at room temperature and that proteins sample an ensemble of configurations near the experimental crystal structure. This insight was gained from what was a remarkable simulation of thermal fluctuations over a duration ~4 ps (4 x 10^{-12} seconds)[2]. Their seminal paper demonstrated that proteins are dynamic machines whose behaviour is governed by diffusive motions of atoms according to the local and long-range interactions within the protein. Martin Karplus shared the Nobel prize in 2013 for his advancement of the field of computational chemistry, which laid the groundwork for all subsequent biomolecular simulations.

As the field progressed there was a need for standardized MD engines – software that propagates molecular motions through force integrations. One such pioneer in this regard, Klaus Schulten – whose contributions to the field of molecular dynamics are nearly innumerable – is perhaps best known for his development of the NAnoscale Molecular Dynamics (NAMD)[4,5] and Visual Molecular Dynamics (VMD)[6] software suites. NAMD & VMD streamlined the preparation, execution, and analysis of MD simulations, broadening the accessibility of the field. The first simulation with NAMD/VMD was of a biological membrane, lasting 247 ps, a ground-breaking record at the time[7]. To date there are many open source, and proprietary MD engines: CHARMM[8], Amber[9], Gromacs[10], OpenMM[11], and Desmond[12] to name some of the prominent competitors in this field.

In 1978, Bertil Hille, in his seminal review of ion channels predicted that MD simulations would be a critical tool in the elucidation of mechanisms underlying gating, selectivity and conductivity[13]. Since, there have been a number of MD studies that sought to explain the mechanism and energetics underlying the permeation of water and ions

through simple channels like gramicidin-A[14–17], and the pore of KcsA[18–20]. In 2001 Simon Bernéche and Benoit Roux described the energetics, and mechanism of ion conduction through the highly selective KcsA channel[18]. Using equilibrium free-energy methods to calculate the multi-ion free energy landscape describing permeation – their work determined the dependence on inter-ionic repulsion for rapid conductance.

In 2005, Alexij Aksimentiev and Klaus Schulten utilized an external electric field, to simulate ion conductance through the large-pore channel α-hemolysin[21]. While this was not the first use of an external electric field, it was the first time used as a standard implementation of NAMD. An alternate computational electrophysiology implementation was introduced in 2011, via the Gromacs MD engine, which utilizes charge imbalances to generate voltages[22]. These new tools enable the visualization of the non-equilibrium process of conduction – with atomic resolution.

One limiting factor in computationally studying membrane channels is that membrane channels are difficult to crystalize and therefore difficult to determine a structure, leaving the field with few models to examine and extract mechanistic understanding. That is until the resolution revolution of 2014, which brought CryoEM to the forefront of membrane channel elucidation with an explosion of near-atomic structures[23–26]. CryoEM avoids the issue of crystallization all together, allowing for the visualization of membrane bound channels in near physiological conditions. Furthermore, with recent advancements in graphics processing units (GPUs), MD algorithms, and supercomputers such as *Anton2*, the field of membrane protein structural biology has entered an exciting new era of discovery. In the following, I provide a description of the

theory and methodologic principles used to apply MD toward study the dynamics and function of ion channels.

2.2 Principles of Molecular Dynamics Simulations

Molecular dynamics simulations calculate the 3D trajectories of atoms, through integration of Newton's equations of motion, i.e., Newton's 2nd law.

$$\hat{F}(\hat{R}) = m \times \frac{d^2\hat{R}}{d^2t} = -\nabla U(\hat{R}). \quad \text{Eq. 1}$$

$$U(\hat{R}) = \sum k_i^{bond}(r_i - r_o)^2 + \sum k_i^{angle}(\theta_i - \theta_o)^2 + \quad \text{Eq. 2}$$

$$\sum k_i^{dihedral}[1 + \cos(n_i\phi_i + \delta_i)] +$$

$$\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}$$

Equation 1 demonstrates the relationship of the position of each atom $(\hat{R})$ and the underlying potential energy ($\widehat{U}(\hat{r})$, Eq. 2) that governs the dynamics. The potential energy function is composed of bonded (i.e., spring approximations), and non-bonded potential functions. The non-bonded VdW and electrostatic potentials are calculated from the 6-12 Lennard-Jones potential[27], and Coulombs law respectively (Eq. 2)[28]. The force that an atom feels, which is derived from the potential ($\widehat{U}(\hat{R})$), are dependent on the mass and charge of all other atoms within a predefined cutoff-radius, typically 14 Å. These charges come from an empirically determined force-field (FF). Charge distribution over a molecule is governed by its electronic structure and resonance, which require quantum-mechanics to fully understand. Such quantum-mechanical approaches are used to calculate the

charges of a molecules atoms – MD simulations then assume the charges to remain unmoving throughout the simulation.

While there are many different FFs, such as gromos[29], Amber[30], and OPLS[31], the work in this thesis utilizes the FF developed by the Chemistry at Harvard Molecular Mechanics (CHARMM) group. The CHARMM FF is under continuous development, using theoretical and experimental methods to parameterize each atom, group, molecule, etc. Essentially, MD simulations The FF and underlying potential energy function govern the dynamics (i.e., movement in time) of biomolecules in molecular simulations.

The conventional method of calculating long-range electrostatics was a simple summation of the coulombic interactions in real space, a computationally expensive approach. In 1993 came the development of the Particle Mesh Ewald (PME) summation, which made long-range electrostatics calculations significantly more efficient. This is accomplished by splitting the electrostatic calculation into a short-range and long-range term, the latter is summed in Fourier-space, which converges faster than real-space summations[32]. In 1995, the smoothed PME method was developed, and has since been the standard electrostatic calculation method in the NAMD/VMD packages[33]. These summations take place in Fourier (i.e., inverse) space, which requires the use of periodic boundary conditions.

Dynamics in reality are continuous in time and should be therefore integrated using infinitesimally small steps. In a simulation, a discrete approach to the integration must be taken. There are many numerical algorithms that are applied to this problem, with varying accuracy and efficiency – NAMD utilizes the Verlet discrete integration method[34] (Eq. 3).

$$\hat{r}(t + \Delta t) = 2\hat{r}(t) - \hat{r}(t - \Delta t) + \Delta t^2 \frac{d^2\hat{r}(t)}{d^2t}. \quad \text{Eq.3}$$

This equation predicts the position of an atom given its previous position (known) and acceleration (derived from $\widehat{U}(\hat{R})$). The error introduced in this discretization is on the order of Δt^4. This equation is used at every time-step (Δt), which is chosen prior to the simulation. It is important that the chosen time-step is about an order of magnitude shorter than the highest frequency mode to be modeled in the simulation, e.g., the C-H bond (C-H) vibration, which is approximately 9 x 10^{13} Hz (1.1 x 10^{-14} s ~11 fs)[35]. Thus, in all atom simulations of biomolecules, a standard time-step of 1-2 fs is commonly applied. Larger Δt leads to an under-sampling of these dynamics and unstable simulation. There are, however, approximations that can be made to increase the time-step, such as applying rigid-bonds to hydrogen[36,37]. A more recent implementation is Hydrogen-mass repartitioning, which moves mass from heteroatoms to their bound hydrogens – this conserves mass in the system while reducing high-frequency vibrations and rotations, ultimately allowing for a 4-fs time-step[38].

Just as experimental conditions vary (e.g., energy, temperature, pressure, volume etc.), MD simulations can be performed in different thermal configurations known as ensembles. The constant volume & energy (NVE) microcanonical ensemble, constant volume & temperature (NVT) canonical ensemble, and constant pressure & temperature (NPT) or Isobaric-Isothermal ensemble. The NVE ensemble conserves energy, however, doesn't allow for any processes which exchange heat with their environment. The NVT, and NPT ensemble both require the use of a thermostat (heat bath), which add and

remove heat from the system to maintain constant temperature. The NPT ensemble, unlike NVT, also requires the use of a barostat to maintain pressure, and perhaps most closely corresponds to experiments in a laboratory, which has controlled temperature and atmospheric pressure[3]. The choice of thermostat and barostat are important, as different algorithms/implementations vary in terms of efficiency and accuracy. The different thermostats are: Langevin[39], Anderson[40], Nosé-Hoover[41], and Velocity-Rescaling[42]. The different barostats are: Berendsen[43], Nosé-Hoover[44], and Parinello-Rahman[44]. Over the last decade or so, it was determined that the Berendsen barostat inadequately reproduces equilibrium ensembles. All simulations done in this thesis utilize Langevin dynamics for the thermostat (NVT & NPT), and the Nosé-Hoover barostat (NPT).

Perhaps the biggest limitation to the field of MD simulation is the cost of computational calculations, which implicitly limit the time-scale and/or size of the system that can be modeled. In the general workflow for an MD simulation the atomic positions, $\hat{R}(t)$, are first assessed, and pair-lists for every atom and their neighbors within 14 Å are generated. Second, bonded and non-bonded forces for each atom are calculated, such that a net force vector is calculated for each atom. Third, verlet integration is used to propagate the positions of each atom after some time-step (1-4 fs). Fourth, the atomic positions are re-assessed and the cycle is repeated. In order to simulate one nanosecond for a system containing 10,000 atoms, with approximately 1,000 atoms within 14 Å of any other atom, a computer must make on the order of 10 trillion calculations (using a 1 fs time-step). As the number of atoms are increased, so goes the computational cost. The

need for longer simulations drove the development of algorithms that can leverage the massively parallel nature of graphics processing units (GPUs)[5]. Such technological advancements have brought MD simulation, which once required specialized super-computers, to the desktop – broadening access to anyone with a gaming computer.

2.3 MD Timescales and Methods of Enhanced Sampling

As previously discussed, there is a large computational cost to MD simulations, which ultimately limit the time scales that can be achieved for modestly sized biomolecular systems to the microsecond regime. While this is nearly 6 orders of magnitude greater than what could be achieved just 30 years ago, it is still below the time scales of many biologically interesting mechanisms – for example, conformational changes associated with gap junction channel gating or permeation of small-molecules occur in the millisecond to second range. There are, however, approaches to overcome these barriers. Enhanced sampling simulation techniques can generally be separated into two categories, biased and unbiased methods. Here some of the prominent methods in both categories are presented, together with their general benefits and limitations.

Of the biased MD methods, “directed” MD is a technique which applies an external potential to a single atom, or a set of atoms to drive them along some reaction coordinate. There are two implementations of directed MD simulations in NAMD, namely steered MD (SMD) and targeted MD (TMD)[45]. Both of these methods employ an external potential in the form of a virtual ball and spring tugging on atoms of interest. In SMD, this can be used to overcome molecular permeation time scales by pulling a molecule along its reaction

coordinate, such as glycerol through a GlpF channel[46]; a process which naturally occurs on the microsecond time scale[47]. TMD can be used to simulate large-scale conformational changes, by adding an external force that drives atoms between previously defined initial and final states – generating an atomic trajectory describing the interconverting motions[48]. The drawback of these directed simulation techniques is that they are prone to misrepresenting physiologically meaningful trajectories by washing over slower, important dynamics. Furthermore, kinetic information that could otherwise be used to compare against experiment for validation is lost. Despite these limitations, SMD and TMD are immensely useful in generating structures and transition-pathways, which then act as starting points for more rigorous equilibrium and non-equilibrium techniques.

One commonly used equilibrium-state approach to enhanced sampling, known as umbrella sampling, allows for the efficient calculation of a PMF along a pre-determined reaction coordinate. Provided that the system can be defined at multiple stages along a reaction coordinate (*a.k.a.* collective variable) such that there is significant overlap between the windows, for example a calcein molecule at multiple windowed-regions along the pore of the Cx26 hemichannel, as applied by Zonta et al.[49]. In this work, each state along the reaction coordinate was simulated within defined overlapping windows, with boundary potentials at either end of the window (Fig. 1B (bottom)), restricting calcein to specific windows within the channel pore. This allows sampling of the substrate throughout the entire pore and estimation of a free energy landscape, after removing the bias of the added boundary potentials via the weighted histogram analysis method[50]. In principle, any collective variable can be chosen, whether it's a calcein along the pore of a

channel[51], a twist of a helix[52], or the RMSD to a known folded state[53] (Fig. 1B). This method substantially reduces the computational cost to calculate a PMF and is therefore quite popular for systems where the kinetics are prohibitively slow. The technique is a double-edged sword, for what is gained in efficiency of the thermodynamic calculation, is exchanged for a loss of kinetics, orthogonal dynamics and therefore mechanistic information.

A final, and important caveat with any technique which places a bias along a reaction coordinate, is that these methods are likely to calculate a converged PMF. This may not sound like an issue at first, but it underscores the fact that other degrees of freedom which could be pertinent for the reaction to occur are averaging out[54]. Ultimately it is up

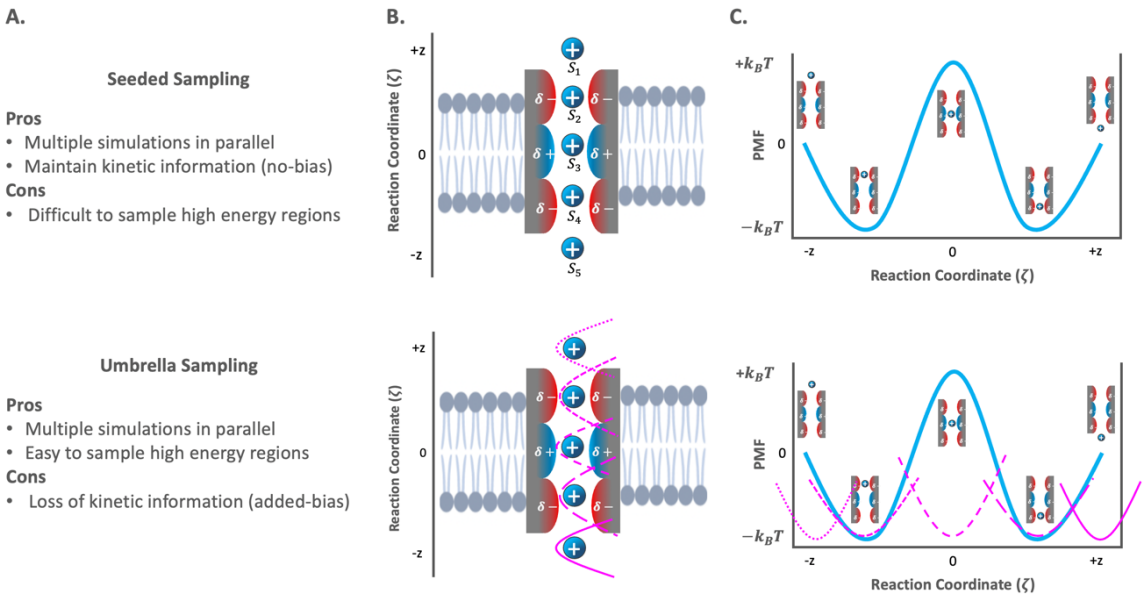


Figure 1. Enhanced sampling simulation techniques. *A) Pros and Cons of Seeded sampling vs. Umbrella sampling. B) Here the 'toy' simulation system is a positively charged molecule passing through a channel with negative charges at the openings and a positive charge at the center. C) is the associated PMF (free energy profile) for the reaction coordinate (ζ), the distance of the molecule along the z-axis. Seeded sampling (Top) starts multiple independent simulations, initialized at different points along ζ. Umbrella sampling (Bottom) runs multiple independent simulations, each with an additional parabolic biasing-potential (magenta) restraining each simulation to sample specific, overlapping regions along ζ.*

to the researcher to choose a reaction coordinate, which takes careful consideration of the system. More so, the researcher must be particularly critical of the information they extract from the bias-generated PMFs. There is therefore a desire for unbiased enhanced sampling techniques.

Seeded sampling is an unbiased simulation technique where multiple simulations are generated, with the system at different points along the chosen reaction coordinate (Fig. 1B). Like umbrella sampling, the individual simulations are combined through a Markov State Model (MSM, discussed in section **2.5**). In this way, a reasonable approximation to the unbiased kinetics of the system can be extracted; however, this method is still limited by the timescale of the process and require significant sampling to get a converged estimate of the kinetic rate[54]. A major limitation of seeded sampling is that the majority of simulation time is spent on common events, such as a molecule in an energetic well, as opposed to rare events, such as a barrier crossing (Fig. 1).

Weighted ensemble (WE) is an unbiased simulation technique that amplifies the simulation of rare events (Fig. 2). In principle, WE divides the reaction coordinate into connected bins, and initiates a pre-determined number of short unbiased simulations (*e.g.,* 3) in the initial bin. Simulations are evaluated for their progress along the reaction coordinate at the end of each run. Simulations which advance to unoccupied bins are then split or combined to reach the requisite number of instances (3 per bin in this example). Simulations may stay or leave the bin they are in, but the WE scheme attempts to ensure that once a bin is visited it will be continually sampled (Fig. 2). The key to the WE method is that it keeps track of the probability weights for each simulation, such that the 3

simulations near a statistically rare event (*i.e.,* a high energy region of the underlying PMF) will have miniscule weights that corresponds to the rare nature of the reaction at that point[55,56]. By keeping track of the statistical weights, the unbiased PMF can be calculated in principal[56]. Beyond the PMF, WE readily calculates the rate at which the process occurs[57], and extract individual, unbiased trajectories revealing the underlying atomic mechanism[58]. Like any of the enhanced sampling techniques, WE is not without its caveats. There is a substantial computational cost to spawning and maintaining multiple simulations per bin, especially when many bins might be required to sample the high energy region. WE greatly benefits from parallelization, as the spawned simulations are independent, and can therefore be used to simulate large complex systems.

2.4 Statistical Mechanics of Biomolecules (a primer)

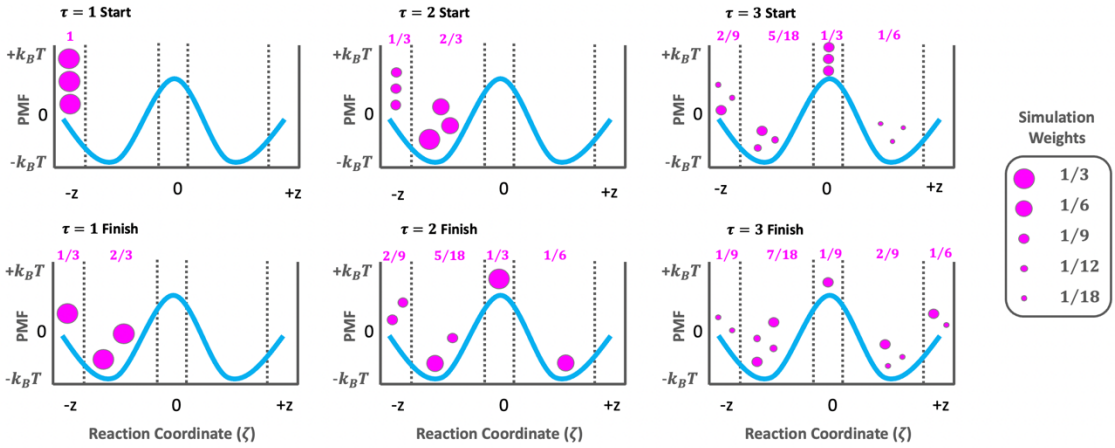


Figure 2. Weighted Ensemble (WE) enhanced sampling technique. *Underlaying PMFs which are being sampled from are in Blue. Magenta circles indicate independent simulations, their sizes correspond to their weight, or their contribution to the energy. The simulation starts (top) with our pre-determined number of simulations initialized in bin-1. After a time τ, some simulations advance to bin-2, which are then divided (and weights tracked) to satisfy the 3-sims/bin rule (middle). This process is repeated until there is sufficient sampling in each bin. The tracked weights allow for the reconstruction of an unbiased PMF. Figure adapted from "Weighted Ensemble Simulations: Review of Methodology, Applications, and Software"[56].*

As complex and dynamic as proteins and other biomolecules are, they are just that, molecules. Statistical mechanics was born from the idea that our macroscopic world is composed of many microscopic components, *i.e.,* atoms. Ludwig Boltzmann was a pioneer of statistical mechanics and believed that macroscopic quantities, such as the temperature of a gas, could be described by the microscopic collisions of individual gas molecules[59]. The ideas put forth by Boltzmann, and many other contemporary physicists and mathematicians, laid the foundation for a new paradigm by which the world could be understood.

Ludwig Boltzmann is perhaps most famous for his mathematical relation of the disorder of a system (entropy), and the ability for the system to do work (energy).

$$S = k_b \times \ln(\mathrm{W}) \qquad \text{Eq. 5}$$

This equation (Eq. 5), which is immortalized on Ludwig Boltzmann's tombstone, equates entropy (S) to the product of the natural log of degenerate states of a system (W) and the Boltzmann constant (k_b). In simpler terms, W represents the number of configurations (or states) which contribute equally to the overall energy. The Boltzmann constant is further used to calculate the energetic difference between two distinct macro-states.

$$\frac{P_A}{P_B} = e^{-\left(\frac{\Delta G_{A-B}}{k_b T}\right)} \qquad \text{Eq. 6}$$

$$\Delta G_{A-B} = -k_b T \times \ln\left(\frac{P_A}{P_B}\right) \qquad \text{Eq. 7}$$

Here, P_A and P_B are the probabilities that a system will reside within macro-states A and B respectively, ΔG_{A-B} is the free energy difference between the two states, k_b is the Boltzmann constant, and T is temperature. From Eq. 7 (Eq. 6 rearranged), it is shown that

the free energy difference between states A and B can be directly calculated, simply by looking at the ratio of their absolute probabilities. That is to say that thermodynamic quantities such as a PMF may be extracted from looking at the number of times the protein enters and leaves a specific state. As simple as these equations are, they have far reaching capabilities once applied with a statistical lens to protein function and dynamics.

It is not enough, however, to understand the energetics of a process. It is desirable to know how fast a reaction occurs, however, thermodynamics do not (necessarily) dictate kinetics. Understanding the rates of a reaction or process requires knowledge of the underlying free energy landscape, it is important to know the energetic barrier separating the two states. Given a simple energy landscape with a single barrier separating the initial and final state, the rate can be calculated from the Arrhenius equation.

$$k_{A\to B} = Ae^{-\left(\frac{E_A}{k_b T}\right)} \qquad \text{Eq. 8}$$

Here, the right-hand side of the equation is similar to Eq. 6 except for a pre-exponential factor (A), which is effectively the attempt-frequency of the $A \to B$ transition, E_A is the activation energy separating the two states. This picture of rates is perhaps too simplistic to understand the kinetics of a complex process, which requires more information such as the diffusivity of the system as a function of the reaction coordinate.

What is diffusion? In the context of ion channels, diffusion describes the movement of ions through a charged pore as in Figure 1. Recall Boltzmann's idea of the macroscopic world being composed of microscopic atomic and molecular collisions.

Diffusion is the random motion of molecules subject only to collisions due to thermal fluctuations. What results is a tendency for molecules, such as water or its solutes, to spread out radially in the absence of an external driving force, such as a sinkhole. This phenomenon is at play when a drop of dye is added to water and spreads (or diffuses) spontaneously. The mathematics describing diffusion were initially developed by Adolf Fick in 1855, famously know as Fick's law[60].

$$\hat{J} = -D\nabla C \qquad \text{Eq. 9}$$

Here, $\hat{J}$ is the flux of a particle due to diffusion, *D* is the diffusion coefficient with units $\frac{length^2}{time}$, and ∇C is the gradient of concentration with respect to space. A well-mixed, homogeneous solution would expect to yield $\hat{J} = 0$, which holds here as ∇C would be 0 and thus no net movement would occur. Regardless of the flux, diffusion underlies molecular motions. In 1905 Albert Einstein investigated the motion of Brownian (or diffusive) particles by modelling diffusion as a gaussian (i.e., bell-curve) distribution which broadens with time (Eq. 10) – he demonstrated that the diffusion coefficient is proportional to the second moment (i.e., variance)[61].

$$p(x,t) = \frac{1}{\sqrt{4\pi D}} \frac{e^{-x^2/4Dt}}{\sqrt{t}} \qquad \text{Eq. 10}$$

$$\sigma^2 = 2Dt \qquad \text{Eq. 11}$$

Eq. 10 & 11 are the probability distribution of a particle in space (x) and time (t) and the variance of the distribution (σ^2), and *D* is the diffusion coefficient. Einstein further concluded that the variance in the distribution is also equal to the mean squared

displacement (MSD) of that particle, a powerful observation that allows for the efficient calculation of the diffusion coefficient in *n*-dimensions:

$$\mathrm{MSD} \equiv \langle |\hat{x} - \hat{x}_0|^2 \rangle = 2nDt \qquad \text{Eq. 12}$$

2.5 An Introduction to Markov State Models

A Markov state model (MSM) is a formalism that allows for the discrete examination of the energetics and kinetics of a continuous process. In application to MD simulation, imagine a process of a molecule permeating through a pore, which may be rare and therefore difficult to simulate from a single trajectory. The seeded sampling approach, from section **2.3**, can be used to initiate simulations of the molecule at different positions along the pore-axis (the reaction coordinate). This coordinate system can then be discretized to just consider transitions between each discrete position along the coordinate (pore-axis). The MSM can be populated with data from multiple independent trajectories to model the process as if it were continuous.

There are certain assumptions that must be made with the MSM. The Markovian assumption states the following: The system evolves without memory, *i.e.,* its next step depends only on its current position, not where it was any number of steps prior. In other words, each step is independent of the last, or they are un-correlated in time. It follows that the discrete states should be designed such that the time it takes to move within the state is significantly shorter than the time required to move between states. As an MSM is not continuous in time, the evolution of the system in discrete time-jumps, known as a lag-time (λ), must be considered. The lag time is chosen such that it is small enough to

resolve the phenomenon of interest and large enough to maintain the Markovian assumption, however this is not always possible.

The following master equation is used to understand the evolution of a discretized system:

$$\frac{d\hat{P}}{dt} = \boldsymbol{T}\hat{P} \qquad \text{Eq. 13}$$

Where $\hat{P}$ is a column vector describing the population of each discrete state, and T is the transition matrix (or MSM) which describes the evolution of $\hat{P}$ for a given lag time (λ). The transition matrix is an $n \times n$ matrix, where n is the dimensionality or number of states, which has as its elements the conditional probability of the system transitioning from one state to another, or itself. Once a system has reached equilibrium, $\hat{P}$ is expected to be constant:

$$\boldsymbol{T}\hat{P}^{\mathrm{Eq}} = \hat{P}^{\mathrm{Eq}} \qquad \text{Eq. 14}$$

In other words, $\hat{P}^{\mathrm{Eq}}$ is an eigenvector of $\boldsymbol{T}$ – resulting in a trivial calculation for $\hat{P}^{\mathrm{Eq}}$ from the populated transition matrix. Since $\hat{P}^{\mathrm{Eq}}$ is a list of individual state probabilities, the Boltzmann equation (Eq. 6), used to calculate the energy between any two states, can be readily applied.

In addition to extracting thermodynamics, Markov models can be used to model many aspects of biological phenomena, including ionic selectivity. The lab of Anatoly Kolomeisky used discrete state MSMs to model simplified channels, which conferred selectivity between two arbitrary solutes[62]. By controlling the rate constants for each respective solute at each step throughout the channel, they could effectively change the

affinity for the two solutes. They performed fast numerical simulations on many initial configuration, comparing the ability for their channel-model to select for one solute or the other. Intriguingly, their models predict that the highest selectivity is achieved when specific binding sites (i.e., selectivity filters) are near the pore entrance. Though these are not atomistic simulations, the tunable nature of MSMs with their fast numerical solutions make them beneficial for understanding the fundamentals of ionic permeation.

2.6 Computational Electrophysiology

Perhaps one of the most common techniques used to study ion-channels is by 'patch-clamp' electrophysiology. Ion channels are primarily responsible for electrically coupling cells[63], maintaining osmotic pressure[64,65], and maintaining the resting membrane potential[66]. The patch clamp can be used control the electrical environment around a channel and record its response. In the voltage-clamp technique, this is accomplished by "clamping" the membrane potential with an electrode and measuring the current through an ion channel(s). This technique can be used to calculate the conductance of a channel, observe gating, and with certain formulaic assumptions, the relative ionic selectivity can be determined using the GHK equation[67]. While electrophysiology can see these long-time scale events, it has a time-resolution limited to a few milliseconds – furthermore, while the proteins functional response can be measured with exquisite detail, this technique provides limited information on the atomistic mechanisms involved. To gain such insights, methods of computational electro-physiology have been developed in the MD field. The formulations of computational electrophysiology approaches used to

understand the molecular mechanisms underlying an active conducting channel are described below.

Different computational methods can be generally divided by how they treat solvent properties. In the case of the GCMC/BD method, as applied to generate I-V curves and characterize the conductive properties of the Cx26 hemichannel[68] described in Section **1.3.5**, a continuum (or implicit) solvent model is applied. Here, the solvent is treated as a constant dielectric, and the Poisson-Boltzmann equation is applied to account for the effect of the protein on the overall electrostatic environment.

$$\nabla^2\psi \equiv \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2} = -\frac{\rho_e}{\epsilon_r\epsilon_0} \qquad \text{Eq. 15}$$

$$\rho_e = c_0 e\left[\mathrm{e}^{\frac{-e\psi(x,y,z)}{k_B T}} - \mathrm{e}^{\frac{e\psi(x,y,z)}{k_B T}}\right]$$

In Eq. 15, the 3D distribution of the electric potential (ψ), described by the Laplacian of the potential $(\nabla^2\psi)$, is equal to the negative local charge density (ρ_e) divided by the product of the local dielectric constant (ϵ_r) and the permittivity of free space (ϵ_0). The local charge density of a continuum can be modelled via the Boltzmann factor – this equation looks at the probability that an ion would be at a point, relative to the bulk concentration (c_0) given the local electric potential $(\psi(x, y, z))$. The GCMC/BD formulation allows for extremely efficient simulations of ions permeating a protein pore, however there are many caveats involved[69]. First and foremost, the simulation is in an implicit solvent and loses the discrete effects of water on the electric potential and the structure of the protein are lost. Second, the protein structure remains static throughout the simulation, ignoring the effects of voltage and conductance on the structural and

dynamical properties of the channel. As previously stated, *biology is dynamic*, and it's imperative to study these channels in the most physiologically relevant state possible.

Two implementations of computational electrophysiology in which atomistic details of the solvent and protein dynamics are preserved have been developed using either a constant charge separation[70] (Fig. 3A), or an externally applied constant electric field[71] (Fig. 3B). Each of these methods utilize classical mechanics force-fields, as opposed to the Brownian dynamics used in GCMC/BD. Importantly, both of these methods use explicit solvent models, which more accurately capture the dynamics of ions, and their mediated interactions with the protein. These effects have been shown to be critical component of ionic permeation[18,46].

In a voltage-clamp electrophysiological experiment, the membrane potential (V_m) is held constant and the resulting current through the channel is recorded. Each ion that transits the channel will alter V_m, bringing it closer to equilibrium. To prevent the attenuation of V_m, electrodes placed in the cell replenish the lost charge. Now consider a simulation box with an ion-channel sitting in a membrane with a 50 mV transmembrane potential. This potential enacts a force on the ions in the system and will drive them through the ion-channel down their electrochemical gradient – however, in the small volume of a MD simulation box only a few permeation events are needed to alter V_m[72]. To overcome this challenge computationally, a scheme was devised to randomly swap an ion for a water in the opposite compartment with a probability of success proportional to the Boltzmann-weighted transmembrane potential[73]. In other words, deviations from the desired transmembrane potential will increase the likelihood of a swap, restoring the

desired charge imbalance (Fig. 3A). One major caveat to this method is its requirement that the intracellular and extracellular baths only connect through the protein, otherwise ions would equilibrate through a periodic boundary[73].

Modern simulations leverage periodic boundary conditions for efficient electrostatic calculations. The labs of Ulrich Zacharia and Bert DeGroot devised a dual-membrane simulation scheme. With two membranes, there is a central bath that is disjointed from the baths above and below (Fig. 3A). The dual-membrane implementation has two benefits. First, it allows for the charge imbalance scheme to be used in tandem with periodic boundaries. Second, it allows the researcher to study the same system at both polarities of their desired voltages. Currently the dual-membrane simulation scheme, known as CompEL, is implemented within the Gromacs MD-engine in a fully automated fashion[73]. A drawback from this approach is the increased computational cost of simulating an additional membrane – furthermore, to simulate a gap junction with this method would require four membranes in total. There is, however, another method that avoids the dual-membrane requirement altogether.

In contrast to the charge-imbalance method, the constant external electric field (CEF) simulation method takes advantage of periodic boundary conditions to emulate an electrophysiology experiment[71,74]. In addition to the underlying force-field, there is an added electric field (Eq. 16), which imposes a force on all charged particles.

$$E = {}^{\Delta V}/_{L_z} \qquad \text{Eq. 16}$$

Here the externally applied electric field (E), with units Kcal $mol^{-1}e^{-1}$, is equal to the voltage (V) divided by the length (L_z) of the cell to which the voltage is applied. Typically,

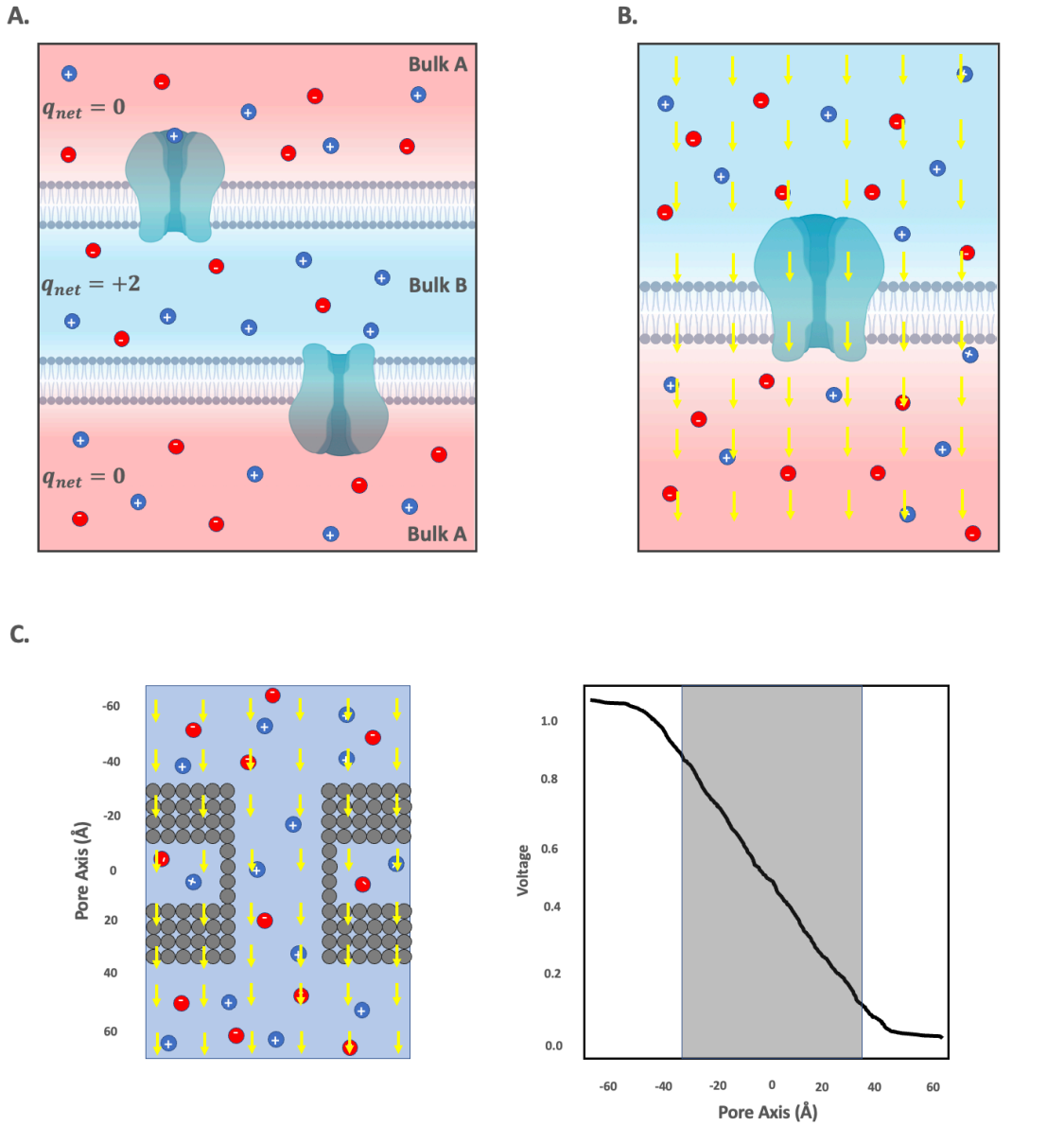


Figure 3. Computational Electrophysiology Implementations. *A) Dual membrane charge-imbalance (CompEL[22]) technique generates a potential across each membrane according to the charge imbalance, Δq ($V_m = \Delta q/C$) where V_m is the potential and C is the capacitance of the channel. The charge imbalance is maintained, periodically reinforcing it by swapping ions with water from Bulk-A and Bulk-B respectively. B) Constant external electric field (CEF) technique generates a V_m by applying a uniform electric field (yellow arrows). In both cases (A & B) the negative (red) potential and positive potential (blue) are the resulting total potentials. C) Cartoon depiction of a carbon-nanotube model of a gap junction in explicit solvent subject to an electric field (left), the total potential drop across the channel. Panel C is adapted from "Exploring the Membrane Potential of Simple Dual- Membrane Systems as Models for Gap-Junction Channels"[75].*

E is applied orthogonal to the membrane, along the pore axis, i.e., the electric field vector, $\hat{E} = (0,0,\pm E)$. The solvent and ions will rapidly respond to the external field and rearrange themselves to oppose and minimize the applied potential, this is known as the reaction field. The effect is such that the linear applied voltage is localized to the membrane, and the drop occurs through the ion-channel pore (Fig. 3B). Ions driven through the periodic boundary from the extracellular solution immediately appear in the intracellular bulk in a manner analogous to a circuit that replenishes ions to maintain a constant voltage via an EMF in the circuit[71]. The sum of E and the reaction field yield the total electric field felt by the ions, and protein in the system (Fig. 3B). The external electric field method has been shown, theoretically, to be a reliable model of an electrophysiology experiment.

The gap junction is a unique system due in part to the fact that it spans two membranes, connecting the cytosols of two cellular compartments. In 2016, Escalona et al., verified the use of the CEF method on simple dual membrane systems, and demonstrated its effectiveness on a Cx26 gap junction[75]. In this study, the authors first modelled a gap junction as two hydrophobic slabs connected by a tunnel (Fig. 3C left), in an explicit solvent with one molar KCl. An analysis of the total electrostatic potential revealed two bulk solvent regions with flat potentials, and a linear drop across the length of the carbon nanotube (Fig. 3C right). They then tested the approach with the Cx26 gap junction model (PDBID: 2ZW3) and found two flat-potential bulk solvents, though a non-linear drop of the voltage across the gap junction pore. The discrepancies between the

carbon nanotube and the Cx26 model comes from the shape, and dynamics of the gap junction pore, highlighting the need for full atomic MD simulations.

When analyzing computational electrophysiology, it is critical to measure effective transmembrane – or transjunctional in the case of gap junctions – potential, and the current through the channel. The Poisson-Boltzmann equation can be used to calculate the electrostatic potential (Eq. 15), though there are alternative methods provided by the VMD software suite, *pmepot*, which calculates the electrostatics of an MD system through a PME solver (section **2.2**)[6,21]. Likewise, there are multiple methods to calculate the current throughout the simulation. For example, by simply counting the number of ion crossing events.

$$\langle I \rangle = {}^{Nq}/_{\tau} \qquad \text{Eq. 17}$$

Where $\langle I \rangle$ is the total average current, N is the total number of ions permeated, q is the charge of the permeating ion, and τ is the length of the simulation. In an alternative method, demonstrated in Schulten and Aksimentiev[21], the instantaneous current is estimated by integrating the total charge transfer over time.

$$I(t) = \frac{1}{\Delta t L_z} \sum_{i=1}^{N} q_i [z_i(t + \Delta t_i) - z_i(t)] \qquad \text{Eq. 18}$$

Here, $I(t)$ is the instantaneous current and time t, Δt is the time-step, q_i and z_i are the charge and the unwrapped (i.e., non-periodic) z-coordinate of the *i*th ion respectively. These tools enable computational biophysicists to generate I-V curves in a similar fashion to *in situ* electrophysiology, providing a direct link between simulation and experiment, and gathering insight to the molecular underpinnings of conductance.

Computational electrophysiology has helped to elucidate the mechanism of ion conductance through highly selective channels like KcsA[76], as well as more non-selective channels like OmpF[77]. The methods and techniques outlined in this chapter combine together to make a computational microscope that enables investigators to visualize the dynamics of proteins at an atomistic-level and to understand how theses dynamics relate to function and physiology. This sentiment was perhaps best described by the physicist Richard Feynman, who said, "…everything that living things do can be understood in terms of the jigglings and wigglings of atoms."[78]

References

1. Kendrew JC. *Myoglobin and the Structure of Proteins*. Vol 139.; 1963.
2. Mccammon JA, Gel BR, Karp Ius M. *Dynamics of Folded Proteins*. Vol 267. Nature Publishing Group; 1977.
3. Frenkel D, Smit B. *Understanding Molecular Simulation: From Algorithms to Applications*. Vol 1. Elsevier; 2001.
4. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26(16):1781-1802. doi:10.1002/jcc.20289
5. Phillips JC, Hardy DJ, Maia JDC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys*. 2020;153(4). doi:10.1063/5.0014475
6. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. *J Mol Graph*. 1996;(14):33-38.
7. Helmut Heller P, Schaefer M, Schulten K. Molecular Dynamics Simulation of a Bilayer of 200 Lipids in the Gel and in the Liquid-Crystal Phases. *J Phys Chem*. 1993;97:41. Accessed May 23, 2021. https://pubs.acs.org/sharingguidelines
8. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*. 1983;4(2):187-217. doi:10.1002/jcc.540040211
9. Pearlman DA, Case DA, Caldwell JW, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun*. 1995;91(1-3):1-41.

doi:10.1016/0010-4655(95)00041-D

10. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun*. 1995;91(1-3):43-56. doi:10.1016/0010-4655(95)00042-E

11. Eastman P, Pande VS. OpenMM: A Hardware-Independent Framework for Molecular Simulations. *Comput Sci Eng*. 2010;12(4):34-39. doi:10.1109/MCSE.2010.27

12. Bowers KJ, Chow E, Xu H, et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proc 2006 ACM/IEEE Conf Supercomput SC'06*. 2006;(November). doi:10.1145/1188455.1188544

13. Hille B. Ionic Channels in Excitable Membranes Current Problems and Biophysical Approaches. *Biophys J*. 1978;22(2):283-294. http://dx.doi.org/10.1016/S0006-3495(78)85489-7

14. Mackay DH, Berens PH, Wilson KR, Hagler AT. Structure and dynamics of ion transport through gramicidin A. *Biophys J*. 1984;46(2):229-248. doi:10.1016/S0006-3495(84)84016-3

15. Roux B, Karplus M. Molecular dynamics simulations of the gramicidin channel. *Annu Rev Biophys Biomol Struct*. 1994;23(November):731-761. doi:10.1146/annurev.bb.23.060194.003503

16. Roux B, Karplus M. *MOLECULAR DYNAMICS SIMULATIONS OF THE GRAMICIDIN CHANNEL*.; 1994. Accessed May 23, 2021. www.annualreviews.org

17. Jensen M, Jogini V, Eastwood MP, Shaw DE. Atomic-level simulation of current-

voltage relationships in single-file ion channels. *J Gen Physiol*. 2013;141(5):619-632. doi:10.1085/jgp.201210820

18. Bernèche S, Roux B. Energetics of ion conduction through the K+ channel. *Nature*. 2001;414(6859):73-77. doi:10.1038/35102067

19. Noskov SY, Bernéche S, Roux B. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature*. 2004;431(7010):830-834. doi:10.1038/nature02943

20. Shrivastava IH, Sansom MSP. Simulations of ion permeation through a potassium channel: Molecular dynamics of KcsA in a phospholipid bilayer. *Biophys J*. 2000;78(2):557-570. doi:10.1016/S0006-3495(00)76616-1

21. Aksimentiev A, Schulten K. Imaging α-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys J*. 2005;88(6):3745-3761. doi:10.1529/biophysj.104.058727

22. Kutzner C, Grubmüller H, De Groot BL, Zachariae U. Computational electrophysiology: The molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophys J*. 2011;101(4):809-817. doi:10.1016/j.bpj.2011.06.010

23. Llácer JL, Hussain T, Marler L, et al. Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Mol Cell*. 2015;59(3):399-412. doi:10.1016/j.molcel.2015.06.033

24. Liu Y, Hill MG, Klose T, et al. Atomic structure of a rhinovirus C, a virus species linked to severe childhood asthma. *Proc Natl Acad Sci U S A*. 2016;113(32):8997-

9002. doi:10.1073/pnas.1606595113

25. Zhao J, Benlekbir S, Rubinstein JL. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature*. 2015;521(7551):241-245. doi:10.1038/nature14365

26. Callaway E. The Revolution Will Not Be Crystallized. *Nature*. 2015;(525):172-174. doi:10.1215/01636545-2009-008

27. April R. On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature. *Proc R Soc London Ser A, Contain Pap a Math Phys Character*. 1924;106(738):441-462. doi:10.1098/rspa.1924.0081

28. Coulomb A. Histoire de l'Académie [royale] des sciences avec les mémoires de mathématiques et de physique, partie "Mémoires" [1785] , 1788,. *A Source B Physics, 408-413*. Published online 1935.

29. Breuer M, Hägele G. Determination of approximate force field parameters. *Comput Chem*. 1989;13(2):123-128. doi:10.1016/0097-8485(89)80004-X

30. Manjunatha Kini R, Evans HJ. Molecular modeling of proteins: A strategy for energy minimization by molecular mechanics in the amber force field. *J Biomol Struct Dyn*. 1991;9(3):475-487. doi:10.1080/07391102.1991.10507930

31. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc*. 1996;118(45):11225-11236. doi:10.1021/ja9621760

32. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys*. 1993;98(12):10089-10092.

doi:10.1063/1.464397

33. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103(19):8577-8593. doi:10.1063/1.470117

34. Brooks CL. Computer simulation of liquids. *J Solution Chem*. 1989;18(1):99. doi:10.1007/BF00646086

35. Lin-Vien D, Colthup NB, Fateley WG, Grasselli JG. *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*. Elsevier; 1991.

36. Lambrakos SG, Boris JP, Oran ES, Chandrasekhar I, Nagumo M. A modified shake algorithm for maintaining rigid bonds in molecular dynamics simulations of large molecules. *J Comput Phys*. 1989;85(2):473-486. doi:https://doi.org/10.1016/0021-9991(89)90160-5

37. Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem*. 1992;13(8):952-962. doi:10.1002/jcc.540130805

38. Balusek C, Hwang H, Lau CH, et al. Accelerating Membrane Simulations with Hydrogen Mass Repartitioning. *J Chem Theory Comput*. 2019;15(8):4673-4686. doi:10.1021/acs.jctc.9b00160

39. Davidchack RL, Handel R, Tretyakov M V. Langevin thermostat for rigid body dynamics. *J Chem Phys*. 2009;130(23):234101. doi:10.1063/1.3149788

40. Sturgeon JB, Laird BB. Symplectic algorithm for constant-pressure molecular dynamics using a Nosé-Poincaré thermostat. *J Chem Phys*. 2000;112(8):3474-

3482. doi:10.1063/1.480502

41. Evans DJ, Holian BL. The Nose-Hoover thermostat. *J Chem Phys*. 1985;83(8):4069-4074. doi:10.1063/1.449071

42. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys*. 2007;126(1):14101. doi:10.1063/1.2408420

43. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys*. 1977;23(3):327-341. doi:10.1016/0021-9991(77)90098-5

44. Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *J Chem Phys*. 1994;101(5):4177-4189. doi:10.1063/1.467468

45. Isralewitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol*. 2001;11(2):224-230. doi:10.1016/S0959-440X(00)00194-9

46. Park S, Tajkhorshid E, Schulten K, Jensen MØ. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc Natl Acad Sci*. 2002;99(10):6731–6736. doi:10.1073/pnas.102649299

47. Borgnia MJ, Agre P. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from Escherichia coli. *Proc Natl Acad Sci U S A*. 2001;98(5):2888-2893. doi:10.1073/pnas.051628098

48. Li Y, Barbault F, Delamar M, Zhang R, Hu R. Targeted molecular dynamics (TMD) of the full-length KcsA potassium channel: On the role of the cytoplasmic domain in the opening process. *J Mol Model*. 2013;19(4):1651-1666. doi:10.1007/s00894-

012-1726-3

49. Zonta F, Polles G, Zanotti G, Mammano F. Permeation pathway of homomeric connexin 26 and connexin 30 channels investigated by molecular dynamics. *J Biomol Struct Dyn*. 2012;29(5):985-998. doi:10.1080/073911012010525027

50. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem*. 1992;13(8):1011-1021. doi:10.1002/jcc.540130812

51. Zonta F, Polles G, Sanasi MF, Bortolozzi M, Mammano F. The 3.5 ångström X-ray structure of the human connexin26 gap junction channel is unlikely that of a fully open channel. *Cell Commun Signal*. 2013;11(1):1-9. doi:10.1186/1478-811X-11-15

52. Fiorin G, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. *Mol Phys*. 2013;111(22-23):3345-3362. doi:10.1080/00268976.2013.813594

53. Salvalaglio M, Tiwary P, Parrinello M. Assessing the reliability of the dynamics reconstructed from metadynamics. *J Chem Theory Comput*. 2014;10(4):1420-1425. doi:10.1021/ct500040r

54. Lyman E, Zuckerman DM. On the structural convergence of biomolecular simulations by determination of the effective sample size. *J Phys Chem B*. 2007;111(44):12876-12882. doi:10.1021/jp073061t

55. Huber GA, Kim S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys J*. 1996;70(1):97-110. doi:10.1016/S0006-3495(96)79552-8

56. Zuckerman DM, Chong LT. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu Rev Biophys*. 2017;46(1):43-57. doi:10.1146/annurev-biophys-070816-033834

57. Adhikari U, Mostofian B, Copperman J, Subramanian SR, Petersen AA, Zuckerman DM. Computational Estimation of Microsecond to Second Atomistic Folding Times. *J Am Chem Soc*. 2019;141(16):jacs.8b10735. doi:10.1021/jacs.8b10735

58. Zwier MC, Pratt AJ, Adelman JL, Kaus JW, Zuckerman DM, Chong LT. Efficient Atomistic Simulation of Pathways and Calculation of Rate Constants for a Protein-Peptide Binding Process: Application to the MDM2 Protein and an Intrinsically Disordered p53 Peptide. *J Phys Chem Lett*. 2016;7(17):3440-3445. doi:10.1021/acs.jpclett.6b01502

59. Boltzmann L. On gas-friction, and on the velocity of electricity in the electric current. *London, Edinburgh, Dublin Philos Mag J Sci*. 1880;9(56):307-309. doi:10.1080/14786448008626844

60. Fick A. Ueber Diffusion. *Ann Phys*. 1855;170(1):59-86. doi:10.1002/andp.18551700105

61. Einstein A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann Phys*. 1905;322(8):549-560. doi:10.1002/andp.19053220806

62. Agah S, Pasquali M, Kolomeisky AB. Theoretical analysis of selectivity mechanisms in molecular transport through channels and nanopores. *J Chem Phys*. 2015;142(4). doi:10.1063/1.4906234

63. Harris AL. Electrical coupling and its channels. *J Gen Physiol*. Published online 2018. doi:10.1085/jgp.201812203

64. Jentsch TJ. VRACs and other ion channels and transporters in the regulation of cell volume and beyond. *Nat Rev Mol Cell Biol*. 2016;17(5):293-307. doi:10.1038/nrm.2016.29

65. Nakamura R, Numata T, Kasuya G, et al. Cryo-EM structure of the volume-regulated anion channel LRRC8D isoform identifies features important for substrate permeation. *Commun Biol*. 2020;3(1). doi:10.1038/s42003-020-0951-z

66. Vassalle M. Contribution of the Na+/K+-pump to the membrane potential. *Experientia*. 1987;43(11-12):1135-1140. doi:10.1007/BF01945511

67. Sterratt DC. Goldman-Hodgkin-Katz Equations. In: *Encyclopedia of Computational Neuroscience*. Springer New York; 2015:1300-1302. doi:10.1007/978-1-4614-6675-8_229

68. Kwon T, Harris AL, Rossi A, Bargiello TA. Molecular dynamics simulations of the Cx26 hemichannel: Evaluation of structural models with Brownian dynamics. *J Gen Physiol*. 2011;138(5):475-493. doi:10.1085/jgp.201110679

69. Lee K Il, Jo S, Rui H, et al. Web interface for brownian dynamics simulation of ion transport and its applications to beta-barrel pores. *J Comput Chem*. 2012;33(3):331-339. doi:10.1002/jcc.21952

70. Sachs JN, Crozier PS, Woolf TB. Atomistic simulations of biologically realistic transmembrane potential gradients. *J Chem Phys*. 2004;121(22):10847-10851. doi:10.1063/1.1826056

71. Roux B. The membrane potential and its representation by a constant electric field in computer simulations. *Biophys J*. 2008;95(9):4205-4216. doi:10.1529/biophysj.108.136499

72. Roux B, Allen T, Bernèche S, Im W. Theoretical and computational models of biological ion channels. *Q Rev Biophys*. 2004;37(1):15-103. doi:10.1017/S0033583504003968

73. Kutzner C, Grubmüller H, De Groot BL, Zachariae U. Computational electrophysiology: The molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophys J*. 2011;101(4):809-817. doi:10.1016/j.bpj.2011.06.010

74. Gumbart J, Khalili-Araghi F, Sotomayor M, Roux B. Constant electric field simulations of the membrane potential illustrated with simple systems. *Biochim Biophys Acta - Biomembr*. 2012;1818(2):294-302. doi:10.1016/j.bbamem.2011.09.030

75. Escalona Y, Garate JA, Araya-Secchi R, Huynh T, Zhou R, Perez-Acle T. Exploring the Membrane Potential of Simple Dual-Membrane Systems as Models for Gap-Junction Channels. *Biophys J*. 2016;110(12):2678-2688. doi:10.1016/j.bpj.2016.05.005

76. Allen TW, Kuyucak S, Chung SH. Molecular dynamics study of the KcsA potassium channel. *Biophys J*. 1999;77(5):2502-2516. doi:10.1016/S0006-3495(99)77086-4

77. Pezeshki S, Chimerel C, Bessonov AN, Winterhalter M, Kleinekathöfer U. Understanding ion conductance on a molecular level: An all-atom modeling of the

bacterial porin OmpF. *Biophys J*. 2009;97(7):1898-1906. doi:10.1016/j.bpj.2009.07.018

78. Feynman RP, Leighton RB, Sands M. *The Feynman Lectures on Physics, Vol. I: The New Millennium Edition: Mainly Mechanics, Radiation, and Heat*. Vol 1. Basic books; 2011.

Chapter 3

Structure of native lens connexin 46/50 intercellular channels by Cryo-EM

This chapter has been published

Structure of native lens connexin-46/50 intercellular channels by cryo-EM.

Nature 654, 372-377 (2018). Janette B. Myers[1, †], Bassam G. Haddad[1, †], Susan E. O'Neill[1], Dror S. Chorev[2], Craig C. Yoshioka[3], Carol V. Robinson[2], Daniel M. Zuckerman[3] & Steve L. Reichow[1]*

[1]Department of Chemistry, Portland State University, Portland, OR, USA.

[2]Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford, UK.

[3]Department of Biomedical Engineering, Oregon Health and Sciences University, Portland, OR, USA.

[†]These authors contributed equally: Janette B. Myers, Bassam G. Haddad

*e-mail: reichow@pdx.edu

The success of this project is due to the work and effort many individuals. Bassam Haddad conducted and analyzed all of the molecular dynamics simulations, under the guidance of Daniel Zuckerman. Janette Myers prepared, screened, collected and CryoEM grids with the help of Craig Yoshioka. Janette Myers processed all CryoEM data. Susan O'Neil performed biochemical characterization and isolation of the lens gap junctions, without which there wouldn't be a project. Dror Chorev, of Carol Robinson's lab performed and analyzed mass spectrometry data. Steve Reichow directed the project.

Abstract

Gap junctions establish direct pathways for cell-to-cell communication through the assembly of twelve connexin subunits that form intercellular channels connecting neighbouring cells. Co-assembly of different connexin isoforms produces channels with unique properties and enables communication across cell types. Here, we used single-particle cryo-electron microscopy to investigate the structural basis of connexin co-assembly in native lens gap junction channels composed of connexin 46 and connexin 50 (Cx46/50). We provide the first comparative analysis to connexin 26 (Cx26), which, together with computational studies, elucidates key energetic features governing gap junction permselectivity. Cx46/50 adopts an open-state conformation that is distinct from the Cx26 crystal structure, yet it appears to be stabilized by a conserved set of hydrophobic anchoring residues. 'Hot spots' of genetic mutations linked to hereditary cataract formation map to the core structural–functional elements identified in Cx46/50, suggesting explanations for many of the disease-causing effects.

Main Text

Cell-to-cell communication directed by gap junctions is essential to neuronal function and cardiac coupling, and for coordinating intercellular signalling and metabolic activity in most tissues (for example, heart, skin, liver and eye lens)[1]. Genetic mutation or aberrant regulation of gap junctions is linked to a variety of pathological conditions, including cardiac arrhythmia, stroke, blindness, deafness, skin disease and cancers[2–4].

Intercellular channel formation occurs through an assembly of twelve connexin subunits[5]. Within the plasma membrane, six connexins are organized into a hemichannel structure. Hemichannels from neighbouring cells dock together to form complete cell-to-cell channels, which cluster to form large gap junction plaques. A remarkably large channel pore provides passage to diverse chemical messages; these include ions, metabolites, hormones, and other small signalling molecules less than about 1 kDa in size (for example, K^+, cyclic AMP (cAMP), inositol triphosphate ($Ins(1,4,5)P_3$) and glucose). In this way, interconnected cells can exchange electrical and chemical information across an entire tissue or organ.

Humans express 21 connexin isoforms in a cell-type-specific fashion[6]. Most cells express multiple isoforms, and certain connexins display an ability to co-assemble, either by docking two hemichannels composed of different isoforms (heterotypic), or through mixed isoform assembly within the same hemichannel (heteromeric). This complexity is thought to allow cells to fine-tune the conductance of chemical messages and support coupling across different cell types[7]. However, our understanding of the physical basis of connexin isoform compatibility, conductance, substrate selectivity and channel gating remains limited[8,9], as high-resolution structural information obtained by crystallographic analysis has so far been restricted to just a single model system, Cx26[10,11].

To gain further insight into the mechanistic effects of gap junction isoform diversity and heteromeric assembly, we applied single-particle imaging methods by cryo-electron microscopy (cryo-EM) to elucidate the structure of native channels made up of Cx46 and Cx50 (Cx46/50), isolated from the eye lens, in which connexin-mediated communication

is required for growth, differentiation and maintenance of lens transparency to support vision[12]. Comparative molecular dynamics simulations reveal key features of ion permeation and selectivity, and suggest that Cx46/50 adopts a more stable open-state conformation compared to the previously described Cx26 crystal structure[10].

Structural overview of Cx46/50

Cx46 and Cx50 form intercellular channels in the mammalian lens, which are potentially heteromeric or heterotypic[13,14] . We isolated native Cx46/50 intercellular channels from core lens tissue (sheep Cx44/49), and verified heteromeric co-assembly by biochemical analysis and chemical cross-linking mass spectrometry (Extended Data Fig. 1). The structure of these Cx46/50 intercellular channels was resolved by single-particle cryo-EM to near-atomic resolution (3.4 and 3.5 Å, from two independent datasets) (Fig. 1a, b, Extended Data Figs. 1–3). The resulting density maps revealed a 15-nm long dodecameric (12-mer) channel with a girdled waist (about 6–9 nm wide). There is a large unobstructed pore, approximately 1.4 nm in diameter, along the channel axis, consistent with the proposed open-state conformation (Fig. 1b).

We were unable to resolve a specific pattern of Cx46/50 heteromeric or heterotypic co-assembly using 3D classification or refinement strategies (Methods and Extended Data Figs. 4, 5). Nevertheless, high-resolution features corresponding to side-chain densities are observed throughout the reconstructions following 12-fold symmetry refinement (that is, by averaging signal contributed by both Cx46 and Cx50). Therefore, these two

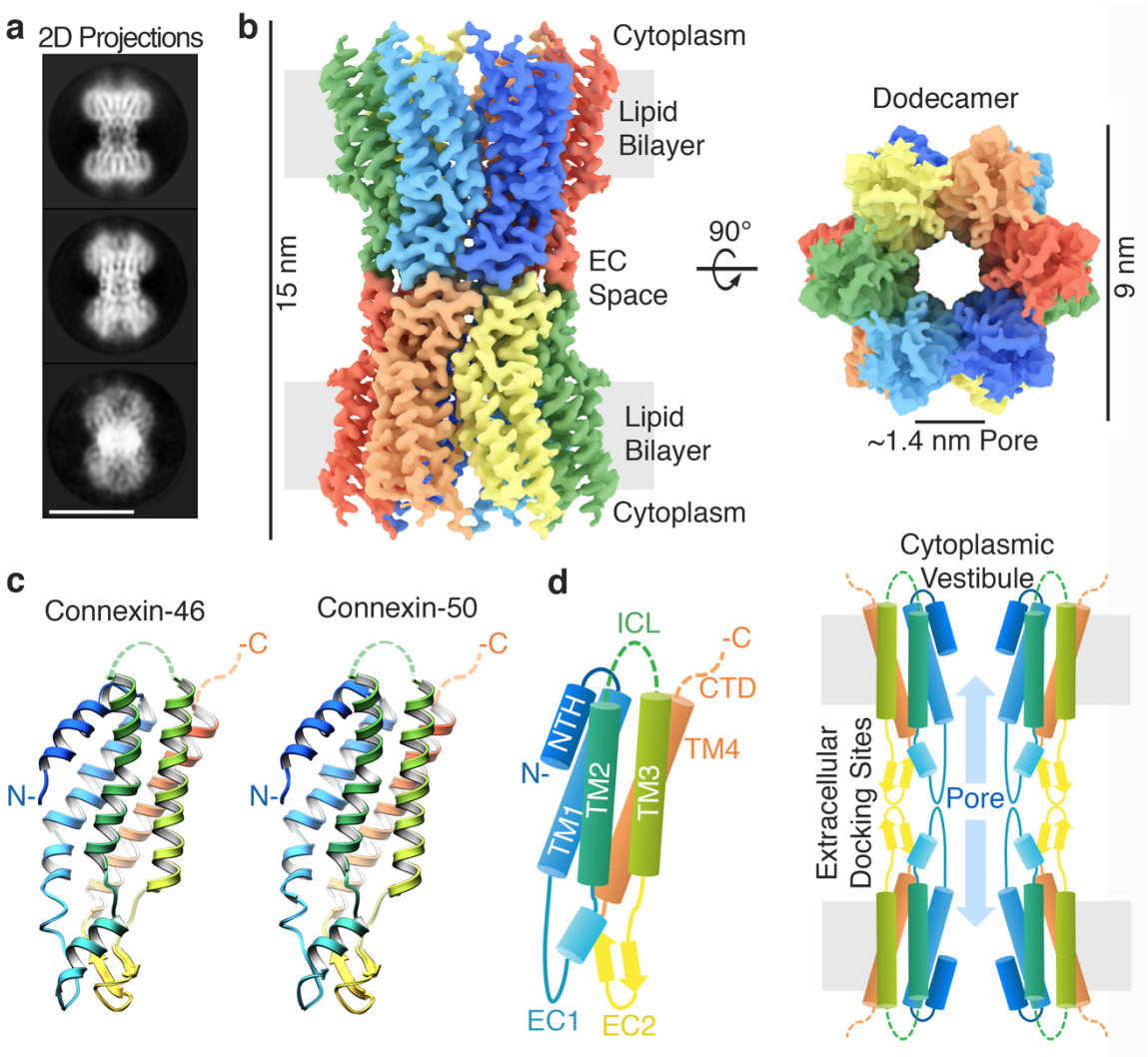


Fig. 1 | Structure of Cx46/50 intercellular channels. a, Representative projection averages of native Cx46/50 intercellular channels (selected from 165 classes). Scale bar, 10 nm. **b**, 3D cryo-EM reconstruction displaying the twelve subunits coloured independently. Regions corresponding to the lipid bilayer, cytoplasm and extracellular (EC) space are indicated. **c**, Atomic models of Cx46 (left) and Cx50 (right) monomers displayed in ribbon representation. **d**, Schematic of the connexin fold (left) and cross section of the assembled intercellular channel (right), with domains TM1–TM4, EC1, EC2 and the NTH domain. The ICL and CTD were not visualized in the cryo-EM density (dotted lines). EC1 and EC2 form docking sites, establishing a continuous channel pore of ~1.4 nm diameter connecting the cytoplasms of neighbouring cells.

highly similar 3D structure (Fig. 1b, c, Extended Data Figs. 4–6), consistent with the ability of Cx46/50 to co-assemble in a variety of heteromeric and/or heterotypic states.

Atomic models for Cx46 and Cx50 were built into the averaged cryo-EM density map, and various heteromeric and heterotypic channels were constructed for comparative analysis.

Whereas the structures display excellent validation statistics (Extended Data Fig. 2), local resolution assessment of the atomic models and experimental density identified features of both models that were less well defined by the density map, in particular at sites where the two isoforms differ in sequence (Extended Data Figs. 3–5 and Supplementary Tables 2, 3). Analysis of the presented models should be approached with caution owing to intrinsic limitations of our heterogeneous dataset, which may extend beyond local differences in primary sequence; for example, owing to the possibility of one isoform being more well-ordered and potentially biasing interpretation.

The refined Cx46/50 structures comprise four alternating transmembrane □-helices (TM1–TM4), two extracellular domains (EC1 and EC2) connecting TM1–TM2 and TM3–TM4, respectively, and an N-terminal helix (NTH) domain that folds into the channel vestibule and is connected to the pore-lining TM1 helix via a short linker (Fig. 1c, d). Density for Met1 is not observed in the cryo-EM maps and was shown, by tandem mass spectrometry (MS/MS), to be removed in both Cx46 and Cx50. The resulting N-terminal glycine (G2) is partially acetylated (Extended Data Fig. 1), as shown for the bovine isoforms[15,16]. The intracellular loop (ICL) connecting TM2–TM3 and the cytoplasmic C-terminal domain (CTD) harbouring the native cleavage sites of Cx46 and Cx50 are also not resolved[16]. The ICL and CTD were also not observed in the crystallographic structures of Cx26[10,11], presumably due to intrinsic disorder in these regulatory domains.

The close structural similarity between Cx46 and Cx50 results in highly similar interfacial interactions that include conserved regions of hydrophobic packing over the transmembrane region, and a highly similar hydrogen-bond–ion-pair network between

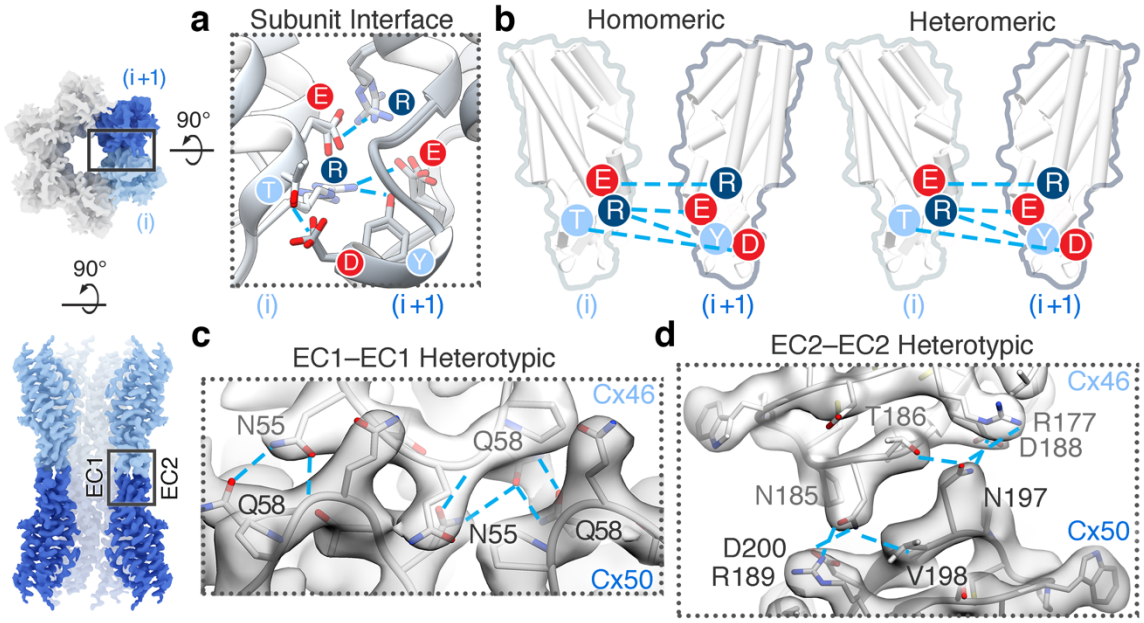


Fig. 2 | Heteromeric and heterotypic interactions between Cx46 and Cx50. a, **b**, Structural overlay (**a**) and illustration (**b**) of the conserved electrostatic network between neighbouring subunits (i and i + 1) identified in homomeric Cx46 and Cx50 and heteromeric Cx46/50 models. Labels are coloured according to amino acid charge characteristics (red, negative; dark blue, positive; light blue, polar). **c**, **d**, Magnified view of EC1–EC1 (**c**) and EC2–EC2 (**d**) docking-site interactions, with atomic models of Cx46 and Cx50 hemichannels built into the cryo-EM density in a heterotypic configuration. Conserved amino acids involved in hydrogen-bond pairing (cyan lines) are labelled.

adjacent (heteromeric interface, Fig. 2a, b, Extended Data Fig. 6) and opposed subunits (heterotypic interface, Fig. 2c, d, Extended Data Fig. 6). Most of these stabilizing interactions are present in Cx26[10,17], including the EC1 Q/N motif (Fig. 2c) and the EC2 pairing involving the K/R–N–D motif (Fig. 2d), a conserved element among group I heterotypic compatible isoforms[18]. Although Cx46/50 and Cx26 are not classified as heteromeric compatible channels[19], the conserved features at the heteromeric interface are congruent with the current understanding that heteromeric co-assembly of connexins is established during biogenesis in the ER–Golgi network[19].

Overall, despite significant sequence differences, Cx46 and Cx50 (□-family connexins) display core structural features that are very similar to the □-family member, Cx26[10] (pair-

wise C_α root mean squared deviation (r.m.s.d.) = 2.18 Å and 2.14 Å versus Cx46 and Cx50, respectively). These different connexin family members thus share a conserved connexin fold and gap junction channel architecture, as presented in Fig. 1d, and our structures are consistent with early low-resolution electron diffraction studies on Cx43 obtained in a lipid bilayer[20,21]. Despite these general similarities, however, we uncovered substantial differences between Cx46/50 and Cx26 localized to key functional sites, which we expect to contribute to isoform-specific permeation and selectivity properties and provide insight into the interactions responsible for fully stabilizing the open-state conformation of these channels, detailed below.

Energetics of ion permeation and selectivity

Comparisons between Cx50, Cx46 and Cx26 intercellular channels reveal distinct electrostatic pore pathways, with shared regions of negative charge potential and steric constriction sites formed by the NTH domains that narrow the cytoplasmic vestibule to around 10–12 Å at both ends of the channel (Fig. 3a, b). The pore diameters are within the range determined for other connexin channels[22], and fitting with the general ability of gap junctions to enable a variety of molecules of less than about 1 kDa in size (such as those in Fig. 3b) to cross between cells. However, these channels can display a substantial level of isoform-specific selectivity for molecules below this size cut-off, including discrimination between small charged ions[23].

To validate our structural models and gain insight into the mechanism of ion selectivity, we conducted comparative all-atom molecular dynamics simulations and potential-of-mean-force (PMF) calculations to define the free-energy landscape of potassium (K^+) and

chloride (Cl^-) permeation for Cx50, Cx46 and Cx26 (Fig. 3c, d, Extended Data Figs. 7, 8). PMFs obtained for Cx26 should be interpreted cautiously owing to significant dynamical behaviour observed for the NTH domain during molecular dynamics simulations (Extended Data Fig. 7), described in detail in the following section.

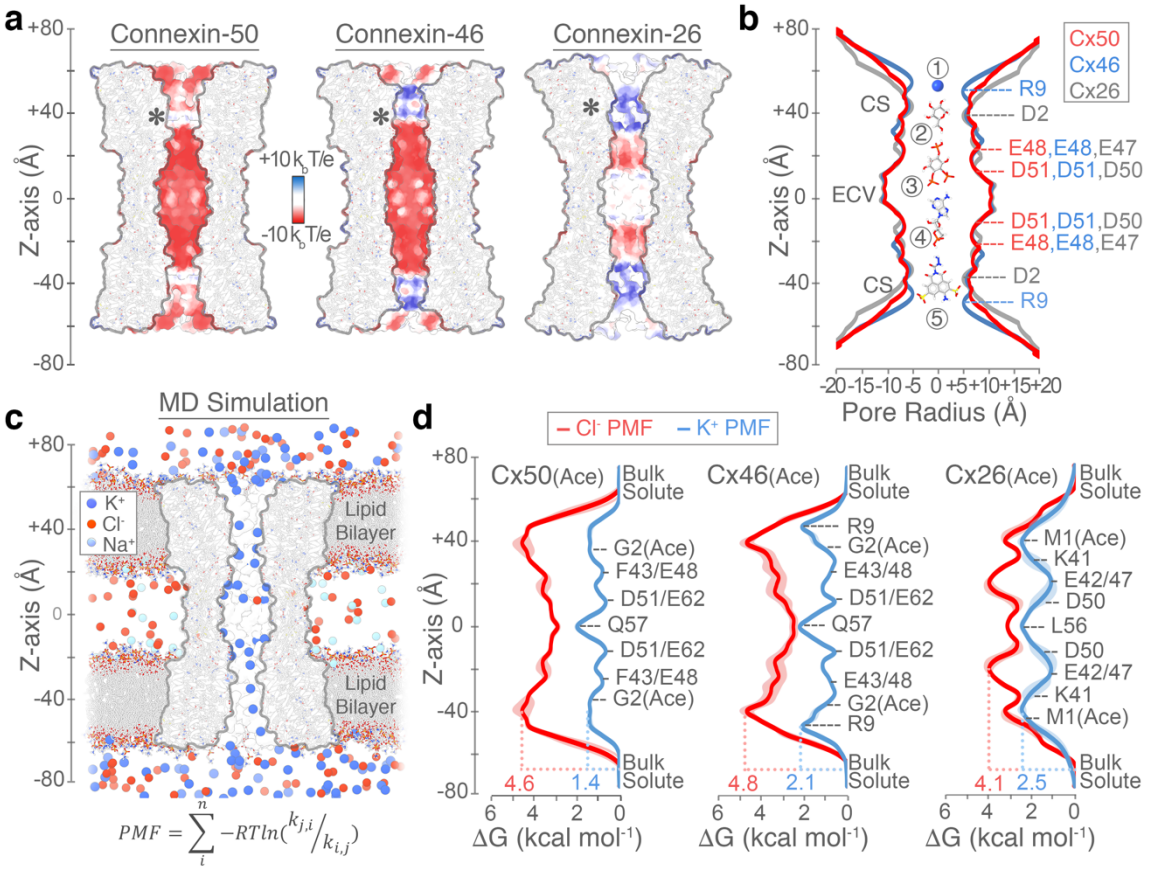


Fig. 3 | Comparative pore profile and energetics of ion permeation. a, Cut-away surface representation of Cx50 (left), Cx46 (centre) and Cx26 (right) (PDB 2ZW3[10]; residues 2–96 and 132–217), coloured by coulombic potential (red, negative; white, neutral; blue, positive). *T*, temperature; k, Boltzmann constant; e, charge of an electron; *, location of N terminus. **b**, Pore radius determined for experimental structures of Cx50 (red), Cx46 (blue) and Cx26 (grey, PDB 2ZW3[10]). Locations of constriction sites (CS) and ECV are indicated. Structures of representative substrates are displayed to scale: K^+ ion (1), glucose (2), Ins(1,4,5)P_3 (3), cAMP (4) and Lucifer yellow (5). **c**, Snapshot of the Cx50 molecular dynamics simulation, showing the membrane channel (white) embedded in two lipid bilayers and solvated in the presence of intracellular K^+, and extracellular Na^+ and Cl^- ions. Water molecules not shown. **d**, PMF describing the free-energy landscape (ΔG) experienced by K^+ ions (blue trace) and Cl^- ions (red trace) permeating the channel pore. Symmetrized values are shown for acetylated models of Cx50ac (left), Cx46ac (centre) and Cx26ac (right), with non-symmetrized values in lighter shading. Amino acid positions are presented for correlation purposes, and do not represent deconvolution of the free-energy components.

Cx50, Cx46 and Cx26 form high-conductance ion channels, with preference for conductance of cations over anions. For molecular dynamics simulation, the N terminus of each of the models was acetylated (Cx50ac, Cx46ac and Cx26ac), as this form is expected to represent the predominant species in vivo[24], and previous molecular dynamics studies suggest this co-translational modification is required to obtain physiologically relevant charge selectivity of Cx26[22]. PMFs for K^+, the major permeant ion, reveal peak energetic barriers within the constriction site, ranging from 1.4 kcal mol^{-1} for Cx50ac, to 2.1 kcal mol^{-1} in Cx46ac and 2.5 kcal mol^{-1} in Cx26ac (Fig. 3d). These relatively low barriers are similar to the peak energetic barrier determined for other high-conductance Na- and K-channels (~2–3 kcal mol^{-1})[25–27], and are consistent with the range of experimental unitary conductance values of these channels (around 220 pS for Cx50[28] versus 140–135 pS for Cx46[29] and Cx26[30], in 130–140 mM CsCl).

The differences in K^+ PMF correlate with isoform-specific differences in both steric and electrostatic environments. The constriction site of Cx50ac displays the lowest barrier and is characterized by a nearly completely electronegative coulombic potential, owing in part to neutralization of the N terminus by acetylation (shown by the asterisk in Fig. 3a and Extended Data Fig. 8). The major K^+ energy barrier of Cx46ac correlates with the position of the positively charged residue R9 ($z \approx 50$ Å, where z is distance along the pore axis) (Fig. 3d, Extended Data Fig. 8), which also limits the constriction site of Cx46 to about 10 Å in our model (versus about 12 Å for Cx50) (Fig. 3b). However, the cryo-EM density map is not well defined at this site (Supplementary Table 2), probably because (at least in part)

of the conformational flexibility of this residue, as dynamical behaviour is observed during molecular dynamics simulation. These dynamics of R9 effectively modulate the steric barrier of Cx46 (between about 10 and 12 Å). The constriction-site K^+ energy barrier of Cx26ac correlates with the location of the basic residue K41 ($z \approx 50$ Å) (Fig. 3d), located on TM1 just below the NTH domain, as previously reported[22,31].

Free energy minima for K^+ are localized within the extracellular vestibule (ECV; $z \approx 10$ Å and $z \approx 30$ Å) of all three isoforms (Fig. 3d), supporting the role of EC1 in establishing charge selectivity and conductance[31–33]. In Cx46 and Cx50, several negatively charged residues (for example, E48, D51 and E62) localize with regions of high K^+ ion density (Fig. 3c, d, Extended Data Fig. 9). Notably, charge substitutions at D51 resulted in decreased unitary conductance in Cx46 hemichannels[34]. E48 and D51 are conserved in Cx26 (equivalent to E47 and D50) (Fig. 3d), and establish transient binding interactions with K^+ ions during molecular dynamics simulation. These sites have also been implicated in Ca^{2+} regulation in Cx26 by X-ray crystallography[11], molecular dynamics studies[35,36] and by functional mutation studies of Cx46[36]. Therefore, competitive K^+ binding at these sites may contribute to the mechanism of Ca^{2+} regulation or sensitivity. E62 (in Cx46/50) appears to form an additional cation-binding site, through coordination between the carboxylate side chain and nearby backbone-carbonyl oxygens (Extended Data Fig. 9). E62 is not conserved in other human connexin isoforms (with the exception of Cx43), and may therefore constitute an isoform-specific regulatory site. Extracellular Ca^{2+} is involved in the mechanism of closing (or gating) connexin hemichannels[37], and competition by K^+

binding at this putative site may contribute to the mechanism of potentiation of Cx50 and Cx46 hemichannels by extracellular K^+ ions[38].

Cx50, Cx46 and Cx26 display an appreciable level of selectivity towards positively charged small ions, with permeability ratios of K^+ to Cl^- (P_{K+}/P_{Cl-}) ranging between around 2.5 to 10 (refs[28,32,39–41]). Hydrated K^+ and Cl^- ions, with a diameter of about 7 Å, would pass unobstructed through a 10–12 Å steric constriction site; yet for all three isoforms, the peak energy barriers to Cl^- are considerably larger than for K^+ (Fig. 3d, Extended Data Fig. 8). Peak Cl^- barriers localize within the constriction site region of Cx46ac and Cx50ac (4.8 kcal mol^{-1}, $z \approx 40$ Å; and 4.6 kcal mol^{-1}, $z \approx 50$ Å, respectively), and slightly deeper into the channel pore for Cx26ac, near the constriction-site–ECV border (4.1 kcal mol^{-1}, $z \approx 20$ Å). As a proxy for degree of P_{K+}/P_{Cl-} selectivity, we assessed the difference in peak K^+ and Cl^- PMF barriers ($\Delta\Delta G$ = 3.2 kcal mol^{-1} for Cx50ac, 2.7 kcal mol^{-1} for Cx46ac and 1.6 kcal mol^{-1} for Cx26ac). These relatively small differences in free energy are in accordance with their moderate P_{K+}/P_{Cl-} selectivity ratios, and on the order of those defined for bacterial sodium channels (around 3.0–3.5 kcal mol^{-1})[26,27], which display only modest selectivity for Na^+ over K^+ ($P_{Na+}/P_{K+} \approx 10–30$). By contrast, voltage-gated K^+ channels display almost ideal selectivity for K^+ over Na^+ ($P_{K+/Na+} \approx 1{,}000$), with energetic barrier differences to these ions reported to be about 6.6 kcal mol^{-1} for KcsA[25].

Diffusion of Cl^- ions across the constriction-site energy barriers was relatively rare on the timescale of our equilibrium molecular dynamics simulations, which necessitated enhanced sampling methods to construct robust Cl^- PMF calculations (see Methods and Extended Data Fig. 8). Nevertheless, a few Cl^- entry events were observed in our

simulation data for Cx50 and Cx46, and in these cases, Cl^- ions appear to co-migrate across the high energy barrier of the constriction site alongside a K^+ counter ion. It is possible that similar mechanisms involving ionic-charge neutralization enable cation-preferring gap junction channels to permit passage of negatively charged signalling molecules (for example, cAMP and Ins(1,4,5)P_3). However, it is difficult to provide a general mechanism for selectivity, as conductance properties for ions do not always correlate well with conductance properties for larger molecules[41,42].

The analysis above supports models proposing that substrate selectivity and conductance properties of gap junctions are established by complex mechanisms involving both steric aperture and the unique pattern of electrostatic features contributed by isoform-specific amino acid composition[42,43]. In this way, Cx46/50 heteromeric or heterotypic channels confer distinct conductance properties of potential functional significance. For example, rectification observed in Cx46/50 heterotypic channels can be explained by the resulting asymmetric free-energy landscape (Extended Data Fig. 8) induced by the uneven distribution of fixed charges[30,40]. Cx46/50 heteromeric assemblies also produced unique K^+–Cl^- PMF profiles, with peak barriers that were intermediate to their homomeric counterparts (Extended Data Fig. 8), supporting observations made from single-channel measurement[29,44].

Additional fine tuning of gap junction permeation properties may be achieved through co-translational and/or post-translational modification of pore-lining residues[22]. In our studies, N-terminal acetylation was found to enhance the cation-to-anion specificity of Cx50, Cx46 and Cx26 intercellular channels (Extended Data Fig. 8). Although N-terminal

acetylation is irreversible, the effect of this co-translational modification illustrates how other dynamic and reversible charge-modifying post-translational modifications may serve to spatially and temporally modulate the behaviour of intercellular communication.

Open-state stabilization of the NTH domain

Despite general similarities, substantial differences between the Cx46/50 cryo-EM structures and the Cx26 crystal structure are localized to the NTH domain (Fig. 4, Extended Data Fig. 7). The connexin NTH domain contributes to ion selectivity and 'fast' trans-junctional voltage gating that is common to all connexin isoforms[45]. The NTH folds into the cytoplasmic vestibule where it forms the constriction site, and is well positioned to function as a selectivity filter or gating domain (Fig. 1d and Fig. 4a).

In the proposed open-state conformation of Cx46/50 described here, the NTH domain adopts a regular amphipathic □-helix and ordered loop connecting to TM1 (Fig. 4a–c). The hydrophobic face is established by a set of aromatic and hydrophobic residues that are conserved across various connexin isoforms (W4, L7, I10, L11 and V14 in Cx46 and Cx50) (Fig. 4b). These anchoring sites pack against the pore-lining helices (TM1–TM2), and along the interface of neighbouring subunits. Despite sequence conservation at these sites, the NTH domain modelled in the crystal structure of Cx26 is in a distinctively different conformation and overall arrangement with respect to the transmembrane domains compared to Cx46/50 ($C_{□}$ r.m.s.d. = 5.2 Å, after alignment of the transmembrane and extracellular domains) (Fig. 4d, Extended Data Fig. 7e). In Cx26, the NTH domain and loop

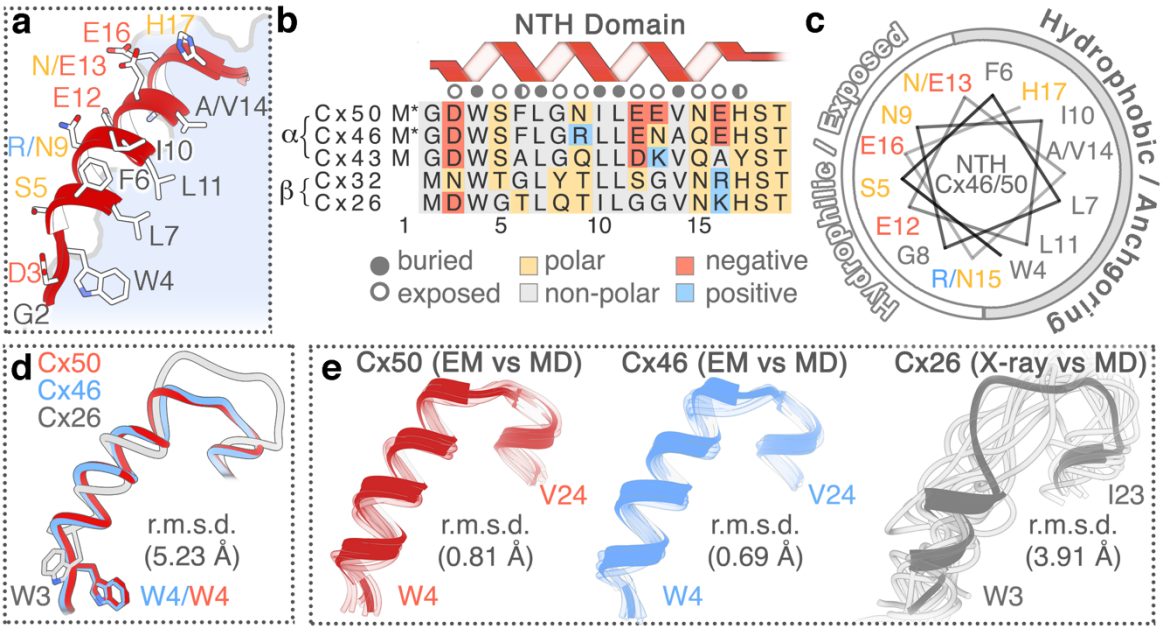


Fig. 4 | Open-state stabilization of the Cx46/50 NTH domain. a, Magnified view of the Cx46/50 NTH domain. **b**, Sequence alignment of the NTH domain from representative α- and β-family connexins. Residues are annotated as buried, exposed, or partially exposed. *, co-translational removal of M1, as confirmed by MS/MS. **c**, Helical wheel representation of the Cx46/50 NTH domain. In **a**–**c**, amino acids are labelled with a single letter for residues that are the same for both Cx46 and Cx50, and separately in the format (Cx46/Cx50), where they differ, and coloured by chemical properties (grey, hydrophobic; blue, positively charged; red, negatively charged; and yellow, hydrophilic). **d**, Overlay of NTH domains from experimental structures of Cx50 (red), Cx46 (blue) and Cx26 (grey; PDB 2ZW3[10]), after super-positioning of TM1–TM4, EC1 and EC2 domains. **e**, Superposition of NTH domains of each monomer captured from the molecular dynamics simulation (MD, faded tube), and aligned against the initial starting structures (X-ray or cryo-EM (EM)), Cx50ac (left), Cx46ac (centre) and Cx26ac (right), displayed as ribbons. C_α r.m.s.d. of the NTH domains are shown in **d** and **e**.

connecting TM1 is less regular, and with the exception of W3 (Cx26 numbering) the conserved hydrophobic residues are modelled towards the solvent[10].

We propose that the network of hydrophobic anchoring observed in the cryo-EM structure of Cx46/50 supports a stabilized open-state conformation. Accordingly, analyses of our molecular dynamics simulations show that the NTH domains of Cx50 and Cx46 are conformationally stable in both acetylated and non-acetylated states, with only small-amplitude backbone fluctuations (root mean square fluctuation (r.m.s.f.) ≈ 1.0–1.2 Å) (Fig. 4e, Extended Data Fig. 7). By contrast, the NTH domain of Cx26 (and Cx26ac)

becomes rapidly disordered (that is, unfolded), and remains conformationally dynamic throughout the production phase of our molecular dynamics simulations (using Protein Data Bank (PDB) code 2ZW3[10] as the starting structure; Fig. 4e, Extended Data Fig. 7). The dynamical behaviour of the Cx26 NTH domain is consistent with previous molecular dynamics studies[35,46,47]. The functional significance of the differences in NTH domain structure and dynamic stability is currently unclear. Indeed, instability of the Cx26 NTH domain may be an intrinsic feature. In a more recent X-ray crystallographic study of Cx26, the NTH domain was completely unresolved, presumably due to local disorder[11]; however, potential effects of the conditions required for crystallization cannot be ruled out.

The amphipathic nature of the Cx46/50 NTH positions hydrophilic residues implicated in voltage sensing and ion selectivity at the solvent-exposed face, forming the cytoplasmic vestibule[45] (Fig. 4a–c). A network of hydrogen-bond interactions appears to contribute to the precise localization of some of these key residues, including the site of N-terminal acetylation. The carbonyl group at the acetylated-G2 site appears to be oriented, at least transiently, through hydrogen bonding to the indole ring of W4 in the same subunit, whereas in the non-acetylated state, G2 forms a transient intermolecular ion pair with D3 of a neighbouring subunit (Extended Data Fig. 10). The side chain of D3 is oriented by a relatively stable intramolecular hydrogen bond with the hydroxyl of S5, in both acetylated and non-acetylated forms of both channels (Extended Data Fig. 10). D3 (in Cx46/50, D2 in Cx26) has been identified as a critical site for establishing polarity and/or sensitivity to trans-junctional voltage[48–50]. The precise spatial orientation imposed by these

interactions may contribute to the strict conservation at this site, as replacement of D3 to a similarly negatively charged residue, glutamate, results in significant perturbation to gating properties, conductance and free energy of the open–closed state in Cx50 gap junctions[51].

Mutational hot spots linked to hereditary cataracts

Cx46/50 gap junctions have a critical role in maintaining the transparency of the eye lens by establishing a pathway for water, ion and nutrient circulation and removal of metabolic waste in this avascular organ[12]. Consequently, a variety of human genetic variations in

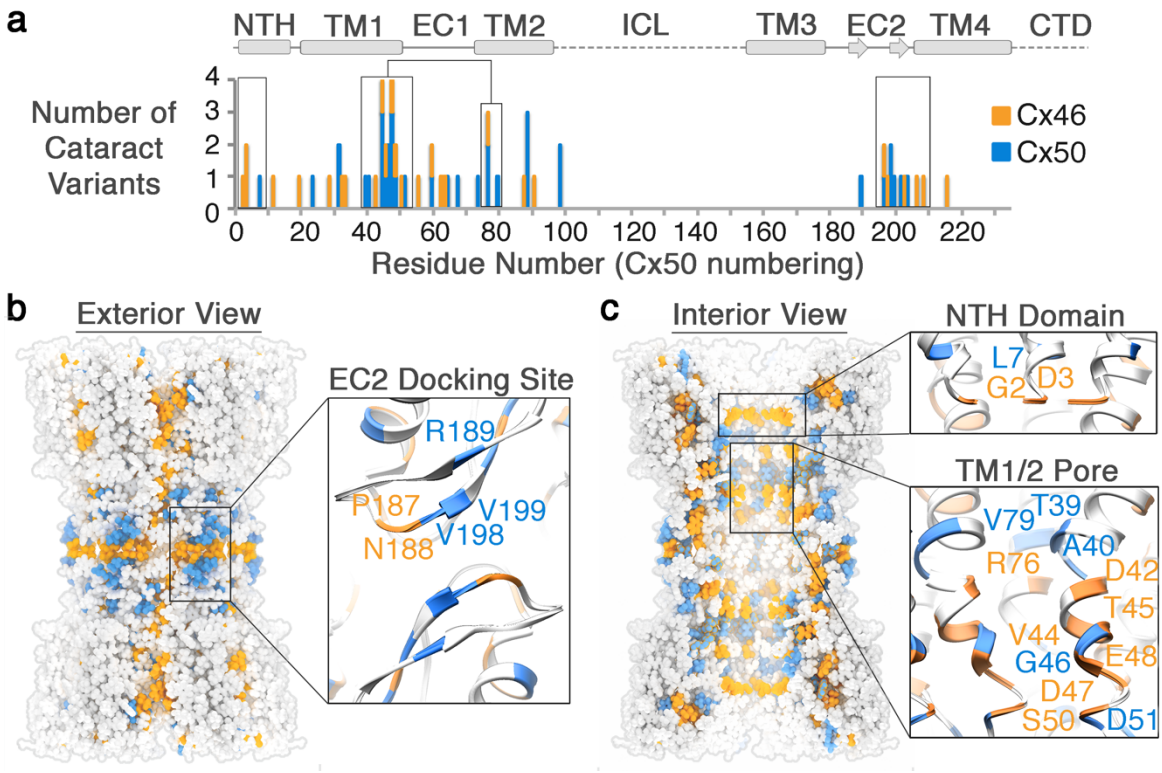


Fig. 5 | Mutational hot spots in Cx46 and Cx50 linked to congenital cataracts. a, Histogram of genetic variants of Cx46 and Cx50 linked to hereditary (congenital) cataracts. Sites of Cx46 (orange) and Cx50 (blue) mutations are overlaid using Cx50 amino acid numbering, with secondary structure and domain elements indicated. Hot spots, regions of high genetic variation, within the NTH, TM1, TM2 and EC2 domains are boxed. **b**, **c**, Exterior (**b**) and interior (**c**) view of the Cx46/50 gap junction channel with cataract mutation sites mapped for Cx46 (orange) and Cx50 (blue), with magnified views of the EC2 domain (**b**, inset), NTH domain (**c**, top inset) and TM1–TM2 pore-lining helices (panel **c**, bottom inset), with representative mutation sites labelled.

Cx50 and Cx46 have been linked to hereditary cataract formation[52]. Age-related cataracts are currently incurable (except by surgery) and remain the leading cause of blindness in the world[53]. The rarer congenital forms of this disease have been linked to genetic mutation of various lens proteins, including Cx46/50—offering critical insight into the mechanisms of maintaining lens transparency throughout life[54]. We mapped 46 mutation sites in Cx46/50, currently reported on the Cat-Map database[54], that are linked to congenital cataracts (Fig. 5a). This analysis suggests explanations for many of the disease-causing effects induced by these polymorphisms, as the mutational hot-spots localize to functionally important regions of the Cx46/50 gap junction structure. These include a cluster of residues localized within the EC2 docking site—for example, Cx46(N188T/I) and Cx50(R189Q/W)—in regions that deviate markedly from the Cx26 structure, such as the NTH gating or selectivity domain—such as Cx46(G2D) and Cx46(D3Y/H), and Cx50 L7P)—and several sites localized to the TM1–TM2 pore-lining helix that form an interaction network with the NTH domain, where mutation is expected to affect the permeation pathway or impact folding and stability within these regions (Fig. 5b, c). The localization of disease-causing mutations underscores the functional significance of these core structural–functional elements, and the importance of proper cell-to-cell communication through Cx46/50 gap junctions for the maintenance of lens transparency. The ability of cryo-EM to provide high-resolution structural information on gap junctions may finally enable detailed mechanistic investigation of these disease-causing mutations in Cx46/50 and in other isoforms responsible for a diverse range of connexinopathies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://www.nature.com/articles/s41586-018-0786-7#additional-information

Received 11 September 2017; accepted 29 October 2018

Acknowledgements We thank T. Gonen for early support of this work; T. White, L. David, U. Adhikari and B. Mostofian for helpful discussions; the staff at the OHSU Multiscale Microscopy Core and Advanced Computing Center, and W. Garrick (PSU) for their assistance and training. C.V.R. and D.S.C. are supported by funding from the European Research Council (No. 695511-ENABLE). D.M.Z. and C.C.Y. are supported by the Center for Spatial Systems Biomedicine at OHSU. S.L.R. is supported by the Medical Research Foundation of Oregon and the National Institutes of Health (R35-GM124779).

Author Contributions J.B.M. and B.G.H. contributed equally. J.B.M. and C.C.Y. collected the cryo-EM datasets. J.B.M. performed image processing and atomic modelling of Cx46/50. S.E.O. performed protein purification and negative-stain electron microscopy studies. B.G.H. conducted and analyzed the molecular dynamics simulations. D.S.C conducted the cross-linking studies and MS/MS analysis. C.V.R. contributed to the experimental design of MS/MS studies. D.M.Z. contributed to the experimental design and statistical analysis of the molecular dynamics simulations. All authors contributed to

manuscript preparation. S.L.R. provided overall guidance to the design and execution of this work.

Competing Interests

The authors declare no competing interests.

Corresponding Author

Correspondence and requests for materials should be addressed to S.L.R.

reichow@pdx.edu

METHODS

Cx46/50 purification and amphipol reconstitution

Lamb eyes were obtained from the Wolverine Packers slaughterhouse (Detroit, MI), and the lenses were removed using a surgical blade and stored at −86 ° C. Gap junction intercellular channels were isolated from the core lens fibre tissue, containing C-terminal truncation variants of Cx46 and Cx50 (also known as MP38)[16,55–57] (Extended Data Fig. 1). Details of the purification procedure are provided below.

Core lens-fibre cell tissue was dissected from cortical tissue using a surgical blade, and stripped membranes were prepared as described[58–60]. Total protein concentration was determined by BCA (Pierce) and membranes were stored at −86 °C suspended in storage buffer containing 10 mM Tris pH8.0, 2 mM EDTA, 2 mM EGTA, at a total protein concentration of ~2 mg ml^{-1}. Stripped membranes were thawed from −86 °C and solubilized in 10 mM Tris pH8.0, 2 mM EDTA, 2 mM EGTA, 1% (w/v) *n*-decyl-□-d-maltoside

(DM) for 30 min at 37 °C. Unsolubilized debris was cleared by ultracentrifugation at 150,000*g* for 30 min at 4 °C. The solubilized fraction was applied to an anion-exchange chromatography column (UnoQ, BioRad) equilibrated with buffer A (10 mM Tris pH 8.0, 2 mM EDTA, 2 mM EGTA, 0.3% DM (w/v)). Protein was eluted with buffer B, which additionally contained 500 mM NaCl. Elution peaks containing Cx46/50, as determined by SDS–PAGE, were pooled and applied to a size-exclusion chromatography (SEC) column (ENC650, BioRad) equilibrated with SEC buffer (20 mM HEPES, 150 mM NaCl, 2 mM EDTA, 2 mM EGTA and 0.3% DM (w/v)). Peak fractions containing purified Cx46/50 were pooled and protein concentration was determined by UV absorbance. All chromatography steps were performed at 4 °C. The presence of both Cx46 and Cx50 was confirmed by western blot analysis using polyclonal antibodies directed against the N-terminal domain of Cx46 (AP11570PU-N, Acris) and C-terminal domain of Cx50 (sc-50432, Santa Cruz) (Extended Data Fig. 1a) and by mass-spectrometry analysis, described below.

Purified Cx46/50 was exchanged from DM to amphipol A8-35 (Anatrace), as follows. Amphipol was added to freshly purified protein in an 5:1 amphipol:protein (w/w) ratio using a stock solution prepared at 5% (w/v) in water. This mixture was incubated for 2.5 h at 4 °C with rotation. Detergent was then removed by application of SM-2 Biobeads (BioRad) at a ratio of 30:1 (w/w) beads:detergent. Biobeads were incubated overnight at 4 °C with rotation. Biobeads were then removed by running samples over a Polyprep column (BioRad) that had been washed with detergent-free SEC buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 2 mM EDTA, 2 mM EGTA). Samples were further cleared by ultra-centrifugation at 150,000*g* for 20 min at 4 °C. The clarified sample was then applied to an

SEC column (ENC650, BioRad) equilibrated with detergent-free SEC buffer to remove excess amphipol. Peak fractions corresponding to amphipol-stabilized Cx46/50 were pooled and concentrated for single-particle electron microscopy studies (Extended Data Fig. 1b). Final protein concentration was determined by UV absorbance at 280 nm.

Chemical cross-linking and mass spectrometry

Cx46/50 was prepared for cross-linking and mass-spectrometry analysis to confirm heteromeric assembly of the two lens isoforms (Extended Data Fig. 1d–f). Crosslinking was carried out using the amine-to-amine cross-linking reagents bis(sulfosuccinimidyl)suberate (BS3; Thermo Fischer Scientific) and disuccinimidyl suberate (DSS H12:D12; Creative Molecules). For BS3 cross-linking, 0.5 µl of 12.5 mM BS3 dissolved in deuterium-depleted water (DDW) were added to 10 µl of purified 20 µM Cx46/50 complexes and incubated for 2 h on ice. For DSS cross-linking, 0.5 µl of a 1:1 mixture of 25 mM non-deuterated (d0) and deuterated (d12) DSS cross-linker dissolved in DMSO was added to 10 µl purified 20 µM Cx46/50 complexes. A control sample, in which only DMSO or DDW were added to the proteins was also prepared. DSS cross-linked samples were incubated at room temperature for 2 h in a thermomixer at 300 revolutions per min (r.p.m.), whereas BS3 cross-linked samples were incubated for 2 h on ice. The cross-linking reaction was quenched by adding Tris pH7.4 at a final concentration of 100 mM for 15 min at room temperature in a thermomixer at 300 r.p.m.

The quenched reaction mixtures were separated on a NuPAGE gel (Thermo Fischer Scientific) and protein bands stained with InstantBlue (Expedeon). Cross-linked protein

bands were excised and digested with trypsin (Promega) as described[61]. Peptides were resuspended in 0.1% formic acid and separated on an Ultimate 3000 UHPLC system (Thermo Fischer Scientific) and electrosprayed directly into a QExactive mass spectrometer (Thermo Fischer Scientific) through an EASY-Spray nano-electrospray ion source (Thermo Fischer Scientific). The peptides were trapped on a C18 PepMap100 pre-column (300 µm i.d. × 5 mm, 100 Å, Thermo Fisher Scientific) using solvent A (0.1% formic acid in water) at a pressure of 500 bar. The peptides were separated on an in-house packed analytical column (75 µm internal diameter packed with ReproSil-Pur 120 C18-AQ, 1.9 µm, 120 Å, Dr.Maisch) using a gradient (length:50 min, 15% to 38% for 30 min followed by 38% to 58% solvent B (0.1% formic acid in acetonitrile, flow rate: 200 nl/min) for 15 min. Raw data was acquired on the mass spectrometer in a data-dependent mode (DDA). Full-scan mass spectra were acquired in the Orbitrap (scan range 350–2000 *m*/*z*, resolution 70000, AGC target 3×10^6, maximum injection time 50 ms). After the mass spectrometry scans, the 10 most intense peaks were selected for HCD fragmentation at 30% of normalized collision energy. HCD spectra were also acquired in the Orbitrap (resolution 17500, AGC target 5×10^4, maximum injection time 120 ms) with first fixed mass at 180 *m*/*z*. Charge exclusion was selected for 1+ and 2+ ions. The dynamic exclusion set to 5 s. Cross-linking identification and analysis was done using pLink[62] and Xcalibur 2.2 (Thermo Scientific). All peptides were manually validated.

For identification of proteins and post-translational modifications, protein bands were excised from the gel and processed as described above. Mass-spectrometry analysis was carried out similarly with a gradient of 15–38% for 30 min and the Orbitrap set to 350–

1500 *m*/*z*. Charge exclusion was selected for 1+ and unassigned ions, dynamic exclusion was set to 5 s. PTM identification was done using the MASCOT Daemon client program.

Negative-stain electron microscopy

Amphipol-stabilized Cx46/50 was prepared for negative-stain electron microscopy as described[58,63]. In brief, 3 □l sample (~0.02 mg ml^{-1}) was applied to a glow-discharged continuous carbon coated electron microscopy specimen grid (Ted Pella), blotted with filter paper and washed two times with detergent-free SEC buffer. The specimen was then stained with freshly prepared 0.75% (w/v) uranyl formate (SPI-Chem).

Negatively stained specimens were visualized on a 120 kV TEM (iCorr, FEI) at a nominal magnification of 49,000× at the specimen level (Extended Data Fig. 1c). Digital micrographs were recorded on a 2k × 2k CCD camera (FEI Eagle) with a calibrated pixel size of 4.37 Å. A total of 75 micrographs were collected. Contrast transfer function (CTF) parameters were determined in EMAN2[64] and micrographs free of significant astigmatism and drift were selected based on Thon rings in the power spectra. A total of 5,330 particles were hand selected in EMAN2 and extracted with a box size of 84 × 84 pixels. Reference-free 2D class averages were generated using CTF-corrected (phase-flipped) images without applied symmetry (Extended Data Fig. 1c). A subset of 3,952 'good' particles was selected following multiple rounds of 2D classification, and an initial model was generated de novo in EMAN2 using a subset of 12 class averages as input. This model was refined against the 'good' particle image dataset in EMAN2 with applied D6 symmetry to a final resolution of ~20 Å (Extended Data Fig. 2).

Cryo-EM data collection, image processing and 3D reconstruction

Samples were prepared for cryo-EM by applying 5 □l of amphipol-stabilized Cx46/50 (2.35 mg ml^{-1}) to a glow-discharged holey carbon grid (Quantifoil R 1.2/1.3) for 10 s. The grid was blotted for 4.0 s and plunge frozen in liquid ethane using a Vitrobot (FEI) at 100% humidity and stored under liquid nitrogen.

Cryo-EM specimen grids were imaged on a Titan Krios (FEI) operated at 300 kV. Image stacks were recorded using a K2 summit direct electron detector (Gatan) in counting mode with a super-resolution pixel size of 0.665 Å/pix. The dose rate was 3.2 electrons $pixel^{-1}$ s^{-1}, with 4 frames s^{-1} collected for a total exposure time of 10 s. A Gatan energy filter with a slit width of 30 eV was used during data collection. An initial dataset of 1,104 micrographs (dataset 1) was obtained by automated data collected using SerialEM[65], with nominal defocus values from 1.25 to 2.5 µm.

Drift correction and dose weighting was performed using MotionCor2[66] and CTF correction was performed using GCTF[67]. 261,206 particles were picked from dataset 1 using DoGPicker[68]. Particles were extracted with 2× binning (resulting in a pixel size of 1.3 Å per pixel). Five rounds of 2D classification in Relion 2.0[69] left 53,791 'good' particles. These particles were then subjected to 3D classification in Relion with four classes and no imposed symmetry. The most populated class contained 33,967 particles. These particles were unbinned and another round of 3D classification was performed, reducing the population of particles to 30,128. 3D auto-refinement was then performed on this set of

particles with D6 symmetry imposed. After masking and post processing in Relion, the final map had a resolution of 3.4 Å by gold standard FSC (Extended Data Figs. 2, 3a).

An additional 1,093-micrograph dataset was collected and processed as above. Particles were picked from this set with DoGPicker, 2× binned and pooled with the original set of 261,206 particles for a total of 398,066 particles (dataset 2). A set of 66,480 'good' particles was obtained after five rounds of 2D classification. These particles were subjected to 3D classification with four classes and no imposed symmetry. The most populated class contained 55,475 particles. This dataset was further culled by removing particles extracted from micrographs with Thon rings that did not extend beyond 3.5 Å, resulting in a final dataset of 44,547 particles. These particle images were subjected to 3D auto-refinement in Relion with D6 symmetry, resulting in an overall 3.5 Å resolution 3D reconstruction after post-processing as judged by gold-standard FSC (Extended Data Fig. 3b). Local resolution analysis using BlocRes[70] showed the 3.4 Å map possessed exceptionally high resolution features within the central regions of the structure, whereas the 3.5 Å reconstruction contained more uniformly defined features throughout the density map, consistent with visual inspection (Extended Data Fig. 3c, d). An overview of cryo-EM data collection and 3D refinement statistics is provided in Extended Data Fig. 2.

Cx46/50 symmetry analysis

In an attempt to uncover a specific pattern(s) of Cx46/50 heteromeric/heterotypic co-assembly, 3D auto-refinement was also pursued in Relion using C1, C3, C6, and D3 symmetries, using the final 30,128-particle (dataset 1) and 3.4 Å map (filtered to 15 Å) as

input. These refinements converged to 4.1 Å (C1), 3.9 Å (C3), and 3.7 Å (D3 and C6). Examination of the resulting maps provided no indication that the Cx46 and Cx50 subunits were being separately resolved (Extended Data Figs. 4, 5). Further attempts were performed using 3D classification in Relion with C3, C6 and D3 symmetry, using the larger 55,475 particle set (which had already been subjected to one round of 3D classification with no imposed symmetry). The initial model was the 3.4 Å map filtered to 25 Å. No resolution limit was enforced, and classification was attempted with and without image alignment. Some classifications converged to a single class, whereas others maintained a more even distribution of particles throughout 3D classification. 3D auto-refine was attempted with the most populated class from each attempted symmetry group. C3 symmetry refined to 3.9 Å from a set of 47,074 particles; C6 symmetry refined to 3.8 Å from a set of 38,404 particles; D3 symmetry refined to 4.2 Å from a set of 16,520 particles. Inspection of the resulting maps provided no indication that isoform-specific features were being separately resolved into any specific symmetric arrangements (not shown). Finally, focused refinement strategies with signal subtraction were also explored using Relion, by masking a single hemichannel or just a single subunit. However, these procedures did not produce isoform-specific features, or improved results compared to the D6-symmetrized maps.

As we were unable to identify a specific pattern of co-assembly for the Cx46/50 dodecameric channel, all further analysis and model building was performed using the 3D maps generated with imposed D6 symmetry. Both pre-processed and post-processed

maps and associated masks generated from datasets 1 and 2 have been deposited to the Electron Microscopy Data Bank under accession code EMD-9116.

Atomic modelling, refinement and validation

The post-processed maps obtained with D6 symmetry were used to build and stereochemically refine atomic models for both Cx46 and Cx50, following similar procedures. An initial C_{α} model was generated using the available crystal structure of Cx26 (PDB 2ZW3[10]) and placed into the post-processed 3.4 Å density map using rigid-body fitting. Starting from this template, all atom models of Cx46 and Cx50 were built separately into the cryo-EM density using COOT[71]. Disulfide bonds were modelled for Cx50 (C54–C201, C61–C195 and C65–C190) and Cx46 (C54–C189, C61–C183 and C65–C178). Models were subjected to real-space refinement in Phenix[72] with non-crystallographic symmetry (D6-symmetry) and secondary structure restraints imposed. Successive rounds of modelling and refinement were conducted until refinement statistics converged, as judged by Molprobity[73] (Extended Data Fig. 2). The Fourier shell correlation (FSC) of the model versus map dropped below 0.5 at 3.4 Å (dataset 1) and 3.5 Å (dataset 2) for both Cx46 and Cx50, judged by the output of Phenix real-space refine (Extended Data Fig. 3a, b). The NTH domain of Cx46 and Cx50 (residues 2–20) were further refined using the post-processed 3.5 Å density map (dataset 2), as this region of the map was more well defined compared to the original 3.4 Å map. Over areas of the density maps where the sequence of Cx46 and Cx50 are identical or similar (80% identical and 8% similar) both models fit well into the D6-symmetrized map, and these regions tend to

display well-resolved side-chain density. Over regions where the sequence of Cx46 and Cx50 differ, side-chain density is sometimes weaker. This observation is possibly due to the imposed D6 symmetry averaging the density of two different side chains in these areas, or relative flexibility as many of these residues contain solvent-exposed side chains. In these areas of difference, where electron microscopy density is observed, both Cx46 and Cx50 can be fit into the density equally well (Extended Data Figs. 4, 5). Fit of the models to the cryo-EM density map were assessed quantitatively by local resolution analysis using BlocRes[70], comparing the calculated maps of Cx50 and Cx46 atomic models to the 3.4 Å experimental cryo-EM map (Extended Data Fig. 3). This analysis was tabulated by assigning each residue a range of resolution values corresponding to the output of this analysis, including the alpha carbon and extending to the end of the side chain (Supplementary Tables 2, 3).

Completed models of the dodecameric structures, corresponding to residues 2–97, 142–222 (Cx46), and 2–97, 154–234 (Cx50) , have been deposited to the Protein Data Bank with accession numbers 6MHQ and 6MHY, respectively). Additional density is observed for the region of TM2 that extends towards the cytoplasm; however, we did not model this region (corresponding to ~1–2 turns of an □-helix) due to the lack of identifiable side-chain density. Various heterotypic/heteromeric models of Cx46/50 were generated for analysis by applying appropriate symmetry operations to the monomeric subunits and combined to form a complete gap junction structure. Coulombic surface potentials were calculated and displayed using Chimera[74].

Molecular dynamics simulations

Visual Molecular Dynamics (VMD) v.1.9.3[75] was used to build systems for the Cx50, Cx46, Cx46/50 heteromeric and heterotypic models, and for Cx26 (PDB 2ZW3)[10]. Representative Cx46/50 heteromeric models (heteromeric models I and II, with C3 or D3 point group symmetry, respectively) were constructed by applying the appropriate symmetry operations to the coordinates of the individual subunits. The Cx46/50 heteromeric and heterotypic channels were run through a steepest descent minimization routine using Phenix[72] to ensure no clashes were introduced in the preparation of these models. Each system comprised the full dodecameric gap junction, and was prepared in explicit solvent and embedded in two lipid bilayers composed of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC), mimicking a cell–cell junction. The Cx26 crystal structure was prepared for molecular dynamics by completing the side chains at residues Lys15, Ser17 and Ser19 and missing protons were added to all amino acids at standard positions. Side chains were protonated according to neutral conditions, and the HSD model was used for all histidine residues. To facilitate comparison to the Cx46/50 models, the Cx26 model was constructed with a Met1 residue added, which was missing in the published crystal structure, but was expected to be present in the protein based on proteomic analysis[76], as previously described[47]. Disulfide bonds identified in the experimental structures were enforced for Cx50 and Cx46 (as described above), and for Cx26[10] (C53–C180, C60–C174 and C64–C169). Amino acids corresponding to the intracellular loop (ICL) connecting TM2–TM3, and the C-terminal domain (CTD) of Cx50, Cx46 and Cx26 and were not included for molecular dynamics simulation, as experimental data describing the

structure of these large domains (~50 residue ICL and ~200 residue CTD in Cx46/50) are missing. The introduced N- and C-terminal residues resulting from the missing ICL segment (Cx46 L97 and L142; Cx50 V97 and L154; and Cx26 G109 and K125) were neutralized. N-terminal acetylation sites were introduced in VMD through an all-atom acetylation patch in the automated PSF-Builder. A complete list of modelled residues for each system is provided in Supplementary Tables 1.

The prepared protein structures were submerged in a hydration shell using Solvate v.1.0.1[77]. Water was removed from sections of the channel corresponding to transmembrane domains, based on hydrophobic character and localization of amphipol observed in the experimental cryo-EM data (~20–50 Å from the centre of the channel). The VMD membrane-builder plugin was used to add two POPC bilayers, with dimensions of 152 × 152 Å for Cx46, Cx50 and Cx46/50 models, and 155 × 155 Å for Cx26, and lipids overlapping with protein were removed. The entire system was then placed in a water box with dimensions 150 × 150 × 180 Å for Cx46, Cx50 and Cx46/50 models, and 150 × 150 × 183 Å for Cx26, using VMD's Solvate plugin. The system was neutralized using the Autoionize plugin, then 150 mM KCl and 150 mM NaCl were added to the solvent areas corresponding to intracellular and extracellular regions of the simulation box, respectively (see Fig. 3c). A summary of atoms counts for each system is provided in Supplementary Table 1.

GPU-accelerated nanoscale molecular dynamics (NAMD) v.2.12[78] was used for all classical molecular dynamics simulations, using the CHARMM36 force field[79,80] for all atoms and TIP3P explicit model for water. Each system was prepared following the same

minimization and equilibration protocol as follows. An initial minimization of the lipid tails, with all other atoms fixed, was performed for 1 ns with a 1-fs time step, allowing the tails to 'melt'. Next, the system, including lipids, solvent and ions were allowed to minimize around the protein, with the protein harmonically constrained for 1 ns. For the Cx46/50 heteromeric/heterotypic and acetylated models, a second minimization step was applied, where the system was free to minimize with a harmonic constraint on the protein backbone to ensure stable quaternary structure. The entire system was then released from restraints and subjected to all-atom equilibration runs using Langevin thermostat, with a constant temperature of 310 K and constant pressure of 1 atm, with 1 or 2-fs time steps and allowed to proceed for 30 ns (see Supplementary Table 1). Periodic boundary conditions were used to allow for the particle mesh Ewald (PME) calculation of electrostatics. Finally, all of the models were continued for a minimum of 50 ns of production. Root mean squared deviations (r.m.s.d.) and root mean squared fluctuations (r.m.s.f.) were calculated using VMD. All three gap junctions approached a steady r.m.s.d. within 20 ns of the equilibration phase (Extended Data Fig. 7a, b). All of these systems maintained an electro-chemical seal to extracellular sodium ions (Na^+) during molecular dynamics simulation (e.g., Fig. 3c), validating the stability of intermolecular docking-site interactions and the various heteromeric/heterotypic models generated for analysis.

Calculation of the PMF with respect to K^+ and Cl^- was performed using the fundamental principle of detailed balance via a one-dimensional Markov state model (MSM). Configuration space was subdivided based on a natural coordinate, the channel pore (*z* axis), and segmented into bins of 4 Å in length. Using a lag-time of 2 ps, a transition matrix

was calculated from the trajectories of individual ions within the simulation. The $i \rightarrow j$ transition probability $k_{i,j}$ is computed using equation (1):

$$k_{i,j} \cong \frac{N_{i,j}}{N_i} \quad \text{Eq. 1}$$

where $N_{i,j}$ is the count of transitions during the lag interval and N_i is the count of ions in bin *i* at the beginning of each lag interval. PMFs were constructed using the principal of detailed balance:

$$P_i^{\mathrm{eq}} k_{i,i+1} = P_{i+1}^{\mathrm{eq}} k_{i+1,i} \quad \text{Eq. 2}$$

$$e^{\frac{-\Delta G_{i,i+1}}{RT}} = \frac{P_{i+1}^{\mathrm{eq}}}{P_i^{\mathrm{eq}}} = \frac{k_{i,i+1}}{k_{i+1,i}} \quad \text{Eq. 3}$$

$$\mathrm{PMF}(i) = \Delta G(i) = \sum_{n=1}^{i-1} -RTln\left(\frac{k_{n,n+1}}{k_{n+1,n}}\right) \quad \text{Eq. 4}$$

Here, P^{eq} is the equilibrium probabilities for an ion to occupy the respective bin (equation (2)), $\Delta G_{i \rightarrow i+1}$ is the free energy difference from bin *i* to bin *i*+1, R is the gas constant (1.986 cal mol^{-1} K^{-1}), and *T* is temperature (310 K) (equations (3), (4)). Final PMF values were adjusted so that the values of the bulk solvent were zero. PMF curves in Fig. 3 and Extended Data Fig. 7 were derived by mapping *z* values to the corresponding bin index *i* and subsequently smoothed using Microsoft Excel. The detailed-balance (rates-based) approach is justified by the high mobility of ions within the channel pore, and the timescales used for analysis were validated by assessing the convergence of the unsymmetrized data to the symmetrized values presented in Fig. 3d and Extended Data Fig. 8. The results were shown to closely match PMFs constructed by taking the

population profile, or average counts, $\langle N_i \rangle$, of the K^+ ions along the channel pore (*z* axis) and solving: $\Delta G = -RT\ln(\langle N_i \rangle)$ [11,81,82] (Extended Data Fig. 8d).

Because the detailed-balance approach requires only local equilibrium sampling, we were able to apply a distributed seeding protocol to construct PMFs for Cl^- ions. Initial analysis of Cl^- trajectories revealed this ion to be poorly sampled inside the channel pore of the Cx50, Cx46 and Cx26 models, presumably owing to an energetic barrier presented by significant regions of negative coulombic potential for each of these systems (see Fig. 3a). Therefore, a distributed seeding approach was employed, where a single Cl^- ion was randomly introduced (seeded) by replacing a K^+ ion within the pore of the equilibrated channel. These coordinates were energy minimized and initial velocities were randomized before allowing the simulation to proceed for 10 ns. This procedure was repeated for each system 3–16 times until sufficient sampling was achieved, as determined by monitoring the resulting PMFs. Cl^- PMFs were constructed based on the transition rate (as described above), which is not sensitive to the initial placement of the ion. This seeding approach was validated by showing that the resulting PMF recapitulated the features of a Cl^- PMF obtained for Cx46, where sufficient sampling had been achieved through random diffusion (Extended Data Fig. 8e). Trajectories from these distributed seeding simulations were combined with the production phase data and included in the MSM for calculation of final Cl^- PMFs.

Analysis of hydrogen bonding within the NTH domains of Cx46 and Cx50 models was performed by recording the distance versus time of potential donor-acceptor pairs. The three sets of interactions probed were potential intermolecular hydrogen bonds between

D3 and the neighbouring N-terminal G2 residue, the intramolecular hydrogen bonding between D3 and S5, and the intramolecular hydrogen bonds between N-terminal acetyl and W4 for acetylated models of Cx46 and Cx50. To simplify the analysis of hydrogen-bonding interactions involving equivalent rotameric donor-acceptor configurations, heavy atoms were selected for D3 (C_{γ}) and G2 (N) for analysis (Extended Data Fig. 10). For comparison, equivalent analysis for Cx26 was conducted (between D2 and M1 and T5 of the adjacent subunit). For Cx26, D2 appeared to form intramolecular hydrogen-bonding pairing with T5; however, stable intermolecular-pairing interactions with Met1 or T5 (as indicated in the crystal structure[10]) were not identified during the production phase of these simulations (data not shown), as previously reported[35].

The NTH domain of Cx26 was found to be unstable (that is, rapidly unfolding) during molecular dynamics simulation in either the acetylated or non-acetylated states (Fig. 4e, Extended Data Fig. 7). The significance of this dynamical behaviour is not clear; however, we attribute this feature as a potential reason for the variation of our calculated PMFs of Cx26 compared to previous studies[47]. Notably, Kwon et al. reported the Cx26 channel to be anion selective in the absence of N-terminal acetylation[47]. These authors conducted an elegant set of experiments employing a grand canonical Monte-Carlo Brownian dynamics (GCMC/BD) based approach for modelling ion conductance using a model of the Cx26 hemichannel (by extracting a single hexamer of the Cx26 intercellular channel). However, a limitation of GCMC/BD method is that the protein structure is held static, and the resulting PMFs obtained by this approach would therefore be strongly influenced by

the selected conformational state of the Cx26 NTH domain. These caveats should be considered when comparing results presented in this work.

Statistical analysis

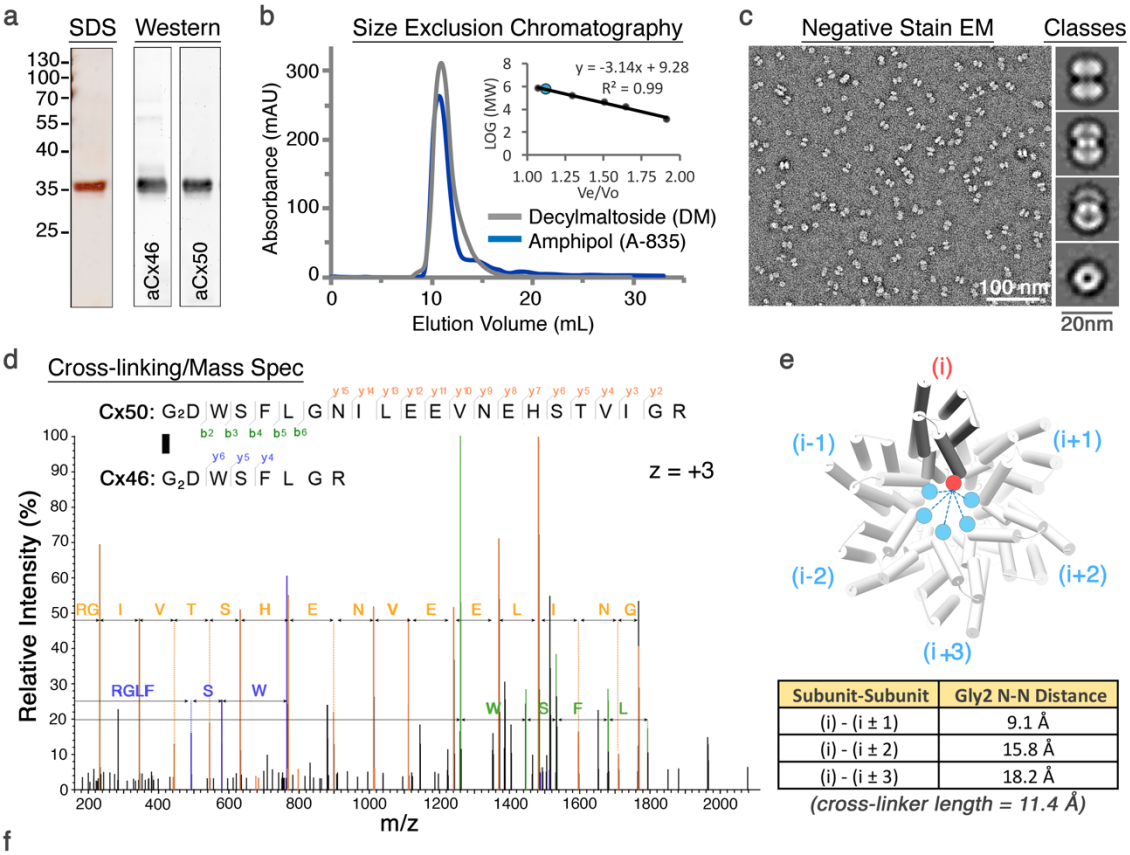
95% confidence intervals for comparison of Cα r.m.s.f. values were calculated using a two-tailed student's *t*-test. No statistical methods were used to predetermine sample size for the cryo-EM datasets. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

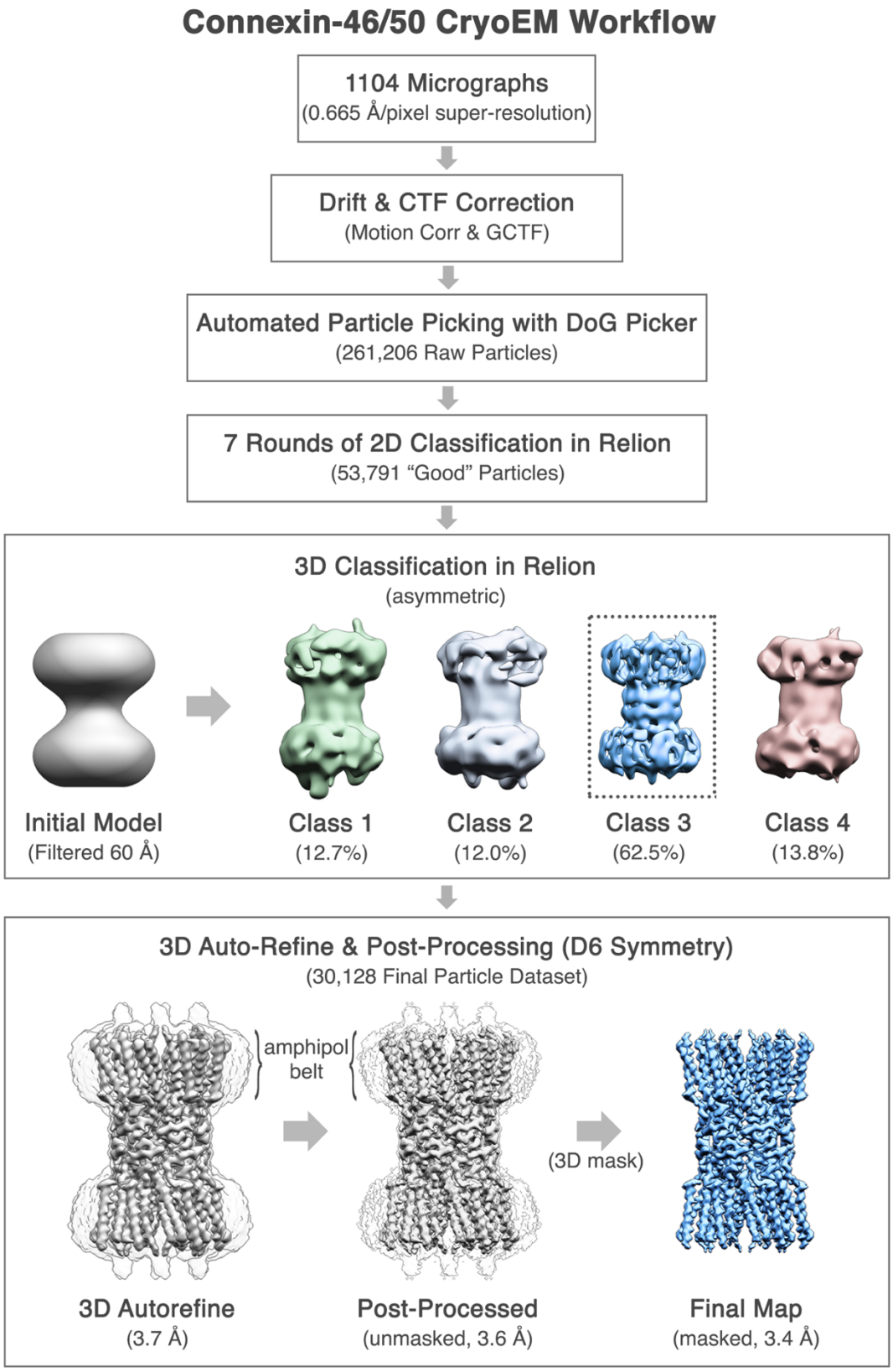
Data availability

Cryo-EM density maps have been deposited to the Electron Microscopy Data Bank under accession number EMD-9116. Coordinates for Cx46 and Cx50 atomic models have been deposited to the Protein Data Bank under accession codes 6MHQ and 6MHY. The original multi-frame micrographs have been deposited to the Electron Microscopy Public Image Archive under accession code EMPIAR-10212.

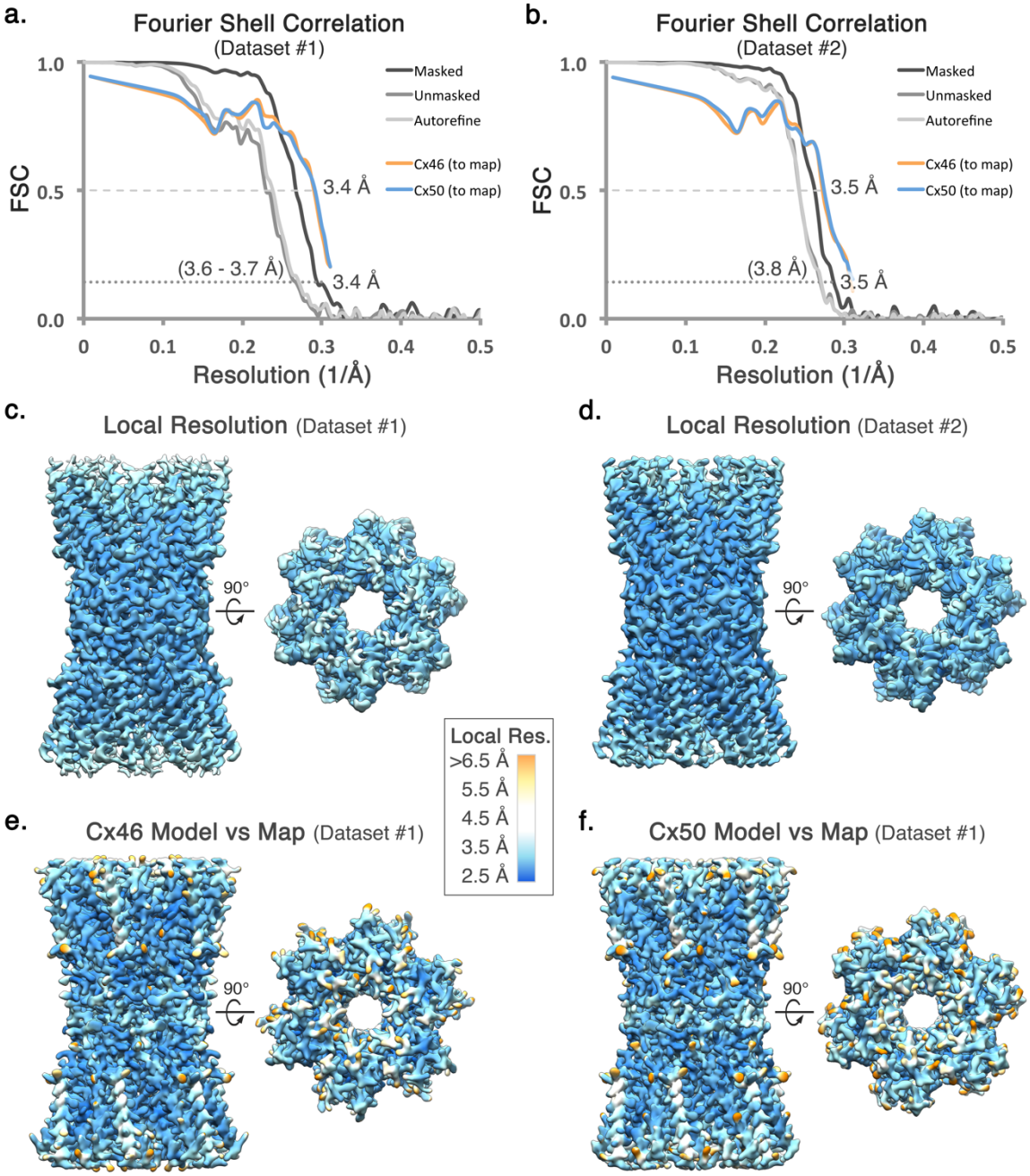


Protein 1	AA position	Peptide	Protein 2	AA position	Peptide	Cross-linker
Cx46	115	EEEPPKAAGPAEEHQDPAPVRDDR	Cx46	115	EEPPKAAGPAEEHQDPAPVR	DSS
Cx50	2	GDWSFLGNILEEVNEHSTVIGR	Cx50	2	GDWSFLGNILEEVNEHSTVIGR	DSS
Cx50	2	GDWSFLGNILEEVNEHSTVIGR	Cx46	2	GDWSFLGR	DSS
Cx46	2	GDWSFLGR	Cx46	2	GDWSFLGR	DSS
Cx50	271	AKGYQLLEEEK	Cx46	2	GDWSFLGR	BS3
Cx46	2	GDWSFLGR	Cx46	2	GDWSFLGR	BS3
Cx46	2	GDWSFLGR	Cx46	105	MEEKRK	BS3
Cx46	2	GDWSFLGR	Cx50	143	SSSSSKGTK	BS3

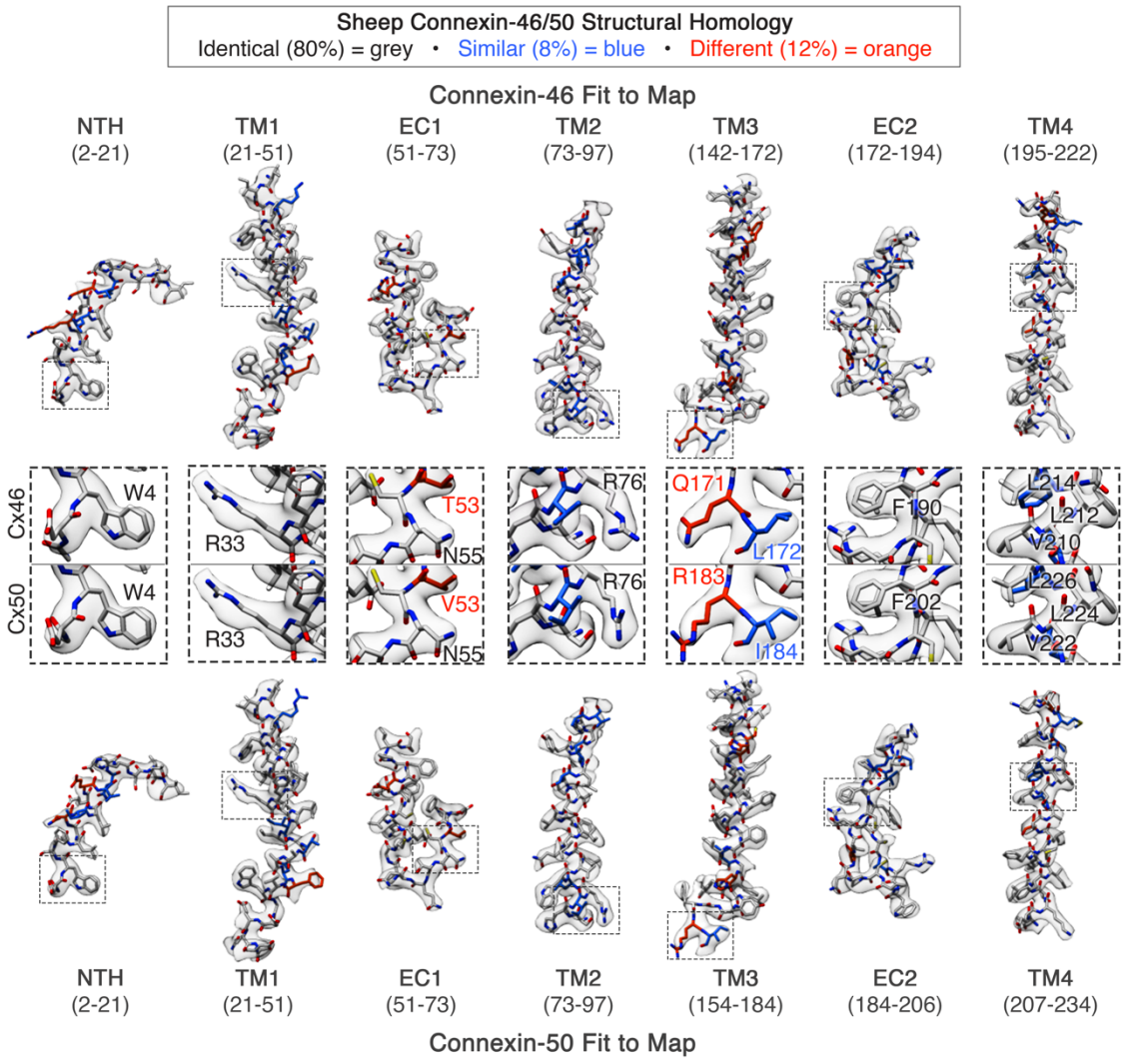
Extended Data Fig. 1 | Preliminary structural analysis of native lens Cx46/50 gap junction intercellular channels. a, Denaturing protein gel and western blot analysis of purified Cx46/50 (also known as MP38) isolated from lens core tissue. Protein bands corresponding to Cx46 and Cx50 co-migrate at a molecular weight of ~38 kDa, consistent with age-dependent proteolytic cleavage within the C-terminal domain of both isoforms[16]. Experiment performed 3 times with similar results. **b**, Size-exclusion chromatography elution profile of Cx46/50 gap junctions reconstituted in decyl-maltoside (DM, grey trace) or amphipol (A-835, blue trace), monitored by UV absorbance. Experiment performed more than 3 times with similar results. Inset, calibration curve (n = 3 runs) demonstrating that Cx46/50 elutes at an apparent molecular weight of ~560 kDa, consistent with the size of a dodecameric protein complex (12 × ~38 kDa) and two micelles (2 × ~50 kDa). **c**, Electron micrograph of negatively stained Cx46/50 gap junctions reconstituted into amphipol. Scale bar, 100 nm. Inset, representative 2D class averages of negatively stained particles (selected from 25 classes). Scale bar, 20 nm. **d–g**, Chemical cross-linking and mass spectrometry. **d**, Representative MS/MS m/z spectrum, identifying inter-subunit cross-linking at the N-terminal Gly2 positions of Cx50 and Cx46. Identified peaks in the m/z spectrum and amino acid identities are indicated (Cx50 b-ions, green; Cx50 y-ions, yellow; Cx46 y-ions, blue). MS/MS data represent the consensus of 3 independent runs. **e**, Structural analysis of cross-linking results, showing inter-subunit distances between the symmetrically related N-terminal Gly2 positions within the connexin hemichannel, ranging from 9.1 Å (i to $i \pm 1$), 15.8 Å (i to $i \pm 2$) and 18.2 Å (i to $i \pm 3$). The cross-linker spacer length is 11.4 Å, indicating a probable (i to $i \pm 1$) arrangement of Cx50 and Cx46 within the same hemichannel, although other arrangements cannot be ruled out. **f**, Overview of identified inter-subunit cross-links between Cx50 and Cx46 assembled gap junctions. Residues in red indicate the site of primary amines involved in the cross-linking reaction using either DSS or BS3. All detected inter-subunit crosslinks are between cytoplasmic domains. **g**, Schematic showing sites of inter-subunit cross-linking between Cx46 and Cx50 (black lines) and post-translational modifications identified during proteomics analysis (yellow, phosphorylation; purple, N-terminal acetylation). Met1 was determined to be removed in both Cx46 and Cx50 and the resulting N-terminal Gly2 position was identified in both acetylated and non-acetylated forms of Cx46 and Cx50, consistent with the specificity of the NatA acetylation complex[24]. Also shown are the predicted CTD cleavage sites in Cx46 and Cx50 (orange), based on previous analysis of bovine Cx46/50 isolated from lens core tissue[16]. Secondary structure and domain labels are indicated for the NTH, TM1–TM4, EC1, EC2, ICL and CTD.


Connexin-46/50 CryoEM Workflow
1104 Micrographs
(0.665 Å/pixel super-resolution)
Drift & CTF Correction
(Motion Corr & GCTF)
Automated Particle Picking with DoG Picker
(261,206 Raw Particles)
7 Rounds of 2D Classification in Relion
(53,791 "Good" Particles)
3D Classification in Relion
(asymmetric)
Initial Model
(Filtered 60 Å)
Class 1
(12.7%)
Class 2
(12.0%)
Class 3
(62.5%)
Class 4
(13.8%)
3D Auto-Refine & Post-Processing (D6 Symmetry)
(30,128 Final Particle Dataset)
amphipol
belt
(3D mask)
3D Autorefine
(3.7 Å)
Post-Processed
(unmasked, 3.6 Å)
Final Map
(masked, 3.4 Å)

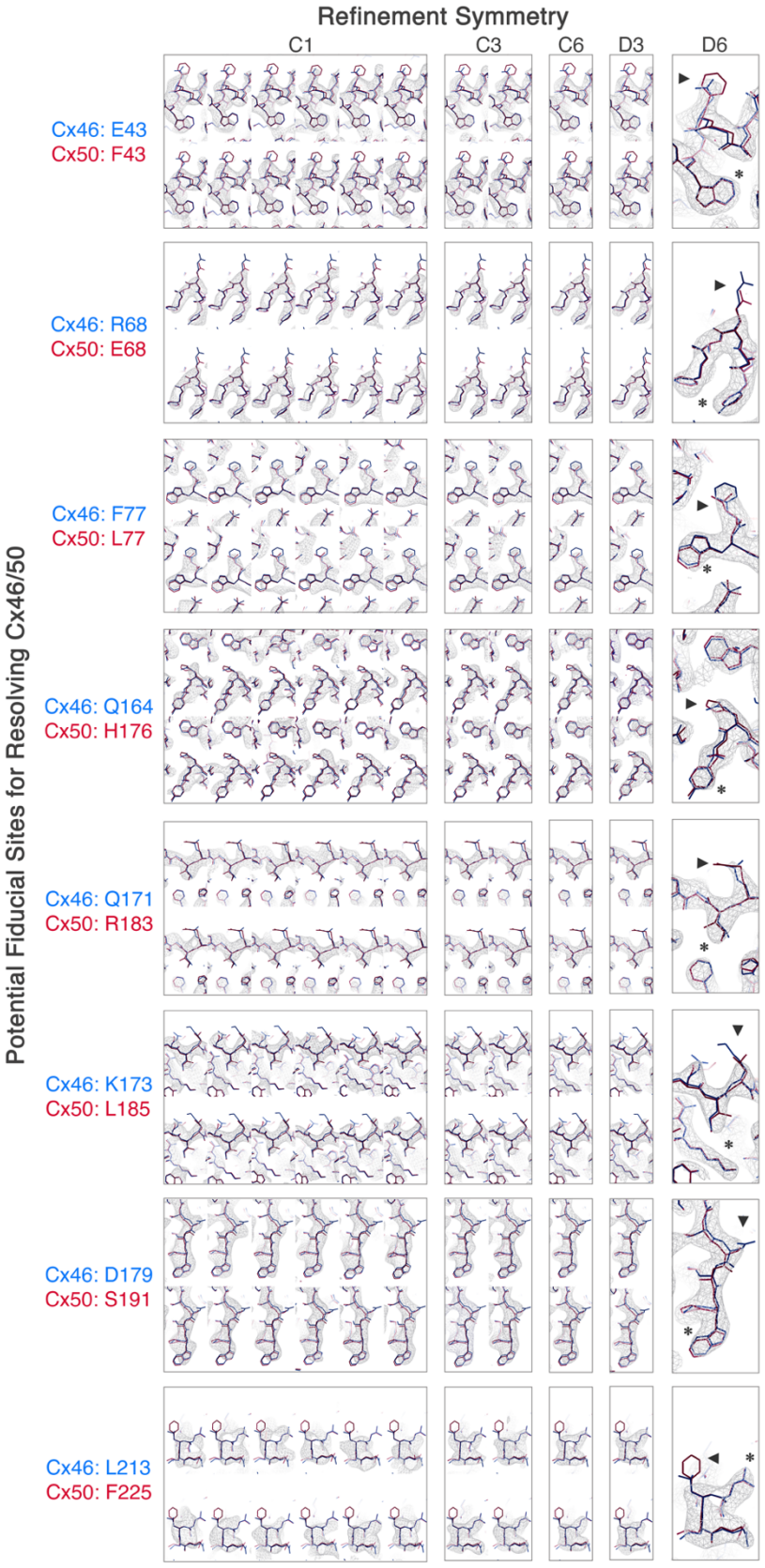
Extended Data Fig. 2 | Overview of cryo-EM image processing and 3D reconstruction. a, A total of 1,104 micrographs were collected in an automated fashion using SerialEM[65] on a 300-kV Titan Krios (dataset 1). Movie stacks were recorded using a K2 summit-direct electron detector operated in super-resolution mode and acquired with an effective pixel size of 0.665 Å. Movie stacks were corrected for drift and CTF using MotionCor2[66] and GCTF[67], respectively. An initial dataset of 261,206 raw particles was obtained using unbiased autopicking procedures in DoG Picker[68]. A refined dataset of 53,791 'good' particles was obtained following several rounds of 2D classification and removal of 'bad' particles (or ice contamination) was done in Relion[69]. 3D classification was seeded using an initial model obtained by negative-stain electron microscopy, filtered to 60 Å. A majority of particles fell into a single 3D class (~62.5% of the "good" particles). These 30,128 particles were used for final 3D auto-refinement and post processing, yielding a final map at 3.4 Å resolution by gold-standard FSC (dataset 1). Dataset 2 was processed in a similar fashion from a total of 2,197 micrographs and 44,547 'good' particles, resulting in a final map at 3.5 Å resolution by gold-standard FSC. **b**, Summary of cryo-EM data collection, refinement and model validation statistics. Dataset 1 was used to obtain the 3.4 Å resolution reconstruction (map 1). Dataset 2 was used to obtain the 3.5 Å resolution reconstruction (map 2). Pre-processed and post-processed maps and associated masks from both datasets have been deposited to the electron microscopy databank (EMD-9116). The original multi-frame micrographs have been deposited to EMPIAR (EMPIAR-10212). Coordinates for Cx50 and Cx46 atomic models have been deposited to the Protein Data Bank (6MHY and 6MHQ, respectively).


a.
Fourier Shell Correlation
(Dataset #1)
1.0
0.5
0.0
FSC
0
0.1
0.2
0.3
0.4
0.5
Resolution (1/Å)
Masked
Unmasked
Autorefine
Cx46 (to map)
Cx50 (to map)
3.4 Å
(3.6 - 3.7 Å)
3.4 Å
b.
Fourier Shell Correlation
(Dataset #2)
3.5 Å
(3.8 Å)
3.5 Å
c.
Local Resolution (Dataset #1)
90°
d.
Local Resolution (Dataset #2)
Local Res.
>6.5 Å
5.5 Å
4.5 Å
3.5 Å
2.5 Å
e.
Cx46 Model vs Map (Dataset #1)
f.
Cx50 Model vs Map (Dataset #1)

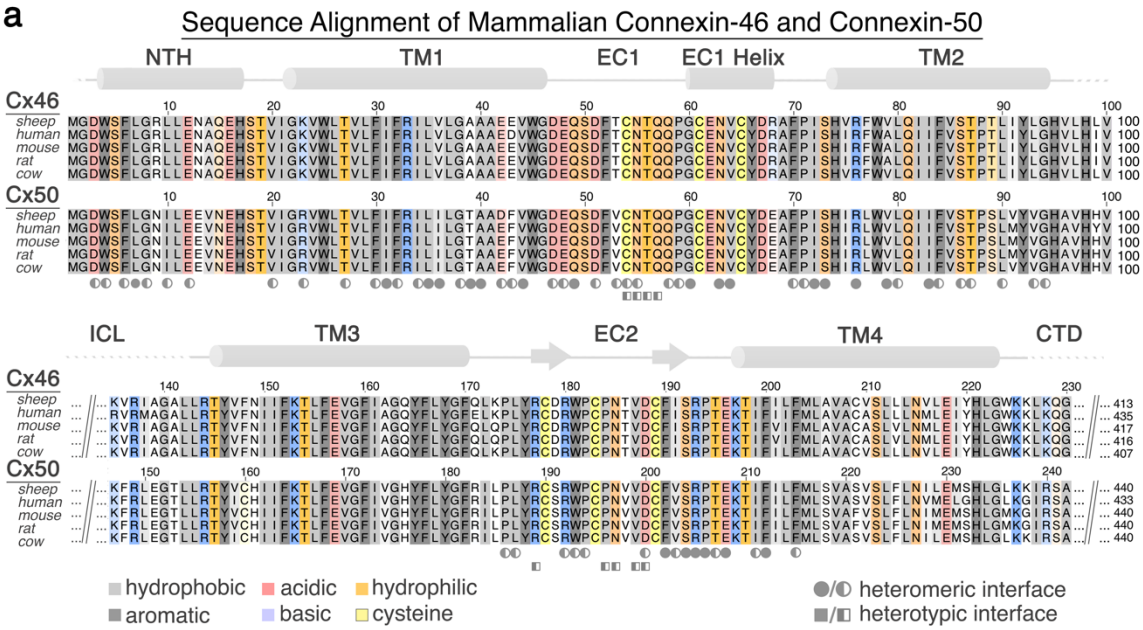
Extended Data Fig. 3 | Global and local resolution analysis. a, **b**, FSC analysis obtained from dataset 1 (**a**) and dataset 2 (**b**). Gold-standard FSC curves following auto-refinement (light grey), post-processing (grey), and masking (dark grey). The final masked maps display an overall resolution of ~3.4 Å (dataset 1) and ~3.5 Å (dataset 2), using a 0.143 cut-off. FSC curves comparing atomic models of Cx46 (orange) and Cx50 (blue) fit to the cryo-EM maps display correlation at 0.5 cut-off to a resolution of 3.4 Å (dataset 1) and 3.5 Å (dataset 2). **c**, **d**, Local resolution analysis using BlocRes[70], obtained for the half-maps for dataset 1 (**c**) and 2 (**d**). **e**, **f**, Local resolution analysis comparing the experimental density map (dataset 1) to the calculated maps of Cx46 (**e**) and Cx50 (**f**). Local resolution ranges in **c**–**f** are indicated by colour (2.5–4.0 Å, blue–cyan; 4.0–5.0 Å, white; 5.0–6.5 Å, yellow–orange). Values obtained for local resolution of Cx46 and Cx50 models compared to the experimental density map are shown in Supplementary Tables 2, 3. Local resolution assessment comparing the density map to the two models indicates that the sites where the two isoforms differ in sequence were generally less well resolved as compared to equivalently exposed residues where Cx46 and Cx50 share conserved sequence.

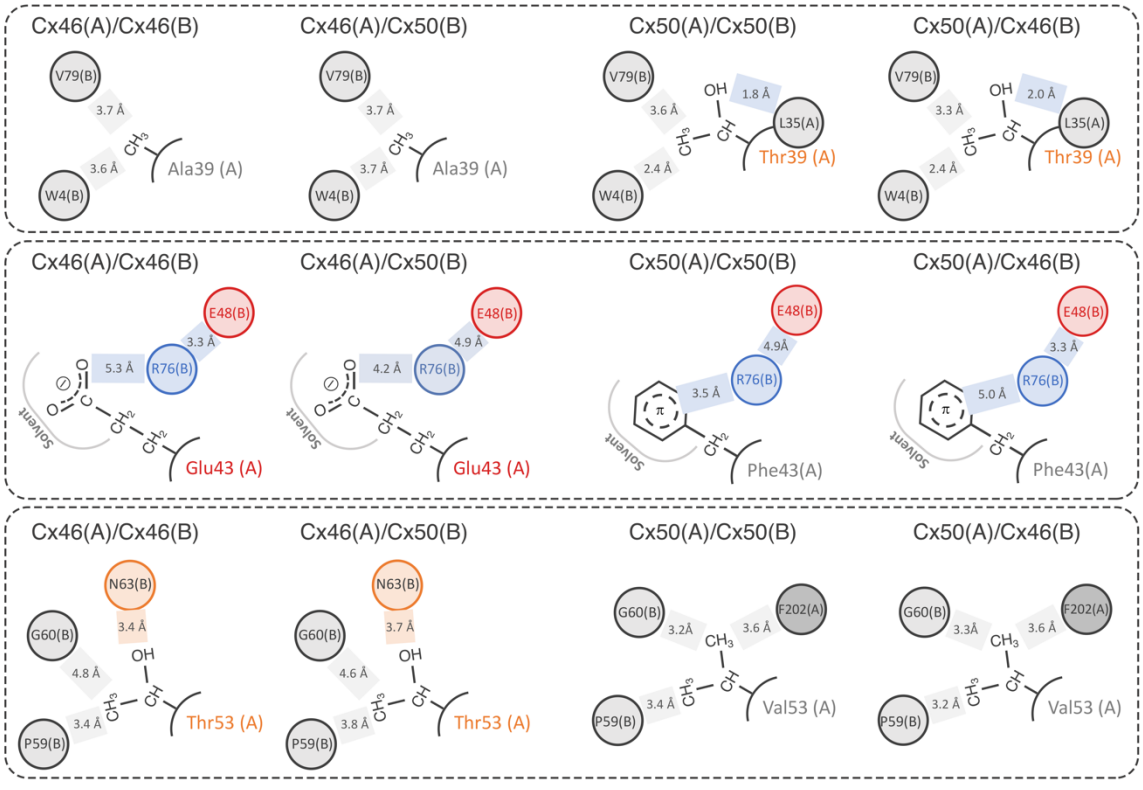

Sheep Connexin-46/50 Structural Homology
Identical (80%) = grey • Similar (8%) = blue • Different (12%) = orange
Connexin-46 Fit to Map
NTH
(2-21)
TM1
(21-51)
EC1
(51-73)
TM2
(73-97)
TM3
(142-172)
EC2
(172-194)
TM4
(195-222)
Cx46
Cx50
W4
W4
R33
R33
T53
N55
V53
N55
R76
R76
Q171
L172
R183
I184
F190
F202
L214
L212
V210
L226
L224
V222
NTH
(2-21)
TM1
(21-51)
EC1
(51-73)
TM2
(73-97)
TM3
(154-184)
EC2
(184-206)
TM4
(207-234)
Connexin-50 Fit to Map

Extended Data Fig. 4 | Cx46 and Cx50 atomic models fit to the cryo-EM density maps. Segmented cryo-EM map with atomic models for sheep Cx46 and Cx50 fit to the experimental densities derived from dataset 1 (3.4 Å, D6 symmetry), including regions for TM1–TM4, EC1 and EC2. The NTH domain is fit into the map from dataset 2 (3.5 Å, D6 symmetry), which was more well defined in this region. Cx46 (top) and Cx50 (bottom) models are coloured according to their pair-wise sequence homology, as being identical (grey, 80%), similar (blue, 8%) and different (orange, 12%). Windows show magnified views corresponding to boxed regions of the segmented maps, highlighting representative side-chain densities and fit to the atomic models. Regions of identical or similar amino acids are fit equally well by both models (for example, Cx46 L172 versus Cx50 I184, blue labels). Over regions where the sequence of Cx46 and Cx50 differ, side-chain density is typically weaker (see also Extended Data Fig. 5). This is possibly due to the imposed averaging of two different side chains in these areas, or relative flexibility as many of these residues correspond to solvent-exposed side chains. In these areas of difference, and where electron microscopy density is present, both Cx46 and Cx50 models were typically fit equally well into the density map (for example, Cx46 T53 versus Cx50 V53 and Cx46 Q171 versus Cx50 R183, orange labels).

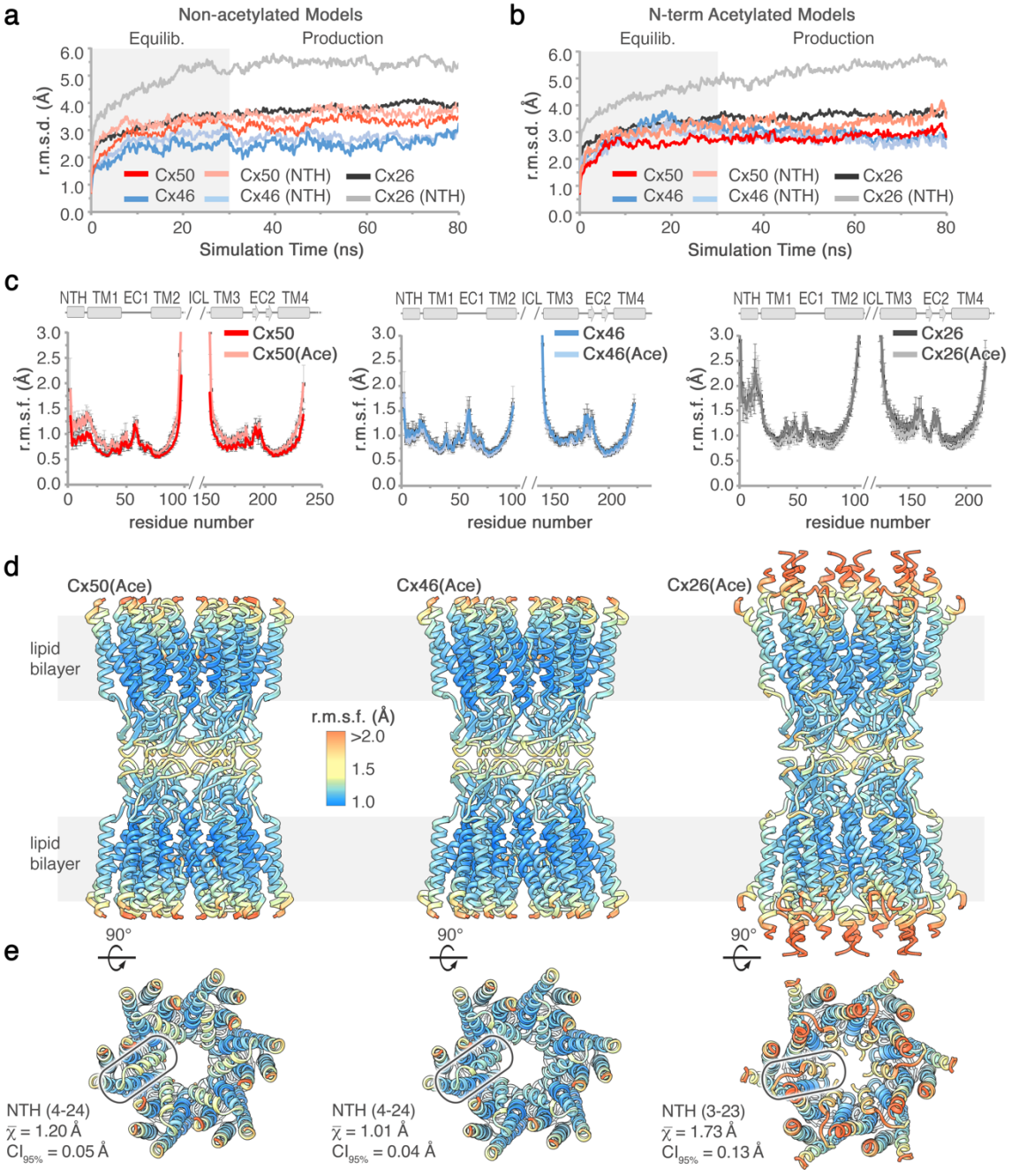

Refinement Symmetry
C1
C3
C6
D3
D6
Potential Fiducial Sites for Resolving Cx46/50
Cx46: E43
Cx50: F43
Cx46: R68
Cx50: E68
Cx46: F77
Cx50: L77
Cx46: Q164
Cx50: H176
Cx46: Q171
Cx50: R183
Cx46: K173
Cx50: L185
Cx46: D179
Cx50: S191
Cx46: L213
Cx50: F225

Extended Data Fig. 5 | Analysis of different symmetry refinements on the ability to resolve differences between Cx46 and Cx50. Eight sites of sequence differences involving bulky amino acids (labelled, and indicated by arrowhead) were selected as potential fiducial markers for resolving the two different isoforms following 3D refinement with various applied symmetries (C1, 4.1 Å resolution; C3, 3.9 Å resolution; C6, 3.7 Å resolution; D3, 3.7 Å resolution; D6, 3.4 Å resolution). For the applied symmetries, views are presented for each unique asymmetric subunit (boxed). Despite the modest resolution of the asymmetric (C1) reconstruction, side-chain density for bulky amino acids is typically observed at sites where the two isoforms are conserved (asterisk). However, at the selected sites of sequence variation (arrow head) the side-chain densities are either not well resolved, or there was no systematic variation that indicated an ability to distinguish the two isoforms. The most resolved features at these sites of variation were obtained with D6 symmetry, and typically corresponded to regions where these different amino acids share similar structure (such as C_{α} positions). Although the cryo-EM density at these sites of variation were typically weak, the resolvable side-chain features throughout the rest of the map were generally enhanced when higher symmetry was applied during map refinement (indicated by asterisk), suggesting that regions of sequence similarity between Cx46 and Cx50 also share a high level of structural similarity.

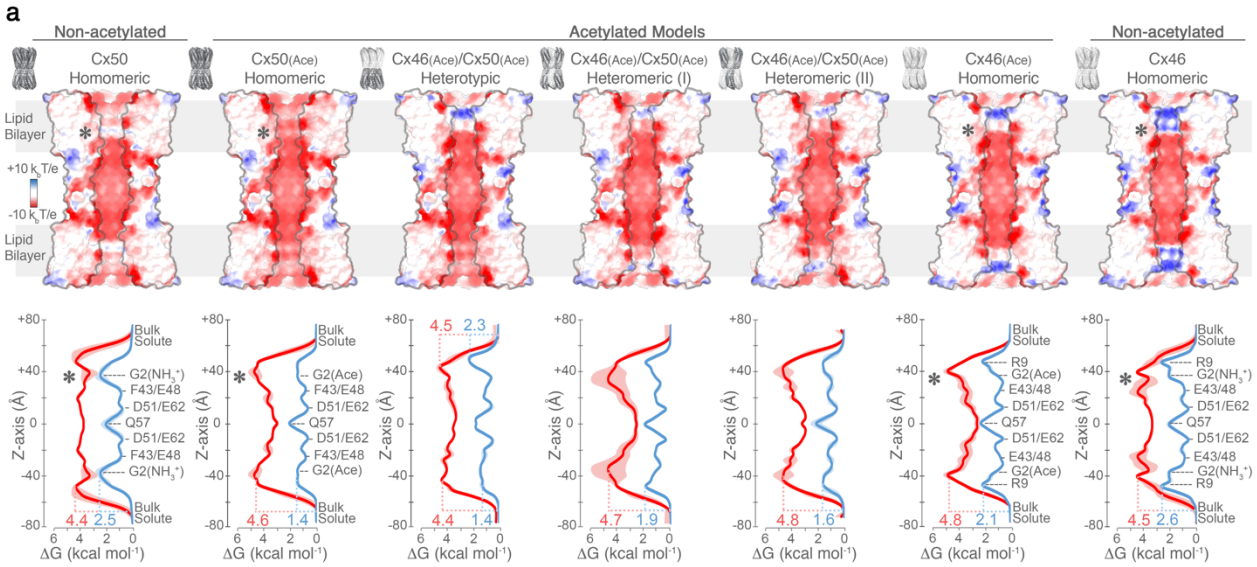

a
Sequence Alignment of Mammalian Connexin-46 and Connexin-50
NTH
TM1
EC1
EC1 Helix
TM2
Cx46
Cx50
sheep
human
mouse
rat
cow
ICL
TM3
EC2
TM4
CTD
hydrophobic
aromatic
acidic
basic
hydrophilic
cysteine
heteromeric interface
heterotypic interface

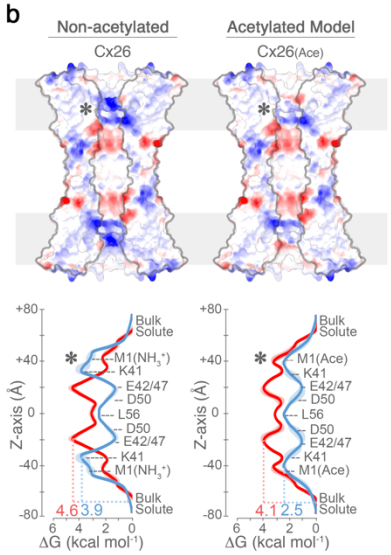

b
Variable Sites of Heteromeric Interface Interactions
Cx46(A)/Cx46(B)
Cx46(A)/Cx50(B)
Cx50(A)/Cx50(B)
Cx50(A)/Cx46(B)
V79(B)
W4(B)
Ala39 (A)
Thr39 (A)
L35(A)
E48(B)
R76(B)
Glu43 (A)
Phe43(A)
N63(B)
G60(B)
P59(B)
Thr53 (A)
F202(A)
Val53 (A)
Solvent

Extended Data Fig. 6 | Sequence and structure conservation of Cx46 and Cx50 heteromeric/heterotypic interfaces. a, Multiple sequence alignment of mammalian Cx46 and Cx50 isoforms with residues contributing to heteromeric and heterotypic interfaces annotated[83]. Circle, heteromeric interface; square, heterotypic interface; filled, ≥70% buried; half-filled, 20–70% buried. Colouring corresponds to amino acid type (grey, hydrophobic; dark grey, aromatic; red, acidic; blue, basic; orange, hydrophilic; yellow, cysteine). Regions of sequence homology are indicated by the level of shading. Secondary structure and domain labels are indicated for the NTH domain, TM1–TM4, EC1 and EC2. Regions lacking defined structure and with poor sequence homology within the intracellular loop (ICL) and C-terminal domain (CTD) have been omitted for clarity. Sheep and human Cx46 and Cx50 orthologues contain ~95% sequence identity (~98% similarity) over the structured regions of the protein. Numbering corresponds to the amino acid sequence of sheep Cx44 and Cx49 used in the main text. **b**, Illustration of homomeric and heteromeric interface interactions involving the three sites lacking conservation between Cx44 and Cx49 at this interface (positions 39, 43 and 53). Despite these sequence differences, the interactions involving these residues are generally similar (hydrophobic, grey; hydrogen bonding, orange; ion-pairing, blue).


a
Non-acetylated Models
Equilib.
Production
r.m.s.d. (Å)
Cx50
Cx50 (NTH)
Cx26
Cx46
Cx46 (NTH)
Cx26 (NTH)
Simulation Time (ns)
b
N-term Acetylated Models
Equilib.
Production
r.m.s.d. (Å)
Cx50
Cx50 (NTH)
Cx26
Cx46
Cx46 (NTH)
Cx26 (NTH)
Simulation Time (ns)
c
NTH TM1 EC1 TM2 ICL TM3 EC2 TM4
r.m.s.f. (Å)
Cx50
Cx50(Ace)
residue number
NTH TM1 EC1 TM2 ICL TM3 EC2 TM4
r.m.s.f. (Å)
Cx46
Cx46(Ace)
residue number
NTH TM1 EC1 TM2 ICL TM3 EC2 TM4
r.m.s.f. (Å)
Cx26
Cx26(Ace)
residue number
d
Cx50(Ace)
Cx46(Ace)
Cx26(Ace)
lipid bilayer
lipid bilayer
r.m.s.f. (Å)
>2.0
1.5
1.0
e
90°
90°
90°
NTH (4-24)
χ̄ = 1.20 Å
CI95% = 0.05 Å
NTH (4-24)
χ̄ = 1.01 Å
CI95% = 0.04 Å
NTH (3-23)
χ̄ = 1.73 Å
CI95% = 0.13 Å

Extended Data Fig. 7 | Analysis of protein backbone dynamics during molecular dynamics equilibration and production. a, C_α r.m.s.d. analysis of equilibrium (0–30 ns) and production phases (30–80 ns) of the molecular dynamics simulations, calculated with respect to the experimental starting structure for non-acetylated models of Cx50 (red traces), Cx46 (blue traces) and Cx26 (with Met1 added; grey traces). Separate analysis for the NTH domains are shown in lighter shades. **b**, Same analysis as in **a**, for models with N-terminal acetylation added. The NTH domain of Cx26 (light grey traces) shows significantly higher r.m.s.d. values, for both non-acetylated and acetylated models. **c**, Plot of average C_α r.m.s.f. during the production phase of the molecular dynamics simulations for Cx50 (left, red traces), Cx46 (centre, blue traces) and Cx26 (right, grey traces). Data obtained for the N-terminal acetylated models are shown in lighter shades. Averages are determined for the 12 subunits composing the intercellular channel. Error bars represent 95% confidence intervals (n = 12 subunits). Secondary structure and domain labels are indicated for the NTH, TM1–4, EC1 and EC2, and ICL (not modelled). **d**, **e**, Average r.m.s.f. values of the acetylated models mapped to the experimental starting structures of Cx50 (left), Cx46 (centre) and Cx26 (right). Colours correspond to r.m.s.f. amplitudes: 0–1.0 Å (cyan); 1.0–2.0 Å (yellow–orange), 2.0 Å (red). In **e**, a single NTH domain is circled and the average r.m.s.f. values and 95% confidence intervals (student t-test) calculated over the NTH domain of each isoform are displayed (n = 12 subunits). The NTH domain of Cx26 shows significantly higher r.m.s.f. values, for both acetylated and non-acetylated models (P 0.0001).



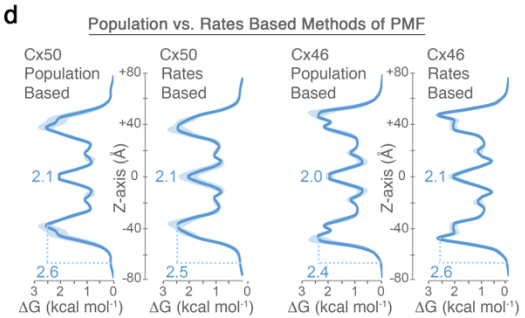


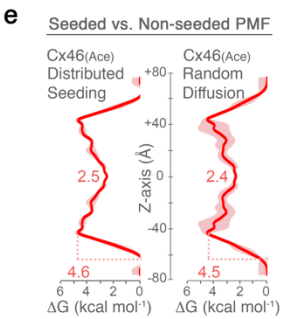
c

Table of Peak PMF Values

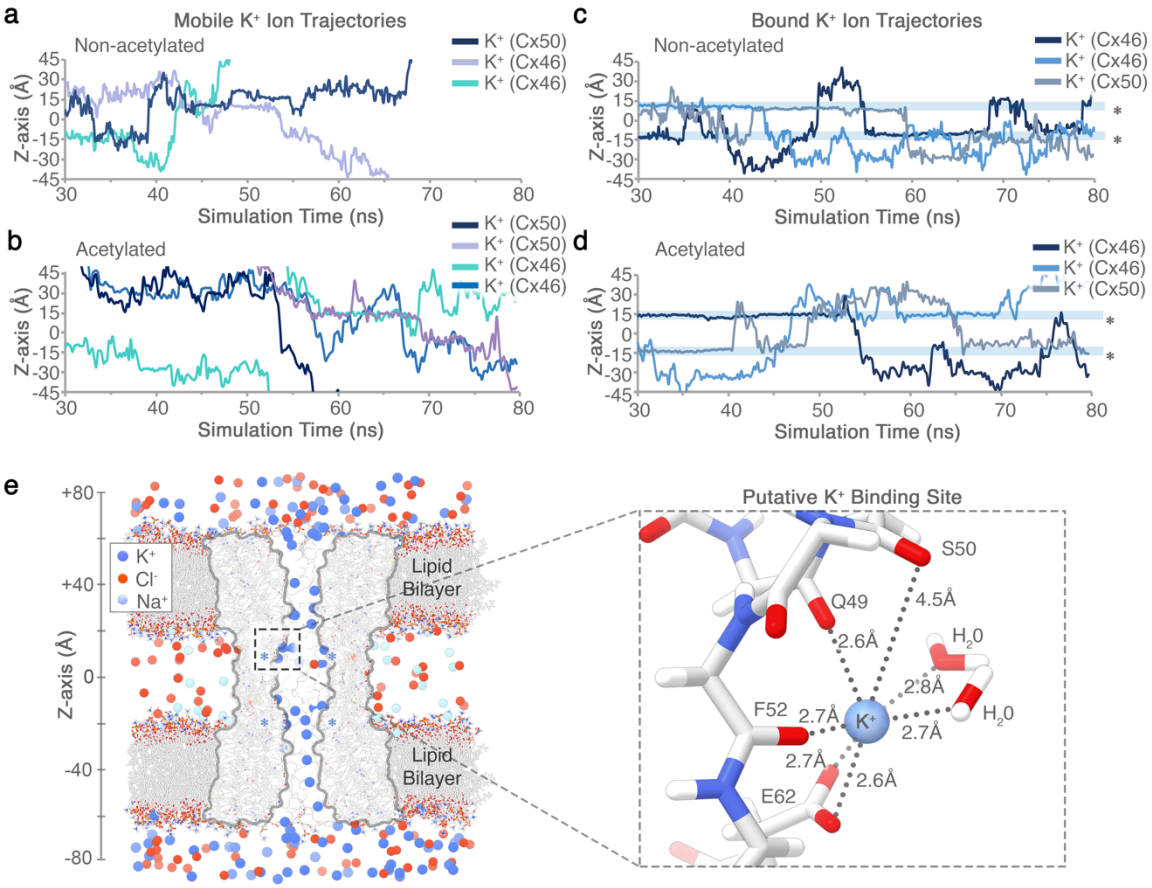
Peak Barrier	Cx50 Homomeric	Cx50* Homomeric	Cx46*/Cx50* Heterotypic	Cx46*/Cx50* Heteromeric (I)	Cx46*/Cx50* Heteromeric (II)	Cx46* Homomeric	Cx46 Homomeric	Cx26 Homomeric	Cx26* Homomeric
ΔGK+ (kcal mol^{-1})	2.5	1.4	2.3/1.4	1.9	1.6	2.1	2.6	3.9	2.5
ΔGCl- (kcal mol^{-1})	4.4	4.6	4.5/4.4	4.7	4.8	4.8	4.5	4.6	4.1
ΔΔG (kcal mol^{-1})	1.9	3.2	2.2/3.0	2.8	3.2	2.7	1.9	0.7	1.6

*n-terminal acetylation

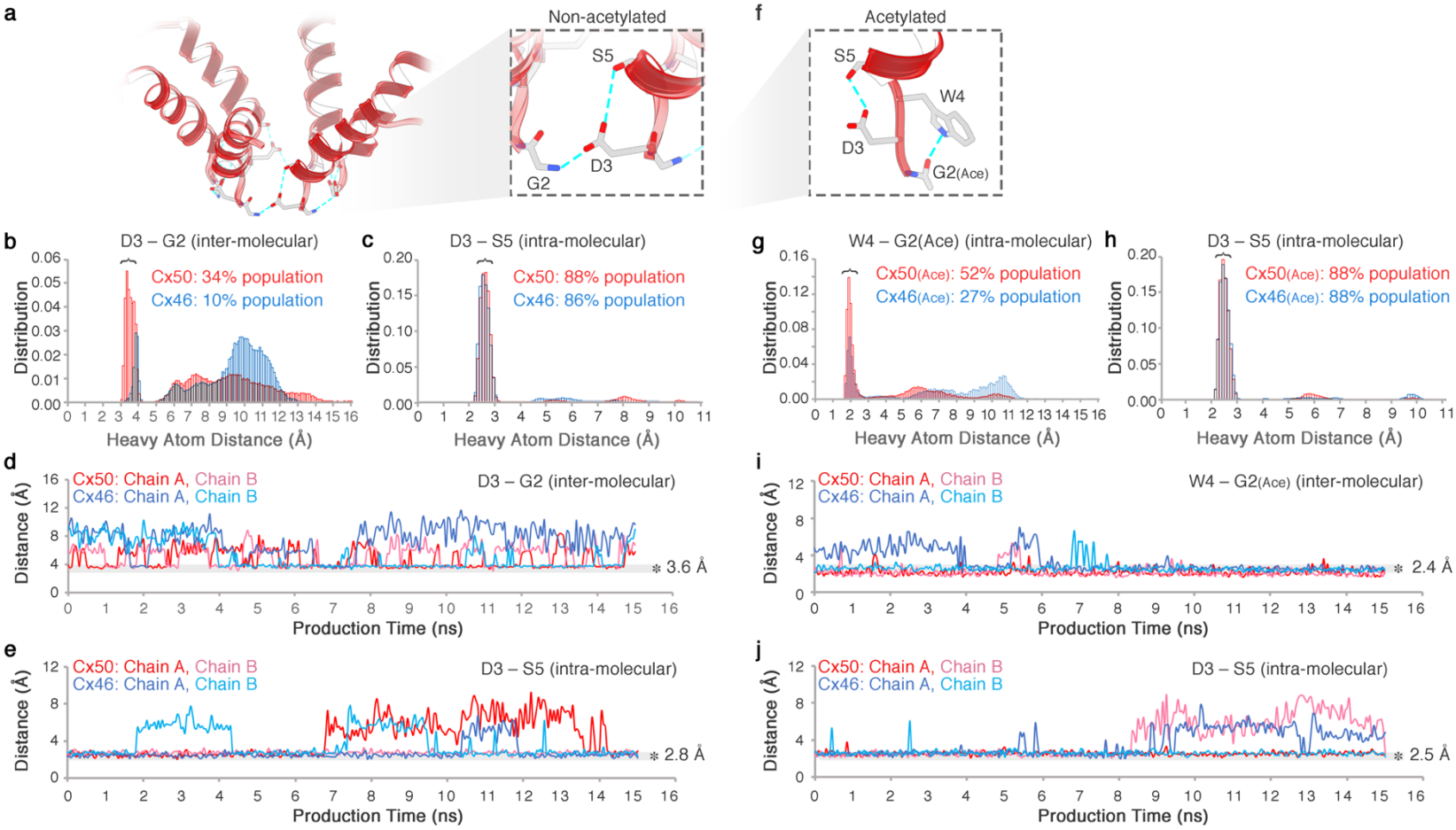




Extended Data Fig. 8 | Modulation of coulombic surface potential and K^+/Cl^- PMFs resulting from N-terminal acetylation and Cx46/50 heterotypic/heteromeric assembly. **a**, Coulombic surface potential maps (top) and PMF (bottom) obtained for a set of non-acetylated and acetylated Cx50–Cx50ac and Cx46–Cx46ac models characterized by molecular dynamics simulation. Acetylated Cx50ac and Cx46ac monomers were used to construct a heterotypic channel and two different heteromeric channels (labelled, I and II). Icons at the top of each structure show the relative configurations of Cx50 (black) and Cx46 (white). The resulting coulombic surface potentials are coloured as in Fig. 3a (negative, red; neutral, white; positive, blue). Only eight subunits are shown to portray both the channel pore and subunit interfaces. An asterisk (shown in **a** and **b**) indicates the site of N-terminal acetylation, which neutralizes the positively charged N terminus. PMFs obtained for K^+ (blue traces) and Cl^- ions (red traces) are displayed directly beneath each model displayed in **a**. Free energy maxima are labelled and pore axis (*z* axis) is indicated. Traces indicate symmetrized values, with unsymmetrized values in lighter shading. In the case of the asymmetric Cx46/50 heterotypic model (middle left), PMFs represent the average from the first and last ~60 ns of simulation. **b**, Coulombic surface potential (top) and PMFs (bottom) obtained for non-acetylated Cx26 (left) and Cx26ac (right) are displayed as in **a**. **c**, Table of peak free energy barriers for K^+ (ΔG_{K+}) and Cl^- (ΔG_{Cl-}) and corresponding $\Delta\Delta G$, reported as a proxy for charge selectivity. Asterisk indicates models with acetylated N terminus. **d**, **e**, Validation of methods used to construct the PMFs. **d**, Comparison of K^+ PMFs obtained for Cx50ac and Cx46ac using population states (left) or transition rates (right) (see Methods). Both methods yielded similar PMF profiles. All other PMFs were constructed using the transition rates method. **e**, Comparison of Cl^- PMFs obtained for Cx46ac using transition rates of Cl^- ions that diffused into the pore (left) and those that were randomly seeded within the pore (right). All other Cl^- PMFs were constructed using the distributed seeding approach to enhance sampling (see Methods).


a
Mobile K+ Ion Trajectories
Non-acetylated
K+ (Cx50)
K+ (Cx46)
K+ (Cx46)
Z-axis (Å)
Simulation Time (ns)
b
Acetylated
K+ (Cx50)
K+ (Cx50)
K+ (Cx46)
K+ (Cx46)
c
Bound K+ Ion Trajectories
Non-acetylated
K+ (Cx46)
K+ (Cx46)
K+ (Cx50)
d
Acetylated
e
K+
Cl-
Na+
Lipid Bilayer
Lipid Bilayer
Putative K+ Binding Site
S50
Q49
4.5Å
2.6Å
H20
2.8Å
F52
2.7Å
K+
2.7Å
H20
2.7Å
2.6Å
E62

Extended Data Fig. 9 | Analysis of K^+ trajectories and putative binding site observed during molecular dynamics simulation. a–d, K^+ ion trajectories obtained for Cx50 and Cx46 along the channel pore (*z* axis). **a**, **b**, Representative traces of mobile ions transiting and exiting or entering the channel pore in both acetylated (**a**) and non-acetylated (**b**) models of Cx46 and Cx50. **c**, **d**, K^+ ions displaying long dwell times (~10–20 ns) localized at one or more putative binding sites within the channel pore (asterisk at *z* = ~14 Å) in both models of Cx46 and Cx50. In **a**–**d**, Similar results were observed from 6 independent runs using non-acetylated models (1 × 80 ns and 2 × 10 ns runs for both Cx50 and Cx46) and 13 independent runs using the acetylated models (1 × 80 ns and 6 × 10 ns runs for Cx50; and 1 × 80 ns and 7 × 10 ns runs for Cx46). **e**, Representative snapshot showing an enlarged view of the putative K^+ binding site identified for Cx50 and Cx46, corresponding to the region indicated by the asterisk in **c** and **d**. A single K^+ ion is bound by a conserved set of amino acids (among Cx46/50 orthologues), coordinated by the side-chain carboxylate of Glu62 and backbone carbonyls of Gln49, Ser50 and Phe52 (identical in Cx46 and Cx50). Two transient water molecules observed coordinating the bond K^+ ion are shown. Twelve binding sites are present within the dodecameric channel. Similar behaviour was observed from simulations using both non-acetylated and acetylated models (19 independent simulations). A functional role for this putative binding site is not yet clear, but may represent a physiologically relevant cation-binding site similar to the recently proposed Ca^{2+}-binding site in Cx26[11].


a
Non-acetylated
S5
D3
G2
f
Acetylated
S5
W4
D3
G2(Ace)
b
D3 – G2 (inter-molecular)
Cx50: 34% population
Cx46: 10% population
Distribution
Heavy Atom Distance (Å)
c
D3 – S5 (intra-molecular)
Cx50: 88% population
Cx46: 86% population
Distribution
Heavy Atom Distance (Å)
g
W4 – G2(Ace) (intra-molecular)
Cx50(Ace): 52% population
Cx46(Ace): 27% population
Distribution
Heavy Atom Distance (Å)
h
D3 – S5 (intra-molecular)
Cx50(Ace): 88% population
Cx46(Ace): 88% population
Distribution
Heavy Atom Distance (Å)
d
Cx50: Chain A, Chain B
Cx46: Chain A, Chain B
D3 – G2 (inter-molecular)
Distance (Å)
* 3.6 Å
Production Time (ns)
i
Cx50: Chain A, Chain B
Cx46: Chain A, Chain B
W4 – G2(Ace) (inter-molecular)
Distance (Å)
* 2.4 Å
Production Time (ns)
e
Cx50: Chain A, Chain B
Cx46: Chain A, Chain B
D3 – S5 (intra-molecular)
Distance (Å)
* 2.8 Å
Production Time (ns)
j
Cx50: Chain A, Chain B
Cx46: Chain A, Chain B
D3 – S5 (intra-molecular)
Distance (Å)
* 2.5 Å
Production Time (ns)

Extended Data Fig. 10 | Dynamic hydrogen-bond network within the NTH domain observed by molecular dynamics simulation. **a**–**e**, Analysis of hydrogen-bond interactions for non-acetylated models of Cx46 and Cx50 observed during molecular dynamics simulation. **a**, Inset, magnified view of D3 pairing with the positively charged N-terminal G2 position from a neighbouring subunit (intermolecular) and with the hydroxyl of S5 within the same subunit (intramolecular). The D3–G2 interactions are dynamically formed and broken during molecular dynamics simulation, whereas the intramolecular D3–S5 hydrogen bond is relatively stable (as shown in **b**–**e**). **b**, **c**, Population statistics of inter-atomic distances involving D3 ($C_{\square}$) and G (N) of the neighbouring chain (**b**) and Ser5 ($H_{\square}$) of the same chain (**c**), extracted from molecular dynamics simulation production runs of Cx46 (blue histogram) and Cx50 (red histogram). For D3 and G2, heavy atoms were chosen as proxies to monitor hydrogen-bonding interactions involving equivalent rotameric donor-acceptor configurations. The population centred at ~3.6 Å (**b**) and ~2.8 Å (**c**) are considered to be within hydrogen-bond distance. **d**, **e**, Trajectories extracted from molecular dynamics simulation of Cx46 (blue traces) and Cx50 (red traces) showing the dynamical behaviour of the D3–G2 intermolecular charge paring (**d**) and D3–S5 intramolecular hydrogen bonding (**e**). The dwell times showing hydrogen bond pairing (~3.6 Å in **d**; and ~2.8 Å in **e**) are indicated with transparent grey shading. In the Cx26 crystal structure, the equivalent D2 site is modelled in hydrogen-bond distance to a neighbouring T5 site (Cx26 numbering)[10], but this intermolecular interaction is rapidly broken during molecular dynamics simulations and does not appear to reform within the timescale of our molecular dynamics experiments, and instead forms a stable intramolecular interaction with T5, as described previously[35] (data not shown). **f**–**j**, Analysis of hydrogen-bond interactions observed during molecular dynamics simulation for Cx46ac and Cx50ac modelled with the N-terminal G2 position acetylated. **f**, Inset, magnified view of acetylated G2ac position hydrogen bonded to the indole ring of W4 from the same subunit (intramolecular) and the same intramolecular D3–S5 hydrogen-bond interaction observed in the non-acetylated channel. **g**, **h**, Population statistics of inter-atomic distances involving W4 ($N_{\square}$) and G2ac (acetyl carbonyl) (**g**) and D3 ($C_{\square}$) distance to Ser5 ($H_{\square}$) of the same chain (**h**), extracted from molecular dynamics simulation production runs of Cx46ac (blue histogram) and Cx50ac (red histogram). **i**, **j**, Trajectories extracted from molecular dynamics simulation of Cx46ac (blue traces) and Cx50ac (red traces) showing the dynamical behaviour of the W4–G2ac hydrogen-bond pairing (**i**) and D3–S5 intramolecular hydrogen bonding (**j**). The dwell times showing hydrogen-bond pairing (~2.4 Å in **i**, and ~2.5 Å in **j**) are indicated with transparent grey shading. For clarity, only the first 15 ns of the production period is shown (**d**–**e**, **i**–**j**). Similar results were observed from 6 independent runs using non-acetylated models (1 × 80 ns and 2 × 10 ns runs for both Cx50 and Cx46) and 13 independent runs using the acetylated models (1 × 80 ns and 6 × 10 ns runs for Cx50; and 1 × 80 ns and 7 × 10 ns runs for Cx46).

Summary of molecular dynamics simulation setup and conditions for Cx50, Cx46, Cx26 and Cx46/50 heteromeric/heterotypic models.

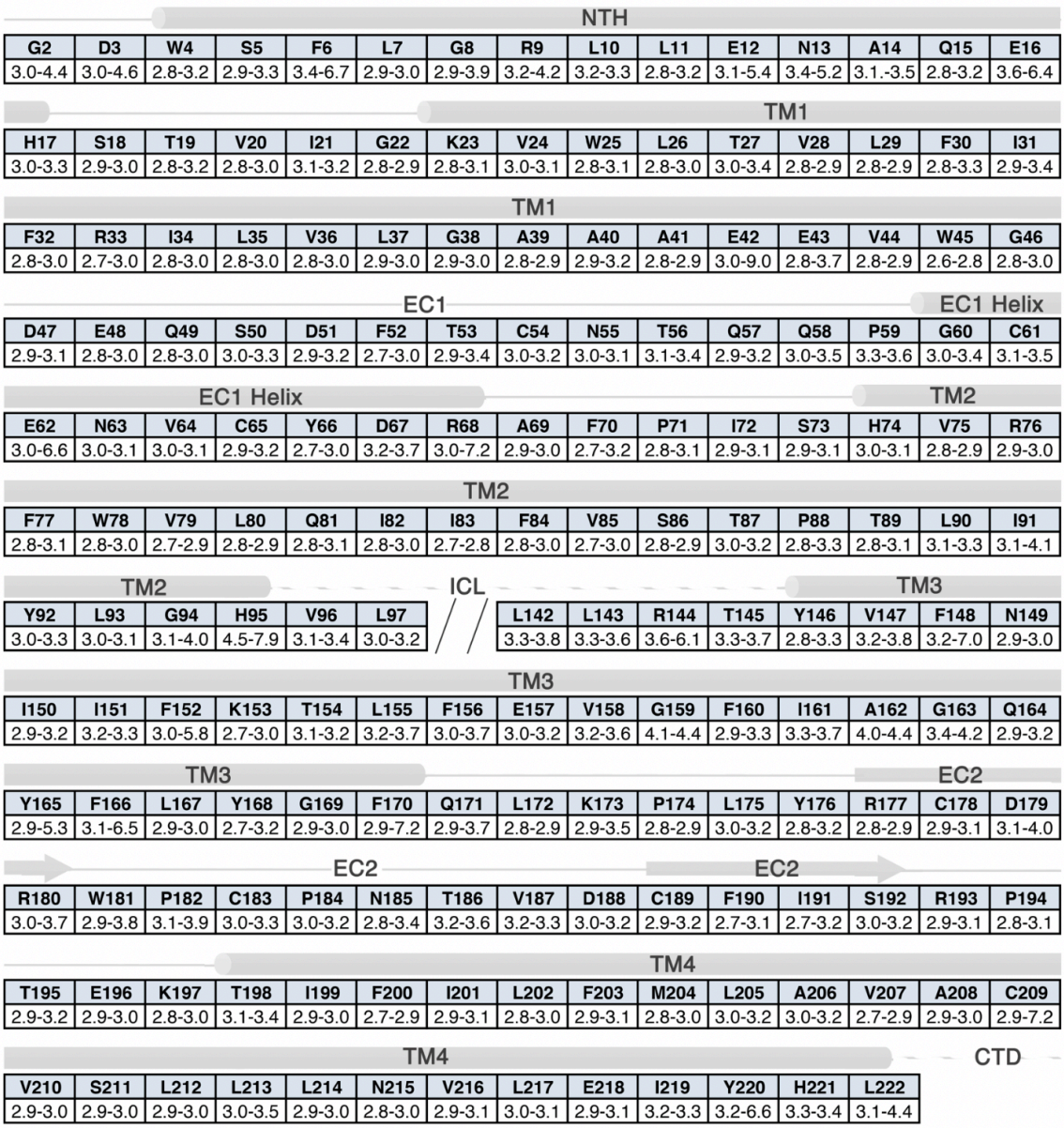
(* n-terminal acetylation)

	Connexin-50	Connexin-50*	Connexin-46	Connexin-46*	Connexin-26	Connexin-26*
Total Atoms	364,952	363,722	360,190	355,524	368,756	367,346
Solvent	196,248	194,928	191,346	191,310	194,856	193,548
Lipids	133,866	133,866	133,732	129,042	133,464	133,330
Protein	34,440	34,488	34,716	34,764	39,972	40,020
Ions	398	440	396	408	464	448
Modelled Residues	2-97; 154-234	2-97; 154-234	2-97; 142-222	2-97; 142-222	1-109; 125-217	1-109; 125-217
Simulation Conditions						
Simulation Box (Å)	152 x 152 x 176	152 x 152 x 176	152 x 152 x 176	152 x 152 x 176	155 x 155 x 176	155 x 155 x 176
Pressure (atm)	1	1	1	1	1	1
Temperature (K)	310	310	310	310	310	310
Time Step (fs)	1	2	1	2	2	2
Equilibration Time (ns)	30	30	30	30	30	30
Production Time (ns)	100	70	70	50	50	70
seed # (10ns/seed)	2	6	2	5	3	4
Total Time (ns)	150	160	120	130	110	140

Connexin-46*/50* Models		Heterotypic	Heteromeric (I)	Heteromeric (II)
Total Atoms		363,884	361,412	361,412
Solvent		194,964	192,498	192,498
Lipids		133,866	133,866	133,866
Protein		34,626	34,626	34,626
Ions		428	422	422
Modelled Residues	Cx50:	2-97; 154-234	2-97; 154-234	2-97; 154-234
	Cx46:	2-97; 142-222	2-97; 142-222	2-97; 142-222
Point Group Symmetry		C6	C3	D3
Simulation Conditions				
Simulation Box (Å)		152 x 152 x 176	152 x 152 x 176	152 x 152 x 176
Pressure (atm)		1	1	1
Temperature (K)		310	310	310
Time Step (fs)		2	2	2
Equilibration Time (ns)		30	30	30
Production Time (ns)		80	70	70
seed #		16	4	4
Total Time (ns)		270	140	140

Supplementary Table 1. Summary of molecular dynamics simulation setup and conditions for Cx50, Cx46, Cx26 and Cx46/50 heteromeric/heterotypic models. Asterisk indicates models that included modification by n-terminal acetylation. Connexin-26 models were derived from PDB 2ZW310.

Local resolution assessment of a calculated map based on the Cx46 atomic model versus the 3.4 Å resolution CryoEM density map.



G2	D3	W4	S5	F6	L7	G8	R9	L10	L11	E12	N13	A14	Q15	E16
3.0-4.4	3.0-4.6	2.8-3.2	2.9-3.3	3.4-6.7	2.9-3.0	2.9-3.9	3.2-4.2	3.2-3.3	2.8-3.2	3.1-5.4	3.4-5.2	3.1.-3.5	2.8-3.2	3.6-6.4

H17	S18	T19	V20	I21	G22	K23	V24	W25	L26	T27	V28	L29	F30	I31
3.0-3.3	2.9-3.0	2.8-3.2	2.8-3.0	3.1-3.2	2.8-2.9	2.8-3.1	3.0-3.1	2.8-3.1	2.8-3.0	3.0-3.4	2.8-2.9	2.8-2.9	2.8-3.3	2.9-3.4

F32	R33	I34	L35	V36	L37	G38	A39	A40	A41	E42	E43	V44	W45	G46
2.8-3.0	2.7-3.0	2.8-3.0	2.8-3.0	2.8-3.0	2.9-3.0	2.9-3.0	2.8-2.9	2.9-3.2	2.8-2.9	3.0-9.0	2.8-3.7	2.8-2.9	2.6-2.8	2.8-3.0

D47	E48	Q49	S50	D51	F52	T53	C54	N55	T56	Q57	Q58	P59	G60	C61
2.9-3.1	2.8-3.0	2.8-3.0	3.0-3.3	2.9-3.2	2.7-3.0	2.9-3.4	3.0-3.2	3.0-3.1	3.1-3.4	2.9-3.2	3.0-3.5	3.3-3.6	3.0-3.4	3.1-3.5

E62	N63	V64	C65	Y66	D67	R68	A69	F70	P71	I72	S73	H74	V75	R76
3.0-6.6	3.0-3.1	3.0-3.1	2.9-3.2	2.7-3.0	3.2-3.7	3.0-7.2	2.9-3.0	2.7-3.2	2.8-3.1	2.9-3.1	2.9-3.1	3.0-3.1	2.8-2.9	2.9-3.0

F77	W78	V79	L80	Q81	I82	I83	F84	V85	S86	T87	P88	T89	L90	I91
2.8-3.1	2.8-3.0	2.7-2.9	2.8-2.9	2.8-3.1	2.8-3.0	2.7-2.8	2.8-3.0	2.7-3.0	2.8-2.9	3.0-3.2	2.8-3.3	2.8-3.1	3.1-3.3	3.1-4.1

Y92	L93	G94	H95	V96	L97	L142	L143	R144	T145	Y146	V147	F148	N149
3.0-3.3	3.0-3.1	3.1-4.0	4.5-7.9	3.1-3.4	3.0-3.2	3.3-3.8	3.3-3.6	3.6-6.1	3.3-3.7	2.8-3.3	3.2-3.8	3.2-7.0	2.9-3.0

I150	I151	F152	K153	T154	L155	F156	E157	V158	G159	F160	I161	A162	G163	Q164
2.9-3.2	3.2-3.3	3.0-5.8	2.7-3.0	3.1-3.2	3.2-3.7	3.0-3.7	3.0-3.2	3.2-3.6	4.1-4.4	2.9-3.3	3.3-3.7	4.0-4.4	3.4-4.2	2.9-3.2

Y165	F166	L167	Y168	G169	F170	Q171	L172	K173	P174	L175	Y176	R177	C178	D179
2.9-5.3	3.1-6.5	2.9-3.0	2.7-3.2	2.9-3.0	2.9-7.2	2.9-3.7	2.8-2.9	2.9-3.5	2.8-2.9	3.0-3.2	2.8-3.2	2.8-2.9	2.9-3.1	3.1-4.0

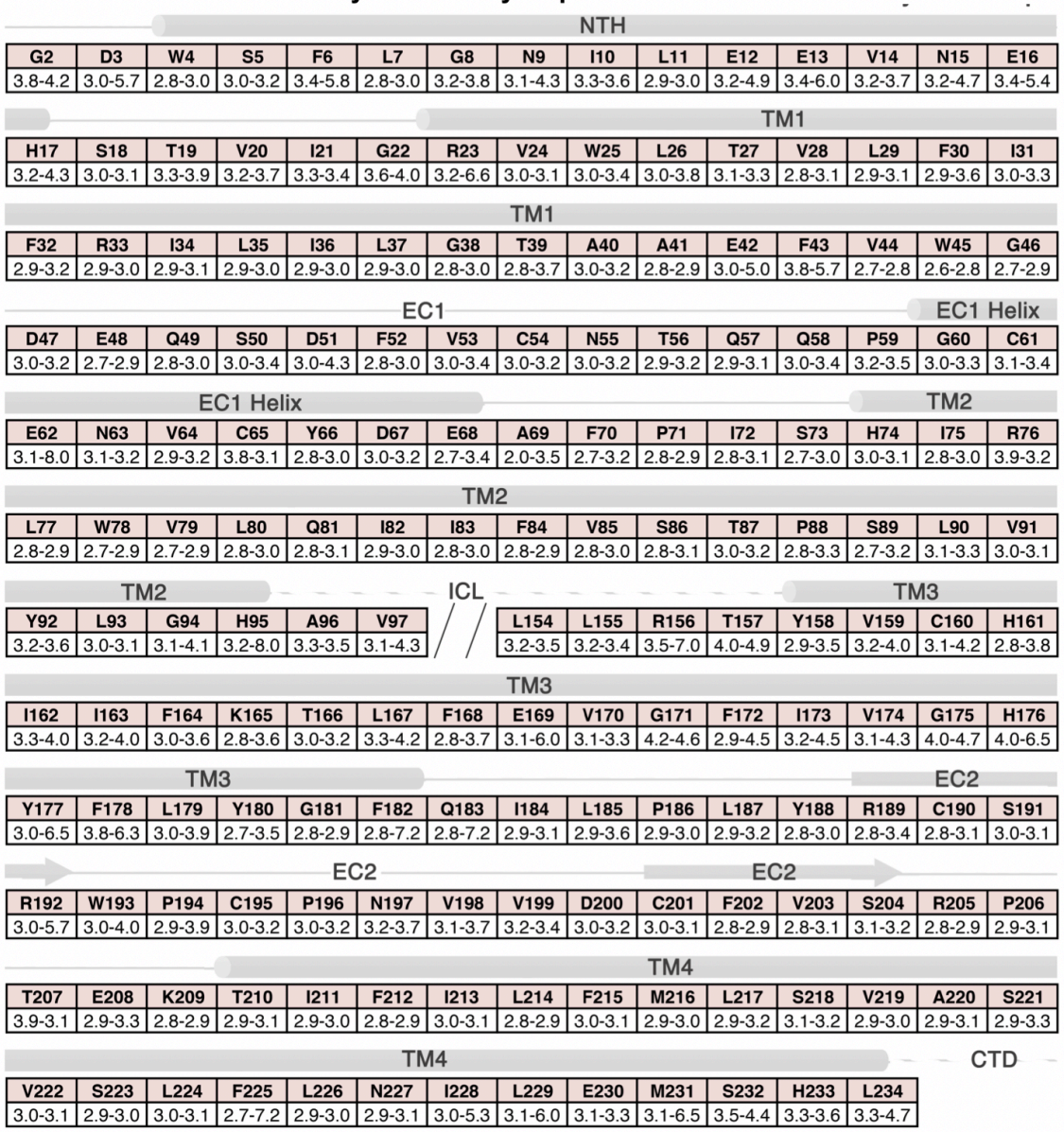
R180	W181	P182	C183	P184	N185	T186	V187	D188	C189	F190	I191	S192	R193	P194
3.0-3.7	2.9-3.8	3.1-3.9	3.0-3.3	3.0-3.2	2.8-3.4	3.2-3.6	3.2-3.3	3.0-3.2	2.9-3.2	2.7-3.1	2.7-3.2	3.0-3.2	2.9-3.1	2.8-3.1

T195	E196	K197	T198	I199	F200	I201	L202	F203	M204	L205	A206	V207	A208	C209
2.9-3.2	2.9-3.0	2.8-3.0	3.1-3.4	2.9-3.0	2.7-2.9	2.9-3.1	2.8-3.0	2.9-3.1	2.8-3.0	3.0-3.2	3.0-3.2	2.7-2.9	2.9-3.0	2.9-7.2

V210	S211	L212	L213	L214	N215	V216	L217	E218	I219	Y220	H221	L222
2.9-3.0	2.9-3.0	2.9-3.0	3.0-3.5	2.9-3.0	2.8-3.0	2.9-3.1	3.0-3.1	2.9-3.1	3.2-3.3	3.2-6.6	3.3-3.4	3.1-4.4

Supplementary Table 2. Local resolution assessment of a calculated map based on the Cx46 atomic model versus the 3.4 Å CryoEM density map. Values indicated by each residue correspond to the range of resolutions (Å) reported for amino acid sidechain densities using BlocRes70. Secondary structure and domain labels are indicated for the n-terminal helix (NTH), transmembrane helices (TM1-4) and extracellular domains (EC1-2) intracellular loop (ICL) and C-terminal domain (CTD).

Local resolution assessment of a calculated map based on the Cx50 atomic model versus the 3.4 Å resolution CryoEM density map.



Supplementary Table 3. Local resolution assessment of a calculated map based on the Cx50 atomic model versus the 3.4 Å CryoEM density map. Values indicated by each residue correspond to the range of resolutions (Å) reported for amino acid sidechain densities using BlocRes70. Secondary structure and domain labels are indicated for the n-terminal helix (NTH), transmembrane helices (TM1-4) and extracellular domains (EC1-2) intracellular loop (ICL) and C-terminal domain (CTD).

References

1. Goodenough, D. A. & Paul, D. L. Gap junctions. *Cold Spring Harb. Perspect. Biol.* **1**, a002576 (2009).
2. Delmar, M. et al. Connexins and Disease. *Cold Spring Harb. Perspect. Biol.* (2017).
3. Garcia, I. E. et al. Connexinopathies: a structural and functional glimpse. *BMC Cell Biol.* **17** (Suppl 1), 17 (2016).
4. Aasen, T., Mesnil, M., Naus, C. C., Lampe, P. D. & Laird, D. W. Gap junctions and cancer: communicating for 50 years. *Nat. Rev. Cancer* **16**, 775–788 (2016). </jrn>
5. Sosinsky, G. E. & Nicholson, B. J. Structural organization of gap junction channels. *Biochim. Biophys. Acta* **1711**, 99–125 (2005).
6. Sohl, G. & Willecke, K. Gap junctions and the connexin protein family. *Cardiovasc. Res.* **62**, 228–232 (2004).
7. Cottrell, G. T. & Burt, J. M. Functional consequences of heterogeneous gap junction channel formation and its influence in health and disease. *Biochim. Biophys. Acta* **1711**, 126–141 (2005).
8. Beyer, E. C. & Berthoud, V. M. Gap junction structure: unraveled, but not fully revealed. *F1000 Res.* **6**, 568 (2017).
9. Grosely, R. & Sorgen, P. L. A history of gap junction structure: hexagonal arrays to atomic resolution. *Cell Commun. Adhes.* **20**, 11–20 (2013).
10. Maeda, S. et al. Structure of the connexin 26 gap junction channel at 3.5 Å resolution. *Nature* **458**, 597–602 (2009).

11. Bennett, B. C. et al. An electrostatic mechanism for Ca^{2+}-mediated regulation of gap junction channels. *Nat. Commun.* **7**, 8770 (2016). </jrn>

12. Mathias, R. T., White, T. W. & Gong, X. Lens gap junctions in growth, differentiation, and homeostasis. *Physiol. Rev.* **90**, 179–206 (2010). </jrn>

13. Konig, N. & Zampighi, G. A. Purification of bovine lens cell-to-cell channels composed of connexin44 and connexin50. *J. Cell Sci.* **108**, 3091–3098 (1995).

14. Jiang, J. X. & Goodenough, D. A. Heteromeric connexons in lens gap junction channels. *Proc. Natl Acad. Sci. USA* **93**, 1287–1291 (1996).</jrn>

15. Shearer, D., Ens, W., Standing, K. & Valdimarsson, G. Posttranslational modifications in lens fiber connexins identified by off-line-HPLC MALDI-quadrupole time-of-flight mass spectrometry. *Invest. Ophthalmol. Vis. Sci.* **49**, 1553–1562 (2008).

16. Wang, Z. & Schey, K. L. Phosphorylation and truncation sites of bovine lens connexin 46 and connexin 50. *Exp. Eye Res.* **89**, 898–904 (2009). </jrn>

17. Kwon, T. et al. Molecular dynamics simulations of the Cx26 hemichannel: insights into voltage-dependent loop-gating. *Biophys. J.* **102**, 1341–1351 (2012). </jrn>

18. Bai, D. Structural analysis of key gap junction domains–lessons from genome data and disease-linked mutants. *Semin. Cell Dev. Biol.* **50**, 74–82 (2016).

19. Koval, M., Molina, S. A. & Burt, J. M. Mix and match: investigating heteromeric and heterotypic gap junction channels in model systems and native tissues. *FEBS Lett.* **588**, 1193–1204 (2014).

20. Unger, V. M., Kumar, N. M., Gilula, N. B. & Yeager, M. Three-dimensional structure of a recombinant gap junction membrane channel. *Science* **283**, 1176–1180 (1999).

21. Fleishman, S. J., Unger, V. M., Yeager, M. & Ben-Tal, N. A C^{α} model for the transmembrane alpha helices of gap junction intercellular channels. *Mol. Cell* **15**, 879–888 (2004).

22. Gong, X. Q. & Nicholson, B. J. Size selectivity between gap junction channels composed of different connexins. *Cell Commun. Adhes.* **8**, 187–192 (2001).

23. Goldberg, G. S., Valiunas, V. & Brink, P. R. Selective permeability of gap junction channels. *Biochim. Biophys. Acta* **1662**, 96–101 (2004).

24. Varland, S., Osberg, C. & Arnesen, T. N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics* **15**, 2385–2401 (2015).

25. Berneche, S. & Roux, B. Energetics of ion conduction through the K^{+} channel. *Nature* **414**, 73–77 (2001).

26. Corry, B. & Thomas, M. Mechanism of ion permeation and selectivity in a voltage gated sodium channel. *J. Am. Chem. Soc.* **134**, 1840–1846 (2012).

27. Ulmschneider, M. B. et al. Molecular dynamics of ion transport through the open conformation of a bacterial voltage-gated sodium channel. *Proc. Natl Acad. Sci. USA* **110**, 6364–6369 (2013).

28. Srinivas, M. et al. Voltage dependence of macroscopic and unitary currents of gap junction channels formed by mouse connexin50 expressed in rat neuroblastoma cells. *J. Physiol.* **517**, 673–689 (1999).

29. Hopperstad, M. G., Srinivas, M. & Spray, D. C. Properties of gap junction channels formed by Cx46 alone and in combination with Cx50. *Biophys. J.* **79**, 1954–1966 (2000).

30. Oh, S., Rubin, J. B., Bennett, M. V., Verselis, V. K. & Bargiello, T. A. Molecular determinants of electrical rectification of single channel conductance in gap junctions formed by connexins 26 and 32. *J. Gen. Physiol.* **114**, 339–364 (1999).

31. Tong, X. et al. The first extracellular domain plays an important role in unitary channel conductance of Cx50 gap junction channels. *PLoS One* **10**, e0143876 (2015).

32. Trexler, E. B., Bukauskas, F. F., Kronengold, J., Bargiello, T. A. & Verselis, V. K. The first extracellular loop domain is a major determinant of charge selectivity in connexin46 channels. *Biophys. J.* **79**, 3036–3051 (2000).

33. Oh, S., Verselis, V. K. & Bargiello, T. A. Charges dispersed over the permeation pathway determine the charge selectivity and conductance of a Cx32 chimeric hemichannel. *J. Physiol.* **586**, 2445–2461 (2008).

34. Kronengold, J., Trexler, E. B., Bukauskas, F. F., Bargiello, T. A. & Verselis, V. K. Pore-lining residues identified by single channel SCAM studies in Cx46 hemichannels. *Cell Commun. Adhes.* **10**, 193–199 (2003).

35. Zonta, F., Polles, G., Zanotti, G. & Mammano, F. Permeation pathway of homomeric connexin 26 and connexin 30 channels investigated by molecular dynamics. *J. Biomol. Struct. Dyn.* **29**, 985–998 (2012).

36. Lopez, W. et al. Mechanism of gating by calcium in connexin hemichannels. *Proc. Natl Acad. Sci. USA* **113**, E7986–E7995 (2016).

37. Harris, A. L. & Contreras, J. E. Motifs in the permeation pathway of connexin channels mediate voltage and Ca^{2+} sensing. *Front. Physiol.* **5**, 113 (2014).

38. Srinivas, M., Calderon, D. P., Kronengold, J. & Verselis, V. K. Regulation of connexin hemichannels by monovalent cations. *J. Gen. Physiol.* **127**, 67–75 (2006).

39. Trexler, E. B., Bennett, M. V., Bargiello, T. A. & Verselis, V. K. Voltage gating and permeation in a gap junction hemichannel. *Proc. Natl Acad. Sci. USA* **93**, 5836–5841 (1996).

40. Suchyna, T. M. et al. Different ionic selectivities for connexins 26 and 32 produce rectifying gap junction channels. *Biophys. J.* **77**, 2968–2987 (1999).

41. Veenstra, R. D. Size and selectivity of gap junction channels formed from different connexins. *J. Bioenerg. Biomembr.* **28**, 327–337 (1996).

42. Veenstra, R. D. et al. Selectivity of connexin-specific gap junctions does not correlate with channel conductance. *Circ. Res.* **77**, 1156–1165 (1995).

43. Nicholson, B. J. et al. The molecular basis of selective permeability of connexins is complex and includes both size and charge. *Braz. J. Med. Biol. Res.* **33**, 369–378 (2000).

44. Ebihara, L., Xu, X., Oberti, C., Beyer, E. C. & Berthoud, V. M. Co-expression of lens fiber connexins modifies hemi-gap-junctional channel behavior. *Biophys. J.* **76**, 198–206 (1999).

45. Xin, L. & Bai, D. Functional roles of the amino terminal domain in determining biophysical properties of Cx50 gap junction channels. *Front. Physiol.* **4**, 373 (2013).

46. Luo, Y., Rossi, A. R. & Harris, A. L. Computational studies of molecular permeation through connexin26 channels. *Biophys. J.* **110**, 584–599 (2016).

47. Kwon, T., Harris, A. L., Rossi, A. & Bargiello, T. A. Molecular dynamics simulations of the Cx26 hemichannel: evaluation of structural models with Brownian dynamics. *J. Gen. Physiol.* **138**, 475–493 (2011).

48. Verselis, V. K., Ginter, C. S. & Bargiello, T. A. Opposite voltage gating polarities of two closely related connexins. *Nature* **368**, 348–351 (1994).

49. Peracchia, C. & Peracchia, L. L. Inversion of both gating polarity and CO_2 sensitivity of voltage gating with D3N mutation of Cx50. *Am. J. Physiol. Cell Physiol.* **288**, C1381–C1389 (2005).

50. Srinivas, M., Kronengold, J., Bukauskas, F. F., Bargiello, T. A. & Verselis, V. K. Correlative studies of gating in Cx46 and Cx50 hemichannels and gap junction channels. *Biophys. J.* **88**, 1725–1739 (2005).

51. Xin, L., Nakagawa, S., Tsukihara, T. & Bai, D. Aspartic acid residue D3 critically determines Cx50 gap junction channel transjunctional voltage-dependent gating and unitary conductance. *Biophys. J.* **102**, 1022–1031 (2012).

52. Beyer, E. C., Ebihara, L. & Berthoud, V. M. Connexin mutants and cataracts. *Front. Pharmacol.* **4**, 43 (2013).

53. Pascolini, D. & Mariotti, S. P. Global estimates of visual impairment: 2010. *Br. J. Ophthalmol.* **96**, 614–618 (2012).

54. Shiels, A. & Hejtmancik, J. F. Mutations and mechanisms in congenital and age-related cataracts. *Exp. Eye Res.* **156**, 95–102 (2017).

55. Kistler, J., Christie, D. & Bullivant, S. Homologies between gap junction proteins in lens, heart and liver. *Nature* **331**, 721–723 (1988).

56. Kistler, J., Schaller, J. & Sigrist, H. MP38 contains the membrane-embedded domain of the lens fiber gap junction protein MP70. *J. Biol. Chem.* **265**, 13357–13361 (1990).

57. White, T. W., Bruzzone, R., Goodenough, D. A. & Paul, D. L. Mouse Cx50, a functional member of the connexin family of gap junction proteins, is the lens fiber protein MP70. *Mol. Biol. Cell* **3**, 711–720 (1992).

58. Reichow, S. L. et al. Allosteric mechanism of water-channel gating by Ca^{2+}-calmodulin. *Nat. Struct. Mol. Biol.* **20**, 1085–1092 (2013). 10.1038/nsmb.2630

59. Gold, M. G. et al. AKAP2 anchors PKA with aquaporin-0 to support ocular lens transparency. *EMBO Mol. Med.* **4**, 15–26 (2012). 10.1002/emmm.201100184

60. Reichow, S. L. & Gonen, T. Noncanonical binding of calmodulin to aquaporin-0: implications for channel regulation. *Structure* **16**, 1389–1398 (2008). 10.1016/j.str.2008.06.011

61. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protocols* **1**, 2856–2860 (2006). 10.1038/nprot.2006.468

62. Yang, B. et al. Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012). 10.1038/nmeth.2099

63. Myers, J. B. et al. The CaMKII holoenzyme structure in activation-competent conformations. *Nat. Commun.* **8**, 15742 (2017).

64. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).

65. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

66. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

67. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

68. Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.* **166**, 205–213 (2009).

69. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

70. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).

71. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

72. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

73. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

74. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

75. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 27–38 (1996).

76. Locke, D., Bian, S., Li, H. & Harris, A. L. Post-translational modifications of connexin26 revealed by mass spectrometry. *Biochem. J.* **424**, 385–398 (2009).

77. Grubmuller, H., Heymann, B. & Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin–biotin rupture force. *Science* **271**, 997–999 (1996).

78. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

79. Klauda, J. B. et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).

80. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).

81 Zuckerman, D. M. *Statistical Physics of Biomolecules: An Introduction* 1st edn (CRC Press; Boca Raton, 2010).

82. Im, W., Seefeld, S. & Roux, B. A grand canonical Monte Carlo–Brownian dynamics algorithm for simulating ion channels. *Biophys. J.* **79**, 788–801 (2000).

83. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

Chapter 4

Connexin-46/50 in a dynamic lipid environment resolved by CryoEM at 1.9 Å

This chapter has been published

Connexin-46/50 in a dynamic lipid environment resolved by CryoEM at 1.9 Å.

Nat. Commun. 11, 4331 (2020). Jonathan A. Flores[1,2,†], Bassam G. Haddad[1,†], Kimberly A. Dolan[1,3,†], Janette B. Myers[1], Craig C. Yoshioka[4], Jeremy Copperman[4], Daniel M. Zuckerman[4] and Steve L. Reichow[1,2,*]

Affiliations:

[1] Department of Chemistry, Portland State University, Portland OR 97201, U.S.A.

[2] Department of Chemical Physiology and Biochemistry, Oregon Health and Science University, Portland OR 97239, U.S.A.

[3] Current Address: Biophysics Graduate Group, University of California, Berkeley, CA 94720, U.S.A.

[4] Department of Biomedical Engineering, Oregon Health & Science University, Portland OR 97239, U.S.A.

† equal contribution

* corresponding author (reichow@pdx.edu)

Bassam Haddad designed, conducted, and analyzed the MD simulations with the guidance of Jeremey Copperman and Daniel Zuckerman. Jonathan Flores processed and analyzed CryoEM grids with guidance from Janette Myers and Craig Yoshioka. Kimberly Dolan prepared and characterized MSPs (nanodisc), and prepared CryoEM grids with Jonathan Flores. Steve Reichow directed the project.

Abstract

Gap junctions establish direct pathways for connected cells and tissues to transfer metabolic and electrical messages[1]. The local lipid environment is known to affect the structure, stability and intercellular channel activity of gap junctions[2-5]; however, the molecular basis for these effects remains unknown. To gain insight toward how gap junctions interact with their local membrane environment, we used lipid nanodisc technology to incorporate native connexin-46/50 (Cx46/50) intercellular channels into a dual lipid membrane system, closely mimicking a native cell-to-cell junction. Structural characterization of Cx46/50 lipid-embedded channels by single particle CryoEM revealed a lipid-induced stabilization to the channel, resulting in a 3D reconstruction at 1.9 Å resolution. Together with all-atom molecular dynamics (MD) simulations and 3D heterogeneity analysis of the ensemble CryoEM data, it is shown that Cx46/50 in turn imparts long-range stabilization to the dynamic local lipid environment that is specific to the extracellular lipid leaflet of the two opposed membranes. In addition, nearly 400 water molecules are resolved in the CryoEM map, localized throughout the intercellular permeation pathway and contributing to the channel architecture. These results illustrate how the aqueous-lipid environment is integrated with the architectural stability, structure and function of gap junction communication channels, and demonstrates the ability of CryoEM to effectively characterize dynamical protein-lipid interactions.

Main Text

The connexins are a family of transmembrane proteins (21 isoforms in human) that form intercellular channels for cell-to-cell communication[6]. These intercellular channels establish a ~1.4 nm pore that couples the cytoplasms of neighboring cells, and enable direct passage of electrical and small molecule signals (such as, ions, second messengers, hormones and metabolites)[7] and therapeutic agents[8]. 10's – 1000's of connexin channels may assemble together to form large hexagonally packed arrays, *a.k.a*. plaques, known as gap junctions. In this way, gap junctions enable the near instantaneous response of electrical synapses in the brain and heart, and contribute to the long-range signaling and metabolic coupling of most tissues. Because of these fundamental roles, aberrant gap junctional coupling is associated with a variety of human diseases, including blindness, deafness, skin disorders, arrhythmia, stroke and cancers[9-11].

Gap junction intercellular communication is facilitated by a unique macromolecular architecture, where intercellular channels directly couple the cytoplasms of two neighboring cells. The lipid bilayers of opposing cells are separated by a characteristic gap of ~3.5 nm[12], a feature for which these structures were first recognized in electron micrographs of cell sections[5,13]. Furthermore, large-scale gap junctional plaque formation is dependent upon a dense mosaic of protein-lipid interactions. *In vitro* reconstitution studies have established that plaque assembly and intercellular channel function are dependent on the lipid environment[2,14,15]. However, the molecular basis for these effects remain largely unknown, due to the lack of high-resolution structural information within a lipid bilayer.

Here, we present a CryoEM structure of native connexin-46/50 (Cx46/50) intercellular channels stabilized in a dual lipid nanodisc system at 1.9 Å resolution – providing an unprecedented level of detail for this class of membrane channels. These structural results are coupled with all-atom molecular dynamics (MD) simulation studies, which together reveal many new features of the connexin channels. Cx46/50 is shown to have a remarkable influence on the local lipid environment, effectively inducing a phase separation (to the gel state) that is specific to the extracellular lipid leaflet of the two opposed membranes. 3D heterogeneity analysis of the CryoEM data identified multiple lipid configurations that co-exist within the dynamic lattice of stabilized lipids, which is further detailed by MD. In addition, ~400 water molecules are resolved in the CryoEM map, localized at architectural and functionally important sites. Together this work uncovers previously unrecognized roles of the aqueous-lipid environment in stabilizing the structure and assembly of the gap junctions, and suggest Cx46/50 plays an important role in shaping the properties of local membrane environment.

Structural overview of connexin-46/50 in a dual lipid bilayer

Native (heteromeric/heterotypic) connexin-46/50 intercellular channels were purified from mammalian lens tissue (obtained from sheep), as previously described[16]. Freshly purified channels were reconstituted into self-assembling lipid nanodiscs containing pure

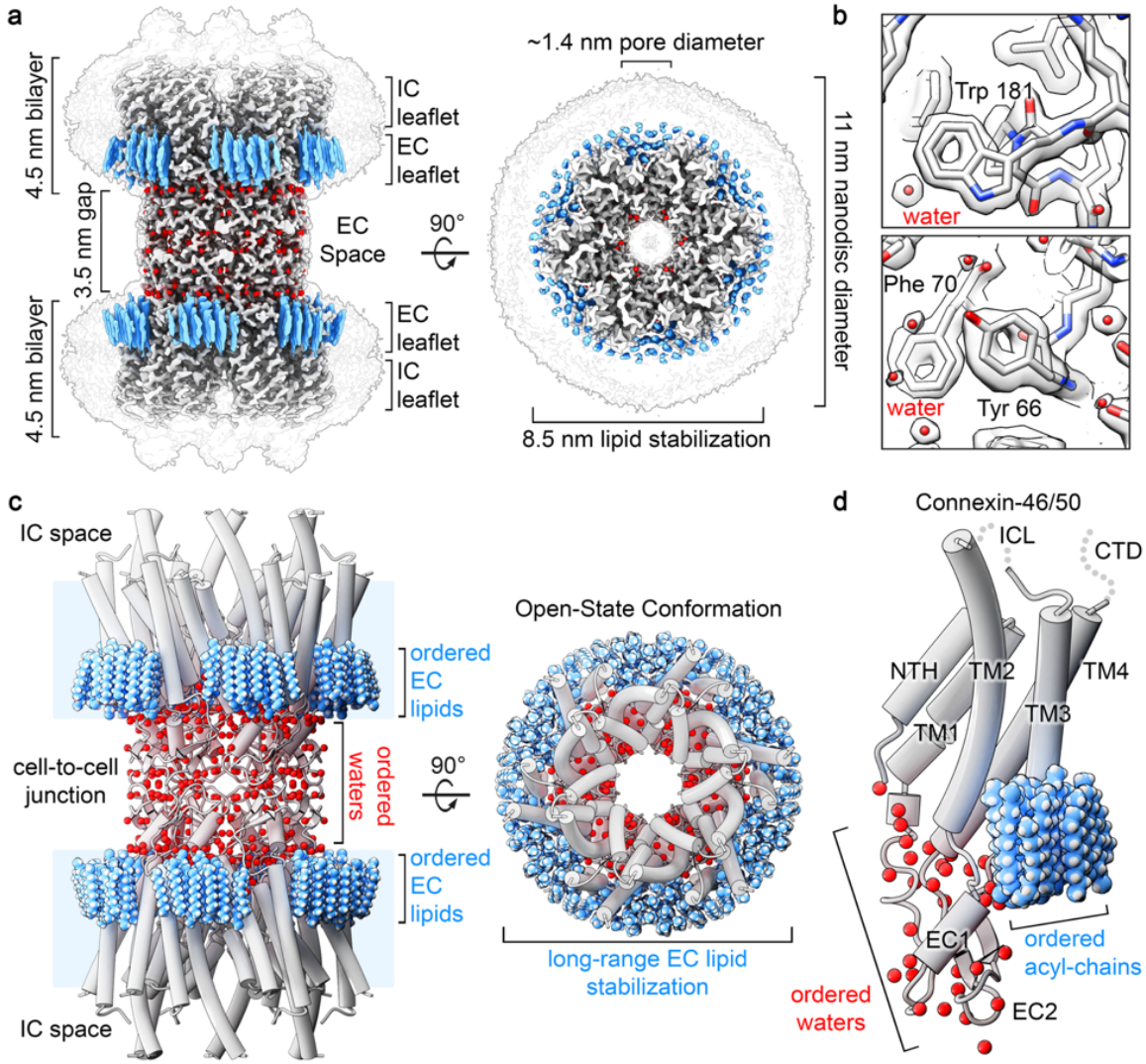


Figure 1. Structure of Connexin-46/50 in lipid nanodiscs by CryoEM. a) CryoEM 3D reconstruction of Cx46/50 (white) in an open state conformation, with resolved lipid acyl-chains (blue) and water molecules (red). Transparent silhouette displays the map at low-contour to illustrate the dimensions of the lipid nanodisc densities, with intracellular (IC) and extracellular (EC) lipid leaflets indicated. **b)** Zoom views of the CryoEM map and fitted atomic models, showing high-resolution features observed at 1.9 Å resolution. **c)** Model of Cx46/50 (cylinder representation) with extracellular (EC) lipids and ordered water molecules displayed (spheres). **d)** Cx46/50 monomer, and 15 bound lipids and 33 waters associated with each subunit. Domains labeled for transmembrane helices (TM1-4), extracellular loops (EC1-2) and n-terminal helix (NTH). The intracellular loop (ICL) and c-terminal domain (CTD) are not resolved, indicated by dotted lines.

dimyristoyl phosphatidylcholine (DMPC) at room temperature (~25° C), supported by the membrane scaffold protein MSP1E1[17] (see Methods). Under optimized conditions, the reconstitution resulted in a monodispersed population of intercellular channels

embedded into a pair of lipid-nanodiscs, as assessed by size-exclusion chromatography and negative stain EM (Extended Data Fig. 1).

Structure determination by high-resolution single particle CryoEM resulted in a high-quality 3D reconstruction, with an overall resolution of 1.9 Å (gold-standard FSC) (Fig. 1a,b, Extended Data Fig. 2,3 and Supplemental Movie 1). The quality of the CryoEM map allowed for detailed stereochemical structural refinement of both Cx46 and Cx50 (Fig. 1b, Extended Data Table 1 and Extended Data Fig. 3). The heteromeric pattern(s) of Cx46/50 co-assembly remain unresolved, following various attempts at computational image classification (see Methods). Nevertheless, atomic models of both Cx50 and Cx46 isoforms were equally well-fit into the D6-symmetrized CryoEM map, reflecting their close sequence and structural similarities, 89% sequence similarity over the structured regions and a resulting 0.17 Å backbone r.m.s.d. (see Methods and Extended Data Fig. 3 for details and limitations regarding the heterogeneity of the natively isolated specimen).

Cx46/50 is captured in the stabilized open-state, as previously described[16] (backbone C□ r.m.s.d. = 0.49–0.56 Å and Extended Data Fig. 5), and exposes many new features of the connexin channels that are detailed below. Intercellular channels are constructed by a dodecameric (12-mer) assembly, with six subunits assembled into 'hemi-channels' that dock together through extracellular domains, resulting in a continuous ~1.4 nm pore for intercellular permeation (Fig. 1a,c and Supplemental Movie 2). The distance separating

the two lipid nanodisc densities is ~3.5 nm (Fig. 1a,c), matching that observed by x-ray diffraction on native gap junctional plaques[12].

Each monomer consists of four transmembrane helices (TM1-4), two extracellular loops (EC1-2) that form the sites of docking interaction and an amphipathic n-terminal helix (NTH), implicated in channel selectivity/gating, is well resolved in the stabilized open-state, as previously described[16] (Fig. 1c,d and Supplemental Movie 3). However, the significant enhancement in resolution allowed for detailed refinement of sidechain conformations and notable improvement in precision at functional sites, including the NTH domain and the EC1/2 docking sites (Extended Data Fig. 3,5). Furthermore, the quality of the CryoEM map allowed for modeling previously un-resolved regions of TM2 and TM3, which effectively extend the cytoplasmic vestibule of the channel by ~20 Å, as compared to our previous model (Fig. 1c,d), significantly augmenting the electrostatic environment of the pore entrance (Extended Data Fig. 5). The intracellular loop (ICL) and c-terminal domain (CTD) remain unresolved, presumably due to intrinsic disorder of these regulatory domains[16,18,19].

Perhaps the most remarkable features of the CryoEM map, however, are the non-protein components of the cell-to-cell junction that are now resolved. A bouquet of 15 ordered lipid acyl-chains is held in place by each of the 12 connexin subunits, which appear to buttress the channel assembly by filling a cavity formed at the lateral subunit interfaces (Fig. 1a,c,d; *blue*). Surprisingly, acyl-chain densities are observed well beyond the first

layer of annular lipids that directly interact with the TM domains (primarily TM4 and TM3 of a neighboring subunit) (Fig. 1c,d and Extended Data Fig. 6), suggesting Cx46/50 has a long-range effect on the stability and biophysical properties of the membrane. Remarkably, all of the resolved lipid densities in the CryoEM map are specifically localized to the extracellular leaflet of the bilayer, indicating a selective interaction with the local lipid environment. In contrast, individual lipids are not resolved in the intracellular leaflet, even at lower contour levels (Fig. 1a and Extended Data Fig. 2) presumably due to intrinsic disorder and/or lack of specific interaction with this region of the channel.

In addition to stabilized lipids, 396 ordered water molecules are resolved throughout the channel (33 waters per subunit) (Fig. 1,2; *red* and Extended Data Fig. 6). Waters are found at both solvent accessible and buried sites within the core of the channel, apparently contributing to the permeation pathway and structural integrity of the channel assembly (Fig. 1,2). The assignment of water densities was validated by assessment of hydrogen bonding patterns (< 4 Å donor-acceptor distances) and supported by comparison to all-atom equilibrium MD simulations conducted in the presence of explicit water and 150mM NaCl or KCl (see Methods and Extended Data Fig. 7,8). There was no clear evidence that the resolved solvent sites correlated with low-affinity ion binding sites observed by MD (not shown). In the following sections, we describe these newly resolved features in further detail and discuss their potential structural and functional roles.

Stabilized waters contribute to the permeation pathway and core architecture of Cx46/50

Gap junctions establish aqueous pathways that allow a variety of cytosolic substrates, less than ~1 kDa in size, to permeate from cell to cell[20]. The permeation pathway is established by the pore-lining NTH domain, TM1/2 and EC1 domains (Fig. 1c,d). Within the channel pore of Cx46/50, there are 108 waters bound at solvent-exposed sites (9 per subunit). Pore-bound waters localize to regions of the EC1 domain and TM1 parahelix, and mediate an extensive network of h-bonding interactions, involving, D42, F43 (bonding), E48, Q49, N63 and R76 in Cx50 (positions 42 and 43 are Glu in Cx46), and several protein backbone interactions (Fig. 2a,b and Extended Data Fig. 6).

EC1 and the TM1 parahelix contribute to the selectivity, conductance and slow (loop) voltage-gating mechanisms of Cx46/50[16,21-23] and other connexins[24-29], and are implicated in Ca^{2+}-regulation in Cx26 by X-ray crystallography[19], MD studies[27,30], and by functional mutation studies of Cx46[30]. As such, these pore-lining waters may functionally contribute to these mechanisms, for example, by orienting or extending the hydrogen-bonding potential of amino-acid sidechains involved in the coordination of substrates (or regulatory ions), buffering the electrostatic properties of the channel pore, or integrating the electrostatic network that is proposed to couple EC1/TM1 to the fast (NTH) voltage-gating domain[31,32].

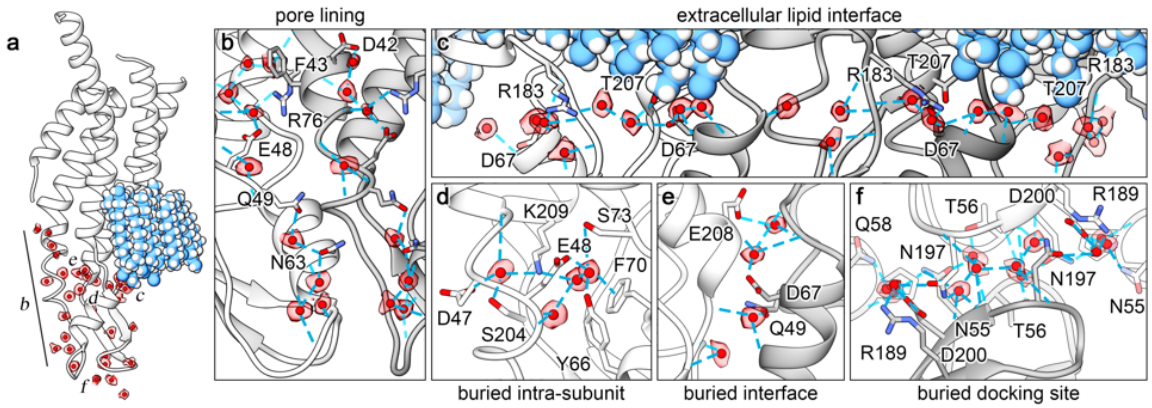


Figure 2. Ordered water molecules resolved in Cx46/50 by CryoEM. a) Cx46/50 subunit with segmented CryoEM density of waters overlaid in transparency (colored as in Fig. 1). Labels in panel a indicate position of the various zoom views, presented in panels **b–f**, showing water molecules bound to **b)** pore-lining sites, **c)** extracellular lipid interface, **d)** buried intra-subunit sites, **e)** buried subunit interface sites, and **f)** buried cell-to-cell docking sites. In panels b-f, amino acids sidechains forming hydrogen bonds to water are displayed (blue dotted lines) and labeled using Cx50 numbering.

On the extracellular surface of the channel, symmetry-related rings of tightly bound water molecules are organized at the extracellular aqueous-lipid boundary (Fig.1a,c and Fig. 2a,c). In the ensemble CryoEM map, the PC lipid head groups are not resolved (due to local disorder described in the following sections). Nevertheless, these stabilized rings of water are nominally positioned at the acyl-headgroup boundary of the extracellular lipid leaflet. These waters are stabilized by hydrogen bonds with EC1/2 residues (D67, R183/Q171 and T207/T195 in Cx50/Cx46, respectively) and expected to be further coordinated through non-specific interactions with the phospho-glycerol backbone of the extracellular PC lipids (Fig. 2c; and discussed below).

The EC1/2 domains appear to be the most well-ordered region of the channel, as reflected by local-resolution of the CryoEM density map (Extended Data Fig. 3) and root-mean-square-fluctuation (r.m.s.f.) analysis of MD-trajectories (Extended Data Fig. 7) This high-degree of stability reflects the important functional role of the EC1/2 domains in

maintaining an electro-chemical seal at the cell-to-cell junction. Several clusters of water molecules are found buried at sites located both within and between the EC1/2 domains of individual subunits (Fig. 2a, d-f). A cluster of four waters are buried within the EC1/2 domains is coordinated by residues D47, E48, Y66, F70 (□ bonding), S73, S204/S192 and K209/K197, in Cx50/46 respectively (Fig. 2d). Four additional waters are buried at the lateral EC domain interface formed by neighboring subunits, primarily coordinated by hydrogen bonding interactions with the peptide backbone and sidechains of Q49, D67 and E208/E196, in Cx50/Cx46 respectively (Fig. 2e). The degree of coordination of these buried waters suggest they contribute to the architectural integrity of EC1/2 docking domains, and may in part explain why deleterious mutations at D47, E48 and D67 in Cx50 linked to cataract formation disrupt junctional coupling and/or biogenesis[33-35].

The EC1/EC2 domains also play important roles in establishing the specificity of hemi-channel docking interactions formed between different connexin isoforms, and the ability to establish so-called homotypic or heterotypic channels[36-38]. Elucidating the determinants of hemi-channel recognition is therefore critical to understanding the principles dictating cell-type specificity of gap junctional coupling[39]. It has been proposed that isoform-specific hydrogen bonding patterns that bridge the EC1/EC2 interface govern hemi-channel docking compatibility[38,40]. Contributing to this bridging site in Cx46/50 is a cluster of 12 water molecules (per subunit pair) that are deeply integrated within a dense network of hydrogen bonds between EC1/EC2 residues of opposed subunits (Fig. 2f). At the center of this network is the highly conserved K/R-N-D motif

found in EC2 of Group I compatible connexins (including Cx50, Cx46, Cx32 and Cx26). Genetic mutations of this motif in Cx46/50 are linked to congenital cataracts[16,41], as well as other genetic disorders (*e.g.,* Charcot-Marie-Tooth disease[42] and non-syndromic deafness[43]), when mutated in other Group I connexins. These observations suggest interfacial waters may play previously unappreciated and functionally important roles in establishing the structural integrity of the intercellular channel and contribute to the specificity of hemi-channel docking interactions involved in regulating the formation of intercellular communication pathways.

Cx46/50 induces long-range ordering at the extracellular lipid leaflet

The degree of long-range stabilization to the local lipid environment observed in the Cx46/50 nanodisc reconstruction, extending several solvent layers away from the protein, is (to our knowledge) unprecedented. DMPC was selected as a model lipid because of the high PC content of mammalian (sheep) lens[44], and reconstitution studies show DMPC produces Cx46/50 assemblies that are indistinguishable from those formed with native lipids[14,45]. Due to its complete saturation DMPC has a relatively high phase-transition (*i.e.,* melting) temperature (T_m) compared to other biological lipids (T_m ~24° C in pure lipid vesicles[46]). This value is close to the temperature at which reconstitution was performed (~25° C, room temperature). However, in nanodiscs the melting temperature of DMPC is reportedly higher (~28° C), due to compartmentalization effects by the MSP scaffold[47]. Nevertheless, the specific localization of stabilized lipids to the extracellular leaflets

observed by CryoEM (and also by MD studies, described below) suggested long-range lipid stabilization is induced through interactions with Cx46/50 (Fig. 1, 3a).

To gain further insight into the lipid-stabilization observed by CryoEM, we analyzed time-averaged densities of DMPC acyl-chain positions obtained by unbiased all-atom MD simulations for both Cx50 and Cx46, conducted at 37° C, where the starting positions of DMPC molecules had been randomly placed into a 15.4 x 15.4 nm lipid bilayer (see Methods and Extended Data Fig. 7a). Following equilibration, the resulting acyl-lipid density profiles displayed remarkable similarity to what was resolved by CryoEM (Fig. 3a). In both cases, lipids within the extracellular leaflets are specifically stabilized, as compared to the intracellular lipid leaflet (Fig. 3a). Furthermore, the resolved clusters of acyl-chain densities obtained by MD display the same hexagonal packing pattern that extends 3-4 orders beyond the annular shell, as observed by CryoEM (Fig. 3a, *inset*).

The corroborating results obtained by MD imply that the lipid stabilization observed by CryoEM is specifically induced by structural features of the Cx46/50 TM domains, and not an artifact of the nanodisc. Each cluster of lipids is bound by a shallow pocket of hydrophobic and aromatic residues, displayed by TM2/3 and TM4 of adjacent subunits (Fig. 3a,b). A cleft, rich in aromatic sidechains (formed by F32, F84, L167/L155 and F168/F156 in Cx50/C46, respectively) intercalates into the bilayer, appearing to bisect the extracellular leaflet from the more disordered intracellular leaflet (Fig. 3b). In this way, it appears that the acyl-lipid binding pocket selectively grasps a large bouquet of lipids from

the extracellular leaflet, inducing long-range stabilization to the membrane through extensive Van der Waals interaction.

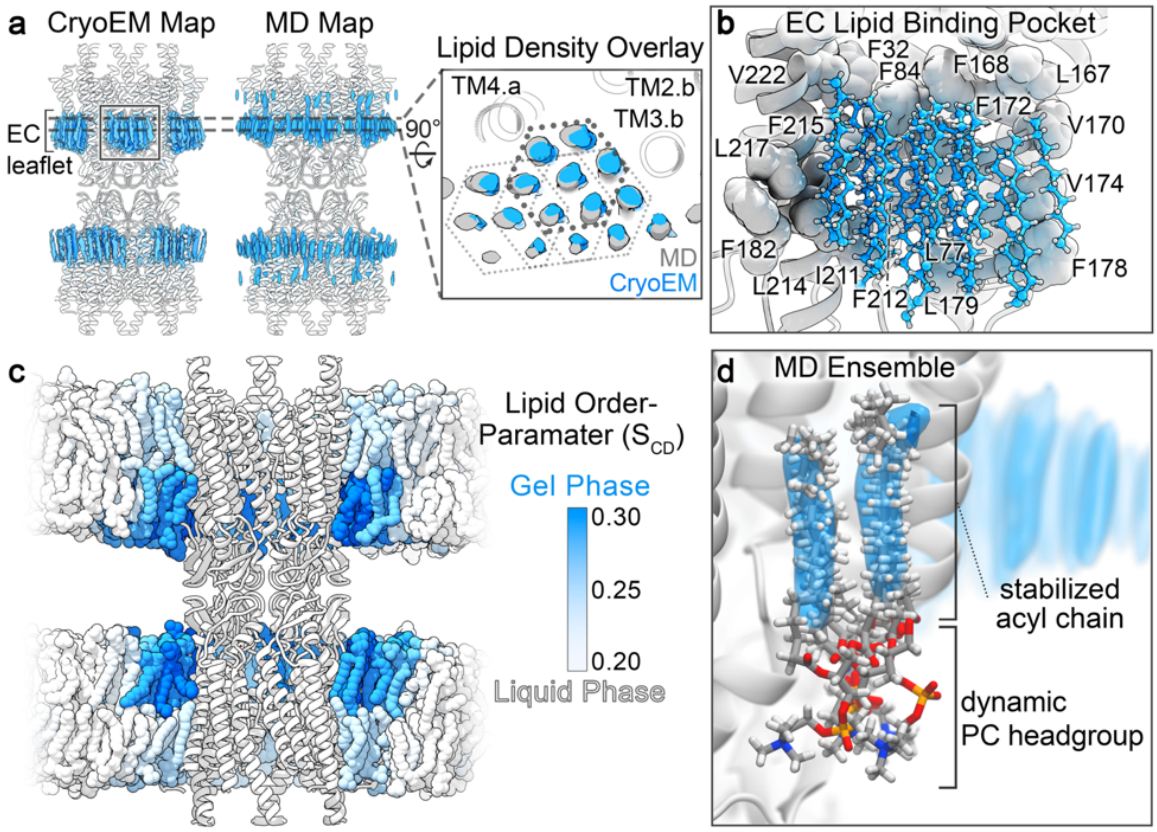


Figure 3. Cx46/50 induces a local phase-separation to the extracellular lipid leaflet. a) Comparison of acyl-lipid density maps (blue) obtained by CryoEM and time-averaged all-atom MD-simulation, overlaid onto the Cx46/50 ribbon structure (white). *Inset,* shows a slice-view (rotated 90°) of overlaid acyl-lipid densities by CryoEM (blue) and MD-simulation (grey). The hexagonal packing pattern of acyl-chains is indicated (solid and dotted lines), and TM helices interacting with lipid are labeled. Different subunits are indicated by suffix (*a* or *b*). **b)** Zoom-view of the acyl-lipid binding pocket, with lipid-binding residues displayed (spheres) and labeled (Cx50 numbering). **c)** MD-snapshot of Cx50 in phosphatidylcholine (PC) lipid bilayers, with time-averaged lipid order parameter (S_{CD}) for each lipid indicated by shading (blue = 0.30 to white = 0.20). **d)** Zoom view, showing an ensemble super-positioning of symmetry-related lipids obtained by MD-simulation (displayed as all atom representation) occupying the MD-based lipid acyl-chain density map (blue).

The extended acyl-lipid chain conformation and hexagonal packing adopted by the bouquet of bound lipids are indicative of a quasi phase-transition to the liquid-ordered (or gel-like) state. To obtain a more quantitative assessment of the degree of lipid stabilization, we extracted SN1 and SN2 lipid order parameters (S_{CD}) from the MD-

simulations, which have been parameterized to fit well to experimental NMR-based order parameters[48]. These results are consistent with the notion that lipids in the intracellular leaflet are maintained in a fluid state, while Cx46/50 induces a phase transition from a fluid to a gel-like state that is specific to the extracellular lipid leaflet, as indicated by a shift in order parameters to above ~0.25[49] (acyl-chain carbons 4–11; Fig. 3c and Extended Data Fig. 9), which extend ~10–20 Å from the protein surface, as observed by CryoEM.

Although this degree of stabilization to the local lipid environment is likely to depend on lipid type, the general effects may be functionally important. For example, by contributing to the architectural integrity at the gap junctional interface, partitioning specific types of lipids, or even templating long-range hexagonal packing interactions found in plaque assemblies[45,50]. In this context, it is noteworthy that connexins localize to lipid raft domains[51,52], which are rich in high T_m lipids (e.g., sphingomyelin) and characterized as forming a liquid-ordered state.

Annular PC lipids adopt a dynamic ensemble of conformational and configurational states

Another notable feature of the lipid densities observed in the CryoEM map is that PC head groups are not observed, despite sufficient resolution to expect such features (Fig. 1, 3a). Super-positioning of representative lipid conformations obtained by MD show that, although the annular lipid acyl-chains were relatively well ordered and superimpose, their corresponding head groups remain conformationally dynamic and/or heterogeneously

positioned (Fig. 3d and Supplemental Movie 4). Such behavior would rationalize the lack of resolvability in the averaged CryoEM density map. In an attempt to resolve this heterogeneity, we conducted 3D-classification analysis on the ensemble CryoEM data (Methods and Extended Data Fig. 4), which resulted in three distinct 3D reconstructions resolved at resolutions of ~2.5 Å (gold-standard FSC) (Fig. 4a-c and Extended Data Fig. 4).

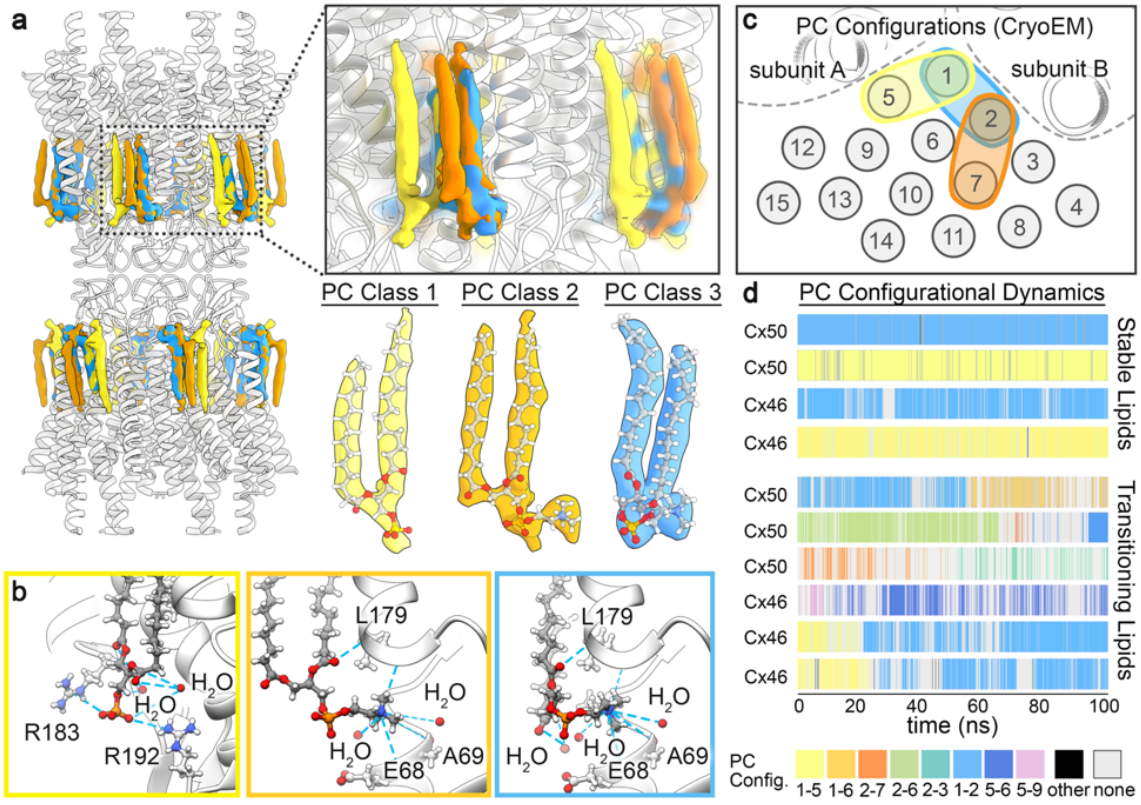


Figure 4. PC lipid configurational heterogeneity and dynamics resolved by CryoEM and MD. a) Segmented phosphatidylcholine (PC) density maps obtained by CryoEM 3D heterogeneity analysis and classification (PC Class 1 – yellow, PC Class 2 – orange, PC Class 3 – blue). *Insets*, show a zoom-view displaying the overlapping features of resolved lipid configurations, and segmented densities with fitted atomic-models obtained from the three PC classes. CryoEM density for all other non-unique acyl-lipid chains, with unresolved head groups, have been omitted for clarity. **b)** Zoom-view, showing Cx50 hydrogen bond interactions (blue dotted line) between the PC lipid headgroup and phospho-glycerol backbone. Interacting amino-acids and stabilized water molecules are labeled. Yellow box – PC Class 1, orange box – PC Class 2, blue box – PC Class 3. **c)** Illustration, showing acyl-chain positions and configurational assignments resolved by CryoEM (represented as grey circles and numbered 1 – 15). **d)** PC configurational classification and dwell times obtained by all-atom MD-simulation, showing representative populations of stable (non-transitioning) and dynamic (transitioning) lipids. PC configurations were classified by acyl-chain occupancy in densities numbered as in panel c, and colored uniquely (as indicated, bottom of panel d).

In each of the 3D classes, PC head groups and/or phospho-glycerol backbone of individual annular lipids were uniquely resolved (PC Class 1–3, Fig. 4a). The conformational state of Cx46/50 is very similar in all three classes, and essentially indistinguishable from models derived from the ensemble density map (Cα r.m.s.d.'s = 0.24 – 0.34 Å). Notably, structural features of each of these fully-resolved lipids are shared amongst these classes. For example, the SN2 acyl-chain of PC Class 1 overlays with the SN1 chain of PC Class 3 (Fig. 4a,c; *yellow and blue*). Likewise, the SN1 chain of PC Class 2 also overlays with the SN1 chain of PC Class 3 (Fig. 4a,c; *orange and blue*). This suggested that multiple, overlapping, configurational states are capable of supporting the same lattice of acyl-chain positions observed in the CryoEM reconstructions, consistent with non-specific and/or transient binding interactions.

The choline head group of PC Class 1 remained unresolved, however the negatively charged phospho-glycerol backbone is clearly visualized and appears to be stabilized by flanking positively charged arginine residues R183 and R192 in Cx50 (Q171 and R180 in Cx46) (Fig. 4b; *left*), and hydrogen bonding with two water molecules (part of the belt of extracellular waters described above, see Fig. 2c). PC Class 2 and 3 resolve distinct acyl-chain configurations, yet, both of these states share a similar placement of their positively charged choline head groups. Head group placement of these lipids is supported by non-specific hydrogen-bond interactions with backbone carbonyls presented by EC1 and TM3 (involving residues E68, A69 and L179 in Cx50; position 68 is an Arg in Cx46) and a buried water molecule (Fig. 4b; *center and right*). The phospho-glycerol backbones of PC Class

2/3 lipids are coordinated by hydrogen-bonding to local waters and the backbone amide of L179 in Cx50 (L167 in Cx46). Remarkably, the SN2 acyl chain and glycerol backbone of the PC Class 2 lipid is completely resolved, despite lacking any direct contact with the Cx46/50 protein interface (Fig. 4a,c).

Collectively, these observations support the notion that while the Cx46/50 acyl-chain interactions appear to be high-affinity, the lipid head group interactions are nonspecific and adopt a variety of configurational/conformational states. This is reinforced by our MD-simulation studies for both Cx50 and Cx46, where mapping of PC arrangements at resolved acyl-chain densities show a variety of configurational states that co-exist within the dodecameric assembly. Furthermore, during the timescale of the simulations (100 ns), time-resolved PC configurations could be classified as being either stable or dynamically transitioning between multiple configurational states (Fig. 4d and Supplemental Movie 5,6). Notably, the most stable (yet overlapping) configurations (*e.g.,* 1-2 and 1-5 configurations) are the same as those resolved by CryoEM 3D classification (Fig. 4a,c,d; *blue and yellow respectively*). Yet, other lipid trajectories were observed interconverting between these same configurations over this relatively short time-scale (Fig. 4d). The degree of configurational preference diminishes beyond the first two solvent shells, presumably due to the loss of energetic influence induced by protein interactions (Fig. 4d), and reflect the randomized head group arrangements expected of a bulk lipid population. Taken together, these data show Cx46/50 stabilizes the dynamic local lipid environment through non-specific interactions with the extracellular leaflet,

with multiple configurational PC lipid states existing at the annular interface and effectively captured by CryoEM.

Concluding Remarks

The structure and function of membrane proteins are deeply integrated with their lipid environment. Our mechanistic understanding of protein-lipid interactions have been largely shaped by high-resolution structures of membrane proteins where specifically bound lipids have been captured at well-defined binding sites[53]. Yet, most interactions made between membrane proteins and their local membrane environment are relatively non-specific and highly dynamic. The mechanistic principles and biophysical consequences underlying such interactions remains poorly understood, as these interactions are typically lost during protein purification, or remain too dynamic to resolve by traditional structural methods. By exploiting the potential of lipid nanodisc technologies coupled with single particle CryoEM and MD simulation, we show that Cx46/50 intercellular communication channels form dynamic interactions with annular lipids. These non-specific interactions have long-range stabilizing effects capable of inducing a phase separation to high T_m lipids, which may extend ~20 Å from the protein surface. These interactions appear selective toward the extracellular leaflet of pure PC membranes, which may have significant consequences on the biomechanical properties and lipid composition of gap junctional domains. In fact, the lack of resolved lipids in the intracellular leaflet may reflect the selectivity at this leaflet toward non-PC lipid types, as suggested for Cx26/32[15]. The methods developed here provide a valuable high-resolution

platform for developing our deeper understanding of the specificity and physiological role lipids play in gap junction biology, and how aberrant lipid environments may contribute to connexin-related pathologies. Indeed, the capability of resolving connexin channels beyond the critical threshold of ~2.0–2.5 Å resolution, the precision desired for structure-based drug design – *e.g.,* providing detailed stereochemical models and placement of architectural water molecules – now opens the door to rational development of selective high-affinity pharmacological tools that are desperately needed in this field to better understand and potentially treat a wide range of connexin-opathies[54].

Acknowledgements

We thank Dror Chorev and Carol V. Robinson for help analyzing specimens prepared for CryoEM. We are grateful to the staff at the OHSU Multiscale Microscopy Core and Advanced Computing center, and to the Pacific Northwest Center for CryoEM (supported by NIH Grant U24GM129547) and accessed through EMSL (grid.436923.9). J.A.F. is supported by the National Institutes of Health NRSA (F31-EY030409). K.A.D. is supported by National Institutes of Health BUILD EXITO Program (TL4-GM118965) and Berkeley Molecular Biophysics Training Grant (T32-GM008295). C.C.Y. is supported by OHSU. J.C. and D.M.Z. are supported by the OHSU Center for Spatial Systems Biomedicine, by the National Science Foundation (MCB 1715823) and by the National Institutes of Health (R01-GM115805). S.L.R. is supported by the National Institutes of Health (R35-GM124779).

Author Contributions

J.A.F, B.G.H and K.A.D. contributed equally. K.A.D. and J.A.F. conducted the protein purification and reconstitution of CryoEM specimens. J.A.F. collected the CryoEM datasets, performed image analysis and atomic modeling. K.A.D., J.B.M., C.C.Y. contributed to image analysis. B.G.H. conducted and analyzed the MD simulations. B.G.H., J.C. and D.M.Z. contributed to the experimental design and analysis of MD simulations. All authors contributed to manuscript preparation. S.L.R. provided overall guidance to the design and execution of the work.

Author Information

These authors contributed equally: Jonathan A. Flores, Bassam G. Haddad and Kimberly A. Dolan.

Affiliations

Department of Chemistry, Portland State University, Portland OR 97201, U.S.A.

Jonathan A. Flores, Bassam G. Haddad, Kimberly A. Dolan, Janette B. Myers and Steve L. Reichow

Department of Chemical Physiology and Biochemistry, Oregon Health and Science University, Portland OR 97239, U.S.A.

Jonathan A. Flores and Steve L. Reichow

Biophysics Graduate Group, University of California, Berkeley, CA 94720, U.S.A.

Kimberly A. Dolan (current address)

Department of Biomedical Engineering, Oregon Health and Science University, Portland OR 97239, U.S.A.

Craig C. Yoshioka, Jeremy Cooperman and Daniel M. Zuckerman

Competing Interest

The authors declare no competing interests.

Corresponding Author

Correspondence and requests for materials should be addressed to reichow@pdx.edu

Methods

MSP expression and purification. A plasmid containing the coding sequence for membrane scaffold protein 1E1 (MSP1E1) was obtained from Addgene[17] and the protein was expressed and purified as described[55], with minor modification. Freshly transformed *E. coli* cells (BL21Gold-DE3) were grown in LB medium containing 50 μg mL^{-1} kanamycin at 37° C with shaking (250 rpm). Induction with 0.5 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) was performed when OD_{600} reached 0.5–0.6, and allowed to express for 3–5 hours post-induction at 37° C. Cells were harvested by centrifugation at 4,000 x *g* for 20 minutes at 4° C, and cell pellets were resuspended in MSP Lysis Buffer (40 mM Tris [pH 7.4], 1% Triton X-100, 1 mM PMSF) at a density of ~20 mL of Lysis Buffer per Liter of culture. Cell suspensions were flash frozen in liquid nitrogen and stored at –86° C for up to several months.

Frozen cell suspensions were thawed from –86° C storage, supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) and lysed by sonication on ice. Crude lysate was cleared by ultra-centrifugation at 146,550 x *g* for 30 minutes at 4° C. The supernatant was filtered (Millipore; 0.22 μm) and applied to a gravity column with 5 mL of HisPur Ni-NTA resin (Thermo Fisher Scientific) prepared in equilibration buffer (40mM Tris [pH 7.4]). MSP-bound resin was washed with 5 column volumes (CV) of equilibration buffer, followed by 5 CVs of each of the following: Triton buffer (40 mM Tris [pH 8.0], 300 mM

NaCl, 1% TX-100), Cholate buffer (40 mM Tris [pH 8.0], 300 mM NaCl, 50mM cholate), Imidazole Wash Buffer (40 mM Tris [pH 8.0], 300 mM NaCl, 50 mM imidazole). MSP1E1 was eluted with 3 CVs of Elution Buffer (40 mM Tris [pH 8.0], 300 mM NaCl and 750 mM imidazole). The eluate was filtered (Millipore; 0.22 μm) and applied to a size exclusion chromatography (SEC) column (ENC70; BioRad) equilibrated in 20 mM HEPES (pH 7.4), 150 mM NaCl and 1 mM EDTA using an FPLC (NGC system; BioRad). Peak fractions were monitored by UV_{280}, pooled and concentrated to 400-600 μM using a centrifugal device. Final protein concentration was determined by UV absorbance at 280 nm. Samples were aliquoted, flash frozen in liquid nitrogen and stored at –86° C for up to several months.

Cx46/50 purification and nanodisc reconstitution. Native Cx46/50 intercellular channels were isolated as previously described[16]. Briefly, lamb eyes were obtained from the Wolverine Packers slaughterhouse (Detroit, MI), and the lenses were removed using a surgical blade and stored at –86° C. Gap junction intercellular channels were isolated from the core lens fiber tissue, containing c-terminal truncation variants of Cx46 and Cx50 (*a.k.a.* MP38)[56-59]. Details of the purification procedure are provided below.

Lenses were thawed from –86° C, core lens fiber cell tissue was dissected from the outer cortical tissue using a surgical blade and stripped core membranes were prepared as described[60-62]. Total protein concentration was determined by BCA (Pierce) and membranes were stored at –86° C, in storage buffer (10 mM Tris [pH 8.0], 2 mM EDTA, 2 mM EGTA) at a total protein concentration of ~2 mg mL^{-1}. Stripped membranes were

thawed from –86° C and solubilized with 10 mM Tris (pH 8.0), 2 mM EDTA, 2 mM EGTA, 1% (wt vol^{-1}) n-decyl-β-D-maltoside (DM) for 30 minutes at 37° C. Insoluble material was cleared by ultra-centrifugation at 146,550 x *g* for 30 minutes at 4° C. The supernatant was filtered (Millipore; 0.22 μm) and separated by anion-exchange chromatography (UnoQ, BioRad) with buffer A (10 mM Tris [pH 8.0], 2 mM EDTA, 2 mM EGTA, 0.3% DM [wt vol^{-1}]). Protein was eluted with a 20 CV gradient of buffer B that additionally contained 500 mM NaCl. Elution peaks containing Cx46/50, as determined by SDS-PAGE, were pooled and applied to a size exclusion chromatography (SEC) column (Superose 6 Increase 10/300 GL; GE Healthcare) equilibrated with SEC buffer (20 mM HEPES [pH 7.4], 150 mM NaCl, 2 mM EDTA, 2 mM EGTA and 0.3% DM [wt vol^{-1}]). Peak fractions containing purified Cx46/50 were pooled and concentrated to 5–6 mg mL^{-1} with a centrifugal device (Vivaspin 6; 50-kDa cut-off filter; Sartorius). Protein concentration was determined by UV absorbance at 280 nm. All chromatography steps were performed by FPLC at 4° C.

Freshly purified Cx46/50 was reconstituted into MSP1E1 nanodiscs using dimyristoyl phosphatidylcholine (DMPC) lipids, following established procedures[55,63]. Chloroform-solubilized DMPC (Avanti) was dried under nitrogen gas and left under vacuum overnight to remove residual solvent. The resulting thin film was resuspended in 5% DM (wt vol^{-1}) to a final DMPC concentration of 30 mM, and solubilized in a sonicator bath at 37° C. DM-solubilized Cx46/50 (5–6 mg mL^{-1}) was combined with DMPC at a molar ratio of 0.6:90 (Cx46/50:DMPC) and incubated at 25° C with gentle agitation for 60 minutes. Purified

MSP1E1 was then added at a final molar ratio 0.6:1:90 (Cx46/50:MSP1E1:DMPC) and allowed to incubate at 25° C for an additional 20 minutes. Detergent was removed with SM-2 Bio-Beads (BioRad) at a ratio of 30:1 beads:detergent (wt wt^{-1}) by overnight incubation at 25° C with gentle agitation. Bio-Beads were removed by filtration and the sample was ultra-centrifuged at 146,550 x *g* for 15 minutes at 4° C to remove insoluble material. The supernatant was filtered (Millipore; 0.22 μm) and applied to an SEC column (Superose 6 Increase 10/300 GL; GE Healthcare) equilibrated in 20 mM HEPES (pH 7.4) and 150mM NaCl, to separate empty nanodiscs from Cx46/50-embedded nanodiscs. Peak fractions containing both Cx46/50 and MSP1E1, as determined by SDS-PAGE, were collected and concentrated using a centrifugal device (Vivaspin 6; 50-kDa cut-off filter; Sartorius) to a final concentration ~2.5 mg mL^{-1}, as determined by UV absorbance at 280nm (Extended Data Fig. 1a). All chromatography steps were performed by FPLC at 4° C. The presence of both Cx46 and Cx50 in the final sample was confirmed by western blot analysis using polyclonal antibodies directed against the N-terminal domain of Cx46 (AP11570PU-N, Acris) and the N-terminal domain of Cx50 (LS-C116220, LSBio) (Extended Data Fig. 1b).

Negative-stain electron microscopy. Cx46/50-lipid nanodisc complexes were prepared for negative stain EM as described[16]. Briefly, a 3 μl droplet of sample (~0.02 mg mL^{-1}) was applied to a glow-discharged continuous carbon coated EM specimen grid (Ted Pella), blotted with filter paper and washed two times with detergent-free SEC buffer. The

specimen was then stained with freshly prepared 0.75% (wt vol^{-1}) uranyl formate (SPI-Chem).

Negatively stained specimens were visualized on a 120kV TEM (iCorr, Thermo Fisher Scientific) at 49,000x magnification at the specimen level (Extended Data Fig. 1c). A total of 76 digital micrographs were collected on a 2k x 2k CCD camera (Eagle 2K TEM CCD, Thermo Fisher Scientific) with a calibrated pixel size of 4.37 Å and with nominal defocus values ranging from 1.5–3.0 μm. All negative-stain image processing was performed in EMAN2.2[64,65]. After contrast transfer function (CTF) parameters were determined, micrographs with significant astigmatism or drift were excluded based on visual inspection of Thon rings in the power spectrum. 7,598 hand-picked particles were extracted with 84 x 84 pixel box size and subjected to multiple rounds of reference-free 2D classification, resulting in a final dataset of 3,826 "good" particles. Representative class averages are shown in (Extended Data Fig. 1c), which revealed dimensions consistent with the expectation that Cx46/50 intercellular channels had been reconstituted into a pair of lipid-nanodiscs.

CryoEM specimen preparation and data collection. Samples were prepared for CryoEM by applying 5 μl freshly purified Cx46/50-lipid nanodisc complex (~2.5 mg mL^{-1}) to a glow-discharged holey carbon grid (Quantifoil R 1.2/1.3, 400 mesh) for 10 seconds. The grid was blotted for 4.0 seconds and plunge frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific) at 100% humidity and stored under liquid nitrogen.

CryoEM specimen grids were imaged on a Titan Krios (Thermo Fisher Scientific) operated at 300 kV. Dose-fractionated image stacks were recorded on a Falcon 3EC Direct Electron Detector (Thermo Fisher Scientific) at 120,000x nominal magnification in counting mode, with a calibrated pixel size of 0.649 Å $pixel^{-1}$ (Extended Data Fig. 2a). The dose rate was 1.14 e^- $pixel^{-1}$ sec^{-1}, with 5 frames sec^{-1} collected for a total exposure of 30 seconds, resulting in a total dose for each exposure of ~52.5 e^- $Å^{-2}$. A dataset of 2,087 movies was obtained with nominal defocus values ranging from 1.0–2.2 μm, and data collection parameters were controlled in an automated manner using EPU (Thermo Fisher Scientific).

Cryo-EM image processing for high-resolution work-flow. The full dataset of 2,087 movies were corrected for beam-induced motion in RELION-3.0[66] and contrast transfer function (CTF) estimation was performed with Gctf[67] on the non-dose-weighted, aligned micrographs. Laplacian-of-Gaussian autopicking in RELION-3.0 yielded an initial set of 756,374 picks, which after multiple rounds of 2D classification left 183,784 *bona fide* particles (binned to a 64-pixel box, 3.894 Å $pixel^{-1}$). These particles were used to generate a *de novo* initial model in RELION, and subsequent 3D refinement of these particles yielded a map at 8.0 Å resolution (64 pixel box, 3.894 Å $pixel^{-1}$). This map was low-pass filtered to 20 Å and projected in 14 unique orientations to perform 3D template-based autopicking in RELION-3.0 to yield 1,210,797 particle picks. Following multiple rounds of 2D classification, this dataset yielded 379,423 “good” particles (200-pixel box, 1.947 Å

pixel^{-1}) (Extended Data Fig. 2b). Particles that had been translated within 20 Å of their nearest neighbor were removed to prevent invalidation of gold-standard Fourier-shell correlation by duplicate particles. Removal of 120,228 duplicates yielded a 259,195 refined particle set.

This particle set was then re-extracted (1.62 Å pix^{-1}, 280-pixel box) and subjected to 3D refinement (D6 symmetry), yielding a map at 3.3 Å resolution. A subsequent round of de-duplication (20 Å cut-off) yielded 227,618 particles that were again re-extracted (0.974 Å pix^{-1}, 512-pixel box) and subjected to 3D-refinement (D6 symmetry), which improved the resolution to 3.2 Å. Two rounds of Bayesian polishing and CTF refinement (per-particle defocus, per-micrograph astigmatism) with subsequent 3D refinement (D6 symmetry) yielded a map at 2.7 Å resolution. Particles were then completely unbinned (400-pixel box, 0.649 Å pix^{-1}) and subjected to another round of 3D refinement (D6 symmetry), yielding a map that reached the same resolution prior to unbinning (2.7 Å). Bayesian polishing and subsequent 3D refinement of these particles showed no significant improvement. The maps generated up to this stage were obtained without applied masks and there was no evidence of resolved lipids within the intracellular leaflet (see Extended Data Fig. 2, Step 2).

At this stage, the newly-developed tools in RELION-3.1-beta[68] were implemented to estimate the degree of beam tilt and high-order aberrations (3-fold and 4-fold astigmatism) present in the particle images. Subsequent 3D refinement (D6 symmetry)

improved the resolution to 2.2 Å. Particles that had been translated to within 35 Å of their nearest neighbor (6,224 particles) were again removed to prevent invalidation of the gold-standard Fourier-shell correlation from duplicate particles. The remaining 221,394 particles were subjected to 3D classification into 2 classes with D6 symmetry and a tight solvent mask. Approximately ~89% of the particles (196,320) fell into one class that was subsequently refined to 2.2 Å resolution (D6 symmetry and solvent mask applied). The remaining 11% of particles (26,005) yielded a 2.0 Å resolution map after 3D refinement (D6 symmetry and solvent mask applied). There was no clear difference in protein conformation between these two classes, however, the high-resolution class was characterized by a particle set with comparatively lower defocus values (mean defocus = 1.20 ± 0.29 µm (s.d.)) as compared to the lower resolution class (mean defocus = 1.86 ± 0.52 µm (s.d.)) (https://github.com/huwjenkins). All subsequent processing steps were performed on this high-resolution particle set.

Particles were re-extracted with an expanded box size (initially to 448-pixels) to mitigate delocalized CTF signal from particle images with relatively high defocus. New polishing parameters were obtained by running the Bayesian polishing job type in RELION-3.1-beta in “Training mode” on a random 5,000 particle subset of these refined particles. Bayesian polishing was performed with these new parameters and the subsequent 3D refinement (D6 symmetry and solvent mask applied) improved the resolution slightly to 1.97 Å. This process was iterated multiple times with successive increase in box size and incrementally tighter solvent mask applied during Bayesian polishing until no further improvements

were observed, resulting in a final box size of 540 pixels and refined map at 1.94 Å resolution with D6-symmetry and 2.3 Å resolution without symmetry (Gold-Standard, 0.143 cut-off)[69] (Extended Data Fig. 2c, 3a). The asymmetric reconstruction showed no discernable difference in protein conformation as compared to the D6-symmeterized map. Local resolution of the final map was estimated in RELION-3.1-beta[68], and local resolution-filtered maps were generated for model building (Extended Data Fig. 2d, 3b). A schematic illustrating this high-resolution CryoEM workflow is presented in (Extended Data Fig. 2c).

Cryo-EM image processing workflow for lipid classification. For classification and analysis of lipid configurational/conformational heterogeneity, a modified workflow starting from the totally unbinned 227,618 particle set (0.649 Å pix^{-1}, 400-pixel box) which yielded the 2.7 Å resolution map was applied, as described here (and illustrated in Extended Data Fig. 4a). The particle set was subjected to 3D classification (eight classes), with D6 symmetry and a generous solvent mask applied. Two of the eight classes yielded maps in which the lipid configuration was unambiguously resolved: assigned as PC Class 1, containing 9,190 particles (~4% of the data) and PC Class 3, containing 6,944 particles (~3% of the data). Overlapping configurations were resolved in two of the other 3D classes, and so particles from these classes were combined and subjected to a second round of 3D classification with only 2 classes and a tight solvent mask applied. This yielded one class with unresolved lipid configurations, and a second class in which the lipid configuration was unambiguously resolved: assigned PC Class 2, containing 6,075

particles (~3% of the data). Particles assigned to PC Class 1, 2, and 3 were separately subjected to a final round of 3D refinement with a solvent mask and D6 symmetry applied (Extended Data Fig. 4a). The final reconstructions from particles in each of these classes all reached ~2.5 Å resolution (Gold-Standard, 0.143 cut-off) (Extended Data Fig. 4b). Local resolution was estimated in RELION-3.1-beta, and local resolution-filtered maps were generated for model building (Extended Data Fig. 4c).

We note that alternative approaches to heterogeneity analysis were also pursued. For example, using the signal-subtraction and symmetry expansion methods in attempt to characterize the specimen heterogeneity at the single subunit level[70]. However, these approaches did not resolve any additional PC lipid conformations/configurations than those captured by the approach described above. It also did not help to resolve the Cx46 and Cx50 isoform heterogeneity.

Atomic modelling, refinement and validation. For all atomic models of Cx46 and Cx50, initial models were derived from previously reported CryoEM structure of amphipol-stabilized Cx46 and Cx50 (PDB 6MHQ and 6MHY[16], respectively). Initial models were fit as rigid bodies into the D6-symmetrized CryoEM maps with applied local resolution-filtering using UCSF Chimera[71]. All atom models for Cx46 and Cx50 were further built into the CryoEM density maps with COOT[72], and subjected to real-space refinement in PHENIX[73] with secondary structure and non-crystallographic symmetry (D6) restraints applied. Several iterations of manual adjustment of the protein model in COOT followed by real-

space refinement in PHENIX, were performed while monitoring model quality with MolProbity[74] and quality of side chain fit with EMRinger[75]. Coordinate and restraint files for the dimyristoylated phosphatidylcholine (DMPC) ligands were generated with PHENIX eLBOW[76]. DMPC molecules were manually fit into the Cryo-EM density with COOT. Since density for the phosphatidylcholine (PC) head groups was not resolved in the high-resolution ensemble CryoEM map (1.9 Å map), head group and acyl chain atoms that could not be accommodated by the density were deleted. For the PC Lipid Classes 1–3, the postprocessed maps from RELION were low pass-filtered to 3.5 Å resolution to facilitate modeling of the fully-resolved PC lipids. COOT was further used to manually place water molecules into solvent densities of the CryoEM maps. Appropriate placement of waters was determined by the following three criteria: 1) confirmation of at least two hydrogen bond donor/acceptor interactions with the FindHBond tool in UCSF Chimera (< 4 Å donor-acceptor distance) and by visual inspection, 2) confirmation of solvent densities consistently observed in both gold-standard separated half-maps (contoured $\geq 2.5\ \sigma$), and 3) as an additional measure we looked for density overlap between the local resolution-filtered Cryo-EM map (contoured $\geq\ 5.3\ \sigma$) and the time-averaged water density map generated by equilibrium molecular dynamics simulation (contoured $\geq\ 5.0\ \sigma$) to help assign weak experimental water densities (Extended Data Fig. 8, see calculation of water density maps from MD described in Methods below). However, not all of the assigned CryoEM water densities were observed by MD (76% of waters were observed at equivalent positions by CryoEM and MD). Several iterations of real-space refinement on the entire model were completed until refinement statistics converged.

Disclosure of unresolved heteromeric/heterotypic assemblies of Cx46/50. All models of Cx46 and Cx50 were built using D6 (12-fold) symmetrized CryoEM maps. Because native Cx46/50 intercellular channels may form homomeric and/or various patterns of heteromeric/heterotypic configurations[16,77,78], this map most likely represent a heterogeneous mixture of these two isoforms[16]. This approach was chosen because all attempts to separate the heteromeric/heterotypic assembly of these two isoforms using image classification procedures were unsuccessful (presumably due to the close sequence and structural similarity of these two isoforms) (see also[16]). Indeed, Cx46 and Cx50 are 80% identical and 89% similar in sequence over the resolved structural domains, while sites of difference are typically at solvent exposed regions (Extended Data Fig. 6a). Despite this limitation, all atomic-models generated by this approach showed good stereochemical refinement statistics (see Extended Data Table 1), and significant improvements to the previously described amphipol-stabilized models that were refined to 3.4 Å resolution (Extended Data Fig. 5a-e). It is important to note that sites in the density maps where the sequence of Cx46 and Cx50 are identical or similar, both models fit well into the D6 symmetrized map, and these regions tend to display well-resolved sidechain density (Extended Data Fig. 3c,d). Over regions where the sequence of Cx46 and Cx50 differ, sidechain density is sometimes weaker and/or displays appearance of density consistent with a mixture of both isoforms (Extended Data Fig. 3c,d). These observations are possibly due to the imposed D6 symmetry averaging of density belonging to two different sidechains in these areas, or relative flexibility at these sites as many of these

residues are also solvent-exposed. In these areas of difference, where EM density is observed, both Cx46 and Cx50 can be fit into the density equally well (Extended Data Fig. 3c,d). Nevertheless, caution should be used with interpretation of the conformational details at these sites of isoform difference.

Molecular dynamics simulations. Visual Molecular Dynamics (VMD) v1.9.3[79] was used to build systems for sheep Cx46 and Cx50 in a dual lipid-bilayer with varying salt conditions, designed to mimic either the cellular environment (cytoplasmic KCl, extracellular NaCl) or experimental CryoEM conditions (uniform NaCl). To produce unbiased analysis of water and lipid interactions, all water and lipid molecules derived by CryoEM analysis were removed from the Cx46 and Cx50 models prior to the MD setup. Each system comprised the full dodecameric gap junction intercellular channel, prepared in explicit water (model TIP3P) and embedded in two lipid bilayers composed of dimyristoylated phosphatidylcholine (DMPC), mimicking the cell-to-cell junction. For all models, sidechains were protonated according to neutral conditions and the HSD model was used for all histidine residues. Disulfide bonds identified in the experimental structures were enforced. Amino acids corresponding to the intracellular loop (ICL; residues 110–136 in sheep Cx46 and residues 110–148 in sheep Cx50) and c-terminal domain (CTD; residues 225–413 in sheep Cx46 and residues 237–440 in sheep Cx50) were not included for the MD simulations, as experimental data describing the structure of these large flexible domains (~30 residue ICL and ~200 residue CTD in Cx46 and Cx50) are missing. The introduced n- and c-terminal residues resulting from the missing ICL segment (sheep Cx46

R109 and K137; sheep Cx50 R109 and R149) were neutralized. All of the systems were modified with an n-terminal acetylation (at the starting residue Gly 2) in VMD through an all-atom acetylation patch in the automated PSF-Builder, in accordance with previously described proteomics analysis on native Cx46/50[16,59,80], and expectation that this species would predominate in cells[81]. A complete list of modeled residues for each system is provided in Extended Data Fig. 7a.

The prepared protein structures were submerged in a hydration shell using Solvate 1.0.1[82]. Water was removed from sections of the channel corresponding to transmembrane domains, based on hydrophobic character and localization of lipid-nanodisc observed in the experimental CryoEM data (+/- 20–50 Å from the center of the channel). The CHARMM-GUI membrane-builder[83] was used to build the DMPC bilayers (pre-melted), with dimensions of 154 x 154 Å for Cx46 and Cx50, and lipids overlapping with protein were removed. The entire system was then placed in a water box with dimensions 147 x 147 x 174 Å for both Cx46 and Cx50, using VMD's Solvate plugin. The system was neutralized using the Autoionize plugin, then 150 mM KCl and 150 mM NaCl was added to the solvent areas corresponding to intracellular and extracellular regions of the simulation box for the "KCl" systems, while the "NaCl" systems contained 150 mM NaCl for the entire box. A summary of atoms counts for each system is provided in Extended Data Fig. 7a.

CUDA-accelerated nanoscale molecular dynamics (NAMD) 2.13[84] was used for all classical MD simulations, using the CHARMM36 force-field[85] for all atoms and TIP3P explicit model for water. Each system was prepared following the same minimization and equilibration protocol, as follows. An initial minimization step, where the lipids, solvent and ions were allowed to minimize around the protein was performed, with the protein harmonically constrained for 1 ns, with 1 fs timestep and constant pressure (NPT ensemble). A second minimization step was applied, where the system was free to minimize with a harmonic constraint on the protein backbone to ensure stable quaternary structure for 1 ns – lipids relax and compress during minimization steps with minimized dimensions equal to the water box (14.7 x 14.7). The entire system was then released from restraints and subjected to all-atom equilibration runs employing Langevin thermostat, with a constant temperature of 310 K and constant pressure of 1 atm (NPT ensemble), with 2 fs time-steps and allowed to proceed for 30 ns. Periodic boundary conditions were used to allow for the smoothed particle mesh Ewald (PME) calculation of electrostatics. Finally, two independent 100 ns production runs were seeded with randomly initialized velocities from the initial equilibration simulation – providing 200ns of production for each system. Root mean squared deviations (r.m.s.d.) and root mean squared fluctuations (r.m.s.f.) were calculated using VMD, and r.m.s.f. values were displayed to the protein structure using UCSF Chimera (Extended Data Fig. 7b-d). All systems approached a steady r.m.s.d. within 30 ns of the equilibration phase (Extended Data Fig. 7b), and r.m.s.f. values appeared well-behaved over the production periods, including regions corresponding to the NTH domain[16] (Extended Data Fig. 7c,d). The only significant fluctuations (i.e., > 2.5

Å) occurred at the TM2, TM3 and TM4 cytoplasmic termini, which is expected as these regions form the boundary to the intrinsically disordered ICL and CTD regions of the protein (not modeled). All systems maintained an electro-chemical seal to extracellular sodium ions (Na^+) around the ECD docking domains during MD simulation.

Calculation of density maps from MD for water, lipids and ions. The Volmap plugin in VMD was used for the calculation of volumetric density maps, by replacing each atom with a normalized gaussian distribution, whose standard deviation is equal to the radius of the atom. All of the gaussians are summed and distributed on a grid for each frame of the simulation. The grids were re-sampled to a final voxel resolution of 0.649 Å to match the pixel size used in the CryoEM reconstruction. Water, ion, and lipid maps were calculated from each of two 100 ns production runs, and subsequently averaged and symmetrized (D6-symmetry) with the relion_image_handler tool in RELION-3.0[66]. Lipid and water density maps produced from Cx46 and Cx50 MD simulations contained significant overlap to each other and to the CryoEM maps, however, the maps produced from Cx50 MD simulations were of higher quality and were selected for detailed comparative analysis to the CryoEM density maps (Fig. 3 and Extended Data Fig. 8). Ion density maps showed only a few features and did not correspond to densities observed by CryoEM (not shown), and were therefore excluded from further analysis.

MD-based area per lipid (APL) and lipid order parameter (S_{CD}) calculations. Area per lipid (APL) for each membrane, separated by intracellular and extracellular leaflet, were

calculated using the program FATSLiM[86], and used as an indicator of equilibration of the lipid systems (Extended Data Fig. 9a).

Eq.1 $$S_{CD} \equiv -\langle \frac{3\cos^2(\theta_{CD})-1}{2} \rangle$$

The S_{CD} lipid order parameter, as defined by Eq. 1, measures the orientation of the SN1 and SN2 acyl-chains by monitoring the angle that each acyl C-H vector makes with the bilayer normal θ_{CD}. The calculations of S_{CD} were done using the VMD script *calc_op.tcl*[87]. To analyze the distance dependence of S_{CD} time-averaged S_{CD} values were calculated for each lipid (SN1 and SN2 combined values for acyl carbons 4–11), and colored according to this value using UCSF Chimera (Fig. 3c).

MD lipid configuration analysis. Analysis of PC lipid configuration (*i.e.,* acyl-chain positioning) was performed using in-house scripts to assess how phospholipids are organized within the extracellular leaflet of the Cx46 and Cx50 intercellular channels during MD simulation, as compared to the PC configurations classified by CryoEM. This was done by counting the instances of a single DMPC molecule occupying the region bounded by the MD-based lipid-density, contoured at $\sigma_{min} = 8$ (Fig. 3a). The lipid acyl-chain density maps calculated from the Cx50 MD simulations reveal more than 19 resolved rods (i) of density per connexin subunit (12 subunits), and each rod was arbitrarily numbered 1 through 19 (total of 228 acyl-chain positions). A lipid was classified

in a state when both acyl-chains occupied a density, $\text{state} \equiv \text{"i} - \text{j"}\ (\text{where i} \neq \text{j})$. A rod density is considered occupied if at least 5 carbons of a lipid's acyl-chain are within the density, such that $\sigma_{\text{carbon}} \geq \sigma_{\text{min}}$. This classification scheme was applied to every lipid within 15 Å of the protein, over each frame (0.1 ns per frame). To analyze the dynamics of lipids surrounding the protein, the state of each lipid (e.g., "1–2", "1–5", "none", etc.) was monitored and recorded at every frame providing a time series of lipid configurational dynamics in state-space (Fig. 4d).

Statistical analysis. 95% confidence intervals for Cα r.m.s.f. values are reported (n=24) using a two-tailed student t-test. Fourier-Shell Correlation (FSC) was performed using Gold-Standard methods with a 0.143 cut-off criteria[69]. No statistical methods were used to predetermine sample size for the CryoEM dataset. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Figures and movie preparation. Figure panels and movies were created using Chimera[71], ChimeraX[88], VMD[79], and Blender[89]. Figures were composited in Adobe Photoshop. Movies were composited in Blender.

Data Availability

CryoEM density maps, including half-maps, pre- and post-processed maps and masks, have been deposited to the Electron Microscopy Data Bank (EMD-22358, EMD-22382,

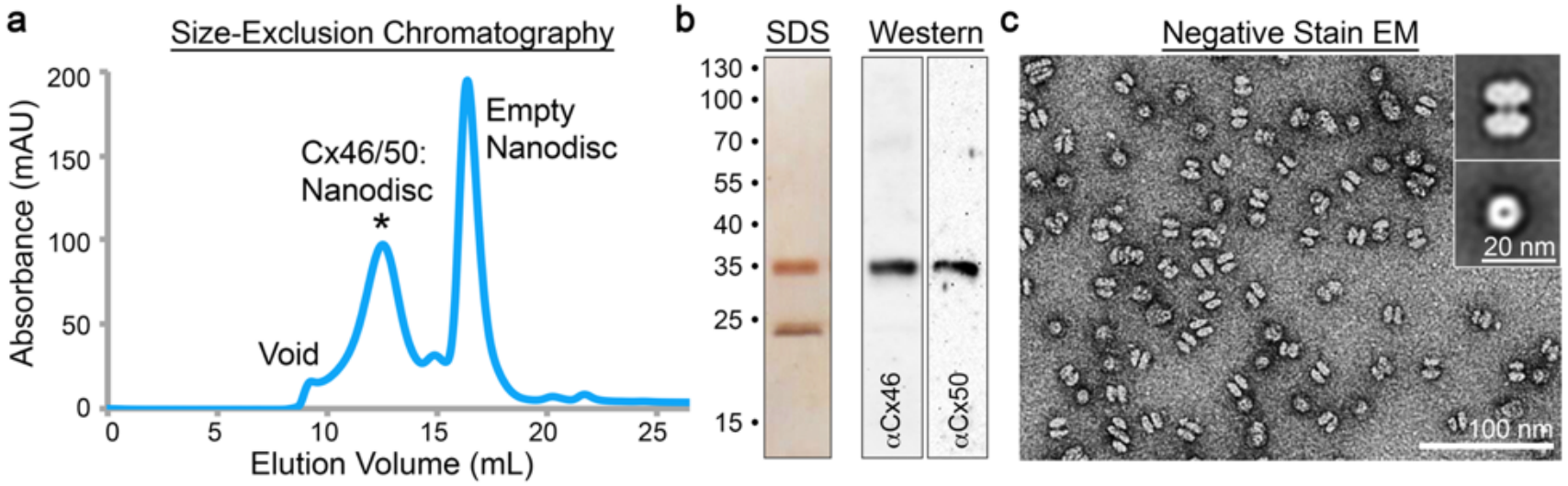
EMD-22390, and EMD-22391). Coordinates for Cx46 and Cx50 atomic models have been deposited to the Protein Data Bank (PDB7JJP and PDB7JKC). The original multi-frame micrographs have been deposited to EMPIAR (EMPIAR-10480). MD trajectory files and MD-based density maps have been deposited to Zenodo (doi: 10.5281/zenodo.3951861).

Code Availability

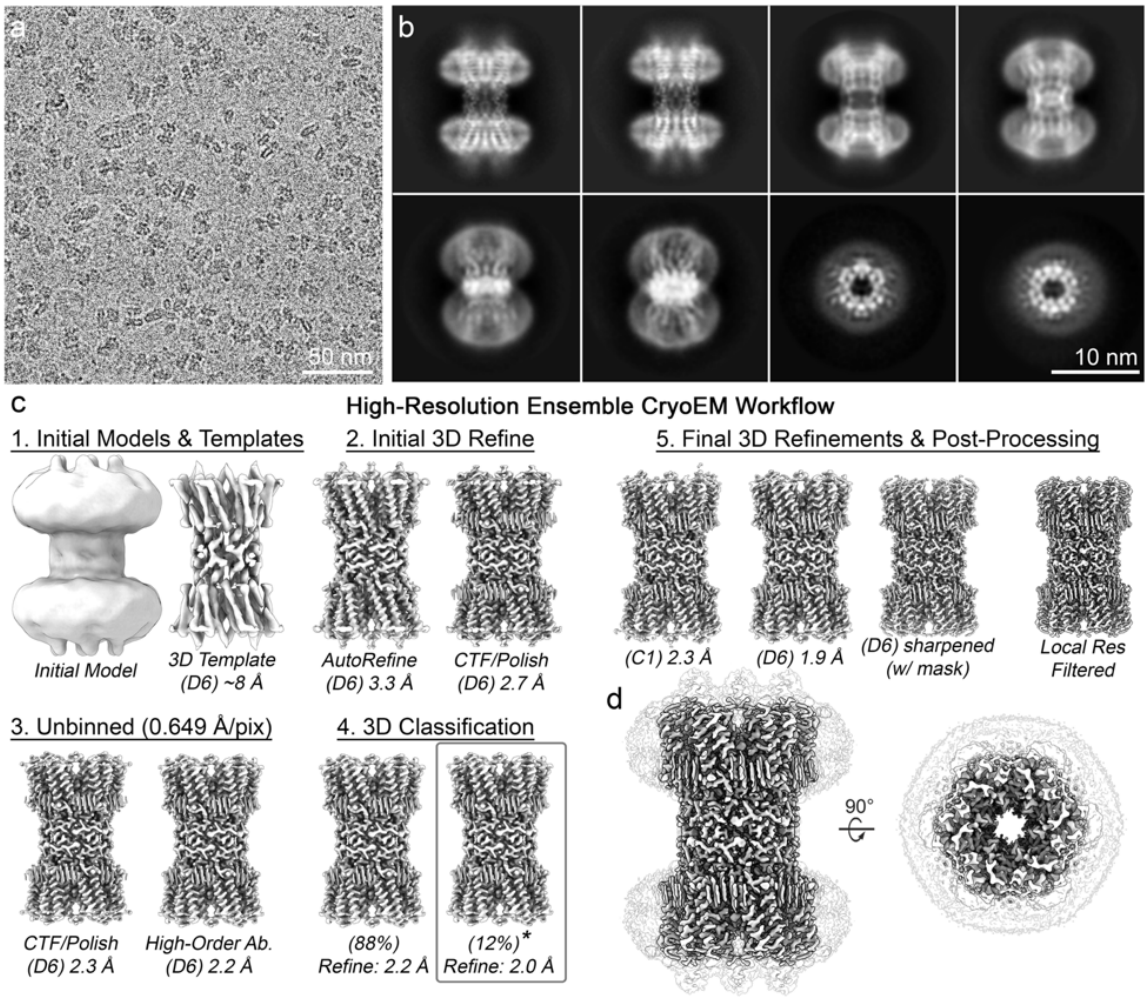
In-house code developed for the analysis of MD trajectories to assess lipid configurational state dynamics have been developed and maintained in our lab GitHub repository (https://github.com/reichow-lab), final codes have been deposited on Zenodo (doi: 10.5281/zenodo.3955131).

	Cx46	Cx50	Cx46 *PC Class 1*	Cx50 *PC Class 1*	Cx46 *PC Class 2*	Cx50 *PC Class 2*	Cx46 *PC Class 3*
Data collection and processing							
Magnification	120,000	120,000	120,000	120,000	120,000	120,000	120,000
Voltage (kV)	300	300	300	300	300	300	300
Electron exposure (e–/Å^2)	52.5	52.5	52.5	52.5	52.5	52.5	52.5
Defocus range (μm)	-1.0 to -2.2	-1.0 to -2.2	-1.0 to -2.2	-1.0 to -2.2	-1.0 to -2.2	-1.0 to -2.2	-1.0 to -2.2
Pixel size (Å)	0.649	0.649	0.649	0.649	0.649	0.649	0.649
Symmetry imposed	D6	D6	D6	D6	D6	D6	D6
Initial particle images (no.)	1,210,797	1,210,797	1,210,797	1,210,797	1,210,797	1,210,797	1,210,797
Final particle images (no.)	26,005	26,005	6,073	6,073	9,188	9,188	6,942
Map resolution (Å)	1.94	1.94	2.52	2.52	2.47	2.47	2.45
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	1.94–3.30	1.94–3.30	2.42–4.56	2.42–4.56	2.33–4.14	2.33–4.14	2.47–4.33
Refinement							
Initial model used (PDB code)	6MHQ	6MHY	6MHQ	6MHY	6MHQ	6MHY	6MHQ
Model resolution (Å)	1.96	1.95	2.55	2.52	2.35	2.35	2.63
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Model resolution range (Å)	-	-	-	-	-	-	-
Map sharpening *B* factor (Å^2)	-26.28	-26.28	-40.49	-40.49	-45.42	-45.42	-43.73
Model composition							
Non-hydrogen atoms	21,876	21,804	22,308	22,176	21,624	21,552	21,312
Protein residues	2352	2352	2,328	2,328	2,316	2,316	2,304
Ligands	180	180	168	168	132	132	144
B factors (Å^2)							
Protein	31.41	29.99	51.09	50.04	41.92	47.13	59.04
Ligand	32.54	30.74	55.25	55.19	43.92	47.64	61.77
R.m.s. deviations							
Bond lengths (Å)	0.011	0.010	0.010	0.007	0.009	0.009	0.007
Bond angles (°)	0.861	0.859	0.825	0.821	0.934	1.073	0.816
Validation							
MolProbity score	0.75	1.04	1.03	1.03	1.13	1.22	0.98
Clashscore	0.70	2.51	2.45	2.46	2.90	3.93	2.07
Poor rotamers (%)	0.57	0.56	0.00	0.56	1.17	1.14	0.00
Ramachandran plot							
Favored (%)	97.92	98.44	98.11	98.95	98.41	98.94	98.40
Allowed (%)	2.08	1.56	1.89	1.05	1.59	1.06	1.60
Disallowed (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00

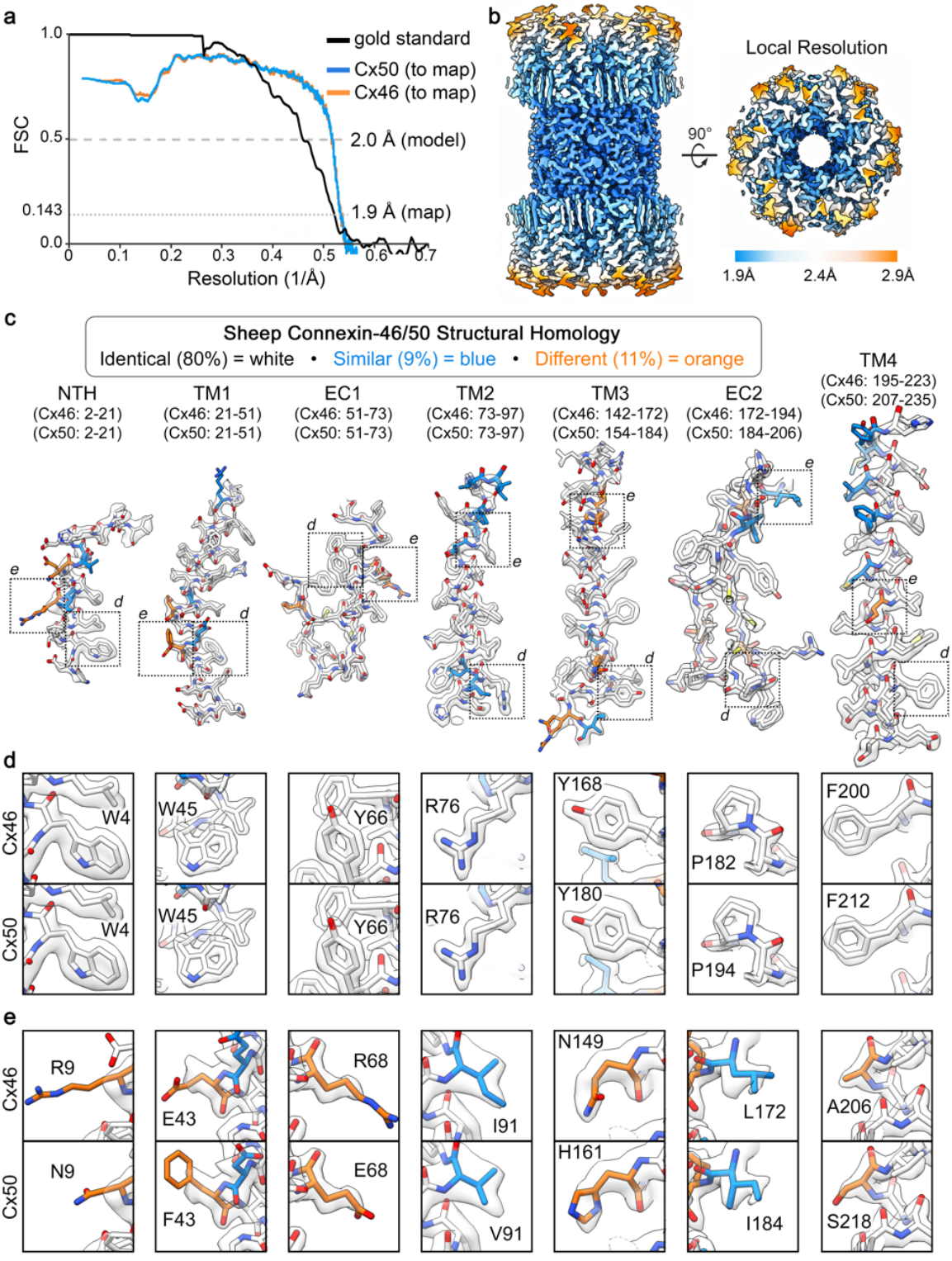
Extended Data Table 1. CryoEM Statistics. Summary of CryoEM data collection, refinement and model validation statistics. The ensemble CryoEM dataset was used to obtain the 1.9 Å resolution reconstruction and atomic models for Cx46 and Cx50, including 396 water molecules and 150 lipid acyl-chains. 3D classification was used to obtain the three PC classes, and associated atomic models for Cx46 and Cx50 (PC Class 1–3). Pre-processed and post-processed maps and associated masks from all datasets have been deposited to the EM databank (EMD-XXXX). The original multi-frame micrographs have been deposited to EMPIAR (EMPIAR-XXXX). Coordinates for Cx50 and Cx46 atomic models have been deposited to the Protein Data Bank (XXXX and XXXX correspond to the ~1.9 Å models, XXXX and XXXX correspond to the ~2.5 Å models from PC Class 1; XXXX and XXXX correspond to PC Class 2; and XXXX and XXXX correspond to PC Class 3).


a
Size-Exclusion Chromatography
Absorbance (mAU)
200
150
100
50
0
0
5
10
15
20
25
Elution Volume (mL)
Void
Cx46/50:
Nanodisc
*
Empty
Nanodisc
b
SDS
Western
130
100
70
55
40
35
25
15
αCx46
αCx50
c
Negative Stain EM
20 nm
100 nm

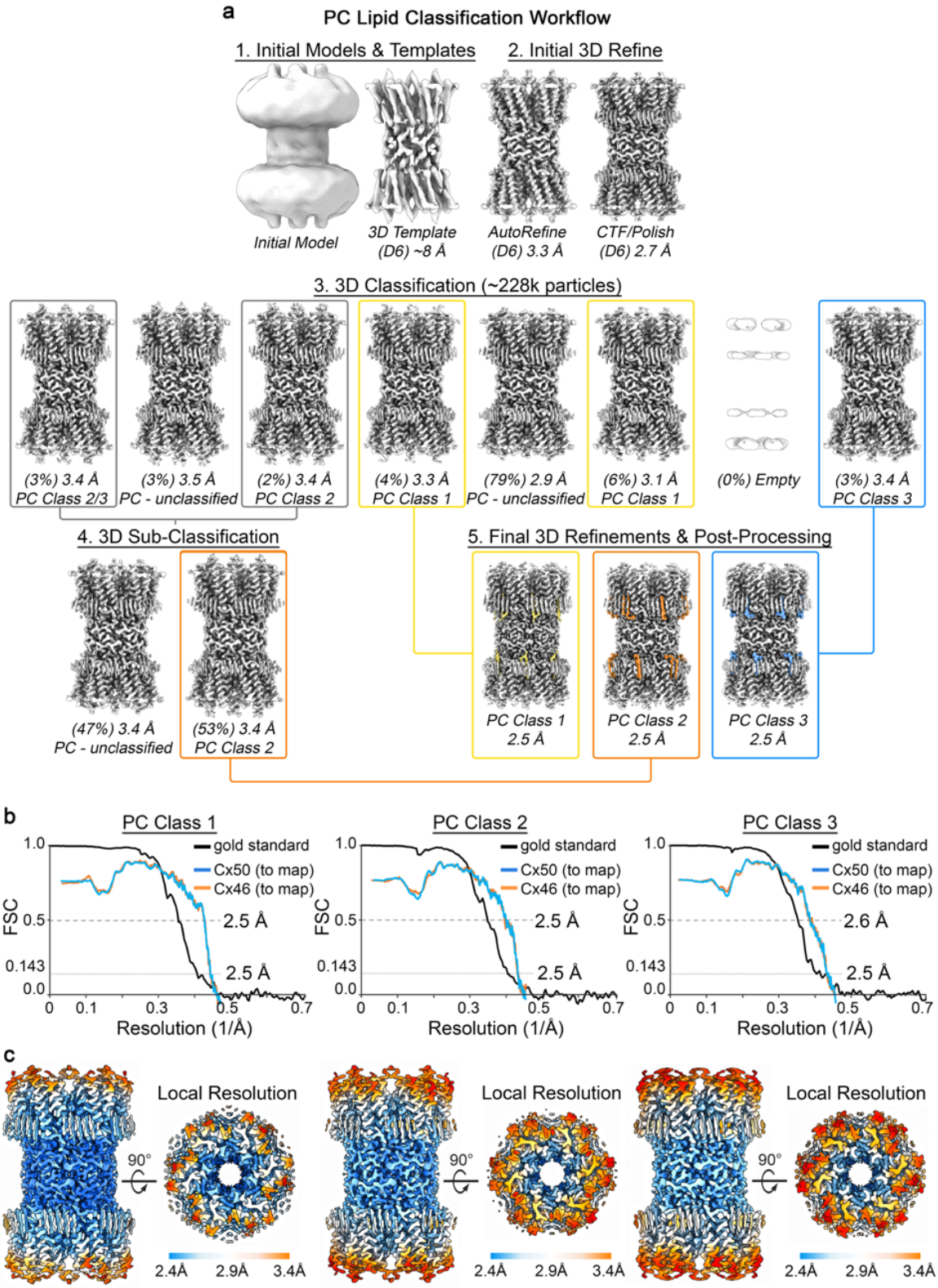
Extended Data Figure 1. Cx46/50 reconstitution into MSP1E1/DMPC lipid-nanodiscs and negative stain EM. a) Size-exclusion chromatography (SEC) trace monitored by UV absorbance at 280 nm. Peaks corresponding to Cx46/50 reconstituted into MSP1E1/DMPC nanodiscs (□), empty nanodisc and void are indicated. **b)** SDS-PAGE (*left*) and western blot (*right*) of peak SEC fraction (labeled □), with molecular weight markers indicated. MSP1E1 migrates as a ~24 kDa band (predicted MW ~27.5 kDa). Cx46 and Cx50 migrate together at ~38 kDa band, as expected from c-terminal truncation from core lens fiber cells[16], and confirmed by western blot (*right*). **c)** Electron micrograph of negatively stained particles from SEC fraction (labeled □), with scale bar = 100 nm. *Inset*, shows representative 2D class averages of sideview (*top*) and top view (*bottom*), with scale bar = 20 nm. Particles display dumbbell-like structures corresponding to Cx46/50 gap junctions intercellular channels[16,90], embedded into a pair of ~10-11 nm wide nanodiscs (MSP1E1 nanodiscs have a predicted diameter of ~10.5 nm[17]).


a
50 nm
b
10 nm
c
High-Resolution Ensemble CryoEM Workflow
1. Initial Models & Templates
Initial Model
3D Template
(D6) ~8 Å
2. Initial 3D Refine
AutoRefine
(D6) 3.3 Å
CTF/Polish
(D6) 2.7 Å
3. Unbinned (0.649 Å/pix)
CTF/Polish
(D6) 2.3 Å
High-Order Ab.
(D6) 2.2 Å
4. 3D Classification
(88%)
Refine: 2.2 Å
(12%)*
Refine: 2.0 Å
5. Final 3D Refinements & Post-Processing
(C1) 2.3 Å
(D6) 1.9 Å
(D6) sharpened
(w/ mask)
Local Res
Filtered
d
90°

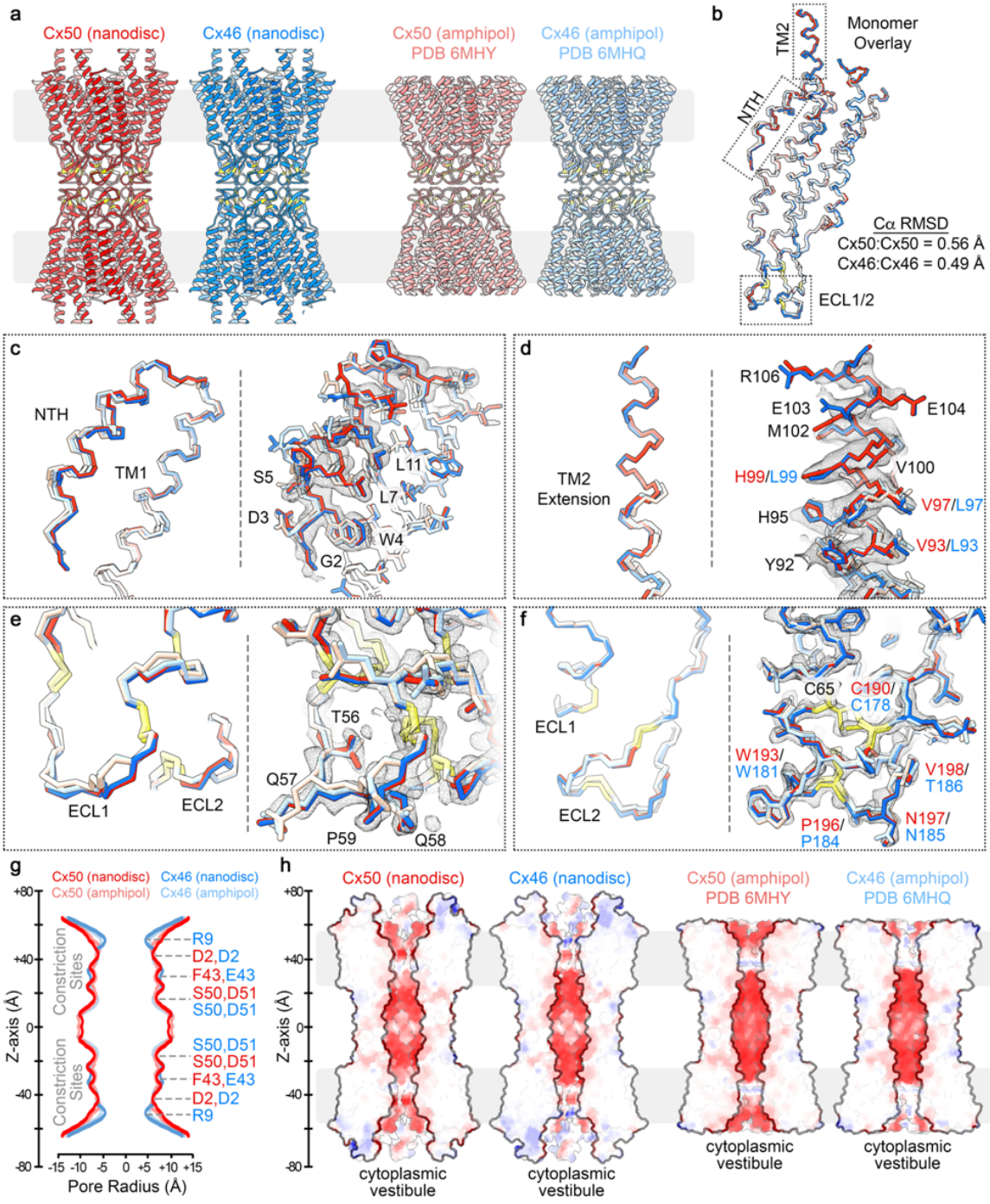
Extended Data Figure 2. CryoEM image processing workflow for 1.9 Å ensemble reconstruction of Cx46/50 in DMPC lipid nanodiscs. a) Representative CryoEM micrograph (dataset of 2088 movies) recorded on a Falcon III detector, with physical pixel size = 0.649 $Å^2$ and total dose of ~60 e^- per $Å^2$. Scale bar = 50 nm. **b)** Representative 2D class averages. Scale bar = 10 nm**. c)** Image processing and 3D reconstruction workflow carried out in Relion[66,68], with representative maps at different stages of the image processing pipeline. Step 1) *De-novo* model generated in Relion (*left*) and initial 3D AutoRefinement with D6-symmetry (~8 Å resolution, 3.9 Å pixel size) (*right*), which was then filtered to 20 Å and used for 3D template auto-picking in Relion (resulting in ~1.2M particle picks, which were culled to ~228k "good" particles following multiple rounds of 2D classification and de-duplication). Step 2) Resulting 3D AutoRefine with D6-symmetry without masking (3.2 Å resolution, 0.97 Å pixel size) (*left*), and resulting map following per particle CTF-refinement and polishing in Relion (2.7 Å) (*right*). Step 3) Particles were unbinned (pixel size 0.649 Å/pix, box size = 400 pix) and refined with per-particle CTF-correction and polishing (2.3 Å) (*left*), and further refinement of high-order aberration parameters in Relion v3.1-beta[68] (2.2 Å) (*right*). Step 4) Particles were de-duplicated, resulting in a set of ~221k particles, and subjected to 3D classification (two classes). Class 1 contained 88% of the particles and was further refined to 2.2 Å resolution (*left*). Class 2 contained 12% of the particles and was further refined to 2.0 Å resolution (*right, asterisk*). It is noted that the major differences between these two classes appears to be the distribution of particle defocus values, were the higher resolution class contains particles with lower defocus range (mean defocus = 1.20 ± 0.29 µm (s.d.)). Step 5) Particles belonging to Class 2 (~26k particles), were then subjected to multiple rounds of 3D Auto-refinement followed by per-particle CTF, aberration-correction and polishing, using successively larger box-sizes until no further improvement, resulting in a final reconstruction at 2.3 Å resolution (C1 symmetry) (*left*) and 1.9 Å resolution (D6 symmetry) (*center, left*). There were no clear conformational differences between the C1 and D6-symmetrized reconstructions. The D6-symmeterized map was then subjected to post-processing (b-factor sharpening) (*center, right*) and local-resolution filtering in Relion (*right*) for downstream analysis. **d)** Final reconstruction of Cx46/50 following local resolution filtering, used for atomic-modeling. Transparent silhouette displays the unmasked map at low-contour to illustrate the dimensions of the lipid nanodisc densities.


a
1.0
0.5
0.143
0.0
FSC
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Resolution (1/Å)
gold standard
Cx50 (to map)
Cx46 (to map)
2.0 Å (model)
1.9 Å (map)
b
Local Resolution
90°
1.9Å
2.4Å
2.9Å
c
Sheep Connexin-46/50 Structural Homology
Identical (80%) = white • Similar (9%) = blue • Different (11%) = orange
NTH (Cx46: 2-21) (Cx50: 2-21)
TM1 (Cx46: 21-51) (Cx50: 21-51)
EC1 (Cx46: 51-73) (Cx50: 51-73)
TM2 (Cx46: 73-97) (Cx50: 73-97)
TM3 (Cx46: 142-172) (Cx50: 154-184)
EC2 (Cx46: 172-194) (Cx50: 184-206)
TM4 (Cx46: 195-223) (Cx50: 207-235)
d
Cx46
Cx50
W4
W45
Y66
R76
Y168
P182
F200
Y180
P194
F212
e
R9
N9
E43
F43
R68
E68
I91
V91
N149
H161
L172
I184
A206
S218

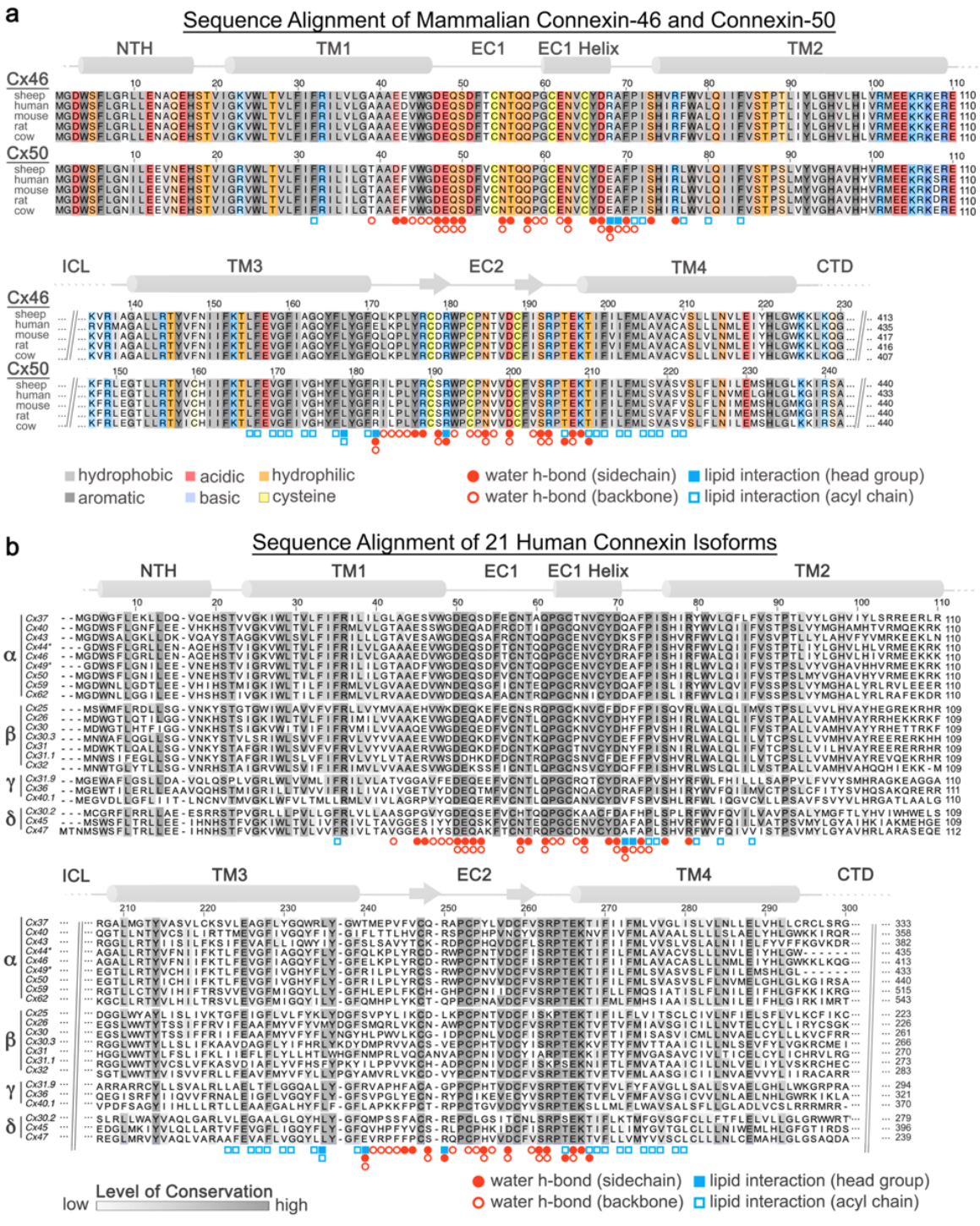
Extended Data Figure 3. Global and local resolution assessment of the 1.9 Å ensemble reconstruction of Cx46/50 in DMPC lipid nanodiscs. a) Fourier Shell Correlation (FSC) analysis obtained from the ensemble CryoEM map of Cx46/50 in DMPC lipid nanodiscs. Gold-standard FSC (black) of the final refined CryoEM map indicates a global resolution of 1.9 Å (0.143 cut-off). FSC curves comparing atomic models for Cx50 (blue) and Cx46 (orange) fit to the final CryoEM map display overall correlation at 2.0 Å (0.5 cut-off). **b)** Local resolution analysis of the final CryoEM map using Relion[68], displayed by colored surface (1.9 – 2.4 Å = blue – white; 2.4 – 2.9 Å = white – orange). **c)** Segmented CryoEM map with regions of the atomic models for sheep Connexin-46 (Cx46) and Connexin-50 (Cx50) fit to the local-resolution filtered map. Residue numbering for Cx46 and Cx50 is displayed above the corresponding segments for the n-terminal helix domain (NTH) domain, the transmembrane domains 1-4 (TM1-4) and extracellular domains 1-2 (EC1-2), Residues are colored according to the pair-wise sequence homology between sheep Cx46 and Cx50, as being identical (white, 80%), similar (blue, 9%) and different (orange, 11%), with all heteroatoms colored by standard scheme (oxygen – red; nitrogen – blue, sulfur – yellow). **d, e)** Windows show zoom-views corresponding to boxed regions of the segmented maps. **d)** Displays fits over representative regions where both Cx46/50 contain identical amino acids, where the high-resolution features are well-resolved. **e)** Displays fits over representative regions where the sequence of Cx46 and Cx50 differ, and where sidechain density is weaker and/or consistent with heterogeneity. This is presumably due to the heteromeric/heterotypic mixture of these isoforms[16,77,78] and the imposed averaging of two different sidechains in these areas, and/or to relative flexibility at these sites, as many of these same residues correspond to solvent/lipid exposed sidechains (e.g., R9/N9; E43/F43, R68/E68, I91/V91 and A206/S218).


a
PC Lipid Classification Workflow
1. Initial Models & Templates
2. Initial 3D Refine
Initial Model
3D Template (D6) ~8 Å
AutoRefine (D6) 3.3 Å
CTF/Polish (D6) 2.7 Å
3. 3D Classification (~228k particles)
(3%) 3.4 Å PC Class 2/3
(3%) 3.5 Å PC - unclassified
(2%) 3.4 Å PC Class 2
(4%) 3.3 Å PC Class 1
(79%) 2.9 Å PC - unclassified
(6%) 3.1 Å PC Class 1
(0%) Empty
(3%) 3.4 Å PC Class 3
4. 3D Sub-Classification
(47%) 3.4 Å PC - unclassified
(53%) 3.4 Å PC Class 2
5. Final 3D Refinements & Post-Processing
PC Class 1 2.5 Å
PC Class 2 2.5 Å
PC Class 3 2.5 Å
b
PC Class 1
PC Class 2
PC Class 3
gold standard
Cx50 (to map)
Cx46 (to map)
FSC
Resolution (1/Å)
2.5 Å
2.6 Å
c
Local Resolution
90°
2.4Å
2.9Å
3.4Å

Extended Data Figure 4. Image processing and resolution assessment for 3D lipid-classification work-flow. a) Image process and 3D reconstruction workflow carried out in Relion for the analysis of PC lipid configuration/conformational heterogeneity, with representative maps at different stages of the image processing pipeline. Steps 1 and 2) are the same as described in Extended Data Fig. 2, which resulted in a 2.7 Å reconstruction from a dataset of ~228k "good" particles (*right*). Step 3) These particles were unbinned and re-extracted (0.649 Å/pix, 400 pixel box), and subjected to 3D classification (eight classes) without image alignment. Two of the eight classes yielded maps in which the lipid configuration was unambiguously resolved: assigned as PC Class 1 (yellow box), containing 9,190 particles (~4% of the data) and PC Class 3 (blue *box*), containing 6,944 particles (~3% of the data). Overlapping configurations were resolved in two of the other 3D classes (grey boxes). Step 4) The particles from these overlapping classes (grey boxes) were combined and subjected to a second round of 3D classification with two classes. This yielded one in which the lipid configuration was unambiguously resolved: assigned PC Class 2 (orange box), containing 6,075 particles (~3% of the data). Step 5) Particles assigned to PC Class 1 (*left*), PC Class 2 (*center*) and PC Class 3 (*right*) were separately subjected to a final round of 3D refinement and per-particle polishing, with D6 symmetry applied, resulting in final reconstructions ~2.5 Å resolution (Gold-Standard, 0.143 cut-off). **b)** Fourier Shell Correlation (FSC) analysis obtained for PC Class 1 (*left*), PC Class 2 (*center*) and PC Class 3 (*right*). Gold-standard FSC (black) of the final refined, masked and post-processed CryoEM map indicates a global resolution of 2.5 Å (0.143 cut-off). FSC curves comparing atomic models for Cx50 (blue) and Cx46 (orange) fit to the final CryoEM maps display overall correlation at 2.5–2.6 Å (0.5 cut-off). **c)** Local resolution analysis of the final CryoEM maps for PC Class 1 (*left*), PC Class 2 (*center*) and PC Class 3 (*right*) using Relion, displayed by colored surface (2.4 – 2.9 Å = blue – white; 2.9 – 3.4 Å = white – orange).


a
Cx50 (nanodisc)
Cx46 (nanodisc)
Cx50 (amphipol)
PDB 6MHY
Cx46 (amphipol)
PDB 6MHQ
b
TM2
Monomer
Overlay
NTH
Cα RMSD
Cx50:Cx50 = 0.56 Å
Cx46:Cx46 = 0.49 Å
ECL1/2
c
NTH
TM1
S5
L11
L7
D3
W4
G2
d
TM2
Extension
R106
E103
E104
M102
V100
H99/L99
V97/L97
H95
V93/L93
Y92
e
ECL1
ECL2
T56
Q57
P59
Q58
f
ECL1
ECL2
C65
C190/
C178
W193/
W181
V198/
T186
P196/
P184
N197/
N185
g
Cx50 (nanodisc)
Cx46 (nanodisc)
Cx50 (amphipol)
Cx46 (amphipol)
Constriction
Sites
Constriction
Sites
R9
D2,D2
F43,E43
S50,D51
S50,D51
S50,D51
S50,D51
F43,E43
D2,D2
R9
Z-axis (Å)
+80
+40
0
-40
-80
-15 -10 -5 0 +5 +10 +15
Pore Radius (Å)
h
Cx50 (nanodisc)
Cx46 (nanodisc)
Cx50 (amphipol)
PDB 6MHY
Cx46 (amphipol)
PDB 6MHQ
Z-axis (Å)
+80
+40
0
-40
-80
cytoplasmic
vestibule
cytoplasmic
vestibule
cytoplasmic
vestibule
cytoplasmic
vestibule

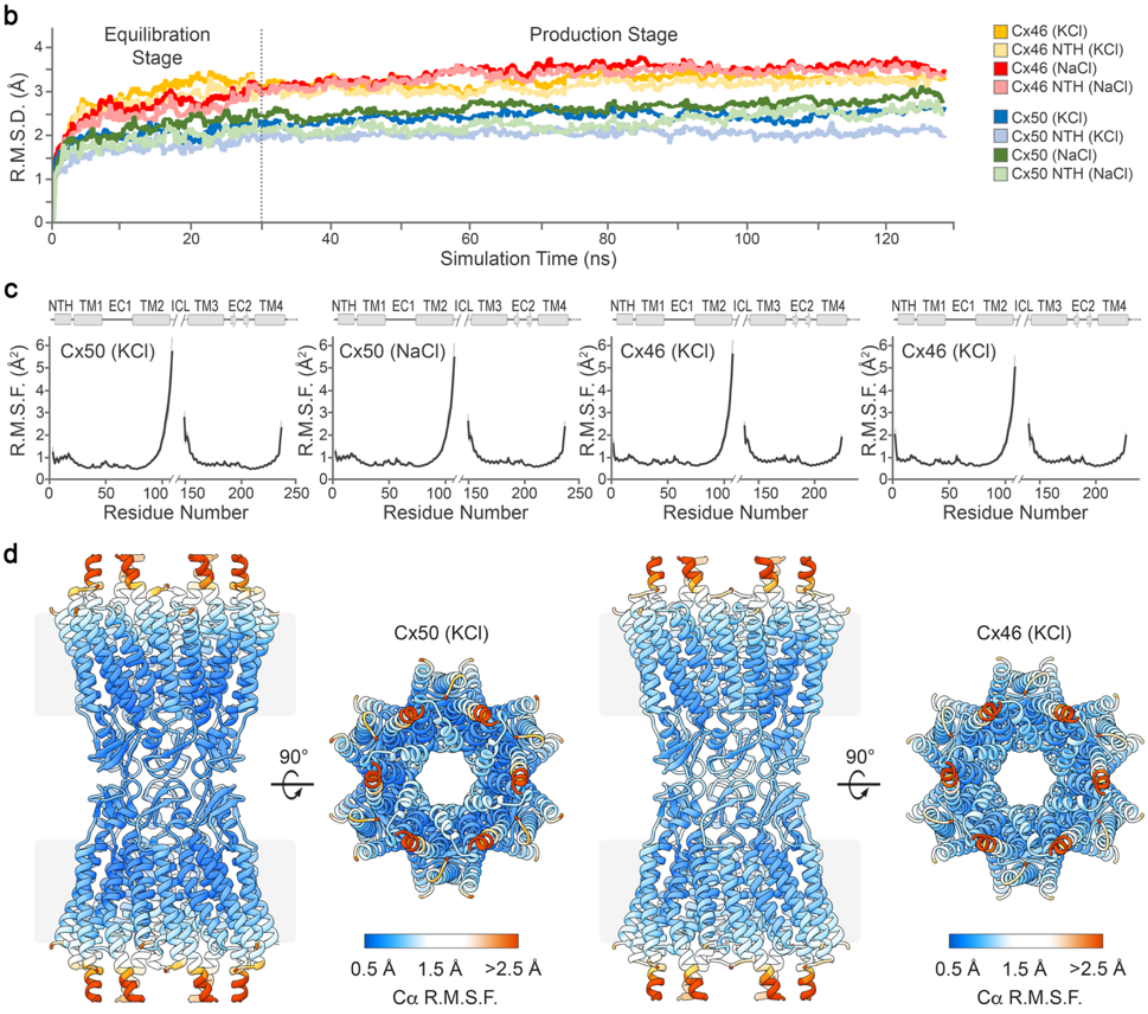
Extended Data Figure 5. Comparison of Cx46/50 structures determined in amphipol and lipid-nanodiscs. a) Ribbon structures of Cx50 (red) and Cx46 (blue) determined by CryoEM in lipid-nanodisc (left) and as previously determined in amphipol (right) with Cx50 (light red, PDB 6MHY) and Cx46 (light blue, PDB 6MHQ)[16]. Regions of lipid bilayer are indicated by light grey box. Conserved cysteine positions within the EC1/2 domains, involved in disulfide formation, are indicated in yellow. **b)** C traces over-laid for these four models, corresponding to a single subunit following super-positing (colored as in panel a). C r.m.s.d. following super-positioning is indicated for Cx50 (nanodisc) vs. Cx50 (amphipol) = 0.56 Å, and Cx46 (nanodisc) vs. Cx46 (amphipol) = 0.49 Å. **c–f)** Shows zoom views corresponding to the boxed regions in panel b. For each panel, (left) shows C trace and (right) shows all atom fit into the 1.9 Å CryoEM density map obtained from the nanodisc embedded structure, to show regions of improved fit to the experimental density map. Highlighted residues are indicated, and labels colored according to identity between the Cx50 and Cx46 isoforms (black – identical, red – Cx50 and blue – Cx46). **g)** Pore radius determined using HOLE[91], for experimental structures of Cx50-nanodisc (red), Cx46-nanodisc (blue), Cx50-amphipol (light red) and Cx46-amphipol (light blue). Locations corresponding to constriction sites are indicated, and residues contributing to these sites of constriction for both isoforms are labeled (Cx50 – red; Cx46 – blue). **h)** Cut-away surface representation of Cx50-nanodisc *(left)*, Cx46-nanodisc *(left center)* and Cx50-amphipol *(right center)* and Cx46-amphipol *(right)*, colored by coulombic potential (negative – red, neutral – white and positive – blue). This comparison illustrates the electrostatic environment of the permeation pathways and the extension of the intracellular vestibule that is resolved in the Cx46/50-nanodisc models, as compared to the previously described Cx46/50-amphipol models.


a
Sequence Alignment of Mammalian Connexin-46 and Connexin-50
NTH
TM1
EC1
EC1 Helix
TM2
Cx46
Cx50
sheep
human
mouse
rat
cow
ICL
TM3
EC2
TM4
CTD
hydrophobic
acidic
hydrophilic
aromatic
basic
cysteine
water h-bond (sidechain)
water h-bond (backbone)
lipid interaction (head group)
lipid interaction (acyl chain)
b
Sequence Alignment of 21 Human Connexin Isoforms
α
β
γ
δ
low
Level of Conservation
high

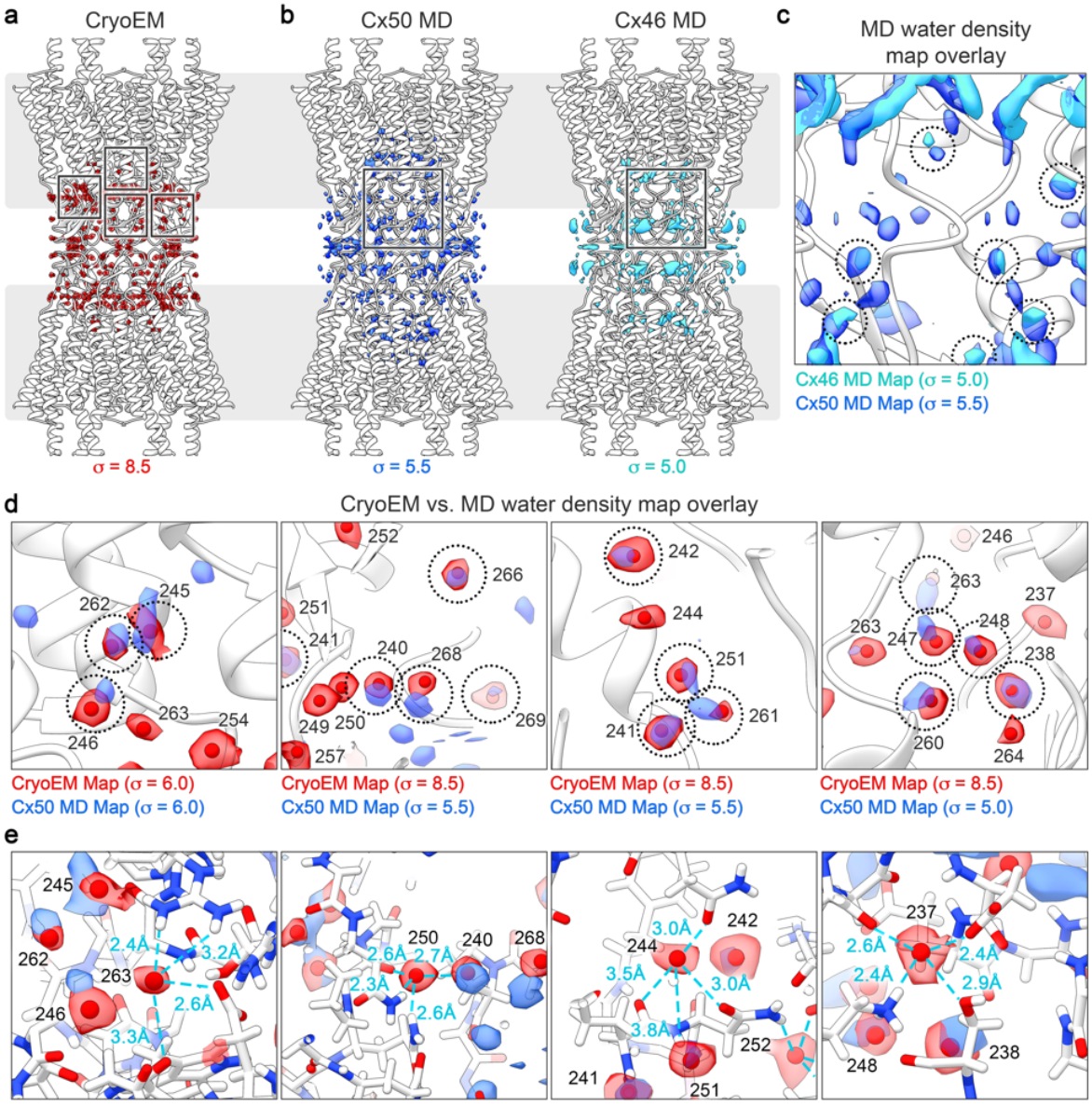
Extended Data Figure 6. Sequence alignment with annotated lipid and water binding sites. a) Multiple sequence alignment of mammalian Cx46 and Cx50 isoforms with residues contributing to lipid and/or water binding sites annotated (filled circle – water h-bonding with amino acid sidechain; open circle – water h-bonding with amino acid backbone) and (filled square – interaction involving lipid headgroup; open square – interaction involving lipid acyl chain). Primary sequence coloring corresponds to amino acid type (grey – hydrophobic; dark grey – aromatic; red – acidic; blue – basic; orange – hydrophilic; yellow – cysteine). Regions of sequence homology are indicated by the level of shading. Secondary structure and domain labels are indicated for the n-terminal helix (NTH), transmembrane helices (TM1-4) and extracellular domains (EC1-2). Regions lacking defined structure and of poor sequence homology within the intracellular loop (ICL) and c-terminal domain (CTD) have been omitted for clarity. Sheep and human Cx46 and Cx50 orthologs contain ~95% sequence identity (~98% similarity) over the structured regions of the protein. Numbering corresponds to the amino acid sequence of sheep Cx44 and Cx49 used in the main text. **b)** Multiple sequence alignment of 20 human connexin isoforms, with sheep Cx44 (Cx46 homolog) and Cx49 (Cx50 homolog) included for comparison. Isoforms are categorized by connexin family □, □, □ and □. The orphan Cx23 was excluded from analysis. Regions of sequence homology are indicated by the level grey of shading. Annotations for lipid and water binding sites and secondary structural elements/domains are indicated as in panel a.

a

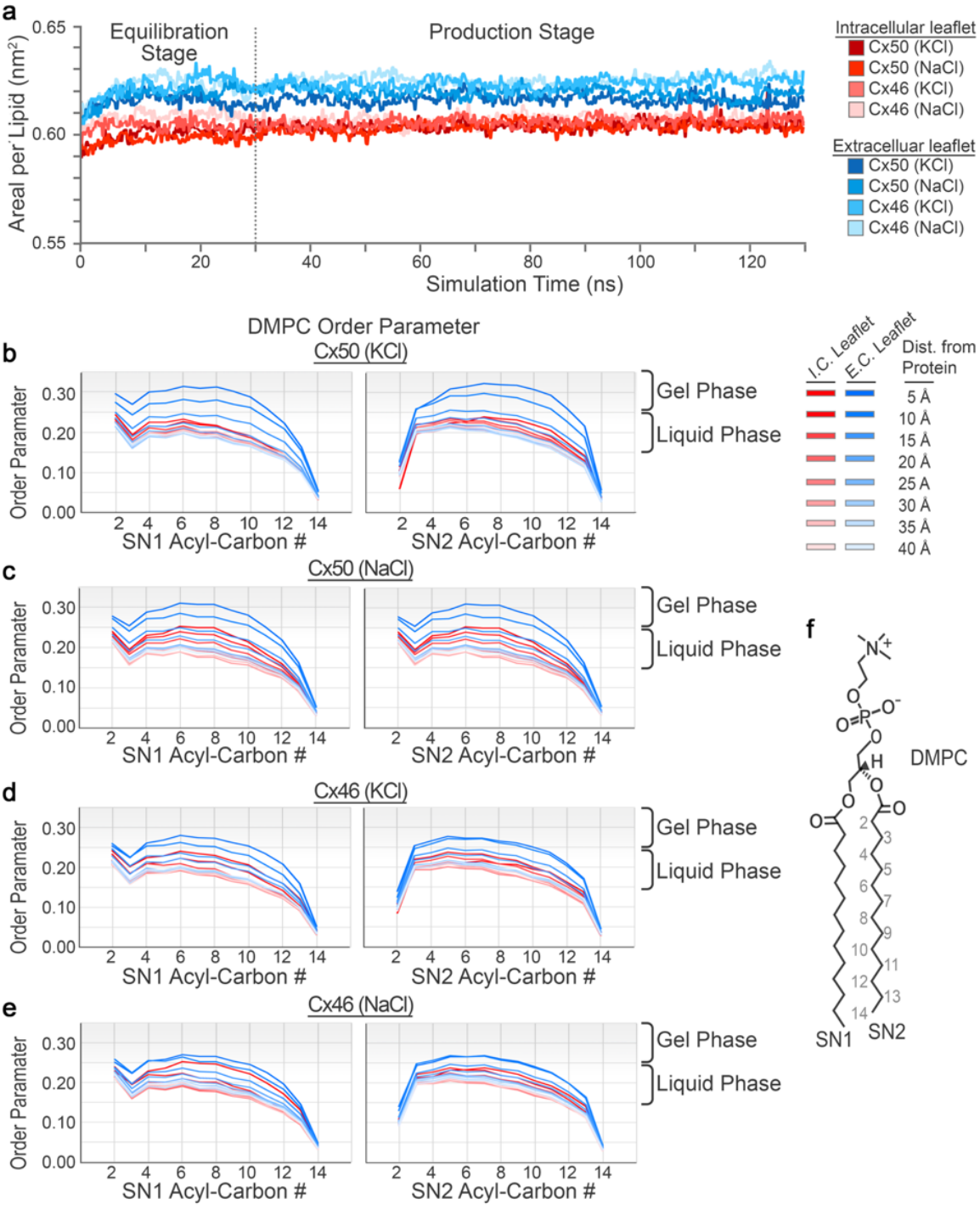
		Connexin-50 (KCl)	Connexin-50 (NaCl)	Connexin-46 (KCl)	Connexin-46 (NaCl)
Total Atoms		356,838	356,838	357,545	357,545
	Solvent	186,150	186,150	186,663	186,663
	Lipids	131,806	131,806	131,806	131,806
	Protein	38,484	38,484	38,712	38,712
	Ions	398	398	364	364
Modelled Residues		2-109 ; 149-236	2-109 ; 149-236	2-109 ; 137-224	2-109 ; 137-224
Modifications to model		n-term acetylation	n-term acetylation	n-term acetylation	n-term acetylation
Simulation Conditions					
Simulation Box (Å)		147 x 147 x 174	147 x 147 x 174	147 x 147 x 174	147 x 147 x 174
Pressure (atm)		1	1	1	1
Temperature (K)		310	310	310	310
Time Step (fs)		2	2	2	2
Equilibration Time (ns)		30	30	30	30
Production Time (ns)		100	100	100	100
Production Replicas		2	2	2	2
Total Production Time (ns)		200	200	200	200



Extended Data Figure 7. Molecular dynamics setup and validation. a) Summary of molecular dynamics (MD) simulation setup and conditions. Each simulation was setup similarly, using and explicit solvent model containing either KCl (Cx50KCl; Cx46 KCl) or NaCl (Cx50 NaCl; Cx46 NaCl) in the cytoplasmic space, to match either cellular or *in vitro* conditions used for CryoEM studies, respectively. All simulations were conducted with NaCl in the extracellular space and using DMPC as the lipid system. Following minimization, all systems were equilibrated for 30 ns at 37° C, and multiple replicates (N=2) of production (100 ns each) were acquired for analysis at 37° C. **b)** Cα root mean squared deviation (r.m.s.d.) analysis of equilibrium (0 – 30 ns) and production phases (30–130 ns) of the MD simulations, calculated with respect to the experimental starting structures, where Cx50 KCl (blue traces); Cx50 NaCl (green traces); Cx46 KCl (orange traces); Cx46 NaCl (red traces). Separate analysis for the n-terminal helix (NTH) domains are shown in lighter shades. **c)** Plot of average Cα root mean squared fluctuation (*r.m.s.f.*) during the production phase of the molecular dynamics (MD) simulations for Cx50 KCl (*left*), Cx50 NaCl *(left center)* Cx46 KCl *(right center)* and Cx46 NaCl *(right)*. Averages are determined for the 12 subunits composing the intercellular channel, analyzed for both independent productions. Error bars (light grey shading) represent 95% confidence intervals (n = 24). Secondary structure and domain labels are indicated for the n-terminal helix (NTH), transmembrane helices (TM1-4), extracellular domains (EC1-2) and intracellular loop (ICL; not modeled). **d**) Representative r.m.s.f. values mapped to the experimental starting structures of Cx50 KCl *(left)* and Cx46 KCl *(right)*. Colors correspond to r.m.s.f. amplitudes: < 0.5 Å (blue) – 1.5 Å (white) – 2.5 Å (red).


a
CryoEM
σ = 8.5
b
Cx50 MD
Cx46 MD
σ = 5.5
σ = 5.0
c
MD water density map overlay
Cx46 MD Map (σ = 5.0)
Cx50 MD Map (σ = 5.5)
d
CryoEM vs. MD water density map overlay
CryoEM Map (σ = 6.0)
Cx50 MD Map (σ = 6.0)
CryoEM Map (σ = 8.5)
Cx50 MD Map (σ = 5.5)
CryoEM Map (σ = 8.5)
Cx50 MD Map (σ = 5.5)
CryoEM Map (σ = 8.5)
Cx50 MD Map (σ = 5.0)
e
2.4Å
3.2Å
2.6Å
3.3Å
2.6Å
2.7Å
2.3Å
2.6Å
3.0Å
3.5Å
3.0Å
3.8Å
2.6Å
2.4Å
2.4Å
2.9Å

Extended Data Figure 8. Analysis of MD-based water density maps. a) Ribbon structure of Cx46/50 with segmented water densities from the ensemble CryoEM map (red density, threshold = 8.5 □). **b)** Ribbon structures of Cx50/46 with overlaid time-averaged and symmetrized water density maps calculated from MD-simulation for Cx50 (*left;* blue density, threshold = 5.5 □) and Cx46 (*right;* cyan density, threshold = 5.0 □). **c)** Zoom view, corresponding to boxed regions in panels b and c, showing overlaid MD-based water densities. Representative regions of overlapping density are circled. **d)** Zoom views of boxed regions in panel a, showing representative regions of CryoEM water densities (red) overlaid the Cx50 MD-based water density map (blue). Identities of modeled waters are indicated (using Cx50 numbering). Representative regions of overlapping density are circled. 76% of waters modeled into the CryoEM map show corresponding density in the MD-based water maps. Density map threshold values (□) used for visualization in each panel are indicated. **e**) Alternative views to those presented in panel d, depicting some of the water densities observed by CryoEM that were not also resolved in the MD-based density maps. To validate these sites, the local coordination to sidechain and/or backbone hydrogen bond partners were assessed (donor/acceptor distances < 4 Å, blue dotted lines).


a
Areal per Lipid (nm²)
0.65
0.60
0.55
Equilibration Stage
Production Stage
0
20
40
60
80
100
120
Simulation Time (ns)
Intracellular leaflet
Cx50 (KCl)
Cx50 (NaCl)
Cx46 (KCl)
Cx46 (NaCl)
Extracelluar leaflet
Cx50 (KCl)
Cx50 (NaCl)
Cx46 (KCl)
Cx46 (NaCl)
DMPC Order Parameter
b
Cx50 (KCl)
c
Cx50 (NaCl)
d
Cx46 (KCl)
e
Cx46 (NaCl)
Order Paramater
0.30
0.20
0.10
0.00
2 4 6 8 10 12 14
SN1 Acyl-Carbon #
SN2 Acyl-Carbon #
Gel Phase
Liquid Phase
I.C. Leaflet
E.C. Leaflet
Dist. from Protein
5 Å
10 Å
15 Å
20 Å
25 Å
30 Å
35 Å
40 Å
f
DMPC
SN1
SN2

Extended Data Figure 9. Analysis of MD-based lipid dynamics. a) Lipid equilibration was monitored by analyzing the averaged area per lipid (nm^2) over the duration of MD-simulation. Traces correspond to lipids from the extracellular leaflets (blue shades) and intracellular leaflets (red shades) for each system (Cx50 KCl, Cx50 NaCl, Cx46 KCl and Cx46 NaCl) are displayed. **b-e)** Averaged lipid order parameters calculated for the SN1 *(left)* and SN2 *(right)* acyl-chain C-H bonds (S_{CD}) for each system (*panel b*, Cx50 KCl; *panel c*, Cx50 NaCl; *panel d,* Cx46 KCl; and *panel e,* Cx46 NaCl). Traces correspond to lipids from the intracellular leaflets (red, *I.C. leaflet*) and extracellular leaflets (blue, *E.C. leaflet*), with dark to light shading showing the radial distance dependence from the surface of the protein (5 Å shells). **f)** Structure of dimyristoyl phosphatidylcholine (DMPC) with SN1 and SN2 acyl-chains labeled.

References

1 Goodenough, D. A. & Paul, D. L. Gap junctions. *Cold Spring Harbor perspectives in biology* **1**, a002576, doi:10.1101/cshperspect.a002576 (2009).

2 Malewicz, B., Kumar, V. V., Johnson, R. G. & Baumann, W. J. Lipids in gap junction assembly and function. *Lipids* **25**, 419-427, doi:10.1007/bf02538083 (1990).

3 Cascio, M. Connexins and their environment: effects of lipids composition on ion channels. *Biochim Biophys Acta* **1711**, 142-153, doi:10.1016/j.bbamem.2004.12.001 (2005).

4 Puebla, C., Retamal, M. A., Acuna, R. & Saez, J. C. Regulation of Connexin-Based Channels by Fatty Acids. *Front Physiol* **8**, 11, doi:10.3389/fphys.2017.00011 (2017).

5 Sosinsky, G. E. & Nicholson, B. J. Structural organization of gap junction channels. *Biochim Biophys Acta* **1711**, 99-125, doi:10.1016/j.bbamem.2005.04.001 (2005).

6 Sohl, G. & Willecke, K. Gap junctions and the connexin protein family. *Cardiovasc Res* **62**, 228-232, doi:10.1016/j.cardiores.2003.11.013 (2004).

7 Harris, A. L. Connexin channel permeability to cytoplasmic molecules. *Prog Biophys Mol Biol* **94**, 120-143, doi:10.1016/j.pbiomolbio.2007.03.011 (2007).

8 Bonacquisti, E. E. & Nguyen, J. Connexin 43 (Cx43) in cancer: Implications for therapeutic approaches via gap junctions. *Cancer Lett* **442**, 439-444, doi:10.1016/j.canlet.2018.10.043 (2019).

9 Delmar, M. *et al.* Connexins and Disease. *Cold Spring Harbor perspectives in biology*, doi:10.1101/cshperspect.a029348 (2017).

10 Garcia, I. E. *et al.* Connexinopathies: a structural and functional glimpse. *BMC Cell Biol* **17 Suppl 1**, 17, doi:10.1186/s12860-016-0092-x (2016).

11 Aasen, T., Mesnil, M., Naus, C. C., Lampe, P. D. & Laird, D. W. Gap junctions and cancer: communicating for 50 years. *Nat Rev Cancer* **16**, 775-788, doi:10.1038/nrc.2016.105 (2016).

12 Makowski, L., Caspar, D. L., Phillips, W. C. & Goodenough, D. A. Gap junction structures. II. Analysis of the x-ray diffraction data. *J Cell Biol* **74**, 629-645 (1977).

13 Revel, J. P. & Karnovsky, M. J. Hexagonal array of subunits in intercellular junctions of the mouse heart and liver. *J Cell Biol* **33**, C7-C12, doi:10.1083/jcb.33.3.c7 (1967).

14 Kistler, J., Goldie, K., Donaldson, P. & Engel, A. Reconstitution of native-type noncrystalline lens fiber gap junctions from isolated hemichannels. *J Cell Biol* **126**, 1047-1058, doi:10.1083/jcb.126.4.1047 (1994).

15 Locke, D. & Harris, A. L. Connexin channels and phospholipids: association and modulation. *BMC Biol* **7**, 52, doi:10.1186/1741-7007-7-52 (2009).

16 Myers, J. B. *et al.* Structure of native lens connexin 46/50 intercellular channels by cryo-EM. *Nature* **564**, 372-377, doi:10.1038/s41586-018-0786-7 (2018).

17 Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J Am Chem Soc* **126**, 3477-3487, doi:10.1021/ja0393574 (2004).

18 Maeda, S. *et al.* Structure of the connexin 26 gap junction channel at 3.5 A resolution. *Nature* **458**, 597-602, doi:10.1038/nature07869 (2009).

19 Bennett, B. C. *et al.* An electrostatic mechanism for Ca(2+)-mediated regulation of gap junction channels. *Nat Commun* **7**, 8770, doi:10.1038/ncomms9770 (2016).

20 Gong, X. Q. & Nicholson, B. J. Size selectivity between gap junction channels composed of different connexins. *Cell Commun Adhes* **8**, 187-192 (2001).

21 Trexler, E. B., Bukauskas, F. F., Kronengold, J., Bargiello, T. A. & Verselis, V. K. The first extracellular loop domain is a major determinant of charge selectivity in connexin46 channels. *Biophys J* **79**, 3036-3051, doi:10.1016/S0006-3495(00)76539-8 (2000).

22 Kronengold, J., Trexler, E. B., Bukauskas, F. F., Bargiello, T. A. & Verselis, V. K. Pore-lining residues identified by single channel SCAM studies in Cx46 hemichannels. *Cell Commun Adhes* **10**, 193-199 (2003).

23 Verselis, V. K., Trelles, M. P., Rubinos, C., Bargiello, T. A. & Srinivas, M. Loop gating of connexin hemichannels involves movement of pore-lining residues in the first extracellular loop domain. *The Journal of biological chemistry* **284**, 4484-4493, doi:10.1074/jbc.M807430200 (2009).

24 Oh, S., Verselis, V. K. & Bargiello, T. A. Charges dispersed over the permeation pathway determine the charge selectivity and conductance of a Cx32 chimeric hemichannel. *J Physiol* **586**, 2445-2461, doi:10.1113/jphysiol.2008.150805 (2008).

25 Kwon, T. *et al.* Molecular dynamics simulations of the Cx26 hemichannel: insights into voltage-dependent loop-gating. *Biophys J* **102**, 1341-1351, doi:10.1016/j.bpj.2012.02.009 (2012).

26 Kwon, T., Harris, A. L., Rossi, A. & Bargiello, T. A. Molecular dynamics simulations of the Cx26 hemichannel: evaluation of structural models with Brownian dynamics. *J Gen Physiol* **138**, 475-493, doi:10.1085/jgp.201110679 (2011).

27 Zonta, F., Polles, G., Zanotti, G. & Mammano, F. Permeation pathway of homomeric connexin 26 and connexin 30 channels investigated by molecular dynamics. *J Biomol Struct Dyn* **29**, 985-998, doi:10.1080/073911012010525027 (2012).

28 Bargiello, T. A., Tang, Q., Oh, S. & Kwon, T. Voltage-dependent conformational changes in connexin channels. *Biochim Biophys Acta* **1818**, 1807-1822, doi:10.1016/j.bbamem.2011.09.019 (2012).

29 Tong, X. *et al.* The First Extracellular Domain Plays an Important Role in Unitary Channel Conductance of Cx50 Gap Junction Channels. *PLoS One* **10**, e0143876, doi:10.1371/journal.pone.0143876 (2015).

30 Lopez, W. *et al.* Mechanism of gating by calcium in connexin hemichannels. *Proc Natl Acad Sci U S A* **113**, E7986-E7995, doi:10.1073/pnas.1609378113 (2016).

31 Garcia, I. E. *et al.* The syndromic deafness mutation G12R impairs fast and slow gating in Cx26 hemichannels. *J Gen Physiol* **150**, 697-711, doi:10.1085/jgp.201711782 (2018).

32 Rubinos, C., Sanchez, H. A., Verselis, V. K. & Srinivas, M. Mechanism of inhibition of connexin channels by the quinine derivative N-benzylquininium. *J Gen Physiol* **139**, 69-82, doi:10.1085/jgp.201110678 (2012).

33 Banks, E. A. *et al.* Connexin mutation that causes dominant congenital cataracts inhibits gap junctions, but not hemichannels, in a dominant negative manner. *J Cell Sci* **122**, 378-388, doi:10.1242/jcs.034124 (2009).

34 Berthoud, V. M. *et al.* Connexin50D47A decreases levels of fiber cell connexins and impairs lens fiber cell differentiation. *Invest Ophthalmol Vis Sci* **54**, 7614-7622, doi:10.1167/iovs.13-13188 (2013).

35 Reis, L. M. *et al.* Whole exome sequencing in dominant cataract identifies a new causative factor, CRYBA2, and a variety of novel alleles in known genes. *Hum Genet* **132**, 761-770, doi:10.1007/s00439-013-1289-0 (2013).

36 White, T. W., Bruzzone, R., Wolfram, S., Paul, D. L. & Goodenough, D. A. Selective interactions among the multiple connexin proteins expressed in the vertebrate

lens: the second extracellular domain is a determinant of compatibility between connexins. *J Cell Biol* **125**, 879-892 (1994).

37 White, T. W., Paul, D. L., Goodenough, D. A. & Bruzzone, R. Functional analysis of selective interactions among rodent connexins. *Mol Biol Cell* **6**, 459-470, doi:10.1091/mbc.6.4.459 (1995).

38 Nakagawa, S. *et al.* Asparagine 175 of connexin32 is a critical residue for docking and forming functional heterotypic gap junction channels with connexin26. *The Journal of biological chemistry* **286**, 19672-19681, doi:10.1074/jbc.M110.204958 (2011).

39 Cottrell, G. T. & Burt, J. M. Functional consequences of heterogeneous gap junction channel formation and its influence in health and disease. *Biochim Biophys Acta* **1711**, 126-141, doi:10.1016/j.bbamem.2004.11.013 (2005).

40 Bai, D. & Wang, A. H. Extracellular domains play different roles in gap junction formation and docking compatibility. *Biochem J* **458**, 1-10, doi:10.1042/BJ20131162 (2014).

41 Schadzek, P. *et al.* The cataract related mutation N188T in human connexin46 (hCx46) revealed a critical role for residue N188 in the docking process of gap junction channels. *Biochim Biophys Acta* **1858**, 57-66, doi:10.1016/j.bbamem.2015.10.001 (2016).

42 Silander, K. *et al.* Spectrum of mutations in Finnish patients with Charcot-Marie-Tooth disease and related neuropathies. *Hum Mutat* **12**, 59-68, doi:10.1002/(SICI)1098-1004(1998)12:1<59::AID-HUMU9>3.0.CO;2-A (1998).

43 Primignani, P. *et al.* A novel dominant missense mutation--D179N--in the GJB2 gene (Connexin 26) associated with non-syndromic hearing loss. *Clin Genet* **63**, 516-521, doi:10.1034/j.1399-0004.2003.00079.x (2003).

44 Deeley, J. M. *et al.* Human lens lipids differ markedly from those of commonly used experimental animals. *Biochim Biophys Acta* **1781**, 288-298, doi:10.1016/j.bbalip.2008.04.002 (2008).

45 Lampe, P. D. *et al.* In vitro assembly of gap junctions. *J Struct Biol* **107**, 281-290 (1991).

46 Mabrey, S. & Sturtevant, J. M. Investigation of phase transitions of lipids and lipid mixtures by sensitivity differential scanning calorimetry. *Proc Natl Acad Sci U S A* **73**, 3862-3866, doi:10.1073/pnas.73.11.3862 (1976).

47 Shaw, A. W., McLean, M. A. & Sligar, S. G. Phospholipid phase transitions in homogeneous nanometer scale bilayer discs. *FEBS Lett* **556**, 260-264, doi:10.1016/s0014-5793(03)01400-5 (2004).

48 Vermeer, L. S., de Groot, B. L., Reat, V., Milon, A. & Czaplicki, J. Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular

dynamics simulations and comparison with 2H NMR experiments. *Eur Biophys J* **36**, 919-931, doi:10.1007/s00249-007-0192-9 (2007).

49 Khakbaz, P. & Klauda, J. B. Investigation of phase transitions of saturated phosphocholine lipid bilayers via molecular dynamics simulations. *Biochim Biophys Acta Biomembr* **1860**, 1489-1501, doi:10.1016/j.bbamem.2018.04.014 (2018).

50 Caspar, D. L., Goodenough, D. A., Makowski, L. & Phillips, W. C. Gap junction structures. I. Correlated electron microscopy and x-ray diffraction. *J Cell Biol* **74**, 605-628 (1977).

51 Schubert, A. L., Schubert, W., Spray, D. C. & Lisanti, M. P. Connexin family members target to lipid raft domains and interact with caveolin-1. *Biochemistry* **41**, 5754-5764, doi:10.1021/bi0121656 (2002).

52 Locke, D., Liu, J. & Harris, A. L. Lipid rafts prepared by different methods contain different connexin channels, but gap junctions are not lipid rafts. *Biochemistry* **44**, 13027-13042, doi:10.1021/bi050495a (2005).

53 Hunte, C. Specific protein-lipid interactions in membrane proteins. *Biochem Soc Trans* **33**, 938-942 (2005).

54 Spray, D. C., Rozental, R. & Srinivas, M. Prospects for rational development of pharmacological gap junction channel blockers. *Current drug targets* **3**, 455-464 (2002).

55 Ritchie, T. K. *et al.* Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods Enzymol* **464**, 211-231, doi:10.1016/S0076-6879(09)64011-8 (2009).

56 Kistler, J., Christie, D. & Bullivant, S. Homologies between gap junction proteins in lens, heart and liver. *Nature* **331**, 721-723, doi:10.1038/331721a0 (1988).

57 Kistler, J., Schaller, J. & Sigrist, H. MP38 contains the membrane-embedded domain of the lens fiber gap junction protein MP70. *The Journal of biological chemistry* **265**, 13357-13361 (1990).

58 White, T. W., Bruzzone, R., Goodenough, D. A. & Paul, D. L. Mouse Cx50, a functional member of the connexin family of gap junction proteins, is the lens fiber protein MP70. *Mol Biol Cell* **3**, 711-720 (1992).

59 Wang, Z. & Schey, K. L. Phosphorylation and truncation sites of bovine lens connexin 46 and connexin 50. *Exp Eye Res* **89**, 898-904, doi:10.1016/j.exer.2009.07.015 (2009).

60 Reichow, S. L. *et al.* Allosteric mechanism of water-channel gating by Ca2+-calmodulin. *Nat Struct Mol Biol* **20**, 1085-1092, doi:10.1038/nsmb.2630 (2013).

61 Gold, M. G. *et al.* AKAP2 anchors PKA with aquaporin-0 to support ocular lens transparency. *EMBO Mol Med* **4**, 15-26, doi:10.1002/emmm.201100184 (2012).

62 Reichow, S. L. & Gonen, T. Noncanonical binding of calmodulin to aquaporin-0: implications for channel regulation. *Structure* **16**, 1389-1398, doi:10.1016/j.str.2008.06.011 (2008).

63 Efremov, R. G., Gatsogiannis, C. & Raunser, S. Lipid Nanodiscs as a Tool for High-Resolution Structure Determination of Membrane Proteins by Single-Particle Cryo-EM. *Methods Enzymol* **594**, 1-30, doi:10.1016/bs.mie.2017.05.007 (2017).

64 Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46, doi:10.1016/j.jsb.2006.05.009 (2007).

65 Ludtke, S. J. Single-Particle Refinement and Variability Analysis in EMAN2.1. *Methods Enzymol* **579**, 159-189, doi:10.1016/bs.mie.2016.05.001 (2016).

66 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, doi:10.7554/eLife.42166 (2018).

67 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).

68 Zivanov, J., Nakane, T. & Scheres, S. H. W. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253-267, doi:10.1107/S2052252520000081 (2020).

69 Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat Methods* **9**, 853-854, doi:10.1038/nmeth.2115 (2012).

70 Scheres, S. H. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-157, doi:10.1016/bs.mie.2016.04.012 (2016).

71 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

72 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr.* **D66**, 486-501 (2010).

73 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, doi:10.1107/S2059798318006551 (2018).

74 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315, doi:10.1002/pro.3330 (2018).

75 Barad, B. A. *et al.* EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat Methods* **12**, 943-946, doi:10.1038/nmeth.3541 (2015).

76 Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* **65**, 1074-1080, doi:10.1107/S0907444909029436 (2009).

77 Konig, N. & Zampighi, G. A. Purification of bovine lens cell-to-cell channels composed of connexin44 and connexin50. *J Cell Sci* **108 (Pt 9)**, 3091-3098 (1995).

78 Jiang, J. X. & Goodenough, D. A. Heteromeric connexons in lens gap junction channels. *Proc Natl Acad Sci U S A* **93**, 1287-1291 (1996).

79 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).

80 Shearer, D., Ens, W., Standing, K. & Valdimarsson, G. Posttranslational modifications in lens fiber connexins identified by off-line-HPLC MALDI-quadrupole time-of-flight mass spectrometry. *Invest Ophthalmol Vis Sci* **49**, 1553-1562, doi:10.1167/iovs.07-1193 (2008).

81 Varland, S., Osberg, C. & Arnesen, T. N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. *Proteomics* **15**, 2385-2401, doi:10.1002/pmic.201400619 (2015).

82 Grubmuller, H., Heymann, B. & Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271**, 997-999 (1996).

83 Wu, E. L. *et al.* CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **35**, 1997-2004, doi:10.1002/jcc.23702 (2014).

84 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802, doi:10.1002/jcc.20289 (2005).

85 Huang, J. & MacKerell, A. D., Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34**, 2135-2145, doi:10.1002/jcc.23354 (2013).

86 Buchoux, S. FATSLiM: a fast and robust software to analyze MD simulations of membranes. *Bioinformatics* **33**, 133-134, doi:10.1093/bioinformatics/btw563 (2017).

87 Piggot, T. J., Allison, J. R., Sessions, R. B. & Essex, J. W. On the Calculation of Acyl Chain Order Parameters from Lipid Simulations. *J Chem Theory Comput* **13**, 5683-5696, doi:10.1021/acs.jctc.7b00643 (2017).

88 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25, doi:10.1002/pro.3235 (2018).

89 Community, B. O. (Stichting Blender Foundation, Masterdam, 2018).

90 Donaldson, P. & Kistler, J. Reconstitution of channels from preparations enriched in lens gap junction protein MP70. *J Membr Biol* **129**, 155-165 (1992).

91 Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J Mol Graph* **14**, 354-360, 376 (1996).

Chapter 5

Connexin 46 and connexin 50 gap junction channel properties are shaped by structural and dynamic features of their N-terminal domains

This chapter has been published

Connexin 46 and connexin 50 gap junction channel properties are shaped by structural and dynamic features of their N-terminal domains. *Journal of Physiology.* (2021)

Benny Yue,[1,†] Bassam G. Haddad,[2,†] Umair Khan,[2] Honghong Chen,[1] Mena Atalla,[1] Ze Zhang,[1] Daniel M. Zuckerman,[3] Steve L. Reichow[2,*] and Donglin Bai[1,*]

[1]Department of Physiology and Pharmacology, University of Western Ontario, London, Ontario, Canada

[2]Department of Chemistry, Portland State University, Portland, OR 97201, USA

[3]Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR 97239, USA

† Indicates equal contribution

* Author of correspondence:

Dr. Donglin Bai, Department of Physiology and Pharmacology, University of Western Ontario, London, N6A 5C1, Canada. Tel. 519-850-2569; Email: donglin.bai@schulich.uwo.ca

Dr. Steve L. Reichow, Department of Chemistry, Portland State University, Portland, OR 97201, USA. Tel. 503-725-7766; Email: reichow@pdx.edu

This project was a collaboration between the Reichow lab and the Bai lab. Benny Yue designed, conducted and analyzed patch-clamp experiments. Bassam Haddad designed

and analyzed MD simulations. Umair Khan built homology models and conducted MD simulations under the guidance of Bassam Haddad and Daniel Zuckerman. Mena Atella performed cell-fluorescence assays. Honghong Chen and Ze Zhang aided in patch-clamp experiments. Steve Reichow and Donglin Bai directed the project.

Abstract

Connexins form intercellular communication channels, known as gap junctions (GJs), that facilitate diverse physiological roles, from long-range electrical and chemical coupling to coordinating development and nutrient exchange. GJs formed by different connexin isoforms harbor unique channel properties that have not been fully defined mechanistically. Recent structural studies on Cx46 and Cx50 defined a novel and stable open state and implicated the amino-terminal (NT) domain as major contributor for isoform-specific functional differences between these closely related lens connexins. To better understand these differences, we constructed models corresponding to wildtype Cx50 and Cx46 GJs, NT domain swapped chimeras, and point variants at the 9^{th} residue for comparative MD simulation and electrophysiology studies. All constructs formed functional GJ channels, except the chimeric Cx46-50NT variant, which correlated with an introduced steric clash and increased dynamical behavior (instability) of the NT domain observed by MD simulation. Single channel conductance correlated well with free-energy landscapes predicted by MD, but resulted in a surprisingly greater degree of effect. Additionally, we observed significant effects on transjunctional voltage-dependent gating (V_j-gating) and/or open state dwell times induced by the designed NT domain variants.

Together, these studies indicate intra- and inter-subunit interactions involving both hydrophobic and charged residues within the NT domains of Cx46 and Cx50 play important roles in defining GJ open state stability and single channel conductance, and establish the open state Cx46/50 structural models as archetypes for structure-function studies targeted at elucidating GJ channel mechanism and the molecular basis of cataract-linked connexin variants.

Introduction

Gap junctions (GJs) are a class of membrane channels that provide a direct passageway between neighboring cells and facilitate the exchange of ions and small molecules (Saez *et al.*, 2003; Goodenough & Paul, 2009). This type of cell-to-cell communication is facilitated by a unique channel architecture, whereby a continuous water-filled pore ~1.5 nm in diameter is formed between two opposing cell membranes, effectively coupling the cytoplasms of adjoined cells (Sosinsky & Nicholson, 2005). These direct passageways are essential for enabling fast transmission of electrical signals in the brain and heart and for facilitating long-range metabolic coupling in most tissues. Because of their important physiological roles, genetic mutations or pathological conditions that lead to aberrant channel function have been linked to a variety of human disease, including deafness, cataracts, peripheral neuropathy, cardiac arrhythmia, stroke, skin disorders, and cancers (Aasen *et al.*, 2016; Garcia *et al.*, 2016; Delmar *et al.*, 2017).

GJ intercellular channels are formed through the assembly of twelve integral membrane proteins called connexins. Six connexins oligomerize to form a hemichannel (also known

as connexon) and two hemichannels from neighboring cells can dock together to form a functional gap junction channel if they are docking compatible (White *et al.*, 1994b; Bai *et al.*, 2018). All connexins are predicted to have the same topological structure, which includes four transmembrane helices (TM1-4), two extracellular loops (EC1 and EC2), an amino-terminal (NT) domain, a carboxyl terminus (CT), and a cytoplasmic loop (CL) (Sohl & Willecke, 2004). Humans express 21 connexin isoforms in a cell-type specific fashion, possibly reflecting the need for unique channel functions that match the physiological demands of their environment. Adding to this diversity, different tissues often express more than one type of connexin, which may co-assemble into so-called heteromeric channels (mixed within the same hemichannel) and/or heterotypic channels (mixed between opposing hemichannels) (Koval *et al.*, 2014; Bai *et al.*, 2018). In this way, cells may fine-tune their GJ channel properties to control the synchronization of physiological activities or to maintain homeostasis.

It has been well-established that GJs formed by different connexins display different channel properties, including distinct permeability to substrates, various rates of ion permeation, and gating control by a variety of factors, including intracellular protons, divalent cations, and transjunctional voltage (V_j) (Bukauskas & Verselis, 2004; Bargiello & Brink, 2009). Transjunctional voltage dependent gating (also known as V_j-gating) exists in all characterized GJs, and depending on the component connexins the resultant GJs could show various levels of gating extent, half deactivation voltage, gating charge (which determines gating sensitivity), and gating kinetics (Harris *et al.*, 1981; Paul *et al.*, 1991; Veenstra *et al.*, 1994; Verselis *et al.*, 1994; White *et al.*, 1994a; Trexler *et al.*, 1996; Oh *et*

al., 1999; Musa *et al.*, 2004). Similarly, different connexin GJs also show drastically different rates of ion permeation measured by single channel conductance (γ_j), from a few pico-Siemens (pS) in Cx30.2 and Cx36 to 200 – 300 pS in Cx50 and Cx37 (Veenstra *et al.*, 1994; Srinivas *et al.*, 1999; Moreno *et al.*, 2005; Bukauskas *et al.*, 2006). Although such distinct channel functions are well-established, the underlying molecular and structural mechanisms for controlling GJ channel gating and ion permeation have not yet been fully defined.

Early functional studies on β-type connexins, such as Cx26 and Cx32, revealed that the charged residues in the beginning portion of the NT domain (within the first 10 amino acid residues) are important in sensing the V_j, determining the V_j-gating polarity, and likely serving as a gate to close the channel (Verselis *et al.*, 1994; Purnick *et al.*, 2000; Oh *et al.*, 2004). Studies on Cx40, Cx46 and Cx50 or their orthologs (members of α connexins) also showed the importance of charged residues in their respective NT domains in determining GJ gating polarity, channel open stability, chemical block, and/or rate of ion permeation (Musa *et al.*, 2004; Tong *et al.*, 2004; Peracchia & Peracchia, 2005; Srinivas *et al.*, 2005; Tong & Ebihara, 2006; Xin *et al.*, 2010; Xin & Bai, 2013). However, studies on residues in Cx45 or Cx50 NT showed that exchange of charged residues with similar charge properties could also alter V_j-gating properties, gating kinetics, and/or the rate of ion permeation (Xin *et al.*, 2012; Santos-Miranda *et al.*, 2020), suggesting features in addition to the charge state of the NT domain play a role in these channel properties. The structural mechanisms for how NT domains are anchored in an open conformation and the molecular interactions to stabilize/destabilize this domain are also not yet clear.

Previous structural studies on a Cx26 M34A variant showed evidence that the NT domain of Cx26 could form a plug-like structure in the center of the pore, likely representing a closed GJ structure (Oshima *et al.*, 2007; Oshima *et al.*, 2011). However, due to the limited resolution of this study, the detailed NT domain interactions responsible for forming the plug were not resolved (Oshima *et al.*, 2011).

In 2009, Maeda and colleagues resolved the first high resolution (3.5Å) crystal structure of human Cx26 GJ (Maeda *et al.*, 2009). In this structure, the NT domains folds into the GJ channel pore forming the narrowest part of the GJ channel, a position that could serve the role of V_j-sensing, regulating ion permeation selectivity and rate, modulation by intracellular substrates, as well as gating to close the GJ channel (Maeda *et al.*, 2009). However, MD simulation studies using this Cx26 GJ structure indicate that the NT domain may not represent the open state, and required significant chemical modification and/or freezing the dynamics of the structure to explain experimental data (Kwon *et al.*, 2011; Villanelo *et al.*, 2017). In 2018, using single particle Cryo-EM, we resolved native sheep lens GJs, Cx46 and Cx50, at 3.4 Å resolution (Myers *et al.*, 2018). The NT domains of these two α connexins also folded into the vestibule of the GJ pore to form the narrowest part of the pore, but in this case the NT domain adopted a more regular amphipathic helical conformation, where hydrophobic resides were positioned to anchor the NT domain through hydrophobic packing interactions with TM1/TM2, resulting in a more stable open state conformation as compared to Cx26 (Myers *et al.*, 2018). In addition to these conformational differences, charged residues could directly influence the electrostatic environment of the permeation pathway and several novel intra- and inter-subunit

interactions were discovered in the NT domain and it is not clear if these interactions play a role in stabilizing/destabilizing the channel open state or V_j gating mechanism.

To explore these findings further and harmonize the functional study with our structural models, we combined MD simulation studies with dual patch clamp technique to characterize GJ properties of sheep Cx46, Cx50, NT domain swapped chimeras (Cx46-50NT and Cx50-46NT), and single point variants (Cx46-R9N and Cx50-N9R). Our results showed well correlated atomic models and experimentally measured channel properties, despite timescale limitations of all-atom MD simulations of this size (typically limited to ns – μs timescales) that prevented full characterization of GJ gating properties measured by patch clamp measurements (ms – s timescales). This combined approach revealed a novel gating model that is dependent not just the charge properties of the NT domain, but also the hydrophobic anchoring sites that we show structurally and functionally serve to stabilize (or destabilize) the NT gating domain, which we believe represents a new paradigm in our understanding of how GJ gating is modulated. In addition, we demonstrate how local salt bridge interactions between neighboring NT domains are likely very important for modulating both the conductance and V_j-gating of Cx46 and Cx50 GJs. More broadly, this type of interaction suggests that cooperativity between neighboring connexin subunits may play a previously unrecognized, yet important role in defining the gating properties of GJs. Taken together, we demonstrate that the combined MD simulation models and electrophysiological functional studies on Cx46/50 are very powerful in providing novel insights in the molecular and structural mechanisms

governing the GJ gating properties, channel open state stability, and energetics of ion permeation.

Methods

Plasmid construction. Sheep Cx46 (Cx46, also known as Cx44) and Cx50 (Cx50, also known as Cx49) cDNA were synthesized and each of them was inserted into an expression vector, pIRES2-EGFP, with an untagged GFP reporter between the restriction enzyme sites, XhoI and EcoRI (NorClone Biotech Laboratories, London, Ontario). Cx46-IRES-GFP was used as a template for polymerase chain reaction cloning to generate the chimera, Cx46-50NT, in which the amino terminal (NT) domain of Cx46 was replaced by Cx50 NT domain and a missense variant Cx46-R9N. Similarly, Cx50-IRES-GFP was used as a template to generate Cx50-46NT and Cx50-N9R. The primers used to generate these chimeras and point variants are listed below.

Cx46-50NT: forward 5’ ATG GGC GAC TGG AGC TTC CTG GGG AAC ATC TTG GAG GAG GTG AAT GAG CAC TCC ACT GTC ATC 3’ and reverse 5’ GAT GAC AGT GGA GTG CTC ATT CAC CTC CTC CAA GAT GTT CCC CAG GAA GCT CCA GTC GCC CAT 3’

Cx50-46NT: forward 5’ ATG GGA GAC TGG AGT TTC CTG GGG AGA CTC CTA GAG AAC GCC CAG GAG CAC TCC ACG GTC ATC 3’ and reverse 5’ GAT GAC CGT GGA GTG CTC CTG GGC GTT CTC TAG GAG TCT CCC CAG GAA ACT CCA GTC TCC CAT 3’

Cx46-R9N: forward 5’ GAC TGG AGC TTC CTG GGG AAC CTC CTA GAG AAC GCC CAG 3’ and reverse 5’ CTG GGC GTT CTC TAG GAG GTT CCC CAG GAA GCT CCA GTC 3’

Cx50-N9R: forward 5’ GAC TGG AGC TTC CTG GGG AAC CTC CTA GAG AAC GCC CAG 3’ and reverse: 5’ CTG GGC GTT CTC TAG GAG GTT CCC CAG GAA GCT CCA GTC 3’

GFP fusion tagged Cx46 (Cx46-GFP), the chimera (Cx46-50NT-GFP), and point variant (Cx46-R9N-GFP) were generated by subcloning these construct into EGFP expression vector (Kim *et al.*, 2013). Mutagenesis was used to remove the stop codon from each of these vectors and ensure that the GFP was linked in frame with a peptide linker (LGILQSTVPRARDPPVAT) between Cx46 (or Cx46-50NT) and GFP.

Cell culture and transient transfections. Gap junction (GJ) deficient mouse neuroblastoma (N2A) cells (American Type Culture Collection, Manassas, VA, USA) were grown in Dulbecco’s Modified Eagle’s Medium (DMEM) (Life Technologies Corporation, Grand Island, NY, USA) containing 4.5 g/L D-(+)-glucose, 584 mg/L L-glutamine, 110 mg/L sodium pyruvate, 10% fetal bovine serum (FBS), 1% penicillin, and 1% streptomycin, in an incubator with 5% CO_2 at 37° C (Sun *et al.*, 2013). N2A cells were transfected with 1.0 µg of a cDNA construct and 2 µL of X-tremeGENE HP DNA transfection reagent (Roche Diagnostics GmbH, Indianapolis, IN, USA) in Opti-MEM + GlutaMAX-I medium for 4 hours. After transfection, the medium was changed back to FBS-containing DMEM and incubated overnight. The next day, N2A cells transfected with Cx46 and its variants were replated onto glass coverslips for 2 hours before transferring to the recording chamber. N2A cells transfected with Cx50 and its variants were replated onto glass coverslips for 10 hours before transferring to the recording chamber. Isolated green fluorescent cell pairs were selected for patch clamp study of homotypic GJs.

Electrophysiological recording. Glass coverslips with transfected cells were placed into a recording chamber on an upright microscope (BX51WI, Olympus). The chamber was filled with extracellular solution (ECS) at room temperature (22 – 24° C). The ECS contained (in mM): 135 NaCl, 2 CsCl, 2 $CaCl_2$, 1 $MgCl_2$, 1 $BaCl_2$, 10 HEPES, 5 KCl, 5 D-(+)-glucose, 2 Sodium pyruvate, pH adjusted to 7.4 with 1M NaOH, and osmolarity of 310-320 mOsm. Dual whole cell patch clamp was performed on green fluorescent cell pairs with MultiClamp 700A amplifier (Molecular Devices, Sunnyvale, CA, USA). Patch pipettes were pulled with a micropipette puller (PC-10, Narishige International USA Inc., Amityville, NY, USA) and filled with intracellular solution (ICS) containing (in mM): 130 CsCl, 10 EGTA, 0.5 $CaCl_2$, 5 Na_2ATP, 10 HEPES, adjusted to pH 7.2 with 1 M CsOH, and osmolarity of 290-300 mOsm. After establishing whole cell recording on the cell pair, both cells were voltage clamped at 0 mV. In one cell of the pair, a series of voltage pulses (±20 to ±100 mV) were applied to establish transjunctional voltage (V_j). The other cell of the pair was constantly held at 0 mV to record gap junctional current (I_j). The current was low-pass filtered (Bessel filter at 1 kHz) and recorded using pClamp9.2 software at a sampling rate of 10 kHz via an AD/DA converter (Digidata 1322A, Molecular Devices, Sunnyvale, CA, USA).

Transjunctional voltage dependent gating. Transjunctional voltage (V_j) dependent gating (V_j-gating) was studied in cell pairs expressing one of the constructs. Voltage pulses (±20 to ±100 mV with 20 mV increment) were delivered in one cell of the pair to establish V_js and transjunctional currents (I_js) were recorded in the other cell. In most cases, I_js peaked

at the beginning and then deactivated (especially with high V_js, ±40 to ±100 mV) to a steady-state near the end of a 7 second V_j pulse. Gap junctional conductance (G_j) was calculated ($G_j = I_j/V_j$). The steady state G_j was normalized to the peak G_j to obtain a normalized steady-state junctional conductance ($G_{j,ss}$) for each tested V_js. The $G_{j,ss}$ were then plotted with V_js to obtain a $G_{j,ss} - V_j$ plot, which could sometimes fit well with a two-state Boltzmann equation for each V_j polarity to obtain gating parameters, V_0, G_{min}, G_{max}, and *A* (Jassim *et al.*, 2016). V_0 is the voltage when the $G_{j,ss}$ is reduced by half [(G_{max} - G_{min})/2], G_{min} is the normalized minimum residual conductance, while G_{max} represents the maximum normalized conductance, and *A* is the slope of the curve which reflects V_j-gating sensitivity (Spray *et al.*, 1981).

$$G_{j,ss} = \frac{G_{max} - G_{min}}{1 + \exp\left[A(V_j - V_0)\right]} + G_{min} \qquad \text{Eq. 1}$$

Single Channel Analysis. Single channel current (i_j) could be observed in cell pairs with few active GJ channels. The recorded currents were further filtered using a low-pass Gaussian filter at 200 Hz in Clampfit9.2 for measuring current amplitude and display in figures. The amplitude i_js for fully open state at different V_js were measured with fitting Gaussian functions on all point current amplitude histograms. The i_js of different cell pairs were averaged under the same V_j, regardless of V_j polarity, to generate $i_j - V_j$ plot. Linear regression of $i_j - V_j$ plot with at least 4 different V_js was used to estimate slope unitary GJ channel conductance (also known as single channel conductance, γ_j).

Single GJ channel open dwell times were analyzed on cell pairs with only one active GJ channel. In this case, single channel currents were filtered with a low-pass Gaussian filter

at 500 Hz (Clampfit9.2). The open dwell time was measured at the V_js from ±40 to ±100 mV using Clampfit9.2. Any single channel open events with less than 2 ms were excluded from the analysis as this is beyond the resolution of our single channel recordings. The average open dwell time at each V_j was calculated and plotted for comparison. The total number of open events for Cx50 and Cx50-N9R GJs was a lot higher than other GJs and suitable for more in depth analysis on the open dwell times. Histograms of Cx50 and Cx50-N9R were plotted in a logarithmic scale with 5 bins/decade and fitted with a two-exponential probability fitting (Ramanan et al. 1999). The two time constant (□□ values represent the corresponding open dwell times for each of the peaks in the fitting. The □$_{mean}$ value was calculated by taking the weighted average of the two □ values using the area under each peak.

Molecular dynamics simulations. Visual Molecular Dynamics (VMD) v.1.9.3 was used to build the dodecameric channels for Cx46 (PDB: 6MHQ) and Cx50 (PDB: 6MHY) wildtype systems for molecular dynamics (MD) simulations (Humphrey *et al.*, 1996; Myers *et al.*, 2018). Sidechains were protonated according to neutral conditions, and the protonated HSD model was used for all histidine residues. Disulfide bonds identified in the experimental structures were enforced for both models. Amino acids corresponding to the cytoplasmic loop (CL) connecting TM2–TM3, and the C-terminal (CT) domain of Cx46 and Cx50 were not included for molecular dynamics simulation, as experimental data describing the structure of these large domains (~50 residue CL and ~200 residue CT domain in Cx46 and Cx50) are missing. The introduced N- and C-terminal residues

resulting from the missing CL segment (L97 and L142 in Cx46; V97 and L154 in Cx50) were neutralized. NT acetylation sites were introduced in VMD through an all-atom acetylation patch using the AutoPSF plugin, to mimic the *in vivo* co-translational modification identified in the native proteins (Myers *et al.*, 2018). The prepared protein structures were submerged in a hydration shell using Solvate v.1.0.177. Water was removed from sections of the channels corresponding to transmembrane domains, based on hydrophobic character and localization of amphipol observed in experimental Cryo-EM data (~20–50 Å from the center of the channel). To mimic a cell-to-cell junction, the VMD Membrane Builder plugin was used to add two lipid bilayers to each system, containing 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), with dimensions of 152 × 152 Å.

The four structural models of designed variants, Cx46-R9N, Cx46-50NT, Cx50-N9R and Cx50-46NT, were also built using VMD, as follows. First, protein-only structures of the respective wildtype models were mutated, as per the sequence differences of the NT domain (defined as residue 2 – 19) for the NT domain chimeras or just the 9th residue for the point variants, using VMD's *mutator* plugin (Humphrey *et al.*, 1996). The variant models were then merged with the lipid bilayers created for their respective wildtype models, described above, using VMD's *mergestructs* plugin. Lipids which overlapped with the protein models were then removed. Each model was then placed in a water box with dimensions 150 × 150 × 180 Å using VMD's *solvate* plugin. The models were neutralized using the Autoionize plugin, then 150 mM KCl and 150 mM NaCl were added to the

solvent areas corresponding to intracellular and extracellular regions of the simulation box, respectively.

GPU-accelerated Nanoscale Molecular Dynamics v.2.13 (Phillips *et al.*, 2005) was used for all classical molecular dynamics (MD) simulations, using the CHARMM36 force field (Huang & MacKerell, 2013) for all atoms and TIP3P explicit model for water. Each model was prepared following the same minimization and equilibration protocol, as follows. First, the lipid tails were allowed to minimize with all other atoms fixed for 1 ns using a 1 fs timestep, allowing the acyl chains to "melt" with constant volume at a temperature of 300 K (NVT). All subsequent simulations were performed using the Langevin piston Nose-Hoover method for pressure control (NPT). Next, the entire system, including lipids, solvent and ions, was allowed to minimize with the protein harmonically constrained (1 kcal mol^{-1}), for 1 ns using a 1 fs timestep. A third 1 ns minimization step was then applied, in which the system was free to minimize with a harmonic constraint applied only to the protein backbone (1 kcal mol^{-1}), to ensure stable quaternary structure while sidechains relax in their local environment. The entire system was then released from all constraints and subject to all-atom equilibration using a Langevin thermostat (damping coefficient of 1 ps^{-1}), with a constant temperature of 310 K and constant pressure of 1 atm, using a 2 fs timestep for 30 ns. Periodic boundary conditions were used to allow for the particle mesh Ewald calculation of electrostatics. Finally, all models were simulated for a total of 100 ns of production, using four 25 ns replicas to facilitate statistical analysis. Each replica started from the end of 30 ns equilibration with velocities reinitialized, using a 2 fs timestep.

Root mean squared deviations (r.m.s.d.), comparing the backbone conformations of MD simulation to the original pre-equilibrated starting structures, and root mean square fluctuations (r.m.s.f.) comparing the amplitudes of backbone fluctuations during MD simulation, were calculated using VMD. All six GJ models approached a steady r.m.s.d. during the first 20 ns of the equilibration phase and maintained stability during all production runs (Figure 1c,d).

Population distribution functions were constructed by monitoring the distance between these residues at the 9^{th} and 12^{th} position in adjacent ($i \rightarrow i+1$) subunits (in the clockwise direction when viewed from the cytoplasmic vestibule). Distances between functional groups were recorded at every tenth step of the trajectory across all four production runs, for Cx50, Cx46 and each of the variants. The points of reference used to measure the interatomic distance were chosen to capture equivalent rotameric states, as follows: $C_{\square}$ for R9, $C_{\square}$ for N9, and $C_{\square}$ for E12. Histograms (bin size = 0.1 Å) were normalized and plotted as probability density functions.

Energetics and analysis of ion permeation pathway. Potentials of mean force (PMF), or the energy landscape, describing the permeation of K^+ and Cl^- were calculated for all systems. To calculate the PMF, a Markov State Model (MSM) was constructed by defining state-space as the position of an ion along the pore-axis (z-axis), which was subdivided into 3 Å bins. A transition matrix (T), which describes the time evolution of the system, was constructed from the conditional probabilities T_{ij} of an ion ending up in state j after a given lag-time, having begun in state i. As in previous work (Myers *et al.*, 2018), we

employed a short lag-time (2 ps), ensuring the vast majority of transition probabilities occurred between nearest neighbors (*i.e.,* i-1 ⟷ i ⟷i+1). The principle of detailed balance guarantees that any connected paring of states (e.g., neighbors only) is sufficient to determine the unique equilibrium distribution independent of lag time. Transition probabilities are estimated by counting the instances of transitions at every lag time and storing the values in a transition count matrix. The count matrix was then row normalized to achieve an approximate transition probability matrix (T)

$$T_{i,i+1} \cong \frac{N_{i,i+1}}{N_i} \qquad \text{Eq. 2}$$

where $N_{i,i+1}$ is the number of transitions from state (*i*) to state (*i+1*) in a lag time τ = 2ps, and N_i is the number of times an ion was found in state *i*. The thermodynamics underlying ionic permeation may then be extracted using the principle of detailed balance (*i.e.* statistical equilibrium) (Eq. 3) and Boltzmann statistics (Eq. 4).

$$P_i^{\mathrm{eq}}\, T_{i,i+1} = P_{i+1}^{\mathrm{eq}}\, T_{i+1,i} \qquad \text{Eq. 3}$$

$$e^{\frac{-\Delta G_{i,i+1}}{RT}} = \frac{P_{i+1}^{\mathrm{eq}}}{P_i^{\mathrm{eq}}} = \frac{T_{i,i+1}}{T_{i+1,i}} \qquad \text{Eq. 4}$$

$$\mathrm{PMF}(i) = \sum_{n=1}^{i-1} \Delta G_{n,n+1} = -\sum_{n=1}^{i-1} RT \ln\left(\frac{T_{n,n+1}}{T_{n+1,n}}\right) \qquad \text{Eq. 5}$$

Here, P_i^{eq} is the probability that an ion will be in each respective bin once equilibrium is achieved, $\Delta G_{i,i+1}$ is the free energy difference between the two states, R is the ideal gas constant (1.986 cal mol^{-1} K^{-1}), and T is temperature (310 K) (Eq. 5). Values from the PMF were mapped to the z-axis, interpolated and smoothed using a b-spline. Final PMFs were symmetrized around the center of the channel and adjusted such that the bulk regions

were at zero. To enable sufficient sampling of Cl^- ions in the channel, a distributed seeding approach was implemented where an individual Cl^- is randomly placed within the channel, followed by short 10 ns simulations, and repeated until sufficient sampling for the MSM is achieved. Further explanation and a detailed justification of the distributed seeding approach, and PMF calculation can be found in (Myers *et al.*, 2018).

Coulombic surface potentials were calculated using the Adaptive Poisson Boltzmann Solver (APBS) (Jurrus *et al.*, 2018) within Chimera (Pettersen *et al.*, 2004), using standard settings. Pore profile analysis of the radius at each point along the pore-axis (Z-axis) was calculated using the program HOLE (Smart *et al.*, 1996). The calculation is done by rolling a sphere with the radius of a water molecule over the Van der Waals surface of the pore, with the beginning/end of the pore being defined as having a maximum radius of 12 Å. To assess the average pore profile of each model obtained by MD simulation, a snapshot of the protein was saved every 1000th frame (i.e. 2 ns), symmetrized, and then averaged together to provide the average pore profile.

Statistical Analysis. Data are expressed as mean ± SD Kruskal Wallis followed by a Dunn's post-hoc test was used to compare multiple groups of non-parametric data, as specified. One-way ANOVA was used to compare multiple groups of data with Gaussian distribution. Statistical significance is indicated on the graphs only for biologically meaningful group comparisons. MD data are expressed as mean ± SD, or ± 95% confidence interval where indicated. Statistical significance for MD data were calculated via a standard two-mean p-value test (Williams, 1991), using the mean values obtained from the symmetry halves of

each system over the four production runs (n = 8). The nature of the distributed seeding protocol used to calculate Cl^- PMF's precluded similar analysis, and so this data was not subjected to comparative statistical analysis and instead reported based on average of the two symmetry halves of each system (n = 2).

Results

Design of variants to probe channel properties of Cx46 and Cx50. The NT domains of Cx46 and Cx50 adopt an amphipathic helix that folds into the GJ pore, contributing to the cytoplasmic vestibules of the channel and differ at only five sites (Fig. 1a,b). This conformational state of the Cx46 and Cx50 NT domain is shown to be stable during multiple replicates of all-atom MD simulation (Fig. 1 c,d). Proteomics analysis on native proteins isolated from the lens indicated that Met1 is removed and the resulting NT Gly2 position is acetylated co-translationally in both isoforms (Shearer *et al.*, 2008; Myers *et al.*, 2018). The first difference in structure occurs at position 9, which is a positively charged arginine in Cx46 and a neutral asparagine in Cx50 (indicated by asterisk in Fig. 1a,b). In the Cryo-EM based models, R9 of Cx46 restricts the pore diameter to ~8.8 Å and was identified as forming the main constriction site within the NT domain (Fig. 1e, and (Myers *et al.*, 2018)). In comparison, Cx50's diameter at this site is significantly wider in the Cryo-EM based model, at 12.4 Å, (with a minimum constriction of ~11.4 Å located lower on NT domain near S5/D3) (Fig. 1e). Further comparisons show the NT of Cx50 lacks any positively charged amino acids (assuming neutralizing of the n-terminal G2 position by acetylation), resulting in an overall net charge of –4, whereas the net charge on the

Cx46 NT domain is –2 (Fig. 1b). Other differences localize to the identity of the hydrophobic positions 10 and 14, which form packing interactions with TM1/2 (Fig. 1a,b).

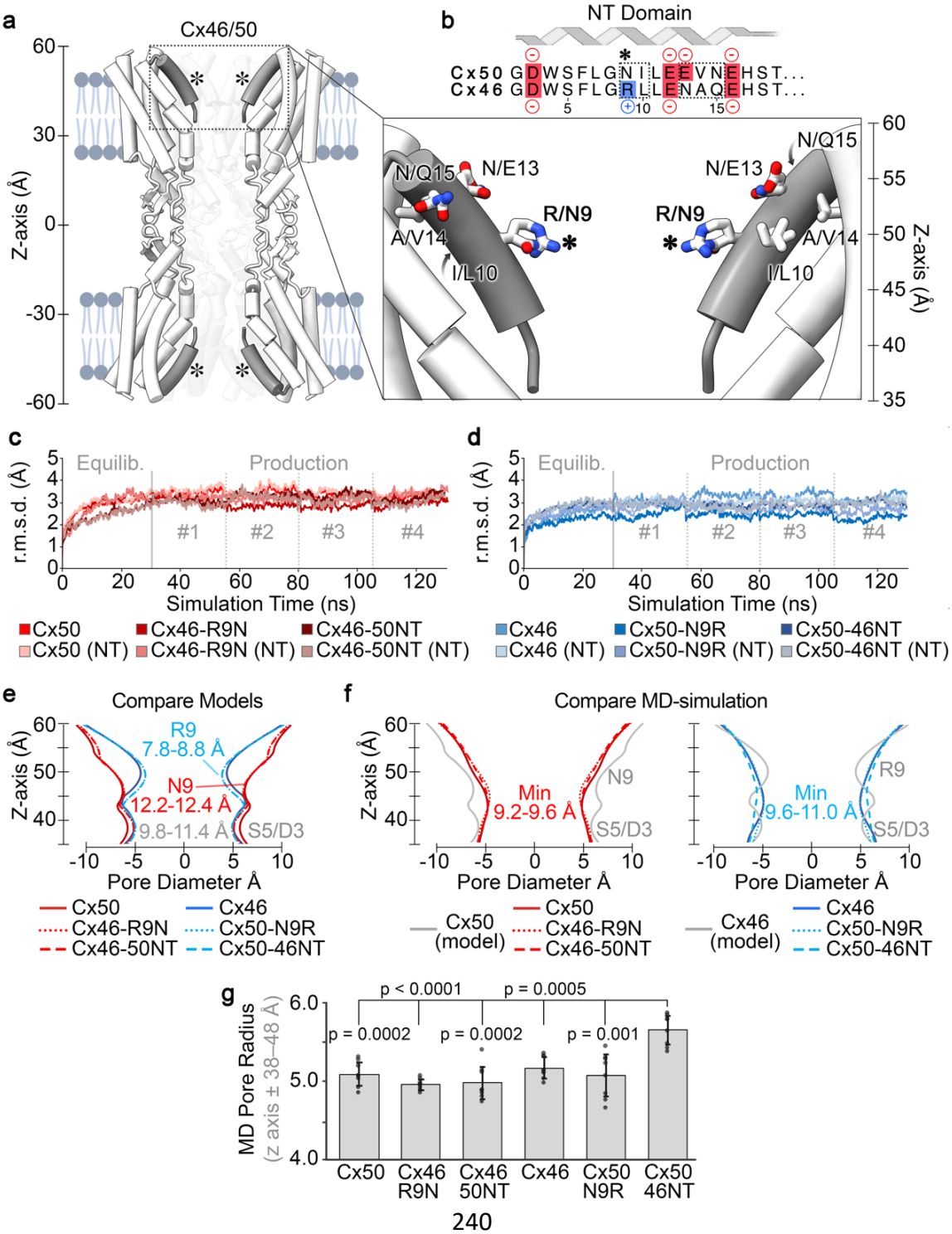


Fig. 1 The NT-domain constriction sites of Cx50, Cx46 and designed variants are augmented following MD equilibration. a. Cut-away view of Cx46/50 GJ channel shown as tube representation with the NT domain in grey. (*inset*) Shows a zoom view of the NT domain with residue sidechains displayed at sites of genetic variability between Cx50 and Cx46, respectively. (Asterisk, ⍰) indicates location of residue 9 (Arginine in Cx46 and Asparagine in Cx50). **b.** Primary sequence alignment of the Cx50 and Cx46 NT domain (residues 2 – 19). Negatively charged residues are colored in red, positively charged residues are indicated in blue. Sites of variation between the two isoforms are indicated (dotted box). Met1 is not included, as this site is co-translationally removed and Gly2 is acetylated in both isoforms (Myers *et al.*, 2018). **c, d.** Backbone root mean squared deviation (r.m.s.d.) analysis of equilibrium (0 – 30 ns) and production phases (30–130 ns) of the MD simulations, calculated with respect to the experimental starting structures, where Cx50, Cx46-R9N and Cx46-50NT (red traces) are shown in panel c, and Cx46, Cx50-N9R and Cx50-46NT (blue traces) are shown in panel d. Separate analysis for the n-terminal (NT) domains are shown in lighter shades. **e.** Pore diameter analysis of Cryo-EM based structures of Cx50, Cx46 and the designed NT domain variants. **f.** Averaged pore diameter of Cx50, Cx46 and designed NT domain variants following MD equilibration (pore profile of the Cx46 and Cx50 starting structures from panel c are shown for comparison in grey). **g.** Bar graph summarizing the NT constriction region (*i.e.,* z-axis ± 38–48 Å) following MD equilibration. Error bars represents SD (n = 8). All models equilibrated with a minimum constriction of ~4.6 – 4.9 Å around residue S5, with the exception of Cx50-46NT which displays a significantly wider pore constriction at S5 of ~5.5 Å ($p_{50\text{-}WT}$ = 0.000217, $p_{46\text{-}R9N}$ < 0.0001, $p_{46\text{-}50NT}$ = 0.000176, $p_{46\text{-}WT}$ = 0.000468, $p_{50\text{-}N9R}$ = 0.00137).

Based on these structural comparisons, and previous MD-studies which indicated that Cx46 introduces a significant energetic barrier to K^+ permeation as compared to Cx50 that is localized to the NT domain (Myers *et al.*, 2018), we constructed four models designed to probe the mechanistic and functional role of the NT-domain and the

residue identity at the 9th position for comparative all-atom equilibrium MD-simulation and electrophysiology studies in GJ deficient N2A cells. To facilitate the most direct comparisons, all models and constructs used for experimental characterization were based on the sheep Cx46 and Cx50 structures, and correspond to NT domain swapped chimeras (Cx46-50NT and Cx50-46NT), and single point variants at the 9th residue (Cx46-R9N and Cx50-N9R). These models and/or constructs were then used to test the hypothesis that the NT domain and/or position-9 are important in defining key

differences in channel properties that exist between the closely related Cx46 and Cx50 GJs.

Position 9 does not form the primary NT domain constriction site in Cx46, Cx50 or their variants during MD simulation. Constructed models of NT domain swapped chimeras and single point variants at position-9 of Cx46 and Cx50 were built by *in silico* mutation, and the NT position (G2 in all models) was acetylated (Myers *et al.*, 2018). Pore profile analysis, using the program HOLE (Smart *et al.*, 1996), show the resulting models produced pore constriction sites as expected based on the identity of position-9 (Fig. 1e), with some differences resulting from the precise conformation selected by the modeling program for the position-9 variants.

Following all-atom MD equilibration in explicit water, the NT domains of all models appeared well behaved and maintained alpha-helical secondary structure; however, the pore profiles of all constructs were modified within the NT domains, as compared to their starting structures, primarily through reorientation of pore-lining sidechain residues (Fig. 1c-f). In contrast to the experimental starting models, the ensemble of structures obtained by MD displayed averaged steric landscapes that were all very similar to each other (Fig. 1f). Cx50, Cx46-50NT and Cx46-R9N all converged to a similar profile, with a primary constriction site diameter of ~9.2 (0.3) – 9.6 (0.4) Å that is substantially smaller than the experimental model of Cx50, and defined primarily by S5/D3 (Z-axis approximately ± 40 Å) (Fig. 1f, red traces). In comparison, Cx46, Cx50-46NT and Cx50-N9R all converged to a similar profile that is larger than the experimental model of Cx46, ranging from ~9.7 (0.3) Å for Cx46 and Cx50-N9R to ~10.9 (0.4) Å for Cx50-46NT (Fig. 1f,

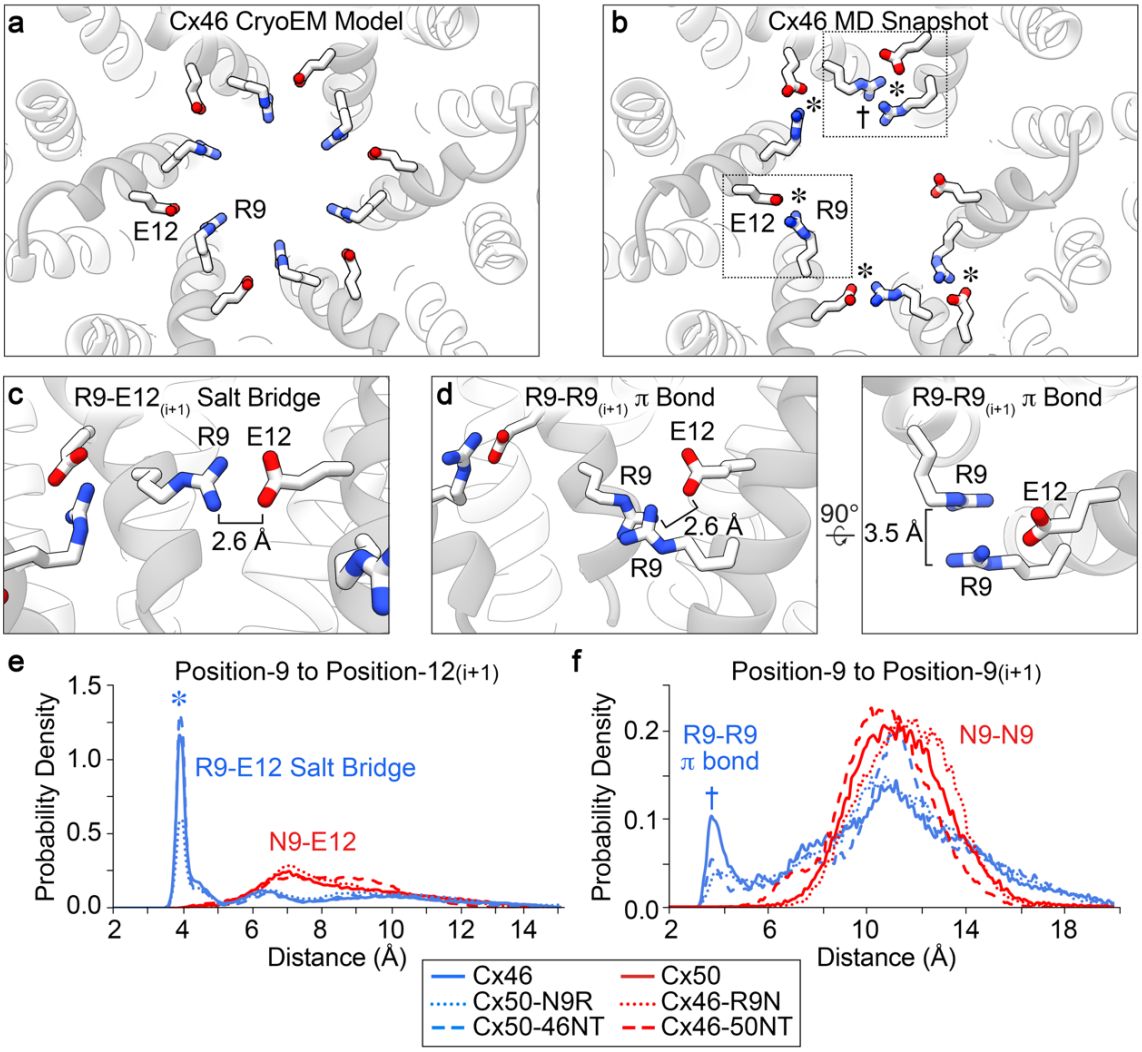


Fig. 2 Arginine-9 of Cx46 and designed variants adopts a dynamic ensemble of stable conformational states that prevent pore constriction. a. Cryo-EM based structure of Cx46 (PDB 6MHQ), viewed along the pore axis with residues R9 and E12 displayed in stick representation. R9 is oriented toward the center of the pore permeation pathway, and forming a primary constriction site. **b.** Representative snap-shot of Cx46 following MD equilibration, viewed as in panel a. R9 adopts alternative conformational states that reorient this sidechain away from the pore permeation pathway, that appear to be stabilized by salt bridge interaction with a neighboring E12 (asterisk, ⍰) and/or ⍰-⍰ interactions with R9 from a neighboring subunit (dagger, †). **c.** Zoom view (boxed in panel b) of a representative R9 salt bridge interaction with a neighboring E12 residue. **d.** Zoom view (boxed in panel b) of a representative ⍰-⍰ interaction between two R9 residues of neighboring subunits. Distances displayed in panels c,d are between nearest heavy atoms. **e.** Probability distribution of the distance between position-9 to position-12 in the neighboring subunit obtained from MD simulation for Cx46, Cx50-N9R and Cx50-46NT (blue traces, represent R9-E12$_{(i+1)}$ distances) and Cx50, Cx46-R9N and Cx46-50NT (red traces, represent N9-E12$_{(i+1)}$ distances). **f.** Probability distribution of the distance of position-9 to position-9 in the neighboring subunit obtained from MD simulation for Cx46, Cx50-N9R and Cx50-46NT (blue traces, represent R9-R9$_{(i+1)}$ distances) and Cx50, Cx46-R9N and Cx46-50NT (red traces, represent N9-N9$_{(i+1)}$ distances). Distances displayed in panels e,f were measured between R9 C⍰ or N9 C⍰ and E12 C⍰ to capture equivalent rotameric states, and are therefore greater than those displayed in panels c,d.

blue traces). Notably, the MD equilibrated diameter of Cx50-46NT is significantly wider than all other models ($p = 0.00468 – < 0.0001$) (Fig. 1f,g), which is of potential interest to channel function and discussed in further detail in the following sections.

In contrast to the Cryo-EM based models, in all related MD models, the R9 position is no longer contributing as the primary constriction site. Upon closer inspection, it is observed that R9 adopts a dynamic exchange of conformational states during MD simulation of Cx46, Cx50-46NT and Cx50-N9R, that reorients the sidechain away from the pore permeation pathway (Fig.2). These conformations include the formation of a salt bridge between the positively charged R9 sidechain and the negatively charged E12 position (heavy atom distance = 2.6 Å) of a neighboring subunit at the (i+1) position (Fig. 2b,c), which is a conserved site in all models (see Fig.1b). Amongst the three variants containing an arginine at position-9, this interaction accounts for ~30-50% of the conformational states observed by MD (Fig. 2e). An alternative, and less populated, conformational state is also observed, where R9 forms an optimal □-□ stacking interaction with guanidinium group of a neighboring R9 (planar distance = 3.5 Å) from an adjacent subunit at the i+1 position (Fig. 2b,d) (Vernon *et al.*, 2018). This state appears to be further stabilized by salt bridge formation between R9 and E12 of the same subunit (Fig. 2b,d), or between neighboring subunits (similar to the state shown in Fig. 2c). This conformation was observed in ~5-10% of the total conformational states observed by MD (Fig. 2f). In all cases, these bridging interactions effectively withdraw the large R9 sidechain away from the permeation pathway and stabilize R9 against the lumen of the channel. Equivalent

interactions were not observed in variants where position-9 is occupied by asparagine, with N9 appearing to adopt more random (fluctuating) conformational states (Fig. 2e,f). Importantly, the reorganization and dynamical behavior of the NT domain sidechain conformations observed during MD simulation are consistent with the experimental Cryo-EM data that was used to build the atomic starting models. This is because, in contrast to the hydrophobic NT anchoring sites, the conformational states of the pore-lining NT domain residues, especially position-9, were not well defined by experimental Cryo-EM density (presumably reflecting the specimen heterogeneity and/or intrinsic dynamic behavior of these residues) (Myers *et al.*, 2018).

Position 9 defines a key energetic difference in K^+ ion permeation in Cx50 and Cx46, and their mutational variants. Given the similarities between Cx50 and Cx46 pore constriction sites observed by MD, we next sought to understand how postion-9 and the NT domain may contribute to differences in their respective energetic barriers to ion permeation by evaluating the potential of mean force (PMF) for Cl^- and K^+, which can be extracted from the MD data to describe the free-energy landscapes of these ions along the permeation pathway (*i.e.,* the Z-axis) (Fig. 3). It is noted that inspection of the Coulombic surface of the Cx50 starting model shows that this isoform lacks any positive charge within the pore, when the NT is acetylated (Fig. 3a). In contrast, Cx46 possesses a prominent ring of positive Coulomb potential, belonging to R9, that lines the cytoplasmic vestibules at both ends of the channel (Z-axis approximately ± 50 Å) (Fig. 3a). The Coulombic surface

potentials of the resulting mutation models are similar to their wildtype counterparts based on the presence or absence of an arginine at the 9th position (not shown). Consistent with their Coulombic surface properties, the MD-derived PMFs for Cl^- showed a large peak energetic barrier, ΔG_{Cl^-} ~5.1 kcal mol^{-1} for both Cx50 and Cx46 (defined as the difference between bulk solvent and peak PMF barrier to Cl^-), which is at least twice as

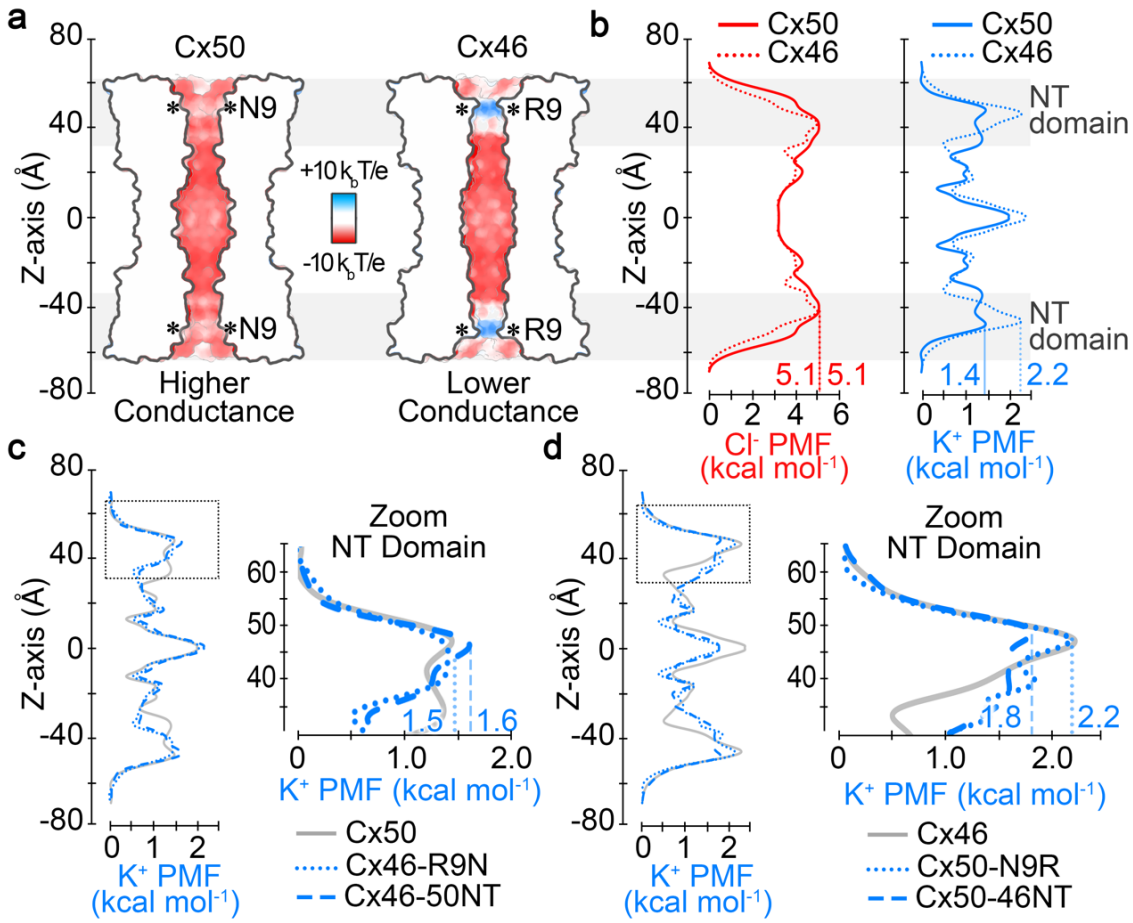


Fig. 3 Electrostatics and energetics of the ion permeation pathways of Cx50, Cx46 and designed NT domain variants. a. Cut-away surface representation of Cx50 and Cx46, colored by Coulombic potential (red, negative; white, neutral; blue, positive). *T*, temperature; *k*, Boltzmann constant; *e*, charge of an electron. **b.** Potential of mean force (PMF) describing the free-energy landscape (Δ*G*) experienced by Cl^- ions (red traces) and K^+ ions (blue traces) permeating the channel pore of Cx50 and Cx46, as compared to the bulk solvent. **c.** PMFs of K^+ ions for designed variants Cx46-R9N and Cx46-50NT, with PMF for Cx50 displayed for comparison (gray trace). (*inset*) Shows a zoom view of K^+ PMFs corresponding to the NT domain in panels c (dotted box region). **d.** PMFs of K^+ ions for designed variants Cx50-N9R and Cx50-46NT, with corresponding PMF for Cx46 displayed for comparison (gray trace). (*inset*) Shows zoom view of K^+ PMFs corresponding to the NT domain in panels f (dotted box region). PMFs were symmetrized along Z-axis, and peak energetic barriers identified within the NT domains are indicated. Asterisk (□) in panel a indicates location of position 9. Grey box in panels a,b indicate the region of the NT domain. **e.** Bar graph showing the peak energy barrier to K+ permeation for all variants. Error bars represents SD (n = 8). Cx50 and Cx46 display significant differences ($p < 0.0001$), and both Cx50 and Cx46 are significantly different from Cx50-46NT ($p = 0.000392$ and 0.00481, respectively).

high as the peak energetic barriers to K^+ (Fig. 3b). The primary energetic barrier to both Cl^- and K^+ are established within the NT-domain region of the channels, and are consistent with this domain acting as the selectivity filter and with the displayed preference for cations that has been previously demonstrated experimentally for both Cx50 and Cx46 (Trexler *et al.*, 1996; Srinivas *et al.*, 1999; Sakai *et al.*, 2003; Tong *et al.*, 2014). For these reasons, we have focused the following comparative analysis to the energetic differences between Cx50, Cx46 and the designed variants to their energetics of K^+ (the major permeant ion).

The most distinct energetic differences between Cx50 and Cx46 correspond to the region of the NT domain that aligns with position-9, where Cx46 has a peak energy barrier to K^+ of ~2.24 (0.24) kcal mol^{-1} and Cx50 has a peak barrier of ~1.42 (0.08) kcal mol^{-1}, values that agree with our previous study using slightly modified simulation conditions (see Methods) (Myers *et al.*, 2018). The difference in peak K^+ PMF barriers in these two models

($\Delta\Delta G_{K+}$ = 0.82 kcal mol^{-1}; $p < 0.0001$) is expected to be primarily due to the positive charge characteristics of R9, and not local differences in channel diameter (as described in the previous section).

The K^+ ion PMF profiles for the designed variants show that replacing the entire NT domain of Cx46 with that of Cx50 (Cx46-50NT) or the 9th residue (Cx46-R9N) substantially reduced/eliminated the differences of NT domain energy barriers, as compared to Cx46 ($\Delta\Delta G_{K+}$ = 0.03 – 0.2 kcal mol^{-1}), by effectively reducing the peak barrier around the 9th position (Fig. 3c,e). Likewise, replacing the entire NT domain of Cx50 with that of Cx46 (Cx50-46NT) or the 9th residue (Cx50-N9R) also reduced the energy difference between Cx46 and Cx50 (ΔG_{K+} = 0.04 – 0.4 kcal mol^{-1}), however, this time by increasing the peak

barrier to a similar level of Cx46 (Fig. 3d,e). Notably, although the K^+ energy barrier obtained for Cx50-46NT (ΔG_{K+} = 1.82 (0.17) kcal mol^{-1}) is significantly larger than wildtype Cx50 ($p = 0.000392$), it remains lower than that of Cx46 ($p = 0.00481$) and Cx50-N9R ($p = 0.0100$) (Fig. 3e). This difference is proposed to reflect the larger pore diameter of this model, resulting in a less constrictive environment for ion permeation (see Fig. 1f,g). Taken together, these data confirm that the primary difference in K^+ energy barriers between Cx46 and Cx50 is due to the placement of a positively charged arginine at the 9th position and led to the hypothesis that swapping these sites/domains would effectively convert the ion conductance properties of these channels, which was tested by the experiments described in the following sections.

Cx46-50NT forms non-functional gap junction channels and is reflected by steric clash and NT-domain instability observed by MD-simulation. In order to integrate the results of our structural models and MD studies to GJ channel properties, we generated sheep Cx46, Cx50, the NT domain swapped chimeras (Cx46-50NT and Cx50-46NT), and point variants at the 9th residue (Cx46-R9N and Cx50-N9R) and expressed each of them in GJ-deficient N2A cells for functional characterization. N2A cells were transfected with expression vectors containing one of our designed constructs and an untagged GFP (*e.g.,* Cx46-IRES-GFP, Cx46-R9N-IRES-GFP, Cx46-50NT-IRES-GFP, Cx50-IRES-GFP, Cx50-N9R-IRES-GFP or Cx50-46NT-IRES-GFP) (Fig. 1a,b). Cell pairs with positive expression of GFP were voltage clamped using dual whole cell patch clamp technique. Approximately half of the cell pairs expressing Cx46 or Cx50 showed junctional currents (I_js) in response to a V_j pulse (Fig. 4a,b), indicating successful formation of functional GJs. Cell pairs expressing the chimera, Cx50-46NT or one of the point variants (Cx46-R9N and Cx50-N9R) were also frequently coupled with similar coupling percentages as those of wildtype lens connexins (Fig. 4a,b). However, none of the cell pairs expressing Cx46-50NT showed any I_js in 6 independent experiments (Fig. 4a,b). The average coupling conductance (G_j) for each expressed construct was calculated from the GJ coupled cell pairs and no significant differences were observed among different constructs with the exception of Cx46-50NT (Fig. 4c).

To test if the failure of Cx46-50NT to form functional GJs was due to impairment in the localization of this chimera, we used GFP fusion tagged at the carboxyl terminus of Cx46-50NT. As shown in Fig. 4d,e, Cx46-50NT-GFP was localized in intracellular compartmen

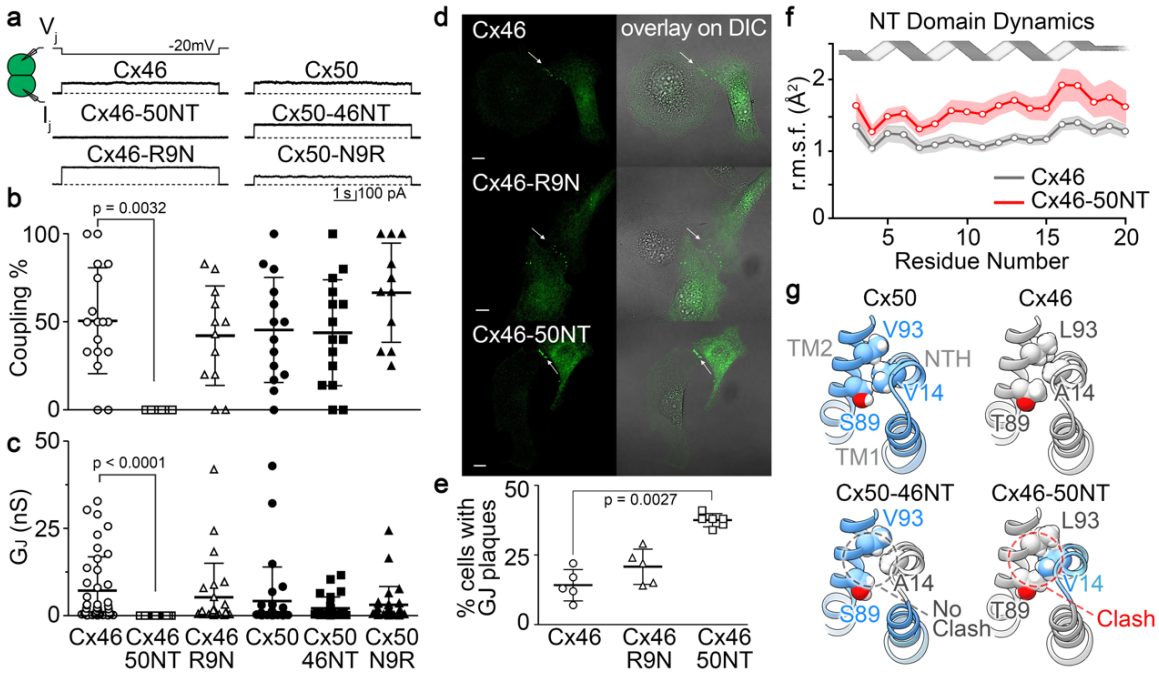


Fig. 4 GJ channel function, cellular localization and NT domain stability of Cx50, Cx46 and designed NT domain variants. a. Dual whole-cell patch clamp technique was used to measure junctional current (I_j) in N2A cell pairs expressing GFP-tagged constructs of sheep Cx50, Cx46 and NT domain variants (Cx50-N9R, Cx50-46NT, Cx46-R9N and Cx46-50NT) in response to a -20 mV V_j pulse. **b.** Bar graph showing the average coupling percentages of cell pairs expressing homotypic sheep Cx50, Cx46 and designed NT domain variants. One-way ANOVA followed by Newman-Keuls post-hoc test was used compare each of the variants with their respective controls. Cx46-50NT showed no coupling, which was significantly different from that of wildtype Cx46 ($p = 0.0032$). The number of transfections is indicated. **c.** Bar graph showing the average coupling conductance (G_j) of coupled cell pairs. Kruskal-Wallis followed by Dunn's post hoc test was used to compare each of the variants with their respective controls. The total number of cell pairs is indicated. **d.** Cell imaging of GFP-tagged Cx46, Cx46-R9N and Cx46-50NT expressed in connexin deficient HeLa cells. GJ plaque-like structures (arrows) similar to that of Cx46-GFP (upper panels) are identified for both Cx46-R9N and Cx46-50NT variants. GFP fluorescent signals are superimposed onto DIC images to show the localization of these fluorescent signals (panels on the right). Scale bars = 10 µm. **e.** Bar graph showing the percentage of cells expressing each of the constructs displaying GJ plaque-like structures at the cell-cell interfaces. Note that the Cx46-50NT showed a significantly higher percentage to observe GJ plaque-like structures ($p = 0.0027$). **f.** Line graph showing NT domain dynamics obtained by MD simulation, as assessed by the backbone root mean square fluctuation (r.m.s.f.) for Cx46 (gray trace) and Cx46-50NT (red trace). Shaded boundaries indicate the 95% confidence intervals (n = 12). Residues 3 – 20 show significant difference between these models ($p = 0.000273$). **g.** Structural comparison of hydrophobic packing interactions involving residue 14 on the NT domain and residues 89 and 93 on TM2 of Cx50, Cx46 and chimeric models for Cx50-46NT and Cx46-50NT. The Cx46-50NT chimera introduces a steric clash involving bulky residues V14 and T89.

and displayed GJ plaque-like clusters at the cell-cell interfaces with a higher percentage than that of Cx46-GFP, suggesting that it is unlikely due to abnormal localization of this

chimera for its failure in forming functional GJs. For comparison, Cx46-R9N-GFP showed a similar percentage of forming GJ plaques as that of Cx46-GFP (Fig. 4d,e).

To gain insight into the molecular basis for the loss of GJ channel function in the Cx46-50NT chimera, we further interrogated the MD simulation data for this construct. Notably, the dynamical behavior of the NT domain in this construct is significantly higher compared to the wildtype Cx46 model, as assessed by the backbone root-mean-square-fluctuation (r.m.s.f.), which describes the amplitude of backbone dynamics (Fig. 4f; $p = 0.000273$ over residues 3 – 20). In comparison, the NT r.m.s.f. values of wildtype Cx50 were not significantly different from the Cx50-46NT model. The reason for the decreased stability of the NT domain in the Cx46-50NT chimera is suspected to be the introduced changes in hydrophobic anchoring residues that dock against the TM domains (Fig. 1a,b). Specifically, placement of the Cx50 NT domain onto the Cx46 channel introduces an apparent steric clash between V14 and T89 (located on TM2) (Fig. 4g). In Cx50, position 89 is occupied by a small serine residue, which can accommodate a bulky V14 anchoring residue. In contrast, Cx46 appears to compensate for the smaller A14 anchoring site with a bulkier T89 residue. Of further comparison, when the Cx46 NT domain (containing the smaller A14 site) is placed onto the Cx50 channel (containing the smaller S89 site), a vacant space is introduced between the NT domain and TM2 of the Cx50-46NT chimera (Fig. 4g). This arrangement allows the NT domain to pack more closely against TM2 during MD simulation, resulting in the larger pore diameter of this construct as compared to Cx46, Cx50 and other variant models (see Fig. 1f,g). It is noted, that the NT r.m.s.f. of this

chimera was also significantly lower (*i.e.,* more stable) as compared to the non-functional Cx46-50NT variant ($p = 0.00526$; over residues 3 – 20).

Swapping the NT-domain or the 9th residue of Cx46 onto Cx50 alters V_j-gating, while V_j-gating of Cx46-R9N is unaffected. To investigate the effects of NT domain variants on transjunctional voltage dependent gating (V_j-gating), cell pairs forming homotypic GJs were recorded by dual whole cell patch clamp and their I_js were measured in response to a series of V_j pulses from ±20 to ±100 mV (Fig. 5a,b). The I_js of Cx46 and Cx50 GJs both showed symmetrical V_j-dependent deactivation, with similar V_j-gating properties (Fig. 5a). For both wildtype GJs, I_js in response to V_js in the range of ±40 to ±100 mV, showed strong deactivation (Fig. 5a). When V_j absolute value was ≤ 20 mV, I_js showed no deactivation (Fig. 5a). The normalized steady state conductance ($G_{j,ss}$) was plotted as a function of V_j (Fig. 5b), which could be well fitted by a Boltzmann equation (Eq. 1) for each V_j polarity, for either Cx46 or Cx50 GJs (Fig. 5b). None of the Boltzmann fitting parameters of Cx46 and Cx50 were significantly different (Table 1).

V_j-gating of Cx46-50NT, Cx46-50R9N, Cx50-46NT, Cx50-N9R GJs were then studied using the same V_j protocol (Fig. 5c,d). As shown in the Fig. 5c, Cx46-50NT was unable to form functional GJs, while Cx46-R9N, Cx50-46NT, and Cx50-N9R all successfully formed functional GJs. Among these functional GJs, Cx46-R9N showed strong symmetrical deactivation when V_js were in the range of ±40 to ±100 mV (Fig. 5c). The $G_{j,ss} – V_j$ plot for Cx46-R9N could be well fitted with the Boltzmann equation for each V_j polarity, which showed no significant difference from those of wildtype Cx46 GJs (Fig. 5d) and Table 1.

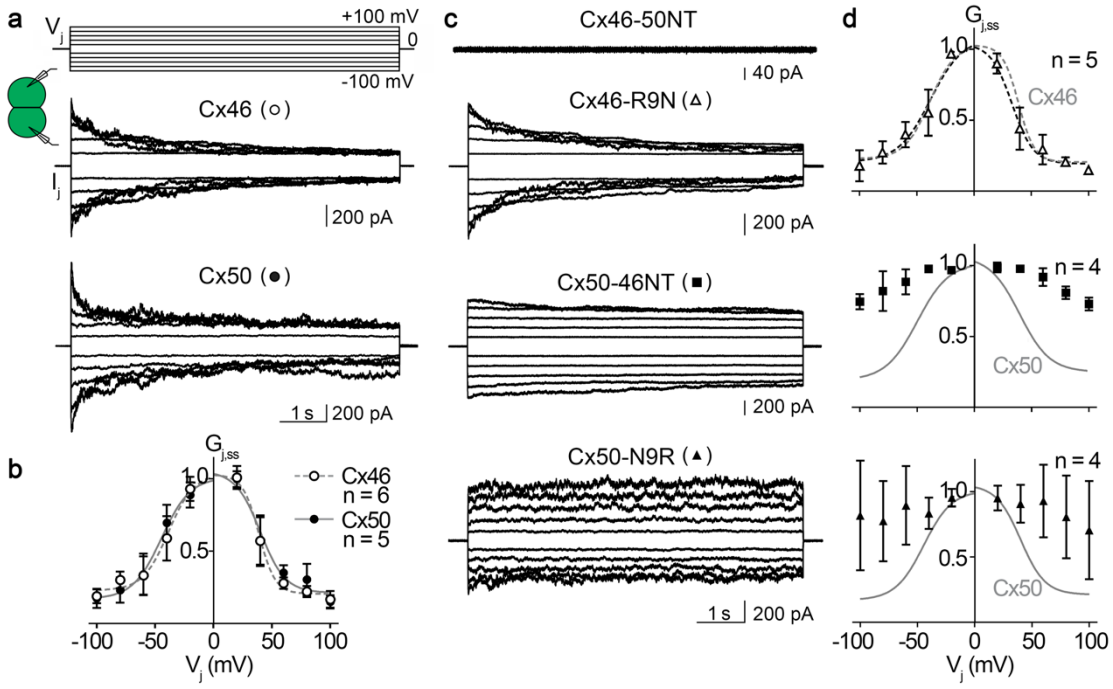


Fig. 5 V_j-gating of homotypic sheep Cx46, Cx50 and NT domain variant gap junction channels. a, c. Superimposed junctional currents (I_js) recorded from cell pairs expressing homotypic sheep Cx46 or Cx50 GJs (panel a) and NT domain variants Cx46-50NT, Cx46-R9N and Cx50-N9R GJs (panel c), in response to a series of V_j pulses (shown on the top of panel a, ±20 to ±100 mV with 20 mV increment). **b.** Normalized steady state junctional conductance, $G_{j,ss}$, of Cx46 (open circles) and Cx50 (filled circles) plotted as a function of Vjs. Boltzmann equations (Eq. 1) were used to fit $G_{j,ss} – V_j$ plots for Cx46 (smooth dashed grey lines) and Cx50 (smooth solid grey lines) GJs. **d.** Normalized steady state junctional conductance, $G_{j,ss}$, of Cx46-R9N (open triangles), Cx50-46NT (filled squares), and Cx50-N9R (filled triangles) plotted as a function of Vjs. Only the $G_{j,ss} – V_j$ plot of Cx46-R9N GJ was fitted well with Boltzmann equations (smooth dashed black lines). For comparison, the Boltzmann fitting curves of wildtype Cx46 (smooth dashed grey lines) or Cx50 GJ (smooth solid grey lines) were superimposed on the respective plot. The number of cell pairs for each construct is indicated in panels b and d.

The I_js of Cx50-46NT GJs showed very weak deactivation in the tested V_js (Fig. 5c), and the $G_{j,ss} – V_j$ plot of these channels could not be fitted with a Boltzmann equation (Fig. 5d). Similarly, I_js of Cx50-N9R GJ also showed no consistent deactivation in the tested V_js (Fig. 5c), and the $G_{j,ss} – V_j$ plot also could not be fitted with a Boltzmann equation for these channels (Fig. 5d). In addition, the I_js of Cx50-N9R showed a lot of fluctuations throughout the V_j pulse (Fig. 5c) resulting in large variations in the $G_{j,ss}$ (Fig. 5d). Taken together, these data indicate that the V_j-response for variants of Cx50 (Cx50-46NT and Cx50-N9R) are

more susceptible to perturbation than the Cx46 variant (Cx46-R9N). A structural basis for these differences was not entirely clear, as the millisecond-second long time-scales of V_j-gating are beyond the limits of traditional MD simulation (see Discussion).

Differences in Cx50 and Cx46 GJ single-channel conductance properties are defined by the 9th residue. MD simulation studies indicated that the incorporation of the positively charged arginine at the 9th position of Cx46, Cx50-N9R and Cx50-46NT introduces a significant energetic barrier to K^+ ion permeation (the major permeant ion of Cx46/Cx50 GJs), as compared to Cx50 and Cx46-R9N (see Fig. 3). To investigate the functional effects of this positively charged residue we characterized the unitary conductance (γ_j) of Cx50, Cx46 and the designed NT domain variants based on single channel current (i_j) recordings at varying V_j values. In Fig. 6a, homotypic sheep Cx46 and Cx50 single channel current (i_j) traces are shown at the V_js indicated. All point histograms and Gaussian fits were used to measure the amplitudes of i_js for the main open state at the tested V_js (Fig. 6b). The averaged i_js were plotted at different V_js and a linear regression $i_j - V_j$ plot for Cx46 or Cx50 GJs was used to estimate the slope unitary channel conductance (γ_j) (Fig. 6c). The γ_j of Cx46 GJ is 166 (14) pS, whereas the γ_j of Cx50 GJ is 208 (14) pS.

Fig. 6d shows i_js of homotypic Cx46-R9N, Cx50-N9R, and Cx50-46NT GJs at the indicated V_js. The averaged i_js for the main open state for each tested V_j of these variants were plotted to create an $i_j - V_j$ plot. Linear regression of $i_j - V_j$ plot for each variant was used to estimate slope γ_j (Fig. 6e). Cx46-R9N GJs showed a significant increase in slope γ_j (261 (23) pS) from that of wildtype Cx46 GJ (Fig. 6e). Conversely, Cx50-N9R GJs displayed

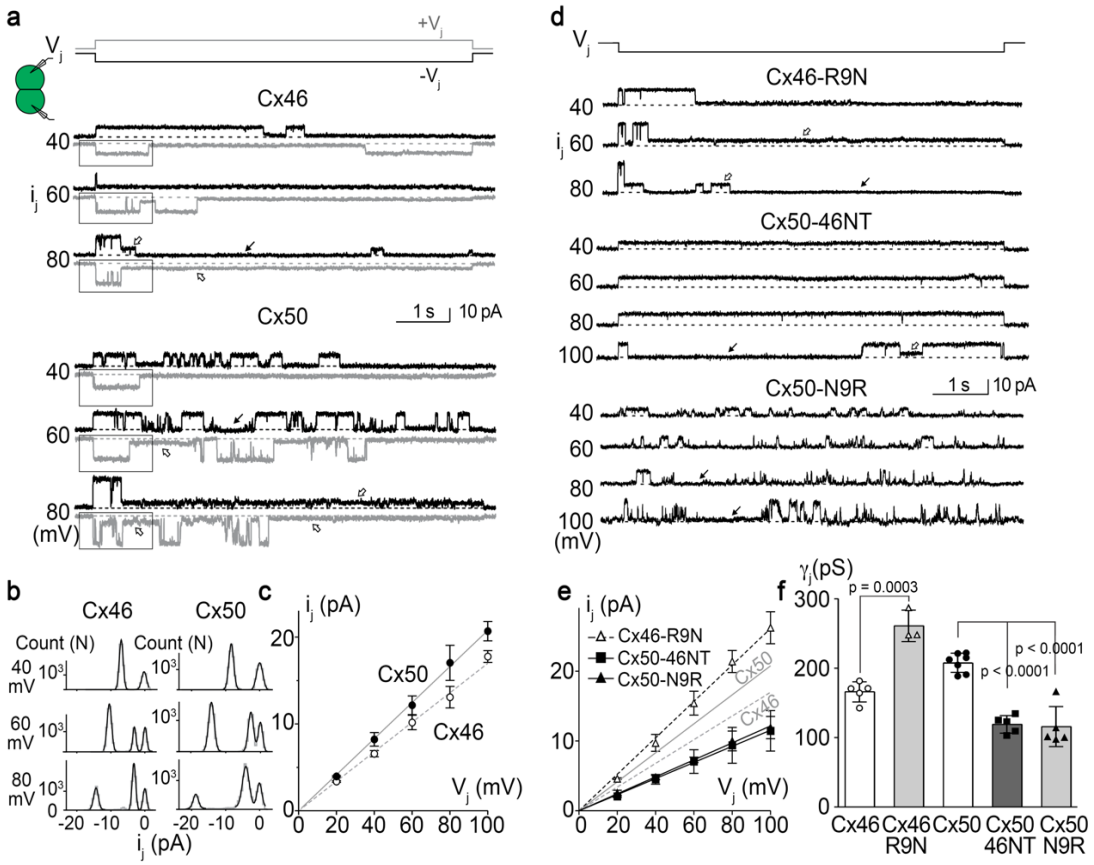


Fig. 6 Single channel properties of homotypic sheep Cx46, Cx50 and NT-domain variant GJs. a. Single channel currents (i_js) recorded from a cell pair expressing wildtype sheep Cx46 (top set) or Cx50 (bottom set) at the indicated V_js. Positive V_j ($+V_j$) or negative V_j ($-V_j$) induced i_js are shown in grey or black, respectively. Both GJs showed fully open state at the beginning of V_j pulses and transitions to either subconductance (open arrows) or fully closed state (solid arrows). Note that more than one subconductance states could be identified at ±80 mV V_js. **b.** All point histogram and Gaussian fits were used to estimate the i_j amplitude at the main open state for both V_j polarity. A selected section as shown in panel a was used to generate all point histogram. Main conductance state and in some cases a subconductance state could be identified at different V_js and averaged from both V_j polarities. **c.** The i_j amplitudes of the main open state were plotted at each tested V_j. The slope of the linear regression lines in the i_j – V_j plot represent the slope unitary channel conductance (γ_j). For Cx46 GJ, the slope γ_j = 166 (14) pS (n = 5) and for Cx50, the slope γ_j = 208 (14) pS (n = 7). **d.** Single channel currents (i_js) recorded from cell pairs expressing sheep Cx46-R9N, Cx50-N9R, or Cx50-46NT at the indicated V_js. The i_js of Cx46-R9N showed a main open state at the beginning of the V_j pulses and transitioned into subconductance state (open arrows) or fully closed state (solid arrow). The i_js of Cx50-46NT preferentially reside at fully open state at most tested V_js (-40 to -80 mV) and transition to fully closed state (solid arrow) or to a subconductance state could be observed at -100 mV V_j. The i_js of Cx50-N9R showed multiple open events with frequent transitions to fully closed state (solid arrows). **e.** The amplitude of i_js (the same way as described in panel b) for the main open state at each tested V_j was plotted against V_j and analyzed as described for panel c. The solid and dashed grey lines represent the slope γ_js of Cx50 and Cx46 respectively. The slope γ_j of Cx46-R9N is 261 (23) pS (n = 3), Cx50-N9R γ_j is 116 (29) pS (n = 5), and Cx50-46NT γ_j is 119 (13) pS (n = 5). **f**. Bar graph to summarize the slope γ_js of all tested variants. The γ_j of Cx46-R9N was significantly higher than that of Cx46 and the γ_js of Cx50-46NT and Cx50-N9R were significantly lower than that of Cx50 (one-way ANOVA followed by Tukey's post hoc test biological meaningful groups). The total number of cell pairs for each variant is shown.

significantly decreased slope γ_j (116 (29) pS) than that of wildtype Cx50 GJ (Fig. 6e). Replacing the NT domain of Cx50 with that of Cx46, Cx50-46NT, also resulted in a pronounced drop in slope γ_j (119 (13) pS). The slope single channel conductance values of all tested variants were plotted as a bar graph for direct comparison (Fig. 6f). These effects are qualitatively consistent with the predicted results obtained by MD, however there was a surprisingly higher degree of effect of the placement/removal of R9 leading to channel variants with correspondingly higher or lower unitary conductance, as compared to wildtype channels.

Cx50-46NT forms GJ channels with an extremely stable open state, while Cx50-N9R open state becomes destabilized. Another surprising result revealed by single channel recordings was the effects on the open state probability of the variant GJs. As shown in Fig. 6d, the i_js of Cx50-46NT were virtually in the fully open state (or probability of open, $P_{open} \approx 1$) at all V_js except at 100 mV, whereas i_js of Cx50-N9R show a much lower P_{open} at all tested V_js (P_{open} in the range of 0.11 - 0.22). The i_js of both Cx50-46NT and Cx50-N9R showed little change in P_{open} in the tested V_js (data not shown).

Upon close inspection of the i_js of Cx46, Cx50 and their variants, we found that the open dwell times were different among different GJs. To quantify open dwell time for each of these GJs, the i_js of these GJs were further analyzed to obtain open dwell time for each open event (Fig. 7a). At V_j of -60 mV, a limited number of open events (transitions from closed to open and return to the closed state) could be identified for Cx50-46NT during a 7-second V_j pulse. A few open events for Cx46 and Cx46-R9N GJs and many more open

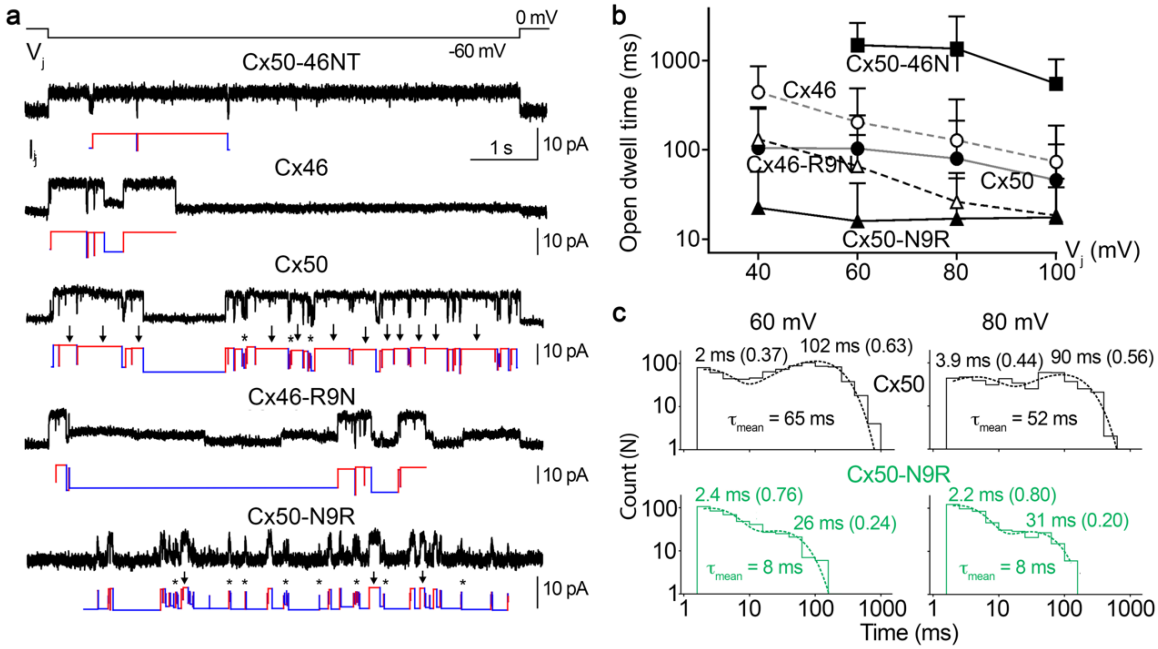


Fig. 7 Single channel open state dwell times of Cx50, Cx46 and NT domain variants. a. The i_js (black) and schematic single channel openings (red) and closings (blue) are shown for sheep Cx50-46NT, Cx46, Cx50, Cx46-R9N, Cx50-N9R GJs. Cx50 and Cx50-N9R GJs showed apparently short (asterisks, *) and long (arrows) open events, as well as more open events than the other GJs. Note that the first and the last open events were excluded as they likely under represent the open dwell time (*e.g.* exist in open state before or after the V_j pulse). Our V_j-pulse duration (7 s) may not be sufficient to reach a true steady state and the open event properties might be slightly different from the fully stable i_js of the studied GJs. **b.** The averaged open dwell time for each GJ were plotted at different V_js. The number of events analyzed ranged from 33-832. Note that the open dwell time is plotted in logarithmic scale to show large differences among these GJs. The ranking of the open dwell times of these GJs from high to low is: Cx50-46NT > Cx46 > Cx50 > Cx46-R9N > Cx50-N9R. **c.** The open dwell time for all open events for Cx50 and Cx50-N9R were plotted in a logarithmic histogram with 5 bins/decade and fitted with two-exponential log components. The two time constant (□□ values represent the corresponding short and long open events for Cx50 (top histograms) and Cx50-N9R (bottom green histograms) are displayed, with their relative proportions in parentheses. The □$_{mean}$ values were calculated by taking the weighted average of the two □ values using the area under each peak for each of these variants at a specific V_j.

events for Cx50 and Cx50-N9R GJs were detected at the same V_j (Fig. 7a). The open dwell time for all events was averaged for each variant GJ (Fig. 7b). In most of the tested V_js, the rank of the open dwell time from long to short is in the following order: Cx50-46NT > Cx46 > Cx50 > Cx46-R9N > Cx50-N9R GJs (Fig. 7b).

It is interesting to note that open event frequency was much higher for both Cx50 and Cx50-N9R GJs and both GJs showed two types of openings, short (labeled by asterisks)

and long (labeled by arrows) opening events (Fig. 7a). Histograms of open dwell time were generated for both of these GJs (Fig. 7c) and each of them could be fitted with two-exponential components. Cx50 showed short (2 ms) and long (102 ms) open events at V_j 60 mV with a higher prevalence of the long open events (63%) (Fig. 7c). At V_j of 80 mV, the average of long open events became shorter (90 ms) with slightly lower percentage (56%), while the average of the short open events is 3.9 ms with an increased percentage (44%). Weighted average of time constants (τ_{mean}) for these two V_js were also calculated (Fig. 7c). Upon mutating the 9th residue to arginine (Cx50-N9R) the long open events became much shorter (26 ms and 31 ms) with reduced percentages (24% and 20%), respectively for V_js of 60 and 80 mV. The Cx50-N9R mean value was much lower than that of Cx50. Taken together, these results suggest that the identity of the 9th residues, in addition to other structural differences within the NT domains of these variants significantly influence the open state stability of these GJs (see Discussion).

Discussion

In this study, structural modeling and molecular dynamics simulations were conducted in combination with macroscopic and single channel electrophysiology studies, to investigate the mechanistic roles of NT domain in two closely related lens GJs, Cx50 and Cx46. We identified differences in structure and dynamics within the NT domains, as the key features differentiating the pore open stability and energetic barriers to ion permeation between these two isoforms (see summary of results presented in Fig. 8). In the following, we discuss the mechanistic insights gained by these results and, in

comparison to previous studies, as well as some unresolved discrepancies between the simulation and experimental data that point to the importance of on-going development in our atomic-level understanding of GJ conductance, V_j-gating and open state stability.

Hydrophobic stabilization of the GJ open state

The mechanistic basis of GJ open state stability has not yet been well defined. All-atom equilibrium MD simulations conducted on wildtype channels and NT domain variants, based the Cryo-EM models of Cx46 and Cx50, all maintained the novel open state conformation that is stabilized by inter- and intra-subunit hydrophobic interactions (a mechanism to anchor the NT domains in an open conformation) that are distinct from the open state originally described for Cx26 (Maeda *et al.*, 2009; Myers *et al.*, 2018). The importance of hydrophobic anchoring residues within the NT domain is further supported by the chimeric construct Cx46-50NT, which displayed significantly destabilized NT domain dynamics as compared to Cx46 ($p = 0.000273$) and this behavior was functionally correlated with a loss of channel activity for this variant (Fig. 8 and see Fig. 4f,g). Remarkably, the non-functional Cx46-50NT GJs appeared to be able to reach the cell membrane and formed morphological GJs (see Fig. 4d,e), indicating that the loss of channel function was unlikely due to defects in trafficking or GJ assembly. Possible mechanistic insights to this phenotype are provided by our structural models, where this chimera introduces a specific steric clash between the introduced hydrophobic residue V14 on the NT domain and T89 located on TM2 (Fig. 8 and see Fig. 4g). Taken together, these data suggest that the NT domain of Cx46-50NT becomes destabilized, likely by steric interactions involving hydrophobic anchoring sites, resulting in loss of channel function

(Fig. 8). While the NT domain does not completely unfold during the relatively short time scale of our MD simulation, improper folding of the Cx50 NT domain in the context of this chimera can also not be ruled out.

In contrast to the Cx46-50NT chimera, placement of the Cx46 NT domain onto Cx50 resulted in more close hydrophobic packing (due to smaller hydrophobic residues at positions 14 and 89), which resulted in an overall wider pore diameter for this construct, as compared to wildtype and the other designed GJs ($p = 0.000468 – < 0.0001$). We suggest that this close packing results in a more stabilized NT domain and contributes, at least in part, to the increased open dwell time and reduced V_j-gating of this GJ. Quantitative measurements of open dwell time of the Cx50-46NT GJ showed they were about 5 – 15 times longer than those of Cx46 and Cx50 GJs (see Fig. 7). It is interesting to note that in previous studies on switching NT domains of orthologous Cx50 and Cx46 GJs and hemichannels, the authors also showed a decrease in unitary conductance and an increased open dwell time in Cx50-46NT as compared to wildtype (Tong *et al.*, 2004; Tong & Ebihara, 2006; Kronengold *et al.*, 2012). Surprisingly, the hemichannel of Cx46-50NT was functional (unlike our result that sCx46-50NT was unable to form functional GJs) with a slightly higher hemichannel unitary conductance (Kronengold *et al.*, 2012). The basis for this functional difference is not apparently clear, as the NT domain and position 89 (the site of steric clash in our Cx50-46NT model) are conserved between sheep and rodent. However, other differences in experimental conditions that may influence NT domain stability cannot be ruled out, such as cell expression system (*Xenopus* Oocyte versus N2A cells) and/or GJ versus hemichannel assemblies.

More broadly, these results reinforce the proposal that hydrophobic anchoring sites within the NT domain play an important functional role in stabilizing the open state conformation resolved in the Cx46/50 Cryo-EM structure (Myers *et al.*, 2018), and likely play a similar functional role in other GJs based on conservation of hydrophobic positions found in other isoforms. This interpretation is consistent with previous studies. For example, the congenital cataract-associated L11S mutation in Cx46 has been show to result in proper trafficking to the plasma membrane, but showed no GJ or hemi-channel activity (Tong *et al.*, 2013). The authors contextualized these results based on the available Cx26 structure; however, we believe that has led to misleading interpretation of this mutation (and other previous mutational studies at these hydrophobic sites), as the hydrophobic sites in the Cx26 GJ structure are displayed toward the permeation pathway of the channel (Maeda *et al.*, 2009).

We suggest the conformational state of the NT captured in Cx46/50, and proposed role of the hydrophobic residues within this domain in stabilizing the open state, are congruent with mutational studies in other isoforms as well. In Cx26, an M34T mutation associated with hereditary deafness significantly reduces channel conductance and has been proposed to disrupt the open stability of the channel via interactions with W3 in the NT domain of Cx26 (equivalent to W4 in Cx46/50) (Maeda *et al.*, 2009; Zonta *et al.*, 2014). The similar M34A mutation in Cx26 results in non-functional channels and has been shown structurally to disrupt the open state conformation of Cx26, inducing an apparent closed-state conformation with a plug density blocking the pore permeation pathway (Oshima *et al.*, 2007; Oshima *et al.*, 2011). Notably, position L35 in Cx46/50 (equivalent

to position M34 in Cx26) also interacts with W4, although the details of this interaction are different in the context of the more stabilized NT domain conformation (Myers *et al.*, 2018). Nevertheless, mutation at this site would predict a similar disruption to the open state NT conformation. In Cx32, a series of mutations have been conducted that scan the functional effect of replacing NT domain residues with a negatively charged aspartate (Purnick *et al.*, 2000). Remarkably, replacement of hydrophobic residues, such as W3D, L6D, L9D or L10D (equivalent to the NT hydrophobic anchoring residues 4, 7, 10 and 11 in Cx46/50), all resulted in non-functional GJs, while mutations performed at other non-hydrophobic sites produced functional channels with altered conductance/gating properties (Purnick *et al.*, 2000).

These correlations all support the over-arching proposal that the open state stability of GJs is driven by inter- and intra-subunit hydrophobic interactions between the NT and TM1/TM2, while genetic mutations that disrupt these hydrophobic packing interactions may lead to significant destabilization of the NT domain, resulting in non-functional channel formation with pathological consequences of disease. We note, however, that the molecular details involved in stabilizing the open-state are likely to be unique among various connexin isoforms, resulting from differences in inter-domain interactions between the NT and TM1/TM2 domains and physiological conditions. Indeed, recent Cryo-EM structures of Cx26 GJ channels (Khan *et al.*, 2020) and Cx31.3 hemi-channels (Lee *et al.*, 2020) resolve NT domains that adopt different orientations as compared to Cx46/50. However, differences in experimental conditions among these studies make direct comparisons challenging (*e.g.,* the presence of lipids inside the pore of Cx31.1).

Role of NT and the 9th residue in V_j-gating of Cx46/50

Both Cx46 and Cx50 GJs displayed prominent V_j-gating in the range of tested V_js (±20 to ±100 mV) that could be well described by a two state Boltzmann equation for each V_j polarity. A classic voltage-gating model for voltage-gated ion channels is that during the gating process, a change in the voltage results in movement of charge or a re-orientation of dipoles (the voltage sensor) relative to the electric field, which leads to either closing or opening of the channel (Bezanilla, 2000). In the case of several GJ channels, the sensor for V_j-gating is believed to reside in the pore lining residues (Harris *et al.*, 1981; Bukauskas & Verselis, 2004; Bargiello & Brink, 2009). In Cx32, charge substitution up to the first 10 residues of the NT domain have been shown to influence V_j-sensing (Purnick *et al.*, 2000). Charge reversal of E9/E13 of Cx40 has also been shown to alter V_j-sensing (Musa *et al.*, 2004). However, reversal of both K9/K13 of Cx43 showed no effect on V_j-sensing, but did alter residual conductance properties (Musa *et al.*, 2004). We show here that removing a positively charged residue in the pore lining NT domain of Cx46 by the variant R9N, showed no change in the V_j-gating sensitivity (*A*, a Boltzmann parameter can be converted into gating charge, z) and in the half deactivation V_j (V_0). These results indicate that 1) the positive charge on R9 is unlikely to play a role in V_j-sensing or V_j-gating in Cx46 GJs, and 2) there is no change in the free energy difference ($\Delta G_0 = zFV_0$) between aggregated open and closed states in the absence of a voltage field (Yifrach & MacKinnon, 2002; Sukhareva *et al.*, 2003; Xin *et al.*, 2012). We suggest that the lack of a voltage-sensing role for R9 in

Cx46 may be due to the charge neutralizing salt bridge interactions formed by R9 and E12 of neighboring subunits (see Fig. 2).

In contrast to the findings on Cx46-R9N, the introduction of a positive charge to the NT in both Cx50-46NT and Cx50-N9R GJs resulted in nearly complete loss of their V_j-gating with very little V_j-dependent deactivation on macroscopic junctional currents (see Fig. 5). We believe that the underlying mechanisms of their apparent loss of V_j-gating are different based on their single channel gating properties. In the case of Cx50-46NT GJ, the single channel currents (i_js) showed long stable opening for all our tested V_js with very brief transitions to a residual or closed state. Open probability of Cx50-46NT was approximately one with longer dwell time. The closed or residual states of Cx50-46NT were not observed unless at the highest V_js (±100 mV). These experimental observations are consistent with our proposed model that switching the NT of Cx46 into Cx50 (Cx50-46NT) stabilize open state of the GJ channel. However, in the case of Cx50-N9R GJs, the opposite effects on i_j were observed with reduced open stability (a much shorter open dwell time than wildtype Cx50) and low P_{open} levels for all tested V_js without any V_j-dependence (Fig. 6). In addition, long lived residue states were lost in Cx50-N9R GJ and during the entire V_j pulses, the i_js displayed frequent transitions from closed state to two open states (see Fig. 7a,c). These data support a model where the Cx50-N9R GJs in our experimental conditions are already gated (or not recovered from the V_j-dependent deactivation during the V_j pulse interval), characterized with a fully closed state and frequent transitions to short-lived open states. Consistent with this model, we observed significant gating of Cx50-N9R GJ during their first large V_j exposure, not the subsequent

V_js (Yue and Bai unpublished observations). Further experiments are needed to test these models fully. It is notable that equivalent mutation (N9R) in rodent or chick Cx50 displayed prominent V_j-gating with altered gating parameters (Tong *et al.*, 2004; Xin *et al.*, 2010), unlike what we observed here in the sheep Cx50. The NT domains of these orthologs are well conserved. Therefore, it is likely that these species-specific differences are due to unique sequence variations within other regions of the channel that may influence V_j-gating properties.

Mechanistic differences in single channel conductance for Cx46 and Cx50 GJs

Previous works have implicated the NT domain and in particular the 9th residue position as being responsible for major differences in single channel conductance (γ_j) in the two closely related isoforms, Cx46 and Cx50 (Tong *et al.*, 2004; Srinivas *et al.*, 2005; Xin *et al.*, 2010; Tong *et al.*, 2013; Xin & Bai, 2013; Myers *et al.*, 2018). We observed that wildtype and variant GJ channels studied here displayed much higher free energy barriers at the NT domain for Cl^- (the major intracellular anion) versus K^+ (the major intracellular cation), making these GJs more permeable to cations with a difference in free energy barriers $\Delta\Delta G_{(Cl^- - K+)}$ = ~2.4 – 3.7 kcal mol^{-1}, consistent with previous electrophysiology studies conducted on rodent Cx50 and Cx46 that demonstrate K^+/Cl^- permeability ratios (P_{K+}/P_{Cl-}) of ~7/1 (Trexler *et al.*, 1996; Srinivas *et al.*, 1999; Trexler *et al.*, 2000; Tong *et al.*, 2014). At the same time, the differences in the peak free energy barrier to K^+ permeation between Cx50 and Cx46 ($\Delta\Delta G_{K+}$ = 0.82 kcal mol^{-1}) is consistent with the relative differences in γ_j for these isoforms, where we showed sheep Cx50 displays higher γ_j

compared to Cx46 (208 pS vs 166 pS, respectively). This energetic difference appears to be primarily related to the positive charge of R9 in Cx46. While R9 appears to also form a constriction site in the Cryo-EM structure of Cx46 that could contribute to the formation of an energetic barrier, this residue equilibrates during MD simulation to form a dynamic ensemble of inter-subunit interactions with E12 and/or R9 of a neighboring subunit that prevents this large residue from narrowing the pore permeation pathway (see Fig. 1 and 2).

The free energy landscape to cation permeation was found to be similar among our designed GJ variants, depending on the presence or absence of a positively charged arginine at the 9th position. For example, the peak ΔG_{K+} for the variants Cx46-R9N and Cx46-50NT were both reduced to that of Cx50, and the peak ΔG_{K+} for Cx50-N9R and Cx50-46NT variants were increased to a degree that was similar to Cx46 and displayed similar R9 conformational dynamics (see Fig. 2 and 3). By functional comparison, our dual patch clamp data on these variants demonstrate that the γ_js of these variants displayed unitary conductances (γ_js) in order from high to low: Cx46-R9N (261 pS) > Cx50 (208 pS) > Cx46 (166 pS) > Cx50-46NT (119 pS) = Cx50-N9R (116 pS), while Cx46-50NT was non-functional (discussed above). These results are consistent with free energy calculations obtained by MD, at least on a qualitative level.

Despite this overall congruence, some functional features displayed by the designed variants were not readily explained by the structural data, which may reflect additional structural difference between Cx46 and Cx50 that will require further investigation, and/or reflect the inherent limitations of the models and/or force-fields used for MD. For

example, γ_j of Cx46-R9N was surprisingly much higher than that of Cx50 and the γ_js of Cx50-N9R and Cx50-46NT were much lower than that of Cx46. Of note, previous studies that swapped the NT domains of orthologous Cx50 and Cx46 GJs and hemichannels reported similar behavior (Tong *et al.*, 2004; Tong & Ebihara, 2006; Kronengold *et al.*, 2012). We suggest a mechanistic basis for this observation is most likely due to other differences between these isoforms located within the pore-lining domains of TM1/EC1 that further modulate the conductance properties of these channels (Trexler *et al.*, 2000; Kronengold *et al.*, 2003; Hu *et al.*, 2006; Tong *et al.*, 2013; Tong *et al.*, 2015). Consistent with these results, we do find isoform-specific differences in the K^+ PMF observed within these regions of the channel, although these have proven more difficult to deconvolute (see Fig. 3) (Myers *et al.*, 2018). Indeed, in the study by Kronengold *et al.*, the NT domain and TM1/EC1 domains were swapped between Cx50 and Cx46 hemichannels (Kronengold *et al.*, 2012), which resulted in chimeras that produced unitary conductance values that were nearly identical to the corresponding wildtype channels from which the NT-TM1-EC1 domains were derived.

Another surprising result was that in our study Cx50-46NT displayed the smallest single channel conductance, despite the results of MD simulation that suggest this chimera forms a larger pore diameter than other variants (*e.g.* as compared to Cx50-N9R). Again, the mechanistic basis for this observation is not entirely clear, as R9 in this construct was found to adopt the same population of stabilizing salt bridge interactions as observed for wildtype Cx46 (*e.g.*, 48% vs. 47% in h-bonding distance to E12, respectively) that prevent this residue from narrowing the pore permeation pathway (see Fig. 2). Other reasons for

the quantitative differences between experiment and simulation could be due to one or combinations of the following factors: 1) contributions to channel conductance not captured by our structural models (*e.g.,* CL and CT domains were not modeled due to intrinsic disorder of these domains); 2) cation/anion permeation ratio could be altered by the variants; 3) inherent limitations in the force-fields and/or time-scales used for MD simulation. Future experiments are needed to resolve some of these possibilities.

Concluding Remarks

Our integrative structural and functional approach was able to provide detailed mechanistic insights on how Cx46 and Cx50 GJs control ion permeation, gating, open state stability, and demonstrate how nuanced differences in the NT domains of Cx50 and Cx46 lead to significantly different channel properties. We uncovered previously under-appreciated roles of hydrophobic anchoring and inter-subunit cooperative interactions (*e.g.,* between R9 and E12) as likely contributors to defining these channel properties. Understanding these differences at a mechanistic level are important to developing our understanding of how various connexins have adapted to their unique physiological roles throughout the body, and how aberrancies introduced by mutation lead to disease. Indeed, the NT domains of Cx46 and Cx50 are hot spots of genetic variation associated with congenital cataracts (Beyer *et al.*, 2013; Mackay *et al.*, 2014; Micheal *et al.*, 2018; Zhang *et al.*, 2018; Ye *et al.*, 2019). A similar integrative approach will be key to providing mechanistic insights into the etiology of these and other cataract-associated variants and possibly other connexin-linked diseases. Lastly, we note that during the completion of this

manuscript, we reported refined structural models of sheep Cx46 and Cx50 that were resolved to 1.9 Å resolution using Cryo-EM, in a near-native lipid environment (Flores *et al.*, 2020). These refined models provide an unprecedented level of detail for connexin family of intercellular communication channels, and are expected to allow even more detailed investigations targeted at understanding the mechanistic principles underpinning the functional differences of these and other closely related GJ channels.

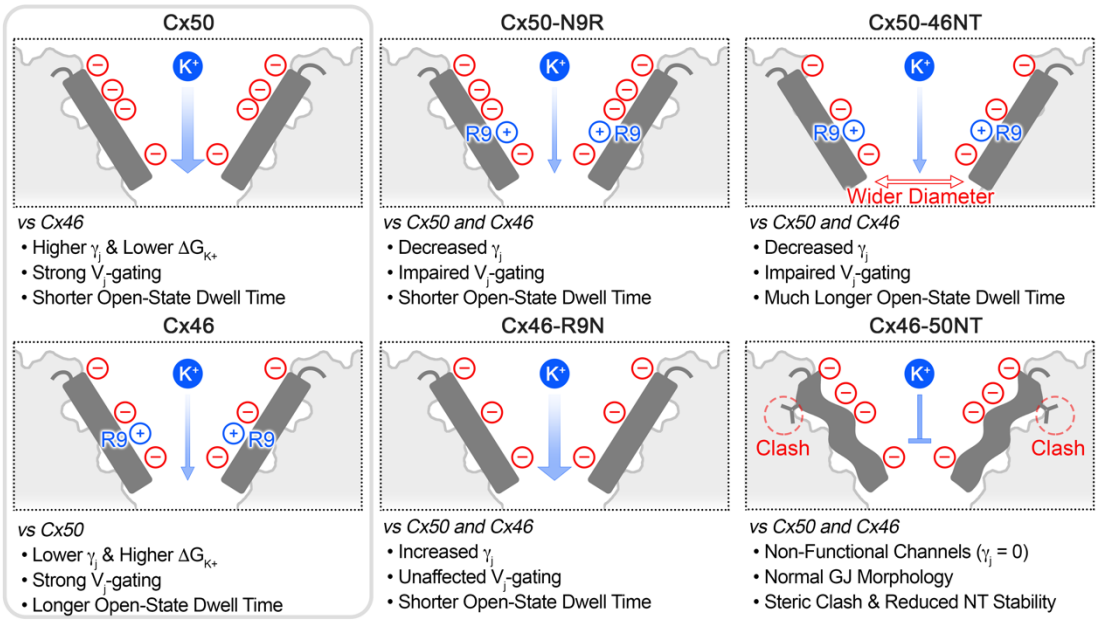


Fig. 8 Overview of significant structural and functional differences between Cx50, Cx46 and NT domain variants. Cartoon illustrating the NT domain regions (shown in dark grey) and structural and/or functional differences between Cx50, Cx46 and the designed variants for this study. Locations of negatively charged residues (red circles) and positively charged R9 residue (blue circles) are indicated (see for comparison Fig. 1b). The relative permeabilities to the major permeant ion (K^+) is indicated by the width of the blue arrow, where Cx50 displays a higher unitary conductance and lower ΔG_{K+}, as compared to Cx46. Both wildtype GJs displayed strong V_j-gating. The single point variants, Cx50-N9R and Cx46-R9N, produced channels with γ_js that were augmented by adding or removing the positively charged R9 position, although these variants affected V_j-gating and open state dwell times differently. The Cx50-46NT chimera displayed remarkably long open state dwell times and impaired V_j sensitivity, as compared to wildtype GJs, which correlated with a tighter packing of the NT domain to TM2 resulting in a wider pore radius compared to Cx46 or Cx50 following MD simulation. The Cx46-50NT construct formed morphological GJs, but these channels were non-functional. The loss of channel activity for this construct is proposed to be due to a steric clash that is introduced between V14 and T89 and reduced stability of the NT domain observed by MD simulation (indicated by the wavy NT domain in the illustration).

Data availability

Data and expression vectors related to the electrophysiology experiments presented in this work will be made available upon request. For MD data, raw, un-aligned protein and ion trajectories have been deposited to Zenodo (DOI: 10.5281/zenodo.4625961). In-house codes developed for analysis of MD trajectory data have been deposited to GitHub (https://github.com/reichow-lab/Yue-Haddad_et-al.JPhysiol).

Competing Interest: none declared.

Author contributions

All authors contributed to the preparation of the manuscript and approved the final version. B.Y. and B.G.H. contributed equally. B.Y. designed, performed and analyzed patch clamp experiments. H.C. designed and generated all variants and designed and performed GFP-tagged Cx46 variant localization experiments. Z.Z. and M.A. participated in patch clamp experiments, localization experiments, and analyzed relevant data. B.G.H. and U.K. developed homology structure models, designed and conducted MD simulations and performed analysis of MD data. D.M.Z. and S.L.R. supervised the MD studies. D.B supervised the electrophysiology and localization studies. S.L.R. and D.B. provided overall oversight to the design and execution of the work. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All persons designated as authors qualify for authorship, and all those who qualify for authorship are listed.

Funding

This work was supported by Natural Sciences and Engineering Research Council of Canada (288241 to D.B.), the National Institutes of Health (R35-GM124779 to S.L.R.) (R01-GM115805 to D.M.Z.) (F31-EY031580 to B.G.H.) and the National Science Foundation (MCB 1715823 to D.M.Z.) (XSEDE EMPOWER, ACI-1548562 to U.K.).

	V_j polarity	G_{min}	V_0	*A*
Cx46	+	0.21 ± 0.06	39.0 ± 3.6	0.14 ± 0.09
(n = 6)	-	0.23 ± 0.08	38.6 ± 7.0	0.09 ± 0.05
Cx50	+	0.21 ± 0.12	40.1 ± 7.1	0.09 ± 0.05
(n = 5)	-	0.17 ± 0.09	44.5 ± 7.7	0.08 ± 0.04
Cx46-R9N	+	0.19 ± 0.06	33.2 ± 5.9	0.11 ± 0.06
(n = 5)	-	0.21 ± 0.11	38.0 ± 9.8	0.07 ± 0.05

Table 2. Boltzmann fitting parameters for homotypic sheep Cx46, Cx50, and Cx46-R9N GJs. Data were fit to the Boltzmann equations (Eq. 1) and presented as mean ± SD, and V_0 are absolute values. One-way ANOVA followed by Tukey post-hoc test was used to compare Boltzmann fitting parameters of all three GJs. None of the parameters showed statistically significant differences for the same V_j polarity.

References

Aasen T, Mesnil M, Naus CC, Lampe PD & Laird DW. (2016). Gap junctions and cancer: communicating for 50 years. *Nat Rev Cancer* **16,** 775-788.

Bai D, Yue B & Aoyama H. (2018). Crucial motifs and residues in the extracellular loops influence the formation and specificity of connexin docking. *Biochim Biophys Acta* **1860,** 9-21.

Bargiello T & Brink P. (2009). Voltage-gating mechanisms of connexin channels. In *Connexins: A Guide*, ed. Harris AL & Locke D, pp. 103-128. Humana Press.

Beyer EC, Ebihara L & Berthoud VM. (2013). Connexin mutants and cataracts. *Front Pharmacol* **4,** 43.

Bezanilla F. (2000). The voltage sensor in voltage-dependent ion channels. *Physiol Rev* **80,** 555-592.

Bukauskas FF, Kreuzberg MM, Rackauskas M, Bukauskiene A, Bennett MV, Verselis VK & Willecke K. (2006). Properties of mouse connexin 30.2 and human connexin 31.9 hemichannels: implications for atrioventricular conduction in the heart. *Proc Natl Acad Sci U S A* **103,** 9726-9731.

Bukauskas FF & Verselis VK. (2004). Gap junction channel gating. *Biochim Biophys Acta* **1662,** 42-60.

Delmar M, Laird DW, Naus CC, Nielsen MS, Verselis VK & White TW. (2017). Connexins and Disease. *Cold Spring Harbor perspectives in biology*.

Flores JA, Haddad BG, Dolan KA, Myers JB, Yoshioka CC, Copperman J, Zuckerman DM & Reichow SL. (2020). Connexin-46/50 in a dynamic lipid environment resolved by CryoEM at 1.9 A. *Nat Commun* **11,** 4331.

Garcia IE, Prado P, Pupo A, Jara O, Rojas-Gomez D, Mujica P, Flores-Munoz C, Gonzalez-Casanova J, Soto-Riveros C, Pinto BI, Retamal MA, Gonzalez C & Martinez AD. (2016). Connexinopathies: a structural and functional glimpse. *BMC Cell Biol* **17 Suppl 1,** 17.

Goodenough DA & Paul DL. (2009). Gap junctions. *Cold Spring Harbor perspectives in biology* **1,** a002576.

Harris AL, Spray DC & Bennett MV. (1981). Kinetic properties of a voltage-dependent junctional conductance. *J Gen Physiol* **77,** 95-117.

Hu X, Ma M & Dahl G. (2006). Conductance of connexin hemichannels segregates with the first transmembrane segment. *Biophys J* **90,** 140-150.

Huang J & MacKerell AD, Jr. (2013). CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34,** 2135-2145.

Humphrey W, Dalke A & Schulten K. (1996). VMD: visual molecular dynamics. *J Mol Graph* **14,** 33-38, 27-38.

Jassim A, Aoyama H, Ye WG, Chen H & Bai D. (2016). Engineered Cx40 variants increased docking and function of heterotypic Cx40/Cx43 gap junction channels. *J Mol Cell Card* **90,** 11-20.

Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, Chun M, Li P, Gohara DW, Dolinsky T, Konecny R, Koes DR, Nielsen JE, Head-

Gordon T, Geng W, Krasny R, Wei GW, Holst MJ, McCammon JA & Baker NA. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Sci* **27,** 112-128.

Khan AK, Jagielnicki M, McIntire WE, Purdy MD, Dharmarajan V, Griffin PR & Yeager M. (2020). A Steric "Ball-and-Chain" Mechanism for pH-Mediated Regulation of Gap Junction Channels. *Cell Rep* **31,** 107482.

Kim MS, Gloor GB & Bai D. (2013). The distribution and functional properties of Pelizaeus-Merzbacher-like disease-linked Cx47 mutations on Cx47/Cx47 homotypic and Cx47/Cx43 heterotypic gap junctions. *The Biochemical journal* **452,** 249-258.

Koval M, Molina SA & Burt JM. (2014). Mix and match: investigating heteromeric and heterotypic gap junction channels in model systems and native tissues. *FEBS Lett* **588,** 1193-1204.

Kronengold J, Srinivas M & Verselis VK. (2012). The N-terminal half of the connexin protein contains the core elements of the pore and voltage gates. *The Journal of membrane biology* **245,** 453-463.

Kronengold J, Trexler EB, Bukauskas FF, Bargiello TA & Verselis VK. (2003). Pore-lining residues identified by single channel SCAM studies in Cx46 hemichannels. *Cell Commun Adhes* **10,** 193-199.

Kwon T, Harris AL, Rossi A & Bargiello TA. (2011). Molecular dynamics simulations of the Cx26 hemichannel: evaluation of structural models with Brownian dynamics. *J Gen Physiol* **138,** 475-493.

Lee HJ, Jeong H, Hyun J, Ryu B, Park K, Lim HH, Yoo J & Woo JS. (2020). Cryo-EM structure of human Cx31.3/GJC3 connexin hemichannel. *Sci Adv* **6,** eaba4996.

Mackay DS, Bennett TM, Culican SM & Shiels A. (2014). Exome sequencing identifies novel and recurrent mutations in GJA8 and CRYGD associated with inherited cataract. *Hum Genomics* **8,** 19.

Maeda S, Nakagawa S, Suga M, Yamashita E, Oshima A, Fujiyoshi Y & Tsukihara T. (2009). Structure of the connexin 26 gap junction channel at 3.5 A resolution. *Nature* **458,** 597-602.

Micheal S, Niewold ITG, Siddiqui SN, Zafar SN, Khan MI & Bergen AAB. (2018). Delineation of Novel Autosomal Recessive Mutation in GJA3 and Autosomal Dominant Mutations in GJA8 in Pakistani Congenital Cataract Families. *Genes* **9**.

Moreno AP, Berthoud VM, Perez-Palacios G & Perez-Armendariz EM. (2005). Biophysical evidence that connexin-36 forms functional gap junction channels between pancreatic mouse beta-cells. *Am J Physiol Endocrinol Metab* **288,** E948-956.

Musa H, Fenn E, Crye M, Gemel J, Beyer EC & Veenstra RD. (2004). Amino terminal glutamate residues confer spermine sensitivity and affect voltage gating and channel conductance of rat connexin40 gap junctions. *J Physiol* **557,** 863-878.

Myers JB, Haddad BG, O'Neill SE, Chorev DS, Yoshioka CC, Robinson CV, Zuckerman DM & Reichow SL. (2018). Structure of native lens connexin 46/50 intercellular channels by cryo-EM. *Nature* **564,** 372-377.

Oh S, Rivkin S, Tang Q, Verselis VK & Bargiello TA. (2004). Determinants of gating polarity of a connexin 32 hemichannel. *Biophys J* **87,** 912-928.

Oh S, Rubin JB, Bennett MV, Verselis VK & Bargiello TA. (1999). Molecular determinants of electrical rectification of single channel conductance in gap junctions formed by connexins 26 and 32. *J Gen Physiol* **114,** 339-364.

Oshima A, Tani K, Hiroaki Y, Fujiyoshi Y & Sosinsky GE. (2007). Three-dimensional structure of a human connexin26 gap junction channel reveals a plug in the vestibule. *Proc Natl Acad Sci U S A* **104,** 10034-10039.

Oshima A, Tani K, Toloue MM, Hiroaki Y, Smock A, Inukai S, Cone A, Nicholson BJ, Sosinsky GE & Fujiyoshi Y. (2011). Asymmetric configurations and N-terminal rearrangements in connexin26 gap junction channels. *J Mol Biol* **405,** 724-735.

Paul DL, Ebihara L, Takemoto LJ, Swenson KI & Goodenough DA. (1991). Connexin46, a novel lens gap junction protein, induces voltage-gated currents in nonjunctional plasma membrane of Xenopus oocytes. *J Cell Biol* **115,** 1077-1089.

Peracchia C & Peracchia LL. (2005). Inversion of both gating polarity and CO2 sensitivity of voltage gating with D3N mutation of Cx50. *Am J Physiol Cell Physiol* **288,** C1381-1389.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605-1612.

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L & Schulten K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem* **26,** 1781-1802.

Purnick PE, Oh S, Abrams CK, Verselis VK & Bargiello TA. (2000). Reversal of the gating polarity of gap junctions by negative charge substitutions in the N-terminus of connexin 32. *Biophys J* **79,** 2403-2415.

Ramanan SV, Brink PR, Varadaraj K, Peterson E, Schirrmacher K & Banach K. (1999). A three-state model for connexin37 gating kinetics. *Biophys J* **76,** 2520-2529.

Saez JC, Berthoud VM, Branes MC, Martinez AD & Beyer EC. (2003). Plasma membrane channels formed by connexins: their regulation and functions. *Physiol Rev* **83,** 1359-1400.

Sakai R, Elfgang C, Vogel R, Willecke K & Weingart R. (2003). The electrical behaviour of rat connexin46 gap junction channels expressed in transfected HeLa cells. *Pflugers Arch* **446,** 714-727.

Santos-Miranda A, Chen H, Chen RC, Odoko-Ishimoto M, Aoyama H & Bai D. (2020). The amino terminal domain plays an important role in transjunctional voltage-dependent gating kinetics of Cx45 gap junctions. *J Mol Cell Cardiol* **143,** 71-84.

Shearer D, Ens W, Standing K & Valdimarsson G. (2008). Posttranslational modifications in lens fiber connexins identified by off-line-HPLC MALDI-quadrupole time-of-flight mass spectrometry. *Invest Ophthalmol Vis Sci* **49,** 1553-1562.

Smart OS, Neduvelil JG, Wang X, Wallace BA & Sansom MS. (1996). HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J Mol Graph* **14,** 354-360, 376.

Sohl G & Willecke K. (2004). Gap junctions and the connexin protein family. *Cardiovasc Res* **62,** 228-232.

Sosinsky GE & Nicholson BJ. (2005). Structural organization of gap junction channels. *Biochim Biophys Acta* **1711,** 99-125.

Spray DC, Harris AL & Bennett MV. (1981). Equilibrium properties of a voltage-dependent junctional conductance. *Journal of General Physiology* **77,** 77-93.

Srinivas M, Costa M, Gao Y, Fort A, Fishman GI & Spray DC. (1999). Voltage dependence of macroscopic and unitary currents of gap junction channels formed by mouse connexin50 expressed in rat neuroblastoma cells. *J Physiol* **517 (Pt 3),** 673-689.

Srinivas M, Kronengold J, Bukauskas FF, Bargiello TA & Verselis VK. (2005). Correlative studies of gating in Cx46 and Cx50 hemichannels and gap junction channels. *Biophys J* **88,** 1725-1739.

Sukhareva M, Hackos DH & Swartz KJ. (2003). Constitutive activation of the Shaker Kv channel. *J Gen Physiol* **122,** 541-556.

Sun Y, Yang YQ, Gong XQ, Wang XH, Li RG, Tan HW, Liu X, Fang WY & Bai D. (2013). Novel germline GJA5/connexin40 mutations associated with lone atrial fibrillation impair gap junctional intercellular communication. *Human mutation* **34,** 603-609.

Tong JJ & Ebihara L. (2006). Structural determinants for the differences in voltage gating of chicken Cx56 and Cx45.6 gap-junctional hemichannels. *Biophysical Journal* **91,** 2142-2154.

Tong JJ, Liu X, Dong L & Ebihara L. (2004). Exchange of gating properties between rat cx46 and chicken cx45.6. *Biophys J* **87,** 2397-2406.

Tong JJ, Sohn BC, Lam A, Walters DE, Vertel BM & Ebihara L. (2013). Properties of two cataract-associated mutations located in the NH2 terminus of connexin 46. *Am J Physiol Cell Physiol* **304,** C823-832.

Tong X, Aoyama H, Sudhakar S, Chen H, Shilton BH & Bai D. (2015). The First Extracellular Domain Plays an Important Role in Unitary Channel Conductance of Cx50 Gap Junction Channels. *PLoS One* **10,** e0143876.

Tong X, Aoyama H, Tsukihara T & Bai D. (2014). Charge at the 46th residue of connexin50 is crucial for the gap-junctional unitary conductance and transjunctional voltage-dependent gating. *J Physiol* **592,** 5187-5202.

Trexler EB, Bennett MV, Bargiello TA & Verselis VK. (1996). Voltage gating and permeation in a gap junction hemichannel. *Proc Natl Acad Sci U S A* **93,** 5836-5841.

Trexler EB, Bukauskas FF, Kronengold J, Bargiello TA & Verselis VK. (2000). The first extracellular loop domain is a major determinant of charge selectivity in connexin46 channels. *Biophys J* **79,** 3036-3051.

Veenstra RD, Wang HZ, Beyer EC, Ramanan SV & Brink PR. (1994). Connexin37 forms high conductance gap junction channels with subconductance state activity and selective dye and ionic permeabilities. *Biophysical Journal* **66,** 1915-1928.

Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P, Lin H & Forman-Kay JD. (2018). Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7**.

Verselis VK, Ginter CS & Bargiello TA. (1994). Opposite voltage gating polarities of two closely related connexins. *Nature* **368,** 348-351.

Villanelo F, Escalona Y, Pareja-Barrueto C, Garate JA, Skerrett IM & Perez-Acle T. (2017). Accessing gap-junction channel structure-function relationships through molecular modeling and simulations. *BMC Cell Biol* **18,** 5.

Wang Z & Schey KL. (2009). Phosphorylation and truncation sites of bovine lens connexin 46 and connexin 50. *Exp Eye Res* **89,** 898-904.

White TW, Bruzzone R, Goodenough DA & Paul DL. (1994a). Voltage gating of connexins. *Nature* **371,** 208-209.

White TW, Bruzzone R, Wolfram S, Paul DL & Goodenough DA. (1994b). Selective interactions among the multiple connexin proteins expressed in the vertebrate lens: the second extracellular domain is a determinant of compatibility between connexins. *J Cell Biol* **125,** 879-892.

Williams WA. (1991). *Statistical Methods (8th ed.)*, vol. 86. American Statistical Association.

Xin L & Bai D. (2013). Functional roles of the amino terminal domain in determining biophysical properties of Cx50 gap junction channels. *Front Physiol* **4,** 373.

Xin L, Gong XQ & Bai D. (2010). The role of amino terminus of mouse Cx50 in determining transjunctional voltage-dependent gating and unitary conductance. *Biophys J* **99,** 2077-2086.

Xin L, Nakagawa S, Tsukihara T & Bai D. (2012). Aspartic acid residue D3 critically determines Cx50 gap junction channel transjunctional voltage-dependent gating and unitary conductance. *Biophys J* **102,** 1022-1031.

Ye Y, Wu M, Qiao Y, Xie T, Yu Y & Yao K. (2019). Identification and preliminary functional analysis of two novel congenital cataract associated mutations of Cx46 and Cx50. *Ophthalmic genetics* **40,** 428-435.

Yifrach O & MacKinnon R. (2002). Energetics of pore opening in a voltage-gated K(+) channel. *Cell* **111,** 231-239.

Zhang L, Liang Y, Zhou Y, Zeng H, Jia S & Shi J. (2018). A Missense Mutation in GJA8 Encoding Connexin 50 in a Chinese Pedigree with Autosomal Dominant Congenital Cataract. *Tohoku J Exp Med* **244,** 105-111.

Zonta F, Buratto D, Cassini C, Bortolozzi M & Mammano F. (2014). Molecular dynamics simulations highlight structural and functional alterations in deafness-related M34T mutation of connexin 26. *Front Physiol* **5,** 85.

Chapter 6

Mechanisms of Cation Permeation and Selectivity through Cx46 and Cx50 Gap Junction Channels Under an Electric Field

Introduction

Gap junctions are a class of membrane channels that facilitate direct cell-cell transport of ions and small molecules[1]. Gap junction mediated intercellular communication is a critical component for electrical coupling in the brain[2,3] and heart[4,5], to maintaining lens transparency via a water microcirculatory system[6,7]. Structurally, gap junctions are formed through the end-to-end docking of two apposed hemichannels, each composed of six subunits known as connexins[8,9]. There are 21 isoforms of connexins in the human genome, each with distinct biophysical properties such as gating, conductance, permeability and selectivity (i.e., perm-selectivity)[1,4,10]. Gap junctions can be composed of identical hemichannels (i.e., homotypic) or different isoforms (i.e., heterotypic)[1,5,11]. Furthermore, connexins are expressed in a tissue specific manner, which in combination with heterotypic selectivity, allow cells to establish cell-type specific intercellular communication pathways[11].

The characteristic topology of a connexin monomer has an n-terminal helix (NTH), four transmembrane helices (TM1-4), two extracellular loops (EC1-2) connecting TM1 & TM2 and TM3 & TM4 respectively, an intercellular loop (ICL) which connects TM2 & TM3, ending with a variably sized c-terminal domain (CTD). The first high resolution structure of a gap junction was of the connexin-26 (Cx26) isoform in a putative open state[12] and was the only model for structural-function studies for almost a decade. Elucidation of

other isoforms via X-ray crystallography remained elusive, as is the case for many membrane channels. The resolution-revolution of 2014[13,14] elevated CryoEM to a high-resolution technique, particularly suited to membrane proteins by resolving them in near-native conditions. Our laboratory used single particle CryoEM to elucidate the structure of native lens connexins Cx46 and Cx50[15,26]. Remarkably, comparative molecular dynamics (MD) simulations demonstrated that the NTH domain resolved in the Cx46/50 structures adopted a more stabilized open-state conformation, as compared to Cx26, where the NTH is unstable and rapidly unfolds[15]. Surface electrostatics and K+/Cl- potentials of mean force (PMF), which describe the free energy for a permeating ion, identified the NTH as the site of selectivity differences between Cx46 and Cx50[15].

To gain further mechanistic insight into the differences in channel properties between the closely related Cx46 and Cx50 isoforms, *in-vitro* and *in-silico* mutagenesis where applied and identified the NTH, and in particular the identity of the residue at position 9 of the NTH, as being key determinant for differences of conductance and ion perm-selectivity between these isoforms[16]. However, a limitation of these studies is that they lacked a complete description of the conduction process under an electro-motive force, and its mechanistic role in ion-selectivity.

There are currently no structures of a gap junction in an actively conducting state, due to current limitations on crystalizing or freezing proteins under a uniform electric field. Though recently, voltage gated ion channels were resolved in a voltage dependent state using CryoEM[17]. Alternatively, MD simulations of ion channels subject to non-zero transmembrane voltages have become a popular and successful technique for elucidating

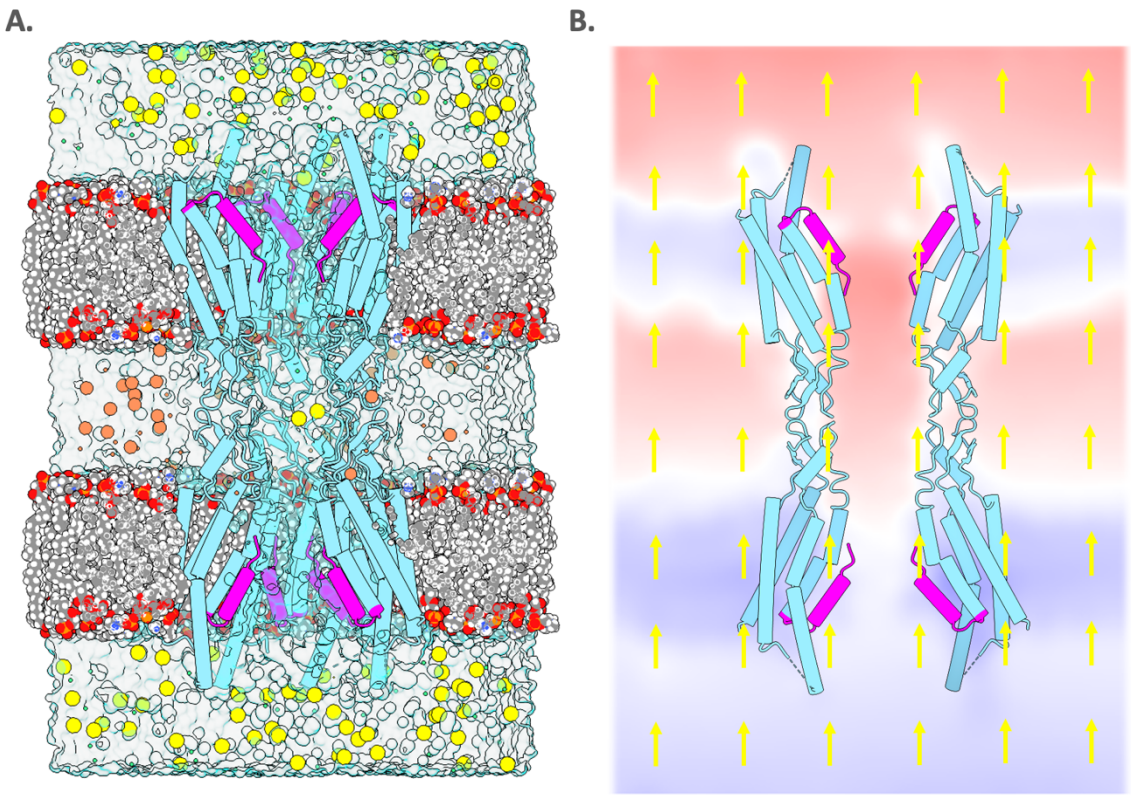


Figure 1. External electric field MD simulation setup*. A) All-atom gap junction MD system in a dual bilayer of POPC, with KCl in the intracellular compartments (above and below the dual membranes), and NaCl in the extracellular compartment. B) Total electrostatic potential map of the gap junction system with a uniform, positive, applied electric field (yellow arrows) which generate a transjunctional potential (negative above, and positive below).*

conductive properties of ion channels, ranging from highly selective voltage gated channels[18,19], to less selective wider channels such as α-hemolysin[20] and molecular transporters[21–23].

Here, computational electrophysiology study of Cx46 and Cx50 (Fig. 1a), using constant electric field MD simulations (Fig. 1B)[24,25] were applied to illuminate the ion permeation mechanism of these channels, as well as provide insights into their selectivity toward various cation species. In addition, Cx46/50 heterotypic channels were simulated to gain insight into the mechanism of current rectification, and shown to provide comparable properties to experimental results.

Methods

*Preparation of Molecular Dynamics Simulations***.** Visual Molecular Dynamics was used to construct the dodecameric gap junction models Cx46 (PDBID: 7JKC), and Cx50 (PDBID: 7JJP)[26]. The Cx46-50 heteromeric channel was built by aligning the two homomeric channels, deleting half of each, and merging them together with VMD plugin, *mergestructs*[27]. Gap junctions were then embedded in two minimized POPC membranes, generated by CHARMM-GUI[28], lipids overlapping with the protein were removed. The system is then submerged in a water box with dimensions 150 Å x 150 Å x 200 Å, using VMDs *solvate* plugin. In order to minimize excess atoms, hexameric boundary conditions applied by cutting a hexagon out of the simulation box, defined as $\begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \\ d & b & 0 \\ 0 & 0 & c \end{bmatrix}$, where a is the longest diameter (D), $b = \frac{D}{2}\sqrt{3}$, $d = \frac{D}{2}$, and c is the height of the simulation box. The intracellular and extracellular compartments were neutralized and filled with 150 mM KCl and NaCl respectively. Additional simulations of the homomeric channels (Cx46 & Cx50) were each prepared with 150 mM NaCl, and CsCl. Hydrogen mass repartitioning (HMR) was employed, using VMDs *hmr* plugin, to allow for a 4 fs time-step[29].

All simulations are performed with GPU accelerated NAMD 3.0[30], using CHARMM36[31] force fields for all atoms, and the TIP3P explicit water model. Minimization began with the lipid melting simulations with all non-lipid atoms fixed in the NVT ensemble, with Langevin dynamics[32] with a damping coefficient of 0.5 for all atoms[29] for 2 ns at a 1 fs timestep. Following, 1 Kcal/mol restraint were applied to protein atoms,

running in the NPT ensemble with the Langevin Piston Nosé-Hoover[33] method for pressure control, for 2ns at a 2 fs timestep. Restraints are then removed from protein side chains, for 2ns at 2 fs timestep. Finally, each system was subject to a 30 ns equilibration with all restraints removed, using a 4 fs timestep.

Computational Electrophysiology Simulations. Each system, Cx46[KCl/NaCl/CsCl], Cx50[KCl/NaCl/CsCl], and Cx46-50, were subject to an applied external electric field defined by

$$E = V/L \qquad \text{Eq. 1}$$

where V is the applied voltage, and L is the height of the simulation-box (i.e., c)[20,24,25,34]. Constant electric field (CEF) simulations for each system – using applied voltages: 0 mV, 75 mV, 100 mV, and 200 mV – were run in quadruplet at constant volume (NVT) with a 4 fs timestep for 250 ns (i.e., 1 μs/system). Due to the asymmetry of the heteromeric channel it was subject to additional CEF simulations at -75 mV, -100 mV, and -200 mV. All CEF simulations relaxed to a steady state within 30 ns, as was evident in the cation fluxes (Fig. 2A).

Average ionic currents ($\langle I \rangle$) were calculated as the total number of ions crossing the gap junction (N_p), over the total time (t) (Eq. 1), additionally currents were calculated as the slope from a linear fit to the cumulative ion counts[20,35]. Ion permeations were defined as an ion having visited intracellular bulk-1 (-90 < z < -60), then pore (|z| < 60), then intracellular bulk-2 (60 < z < 90) for a positive permeation, and the reverse for a negative permeation.

$$\langle I \rangle = q \frac{N_p}{t} \quad \text{Eq. 2}$$

Total potential (ϕ_{tot}) maps are calculated as,

$$\phi_{tot}(z) = \phi_{rxn}(z) + \phi_{app}(z) \quad \text{Eq. 3}$$

$$\phi_{app}(z) = \frac{q\Delta V}{L}(L - z) \quad \text{Eq. 4}$$

Where ϕ_{rxn} is the reaction potential generated from rearrangement of the solvent, and ϕ_{app} is the linear applied potential at each specific z-height, and L is the z-length of the simulation box. ϕ_{rxn} was calculated using the VMD plugin *pmepot*. Junctional conductance was calculated from the slope of the current-voltage (I-V) curve.

Thermodynamics of Ion Permeation. In order to understand the energetics of ions permeating under an electric field, a Markov-state-model (MSM) (see section **2.5**) transition matrix (T), whose elements, ij, are column-normalized rates describing the transition of an ion from z = i to z = j was generated from the z-coordinate trajectories of the ions in the simulation. Solving for the first eigenvector of the matrix T_τ yields the steady state distribution of ions throughout the system.

$$T \cdot \hat{P}_{SS} = \hat{P}_{SS} \quad \text{Eq. 5}$$

$$(T - I)\hat{P}_{SS} = \hat{0} \quad \text{Eq. 6}$$

$$P_z \in \hat{P}_{SS} \;\&\; \sum_z P_z = 1 \quad \text{Eq. 7}$$

$$\mathrm{PMF}(z) = -RT\ln\left(\hat{P}_z\right) \quad \text{Eq. 8}$$

Here, $\hat{P}_{SS}$ is the vector describing the steady-state distribution of the ions along the z-axis of the system, and it's an eigenvector of T. Equation 6 is an algebraic rearrangement of

Eq. 5 and forms an augmented matrix from ($T_\tau - I$), where I is the identity matrix. With the assumption of Eq. 7, any row of the augmented matrix can be replaced with a row of 1's, yielding $\hat{P}_{SS}$ upon reduction of the matrix to upper echelon form. Using Boltzmann statistics, the non-equilibrium energy profile is extracted from $\hat{P}_{SS}$.

Calculation of Diffusion Coefficients. The mechanism of ion permeation through the gap junctions is poorly understood. It is assumed that cations simply diffuse through the channel according to their bulk diffusion coefficients due to the size of the channel pore. To investigate the effect of the gap junction on ion-permeation, we calculate the z-dependent diffusion coefficients of K+, Na+, and Cs+. First a histogram is generated rom single timestep displacements of ions from each respective z-height, with 5 Å bins. The histograms are fit to a gaussian distribution. Using the relation of diffusion as a gaussian (Eq. 10), the diffusion coefficient is related to the variance (σ^2) in the distribution[36] (Eq. 11).

$$p(z,\tau) = \frac{1}{\sqrt{4\pi D}} \frac{e^{-z^2/4D\tau}}{\sqrt{\tau}} \quad \text{Eq. 10}$$

$$\sigma^2(z) = 2D(z)\tau \quad \text{Eq. 11}$$

Equation 10 is the standard form of a gaussian distribution used to model diffusion, z is the displacement, τ is the timestep (100 ps), and $D(z)$ is the z-dependent diffusion coefficient.

Statistical reporting. All values presented are the average plus or minus the standard error of the mean (SEM) with n = 4. Statistical significance is determined through standard two-tailed T-test.

Results and Discussion

Cx46 & Cx50 exhibit isoform specific selectivity toward cations. Gap junctions establish large, 1.4 nm wide, channels that allow a variety of solutes to permeate[12,15,26,37]. Despite the size of the channels, they still exhibit isoform specific selectivity behaviors. For example, Cx46 is permeable to cAMP while Cx50 is impermeable[38]. Most gap junctions are cation selective with exception to Cx32[39,40] and Cx31.1[37] and can even discriminate between monovalent cations[41]. Experimental calculations of ionic specificity are typically done by first calculating reversal potentials of the channels, which can be used with the

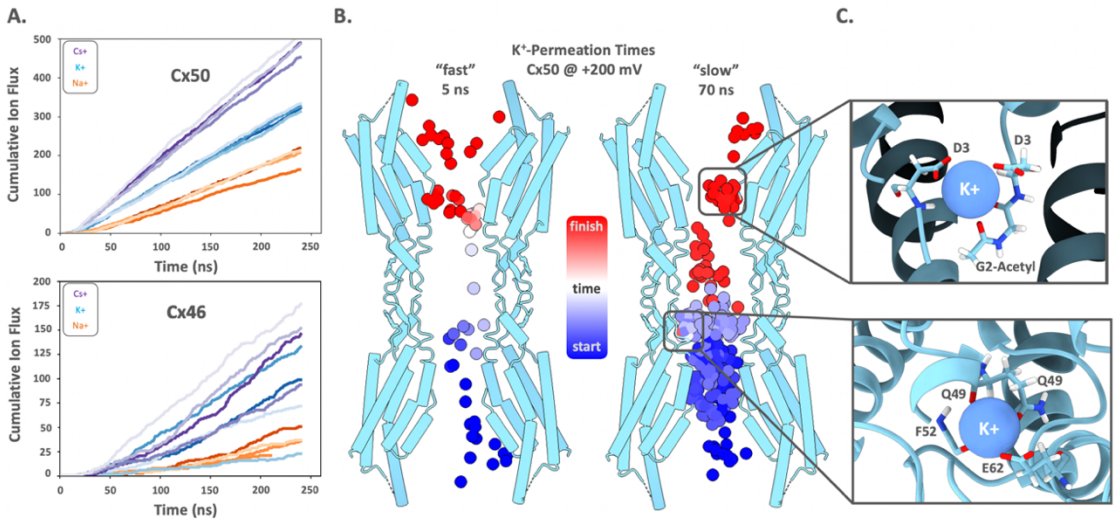


Figure 2. Monovalent cation flux through Cx46 and Cx50. *A) Cumulative flux profiles of Cs+ (purple), K+ (Blue), and Na+ (Orange) under an imposed voltage of 200 mV through Cx50 (top) and Cx46 (bottom). B) Permeation pathway of a fast (left) and slow (right) K+ ion permeating through Cx50. Color indicates time, from dark blue (start) to dark red (finish). Cross-section of protein (sky blue) shown for clarity. Cation binding events of K+ as it permeates through Cx50, stable > 40ns binding (bottom) and less stable interaction with the NTH (top).*

Goldman-Hodgkins-Katz (GHK) equation[42]. Cx46 hemichannels display the following cation series: Cs+ > K+ > Na+ with permeability ratios relative to K+ of 1.19, 1.0, and 0.8–Cx50 exhibits a similar sequence, albeit with different permeability ratios[41]. The diffusion coefficients of K+ and Cs+ are nearly identical, $\sim 0.196\ \text{Å}^2/\text{ps}$ and $\sim 0.206\ \text{Å}^2/\text{ps}$ respectively[43–45], and thus cannot account for the increased current of Cs+ over K+.

Cx46 and Cx50 intercellular channels were simulated under different ionic conditions: CsCl, KCl, and NaCl (150 mM) at equilibrium and with an applied transjunctional voltage of 200 mV (Fig. 2A). All six systems reach a steady state after approximately 30 ns of the applied voltage, as can be seen by their cumulative flux (Fig. 2A). Both Cx46 and Cx50 display higher Cs+ and K+ currents over Na+ currents, consistent with experimental results[41], and expected from the lower diffusion coefficients of Na+ in water ($\sim 0.133\ \text{Å}^2/\text{ps}$)[43]. The cumulative flux profiles for Cx50 show that this isoform is selective for monovalent cations (Fig. 2A). In contrast, Cx46 displayed relatively non-selective behavior, where K+ currents are exhibited to be as low and as high as Na+ and Cs+ currents, respectively (Fig. 2A). Even at high voltages (200 mV) K+ flux through Cx50, exhibit stable and consistent currents, with ion-permeations ranging from 5 ns to >70 ns (Fig. 2B). Multiple transient-binding sites are sampled by cations as they pass through the pore, with fast permeating ions (Fig. 2B left) predominantly permeating through the pore center (Fig. 2B). The slow ion binds to the first extracellular loop (Fig. 2C *bottom*), at the same site(s) predicted by equilibrium studies of Cx46 and Cx50[15], for > 40ns before unbinding and exiting the channel. The difference in the permeation pathways of the two

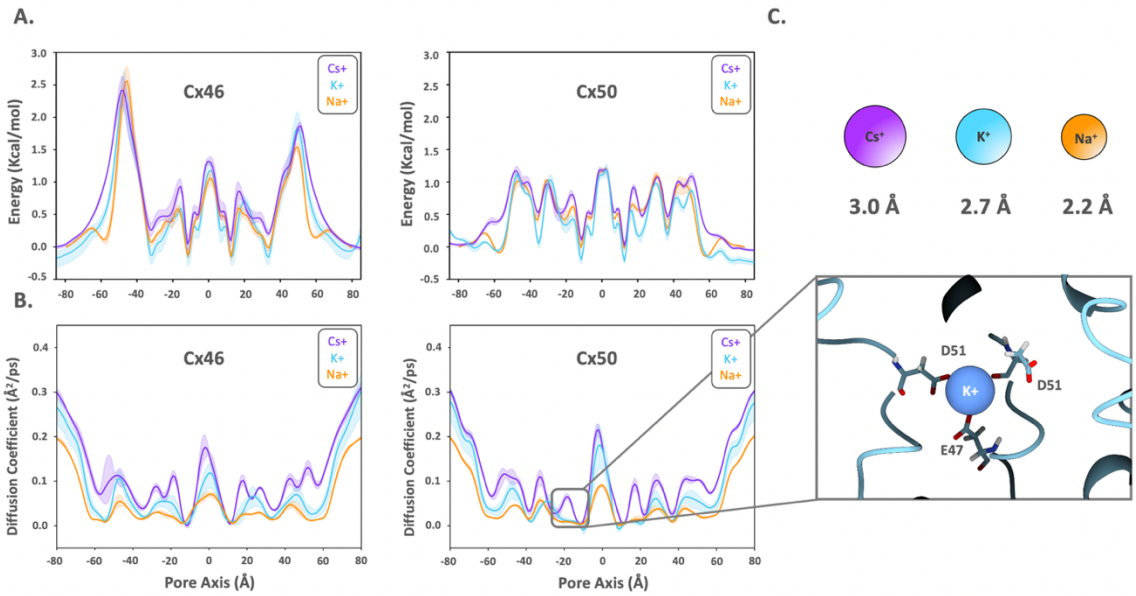


Figure 3. Dynamics of monovalent cations through Cx46 and Cx50. *A) PMFs of Cs+ (purple), K+ (blue), and Na+ (orange) calculated from simulations with an applied voltage of 200 mV, calculated by solving for the steady state probabilities of the ions through the channel. B) Pore-axis dependent diffusion coefficient of Cs+ (purple), K+ (blue), and Na+ (orange) calculated from the mean-squared displacement. C) Representative shot of a stabilized K+ ion bound by E47 and two copies of D51 at z = |17| (below). Ions with their Van der Waals radii (above). shaded area represents 95% C.I*

K+ classes of ions demonstrate a multi-ion dependent conduction mechanism, where a K+ can permeate quickly, provided the transient-binding sites are already occupied.

To better understand the mechanism of ion permeation and selectivity, PMF profiles were calculated for each of the monovalent cation conditions (Fig. 3A). PMFs calculated from MSMs at 200 mV show that overall, there are no major differences in the energy landscape experienced by the different monovalent cations. Both Cs+ and K+ in Cx46 have approximately the same entrance energy barrier (Cs+: $1.9\ \mathrm{kcal/mol}$, K+: $2.1\ \mathrm{kcal/mol}$) (Fig. 3A (left)), while for Na+ there is an energetic well just before it approaches the NTH, giving it an energy barrier of $2.5\ \mathrm{kcal/mol}$, explaining the selectivity for the two other cations (Cs+ & K+). This additional binding site is likely due to the higher charge density of Na+ and its ability to interact with local charged pore-lining residues. In

comparison, for Cx50 both K+ and Na+ experience an energetic well just before reaching the NTH, increasing the size of their entrance barrier by $0.35\ \mathrm{kcal/mol}$ (Fig. 3A (right)) relative to Cs+. Overall, the PMFs show that all three monovalent cations interact with the same binding sites within the pore, albeit with K+ binding more tightly than either Cs+ or Na+ by at least ~ $0.2\ \mathrm{kcal/mol}$.

To gain further insight to the effect of the gap junction pore on the dynamics of the permeating ions, diffusion coefficients were calculated at different points along the pore-axis (Fig. 3B). In both the Cx46 and Cx50 simulations, the diffusion constants for Cs+, K+, and Na+ (i.e., D_{Cs}, D_K, D_{Na}) in bulk are $0.36\ \text{Å}^2/\mathrm{ps}$, $0.34\ \text{Å}^2/\mathrm{ps}$, and $0.23\ \text{Å}^2/\mathrm{ps}$ respectively, in close agreement to experimental calculations ($0.2\ \text{Å}^2/\mathrm{ps}, 0.2\ \text{Å}^2/\mathrm{ps}, \& 0.13\ \text{Å}^2/\mathrm{ps}$)[43]. However, the two channels drastically affect the diffusion coefficients of each cation. In both Cx46 and Cx50, D_{Na} (Cx46: 0.028 ± 0.002, Cx50: 0.027 ± 0.002) is smaller than D_{Cs} (Cx46: 0.079 ± 0.007, Cx50: 0.075 ± 0.005) and D_K (Cx46: 0.049 ± 0.002, Cx50: 0.046 ± 0.004) throughout the entire channel (Fig. 3B). Similarly, in both Cx46 and Cx50 in D_{Cs} is higher on average than D_K – distinct from their predicted behaviour from bulk solvent calculations. In Cx50, the largest difference between D_{Cs} (0.09 ± 0.007) and D_K (0.007 ± 0.003) occurs around z = |17|, near the second extracellular loop – both Na+ (data not shown) and K+ make weak, transient associations with Asp-51 and Glu-47 (Fig. 3C), however Cs+ seamlessly flows past the site. The lack of association is expected to be a function of the cation size, and charge-density preventing Cs+ from forming the optimal coordination geometry with the pore-lining residues.

Mechanism of heterotypic Cx46/50 rectification.

It is well-established that connexin hemichannels rectify, meaning that current at one polarity is higher than at the opposite polarity[46,47]. Furthermore, it is known that Cx46 and Cx50 hemichannels are inward rectifying[48,49] at negative intracellular voltages. Homotypic gap junctions, however, are not rectifying channels due to their symmetric charge distribution about the pore-axis[16,46]. Heterotypic channels formed by Cx46 and Cx50 (Cx46-50) are known to form rectifying channels, with currents at positive voltages (relative to Cx50) being 50% higher than at negative voltages (relative to Cx50)[46]. Interestingly, these results indicate that the Cx50 half of the heterotypic gap junction experiences an outward rectification, in contrast to its hemichannel counterpart.

To gain mechanistic insight into the process of Cx46/50 heterotypic rectification, channels were simulated under constant electric fields of -200, -100, -75, 0, 75, 100, and 200 mV in order to generate the current-voltage (I-V) curves. In addition, Cx46 and Cx50 homotypic channels were simulated at positive 0, 75, 100, and 200 mV, omitting the negative voltages due to the symmetry of the channels. The slope conductance of each channel was calculated based on the simulated I/V curves and show that Cx46/50 indeed exhibits rectification (at a ratio of 50 %), with higher currents at positive voltages (relative to Cx50) in agreement to experiment[46] and with the PMFs of the heterotypic channels simulated under equilibrium conditions[15]. These studies indicate that the rectifying channel behavior is determined by the “selectivity filter” (i.e. NTH) at the pore entrances. Since the free energy barrier at the entrance of the heterotypic channel is lower on the

Cx50 side, it is easier for an ion to traverse from Cx50 to Cx46 (*i.e.,* positive voltages relative to Cx50)[15].

PMFs for each system were calculated at each applied voltage and appeared nearly identical to their equilibrium (0 mV) counterparts (data not shown), indicating a lack of voltage dependent structural changes, at least at the simulated timescales (250 ns). However, in the case of heterotypic Cx46/50 a voltage-polarity dependent difference in the PMF is observed, with a reduction in the peak energy barrier ($0.57 \pm 0.16\ \mathrm{Kcal/mol}$) on the Cx46 half of the channel at positive voltages (relative to Cx50), which indicates potential long-range effects of heterotypic coupling in these channels. In

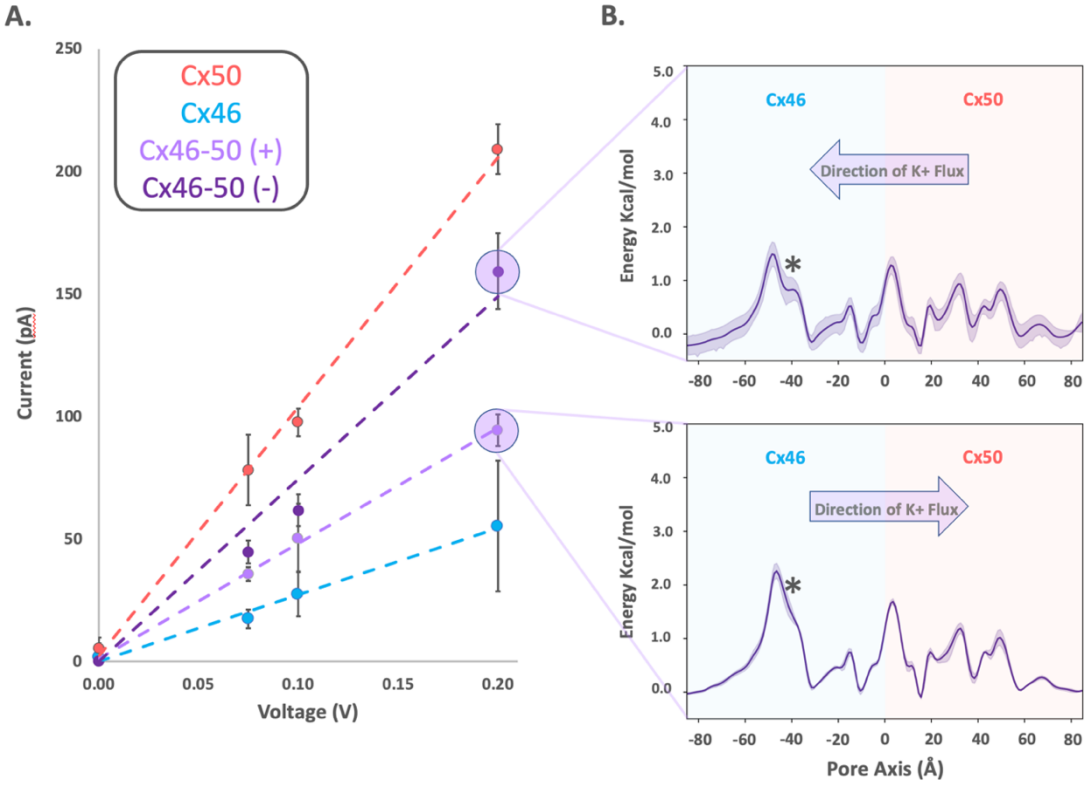


Figure 4. Conductance through Cx46, Cx50 and heterotypic Cx46-50. *Current-Voltage (I-V) curves of Cx46 (blue), Cx50 (red), Cx46-50 at negative potentials relative to Cx46 (dark purple), Cx46-50 at positive potentials relative to Cx46 (light purple). The heterotypic Cx46-50 is displayed using positive voltages for clarity, and comparison to the homomeric channels (left). Error bars show the standard error of the mean (SEM) from 4 replicates. PMFs of Cx46-50 at negative potentials relative to Cx46-half (top-right), and Cx46-50 at positive potentials relative to the Cx46-half (bottom-right). Shading in the PMFs corresponds to the 95% C.I., calculated from 4 independent replicates. Asterisk denotes region of greatest difference between the two systems.*

addition to the reduction in the peak energy barrier, a K+ ion flowing from Cx50 → Cx46 (*i.e.,* positive voltages relative to Cx50) experiences at most $1.2\ \mathrm{kcal/mol}$, compared to $2.3\ \mathrm{kcal/mol}$ flowing from Cx46 → Cx50, accounting for the 50% rectification.

Conclusion

Gap junctions have large pore openings that allow a range of different ions and small molecules to diffuse through them with some specificity. However, unlike canonical membrane channels, such as the voltage gated potassium channels[50,51] or aquaporins[52,53] which have highly selective and narrow pores, the ion permeation pathway cannot be directly resolved from the experimental structures[12,15,26,37]. Here it is demonstrated that even at high voltages (200 mV), K+ ions take multiple distinct paths (Fig. 2B), sampling different binding sites throughout the pore based on their occupancy. Furthermore, it is demonstrated that Cx50 displays a high degree of monovalent cation selectivity, selecting for cations by providing different binding sites throughout the pore, depending on the size/shape of the permeant (Fig. 2A and 3B). The mechanism of heterotypic Cx46/50 intercellular channel rectification was also demonstrated, identifying the pore entrance barrier (*i.e.,* selectivity filter) as establishing the energetic determinant for the direction of rectification. Overall, it is shown that uniform electric field MD simulations of homomeric Cx46, Cx50 and heteromeric Cx46/50 gap junction channels can recapitulate experimental observables to within an order of magnitude – therefor such simulations are a valuable and effective method for investigating the structural dynamics underlying gap junction physiology.

References

1. Goodenough DA, Paul DL. Gap junctions. *Cold Spring Harb Perspect Biol*. 2009;1(1). doi:10.1101/cshperspect.a002576
2. Connors BW, Long MA. Electrical synapses in the mammalian brain. *Annu Rev Neurosci*. 2004;27(Bennett 1977):393-418. doi:10.1146/annurev.neuro.26.041002.131128
3. Harris AL. Electrical coupling and its channels. *J Gen Physiol*. Published online 2018. doi:10.1085/jgp.201812203
4. Simon AM, Goodenough DA. Diverse functions of vertebrate gap junctions. *Trends Cell Biol*. 1998;8(12):477-483. doi:10.1016/S0962-8924(98)01372-5
5. Listan Mes¸e G, Richard G, White TW. Gap Junctions: Basic Structure and Function. *J Invest Dermatol*. 2007;127:2516-2524. doi:10.1038/sj.jid.5700770
6. Mathias RT, Kistler J, Donaldson P. The lens circulation. *J Membr Biol*. 2007;216(1):1-16. doi:10.1007/s00232-007-9019-y
7. Mathias RT, White TW, Gong X. Lens Gap Junctions in Growth, Differentiation, and Homeostasis. *Physiol Rev*. 2010;90(1):179-206. doi:10.1152/physrev.00034.2009
8. Sosinsky GE, Nicholson BJ. Structural organization of gap junction channels. *Biochim Biophys Acta - Biomembr*. 2005;1711(2 SPEC. ISS.):99-125. doi:10.1016/j.bbamem.2005.04.001
9. Perkins G, Goodenough D, Sosinsky G. Three-dimensional structure of the gap junction connexon. *Biophys J*. 1997;72(2 I):533-544. doi:10.1016/S0006-

3495(97)78693-4

10. Veenstra RD. Size and selectivity of gap junction channels formed from different connexins. *J Bioenerg Biomembr*. 1996;28(4):327-337. doi:10.1007/BF02110109

11. Koval M, Molina SA, Burt JM. Mix and match: Investigating heteromeric and heterotypic gap junction channels in model systems and native tissues. *FEBS Lett*. 2014;588(8):1193-1204. doi:10.1016/j.febslet.2014.02.025

12. Maeda S, Nakagawa S, Suga M, et al. Structure of the connexin 26 gap junction channel at 3.5 Å resolution. *Nature*. 2009;458(7238):597-602. doi:10.1038/nature07869

13. Kuhlbrandt W. The resolution revolution. *Science (80-)*. 2014;343(6178):1443-1444. doi:10.1126/science.1251652

14. Callaway E. The Revolution Will Not Be Crystallized. *Nature*. 2015;(525):172-174. doi:10.1215/01636545-2009-008

15. Myers JB, Haddad BG, O'Neill SE, et al. Structure of native lens connexin 46/50 intercellular channels by cryo-EM. *Nature*. Published online 2018. doi:10.1038/s41586-018-0786-7

16. Yue B, Haddad BG, Khan U, et al. Connexin 46 and connexin 50 gap junction channel open stability and unitary conductance are shaped by structural and dynamic features of their N-terminal domains. *J Physiol*. Published online 2021:1-23. doi:10.1101/2020.07.01.182584

17. Dickinson MS, Myasnikov A, Eriksen J, Poweleit N, Stroud RM. Resting state structure of the hyperdepolarization activated two-pore channel 3. *Proc Natl*

Acad Sci U S A. 2020;117(4):1988-1993. doi:10.1073/pnas.1915144117

18. Jensen M, Jogini V, Eastwood MP, Shaw DE. Atomic-level simulation of current-voltage relationships in single-file ion channels. *J Gen Physiol*. 2013;141(5):619-632. doi:10.1085/jgp.201210820

19. Miranda WE, Ngo VA, Wang R, Zhang L, Wayne Chen SR, Yu Noskov S. Molecular Mechanism of Conductance Enhancement in Narrow Cation-Selective Membrane Channels. doi:10.1021/acs.jpclett.8b01005

20. Aksimentiev A, Schulten K. Imaging α-hemolysin with molecular dynamics: Ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys J*. 2005;88(6):3745-3761. doi:10.1529/biophysj.104.058727

21. Kortzak D, Alleva C, Weyand I, et al. Allosteric gate modulation confers K + coupling in glutamate transporters . *EMBO J*. 2019;38(19). doi:10.15252/embj.2019101468

22. Alleva C, Philipp J, Kortzak D, Weyand I, Fahlke C. Molecular Basis of Coupled Transport and Anion Conduction in Excitatory Amino Acid Transporters. *Neurochem Res*. Published online 2021. doi:10.1007/s11064-021-03252-x

23. Machtens JP, Kortzak D, Lansche C, et al. Mechanisms of anion conduction by coupled glutamate transporters. *Cell*. 2015;160(3):542-553. doi:10.1016/j.cell.2014.12.035

24. Escalona Y, Garate JA, Araya-Secchi R, Huynh T, Zhou R, Perez-Acle T. Exploring the Membrane Potential of Simple Dual-Membrane Systems as Models for Gap-Junction Channels. *Biophys J*. 2016;110(12):2678-2688.

doi:10.1016/j.bpj.2016.05.005

25. Roux B. The Membrane Potential and its Representation by a Constant Electric Field in Computer Simulations. doi:10.1529/biophysj.108.136499

26. Flores JA, Haddad BG, Dolan KA, et al. Connexin-46/50 in a dynamic lipid environment resolved by CryoEM at 1.9 Å. *Nat Commun*. 2020;11(1):1-11. doi:10.1038/s41467-020-18120-5

27. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. *J Mol Graph*. 1996;(14):33-38.

28. Wu EL, Cheng X, Jo S, et al. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J Comput Chem*. 2014;35(27):1997-2004. doi:10.1002/jcc.23702

29. Balusek C, Hwang H, Lau CH, et al. Accelerating Membrane Simulations with Hydrogen Mass Repartitioning. *J Chem Theory Comput*. 2019;15(8):4673-4686. doi:10.1021/acs.jctc.9b00160

30. Phillips JC, Hardy DJ, Maia JDC, et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys*. 2020;153(4). doi:10.1063/5.0014475

31. Huang J, Mackerell AD. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J Comput Chem*. 2013;34(25):2135-2145. doi:10.1002/jcc.23354

32. Davidchack RL, Handel R, Tretyakov M V. Langevin thermostat for rigid body dynamics. *J Chem Phys*. 2009;130(23):234101. doi:10.1063/1.3149788

33. Sturgeon JB, Laird BB. Symplectic algorithm for constant-pressure molecular

dynamics using a Nosé-Poincaré thermostat. *J Chem Phys*. 2000;112(8):3474-3482. doi:10.1063/1.480502

34. Gumbart J, Khalili-Araghi F, Sotomayor M, Roux B. Constant electric field simulations of the membrane potential illustrated with simple systems ☆. Published online 2012. doi:10.1016/j.bbamem.2011.09.030

35. Kutzner C, Grubmüller H, De Groot BL, Zachariae U. Computational electrophysiology: The molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophys J*. 2011;101(4):809-817. doi:10.1016/j.bpj.2011.06.010

36. Einstein A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann Phys*. 1905;322(8):549-560. doi:10.1002/andp.19053220806

37. Lee H, Jeong H, Hyun J, et al. Cryo-EM structure of human Cx31 . 3 / GJC3 connexin hemichannel. 2020;(August):1-16.

38. Valiunas V, Brink PR, White TW. Lens Connexin Channels Have Differential Permeability to the Second Messenger cAMP. Published online 2019.

39. Nicholson BJ, Weber PA, Cao F, Chang HC, Lampe P, Goldberg G. The molecular basis of selective permeability of connexins is complex and includes both size and charge. *Brazilian J Med Biol Res*. 2000;33(4):369-378. doi:10.1590/S0100-879X2000000400002

40. Weber PA, Chang HC, Spaeth KE, Nitsche JM, Nicholson BJ. The permeability of gap junction channels to probes of different size is dependent on connexin

composition and permeant-pore affinities. *Biophys J*. 2004;87(2):958-973. doi:10.1529/biophysj.103.036350

41. Trexler EB, Bennett MVL, Bargiello TA, Verselis VK. Voltage gating and permeation in a gap junction hemichannel. *Proc Natl Acad Sci U S A*. 1996;93(12):5836-5841. doi:10.1073/pnas.93.12.5836

42. Sterratt DC. Goldman-Hodgkin-Katz Equations. In: *Encyclopedia of Computational Neuroscience*. Springer New York; 2015:1300-1302. doi:10.1007/978-1-4614-6675-8_229

43. Harned HS. THE DIFFUSION COEFFICIENTS OF THE ALKALI METAL CHLORIDES AND POTASSIUM AND SILVER NITRATES IN DILUTE AQUEOUS SOLUTIONS AT 250*. *J Am Chem Soc*. 1954;76(8):2064-2065. doi:10.1021/ja01637a010

44. Chen DP, Xu L, Tripathy A, Meissner G, Eisenberg B. Selectivity and permeation in calcium release channel of cardiac muscle: Alkali metal ions. *Biophys J*. 1999;76(3):1346-1366. doi:10.1016/S0006-3495(99)77297-8

45. Lee SH. Molecular dynamics simulation of ion mobility. 2. alkali metal and halide ions using the spc/e model for water at 25 °c. *J Phys Chem*. 1996;100(4):1420-1425. doi:10.1021/jp953050c

46. Hopperstad MG, Srinivas M, Spray DC. Properties of gap junction channels formed by Cx46 alone and in combination with Cx50. *Biophys J*. 2000;79(4):1954-1966. doi:10.1016/S0006-3495(00)76444-7

47. Oh S, Rubin JB, Bennett M V, Verselis VK, Bargiello TA. Molecular determinants of electrical rectification of single channel conductance in gap junctions formed by

connexins 26 and 32. *J Gen Physiol*. 1999;114(3):339-364. doi:10.1085/jgp.114.3.339

48. Srinivas M, Calderon DP, Kronengold J, Verselis VK. Regulation of Connexin Hemichannels by Monovalent Cations. *J Gen Physiol*. 2006;127(1):67-75. doi:10.1085/jgp.200509397

49. Srinivas M, Kronengold J, Bukauskas FF, Bargiello TA, Verselis VK. Correlative studies of gating in Cx46 and Cx50 hemichannels and gap junction channels. *Biophys J*. 2005;88(3):1725-1739. doi:10.1529/biophysj.104.054023

50. Bernèche S, Roux B. Energetics of ion conduction through the K+ channel. *Nature*. 2001;414(6859):73-77. doi:10.1038/35102067

51. Doyle DA, Cabral JM, Pfuetzner RA, et al. The structure of the potassium channel: Molecular basis of K+ conduction and selectivity. *Science (80-)*. 1998;280(5360):69-77. doi:10.1126/science.280.5360.69

52. Park S, Tajkhorshid E, Schulten K, Jensen MØ. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc Natl Acad Sci*. 2002;99(10):6731–6736. doi:10.1073/pnas.102649299

53. Zhu F, Tajkhorshid E, Schulten K. Molecular dynamics study of aquaporin-1 water channel in a lipid bilayer. *FEBS Lett*. 2001;504(3):212-218. doi:10.1016/S0014-5793(01)02749-1

Chapter 7:

The Cutting Edge of Computational and Electron Microscopy

The field of bio-molecular dynamics (MD) simulations has made great strides, and advancements in its short 60-years of existence. The field began with short (10's of picoseconds) simulations of soluble bovine trypsin inhibitor (1977), giving us the first atomic-level insight to the inherent 'wigglings and jigglings' within a protein[1]. In 1993, Klaus Schulten's group finished simulating a fully solvated biological membrane (~ 24,000 atoms) for nearly 2 years, achieving 263 ps, demonstrating that contemporary lipid-forcefields were able to reproduce gross-experimental results of lipid membranes, a monumental achievement at the time[2]. Membranes, and therefore membrane bound proteins were an early focal point for MD simulations, as the process of ion-conduction in simple channels occurred at then achievable timescales.

Early membrane simulations of ion channels came from the simple system gramicidin, an anti-bacterial peptide, which forms a single beta-helix which spans and perforates the plasma membrane. The first all-atom simulation study of gramicidin was in 1989, where Aqvist & Warshel calculated the solvation energetics of Na+ transport through the single pass membrane channel, however the membrane was modelled effectively as a hydrophobic slab[3]. Later in the 2001, Schulten's group simulated aquaporin in a fully atomistic lipid bilayer, and characterized, for the first time, the mechanism of water permeation through the channel[4]. Over the next two decades there would be an explosion of novel membrane channel simulations of increasing complexity,

albeit they've been limited by achievable timescales and focused their attention on the fastest components of ion-channel physiology, permeation[5–9].

A comprehensive study on drug targets done in 2006[10], and a follow up in 2016[11] showed that membrane proteins constitute 60% of all drug targets – underscoring the need to understand physiologically relevant mechanisms of membrane-bound receptors, transporters and channels. Membrane channel proteins are more than just passive ion-channels, and have complex biophysical mechanisms and physiology such as allosteric modulation and gating. These mechanisms involve large conformational changes in the protein, or macromolecular complex that occur on the millisecond or longer timescale[12]. Enhanced sampling techniques, such as weighted ensemble (WE), allow for the unbiased simulation of long-timescale motions in increasingly complex systems[13,14]. One particularly exciting and recent application of WE is the simulation of the opening and closing of the SARS-CoV-2 spike protein[15], the membrane bound culprit of the 2019-2021 pandemic. Such strides have already demonstrated the capabilities of computational microscopy.

Concurrent to advances in biomolecular simulations were advances in Cryo-electron microscopy (CryoEM). The resolution revolution of 2014[16] propelled the once low-resolution 'blobology' method to a technique that can achieve < 2 Å resolution structures of membrane bound proteins[17,18]. The technique was the focus of the 2017 Nobel Prize in chemistry, awarded in part for the computational image analyses involved in reconstruction. In recent years, CryoEM computational method developers have used complex statistical image-classification to extract pseudo-dynamics of individual proteins

and complexes[19,20]. The sub-classifications from the pseudo-dynamics provide MD researchers with critical structures along the reaction coordinate of a proteins motions, such as the helical rearrangements in G-protein coupled receptors[21].

As structural and simulation techniques advance so will our understanding of the dynamics underlying larger and more complex systems. So far, MD simulations and CryoEM have been largely limited to looking at single channels within a membrane. Recent developments of lipid nano-discs, which allow for the elucidation of structures in native lipids, aim to utilize endogenous lipid extracts to fully observe proteins in their physiological environment[22]. Furthermore, advances in cryo-electron tomography illuminate our understanding of large biomolecular complexes *in-situ*[23]. This is an exciting time for computational biophysicists, particularly at the cutting edge of structure elucidation and MD simulations. Surely, the next 'resolution-revolution' will be of the computational microscope, revealing the intimate relationship of structure, dynamics and function.

References

1. Mccammon JA, Gel BR, Karp Ius M. *Dynamics of Folded Proteins*. Vol 267. Nature Publishing Group; 1977.
2. Helmut Heller P, Schaefer M, Schulten K. Molecular Dynamics Simulation of a Bilayer of 200 Lipids in the Gel and in the Liquid-Crystal Phases. *J Phys Chem*. 1993;97:41. Accessed May 23, 2021. https://pubs.acs.org/sharingguidelines
3. Mackay DH, Berens PH, Wilson KR, Hagler AT. Structure and dynamics of ion transport through gramicidin A. *Biophys J*. 1984;46(2):229-248. doi:10.1016/S0006-3495(84)84016-3
4. Zhu F, Tajkhorshid E, Schulten K. Molecular dynamics study of aquaporin-1 water channel in a lipid bilayer. *FEBS Lett*. 2001;504(3):212-218. doi:10.1016/S0014-5793(01)02749-1
5. Jensen M, Jogini V, Eastwood MP, Shaw DE. Atomic-level simulation of current-voltage relationships in single-file ion channels. *J Gen Physiol*. 2013;141(5):619-632. doi:10.1085/jgp.201210820
6. Roux B, Allen T, Bernèche S, Im W. Theoretical and computational models of biological ion channels. *Q Rev Biophys*. 2004;37(1):15-103. doi:10.1017/S0033583504003968
7. Noskov SY, Bernéche S, Roux B. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature*. 2004;431(7010):830-834. doi:10.1038/nature02943
8. Allen TW, Kuyucak S, Chung SH. Molecular dynamics study of the KcsA potassium

channel. *Biophys J*. 1999;77(5):2502-2516. doi:10.1016/S0006-3495(99)77086-4

9. Li Y, Barbault F, Delamar M, Zhang R, Hu R. Targeted molecular dynamics (TMD) of the full-length KcsA potassium channel: On the role of the cytoplasmic domain in the opening process. *J Mol Model*. 2013;19(4):1651-1666. doi:10.1007/s00894-012-1726-3

10. Overington JP, Al-Lazikani B, Hopkins AL. How many drug targets are there ? PubMed Commons. *Nat Rev Drug Discov*. 2006;5(12):10. http://www.ncbi.nlm.nih.gov/pubmed/17139284%0Ahttp://dx.doi.org/10.1038/nrd2199

11. Santos R, Ursu O, Gaulton A, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*. 2016;16(1):19-34. doi:10.1038/nrd.2016.230

12. Borgnia MJ, Agre P. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from Escherichia coli. *Proc Natl Acad Sci U S A*. 2001;98(5):2888-2893. doi:10.1073/pnas.051628098

13. Huber GA, Kim S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys J*. 1996;70(1):97-110. doi:10.1016/S0006-3495(96)79552-8

14. Zuckerman DM, Chong LT. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu Rev Biophys*. 2017;46(1):43-57. doi:10.1146/annurev-biophys-070816-033834

15. Sztain T, Ahn S-H, Bogetti AT, et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. *bioRxiv*. Published online 2021:2021.02.15.431212.

doi:10.1101/2021.02.15.431212

16. Kuhlbrandt W. The resolution revolution. *Science (80-)*. 2014;343(6178):1443-1444. doi:10.1126/science.1251652

17. Flores JA, Haddad BG, Dolan KA, et al. Connexin-46/50 in a dynamic lipid environment resolved by CryoEM at 1.9 Å. *Nat Commun*. 2020;11(1):1-11. doi:10.1038/s41467-020-18120-5

18. Nakane T, Kotecha A, Sente A, et al. Single-particle cryo-EM at atomic resolution. *Nature*. 2020;587(7832):152-156. doi:10.1038/s41586-020-2829-0

19. Punjani A, Fleet DJ. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol*. 2021;213(2):107702. doi:10.1016/j.jsb.2021.107702

20. Zhong ED, Bepler T, Berger B, Davis JH. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods*. 2021;18(2):176-185. doi:10.1038/s41592-020-01049-4

21. Marino J, Schertler GFX. A set of common movements within GPCR-G-protein complexes from variability analysis of cryo-EM datasets. *J Struct Biol*. 2021;213(2). doi:10.1016/j.jsb.2021.107699

22. Brown CJ, Trieber C, Overduin M. Structural biology of endogenous membrane protein assemblies in native nanodiscs. *Curr Opin Struct Biol*. 2021;69:70-77. doi:10.1016/j.sbi.2021.03.008

23. Turk M, Baumeister W. The promise and the challenges of cryo-electron tomography. *FEBS Lett*. 2020;594(20):3243-3261. doi:10.1002/1873-3468.13948

Appendix

Codes of Significance and Wayside Projects

GapJAnalysis.py

https://github.com/reichow-lab/Yue-Haddad_et-al.JPhysiol2021/tree/main/GapJ_Analysis

GapJAnalysis is a python program designed to analyze the energetics of ion permeation, and ion conduction through gap junctions. The program has two primary functions: first, to calculate a potential of mean force (PMF) of an ion traversing the pore, using the pore axis (z-axis) as the reaction coordinate. Second, to calculate the current through the channel by counting the ions as they complete a permeation.

Calculating the PMF

The only input to GapJAnalysis is a text file with a 1-dimensional (1D) z-coordinates of each ion in the simulation. The first step in calculating the PMF is to first subdivide the reaction coordinate into bins, then initialize and populate a count-matrix (C^{τ}), whose elements (i,j) are the initial and final z-bins that an ion visits after one timestep (τ) (Fig. 1A, 1B) – 100 ps in many cases. The count matrix is column normalized to produce a transition matrix (T^{τ}) whose elements are rate constants for the i → j transition (k_{ij}) (Fig. 1C). At this stage the transition matrix can be used to calculate the PMF by utilizing detailed balance (Fig. 2A), or by solving for the steady state probability matrix (Fig. 2B) (See Chapter 2 section 2.5). The associated program, Plotter.py (same directory as GapJAnalysis), interpolates the data through a cubic-spline and plots the 95% C.I. when outputs from multiple replicates are used as the input to Plotter.py.

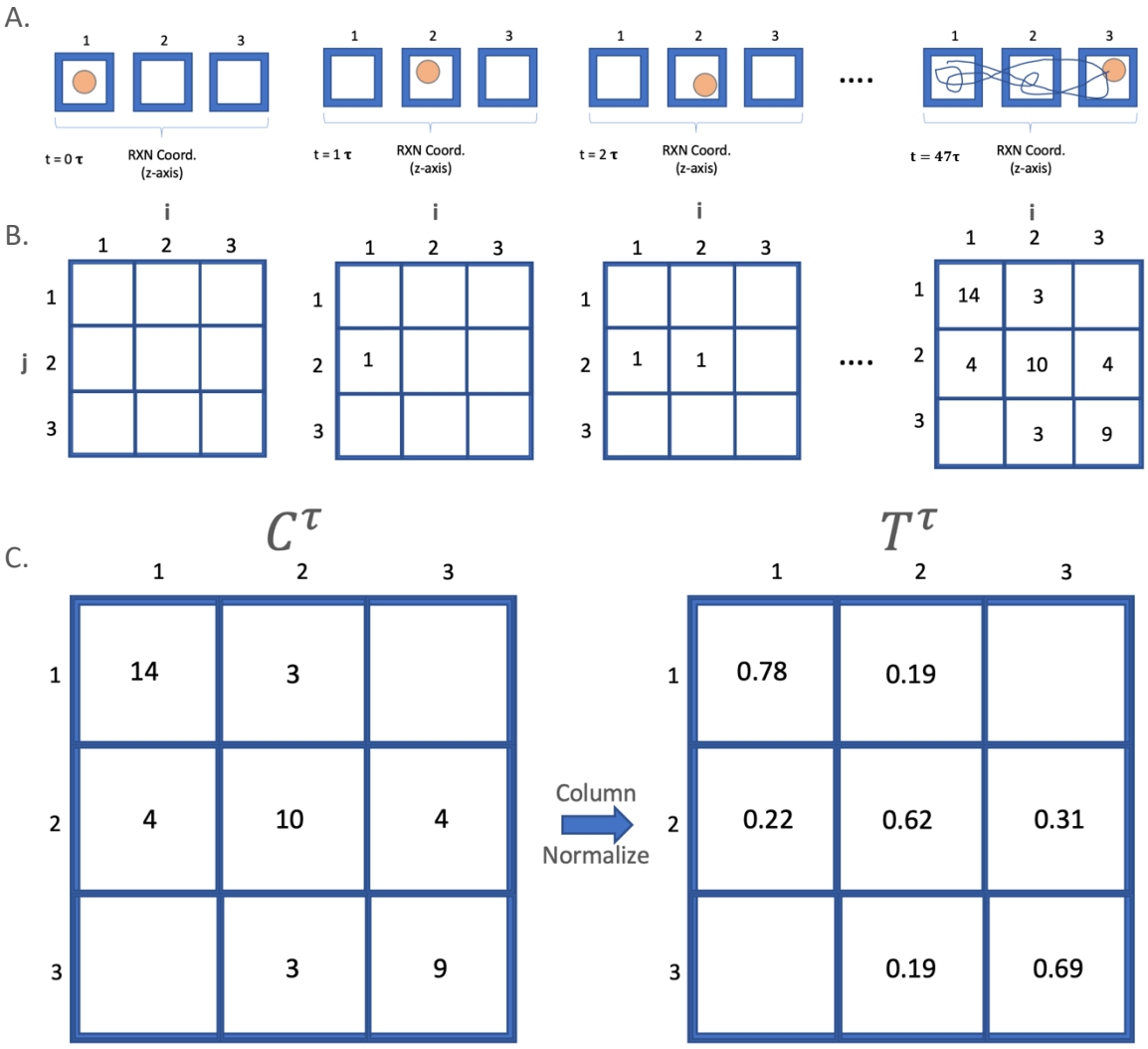


Figure 3. Building the transition matrix (i.e., MSM)*. A) Simplified system with three bins along the reaction coordinate (z-axis), showing the position of the ion (orange circle) traversing through each bin at discrete timesteps (τ). B) The initialized count matrix, being populated after each timestep, with the matrix element i-component being bin-from and the j-component being the bin-to, populated at each timestep. C) Generating a transition (rate) matrix by column normalizing. Each element being a rate constant, or conditional probability describing the movement of the ion.*

A.

1. **Detailed Balance**

2. $$P_i \cdot k_{ij} = P_j \cdot k_{ji}$$

3. $$\frac{P_i}{P_j} = \frac{k_{ji}}{k_{ji}} = e^{\left(-\Delta G_{i\to j}/RT\right)}$$

$$\Delta G_{i\to j} = -RT \cdot \ln\left(\frac{k_{ji}}{k_{ij}}\right)$$

j \ i	1	2	3
1	0.78	0.19	
2	0.22	0.62	0.31
3		0.19	0.69

B.

$$PMF_i = \sum_{n=1}^{i-1} \Delta G_{n\to n+1} = -\sum_{n=1}^{i-1} RT\ln\left(\frac{k_{n+1,n}}{k_{n,n+1}}\right)$$

Solving for Steady-State Probabilities

1. $$T_\tau \cdot \overrightarrow{P_{SS}} = \overrightarrow{P_{SS}}$$

$$(T_\tau - I) \cdot \overrightarrow{P_{SS}} = 0$$

2. $$\sum_i P_i = 1$$

3. $$\Delta G_i = -RT\ln\left(P_i^{SS}\right)$$

Combine equations 1 & 2, modify matrix, solve augmented matrix

$$\begin{bmatrix} (0.78-1) & 0.19 & 0 & 0 \\ 0.22 & (0.62-1) & 0.31 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Modified $(T_\tau - I)$ | $\vec{0}$ with conditional

$$\overrightarrow{P_{SS}}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0.349 \\ 0 & 1 & 0 & 0.403 \\ 0 & 0 & 1 & 0.248 \end{bmatrix}$$

Figure 4. Calculation of the PMF from the transition matrix. *A) Detailed balance describes statistical equilibrium (Eq. A1), where $P_{i/j}$ are the steady state probabilities that the ion exists in bins i and j respectively and $k_{ij/ji}$ are the rate constants (i.e., transition matrix elements). Equations A2. And A3. are rearrangements of equation A1, showing that only rate-constants are required to calculate free energy. To the right is the transition (rate) matrix which can be used to calculate free energy. B) Alternative method which calculates the first eigenvector of the transition matrix, which corresponds to the list of steady state probabilities ($\hat{P}_{ss}$). Equation B1 displays a linear equation with the transition matrix (T^τ) multiplied by $\hat{P}_{ss}$ to yield $\hat{P}_{ss}$. Using equation B2., which states that the sum of state probabilities must equal 1, the transition matrix is modified to be an augmented matrix described by equation B1. Reducing the augmented matrix to echelon form yields $\hat{P}_{ss}$, which can be used to calculate the free energy through the system via equation B3.*

Calculating Current

The current is calculated by counting successful permeations through the channel. Initially, the user defines three regions of the simulation system: Top-Bulk, Pore, Bottom-Bulk (Fig. 3). The program counts a positive permeation as an ion who visits each bin in the following order, Bottom-Bulk → Pore → Top-Bulk – negative permeations occur when the opposite occurs. In addition to tracking what time the ion permeates the program tracks the permeation time, or the time inside the Pore. The associated program Plotter.py calculates the total average current, the instantaneous current and generates plots of ion flux vs. time. Total average current is simply the number of permeations divided by the time passed.

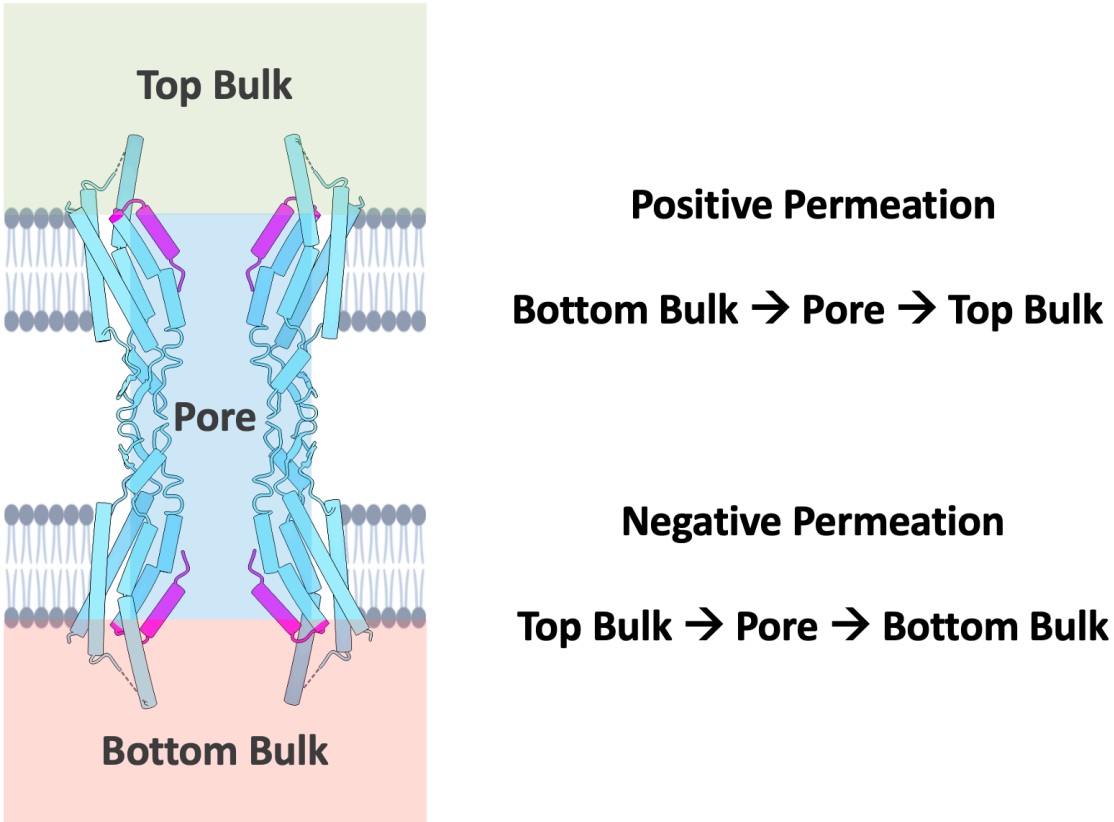


Figure 5. Depiction of the ion tracking algorithm, which assesses where the ion is at any point throughout the trajectory and marks a full transition when one of the two conditions (right) are met.

LipNetwork.tcl

https://github.com/reichow-lab/Flores-Haddad-Dolan_et-al.NatCommun2020/blob/master/LipNetwork.tcl

LipNetwork is a program written in Tcl, the scripting language which VMD[1] is built on, that tracks the dynamics of the lipids surrounding the gap junction. The goal the program was to identify the pattern of lipids that fit into lipid-tail densities calculated both from simulation and experiment[2]. After an initial lipid density map is calculated from the duration of the simulation, LipNetwork tracks any lipid that comes within a pre-defined radius of the gap junction and assess which lipids tails are inside the densities (Fig. 4). A lipid is assigned to a position (i,j) defined by the numbering of the lipid positions if each of that lipids tails are within the lipid densities. The output of this program is a heatmap which defines the most probable lipid-configurations around the gap junction.

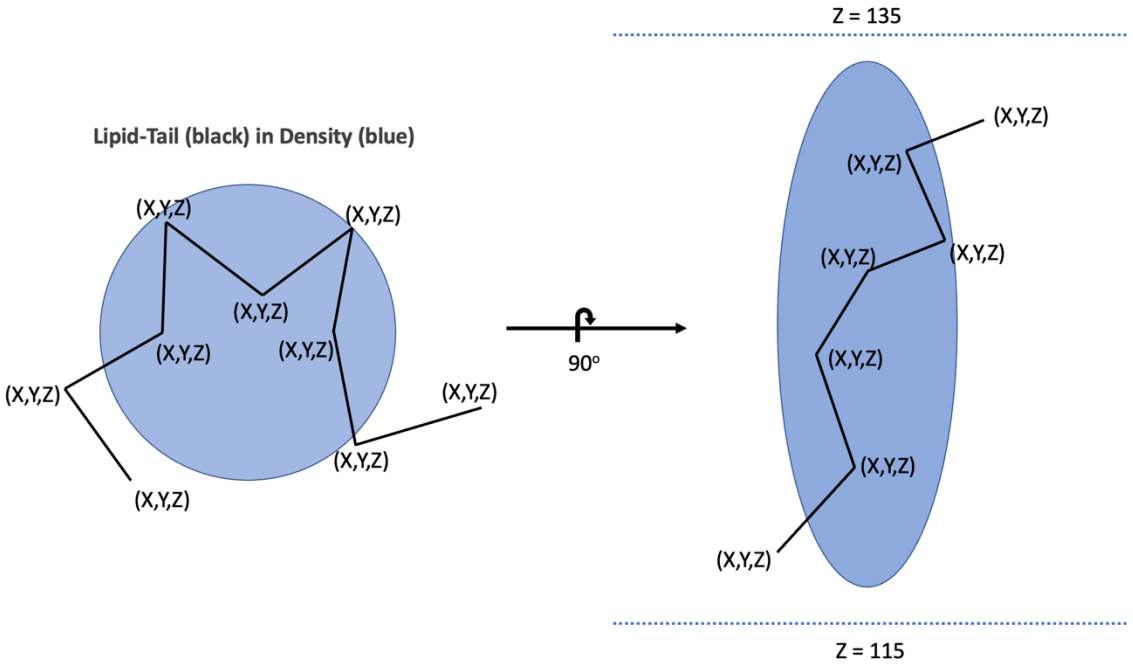


Figure 6. Selection criteria for lipid-tail in density classification. *Lipid tails are depicted as the kinked black lines occupying space within the average lipid density (blue). Each carbon of the lipid chain is assessed by ensuring it is within the bounds of the density (determined by isovalue), and z-height. X,Y coordinates are used to identify the closest lipid density.*

Additionally, the program generates lipid-state trajectories which are used to map the lipid transitions throughout the simulation. A workflow describing how LipNetwork is used is presented in the github page (https://github.com/reichow-lab/Flores-Haddad-Dolan_et-al.NatCommun2020)

Wayside Projects

Perm-selectivity of cAMP and Glutathione through Cx46 and Cx50

The goal of this project was to characterize the permeation and selectivity of Cx46 and Cx50 to biologically relevant permeants, secondary messenger cyclic adenosine monophosphate (cAMP) and antioxidant glutathione (GSH). Cx46 and Cx50 display differential selectivity to cAMP, such that Cx50 is almost entirely impermeable to cAMP[3]. GSH is a reducing agent with high abundance in the eye lens and permeates through Cx46

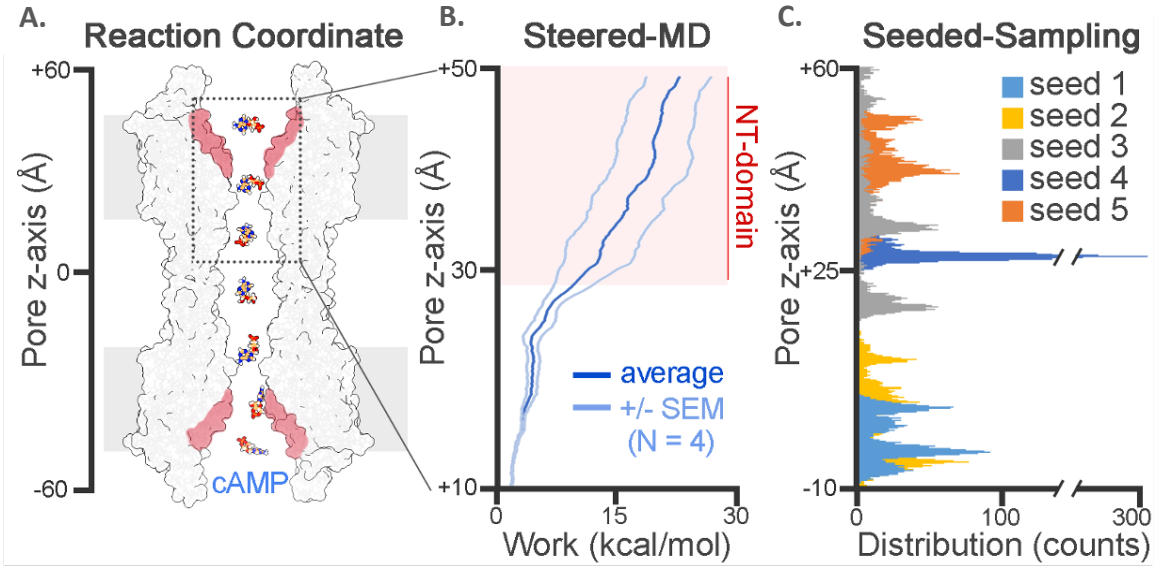


Figure 7. ***Preliminary MD studies on cAMP permeation mechanism. a.*** *Reaction coordinate depicting cAMP permeation through the pore of connexin-46.* ***b.*** *Preliminary steered-MD experiments show a high-energy barrier at the NT-domain (N=4).* ***d.*** *Distribution of cAMP within the channel pore following unbiased seeded sampling (N=5, 10 ns), show cAMP is highly mobile within the pore, providing feasibility of this approach.*

and Cx50[4]. The experimental permeation rate of cAMP through Cx46 is on the order of 2800 molecules/second, therefore brute force equilibrium simulations with physiological concentrations would be infeasible to sample the entire pore pathway. To overcome these barriers, seeded sampling and weighted ensemble (WE) were proposed to characterize cAMP throughout the pore.

To generate initial seeds of cAMP through the channel, steered molecular dynamics (SMD) was used to pull a cAMP through the channel, constrained to the pore-center (Fig. 5A, 5B) In addition to generating initial states, SMD provides an estimate of the force required to pull the molecule (Fig. 5B). Frames from the SMD were extracted as initial "seeds", which were then subject to equilibrium MD simulations. Each seed was subject to 20ns simulations, where they sampled their respective regions of the pore (Fig. 5C). Due to the increased degrees of freedom of cAMP compared to an ion, the seeds proved insufficient for sampling. The next approach was to use weighted ensemble (WE) simulations to propagate the molecule forward through the channel[5].

Selectivity of Hemichannel docking

Gap junctions have the ability to form heterotypic intercellular channels, where there are to hemichannels docking each of a different isoform (see Chapter 1, section **1.3**). There are certain empirical rules that determine the compatibility, or selectivity of gap junction isoforms. While these rules generally work, there are exceptions. Cx50 and Cx46 are both Group-I, meaning they are compatible together, and amongst other members of Group-I; Cx43 on the other hand is in Group-II, yet compatible for docking with Cx46 and

not Cx50. The goal of this project was to identify the molecular determinants of heterotypic docking and specifically the structural reason for Cx46's promiscuity in docking with both Group-I and Group-II gap junctions.

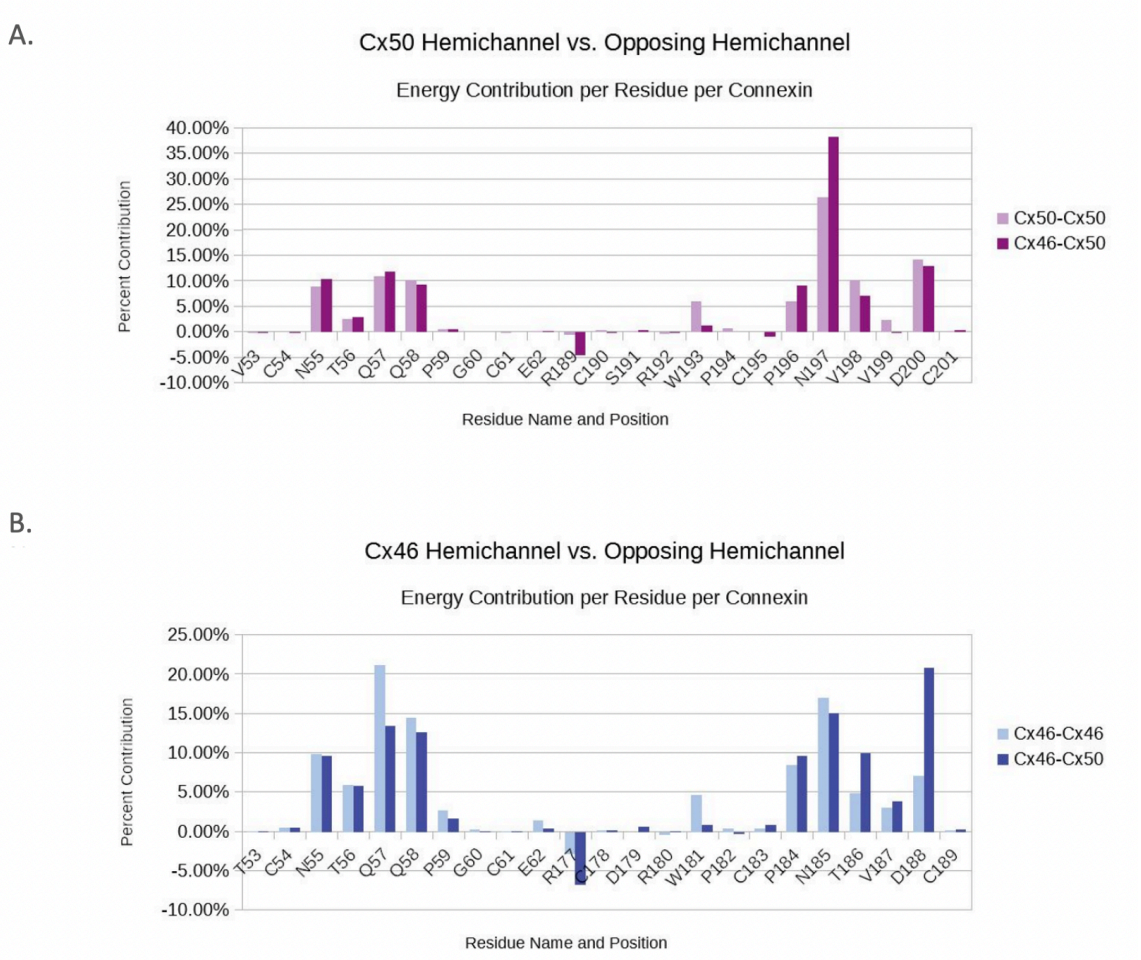


Figure 8. Energetic decomposition of homotypic vs. heterotypic hemichannel docking. *A) comparison of residual interaction energies at the docking interface of homotypic Cx46 (dark) and heterotypic Cx46-50 (light). B) comparison of interaction energies between at the docking interface between homotypic Cx50 (dark) and heterotypic Cx46-50 (light). Positive contributions stabilize the interaction, and negative values destabilize.*

Initially, heterotypic models of Cx46-Cx50 were generated, and subject to equilibrium MD simulations. This was followed by calculations using *namdenergy*[6] to extract the interaction energies between the two hemichannels and decompose them to the residue

level (Fig. 6). Energy decomposition highlighted which residues contributed most to stabilizing the hemichannel interaction.

The next steps in the project would be to generate heterotypic models of Cx46-Cx43, and Cx50-Cx43, the latter being incompatible in nature. These simulations would, ideally, highlight which residues of Cx50 disallow the hemichannel docking. Residues identified to contribute unfavorably to the docking stability would then be targets for mutagenesis. This work would shed light on the specific mechanisms underlying heterotypic compatibility.

Ca^{2+} Binding to Cx46 and Cx50

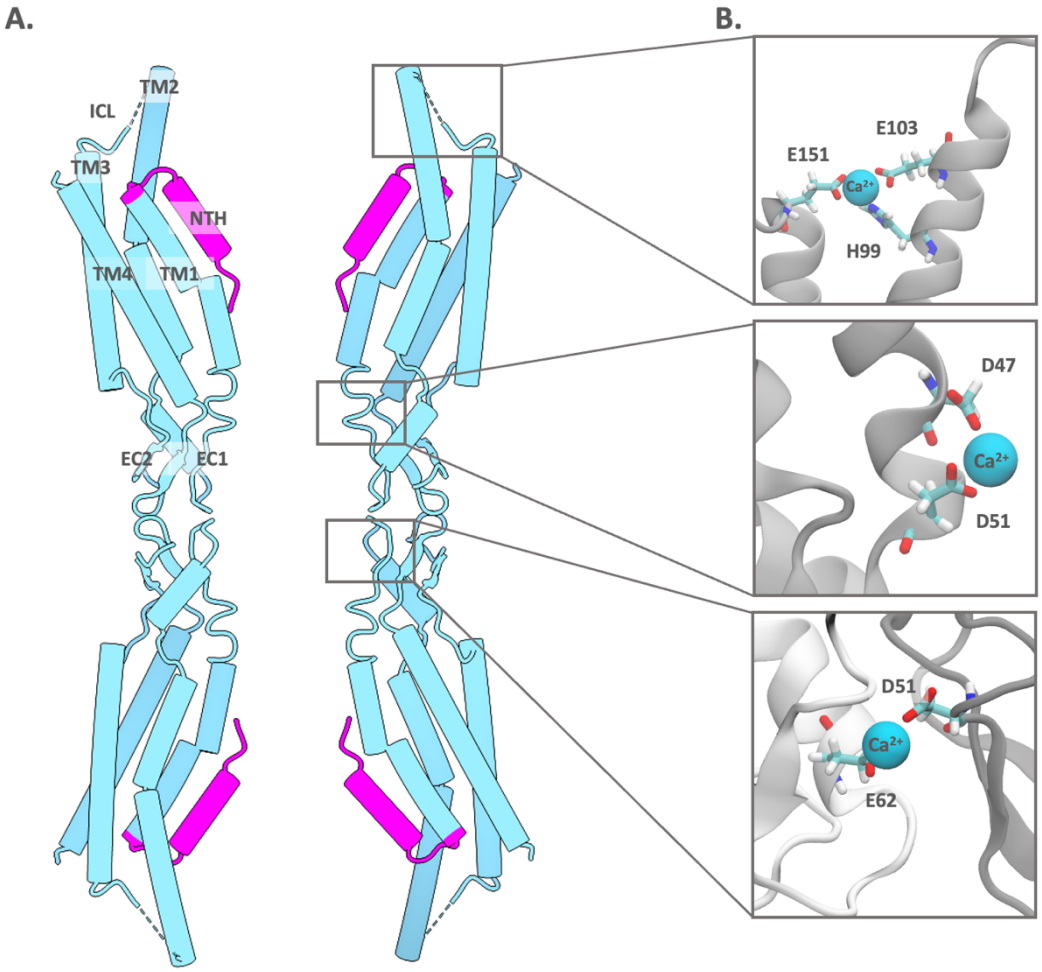


Figure 9. Putative Ca^{2+} binding sites in Cx50 gap junction intercellular channel. *A) Sliced view of Cx50 gap junction with the NTH highlighted in magenta, with topographical segments labelled in one of the connexin-50 monomers. B) Putative Ca^{2+} binding sites. Each Ca^{2+} ion remained bound after initial association.*

Ca^{2+} is a regulator of gap junctions, and hemichannels in particular[7]. Hemichannels are gated by extracellular Ca^{2+}, closing as the concentration increases[8]. Similarly, intracellular Ca^{2+} is known to modify the conductance of gap junction intercellular channels[8]. The Reichow Lab recently reconstructed Cx46 and Cx50 in the presence of Ca^{2+} (data not shown), in order to see the effects on channel structure.

The goal of this project was to analyze the dynamics of Cx46 and Cx50 in the presence of Ca^{2+} using experimental [Ca^{2+}]. Initial simulations of Cx50 in the presence of 50 mM Ca^{2+} revealed multiple binding sites for the divalent cation throughout the pore and intracellular regions (fig. 7). In all cases of Ca^{2+} binding, once it was bound it remained for the duration of the 75 ns exploratory simulation. Both binding sites within the pore are in the same position as two previously identified sodium binding sites, demonstrating that Na^{+} and Ca^{2+} compete for the same binding sites (data not shown).

References

1. Humphrey W, Dalke A, Schulten K. VMD: Visual Molecular Dynamics. *J Mol Graph*. 1996;(14):33-38.
2. Flores JA, Haddad BG, Dolan KA, et al. Connexin-46/50 in a dynamic lipid environment resolved by CryoEM at 1.9 Å. *Nat Commun*. 2020;11(1):1-11. doi:10.1038/s41467-020-18120-5
3. Valiunas V, Brink PR, White TW. Lens Connexin Channels Have Differential Permeability to the Second Messenger cAMP. Published online 2019.
4. Slavi N, Rubinos C, Li L, et al. Connexin 46 (Cx46) gap junctions provide a pathway for the delivery of glutathione to the lens nucleus. *J Biol Chem*. 2014;289(47):32694-32702. doi:10.1074/jbc.M114.597898
5. Huber GA, Kim S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys J*. 1996;70(1):97-110. doi:10.1016/S0006-3495(96)79552-8
6. Phillips JC, Braun R, Wang W, et al. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005;26(16):1781-1802. doi:10.1002/jcc.20289
7. Lopez W, Ramachandran J, Alsamarah A, Luo Y, Harris AL, Contreras JE. Mechanism of gating by calcium in connexin hemichannels. *Proc Natl Acad Sci U S A*. 2016;113(49):E7986-E7995. doi:10.1073/pnas.1609378113
8. Spray DC, White RL, Mazet F, Bennett VL. *Regulation of Gap Junctional Conductance*. Vol 17.; 1985.