

UNIVERSITÀ DEGLI STUDI DI GENOVA
Dipartimento di Neuroscienze, Riabilitazione, Oftalmologia,
Genetica e Scienze Materno-infantili - DINOGMI



ISTITUTO ITALIANO DI TECNOLOGIA, GENOVA
Centro di Neuroscienze e Tecnologie Sinaptiche - IIT NSYN



Corso di Dottorato in Neuroscienze, XXXII Ciclo (2016-2019)

Curriculum in Neuroscienze e Neurotecnologie

Settore Scientifico Disciplinare: FIS/07, Fisica Applicata

*Targeting Tight Junctions in Nanomedicine: a Molecular
Modeling Perspective*

Coordinatore:
Ch.mo Prof. Angelo Schenone

Supervisor:
Dr. Luca Maragliano
Ch.mo Prof. Fabio Benfenati

Studente:
Giulio Alberini,
matricola: 4309583

Anno Accademico 2018/2019

To my family

Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics and Mother and Child Sciences - DINOGMI, Università degli Studi di Genova, Largo P. Daneo, 3, 16132, Genova, Italy.

In agreement with the Foundation of the Italian Institute of Technology (IIT), Via Morego, 30, 16163, Genova, Italy.

Ph.D. Course in: Neurosciences, XXXII Cycle (2016-2019)

Curriculum: Neurosciences and Neurotechnologies

Targeting Tight Junctions in Nanomedicine: a Molecular Modeling Perspective

Candidate:

Giulio Alberini^{1,2}

Senior Ph.D. Student

1 Center for Synaptic Neuroscience & Technology (NSYN@UniGe), Istituto Italiano di Tecnologia, Largo Rosanna Benzi, 10, 16132, Genova, Italy.

2 Department of Experimental Medicine, Università degli Studi di Genova, Viale Benedetto XV, 3, 16132, Genova, Italy.

A Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Neuroscience - XXXII Cycle.

- **Settore Concorsuale di Afferenza:** 02/B3 FISICA APPLICATA

- **Settore Scientifico Disciplinare:** FIS/07 FISICA APPLICATA (A BENI CULTURALI, AMBIENTALI, BIOLOGIA E MEDICINA)

Declaration of Authorship

I, Giulio Alberini, declare that this Thesis titled, *Targeting Tight Junctions in Nanomedicine: a Molecular Modeling Perspective* and the work presented in it are my own.

I confirm that:

1. This work was done wholly while in candidature for a research degree at this Institution.
2. This dissertation has been composed by the author and has not been submitted in whole or in part for any other degree.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
5. I have acknowledged all main sources of help.
6. Where the Thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Date: 18/01/2020



©Copyright, 2020.
Giulio Alberini. All Right Reserved.

Competing Interests

The author declared no competing interests.

Copyright Disclaimer

In this Thesis, when copyrighted material is used, the necessary permissions have been obtained.

In this case, a note is explicitly introduced and further permissions must be obtained from the copyright holder.

We thank the copyright holders for giving permission for free use in this Thesis.

Biographical Sketch

Date and Place of Birth: December 27, 1989, Cremona (Italy).

- **Nov. 2019 - Feb. 2020** *Ph.D. Candidate* at the School of Neurosciences, Università degli Studi di Genova and *Research Fellow* at the Istituto Italiano di Tecnologia, Genova, Italy.

IIT personal webpage <https://www.iit.it/people/giulio-alberini>

- **Sept. 2016 - Dec. 2019** *Ph.D. Student-Fellow* in Applied Physics at the School of Neurosciences, Curriculum: Neurosciences and Neurotechnologies, Università degli Studi di Genova and Istituto Italiano di Tecnologia, Genova.
Ph.D. Project in Computational Biophysics under the supervision of Dr. L. Maragliano and Prof. F. Benfenati.
The Ph.D. topic is focused on the study of transport properties of unconventional paracellular membrane proteins using both standard and biased Molecular Dynamics simulations.
- **Sept. 2015 - July 2016.** M.Sc. Project in Physics (Supervisor Prof. A. Arcovito), Istituto di Biochimica e Biochimica Clinica, Policlinico Agostino Gemelli, Rome (Italy).
Training in Molecular Dynamics (with Dr. S. Della Longa) and in Bioinformatics (with Prof. A. Tramontano).
- **Oct. 2014 - July 2016.** *M.Sc. Degree (Laurea Magistrale) in Physics* (110/110 cum Laude), Università Cattolica del Sacro Cuore, Brescia (Italy).
Training in Matter Physics and Computational Biophysics. Thesis under the supervision of Prof. A. Arcovito and Prof. F. Borgonovi.

- **Oct. 2012 - July 2014.** *M.Sc. Degree (Laurea Magistrale) in Mathematics* (110/110 cum Laude), Università Cattolica del Sacro Cuore, Brescia (Italy).
Training in Mathematical Analysis and Mathematical Physics. Thesis under the supervision of Prof. M. Degiovanni.
- **Oct. 2008 - Oct 2012.** *B.Sc. Degree (Laurea) in Physics, with an additional year in Mathematics* (105/110), Università Cattolica del Sacro Cuore, Brescia (Italy).
- **Sept. 2003 - July 2008.** High School *Maturità Classica* (100/100) Liceo Classico - Scientifico M. G. Vida, Cremona (Italy).

External Reviewers and Thesis Committee

The author acknowledges the *External Reviewers*

- **Mauro Ferrario**
Full Professor of Physics
Università degli Studi di Modena e Reggio Emilia, Modena, Italy
<http://personale.unimore.it/rubrica/dettaglio/ferrario>
email: mauro.ferrario@unimore.it
- **Albert Lau**
Associate Professor of Biophysics and Biophysical Chemistry
John Hopkins School of Medicine, Baltimore, Maryland - USA
<http://biophysics.med.jhmi.edu/people/albert-lau/>
email: alau@jhmi.edu

and the members of the *Thesis Committee* of the Università degli Studi di Genova for taking the time to review the manuscript of the Ph.D. Thesis.

Publications

Peer Reviewed Articles (Chronological Order)

- **Alberini G, Benfenati F, Maragliano L.** *A Refined Model of Claudin-15 Tight Junction Paracellular Architecture by Molecular Dynamics Simulations.* **PLoS One, 2017.**
- **Alberini G, Benfenati F, Maragliano L.** *Molecular Dynamics Simulations of Ion Selectivity in a Claudin-15 Paracellular Channel.* **The Journal of Physical Chemistry B, 2018.**
- **Alberini G, Benfenati F, Maragliano L.** *Structural Mechanisms for ω -pore currents of Kv Voltage Sensor Domains. A Molecular Dynamics Investigation.* **In preparation.**

Chapter Books and Review Articles

- Bramini M, **Alberini G**, Benfenati F, Maragliano L, Cesca F. *Interactions Between 2D Graphene-Based Materials and the Nervous Tissue*. Advanced 2D Materials - Characterization, Production and Applications, **Publisher: Science Publishers (An Imprint of CRC Press/ Taylor & Francis Group), 2018**.
- Bramini M, **Alberini G**, Colombo E, Chiacchiarretta M, DiFrancesco M L, Maya-Vetencourt J F, Maragliano L, Benfenati F, Cesca F. *Interfacing Graphene-Based Materials With Neural Cells*. **Frontiers in Systems Neuroscience, 2018**.

Communications to Conferences

- 62nd Annual Meeting of the Biophysical Society. San Francisco (USA), 17-21 February 2018. [Poster presentation]

Awards (Chronological Order)

- **Graduate Student Fellowship**. Oct 2016. Graduate Student Fellowship for the three year PhD School in Neuroscience (November 2016 - October 2019), Università degli Studi di Genova and Istituto Italiano di Tecnologia (~ 50 000 €).
- **Winner of a National Class B ISCRA Project for High Performance Computing Resources**. Feb 2018. Winner of a national class B ISCRA project for the use of parallel computing resources (~ 1 700 000 hours) at the CINECA consortium (Principal Investigator Dr. Luca Maragliano).
- **Winner of a National Class C ISCRA Project for High Performance Computing Resources** Apr 2019. Winner of a national class C ISCRA project for the use of parallel computing resources (~ 200 000 hours) at the CINECA consortium (as Principal Investigator).
- **Winner of a National Class B ISCRA Project for High Performance Computing Resources**. Jun 2019. Winner of two national class B ISCRA projects for the use of parallel computing resources (~ 3 300 000 hours) at the CINECA consortium (Principal Investigator Dr. Luca Maragliano).

Contributors and Funding Resources

The project involved in this Ph.D. Thesis received funding from the European Union Horizon 2020 Research and Innovation programme under grant agreement No. 696656 and No. 785219.

Computing time allocations were granted by CINECA under the ISCRA initiative.

List of the Main Abbreviations and Notations

- **ADG** α -D-Glucose
- **AFM** Atomic Force Microscopy

- **ANM** Anisotropic Network Model
- **ASICs** Application Specific Integrated Circuits
- **BBB** Blood Brain Barrier
- **Cldn** Claudin
- **CV** Collective Variable
- **CG** Coarse Grained
- **CPE** *Clostridium Perfringens* Enterotoxin
- **DD** Domain Decomposition
- **ECL1** Extracellular Loop 1
- **ECL2** Extracellular Loop 2
- **ENM** Elastic Network Modeling
- **FE** Free Energy
- **FES** Free Energy Surface
- **FRET** Fluorescence Resonance Energy Transfer
- **GJ** Gap Junction
- **GNM** Gaussian Network Modeling
- **GPU** Graphic Processing Unit
- **HB** Hydrogen Bond
- **HPC** High Performance Computing
- **KO** KnockOut
- **KD** KnockDown
- **KV** Voltage Gated Potassium Channels
- **MD** Molecular Dynamics
- **MFEP** Minimum Free Energy Path
- **MIMD** Multiple Instruction Multiple Data
- **MPI** Messaging Passing Interface
- **NAMD** NANoscale Molecular Dynamics (formerly Not Another Molecular Dynamics Program)
- **NaV** Voltage Gated Sodium Channels
- **NMA** Normal Mode Analysis
- **NMR** Nuclear Magnetic Resonance
- **NPT** Isothermal Isobaric Ensemble
- **NVE** Microcanonical Ensemble

- **NVT** Canonical Ensemble
- **PD** Particle Decomposition
- **PDB** Protein Data Bank
- **PE** Processor Element
- **PFS** Potential Functional Site
- **PMF** Potential of Mean Force
- **POPC** 1-Palmitoyl-2-oleoyl-SN-glycero-3-phosphocholine
- **RC** Reaction Coordinate
- **SD** Spacial Decomposition
- **SSw** Single Sweep
- **TAMD** Temperature-Accelerated Molecular Dynamics
- **TM** Trans Membrane
- **TJ** Tight Junction
- **TST** Transition State Theory
- **US** Umbrella Sampling
- **VMD** Visual Molecular Dynamics
- **VTMM** Voronoi Tessellated Markovian Milestoning
- **WHAM** Weighted Histogram Analysis Method
- **WT** Wild Type

Sommario

Negli organismi complessi, le cellule interagiscono formando barriere altamente specializzate, necessarie per il normale sviluppo di tessuti e organi. Queste barriere regolano il passaggio di molecole attraverso la superficie cellulare e nell'ambiente tra cellule adiacenti, detto spazio paracellulare.

Le giunzioni occludenti (Tight Junctions, TJs) sono una classe di complessi proteici che costituiscono uno strato parzialmente permeabile nello spazio paracellulare.

La principale classe di proteine coinvolta nella formazione delle TJs è la famiglia delle Claudine (Cldns), le quali formano complesse strutture oligomeriche tessuto-specifiche all'interfaccia tra due le cellule. Lo studio delle proprietà strutturali delle Claudine costituisce un campo di vasto interesse nella comunità scientifica, soprattutto per le potenziali applicazioni in campo farmaceutico e medico. Tuttavia, i dettagli atomici di questi complessi proteici restano ancora oscuri, a causa di diversi limiti delle tecniche sperimentali attualmente disponibili. Il lavoro di questa Tesi di Dottorato si è concentrato su questo problema, cercando di dare un contributo attraverso la modellistica tridimensionale di queste proteine.

Il punto di partenza della ricerca è stato lo studio di un modello, incompleto, di architettura paracellulare formata da Claudine 15 (Cldn15) [1], responsabili del controllo del trasporto di piccoli cationi nei tessuti intestinali. Diversi approcci computazionali

sono stati utilizzati in questo lavoro per completare e perfezionare la struttura iniziale del primo modello strutturale di canali paracellulari formati da Claudine (denominato *modello di Suzuki*), con particolare riferimento a simulazioni atomistiche in doppio strato lipidico che hanno dimostrato la stabilità della struttura, e evidenziato la presenza di un poro paracellulare con un diametro di $\sim 5\text{-}6 \text{ \AA}$.

Successivamente, è stato analizzato il processo di diffusione di ioni monovalenti e α -D-glucosio (ADG) all'interno della cavità del modello, quantificandone sia le proprietà termodinamiche sia quelle cinetiche. In particolare, il modello descrive un poro permeabile al passaggio di cationi, ma non di cloro. Questo meccanismo di permeabilità è ottenuto attraverso l'azione di un filtro selettivo formato dai residui a carica negativa D55 e D64 che rivestono la parte più stretta del canale, garantendo in particolare il flusso di sodio. Inoltre, il poro non permette il trasporto di ADG, a causa dell'azione di residui polari e carichi.

I risultati esposti in questa Tesi sono in accordo con le conoscenze ottenute precedentemente attraverso studi sperimentali e dimostrano che il modello è in grado di descrivere la dinamica paracellulare sia in termini di dimensione del poro, sia di selettività di carica, offrendo nuove informazioni sulle proprietà delle TJs.

Ancora più importante, il sito di selettività per la Claudina 15 include diversi residui non conservati dagli altri membri della famiglia delle Claudine, la cui variabilità potrebbe determinare una diversa funzione a seconda del tessuto in cui sono espressi.

Per esplorare questa possibilità, abbiamo generato una serie di modelli della Claudina cerebrale 5 (Cldn5), utilizzando la stessa configurazione del modello di Suzuki, attraverso tecniche di bioinformatica strutturale. Le analisi di questi modelli evidenziano la possibilità della trasferibilità del modello Suzuki a un diverso rappresentante della famiglia delle Claudine. In particolare la composizione della cavità paracellulare dimostrerebbe il diverso comportamento della dinamica paracellulare. Specificamente, il residuo D55 della Claudina 15, che è fondamentale per la selettività dei canali paracellulari intestinali, viene sostituito, nel caso della Claudina 5, da un residuo Q neutro, che potrebbe abolire la selettività di carica come previsto per lo spazio paracellulare della Barriera emato-encefalica.

Globalmente, questo lavoro mostra come, con l'utilizzo di una combinazione di strumenti computazionali, sia possibile ottenere modelli tridimensionali di strutture paracellulari con le note proprietà fisiologiche. Complessivamente, il nostro approccio ha fornito il primo modello ottimizzato in dettaglio atomico di un poro paracellulare, che potrà guidare diverse strategie sperimentali, con particolare riferimento agli sviluppi di nuovi vettori artificiali capaci di regolare la permeabilità paracellulare.

Abstract

Outer and inner biological barriers maintain the proper physical and chemical conditions necessary to organisms function. They are designed by nature to keep unsafe material out while allowing necessary molecules to cross.

These cellular architectures include multiprotein junctional complexes belonging to two opposite and adjacent cells, named Tight Junctions (TJs), comprising proteins of the Claudin (Cldn) family. Cldns form the backbone of these junctions and regulate molecular diffusion in the space between cells (paracellular transport) in a tissue-specific manner. Because of their role, they are subject of interest in the field of pharmaceutical and medical applications. Unfortunately, the atomic details of the Cldn structures are still poorly understood.

In this Thesis, we focused on the study of the first structural model of a Cldn based architecture of a paracellular channel [1] made by Cldn15 protomers, named the *Suzuki*

model. Cldn15 proteins mediate the paracellular transport of small monovalent cations along a concentration gradient in the intestinal tissues, playing an important role in Na⁺ homeostasis.

The main result of our work is the first available atom-detailed structural model of a Cldn15-based paracellular pore, refined *via* Molecular Dynamics (MD) simulations. Molecular Modeling was used to complete the complex and MD simulations in double bilayers allow to demonstrate the stability of the model in a realistic environment, showing a stable pore diameter of 5-6 Å.

Then, the pathways of diffusion of monovalent ions and α-D-Glucose (ADG) inside the model were also investigated, *via* atomistic simulations. The Potential of Mean Force (PMF) and kinetics properties of the molecules are calculated using different computational methods, such as Umbrella Sampling (US) simulations, Voronoi Tessellated Markovian Milestoning (VTMM), and the Single Sweep (SSw) algorithm together with the String Method to evaluate the Minimum Free Energy Path (MFEP).

The paracellular pore is able to provide a selective passage of cations and not of Cl⁻ *via* the action of a negatively charged filter, which involves pore-forming D55 residues (*i.e.* residues with the side chain inside the cavity of the channel) and pore-lining D64 residues (*i.e.* residues with the side chain outside the cavity of the channel), consistently with the Cldn15 physiological role. Moreover, the pore is not permissive for ADG transport, because of the action of charged and polar residues.

All these results agree with previous experimental studies and demonstrate that this refined model is able to describe consistent pore sizes with a strong charge-selectivity, offering new elements for the comprehension of TJ functions.

Most importantly, the Cldn15 selectivity filter includes various non-conserved residues whose variability in other Cldns could determine a different functional behavior, depending on the tissue in which they are expressed. Of particular interest is the Cldn5, present in the TJs of the Blood Brain Barrier (BBB), a main target of drug development studies. To investigate this possibility, a set of Cldn5 Suzuki-like pore models were prepared using homology modeling and structural bioinformatics.

In particular, the Cldn15 D55, which is pivotal for the charge selectivity of intestinal Cldns, is replaced by a neutral Glutamine in Cldn5 (Q57), potentially abolishing the charge selectivity of the complex. Furthermore, the pore lining Cldn15 D64 residues are replaced by a group of hydrophobic Valine residues (V66). This surplus of uncharged residues in the narrow region of the Cldn5 pore could result in a smaller size of the cavity, consistent with the enhanced tightening of the BBB TJs. This further analysis highlights the robustness of the transferability of the Suzuki model to different Cldns. More generally, this work shows that, using a combination of MD methods, it is possible to obtain all the main structural properties of a specific Cldn paracellular environment *in silico*, that could be shared by other Cldn-based architectures. These results could be helpful for the rational design of new artificial vectors able to regulate the paracellular permeability, in a reversible and safe manner.

Outline of the Thesis

The Thesis is organized as follows.

- **Chapter 1** illustrates the main features of Cldn-based paracellular architectures, and it is divided in the following sections.

- **Section 1.1** describes the structure of biological barriers, their physiological relevance and the importance of using them as targets for drug design.

- **Section 1.2** introduces the mechanism of transport across barriers.
 - **Section 1.3** describes the architecture of TJs and the basic principles of their function.
 - **Section 1.4** focuses on the family of Cldn proteins which form the tissue-specific backbone of TJs.
 - **Section 1.5** introduces the features of the crystallographic structure of the Cldn15 protomer. This structure is the only available experimental configuration of a Cldn protomer, without the binding of additional molecules. After the publication of the single protomer, a model of a Cldn15 paracellular channel based on the interaction of multiple Cldn15 protomers has been published [1]. This section also reviews the main features of this structural model.
- **Chapter 2** describes the aim of the work.
 - **Chapter 3** illustrates the methods. In particular:
 - **Section 3.1** is focused on the details of standard MD simulations, used to study the equilibrium and transport properties of systems of atoms and molecules.
 - **Section 3.2** shows the main concepts of the theory of Statistical Mechanics, which are at the basis of MD techniques and algorithms.
 - **Section 3.3** deals with methods to calculate dynamic properties and kinetics.
 - **Section 3.4** reviews the main computational techniques to study rare events in biological systems.
 - **Section 3.5** introduces the family of algorithms used in this Thesis for thermodynamics and kinetics calculations of Cldn15 paracellular channels, namely Umbrella Sampling, Markovian Milestoning MD and Single Sweep.
 - **Section 3.6** describes the main features of a family of *coarse-grained* methods named Elastic Network Models (ENMs).
 - **Section 3.7** briefly introduces the basic strategies of Parallel Computing to improve the production of MD simulations.
 - **Chapter 4** defines all the softwares, the procedures and the computational protocols employed.
 - **Section 4.1** illustrates all the computational methodologies.

- **Section 4.2** describes the computational resources used for all the simulations.
- **Chapter 5** reports the results and the associate discussion.
 - **Section 5.1** illustrates the results obtained for the refinement of the Cldn15 model and it is organized as follows:
 1. **Subsection 5.1.1** explains the structural features obtained from standard MD simulations.
 2. **Subsection 5.1.2** explains all the PMF calculations and kinetics calculations with 1 Collective Variable (CV), to study ionic permeation events through the cavity of the Cldn15 paracellular model.
 3. **Subsection 5.1.3** studies the dehydration events observed during ionic permeations.
 4. **Subsection 5.1.4** highlights the results obtained with PMF calculations based on 3 CVs, using the Single Sweep (SSw) algorithm, to study ionic and molecular permeations.
 - **Section 5.2** discusses the previous results.
 1. **Subsection 5.2.1** highlights the structural features from the Cldn15 paracellular channel model, obtained *via* Molecular Modeling and MD simulations.
 2. **Subsection 5.2.2** is devoted to the discussion of the results obtained from PMF and kinetics calculations to evaluate the features of permeation through the pore region of the model.
 3. **Subsection 5.2.3** describes the transferability of the Suzuki model from an intestinal Cld15 TJ to a brain Cldn5 TJ, using further molecular modeling results.
- **Chapter 6** illustrates the conclusions and the different perspectives of the work.
 - the contents of **Section 6.1** suggest routes to validate the possibility of extending the Suzuki model to Cldn complexes of the BBB.
 - **Section 6.2** describes how the structural models presented in this Thesis can be employed together with experimental approaches to design new molecules capable of modulating paracellular permeability in a safe manner.
 - **Section 6.3** discusses the impact of the Thesis.
- **Chapter 7** describes the structural modeling and MD refinement of various Voltage Gated Sodium (NaV) channels I performed within a collaboration with Prof. Cameron F. Abrams (Drexel University, Philadelphia, USA) and Dr. Alexis Paz (Universidad Nacional de Córdoba, Córdoba, Argentina). In particular

- **Section 7.1** summarizes the main functional and structural properties of NaV channels and their role mediating Na⁺ influx in excitable cells;
- **Section 7.2** introduces all the methods and protocols followed for the production of the models.
- **Section 7.3** illustrates the first results of the ongoing collaboration with Prof. Cameron F. Abrams and Dr. Alexis Paz, to perform MD simulations of Na⁺ permeation, through a NaV1.2 model.
To this aim, a TAMD-based PMF reconstruction, named Temperature Accelerated MD/On-The-Fly Parameterization (TAMD/OTFP) [2–6] is used.

Finally, a report is introduced in **Appendix A** to describe the Ph.D. courses I followed during my training.

The material presented in this Thesis is based on the published papers [7] and [8].

Contents

1	Introduction	29
1.1	Biological Barriers and Cell Junctions	29
1.2	Transcellular and Paracellular Transport	29
1.3	The Multiple Functions of Tight Junctions	31
1.4	The Family of Claudin Proteins	31
1.4.1	Channel and Barrier Forming Claudins	33
1.5	Structural Features of Claudins	34
1.5.1	Single Claudin Crystal Structures	34
1.5.2	The Suzuki Model for Claudin-15 Paracellular Channels	36
1.5.3	Further Validation of the Double-row Arrangement Introduced by the Suzuki Model	40
2	Aim of the Thesis	43
3	Methods	45
3.1	Molecular Dynamics Simulations	45
3.1.1	Force Fields	46
3.1.2	Models for Water Molecules	48
3.1.3	The Cutoff Radius	48
3.1.4	Boundary Conditions and the Minimum Image Convention	49
3.1.5	Integrators	49
3.1.6	Statistical Ensembles	51
3.1.7	Constraint Conditions and the Integration Time-step	51
3.1.8	Evaluation of Non-bonded Interactions	52
3.1.9	Energy Minimization	54
3.1.10	Equilibration and MD Simulations	54
3.2	Introduction to Statistical Mechanics for Thermodynamics	57
3.3	Basic Concepts of Kinetic Calculations	58
3.4	Enhanced Sampling Techniques in Biomolecular Simulations	59
3.5	Details of the Methods Applied in This Thesis	61
3.5.1	Umbrella Sampling Simulations and the Weighted Histogram Analysis Method	61
3.5.2	Milestoning MD Simulations	64
3.5.3	Single-Sweep Simulations and Pathway Reconstruction	65
3.5.4	Root-Mean-Square Deviation	67
3.6	Coarse Grained Elastic Network Models	67
3.6.1	Gaussian Network Models	67
3.6.2	Anisotropic Network Models	68
3.7	Basics of Parallel Calculus for Molecular Dynamics Simulations	68

4	Materials	71
4.1	Computational Protocols	71
4.1.1	Preparation of Cldn Monomers	71
4.1.2	Unbiased MD Simulations	75
4.1.3	General protocols for PMF calculations	78
4.1.4	Protocols for Elastic Network Models	80
4.2	Computational Resources	80
5	Results and Discussion	81
5.1	Structural Features of the Refined Cldn15 Paracellular Channel	81
5.1.1	Analysis of the Standard MD Simulations	81
5.1.2	PMF Calculations and Ion Permeation Rates with US and VTMM Simulations	91
5.1.3	Pore Size and Hydration of Na ⁺ during Permeation	93
5.1.4	PMF Calculations with SSw Simulations	94
5.2	Discussion	95
5.2.1	Structural Features of the Models	95
5.2.2	Solute Permeation Features	98
5.2.3	Structural Features of the Cldn5 Suzuki Model	101
6	Conclusions and Future Perspectives	105
6.1	Further Computational Validations	105
6.2	Future Computational Work and Experiments	106
6.3	Thesis Impact	106
7	Further Projects	107
7.1	An Introduction to the Structure-Function Features of Voltage Gated Sodium Channel	108
7.2	Modeling of NaV Channels	109
7.3	MD Simulations of the NaV1.2 Channel in an Open State	110
Appendices		
	Appendix A Academic Transcript	123
	Bibliography	125

List of Figures

- 1.1 **Representation of an intercellular Double Transcellular pathway.** A schematic representation of a Gap Junction which is an example of Double Transcellular protein complex. The flux of solutes passes the surface of the two cells. Pores are colored in magenta and the Connexin protomers are represented by cyan cylinders. Gray lines and plates indicate cell membrane surfaces, and left-right double arrows indicate the pathways and directions of permeating molecules. Modified picture from [1], <https://www.sciencedirect.com/science/article/pii/S0022283614005713>. Copyright (2015), with permission under the terms of the Creative Commons Attribution License. 30
- 1.2 **Representation of an intercellular Paracellular pathway.** Paracellular pores formed by Claudin protomers mediate transport which is parallel with the lipid bilayers. The pore walls are formed by the protein extracellular domains. Pores are colored in magenta and the Claudin protomers are represented by cyan cylinders. Gray lines and plates indicate cell membrane surfaces, and up-down double arrows indicate the pathways and directions of permeating molecules. Modified picture from [1], <https://www.sciencedirect.com/science/article/pii/S0022283614005713>. Copyright (2015), with permission under the terms of the Creative Commons Attribution License. 30
- 1.3 **Freeze-fracture electron micrograph of TJ strands.** Electron microscopy of freeze-fracture replicas of TJs. The arrows point to TJ-like strands. The inset shows a magnified view of the boxed area. The scale bar represents 100 nm. Reprinted from [1], <https://www.sciencedirect.com/science/article/pii/S0022283614005713>. Copyright (2015), with permission under the terms of the Creative Commons Attribution License. 31
- 1.4 **Structure of TJs in epithelial cells.** TJs contain three types of transmembrane proteins: Claudins, Occludins, and Junction Adhesion Molecules (JAMs) as well as scaffold proteins such as Zonula Occludens-1 (ZO). Reprinted from [9], [https://www.allergologyinternational.com/article/S1323-8930\(17\)30159-4/fulltext](https://www.allergologyinternational.com/article/S1323-8930(17)30159-4/fulltext). Copyright (2019), with permission under the terms of the Creative Commons Attribution License. 32
- 1.5 **Structure of a Cldn protomer.** Model of Cldn protein showing predicted topology and secondary structure as well as the different functional domains. Roman numerals indicate the predicted α -helical TM domains. Reprinted from [10], <https://www.physiology.org/doi/full/10.1152/physrev.00019.2012>. Copyright (2013), with permission from the American Physiological Society. 33

1.6	<p>Structure of TJs in epithelial cells. Claudin-based biological systems revealed by Cldn KnockOut (KO) and KnockDown (KD) mice systems. KO and KD mice of specific paracellular channel-forming Cldns mostly show metabolic disorders, including malnutrition due to intestinal malabsorption and decreased bile flow. Abbreviations: WT, Wild Type. Reprinted from [11], https://www.sciencedirect.com/science/article/abs/pii/S0968000418301932. Copyright (2019), with permission from Elsevier.</p>	35
1.7	<p>Crystal structure of Cldn15. Ribbon representation of the Cldn15 structure (PDB ID 4P79). The bars indicate the membrane boundaries with the extracellular (Ext) and cytosolic space (Cyt). Missing residues are highlighted with dashed lines. Adapted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.</p>	37
1.8	<p>The crystallographic Cldn15 linear arrangement. Representation of the crystallographic linear arrangement of Cldn15 protomers (ribbon representation) aligned along the b axis, viewed from the extracellular space (panel A) and from the membrane (panel B). The red square in panel B highlights the region of the lateral interaction between protomers. Panel C shows a zoom of the region where the critical residues are introduced. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.</p>	37
1.9	<p>Double <i>cis</i>- arrangement of Cldn15 protomers. A, Three Cldn15 protomers (green, cyan and purple) are aggregated with another linear group of Cldns (gold) <i>via</i> a second antiparallel <i>cis</i>- interface (<i>face-to-face</i>) formed by the close vicinity of β4 strands and highlighted in B. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.</p>	38
1.10	<p>V1 and V2 regions in the double <i>cis</i>- arrangement of Cldn15 protomers in the Suzuki model. A, Multiple sequence alignment of the extracellular regions in mCldn15 and other classic Cldns in mouse (mCldn1-9, 10a, 10b,14,17,19). Numbering above the sequences is based on the full-length mCldn15 protein. Secondary structure elements seen in the mCldn15 crystal structure are indicated above the sequences as arrows (β-strands; magenta), cylinders (α-helices; cyan), and lines (loops; black). The region that is disordered in the crystal structure is shown as a broken line. The consensus sequence (W-LW-C-C) is highlighted in yellow, and highly conserved residues are highlighted in gray. Red letters in mCldn15 and mCldn2 indicate important residues for their cation selectivity. Magenta and blue boxes indicate regions defined as variable sequence regions, V1 and V2, respectively. B and C, Ribbon diagram of the modeled antiparallel double-row arrangement of mCldn15 molecules in TJ strands, viewed perpendicular to (panel B) and parallel with (panel C) the membrane plane. Magenta and blue ovals indicate the positions of the V1 and V2 regions, respectively. Reprinted from [1], https://www.sciencedirect.com/science/article/pii/S0022283614005713. Copyright (2015), with permission under the terms of the Creative Commons Attribution License.</p>	39

1.11	The Cldn15 <i>cis</i>- arrangement. Side view of the double-row complex. The bars depict the TM boundaries and the lines indicate the half-pipe. Reprinted from [8], https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484 . Copyright (2018), with permission from the American Chemical Society.	40
1.12	The Suzuki model of TJ paracellular channels. Red bars indicate the membrane boundaries and the oval curves indicate pores through which paracellular diffusion occurs. Reprinted from [8], https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484 . Copyright (2018), with permission from the American Chemical Society.	40
1.13	The double-row structure of TJ strand fibrils. A, Cldn11 fibrils visualized by unidirectional platinum shadowing Freeze-fracture. B, Magnification of a single Cldn11 fibril. Reprinted from [12], https://www.nature.com/articles/s42003-019-0319-4 . Copyright (2019), with permission under the terms of the Creative Commons Attribution License.	41
3.1	Illustration of the main bonded interactions described in Force Fields. Bonded interactions described in a Force Field: r represents a bond distance; θ represents the bond angle; ϕ gives the dihedral angle; the out-of-plane angle α may be controlled by an improper dihedral φ . Note that the improper planarity deviation (α) is not used in the potential energy model, but is shown to clarify that improper interactions are used to establish molecular planarity. Reprinted from [13], https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20289 . Copyright (2005), with permission from John Wiley and Sons, Inc; 2005 Wiley Periodicals, Inc.	47
3.2	Illustration of Periodic Boundary Conditions. The supercell positioned of a single peptide (α helix ribbon) in water solvent (three point scheme) at the center of nine cells. Reprinted from [14], https://pubs.acs.org/doi/abs/10.1021/acs.jpcc.7b10830 . Copyright (2018), with permission from the American Chemical Society.	49
3.3	Multiscale simulations for biological processes. Application ranges for molecular modeling techniques at different resolutions: quantum, all-atom, coarse grained, and mesoscale. The plot shows approximate ranges of time scales and system sizes (lengths). Reprinted from [15], https://pubs.acs.org/doi/abs/10.1021/acs.chemrev.6b00163 . Copyright (2016), with permission from the American Chemical Society.	59
3.4	Schematic representation of nodes. Every node is connected to its spatial neighbors by uniform springs. Distance vector between two nodes, i and j , is shown by an arrow and labeled R_{ij} . Equilibrium positions of the i^{th} and j^{th} nodes, R_i^0 and R_j^0 , are shown in the Cartesian coordinate system. R_{ij}^0 is the equilibrium distance between the nodes. ΔR_i and ΔR_j , are the instantaneous fluctuation vectors. Reprinted from [16], https://pubs.acs.org/doi/10.1021/cr900095e . Copyright (2010), with permission from the American Chemical Society.	69

4.1	Model of paracellular TJ channels proposed by Suzuki et al. The model of Cldn-based paracellular channels proposed by Suzuki et al., represented as ribbons and viewed from the apical (A) and the lateral side (B). The pore regions are marked with purple circles, and the V1 and V2 regions (see text) are indicated. Horizontal lines in panel A indicate the membrane boundaries of the two cells. Adapted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	72
4.2	Modeling of Cldn15 loops. Superposition of the Cldn15 structures named <i>Model1</i> , used for the monomer and the single-pore, and <i>Model2</i> for the the double-pore. The identical folding is pictured in pink, while the two different conformations of the loops, ECL1 and ECL2, are colored in brown for <i>Model1</i> and cyan for <i>Model2</i> , respectively. Secondary structure elements are labeled in black and red. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	73
4.3	The single and double-pore channel models. A , structure of the single-pore system viewed from the apical side, in ribbon (left) and surface representation (right). The complex is formed by the trans interaction of two <i>cis</i> - dimers (red-green, P1-P2 dimers; and blue-yellow, P3-P4 dimers). B , Structure of the double-pore system represented and viewed as in panel A . Horizontal lines indicate the membrane boundaries. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	74
4.4	Flow chart of the ModRefiner algorithm. The two step ModRefiner procedure. If backbone and side-chain atoms are already included in the configuration of the model, the algorithm starts with energy minimization. Reprinted from [17], https://www.cell.com/biophysj/references/S0006-3495(11)01245-8 . Copyright (2011), with permission from Elsevier.	75
4.5	The equilibrated Cldn15 monomer. Viewed from the transmembrane domain (left) and from the extracellular environment (right). Cldn15 monomer (rainbow cartoon) is embedded in a POPC bilayer, shown as wire structures. Phosphorus atoms of the membrane are introduced as spheres. Water molecules and ions are not shown for clarity. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	76
4.6	Conformation of the equilibrated single-pore system. The structure of the single pore, after the equilibration protocol. Protomers are shown as cyan spheres. Each <i>cis</i> dimer is embedded in a hexagonal POPC bilayer, shown as wire structures with phosphorus atoms as spheres. Solvent molecules are not shown for clarity.	77

4.7	Legend of the protomers. The nomenclature and color legend for each protomer of the Suzuki model is introduced. For single pore simulations the system is restricted to the four protomers P1-P4, included in the pink rectangle. Horizontal lines indicate the membrane boundaries. Adapted and reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	79
5.1	Structural features of the Cldn15 monomer. A , Superposition of the starting Cldn15 <i>Model1</i> monomer (cyan) and the final configuration from the MD run (brown). B , The side chain of R79 establishes two HBs with the main-chain carbonyl group of L48. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	82
5.2	RMSD of Cldn15 monomer. RMSD values of the protein backbone (A), and of the backbone of the two extracellular loops, ECL1 and ECL2 (B) along the simulation of the Cldn15 monomer. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	83
5.3	RMSD of the single-pore. A , RMSD values of the backbone atoms of each protomer in the single-pore simulation with respect to the initial model. B and C , superposition of the initial (pink ribbon) and final (orange ribbon) configurations of the channel, viewed from the lateral and apical sides, respectively. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	83
5.4	RMSD of the extracellular domains of the single-pore. RMSD values of the backbone of the extracellular loops ECL1 (black) and ECL2 (red) in the four protomers. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	84
5.5	Pore cavity region of the single-pore. Representation of the pore cavity region (blue surface) of the paracellular channel (represented with the Tube style of the VMD software). Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	85
5.6	The pore size and structure in the single-pore. A , time evolution of the minimal pore radius along the single-pore simulation. B , the pore profile along the channel axis for the configuration taken at 130 ns, with red and blue bars indicating the C _α positions of D55 and D64 residues, respectively. C , The starting conformation of the single-pore and three snapshots extracted at 30 ns (D), 80 ns (E), and 130 ns (F). Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	85

5.7	Cross-distances between facing C52 C_α atoms in the single-pore. Measure of cross-distances between facing C52 C _α atoms. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	86
5.8	Hydrophobic contacts in the single-pore. A Contacts between the P1 A152 residue and the conserved residues of the P4 ECH region. B Hydrophobic interactions between the ECL1 segments of diagonally opposed protomers. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	87
5.9	RMSD of the double-pore structure. A, RMSD values of the backbone atoms of each protomer along the double-pore simulation. B and C, superposition of the initial (pink ribbon) and the configuration at 30 ns, viewed from the lateral and apical side, respectively. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	88
5.10	Cross-distances between facing C52 C_α atoms and <i>Cis</i>-linear arrangement in the double-pore. A Cross-distances between facing C52 C _α atoms. B <i>Cis</i> -linear arrangement in the double-pore. Distances between the M68 sulfur atom and the center of mass of F146 benzene ring. In the legend of B , the first protomer name in the pair is that of M68. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	89
5.11	Hydrophobic <i>trans</i>-interactions and pore minimal radius of the double-pore. A, Representation of the <i>trans</i> -ECL2-ECL2 interactions as surfaces, and colored following the Kyte-Doolittle scale: from blue (most hydrophilic), to orange (most hydrophobic). B, Evolution of the minimal radii of the two pores. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	90
5.12	Amino-acid composition of the paracellular pore surface. Representative structure of the single pore model, used for thermodynamic and kinetic calculations, using the same legend of Fig 4.3. A, Lateral, zoomed view of the Cldn15 TJ pore, with the surface representation used for protomers 3 and 4, only. The side chains of acidic residues in protomers 3 and 4 are shown as spheres; the dashed line represents the axis used in the simulations of ion permeation. B, Cut-away view of the channel, obtained with a 90° rotation of the figure in panel A. The section shows only protomers 3 and 4 for clarity and with acidic residues numbering indicated for protomer 4. Note that the β-strands are tilted with respect to the channel axis, as in β-barrel structures.	91

5.13	Energetics of ion permeation (US-WHAM simulations). PMF profiles for the permeation of Na ⁺ (black), K ⁺ (green) and of Cl ⁻ (red), obtained from US-WHAM calculations. The error estimation for each window, obtained with bootstrap analysis, is represented with bars. The position of the C _α atoms of D55 and E46 residues are represented as dashed and dotted lines, respectively. The C _α atoms of D55 residues are located at the narrow region of the pore (windows at 28, 32, 34, 36 Å, respectively, see text). Reprinted from [8], https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484 . Copyright (2018), with permission from the American Chemical Society.	92
5.14	Energetics and kinetic of ion permeation (VTMM MD simulations). A , PMF profiles for the permeation of Na ⁺ (black), K ⁺ (green) and of Cl ⁻ (red). B , MFPTs along the pore axis. Reprinted from [8], https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484 . Copyright (2018), with permission from the American Chemical Society.	93
5.15	Number of oxygen atoms coordinating the Na⁺ ion in US windows. A , Window at 2 Å. B , Window at 32 Å: the black line indicates the coordinating oxygen atoms from water molecules, while the red line indicates those from the D55 sidechains. C , Average of the number of oxygen atoms coordinating the Na ⁺ , as a function of the reaction coordinate, from all the US simulations. Reprinted from [8], https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484 . Copyright (2018), with permission from the American Chemical Society.	94
5.16	Snapshots of partial dehydration events of a sodium ion in the pore. Partial dehydration events of a Na ⁺ ion in the pore region. A , Snapshot extracted from the US simulation at 32 Å. B , Snapshot extracted from the unbiased MD simulation. The ice-blue sphere is the Na ⁺ ion, small cyan spheres represents water molecules in the ion hydration shell, while all other water molecules in the system are not shown. The Na ⁺ ion is hydrated with only 4 and 3 water molecules, respectively. The side chains of the D55 residues are represented as red sticks, and the protein as orange VdW spheres. Reprinted from [8], https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484 . Copyright (2018), with permission from the American Chemical Society.	95
5.17	3D PMF profile of Na⁺ permeation through the channel from SSw simulations. A , Isosurfaces for the 3D PMF map for the coordinates of Na ⁺ inside the channel (blue: 0 kcal/mol; mauve: 1 kcal/mol; azure: 3 kcal/mol; grey: 8 kcal/mol). The D55 C _α atoms are shown as red spheres. The yellow line represents the MFEP. B , The PMF profile along from the MFEP.	96
5.18	PMF profile of additional cations from SSw simulations along their MFEP. PMF profiles obtained from SSw simulations for Li ⁺ , K ⁺ , Rb ⁺ , Cs ⁺	97
5.19	PMF of α-D-Glucose permeation through the channel. The PMF barrier of α-D-Glucose (ADG) permeation through the channel, calculated via SSw simulations.	97
5.20	Pore characterization in Cldn15 and Cldn5 complexes. View of the Cldn15 (brown model) and Cldn5 (green model) Suzuki pore structures, perpendicular to the pore axis. The models are described with the ribbon style and their surface with a dot-mesh. Pore-lining residues located in the narrow region of the filter (Cldn15 D55 and Cldn5 Q57) are highlighted with spheres.	101

5.21	Surface characterization of the <i>cis face-to-face</i> interaction for Cldn15 and Cldn5 dimers. Cut-away view of the Cldn-based complex for Cldn15 (brown model) and Cldn5 (green model) proteins. Only two proteins forming the <i>cis face-to-face</i> interaction are shown. Charged residues located in the cavity are shown with VdW spheres.	102
5.22	Potential Functional Sites for <i>cis- face-to-face</i> Cldn dimers. Potential Functional Sites for <i>cis- face-to-face</i> for the Cldn15 brown dimer (D55, V59, W63) and for the Cldn5 green dimer (T42, Q44, Q57, S58, H61, K65). Residues are shown in stick style for clarity. The potential conserved cysteine bridge is highlighted in purple.	103
7.1	2D schematic map of the NaV α subunit. A The typical α subunit of a Mammalian NaV channel is illustrated as a transmembrane folding diagram in which cylinders represent transmembrane α helices and lines represent loop regions, in proportion to their length. The Roman numerals indicate the four homologous repeats and the Arabic numerals are used to label the six transmembrane helices of each repeat. The S4 helices are colored in red with + signs indicating gating charges. The S5-S6 helices are colored in green and the small white circles indicate key residues in the selectivity filter with + and - signs indicating their charge states. The yellow circle with an h indicates the inactivation gate. B Schematic map of a bacterial channel, which contains all the minimal functional elements of a single mammalian NaV domain. In each panel, the N and C letters indicate the N- terminus and the C- terminus, respectively. Adapted and reprinted from [18], https://www.cell.com/trends/biochemical-sciences/fulltext/S0968-0004(15)00121-8 . Copyright (2015), with permission from Elsevier.	109
7.2	Extracellular view of a human NaV structural model. Ribbon style representation of the Nav1.2 structural model. Repeat I is pictured in gray; repeat II, in yellow; repeat III in green; the linker between repeat III and IV in orange; repeat IV, in blue. The four repeats are assembled in a clockwise order. The pore fenestration is shown with the red ellipse.	111
7.3	Superposition of the hNaV1.2 model and the experimental hNaV1.2 structure. The EEDD motif. Top view of the comparison between the hNaV1.2 model (template hNaV1.4, PDB ID: 6AGF, green ribbons) and the experimental hNaV1.2 structure (PDB ID: 6J8E, cyan ribbons). The conformation of the outer vestibular EEDD motif is highlighted. Each residue is labelled and side chains are shown in sticks.	112
7.4	Superposition of the hNaV1.2 model and the experimental hNaV1.2 structure. The DEKA motif. Top view of the comparison between the hNaV1.2 model (template hNaV1.4, PDB ID: 6AGF, green ribbons) and the experimental hNaV1.2 structure (PDB ID: 6J8E, cyan ribbons). The conformation of the inner DEKA motif is highlighted. Each residue is labelled and side chains are shown in sticks.	112
7.5	The representative configuration with two sodium ions in the selectivity filter. Top view of the representative pore structure, used as a starting configuration for thermodynamic calculations. NaV1.2 is described by green ribbons. Residues belonging to the EEDD motif are illustrated with red sticks and those belonging to the DEKA motif with purple sticks. Na ⁺ ions are represented as yellow spheres. The small panel on the left shows the TM view of the configuration.	114

7.6	Transmembrane view of the hNaV1.2 pore. Ribbon style representation of the hNaV1.2 pore model (green). The POPC molecules are represented using lines and white spheres for the phosphorus atoms. Water molecules and ions are represented as cyan and red spheres, respectively.	114
7.7	FESs from 1D-OTFP simulations. FESs (kcal/mol) obtained from three 1D-OTFP simulations (replicas 1Da, 1Db and 1Dc, respectively). Courtesy of Dr. Alexis Paz (Universidad Nacional de Córdoba, Córdoba, Argentina) and Prof. Cameron F. Abrams (Drexel University, Philadelphia, USA), ©.	116
7.8	CV traces of the 1D-OTFP simulations. The CV traces for three different simulations of the previous figure, together with the z coordinate of the nitrogen atom of the K1422 side chain. The fluctuations of the center of mass of the EEDD and DEKA motifs are also included in the plot. The z -coordinate unit is in Å. Courtesy of Dr. Alexis Paz and Prof. Cameron F. Abrams, ©.	117
7.9	FESs from 2D-OTFP simulations. The FESs obtained from three different 2D-OTFP simulations (replicas 2Da, 2Db and 2Dc, respectively). The coordinate z of two Na ⁺ ions (labeled as SOD:1 and SOD:49) are used as CVs. The plots at the left correspond to the FESs as obtained from the simulation. The plots at the right are obtained after processing the collected statistics to obtain a symmetric CV space. Courtesy of Dr. Alexis Paz and Prof. Cameron F. Abrams, ©.	118
7.10	The CV traces of the 2D-OTFP simulations. The CV traces for the three different 2D-OTFP simulations of the previous figure, together with the z coordinate of the nitrogen atom of the K1422 side chain. The fluctuations of the center of mass of the EEDD and DEKA motifs are included in the plot. The z -coordinate unit is in Å. Courtesy of Dr. Alexis Paz and Prof. Cameron F. Abrams, ©.	119
7.11	The cumulative CV traces of the three 2D-OTFP simulations. The left panel illustrates the FESs obtained from averaging together the statistics of the 3 different 2D-OTFP simulations used in Fig 7.10 . The right plot takes into account the CV-space symmetry. The z -coordinate unit is in Å. Courtesy of Dr. Alexis Paz and Prof. Cameron F. Abrams, ©.	120

List of Tables

5.1	HBs in the <i>face-to-face cis-</i> interactions of the single pore. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	87
5.2	HBs in the <i>trans-</i> interactions of the single pore. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	87
5.3	HBs of the <i>cis- face-to-face interactions</i> of the double pore. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	89
5.4	HBs of the <i>trans-</i> interactions of the double pore. Reprinted from [7], https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190 . Copyright (2017), with permission under the terms of the Creative Commons Attribution License.	90
5.5	Kinetic measures of ion permeation. MFPTs and relative rates of single ion permeation through the Cldn15 paracellular channel. Reprinted from [8], https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484 . Copyright (2018), with permission from the American Chemical Society. . . .	93
7.1	List of the published experimental structures of eukaryotic Voltage Gated Sodium Channels. List of experimental structures of eukaryotic Voltage Gated Sodium Channels, obtained with Single Particle Cryo Electron Microscopy. The atomic resolution spans from 2.6 to 4.2 Å .	108

Chapter 1

Introduction

**There are more things in heaven and earth, Horatio,
Than are dreamt of in your philosophy.**
William Shakespeare, *Hamlet*, Scene V.

1.1 Biological Barriers and Cell Junctions

Biological barriers such as the blood-brain, renal or intestinal barrier, are highly selective structures made of adjoined epithelial or endothelial cells, that separate physiological compartments and maintain their stable chemical-physical conditions.

Three types of junctions can be found between the barrier cells, named Anchoring Junctions, Gap Junctions and Tight Junctions.

- **Anchoring Junctions (Desmosomes).** Anchoring Junctions (AJs) are intercellular junctions of epithelia and cardiac muscle. They are specialized for strong adhesion, also acting as signaling centers, regulating fundamental processes such as cell proliferation, differentiation and morphogenesis.
- **Gap Junctions (Communicating Junctions).** Gap Junctions (GJs) are aggregates of intercellular channels that permit direct cell-cell transfer of ions and small molecules, with a Double Transcellular pathway. GJ channels are primarily homo- or hetero-hexamers of Connexin proteins (**Fig 1.1**).
- **Tight Junctions (Occluding Junctions).** Tight Junctions (TJs) regulate the transport of water and solutes in the paracellular space between adjacent epithelial or endothelial cells. Depending on the tissue of expression, they act as barriers or selective paracellular channels (**Fig 1.2**).

1.2 Transcellular and Paracellular Transport

Biological barriers regulate the transport of solutes, ions and water using both transcellular and paracellular routes. Transcellular diffusion involves primary (ATP-based) and secondary (electrochemical) active transport across the cells membrane, together with diffusion through ion channels. On the contrary, paracellular transport occurs only passively along a concentration gradient, and is regulated by the proteins of the TJs.

As a consequence, the mechanism of paracellular permeation is markedly different from the transcellular one: in the former, the diffusing solutes move parallel to the plane of

Double Transcellular

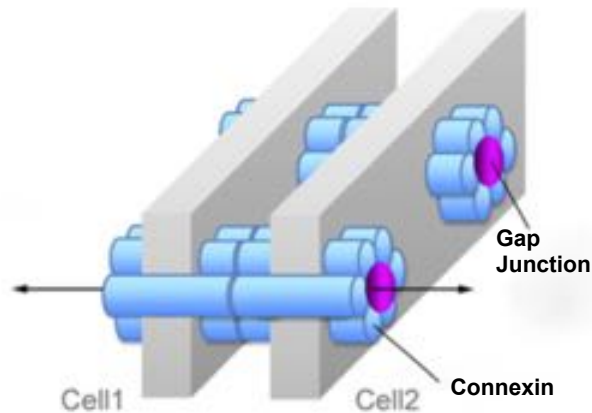


Fig 1.1. Representation of an intercellular Double Transcellular pathway. A schematic representation of a Gap Junction which is an example of Double Transcellular protein complex. The flux of solutes passes the surface of the two cells. Pores are colored in magenta and the Connexin protomers are represented by cyan cylinders. Gray lines and plates indicate cell membrane surfaces, and left-right double arrows indicate the pathways and directions of permeating molecules. Modified picture from [1], <https://www.sciencedirect.com/science/article/pii/S0022283614005713>. Copyright (2015), with permission under the terms of the Creative Commons Attribution License.

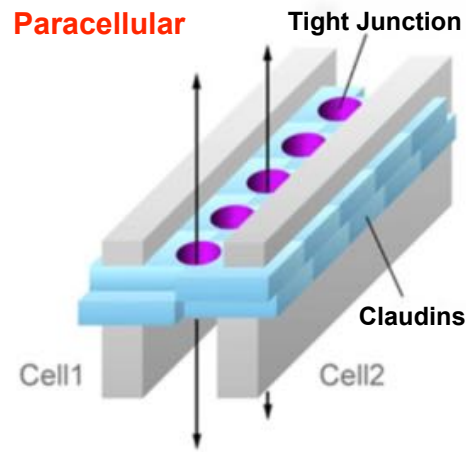


Fig 1.2. Representation of an intercellular Paracellular pathway. Paracellular pores formed by Claudin protomers mediate transport which is parallel with the lipid bilayers. The pore walls are formed by the protein extracellular domains. Pores are colored in magenta and the Claudin protomers are represented by cyan cylinders. Gray lines and plates indicate cell membrane surfaces, and up-down double arrows indicate the pathways and directions of permeating molecules. Modified picture from [1], <https://www.sciencedirect.com/science/article/pii/S0022283614005713>. Copyright (2015), with permission under the terms of the Creative Commons Attribution License.

the lipid bilayers, following a highly hydrated pathway (**Fig 1.2**).
Currently, while the mechanism of transcellular permeation is fairly understood, the

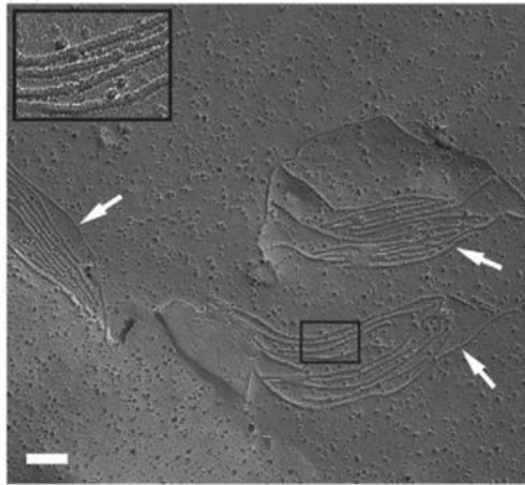


Fig 1.3. Freeze-fracture electron micrograph of TJ strands. Electron microscopy of freeze-fracture replicas of TJs. The arrows point to TJ-like strands. The inset shows a magnified view of the boxed area. The scale bar represents 100 nm. Reprinted from [1], <https://www.sciencedirect.com/science/article/pii/S0022283614005713>. Copyright (2015), with permission under the terms of the Creative Commons Attribution License.

investigation of paracellular pathways is severely hampered by the lack of detailed structural information about TJs. This leaves open many challenges such as the design of TJ penetrating drugs and nanomaterials [19].

1.3 The Multiple Functions of Tight Junctions

TJs are complexes formed by several integral membrane proteins [9, 20–24]. Freeze fracture electron microscopy experiments have revealed that TJs are arranged in strands on the opposing surfaces of two adjoined cells (see **Fig 1.3**).

Proteins in the strands are responsible for regulating the paracellular transport of solutes *via* a size-/charge- selectivity mechanism [25, 26], and their malfunction is associated to many diseases [27–29].

The functions of TJ proteins can be summarized as follows:

- they contribute to holding cells together;
- they maintain the cell polarity by preventing the lateral diffusion of transmembrane proteins;
- they regulate the passage of small molecules and ions through the paracellular environment.

1.4 The Family of Claudin Proteins

Mainly, one family of membrane proteins constitutes the TJ strands, named Claudins (Cldns) (see **Fig 1.4**).

Cldns are small transmembrane, tissue specific, proteins [30, 31]. They span the cellular membrane 4 times (Transmembrane Domains TM1-4), with the N-terminal end and the

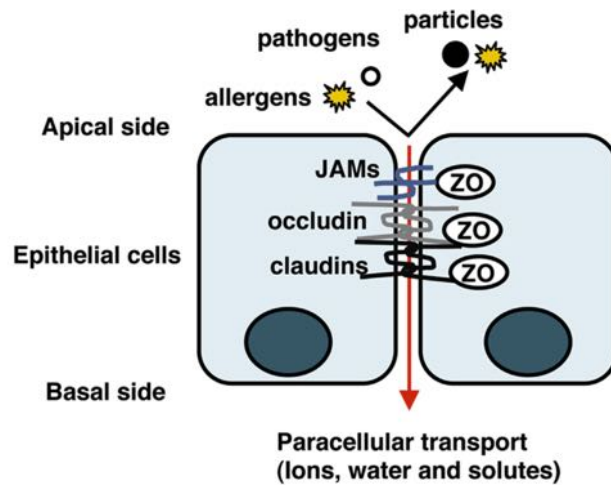


Fig 1.4. Structure of TJs in epithelial cells. TJs contain three types of transmembrane proteins: Claudins, Occludins, and Junction Adhesion Molecules (JAMs) as well as scaffold proteins such as Zonula Occludens-1 (ZO). Reprinted from [9], [https://www.allergologyinternational.com/article/S1323-8930\(17\)30159-4/fulltext](https://www.allergologyinternational.com/article/S1323-8930(17)30159-4/fulltext). Copyright (2019), with permission under the terms of the Creative Commons Attribution License.

C-terminal end both located in the cytoplasm, and two extracellular loops (ECL1-2) which show the highest degree of conservation along the sequence (see **Fig 1.5** for a representation of each domain).

The N-terminal end is usually very short (4-10 amino-acids), while the C-terminal end varies in length from 21 to 63 amino-acids and is necessary for the localization of these proteins.

Further typical features of the Cldn family include a so-called COOH-terminal PDZ-binding motif, through which the majority of human Cldns are able to interact with PDZ domains of other scaffolding proteins.

The ECL1 loop consists of ~ 53 amino-acids and contains a highly conserved signature motif, with a central triplet, G-L-W. Another notable feature of the Cldn ECL1 signature is a pair of conserved Cysteines located shortly after the G-L-W motif. In particular ECL1 participates in regulation of paracellular tightening and permeation.

The shorter ECL2 loop is formed by ~ 24 amino-acids and it is involved in intermolecular interactions, including binding to *Clostridium perfringens* enterotoxin (CPE).

To form TJs, Cldns assemble along the cell membrane *via cis*-interactions and across adjacent cells *via trans*- (intercellular) interactions. In particular, Cldns alone are able to reconstruct TJ-like structures in plasma membrane, recapitulating the TJ structural/functional features with tissue-specific expression [32], and appearing as strands under freeze-fracture electron microscopy [33].

Cldns sequence analysis has led to differentiation into two groups, designated as classic Cldns (1-10, 14, 15, 17, 19) and non-classic Cldns (11-13, 16, 18, 20-24), according to their degree of sequence similarity, in particular for the ECL1/ECL2 regions. Classic Cldns exhibit a much stronger consensus sequence than non-classic Cldns.

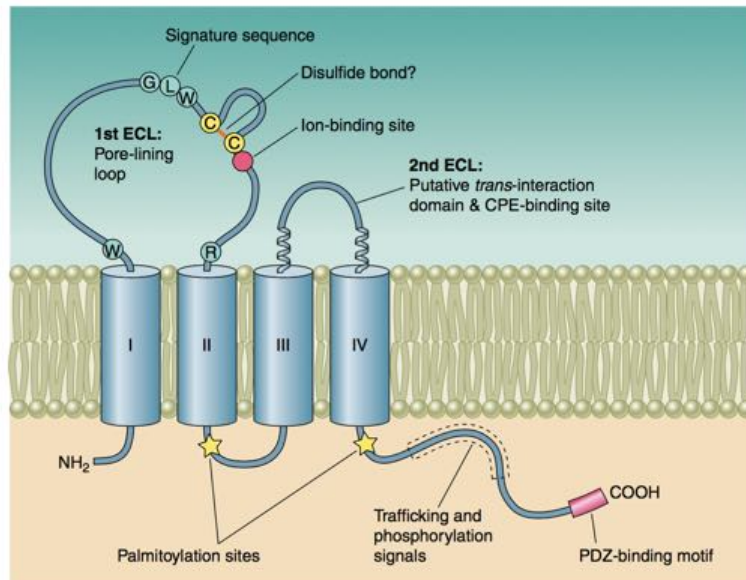


Fig 1.5. Structure of a Cldn protomer. Model of Cldn protein showing predicted topology and secondary structure as well as the different functional domains. Roman numerals indicate the predicted α -helical TM domains. Reprinted from [10], <https://www.physiology.org/doi/full/10.1152/physrev.00019.2012>. Copyright (2013), with permission from the American Physiological Society.

1.4.1 Channel and Barrier Forming Claudins

The Cldn family in mammals consists of 27 proteins, responsible for the paracellular barrier function, and in some cases for an additional paracellular channel function for small solutes [10].

The proteins of this family are typically divided in two main classes, based on their mechanisms:

1. the *barrier* mechanism, where TJ permeability is strongly limited by the absence of paracellular channels,

and

2. the *channel* mechanism, in which solutes pass through Cldn paracellular channels, where permeability is increased for molecules of a certain size/charge but the epithelial barrier function against macromolecules remains intact.

Paracellular channel or tight barrier formation is supported by the spatial encounter of a surplus of repulsing or attracting amino-acid types at the ECL1 domain [34]. In particular, a paracellular channel is likely opened by repulsion of equally charged residues, while an encounter of unequally charged residues or uncharged could lead to the tight interaction of barriers.

However, this classification could be reductive, and two considerations must be introduced:

- firstly, all Cldns that insert into the cell membrane and interact with Cldns from neighboring cells contribute to form a barrier, if compared with the situation of cells without any TJs. For this reason, all Cldns are, in some way, barrier-forming proteins;

- secondly, even though few Cldns unequivocally qualify as paracellular channels forming, thanks to the action of putative charged filters [10], an extensive literature suggests that paracellular space is never completely occluded, and small *tightnesses* could be formed also for barrier Cldn expressions [35–38].

At the moment, the only Cldns qualified as paracellular channels are Cldn2, 10b, and 15, as cation channels, and Cldn10a and 17 as anion channels.

Others have been reported to form channels only when specifically interacting with another Cldn type. Thus a combination of Cldn16/19 has been reported to act as cation channels and Cldn4/8 as anion channels.

Functional studies of paracellular permeability have shown that there are at least two types of channels: a high-capacity pathway with an estimated radius of about 4 Å and a low-capacity pathway with a radius of at least 7 Å [39].

Furthermore, there are reported KnockOut (KO) mouse studies for almost half of the 27 known Cldns [11]. For example, KO and KnockDown (KD) mouse studies of paracellular barrier-forming Cldns have revealed roles of Cldns in preventing water imbalance, inflammation, cancer, and brain disease (see **Fig 1.6**).

Caveat. In this work the term *pore* is used to indicate both paracellular channel and tightness-forming complexes. Often, in a part of the literature, this term is used as a synonym of paracellular channel only.

1.5 Structural Features of Claudins

In this section, I discuss recent advances in our understanding of Cldn structure and function. Cldns were first named in 1998 by Japanese researchers Mikio Furuse and Shoichiro Tsukita at Kyoto University. The name Claudin comes from Latin word *claudere* (*to close*), suggesting the barrier role of these proteins.

All Cldns have a similar membrane topology of four transmembrane (TM) domains, with cytosolic amino and carboxy termini. The first extracellular loop (ECL1) is longer and more hydrophobic than the second extracellular loop (ECL2).

Multiple sequence alignments carried out for some of the Cldns revealed that the amino acid sequence is fairly conserved in the first and fourth transmembrane (TM) domains, but diversified in the second and third TM domains [30].

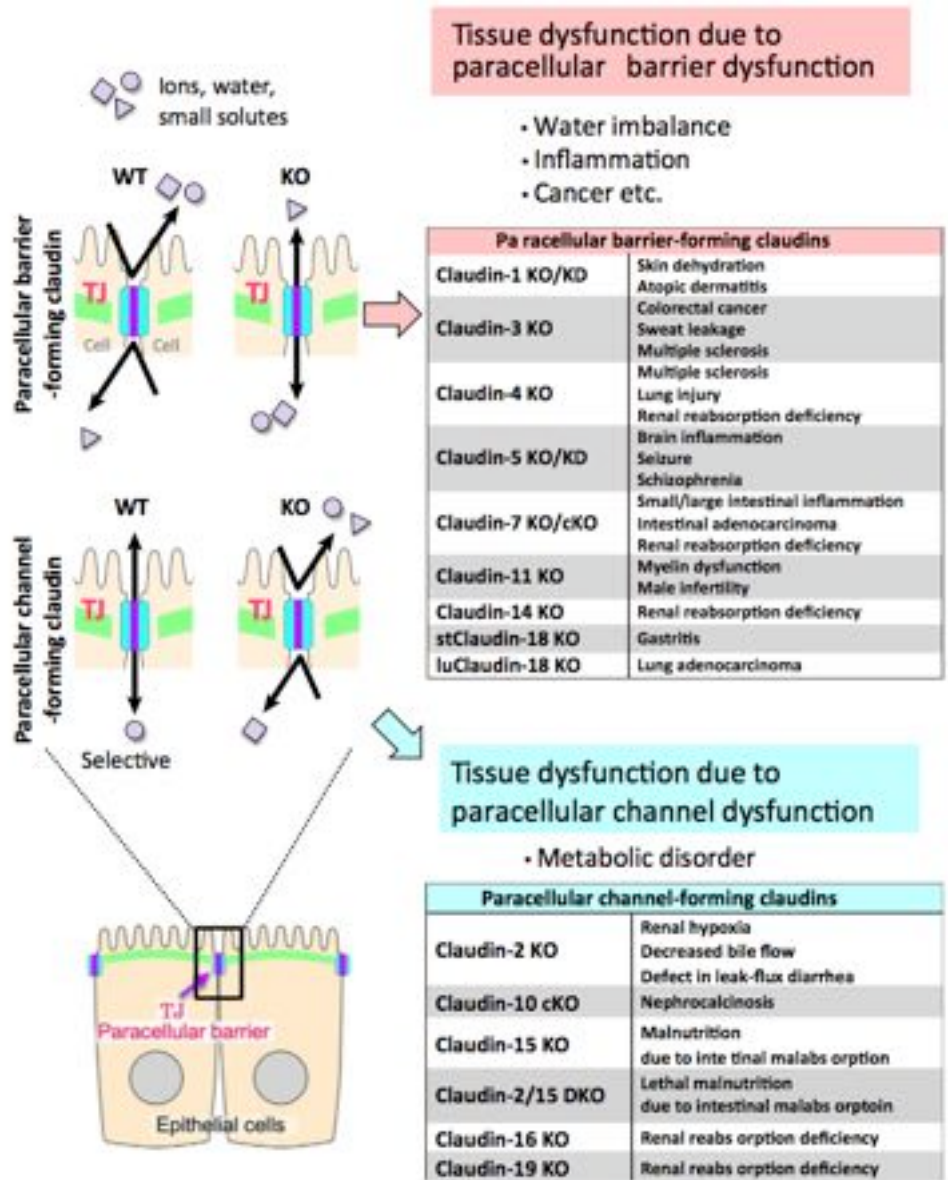
The two loops have a global sequence similarity in almost all the Cldn proteins except for two regions, named V1 in ECL1 and V2 in ECL2 [1].

1.5.1 Single Claudin Crystal Structures

Since many materials can form crystals (such as salts, metals, and semiconductors) X-ray crystallography has been fundamental in the development of many scientific fields. This method also revealed the structure and function of many biological molecules.

Only very recently, crystallographic structures of Cldn protomers have been obtained. In fact, intrinsic membrane proteins remain challenging to crystallize because they require detergents or other means to solubilize them in isolation, and such detergents often interfere with crystallization. The crystal structure of mouse Cldn15 (mCldn15) was resolved in [40] (PDB ID: 4P79). Four other crystal structures were obtained afterwards (mCldn19 [41], PDB ID: 3X29; hCldn4 [42], PDB ID: 5B2G; mCldn3 [43], PDB ID: 6AKE; hCldn19 [44], PDB IDs: 6OV2 and 6OV3) in complex with *Clostridium perfringens* enterotoxin, which illustrate the molecular basis of TJ disruption.

Highly expressed in the intestine, Cldn15 belongs, together with Cldn2, to a family of



Trends in Biochemical Sciences

Fig 1.6. Structure of TJs in epithelial cells. Claudin-based biological systems revealed by Cldn KnockOut (KO) and KnockDown (KD) mice systems. KO and KD mice of specific paracellular channel-forming Cldns mostly show metabolic disorders, including malnutrition due to intestinal malabsorption and decreased bile flow. Abbreviations: WT, Wild Type. Reprinted from [11], <https://www.sciencedirect.com/science/article/abs/pii/S0968000418301932>. Copyright (2019), with permission from Elsevier.

Cldns able to increase the permeability to cations [45,46]. Specifically, Cldn15 protomers form selective pores that are channels for Na^+ and highly resistant barriers to Cl^- ions. For the mCldn15 crystallization, 33 residues at the C terminus were truncated, and the membrane-proximal Cysteines were substituted with Alanines to avoid heterogeneous palmitoyl modification. The modified mCldn15 construct was expressed, purified with

maltose neopentyl glycol detergent, and crystallized in lipidic cubic phase (LCP) [40]. High-resolution diffraction data were obtained from the crystals, and the structure was determined at 2.4 Å resolution using Selenomethionine labeled derivatives by the multiple anomalous dispersion (MAD) method [40]. Crystals of mCldn15 belong to the C2 space group and contain one Cldn monomer per asymmetric unit (**Fig 1.7**).

The TM segments of mCldn15 (TM1 to TM4) form a typical left-handed four-helix bundle, and large portions of the two extracellular segments form a prominent β -sheet structure. The length of the TM helices is consistent with the thickness of a lipid bilayer. The TM region contains many residues with small side chains, ensuring tight packing of the helices [40].

The two extracellular loops (ELC1 and ECL2) comprises five β strands (named β 1 to β 5), four (β 1 to β 4) in the longer ECL1 and one (β 5) at the end of the shorter ECL2, and form a characteristic β -sheet fold. At the C terminus of ECL1, the β 4 strand connects to a short extracellular helix (ECH) that links to the TM2 domain. Remarkably, the loop region between β 1 and β 2 (residues 34 to 41) is not resolved.

In the crystal lattice, Cldn15 protomers assemble in a linear polymer via specific interactions mediated by extracellular residues (see **Fig 1.8**).

In particular, residue M68 of the ECH region of one protomer interacts with residues F146, F147, L158 in TM3 and ECL2 of an adjacent protomer. Some of the corresponding residues in Cldn5 (F147, Y148, and E159) are also reported to contribute to the intermolecular interactions and the formation of TJs [47]. Because mutagenesis experiments indicated that these residues are required for mCldn15 to form TJ-like strands [40], the linear arrangement in the crystal could represent a *cis*-arrangement of Cldns in TJ strands.

Mutagenesis and physiological studies elucidated the residues responsible for the paracellular ion selectivity of the TJs, which are charged and mostly located in the C-terminal half of ECL1. Especially in the structure of mCldn15, the negatively charged residues (D55 and D64) are located on the β 3 to β 4 domains.

The mCldn15 structure and homology models of other Cldns with different charge selectivity exhibit a distinct charge distribution at the distal part of the β -sheet domain, forming a palm-shaped structure, which could be the pore-forming surface of the putative paracellular pathways.

The monomeric structure of Cldn15 has been studied with coarse-grained (CG) simulations [48], and it has been used as a template to build structural models of other Cldns [49, 50].

1.5.2 The Suzuki Model for Claudin-15 Paracellular Channels

From the above description, it is clear that the linear arrangement of Cldn15 protomers in the crystal packing is not sufficient to explain how Cldns arrange themselves in TJ pores.

Soon after the publication of the Cldn15 crystal, an incomplete model of paracellular channels (from now on referred to as *the Suzuki model*) was proposed based on the structure of the Cldn15 monomer [1]. In this subsection, the molecular details of this model will be introduced.

The strand formation based on *cis*-interactions. To form TJs, firstly Cldns assemble along the membrane of a single cell to obtain strands *via cis*-interactions. The configuration of the model combines two distinct *cis*-interfaces, renamed *linear* and *face-to-face*, respectively in [51]. The first is the same observed in the crystal lattice (**Fig 1.8**), while the second is formed through interactions between the edges of β 4-strands of ECL1 of side-by-side protomers (**Fig 1.9**).

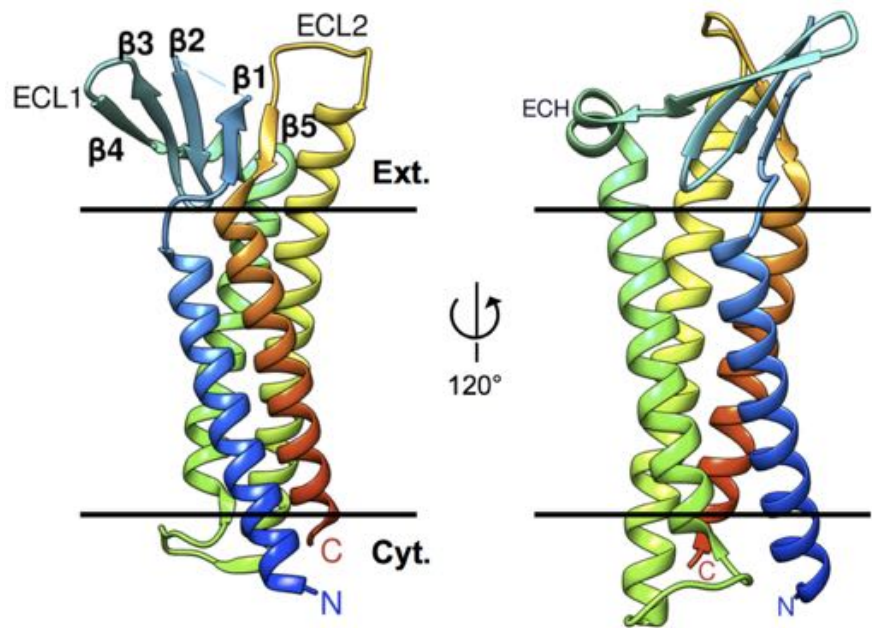


Fig 1.7. Crystal structure of Cldn15. Ribbon representation of the Cldn15 structure (PDB ID 4P79). The bars indicate the membrane boundaries with the extracellular (Ext) and cytosolic space (Cyt). Missing residues are highlighted with dashed lines. Adapted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

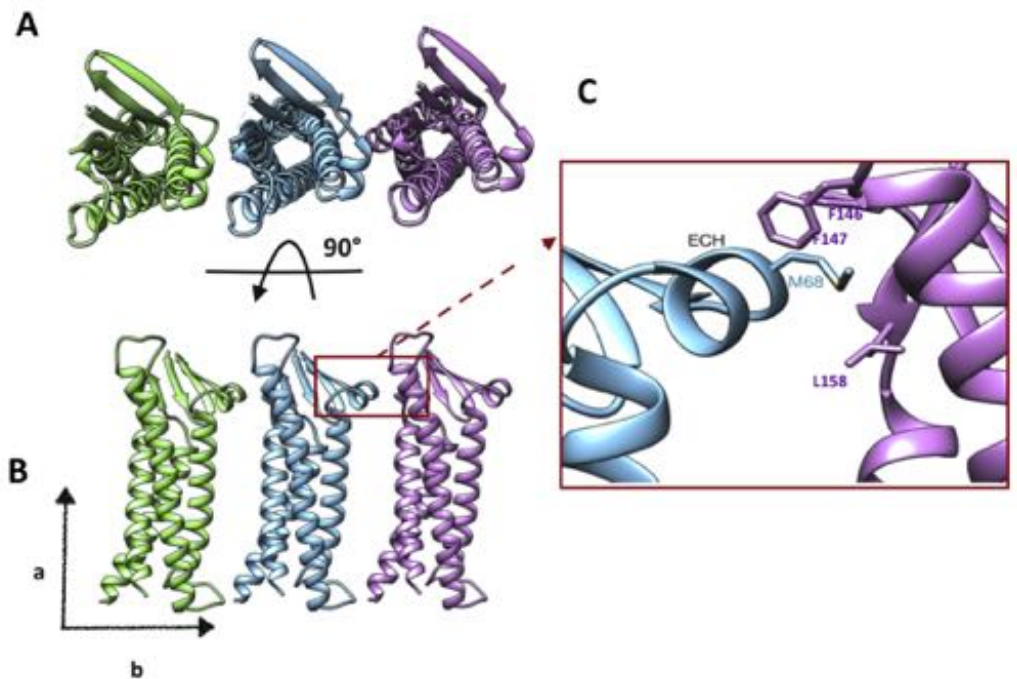


Fig 1.8. The crystallographic Cldn15 linear arrangement. Representation of the crystallographic linear arrangement of Cldn15 protomers (ribbon representation) aligned along the b axis, viewed from the extracellular space (panel A) and from the membrane (panel B). The red square in panel B highlights the region of the lateral interaction between protomers. Panel C shows a zoom of the region where the critical residues are introduced. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

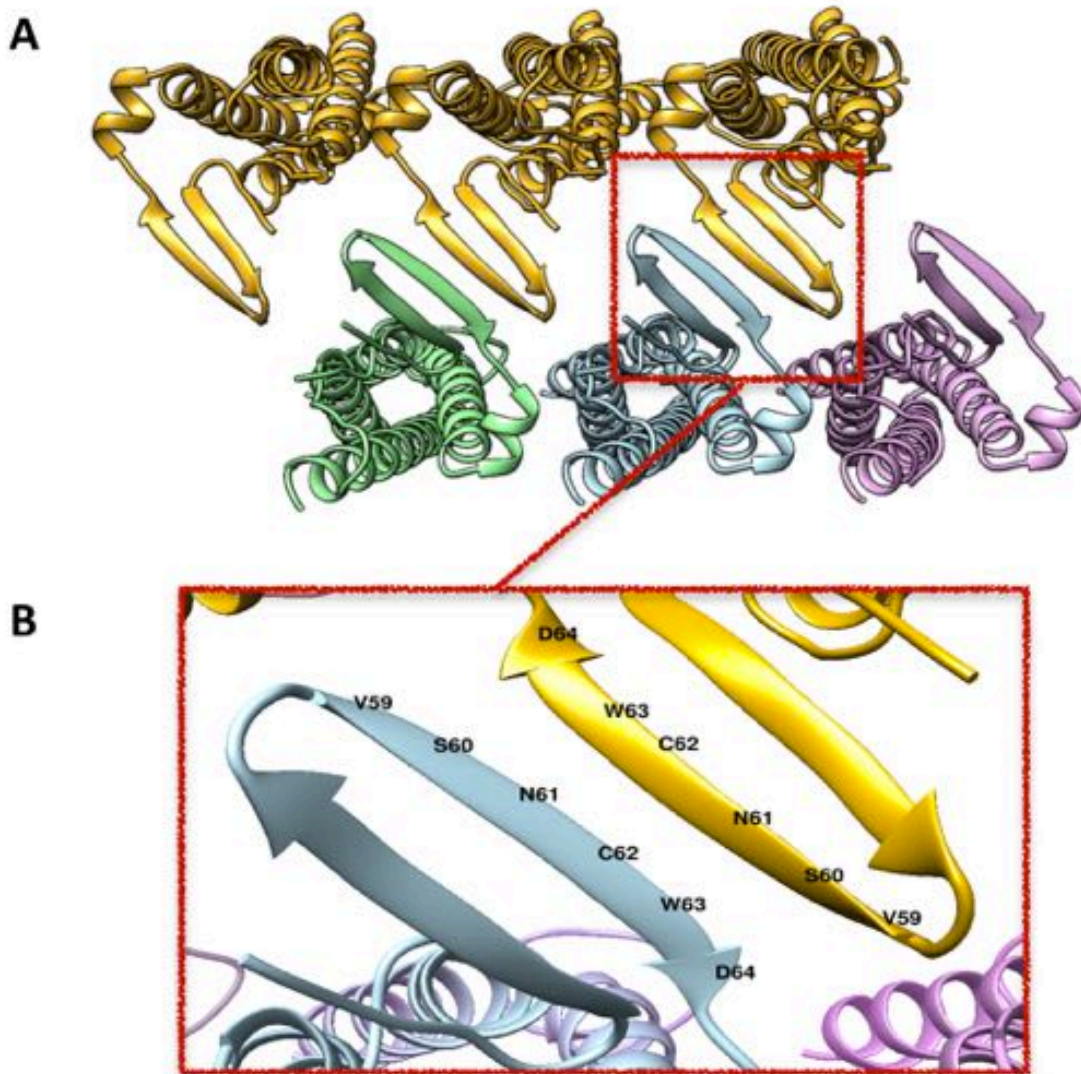


Fig 1.9. Double *cis*- arrangement of Cldn15 protomers. A, Three Cldn15 protomers (green, cyan and purple) are aggregated with another linear group of Cldns (gold) via a second antiparallel *cis*- interface (*face-to-face*) formed by the close vicinity of β 4 strands and highlighted in B. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

This arrangement is supported by Cysteine crosslinking data obtained from mutants of Cldn15. In particular a single Cysteine substitution at N61 resulted in a major shift of the expressed protein to a position consistent with a dimer, while other Cysteine substitutions also resulted in weak dimer bands.

The interaction between two opposite cells. On the basis of the sequence alignment among members of the Cldn family, the crystal structures revealed that poorly conserved sequences are located at the most exposed regions in the extracellular segments, defined as variable regions V1 (Cldn15 residues from 31 to 46) and V2 (Cldn15 residues from 150 to 154), respectively (**Fig 1.10**).

In the crystal of mCldn15, no loop structure in V1 was observed owing to disorder likely caused by the conformational flexibility (see **Fig 1.7**). These variable regions

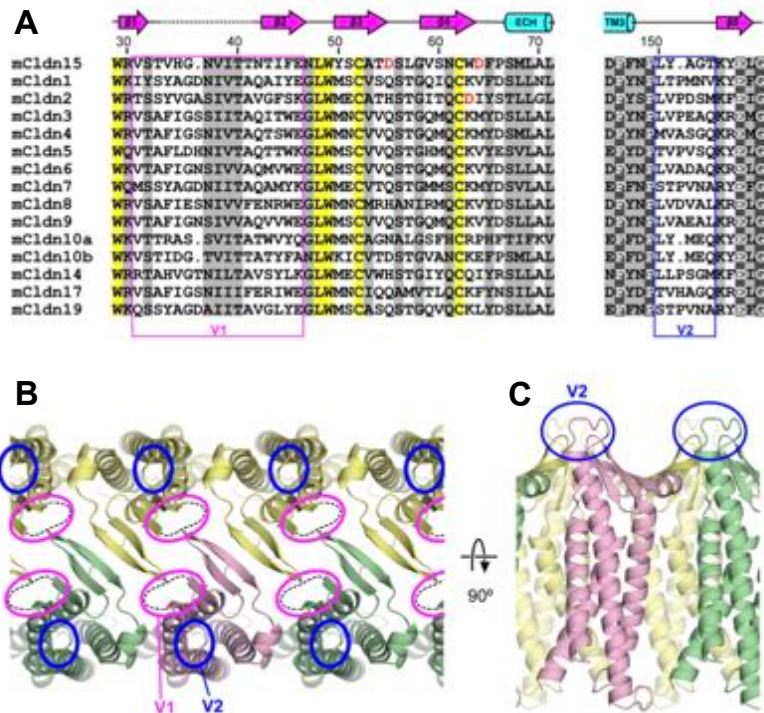


Fig 1.10. V1 and V2 regions in the double *cis*-arrangement of Cldn15 protomers in the Suzuki model. **A**, Multiple sequence alignment of the extracellular regions in mCldn15 and other classic Cldns in mouse (mCldn1-9, 10a, 10b,14,17,19). Numbering above the sequences is based on the full-length mCldn15 protein. Secondary structure elements seen in the mCldn15 crystal structure are indicated above the sequences as arrows (β -strands; magenta), cylinders (α -helices; cyan), and lines (loops; black). The region that is disordered in the crystal structure is shown as a broken line. The consensus sequence (W-LW-C-C) is highlighted in yellow, and highly conserved residues are highlighted in gray. Red letters in mCldn15 and mCldn2 indicate important residues for their cation selectivity. Magenta and blue boxes indicate regions defined as variable sequence regions, V1 and V2, respectively. **B** and **C**, Ribbon diagram of the modeled antiparallel double-row arrangement of mCldn15 molecules in TJ strands, viewed perpendicular to (panel **B**) and parallel with (panel **C**) the membrane plane. Magenta and blue ovals indicate the positions of the V1 and V2 regions, respectively. Reprinted from [1], <https://www.sciencedirect.com/science/article/pii/S0022283614005713>. Copyright (2015), with permission under the terms of the Creative Commons Attribution License.

provide flexibility that might contribute to subtype-specific interactions induced by their conformational changes.

The introduction of the strand model described in **Fig 1.9** and in the panel **B** of **Fig 1.10** suggests the formation of putative *trans*-interactions between two *cis*-strands mediated by the interaction of the V1 and V2 regions.

Considering that the V1 and V2 regions are located at the two edges of the modeled TJ strand (see **Fig 1.10**, panel **C**), these regions of Cldns should form a *half-pipe* (see **Fig 1.11**). These edges may change their conformation to engage, in adjacent cells, *trans*-interactions through an induced fit, thus functioning as pillars that support the space needed for the formation of paracellular pathways [1].

Considering the association of antiparallel Cldn double rows in adjacent membranes, a family of β -barrel-like channels can be formed, *i.e.* groups of β -sheet that twist and coil to form closed structures in which the first strand is bonded to the last strand *via* hydrogen bonds (HBs). This arrangement is represented in **Fig 1.12**. The diameter of the resulting incomplete paracellular channels is smaller than 10 Å.

In the case of Cldn15, this suggested arrangement of protomers is consistent with freeze

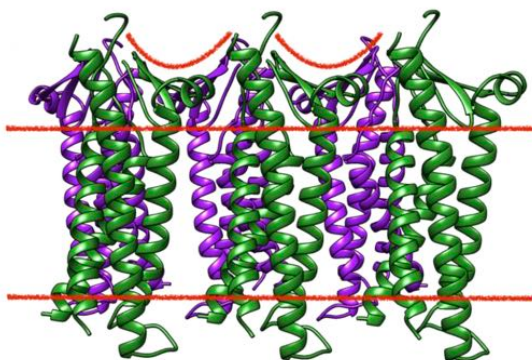


Fig 1.11. The Cldn15 *cis*- arrangement. Side view of the double-row complex. The bars depict the TM boundaries and the lines indicate the half-pipe. Reprinted from [8], <https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484>. Copyright (2018), with permission from the American Chemical Society.

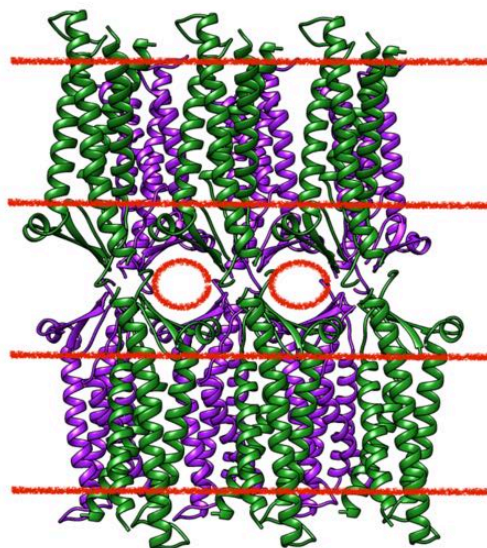


Fig 1.12. The Suzuki model of TJ paracellular channels. Red bars indicate the membrane boundaries and the oval curves indicate pores through which paracellular diffusion occurs. Reprinted from [8], <https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484>. Copyright (2018), with permission from the American Chemical Society.

fracture electron microscopy (EM) images of polymeric TJ strands, and displays pore-forming (D55) and pore-lining (D64) negatively charged residues whose mutation was shown to reverse Cldn selectivity for cations [52]. The pore β -barrel-like structure would be an effective scaffold to line the pathway where the D55 side chains (one for each protomer) point into the interior of the pore.

1.5.3 Further Validation of the Double-row Arrangement Introduced by the Suzuki Model

It is important to note that after the publication of the structure of the Suzuki Cldn arrangement (2015) and of our refined versions of this model (2017-2018), an important work concerning the TJ morphology was published in 2019 [12].

While in the original paper by Suzuki et al. the double row arrangement of the TJ strands was inferred by Freeze-fracture Electron Microscopy, the authors of this recent

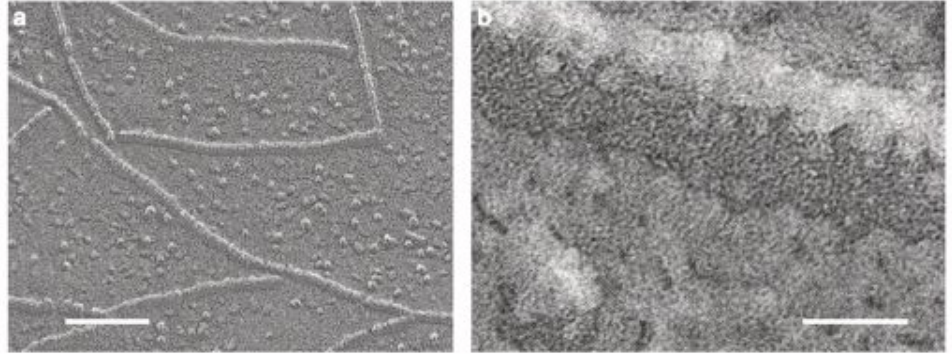


Fig 1.13. The double-row structure of TJ strand fibrils. A, Cldn11 fibrils visualized by unidirectional platinum shadowing Freeze-fracture. **B**, Magnification of a single Cldn11 fibril. Reprinted from [12], <https://www.nature.com/articles/s42003-019-0319-4>. Copyright (2019), with permission under the terms of the Creative Commons Attribution License.

paper describe the direct experimental observation of a double stranded morphology of TJ intramembrane fibrils of a single cell. Combining amorphous carbon replicas [53] with Phase-contrast Electron Microscopy, they provided the surface topography of the TJ strands made by Claudin-11 (Cldn11) in the native context (**Fig 1.13**).

Remarkably, the average distance between the intensity minima of the two strands across the fibril fits the dimension of the Suzuki double-row model, which predicted a diameter of ~ 6 nm [12]. Moreover, this result supports the hypothesis that the double-row Cldn arrangement, introduced originally for the intestinal Cldn15, could be an architecture shared also by strands of other Claudins, expressed in different mammalian tissues.

Chapter 2

Aim of the Thesis

Ignoranti quem portum petat nullus suus ventus est.

Lucius Annaeus Seneca, *Epistulae Morales ad Lucilium*.

In the recent years, there has been an increasing interest in the study of TJ proteins and paracellular selectivity, also thanks to the growing number of X-Ray structures of Claudins [40–44]. However, the fine details of molecular transport through paracellular channels are still uncharted. In this Thesis, using computational tools, we investigated the structural and functional properties of the Cldn15-based model of a paracellular channel proposed in [1] and named the *Suzuki model*, to understand its reliability for the paracellular intestinal tissue and its possible transferability to other Cldn-based systems.

Indeed, the general validity of this structural model, as presented in its original form, has been debated [51]. First of all, barrier-forming Cldns such as Cldn3 and Cldn5 might arrange differently from channel-forming ones as Cldn15. Additionally, the Cldn15 protomers, used to assemble the model, show an incomplete topology, with nine residues missing in ECL1 (eight of which are absent also in the crystal structure), and two in ECL2. About this last issue, it has been suggested [51] that the assembly implies very tight head-to-head packing in the regions of the missing residues, which might cause excessive steric disturbance once these are added.

Furthermore, other mutagenesis studies suggest that TJ strands contain intramembrane *cis*-interactions between Cldn TM helices [54, 55] which are not prominent in the antiparallel double-row model, and that could lead to the formation of a second pore configuration, named the *Nangia model* and described in [49].

On the other hand, Coarse-Grained (CG) models [49, 50] demonstrated that, for Cldn5, a *cis*-arrangement similar to that of Suzuki et al. is among those that are spontaneously formed by monomers. Furthermore, the same recent computational studies suggest that the two competitive pore configurations are not mutually exclusive, since, in a strand of Cldns, both can be formed.

Hence, the structural model proposed by Suzuki et al. could represent one of the pore fenestration of the narrow space in TJ tissues, providing a valuable starting hypothesis for the molecular architecture of Cldn-based TJ pores, which demands further testing in order to assess its validity, even though it cannot fully explain all the available experimental results.

In this Thesis, the use of a variety of computational techniques allows the introduction of the first refined versions of the model [7, 8].

Missing segments were successfully added to the original version of the structure, using Structural Modeling techniques. Then, standard all-atom Molecular Dynamics (MD) simulations were performed, with the aim of assessing the structural stability and conformational properties of this protein complex.

In particular, a refined model of a single paracellular pore made of four interacting monomers embedded in a double lipid bilayer was produced. The tetrameric assembly is based on the *trans*-interactions of two *face-to-face cis*- dimers, and it is the minimal structure consistent with the model of Suzuki et al. that is sufficient to reproduce an individual paracellular pore. It displays inter-monomer residue-residue interactions (both *cis*- and *trans*-) that are described

in the literature as relevant for maintaining TJ integrity and function. The structure preserves the main features of the original model, showing stable *cis*- and *trans*- interactions between monomers, and a stationary minimum pore diameter of 5-6 Å.

An additional refined model of a double-pore, made of eight Cldn15 monomers, allows to observe inter-monomer surfaces not present in the single pore and that are relevant for TJ strand formation.

All these results demonstrate that the missing protein segments in the model of Suzuki et al. can arrange without clashes in multiple different conformations, and confirm that the structure is a possible arrangement of Cldn paracellular pores.

Because a reliable model must describe the main physiological properties of the system it represents, the single-pore model introduced in this work was further validated by verifying that it correctly reproduces the experimentally known selectivity of Cldn15 pores as Na⁺ paracellular channels and Cl⁻ barriers. Therefore, the single paracellular pore model was validated by performing Free Energy (FE, or Potential of Mean Force, PMF) [56] and kinetics calculations for the permeation of ions. Umbrella Sampling (US) calculations [57–59] for Na⁺, K⁺ and Cl⁻ were performed to calculate 1 dimensional (1D)-PMF profiles, using the ion displacement along the axis of the channel as Collective Variable (CV, or Reaction Coordinate, RC). These simulations illustrate a strong selectivity of the channel which allows the passage of cations and not of the chloride ion. Furthermore, the time-scales of the ionic permeation was investigated using Voronoi Tessellated Markovian Milestoning (VTMM) MD simulations [60], to obtain the kinetic properties of the full reaction. Here, the Mean First Passage Times (MFPTs) for Na⁺, K⁺ and Cl⁻ permeation were calculated, with the previous CV, obtaining that the passage of the anion is slowed down by a factor ten with respect to cations.

To further demonstrate that the choice of the single CV is sufficient to reproduce the main features of the model, we performed Single Sweep (SSw) [61–63] simulations to obtain the 1D-PMF calculated along the Minimum Free Energy Path (MFEP) of 3 dimensional (3D)-PMF maps for the diffusion of different ions (Na⁺, K⁺, Li⁺, Ru⁺, Cs⁺) and ADG, using the three Cartesian coordinates of the Center of Mass of the molecule as CVs. Interestingly, the path followed by the ions describe quasi-linear trajectories that can be obtained by using as CV only the ion position along the main axis of the channel.

Overall, these simulations confirm the experimentally known cation-selective nature of the Cldn15 TJ pore, thus validating our refined model based on the Suzuki architecture. They also provide valuable atom-level detail on the ion permeation process.

Finally, using Homology Modeling, a family of Cldn5 complexes belonging to the Blood Brain Barrier (BBB), based on the topology of the Suzuki model, was produced to evaluate if this model can be transferred to other Cldn-based architectures. Interestingly, the variability of the amino-acid composition in the internal cavity of the pore suggests a different structural behavior for the Cldn5 complex, which is consistent with the physiological properties of the brain tissue. In particular the Cldn15 D55, which is pivotal for the charge selectivity of intestinal Cldns, is replaced by a neutral Q, potentially abolishing the charge selectivity provided by a paracellular channel and occluding the intercellular space as expected for the Cldn-based architectures of the BBB.

Chapter 3

Methods

**Per correr miglior acque alza le vele
omai la navicella del mio ingegno,
che lascia dietro a sé mar sì crudele;**
Dante Alighieri, *Divina Commedia, Purgatorio*, Incipit.

Macroscopic physical properties can be distinguished by

1. *equilibrium properties*, such as the binding constant of an inhibitor to an enzyme, or the average potential energy of a system and
2. *non-equilibrium properties*, such as the viscosity of a liquid, diffusion processes in membranes, or the dynamics of phase changes.

Ideally, the study of these processes requires a quantum approach, such as the time-dependent Schrödinger equation (3.1). This equation describes the properties of molecular systems:

$$H\psi = i\hbar \frac{\partial \psi}{\partial t} \quad (3.1)$$

where H denotes the Hamiltonian of the system, ψ the wave function, \hbar the reduced Planck constant. The wave function ψ is a function of the positions of both the nuclei and the electrons, *i.e.* $\psi = \psi(\vec{R}, \vec{r}, t)$. Here \vec{R} denotes positions of the K nuclei, $\vec{R} = (\vec{R}_1, \dots, \vec{R}_K)$, \vec{r} the positions of the M electrons, $\vec{r} = (\vec{r}_1, \dots, \vec{r}_M)$, and t the time. In quantum chemistry and physics, the so-called Born-Oppenheimer (BO) approximation is the assumption that the slow motion of atomic nuclei and the fast motion of electrons in a molecule can be treated separately, assuming that the electronic structure adapts instantaneously to given nuclei positions.

The BO approach holds as long as the potential energy surfaces of distinct excited states do not approach each other. For molecules in the ground state this is usually the case.

However the quantum mechanical approach for biosystems still remains prohibitive because of the computational cost. This opens the way to simplify the description of interactions and average over irrelevant details. As illustrated in this chapter, classical Molecular Dynamics (MD) and Statistical Mechanics provide the theoretical framework for such simplifications. In physics, MD is used to examine the dynamics of atomic-level phenomena, while in structural biology, MD is frequently used to refine three-dimensional structures of proteins and other macromolecules.

In this Chapter, I will describe the theory of classical MD simulations for biological complexes.

3.1 Molecular Dynamics Simulations

In the most frequent situations, classical MD simulations solve Newton's equations of motion for a system of N interacting atoms:

$$m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} = \vec{f}_i, \quad i = 1, \dots, N \quad (3.2)$$

where \vec{r}_i describes here the coordinates of the atom i . The forces are the negative derivatives of a potential function $V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$:

$$\vec{f}_i = -\nabla_{\vec{r}_i} V(\vec{r}) \quad (3.3)$$

Here m_i denotes the mass of atom i and $V(\vec{r}) = E_e^0(\vec{r})$ the potential energy surface of the ground state of the system, as a function of the positions of atoms only.

The equations are solved in small time steps ($\Delta t \sim 1 - 4$ fs), using a discretization scheme described below.

As it will be described in **Subsection 3.1.1**, the potential energy is described *numerically* by a set of parameters named *Force Fields*. A force field refers to the functional form used to calculate the potential energy of the system. In general *all-atom force fields* provide parameters for every type of atom in a system, including hydrogen, while *united-atom interatomic potentials* treat the hydrogen and carbon atoms in each methyl group (terminal methyl) and each methylene bridge as one interaction center. The system studied is followed for some time and the coordinates are written to an output file at regular intervals. Since Newton's equations of motion are second-order Ordinary Differential Equations (ODEs), initial condition basically means $\vec{r}^{3N}(t=0)$ and $\dot{\vec{r}}^{3N}(t=0)$, *i.e.* the initial particle positions and velocities. In general if the experiment starts at $t = t_0$, the coordinates at $t = t_0$ must be known. If velocities are not available, initial atomic velocities $\dot{\vec{r}}_i$, $i = 1 \dots 3N$ can be generated at a given absolute temperature T_0 with a Maxwell-Boltzmann velocity distribution generated from random numbers. The coordinates as a function of time represent a trajectory of the system. After initial changes, the system will usually reach an equilibrium state. By averaging over an equilibrium trajectory, many macroscopic properties can be extracted from the output file.

In different codes, MD simulations solve Langevin equations of motion for a system of N interacting atoms, in place of Eq.s (3.2):

$$m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} = \vec{f}_i - b_i \frac{\partial \vec{r}_i}{\partial t} + \vec{R}_i(t) = -\nabla_{\vec{r}_i} V(\vec{r}) - b_i \frac{\partial \vec{r}_i}{\partial t} + \vec{R}_i(t), \quad i = 1, \dots, N \quad (3.4)$$

The force acting on the particle is written as a sum of a friction force for atom i proportional to the particle's velocity (Stokes' law), $-b_i \frac{\partial \vec{r}_i}{\partial t}$, and a *noise* term $\vec{R}_i(t)$. The force $\vec{R}_i(t)$ has a Gaussian probability distribution with zero mean and correlation function $\langle R_{i,a}(t) R_{i,b}(t') \rangle = 2b_i k_B T \delta_{a,b} \delta(t - t')$ for each couple of \vec{R} components, a and b . Here k_B is the Boltzmann constant, T the specified target temperature, and b_i is the friction coefficient for atom i . Using the Langevin strategy, the molecules interact with a stochastic heat bath via random forces and dissipative forces.

Note. In the following subsections, \vec{r}_i describes the coordinates of the i -th atom.

3.1.1 Force Fields

The potential energy, V , is approximated by a sum of empirical expressions. The empirical approximation is referred to as *force fields*, where the molecules are approximated by a *ball-and-spring-model*. A typical expression for the potential energy reads:

$$V_{tot} = V_{bonded} + V_{nonbond\ pairs} \quad (3.5)$$

$$V_{bonded} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{improper\ dihedrals} \quad (3.6)$$

Explaining the term V_{bonded} , Eq. (3.5) can be re-written as:

$$V_{tot} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{improper\ dihedrals} + V_{nonbond\ pairs} \quad (3.7)$$

Typical expressions for the terms in Eq. (3.7) are:

$$V_{bonds} = \sum_{bonds} k_r (r - r_{eq})^2 \quad (3.8)$$

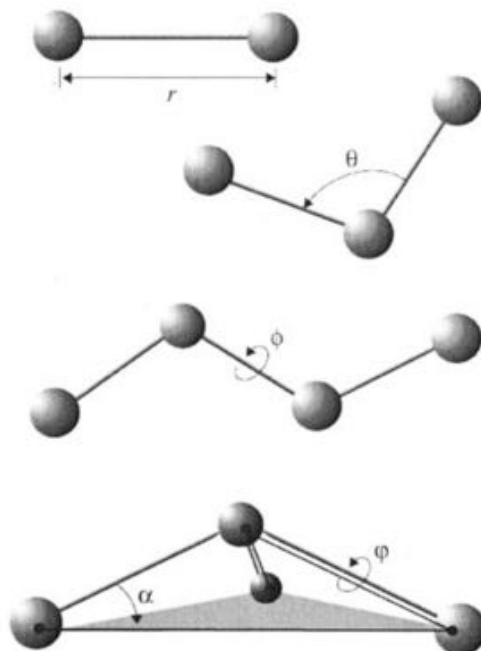


Fig 3.1. Illustration of the main bonded interactions described in Force Fields. Bonded interactions described in a Force Field: r represents a bond distance; θ represents the bond angle; ϕ gives the dihedral angle; the out-of-plane angle α may be controlled by an improper dihedral φ . Note that the improper planarity deviation (α) is not used in the potential energy model, but is shown to clarify that improper interactions are used to establish molecular planarity. Reprinted from [13], <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20289>. Copyright (2005), with permission from John Wiley and Sons, Inc; 2005 Wiley Periodicals, Inc.

$$V_{angles} = \sum_{angles} k_{\theta}(\theta - \theta_{eq})^2 \quad (3.9)$$

$$V_{dihedrals} = \sum_{dihedrals} k_{\phi}[1 + \cos(n\phi) - \gamma] \quad (3.10)$$

$$V_{improper\ dihedrals} = \sum_{improper\ dihedrals} k_i(\varphi - \varphi_{eq})^2 \quad (3.11)$$

$$V_{non\ bonded} = \sum_{i < j} \left\{ \frac{q_i q_j}{4\pi\epsilon_m r_{ij}} + 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \right\} \quad (3.12)$$

where

- V_{bonds} describes the bond stretches, where k_r is the bond force constant and $(r - r_{eq})$ is the difference between the actual value of the bond distance (r) and a reference value (r_{eq}),
- V_{θ} defines the angles term, which accounts for the bond angles, where k_{θ} is the angle force constant and $(\theta - \theta_{eq})$ is the difference between the actual value of the angle (θ), formed by 3 bonded atoms, and a reference value (θ_{eq}),
- $V_{dihedrals}$ represents the dihedrals term, where k_{ϕ} is the dihedral force constant, n is the multiplicity of the function, ϕ is the dihedral angle and γ is the phase shift,

- $V_{improper\ dihedrals}$ characterizes the improper dihedrals terms, out of plane bending, where k_i is the force constant, φ is the out of plane angle and φ_{eq} an associated reference value,
- Nonbonded interactions between pairs of atoms (i, j) are represented by the last term, $V_{non\ bonded}$.

They are calculated between pairs separated by at least three bonds along the molecular chain. They include two terms, which represent the electrostatic interactions *via* the Coulombic energy potential $\frac{q_i q_j}{4\pi\epsilon_m r_{ij}}$ and the van der Waals (VdW) energy calculated with

the standard so-called 12-6 *Lennard-Jones potential* $4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right]$, respectively.

For this expression, q_i and q_j are the charges for the atoms (i, j) at a mutual distance r_{ij} , ϵ_m is the electric permittivity of the environment, ϵ is the depth of the potential well and σ is the finite distance at which the inter-particle potential is zero.

All the bonded interactions and the associated notations are illustrated in **Fig 3.1**.

The bonded potential terms involve 2-, 3-, and 4-body interactions of covalently bonded atoms, with $O(N)$ terms in the summation. The nonbonded potential terms involve interactions between all pairs of atoms (usually excluding pairs of atoms already involved in a bonded term), with $O(N^2)$ terms in the summation, although fast evaluation techniques are used to compute good approximations of their contribution to the potential with a $O(N \log N)$ computational cost, with the Particle Mesh Ewald (PME) method [64] (see below).

The parameters used in this kind of potentials are typically obtained from quantum chemical calculations and experimental data. Very popular force fields are CHARMM (http://mackerell.umaryland.edu/charmm_ff.shtml), AMBER (<http://ambermd.org>), GROMOS (<http://www.gromos.net>) and OPLS (<http://zarbi.chem.yale.edu/oplsam.html>). In addition, the general force fields mentioned above have also developed polarizable versions.

For all the simulations performed in this Thesis, the CHARMM36 (C36) force field was used for the proteins [65] and the lipids [66]. CHARMM36 is the most recent update of the CHARMM force field, where the version number (*i.e.* 36) is based on the version of the programme CHARMM in which the force field version was first released.

3.1.2 Models for Water Molecules

A water model is used to simulate water clusters, liquid water, and aqueous solutions with explicit solvent (see http://www1.lsbu.ac.uk/water/water_models.html for an accurate review of the different water models). In general, they can be classified by following three points:

- the number of interaction points,
- whether the model is rigid or flexible,
- whether the model includes polarization effects.

Popular models for water include SPC (Simple Point Charge) [67], TIP3P (Transferable Intermolecular Potential 3P) [68], and SPC/E (Extended Simple Point Charge) [69].

The TIP3P model implemented in the CHARMM force field was used for all the simulations performed in this work. This model specifies a 3-site rigid water molecule with charges and Lennard-Jones parameters assigned to each of the 3 atoms. Bonded interactions are represented by holonomic constraints (see **Subsection 3.1.7** for a definition of holonomic constraints), the electrostatic interaction is modeled using Coulomb's law, and the dispersion and repulsion forces using the Lennard-Jones potential.

3.1.3 The Cutoff Radius

As introduced in Eq.s (3.5-3.7) the Force Field expression includes non-bonded interactions. These terms are much more computationally costly to calculate in full.

Generally a cutoff radius, \vec{r}_c , is used to speed up the calculation so that atom pairs which distances are greater than \vec{r}_c have a van der Waals interaction energy of zero.

On the other hand, the electrostatic terms, modeled by the Coulomb potential, do not fall off rapidly with distance, and long-range electrostatic interactions are often important features of the system.

To address this problem, a variety of methods are used such as the Particle Mesh Ewald (PME) method [64], that we will introduce in **Subection 3.1.8**.

3.1.4 Boundary Conditions and the Minimum Image Convention

To avoid problems with boundary effects caused by finite size, Periodic Boundary Conditions (PBCs) are often used in MD simulations (**Fig 3.2**). In PBCs, one explicitly keeps track of the motion of N particles in the so-called supercell, but the supercell is surrounded by infinitely replicated, periodic images of itself. Therefore, using PBC scheme, a particle may interact not only with particles in the same supercell, but also with particles in adjacent image supercells (see **Fig 3.2**).

In general, PBCs are used in conjunction with the *minimum image convention*. In this scheme each particle interacts at most with only one image of every other particle in the system, *i.e.* the assumed simulation cell must have a side length of at least as twice as the radius of the cutoff \vec{r}_c , to exclude interactions of a particle with its own images (*i.e.* to exclude self-interactions).

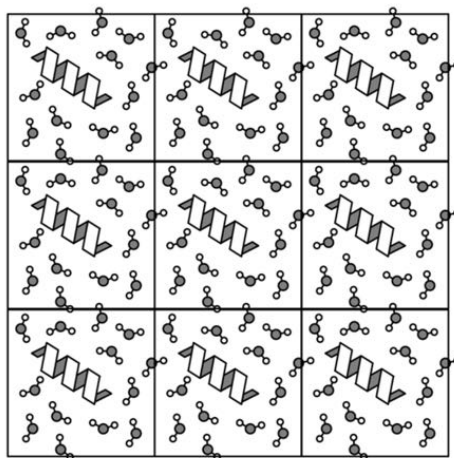


Fig 3.2. Illustration of Periodic Boundary Conditions. The supercell positioned of a single peptide (α helix ribbon) in water solvent (three point scheme) at the center of nine cells. Reprinted from [14], <https://pubs.acs.org/doi/abs/10.1021/acs.jpcc.7b10830>. Copyright (2018), with permission from the American Chemical Society.

3.1.5 Integrators

Eq.s (3.2) or Eq.s (3.4) are a set of second-order ODEs, which can be strongly nonlinear. By converting them to first-order ODEs in the $6N$ -dimensional space of $\{\vec{r}_N, \vec{v}_N\}$, general numerical algorithms for solving ODEs (named *integrators*) can be applied [70].

The Hamiltonian allows for accurate integration algorithms, such as the central difference integrators for Newton Eq.s and associated approaches for Langevin Eq.s.

In general, an integrator advances the trajectory over small time increments Δt : $\vec{r}^{3N}(t_0) \rightarrow \vec{r}^{3N}(t_0 + \Delta t) \rightarrow \vec{r}^{3N}(t_0 + 2\Delta t) \rightarrow \dots \rightarrow \vec{r}^{3N}(t_0 + L\Delta t)$, where L is usually $\sim 10^4 - 10^7$.

Here a brief overview of the so-called *central difference* integrators (Verlet, velocity Verlet) and of the associated integrator of the Langevin Eq. is given.

Integrators for Newton Dynamics. In MD simulations, the most commonly used time integration algorithm is probably the Verlet algorithm.

The basic idea is to write two third-order Taylor expansions for the positions, one forward and

one backward in time, and to sum them. Assuming here $\vec{r}^{3N}(t)$ trajectory is differentiable, perform Taylor expansion

$$\vec{r}_i(t_0 + \Delta t) = \vec{r}_i(t_0) + \dot{\vec{r}}_i(t_0)\Delta t + \frac{1}{2}\ddot{\vec{r}}_i(t_0)(\Delta t)^2 + \frac{1}{6}\vec{j}_i(t_0)(\Delta t)^3 + \mathcal{O}((\Delta t)^4) \quad (3.13)$$

$$\vec{r}_i(t_0 - \Delta t) = \vec{r}_i(t_0) - \dot{\vec{r}}_i(t_0)\Delta t + \frac{1}{2}\ddot{\vec{r}}_i(t_0)(\Delta t)^2 - \frac{1}{6}\vec{j}_i(t_0)(\Delta t)^3 + \mathcal{O}((\Delta t)^4) \quad (3.14)$$

where $\vec{j}_i \equiv \frac{d^3\vec{r}_i}{dt^3}$. Adding these two expansions gives

$$\vec{r}_i(t_0 + \Delta t) + \vec{r}_i(t_0 - \Delta t) = 2\vec{r}_i(t_0) + \ddot{\vec{r}}_i(t_0)(\Delta t)^2 + \mathcal{O}((\Delta t)^4) \quad (3.15)$$

Since $\ddot{\vec{r}}_i(t_0) = \vec{f}_i(t_0)/m_i$ can be evaluated given the atomic position $\vec{r}^{3N}(t_0)$ at $t = t_0$, $\vec{r}^{3N}(t_0 + \Delta t)$ may be approximated by

$$\vec{r}_i(t_0 + \Delta t) = -\vec{r}_i(t_0 - \Delta t) + 2\vec{r}_i(t_0) + \left(\frac{\vec{f}_i(t_0)}{m_i}\right)(\Delta t)^2 + \mathcal{O}((\Delta t)^4) \quad (3.16)$$

By neglecting the $\mathcal{O}((\Delta t)^4)$ term, we obtain a recursion formula to compute $\vec{r}^{3N}(t_0 + \Delta t)$, $\vec{r}^{3N}(t_0 + 2\Delta t)$, ... successively, which is the Verlet algorithm.

The velocities do not participate in the recursion but are needed for property calculations. They can be approximated by $\vec{r}_i(t_0 + \Delta t) - \vec{r}_i(t_0 - \Delta t)$ and using the above Taylor expansion:

$$\vec{v}_i(t_0) \equiv \dot{\vec{r}}_i(t_0) = \frac{1}{2\Delta t}[\vec{r}_i(t_0 + \Delta t) - \vec{r}_i(t_0 - \Delta t)] + \mathcal{O}((\Delta t)^2) \quad (3.17)$$

A related and more commonly used algorithm is the Velocity Verlet algorithm, where the velocity and position are calculated at the same value of the time variable. We start out with $\vec{r}^{3N}(t_0)$ and $\vec{v}^{3N}(t_0)$, then

$$\vec{r}_i(t_0 + \Delta t) = \vec{r}_i(t_0) + \vec{v}_i(t_0)\Delta t + \frac{1}{2}\left(\frac{\vec{f}_i(t_0)}{m_i}\right)(\Delta t)^2 + \mathcal{O}((\Delta t)^3) \quad (3.18)$$

evaluate $\vec{f}^{3N}(t_0 + \Delta t)$, and then,

$$\vec{v}_i(t_0 + \Delta t) = \vec{v}_i(t_0) + \frac{1}{2}\left[\frac{\vec{f}_i(t_0)}{m_i} + \frac{\vec{f}_i(t_0 + \Delta t)}{m_i}\right]\Delta t + \mathcal{O}((\Delta t)^3) \quad (3.19)$$

and we have advanced by one step.

Integrators for Langevin Dynamics. As already introduced in Eq. (3.4), Langevin dynamics consists of adding a random force and subtracting a friction force from each atom during each integration step.

The random force is calculated such that the average force is zero and the standard deviation is: $2k_B T_0 b_i m_i \Delta t$, where k_B is the Boltzmann constant, T_0 the specified target temperature, b_i the friction coefficient for atom i , m_i the mass of atom i and Δt the time-step size. The friction force applied is $b_i \frac{\partial \vec{r}_i}{\partial t}$.

To integrate the Langevin equation, the Brünger-Brooks-Karplus (BBK) method can be used. This is a natural extension of the Verlet method for the Langevin equation. The position recurrence relation of the BBK method is

$$\begin{aligned} \vec{r}_i(t_0 + \Delta t) = & \vec{r}_i(t_0) + \frac{1 - b_i \Delta t / 2}{1 + b_i \Delta t / 2} (\vec{r}_i(t_0) - \vec{r}_i(t_0 - \Delta t)) + \\ & + \frac{1}{1 + b_i \Delta t / 2} (\Delta t)^2 \left[m_i^{-1} \vec{f}_i(t_0) + \sqrt{\frac{2b_i k_B T_0}{(\Delta t) m_i}} \vec{Z}_{t_0} \right] \end{aligned} \quad (3.20)$$

where \vec{Z}_{t_0} is a set of Gaussian random variables of zero mean and variance 1. The BBK integrator requires only one random number for each degree of freedom. The velocities are quantified by variants of Eq. (3.17) or Eq. (3.19).

3.1.6 Statistical Ensembles

In mathematical physics, a *thermodynamic ensemble* (also named *statistical ensemble*) is an idealization consisting of an infinite number of virtual copies of a system, each of which represents a possible state of the system. A thermodynamic ensemble is in statistical equilibrium (defined below), and it is used to derive all the main features of the system. A *microstate* is a configuration that the system may occupy with a certain probability in the course of its thermal fluctuations and corresponding to a fixed configuration (\vec{r}_i, \vec{p}_i) , with coordinates \vec{r}_i and momenta \vec{p}_i .

Because of a large number of particles N , there can be an enormously large number of microstates which collectively form what is known as the *macrostate* of the system. The measure of statistical averages over multiple microstates of the system are called *ensemble averages*.

Different kinds of ensembles exist and they are characterized by fixed values of thermodynamic variables such as the total number of particles N , volume V , temperature T , pressure P , total energy E . Constant NVE defines the *microcanonical ensemble*, and constant NVT defines the *canonical ensemble*, while the *isothermal-isobaric* ensemble is defined by constant NPT .

3.1.7 Constraint Conditions and the Integration Time-step

The choice of a time-step Δt is a crucial issue in MD simulations. An adequate time-step must be small enough to resolve the fastest motions in the system, allowing a numerically stable integration that preserves conservation of system's energy and momentum. However, a value too small will permit only limited exploration of phase space.

A widely used approach to avoid too small time-steps in MD simulations is to eliminate the fastest motions of the system by using constraints. In molecular systems this is done by considering the chemical bonds as rigid, *i.e.* having fixed length. Different algorithms have been developed to implement the constraints, the most widely used is called SHAKE [71].

The SHAKE algorithm. The SHAKE algorithm [71] changes a set of unconstrained coordinates \vec{r}' to a set of coordinates \vec{r}'' that fulfill a list of constraints, using a set of references. Let us assume that the equations of motion must fulfill K holonomic constraints, expressed as:

$$\sigma_k(\vec{r}_1 \dots \vec{r}_N) = 0 \quad (3.21)$$

for $k = 1 \dots K$. The motion of a system subject to a set of holonomic constraints is described *via* the introduction of additional forces expressed in terms of the so-called Lagrange multipliers, *i.e.* unknown parameters to be determined by requiring that the constraints equations are satisfied. SHAKE is a technique to determine these parameters by an iterative procedure which modifies the system's coordinates until all constraint equations are satisfied within a desired accuracy. The total forces on the atom i are defined as:

$$-\nabla_{\vec{r}_i} \left(V + \sum_{k=1}^K \lambda_k \sigma_k \right), \quad (3.22)$$

where λ_k are Lagrange multipliers which must be solved to fulfill the constraint equations. The second part of this sum determines the *constraint forces* \vec{G}_i , defined by

$$\vec{G}_i = - \sum_{k=1}^K \lambda_k \nabla_{\vec{r}_i} \sigma_k \quad (3.23)$$

Therefore, adding Eq.s (3.23) in Eq. (3.2), the $3N$ equations of motion, in which the forces of constraint appear explicitly, are

$$m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} = \vec{f}_i + \vec{G}_i = -\nabla_{\vec{r}_i} V - \sum_1^K \lambda_k \nabla_i \sigma_k \quad (3.24)$$

A solution of Eq. (3.24) can be obtained, following [72], by writing:

$$\lambda_k(t) = \sum_{n=0}^{\infty} \lambda_k^n(t_0) \frac{(t-t_0)^n}{n!} \quad (3.25)$$

and obtaining from Eq. (3.21) an explicit function of the set $\{\lambda^n(t_0)\}$ and then an evaluation of $\vec{r}_i(t, \{\lambda(t)\})$. A possible route to solve these equations with the SHAKE algorithm consists in an iterative process in which all the constraints are handled individually in succession and solved to the first order [72]. To illustrate it, let us focus on a single constraint at a specific iteration cycle. First, we write the Verlet algorithm as:

$$\vec{r}_i^c(t + \Delta t) = \vec{r}_i^u(t + \Delta t) - \frac{\Delta t^2}{m_i} \hat{\lambda}_k \nabla_i \sigma_k(t), \quad (3.26)$$

where $\hat{\lambda}_k$ is the contribution to λ_k at the iteration cycle. We now use a Taylor expansion of the constraint equation to the first order in $\vec{r}_i^c - \vec{r}_i^u$.

$$\sigma_k^c(t + \Delta t) = \sigma_k^u(t + \Delta t) + \sum_{i=1}^N \left(\frac{\partial \sigma_k}{\partial \vec{r}_i} \right)_{\vec{r}_i^u(t+\Delta t)} \cdot [\vec{r}_i^c(t + \Delta t) - \vec{r}_i^u(t + \Delta t)] + O(\Delta t^4) \quad (3.27)$$

By using Eq. (3.26) we obtain

$$\sigma_k^u(t + \Delta t) = \Delta t^2 \hat{\lambda}_k \sum_{i=1}^N \frac{1}{m_i} \nabla_i \sigma_k^u(t + \Delta t) \nabla_i \sigma_k(t) \quad (3.28)$$

and hence

$$\hat{\lambda}_k \Delta t^2 = \frac{\sigma_k^u(t + \Delta t)}{\sum_{i=1}^N \frac{1}{m_i} \nabla_i \sigma_k^u(t + \Delta t) \nabla_i \sigma_k(t)} \quad (3.29)$$

Eq. (3.29) yields the contribution to the Lagrange multiplier λ_k for the constraint k coming from each iteration cycle. Contributions from each cycle to each constraint are summed until all constraint equations have converged. Then, particle velocities are first calculated for the unconstrained system, then modified to meet each constraint. If the Hamiltonian formulation of the dynamics is used, that is, the r^N and p^N are independent variables, we have to impose the constraints on momenta separately, by using the time derivatives of the σ 's.

A method based on SHAKE, is the SETTLE algorithm [73]. SETTLE solves the system of non-linear equations analytically for $K = 3$ constraints in constant time. Although it does not scale to larger numbers of constraints, it is very often used to constrain rigid water molecules, which are present in almost all biological simulations and are usually modeled using three constraints (such as the TIP3P water model).

3.1.8 Evaluation of Non-bonded Interactions

During a classical MD simulation, the evaluation of the right-hand side of Eq. (3.3) is the most time-consuming. Among all the interaction terms, the most time demanding to compute is the one describing non-bonded energy (see Eq. (3.12)), because of the calculation on all copies of atoms (i, j) . In simulations of biological systems it is highly convenient to avoid the calculation of all non-bonded pair interactions, but these interactions primarily governs the behavior of biomolecules, and cannot be simply truncated beyond a given cutoff long ranged.

Ewald Summation and Particle Mesh Ewald (PME). The difference between short and long interactions is the spatial extent of the potential. In general, if a potential function drops down to zero faster than r^d , where r is the separation between two particles and d dimension of the problem, it is called *short ranged*, otherwise it is *long ranged*. Ewald summation, named after Paul Peter Ewald, is a method for computing long-range interactions in periodic systems. It rewrites the original interaction potential as the sum of two different terms, $\phi(\vec{r}) = \phi_{sr}(\vec{r}) + \phi_{lr}(\vec{r})$, where $\phi_{sr}(\vec{r})$ represents the short-range term whose sum quickly converges in real space and $\phi_{lr}(\vec{r})$ represents the long-range term whose sum quickly converges in Fourier (the reciprocal) space. Remarkably, the method requires charge neutrality of the molecular system in order to calculate accurately the total Coulombic interaction. In fact, the

combination of a system with a non-zero net charge and with PBCs would create an infinitely charged configuration which could produce artifacts.

We consider the electrostatic energy of a system of particles for simplicity in a cubic box with PBCs. The treatment is similar for other types of boxes and the strategy introduced here has general validity. At position \vec{r}_i of particle i , the electrostatic potential, ϕ , can be written down as lattice sum:

$$\phi(\vec{r}_i) = \frac{1}{8\pi\epsilon_0} \sum_{\vec{n}=0} \sum_{j=1}^N \frac{q_j}{|\vec{r}_{ij} + \vec{n}L|} \quad (3.30)$$

thus, for N particles, we introduce the energy E :

$$E = \frac{1}{8\pi\epsilon_0} \sum_{\vec{n}=0} \left[\sum_{j=1}^N \sum_{i=1}^N \frac{q_i q_j}{|\vec{r}_{ij} + \vec{n}L|} \right] \quad (3.31)$$

where L is the length of the periodic box, N is the total number of atoms, and $\vec{n} = (n_x, n_y, n_z)$ are the direct lattice vectors (i is not equal to j for $\vec{n} = 0$). This sum extends over infinite number of lattice vectors. Thus, the procedure has to be modified in order to get an absolute convergent sum and to get it fast converging. Mathematically, each point charge q_i can be expressed as a charge density $\rho_i(\vec{r}) = q_i \delta(\vec{r} - \vec{r}_i)$. For every point charge q_i , at the position \vec{r}_i , a gaussian charge distribution of the same magnitude and negative sign is added with charge density:

$$\rho_i^G(\vec{r}) = q_i \left(\frac{\alpha^3}{\sqrt{\pi^3}} \right) \exp(-\alpha^2 |\vec{r} - \vec{r}_i|^2) \quad (3.32)$$

where α is an arbitrary parameter that determines the width of the distribution:

$$\rho(\vec{r}) - \sum_i \rho_i^G(\vec{r}) = \sum_i (q_i \delta(\vec{r} - \vec{r}_i) - \rho_i^G(\vec{r})) \quad (3.33)$$

In order to recover the original charge of the system, another Gaussian charge with magnitude q_i is added and the total charge distribution is:

$$\rho(\vec{r}) = \sum_i q_i \delta(\vec{r} - \vec{r}_i) = \sum_i (q_i \delta(\vec{r} - \vec{r}_i) - \rho_i^G(\vec{r})) + \sum_i \rho_i^G(\vec{r}) \quad (3.34)$$

Therefore if every charge in the system is screened by a counter charge of opposite sign, then the potential of this composite charge distribution is short ranged. In order to compensate for the subtracted charge distribution, this has to be added again, as described in Eq. (3.34). The first sum will only give rise to the short ranged potential that can be calculated in the direct space and it can be truncated, while the second is calculated in the reciprocal space. The corresponding expression for the energy becomes:

$$E = \frac{1}{4\pi\epsilon_0} (\epsilon_{dir} - \epsilon_{self} + \epsilon_{rec}) \quad (3.35)$$

where

$$\epsilon_{dir} = \frac{1}{2} \sum_{\vec{n}=0} \sum_{i \neq j}^N \frac{q_i q_j \operatorname{erfc}(\alpha |\vec{r}_{ij} + \vec{n}L|)}{|\vec{r}_{ij} + \vec{n}L|}, \quad (3.36)$$

$$\epsilon_{self} = \frac{\alpha}{\sqrt{\pi}} \sum_i q_i^2, \quad (3.37)$$

$$\epsilon_{rec} = \frac{1}{2L^3} \sum_{\vec{k} \neq 0} \sum_{i,j=1}^N \frac{4\pi}{k^2} q_i q_j \exp(-i\vec{k} \cdot (\vec{r}_i - \vec{r}_j)) \exp\left(\frac{-k^2}{4\alpha^2}\right) \quad (3.38)$$

where \vec{k} is the reciprocal vector in a cubic box and $k = |\vec{k}|$. Since the complementary error function can be approximated for large arguments by a Gaussian function and the k -space parts decreases like a Gaussian, both terms can be approximated by stopping the sums at a certain lattice vector n and a maximal k -value, *i.e.* k -max. In big O notation, common MD

simulations scale by $O(N^2)$ if all pair-wise electrostatic and van der Waals interactions must be accounted for explicitly. It can be shown that the introduction of the Ewald summation leads to MD simulations with $O(N^{3/2})$.

There are alternative techniques to the Ewald summation for improving the evaluation of long range electrostatic forces, such as Particle Mesh Ewald (PME) [64]. PME is a method to improve the performance of the reciprocal sum in the Ewald summation. Instead of directly summing wave vectors, the charges are assigned to a grid using cardinal B-spline interpolation to arrive at a discretized Poisson equation. The approximate reciprocal PME energy and forces are expressed as convolutions and can thus be evaluated quickly using Fast Fourier Transforms (FFTs), with a $O(N \log N)$ computational cost.

3.1.9 Energy Minimization

Before starting MD simulations, it is desirable to bring the initial molecular structure to a local potential energy minimum. This is mainly done to avoid bad contacts between close atoms, often included in crystal structures or in the *in silico* assembly of proteins with membranes and solvent, that could result in too high forces for the integrator and thus unstable simulations. Various energy minimization algorithms have been devised to this aim, such as the Steepest Descent and the Conjugate Gradient.

Steepest Descent. The steepest descent is the simplest minimization algorithm. It uses the gradient of the function to direct the minimization process toward the nearest local minimum. In this method, one iterates the following equation until it converges or until it reaches a predefined maximum number of steps: $\vec{r}_i^k = \vec{r}_i^{k-1} + \lambda_{k-1} \vec{s}_i^{k-1}$ where \vec{r}_i^k is the actual position of atom i at step k , \vec{r}_i^{k-1} is the position of the step $(k-1)$, λ_{k-1} is the size of the step to be taken at step $k-1$ and \vec{s}_i^{k-1} is the direction of that step. \vec{s}_i^{k-1} is taken to be the opposite of the gradient of the energy function with respect to \vec{r}_i : $\vec{s}_i^{k-1} = -\nabla_i U(\{\vec{r}_i^{k-1}\})$, while λ_k can be adjusted at each iteration step.

Conjugate Gradient. The conjugate gradient method exploits its "memory" of the gradients calculated in previous steps. The first step is taken in the direction of the force: $\vec{s}_i^1 = -\nabla_i U(\{\vec{r}_i^1\})$ but for all subsequent iterations, $k > 1$, the direction of the minimization step is determined as a weighted average of the current gradient and the direction taken in the previous iteration: $\vec{s}_i^k = -\nabla_i U(\{\vec{r}_i^k\}) + b_i^k \vec{s}_i^{k-1}$ where $b_i^k = \frac{|\nabla_i U(\{\vec{r}_i^k\})|^2}{|\nabla_i U(\{\vec{r}_i^{k-1}\})|^2}$. The conjugated gradient method converges usually much better than steepest descent. The default minimizer in NAMD uses a sophisticated conjugate gradient algorithm.

3.1.10 Equilibration and MD Simulations

After energy minimization, the common practice includes equilibration of the solvent. The reason is that the solvent is mostly optimized within itself, and not with the solute and it must be brought to the temperature and pressure required for the simulation. In this part of the simulation, movement of the solute is permitted, but only after overcoming a substantial energy penalty and the utility of position restraints is that they allow us to equilibrate the solvent around the protein, without the added variable of structural changes in the protein. For this subsection the cited references, together with the NAMD (<https://www.ks.uiuc.edu/Research/namd/2.13/ug/>) and GROMACS (<http://manual.gromacs.org/documentation/>) manuals, were used to review the main methods.

Temperature Control. In MD simulations different approaches can be used to obtain a constant temperature (NVT) ensemble such as stochastic randomization through the Andersen thermostat [74], the Berendsen thermostat [75], the *weak-coupling* introduced by [76], the Nosé-Hoover scheme [77, 78] or the Langevin thermostat, based on the Langevin equation (Eq. (3.4)). In general, the fundamental requirement for an integrator is to generate the correct

ensemble distribution for the specified temperature in an appropriate way. For this purpose the equations of motion should be modified.

The simplest possible way to control the temperature would be to set the total kinetic energy at every time-step to match exactly the desired temperature T_0 (*i.e.* scale atom velocities with a suitable factor). However, this may cause significant perturbations of the trajectories.

The Andersen thermostat introduces a stochastic element to the temperature by having *random collisions* of molecules with an imaginary heat bath at T_0 . In particular this method takes an *NVE* integrator and periodically re-select the velocities of the particles from a Maxwell-Boltzmann distribution. This thermostat is also only possible with the velocity Verlet algorithms, because it operates directly on the velocities at each time-step. In the limit of an infinitely long trajectory averaged over many heat bath collisions, it can be shown that the Andersen thermostat rigorously generates the correct canonical ensemble probabilities. However, the presence of random collisions causes the velocities of particles to decorrelate and, therefore, discontinuities in the trajectories can be observed.

On the other hand, the Berendsen algorithm introduces a weak coupling to an external heat bath with a reference temperature T_0 . The temperature of the system is corrected such that the deviation exponentially decays with some time constant τ *i.e.*:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad (3.39)$$

The Berendsen thermostat suppresses the fluctuations of the kinetic energy, generating an improper sampling. However, this error scales with $1/N$, therefore, for very large systems, most ensemble averages could not be affected significantly. The *velocity-rescaling thermostat* [76] is a modified version of the Berendsen thermostat where an additional stochastic term is added to correct the fluctuations of the kinetic energy and obtain the correct distribution:

$$dK = (K_0 - K) \frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}} \frac{dW}{\sqrt{\tau_T}}, \quad (3.40)$$

where K is the kinetic energy, τ_T is a time constant, N_f the number of degrees of freedom of the system and dW a Wiener process *i.e.* a continuous-time stochastic process.

Another approach, to enable canonical ensemble simulations, was derived by Nosé with a reformulation of the equations of motion [77]. After his work, this strategy has been developed by Hoover [78] to establish what is now known as the Nosé-Hoover thermostat, where the Hamiltonian is extended by introducing a thermal reservoir and a friction term. In particular, the Nosé-Hoover thermostat introduces a friction parameter ξ with momentum p_ξ in the particles' equations of motion, *i.e.*:

$$\begin{cases} \frac{\partial \vec{r}_i}{\partial t} = \frac{\vec{p}_i}{m_i} \\ \frac{\partial \vec{p}_i}{\partial t} = \vec{f}_i - \frac{p_\xi}{Q} \vec{p}_i \\ \frac{\partial p_\xi}{\partial t} = T - T_0 \end{cases} \quad (3.41)$$

The strength of the coupling is determined by the constant Q (named the mass parameter of the reservoir) in combination with T_0 .

Another approach is the Langevin thermostat. This method integrates the stochastic Langevin equations of motion in place of Newton's equations. It can be shown that the simulation generated by integrating the Langevin equations (as introduced in Eq. (3.20)) corresponds to a canonical trajectory:

$$\begin{cases} \frac{\partial \vec{r}_i}{\partial t} = \frac{\vec{p}_i}{m} \\ \frac{\partial \vec{p}_i}{\partial t} = \vec{f}_i - \frac{b_i}{m_i} \vec{p}_i + \vec{R}_i(t) \end{cases} \quad (3.42)$$

where \vec{p}_i is the momentum and $\vec{R}_i(t)$ is a Gaussian random process.

Pressure Control. As done with the temperature coupling, the system can also be coupled to a pressure bath to obtain *NPT* simulations. To this aim, we will consider the Berendsen algorithm [75], the extended-ensemble Parrinello-Rhman approach [79], and the mixed Langevin-Hoover [78, 80, 81] algorithm.

The Berendsen algorithm rescales the coordinates and box vectors every step, to obtain a relaxation towards a given reference pressure \mathbf{P}_0 :

$$\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P}_0 - \mathbf{P}}{\tau_p} \quad (3.43)$$

however, it is important to note, as done for the temperature coupling, that the Berendsen pressure control algorithm yields a simulation with the correct average pressure, it does not yield the exact *NPT* ensemble.

The Parrinello-Rahman approach makes each unit vector of the unit-cell independent so that it allows dynamic shape change. The box vectors are represented by the matrix of motion \mathbf{b}

$$\frac{\partial \mathbf{b}^2}{\partial t^2} = V \mathbf{W}^{-1} \mathbf{b}'^{-1} (\mathbf{P} - \mathbf{P}_0) \quad (3.44)$$

The volume of the box is denoted V , and \mathbf{W} is a matrix parameter that determines the strength of the coupling. The equations of motion for the particles are also changed, with:

$$\frac{\partial^2 \vec{r}_i}{\partial t^2} = \frac{\vec{f}_i}{m_i} - \mathbf{M} \frac{\partial \vec{r}_i}{\partial t}, \quad \mathbf{M} = \mathbf{b}^{-1} \left[\mathbf{b} \frac{\partial \mathbf{b}'}{\partial t} + \frac{\partial \mathbf{b}}{\partial t} \mathbf{b}' \right] \mathbf{b}'^{-1} \quad (3.45)$$

For all the simulations produced in this work, the *NPT* ensemble was chosen, using a constant pressure simulation with a modified Nosé-Hoover method in which Langevin dynamics is used to control fluctuations in the barostat [82]. Here, the evolution of both the particle and barostat velocity are described by Langevin dynamics in d dimension, with different friction constants:

$$\left\{ \begin{array}{l} \frac{\partial \vec{r}_i}{\partial t} = \frac{\vec{p}_i}{m_i} + \frac{p_\xi}{Q} \vec{r}_i \\ \frac{\partial \vec{p}_i}{\partial t} = \vec{f}_i(\Lambda) - \gamma \vec{p}_i + \vec{R}_i(t) - \left(1 + \frac{d}{N_f}\right) \frac{p_\xi}{Q} \vec{p}_i \\ \frac{\partial \Lambda}{\partial t} = d \Lambda p_\xi / Q \\ \frac{\partial p_\xi}{\partial t} = d \Lambda (X - P_{ext}) + \frac{d}{N_f} \sum_{i=1}^N \frac{\vec{p}_i^2}{m_i} - \gamma_p p_\xi + R_p(t) \end{array} \right. \quad (3.46)$$

where $\gamma = b_i/m_i$, as already introduced in Eq. (3.42). The forces are the negative derivatives of a potential function $V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N, \Lambda)$: $\vec{f}_i(\Lambda) = -\nabla_{\vec{r}_i} V(\vec{r}, \Lambda)$. These equations introduce the volume Λ as a dynamical-barostat variable. The corresponding momentum variable, p_ξ , is the strain rate $\dot{\xi}$ multiplied by a fictitious mass Q , and R_p is a stochastic force. The scalar X is given by:

$$X = \frac{1}{d\Lambda} \left[\sum_{i=1}^N \frac{\vec{p}_i \cdot \vec{p}_i}{m_i} + \sum_{i=1}^N \vec{r}_i \cdot \vec{f}_i \right] - \frac{\partial}{\partial \Lambda} V(\vec{r}, \Lambda) \quad (3.47)$$

Finally, $\vec{R}_i(t)$ and $R_p(t)$ are drawn from a Gaussian distribution Z_{t_0} of zero mean and unit variance, scaled by $\sqrt{2k_B T \gamma m / \Delta t}$ and $\sqrt{2k_B T \gamma_p Q / \Delta t}$, respectively, where k_B is the Boltzmann's constant. The integrator is introduced in [82].

For *NPT* ensemble simulations, the NAMD software proposes this set of equations, with a little difference respect to [82], *i.e.* the term d/N_f , where $d = 3$ is the dimension and N_f is the number of degrees of freedom, is negligible.

3.2 Introduction to Statistical Mechanics for Thermodynamics

MD simulations generate information at the microscopic level of chemical reactions. Then, with the use of Statistical Mechanics, these microscopic properties are used to obtain macroscopic observables. In particular, thermodynamics gives information regarding the equilibrium conditions of products after the reaction takes place, but does not explain the rate of reaction, which is described by kinetic calculations (see the section below). The main thermodynamic quantity is the energy difference resulting from the free energy given off during a chemical reaction. In this section, the essential concepts of Statistical Mechanics are described for free energy calculations [72, 83–91]. A first branch of Statistical Mechanics, which extends classical thermodynamics, treats equilibrium thermodynamics. However, Statistical Mechanics is also used to study systems that are out of equilibrium, because of irreversible processes.

Equilibrium Thermodynamics Let us consider an observable property O , measured in experiments or MD simulations, as a function of positions \vec{r} and momenta \vec{p} of N particles, *i.e.* $O(\vec{r}(t), \vec{p}(t))$ at a particular time. The value of the following integral approaches the time average of the O feature:

$$\bar{O} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^{\tau} O(\vec{r}(t), \vec{p}(t)) dt \quad (3.48)$$

In principle, a trajectory in time should be generated by numerically solving a set of equations of motion for all interacting particles, from which the average value of the desired property can be computed.

For macroscopic systems containing molecules of the order of the Avogadro constant (10^{23}), it is not feasible to propose the numerical integration to generate a trajectory for $\tau \rightarrow \infty$.

However, the time average of O given in Eq. (3.48), which describes the single system evolving in time can be replaced by the so-called *ensemble average*, where a large number of the replications of the system are considered simultaneously:

$$\langle O \rangle_{ensemble} = \int \int d\vec{p} d\vec{r} O(\vec{p}, \vec{r}) P(\vec{p}, \vec{r}) \quad (3.49)$$

where the brackets $\langle \dots \rangle$ indicate an ensemble average, the average value of property O over the ensemble of configurations generated by the simulation and $P(\vec{p}, \vec{r})$ is the probability density of the ensemble, which is the probability of finding a particular configuration. In the NVT ensemble, $P(\vec{p}, \vec{r})$ is given by:

$$P(\vec{p}, \vec{r}) = \frac{1}{Q} e^{-H(\vec{p}, \vec{r})/k_B T} \quad (3.50)$$

where k_B is the Boltzmann's constant, T is the temperature, and H is the Hamiltonian of the system, which is defined as the sum of kinetic energy K and potential energy U of the system: $H \equiv E = K + U$ and Q is the so-called *partition function*, which is:

$$Q = \int \int d\vec{p} d\vec{r} e^{-H(\vec{p}, \vec{r})/k_B T} \quad (3.51)$$

If the partition function Q is quantified, all the other main thermodynamic properties can be derived from it, such as the Helmholtz free energy A :

$$A = -k_B T \ln Q \quad (3.52)$$

If the pressure, rather than the volume, is kept constant, the Gibbs free energy (usually denoted as G) is obtained. Apart from the change in the ensemble, the formalism is equivalent for A and G . In the condensed phase, which is relevant for most applications, the systems are hardly compressible; so A and G are numerically very similar [59].

The calculation of free energies from atomistic simulations is of great importance in many applications, ranging from the prediction of the phase behavior of a certain substance to the calculation of ligand affinities in drug design.

Nonequilibrium Thermodynamics In the context of nonequilibrium thermodynamics [92], a significant effort has been devoted to the development of efficient free energy calculation algorithms [93,94]. As a consequence of the second law of thermodynamics, the average of work W performed on a system during a nonequilibrium transformation, over many realizations of the non-equilibrium process, $\langle W \rangle$, is larger than the free energy difference ΔG between the equilibrium states corresponding to the transition end points, *i.e.* $\langle W \rangle \geq \Delta G$. The equal sign of this expression holds only if the transformation is carried out reversibly. As shown by Jarzynski in 1997 [94,95], with a result named the *Jarzynski equation*, the work fluctuations resulting for microscopic systems can be described by:

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta G} \quad (3.53)$$

A closely connected result is the Crooks fluctuation theorem [96–98]. These two results can be used in MD simulations for the calculation of *equilibrium* properties, such as free energy calculations [92].

3.3 Basic Concepts of Kinetic Calculations

Chemical reaction kinetics is the area of chemical physics that investigates the rates of chemical reactions. As a general rule, elementary processes involve a transition between two atomic or molecular states (named the *reactants* and the *products*, respectively). The potential barrier constitutes the activation energy of an elementary process, and determines the rate at which it occurs. When the barrier is low, the thermal energy of the reactants will generally be high enough to surmount the barrier and move over to products, and the reaction will be fast.

Atomically detailed MD simulations are available to compute efficiently thermodynamic and equilibrium behavior, but are limited when studying non-equilibrium processes and kinetics [99]. Furthermore, typical trajectories of condensed phase systems are too short to investigate the kinetics of many interesting biophysical systems such as protein conformational transitions, ion permeation and protein folding. Extending the time scale of molecular simulations for kinetic calculations is therefore an important research direction and has attracted the attention of many investigators [99].

The most straightforward computational approach to study kinetics would be to run a large number N trajectories initiated at the reactant. The trajectories are integrated as a function of time using a dynamical model and a numerical algorithm described above. As the time advances, some of the trajectories make it to the product state and complete the reaction, successfully. By convention, an absorbing boundary condition at the product state can be introduced, *i.e.* every trajectory that reaches the product state vanishes.

To measure the rate of the typical time of the reaction, the Mean First Passage Time (MFPT), denoted by $\langle \tau \rangle$, can be introduced. The MFPT is the average time that it takes a collection of trajectories to reach a final state, starting from the initial state. As some of the trajectories disappear the population, $N(t)$, decays as a function of time t . If the decay is of the form $N(t) = N e^{-t/\langle \tau \rangle}$ then a rate coefficient of the reaction is $k \equiv 1/\langle \tau \rangle$. Alternatively, the MFPT can also be estimated as the average time that it takes for the trajectory to reach absorbing boundary $\langle \tau \rangle = 1/N \sum_{i=1, \dots, N} t_i$ where the t_i is the termination time of the i -th trajectory. It is important to comment that while the rate coefficient k is more widely used in the biochemical literature, it is less general than the MFPT. In fact, the rate coefficient is meaningful only for reactions that follow exponential kinetics, while MFPT is a well-defined quantity for any type of kinetics. The strategy described above of running multiple trajectories from reactant to products is the most straightforward way to estimate the MFPT; however it is not feasible in numerous cases. The problem is that the MFPT can be long compared with the timescale accessible to MD simulations.

In the recent years, several approaches have been devised to perform kinetic calculations. Some of these techniques (such as Milestoning, described below) replace the long trajectories between reactants and products by a large number of short trajectories, named *elementary processes*, to estimate local transition rates between nearby positions. The kinetics of the full process is then reconstructed by combining together the transition rates between the intermediate steps. A clear advantage of this approach is that the calculations of the local trajectories can be

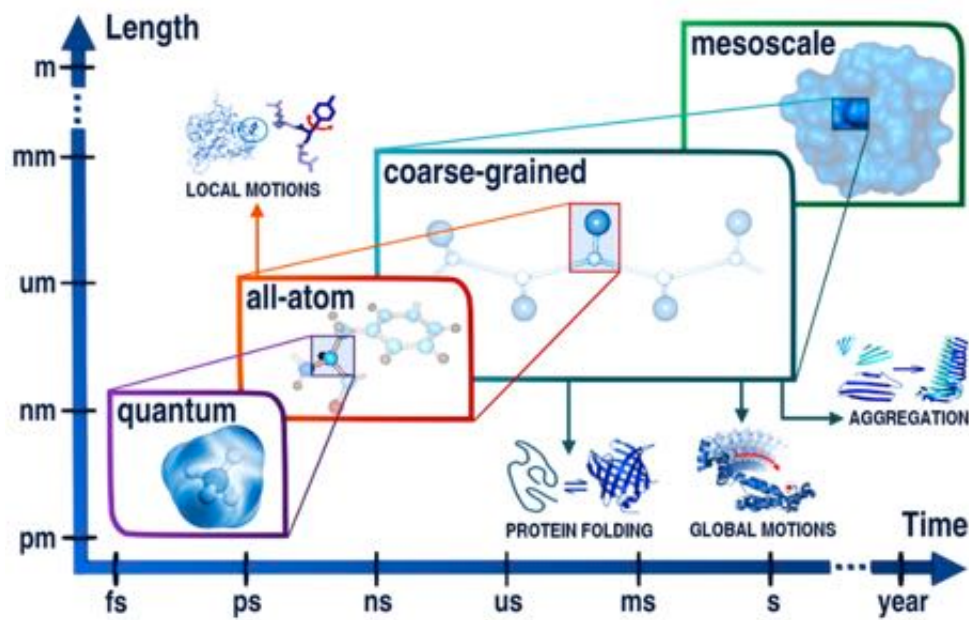


Fig 3.3. Multiscale simulations for biological processes. Application ranges for molecular modeling techniques at different resolutions: quantum, all-atom, coarse grained, and mesoscale. The plot shows approximate ranges of time scales and system sizes (lengths). Reprinted from [15], <https://pubs.acs.org/doi/abs/10.1021/acs.chemrev.6b00163>. Copyright (2016), with permission from the American Chemical Society.

distributed over High Performance Computing (HPC) clusters.

3.4 Enhanced Sampling Techniques in Biomolecular Simulations

In general, as described in the previous sections, MD simulations are used to compute averages of microscopic observables over a finite interval of time. However, many biomolecular motions are separated by timescales that span many orders of magnitude (see **Fig 3.3**) and the challenge for MD is to preserve the details of the accuracy of the resolution recording short timescale processes while observing slow or rare biological events. To achieve this goal, three computational strategies are possible.

1. Atomistic standard MD implemented on advanced computational architectures. As a first example, the development of molecular modeling software packages on Graphic Processing Units (GPUs) [100–104] has made significant impact in accelerating MD calculations. Another remarkable example of a massively parallel machine is Anton computer, developed by the D. E. Shaw Research group in New York. Anton is a special-purpose system for MD simulations of biological macromolecules, with a substantial number of Application Specific Integrated Circuits (ASICs), interconnected by a specialized high-speed, three-dimensional torus network. These hardware machines are customized for running only MD of biomolecules and have provided great opportunity to simulate microsecond to millisecond processes. However, access to such resources is currently limited to few researchers.
2. Since atomistic resolution models alone are presently not efficient enough to handle large system sizes and simulation timescales, the Coarse Grained (CG) approach [15] can be used to investigate complex processes such as the folding of small proteins or

dimerization processes. CG molecules are represented by *pseudo-atoms* approximating groups of atoms, such as a whole amino-acid residue. By decreasing the degrees of freedom, much longer simulation times can be studied at the expense of molecular detail. Although longer timescales than in all-atom simulations can be reached, the limitations in accruing statistics of reactive events mentioned above typically applies also for CG simulations.

3. A different strategy to bridge the timescale gap leverages the development of statistical mechanics-based methods, able to preserve the atomic resolution and to achieve direct sampling of the rare, reactive events. Efforts in this area have resulted in many notable sampling methods ([56], and [105–108]).

- Some approaches fall under the scheme of modifying the equations of motions. For example, Conformational Flooding [109], Puddle Jumping [110], Puddle Skimming [111], Hyperdynamics [112], accelerated MD (aMD) [113] form a family of computational strategies where the potential of the system $V(\vec{r})$ is modified in different ways.

Another family of algorithms such as Umbrella Sampling (US) [57–59], Metadynamics [114–117], the Adaptive Biasing Force (ABF) method [118–120], and Temperature Accelerated Molecular Dynamics (TAMD) [121], requires the introduction of Collective Variables (CVs, also named Reaction Coordinates, RCs). CVs are descriptors that can identify interesting collective phenomena over long timescales, and separate macroscopically different structures or aggregation states. The general idea of these methods is to fill up the potential energy minima with a bias, as a function of CVs, so that less computational time is required to move the system out of the minima. The application of almost all these methods, eventually with the addition of further algorithms, allows calculations of PMF profile, in the framework of equilibrium or quasi-equilibrium thermodynamics.

- A different class of enhanced sampling techniques are those affine to the Replica Exchange Molecular Dynamics (REMD) method.

In the so-called Parallel Tempering Scheme [122], several copies of the system are initiated at different temperatures and configurations sampled in various replicas are allowed to exchange intermittently with the Metropolis criterion. Examples include replica exchange MD [123], replica exchange with solute tempering [124], integrated tempering MD [125], to name a few. Several variants of the traditional temperature dependent REMD (T-REMD) were implemented, such as reservoir REMD (R-REMD) [126]. A more general form of REMD involves exchanges between different Hamiltonians (H-REMD) providing an enhanced sampling in dimensions other than temperature [126,127]. REMD methods are often coupled to the methods listed above, to enhance sampling.

- Free energy calculations from nonequilibrium MD algorithms, as an alternative approach to the computational methods described in the previous groups.

An example is Steered Molecular Dynamics (SMD) [128–132] that accepts *irreversibility*. SMD can be applied restraining a group of atoms to a point in space (restraint point) by an external potential. The restraint point is then shifted in a chosen direction, allowing the selected group to explore new contacts along the MD path. We assume a single RC ξ , and an external harmonic potential $U = \frac{K}{2}(\xi - \xi_0)^2$, where K is the stiffness of the restraint, and ξ_0 is the initial position of the restraint point moving with a constant velocity v . In general, SMD simulations require selection of a path, *i.e.* a series of directions of the applied force. In this way, Jarzynski's equality, introduced in Eq. (3.53), can be used to extract equilibrium free energy differences (ΔG) from work done through nonequilibrium processes [133].

- Other approaches include algorithms for kinetic calculations. Many historical methods for kinetic calculations, such as Transition State Theory (TST) [134,135], apply

to problems where a well-defined transition state (*i.e.* a single, high energy barrier) separates the reactants from products. However, many processes cannot be represented as a transition over a dominant barrier.

For example, in *activated processes*, rare short time trajectories pass over significant free energy barriers. In this context, a significant progress has been made with the introduction of methods such as Transition Path Sampling [136, 137], Milestoning [60, 138, 139], and others [140–142].

On the other hand, progress has been slower for the study of *diffusive processes* (or mixed processes *i.e.* both activated and diffusive processes) in which the times of the individual transitional trajectories are intrinsically long. Computational techniques such as Milestoning and Partial Path Transition Interface Sampling [143] are more suited to study diffusive processes, sharing the idea of breaking the full trajectory into shorter, independent segments.

3.5 Details of the Methods Applied in This Thesis

In this paragraph, I describe in more detail the specific algorithms employed in this thesis for free energy and kinetics calculations.

3.5.1 Umbrella Sampling Simulations and the Weighted Histogram Analysis Method

Umbrella Sampling (US) simulations [57–59] were firstly chosen to investigate the PMF profile of three different ions (Na^+ , K^+ and Cl^-) and understand if the Cldn15 model is able to reproduce the correct charge selectivity of a Cldn15 paracellular channel. The US algorithm is a robust method to calculate PMF profiles. It allows to divide the calculation with different short simulations typically called *windows*.

The algorithm has been widely used to study the permeation of ions through different biological channels. In particular, the ion transport of different conventional channels has been studied *via* US simulations, such as Voltage Gated Potassium channels (KV channels) [144–148] and Voltage Gated Sodium channels (NaV channels) [149–156]. In addition, US simulations have been used to quantify the thermodynamic features of other biological processes such as protein dimerization [157] and protein conformational dynamics [158, 159].

Although US simulations are laborious, they are more accurate than the use of steered MD simulations together with the Jarzynski’s equality. In particular, in [160] both methods were used to calculate the PMF for ion permeation and ligand binding to ion channels. Comparison of those results indicates that Jarzynski’s method suffers from relaxation problems and requires longer simulation times to yield reliable PMFs.

For our purposes, all data generated by these US simulations were combined to obtain the global PMF, using the Weighted Histogram Analysis Method (WHAM) [161–163].

In US calculations [57, 58], a biasing potential energy term is added to the MD potential to ensure efficient sampling along the chosen CV. Each US simulation is performed at different values of the CV, in different independent windows. For a system with n degrees of freedom whose position in configuration space $\Omega \subseteq \mathbb{R}^n$ will be denoted by \vec{r} , we consider a single CV, $z(\vec{r})$ and a biasing harmonic potential expressed as:

$$U_{wi}(\vec{r}) = \frac{1}{2} \hat{\kappa} [z(\vec{r}) - z_i^0]^2 \quad (3.54)$$

is added to the force field, $V(\vec{r})$, where z_i^0 indicates the value at which the CV is restrained in window i and $\hat{\kappa}$ is an elastic constant.

A number N_{wi} of simulations can be performed each with different bias potentials, covering adjacent windows in the CV space. From these simulations, a set of biased (b) probability distributions $\{p_i^{(b)}(z)\}$ will be calculated and an additional family of unbiased (u) probability distributions $\{p_i^{(u)}(z)\}$ can be obtained, as introduced in the review [59]. In this subsection, the superscript b denotes biased quantities, whereas the superscript u describes unbiased quantities

(quantities without superscripts are always unbiased). The unbiased probability distribution of the system at a specific value $z(\vec{r})$, z , is

$$p_i^u(z) = \frac{\int e^{-\beta V(\vec{r})} \delta(z'(\vec{r}) - z) d^N \vec{r}}{\int e^{-\beta V(\vec{r})} d^N \vec{r}} \quad (3.55)$$

where N is the number of degrees of freedom of the system and $\beta \equiv 1/k_B T$. MD simulations of the biased system provides the biased distribution along the reaction coordinate, p_i^b :

$$p_i^b(z) = \frac{\int e^{-\beta(V(\vec{r})+U_{wi}(z'(\vec{r})))} \delta(z'(\vec{r}) - z) d^N \vec{r}}{\int e^{-\beta(V(\vec{r})+U_{wi}(z'(\vec{r})))} d^N \vec{r}} \quad (3.56)$$

Because the integration in the numerator is performed over all degrees of freedom, but z ,

$$p_i^b(z) = e^{-\beta U_{wi}(z)} \frac{\int e^{-\beta V(\vec{r})} \delta(z'(\vec{r}) - z) d^N \vec{r}}{\int e^{-\beta(V(\vec{r})+U_{wi}(z'(\vec{r})))} d^N \vec{r}} \quad (3.57)$$

Therefore, using Eq. (3.55) results in:

$$p_i^u(z) = p_i^b(z) e^{\beta U_{wi}(z)} \frac{\int e^{-\beta V(\vec{r})} e^{-\beta U_{wi}(z(\vec{r}))} d^N \vec{r}}{\int e^{-\beta V(\vec{r})} d^N \vec{r}} \quad (3.58)$$

that can be written as:

$$p_i^u(z) = p_i^b(z) e^{\beta U_{wi}(z)} \langle e^{-\beta U_{wi}(z(\vec{r}))} \rangle \quad (3.59)$$

Therefore, the corresponding unbiased (u) probability distribution, $p_i^{(u)}(z)$, from simulation i is defined as:

$$p_i^{(u)}(z) = e^{\beta[U_{wi}(z)-f_i]} p_i^b(z) \quad (3.60)$$

where $f_i = -(1/\beta) \ln \langle e^{-\beta U_{wi}(z(\vec{r}))} \rangle$ is the free energy coming from the adding of the biasing potential $U_{wi}(z)$ to the reference potential and it is independent of z . In the following, the free energies $\{f_i\}$ are assumed to be known, using a method that provides their computation *a posteriori*.

From Eq.s (3.59) (3.60), the free energy (or PMF) of the i -th window can be quantified, because $p_i^b(z)$ is obtained from an MD simulation of the biased system, $U_{wi}(z)$ is given analytically and f_i is independent of z :

$$G_i(z) \equiv -(1/\beta) \ln p_i^u(z) = -(1/\beta) \ln p_i^b(z) - U_{wi}(z) + f_i \quad (3.61)$$

Therefore, $p_i^b(z)$ is computed as normalized histogram of the values of z occurring the i -th simulation. As long as one window spans the whole range of z , Eq. (3.61) is sufficient. $G(z)$ is in any case only defined up to an additive constant; so in this case, f_i can be chosen arbitrarily. In general, the f_i terms are associated with introducing the bias potential and connect the PMF curves $G_i(z)$ obtained in the different windows:

$$e^{-\beta f_i} = \langle e^{-\beta U_{wi}(z)} \rangle = \int p^u(z) e^{-\beta U_{wi}(z)} dz = \int e^{-\beta(G(z)+U_{wi}(z))} dz \quad (3.62)$$

with $p^u(z)$ being the global unbiased distribution. However, the f_i cannot directly be obtained from sampling and additional methods are required.

For an appropriate choice of $\hat{\kappa}$ in Eq. (3.54), a sufficient overlap between the sampled distributions of adjacent windows is obtained and the full PMF profile $G(z)$ can be calculated with the Weighted Histogram Analysis Method (WHAM) [161–163]. The method is presented here as an extension of the US method for PMF calculations, as introduced in [162]. WHAM has proven to be very efficient and almost free of information loss. The main idea, which goes back to the histogram method developed in [164], consists in writing an optimal unbiased probability distribution $p(z)$, as a linear z -dependent combination of unbiased probability distribution $p_i^u(z)$. Therefore it requires the introduction of a *weighted* sum of the unbiased distribution functions extracted from each window, $p(z)$:

$$p(z) = A \sum_{i=1}^{N_w} \pi_i(z) p_i^u(z). \quad (3.63)$$

The weights π_i are functions of z and A is a normalization constant. The weights are chosen to minimize the variance of $p(z)$, *i.e.* $\sigma^2(p(z))$, such as:

$$\frac{\partial \sigma^2(p(z))}{\partial \pi_i} = 0 \quad (3.64)$$

and subject to normalization $\sum_i \pi_i(z) = 1$. Then $p(z)$ takes the form:

$$p(z) = A \sum_{i=1}^{N_w} \frac{n_i e^{-\beta[U_{wi}(z)-f_i]}}{\sum_{j=1}^{N_w} n_j e^{-\beta[U_{wj}(z)-f_j]}} p_i^u(z) = A \sum_{i=1}^{N_w} \frac{n_i}{\sum_{j=1}^{N_w} n_j e^{-\beta[U_{wj}(z)-f_j]}} p_i^b(z). \quad (3.65)$$

where n_i is the number of sampled points in the i -th window and Eq. (3.60) has been used. Then, Eq. (3.65) can be determined with the use of the Lagrange λ multiplier method:

$$\begin{aligned} \frac{\partial}{\partial \pi_j(z)} \left[\sigma^2(p(z)) - \lambda \left(\sum_i \pi_i(z) - 1 \right) \right] &= \\ = \frac{\partial}{\partial \pi_j(z)} \left[A^2 \sum_i \pi_i^2(z) \sigma^2(p_i^u(z)) - \lambda \left(\sum_i \pi_i(z) - 1 \right) \right] &= \\ = 2A^2 \pi_j(z) \sigma^2(p_j(z)) - \lambda = 0 \end{aligned} \quad (3.66)$$

which, after writing $\lambda/2A^2 = 1/\sum_i [\sigma^2(p_i^u(z))]^{-1}$ (from normalization), gives:

$$\pi_j(z) = \frac{[\sigma^2(p_j^u(z))]^{-1}}{\sum_{i=1}^{N_w} [\sigma^2(p_i^u(z))]^{-1}}. \quad (3.67)$$

In a real simulation is clearly useful to express the weights in terms of the known biasing potential $U_{wi}(z)$, often described as in Eq. (3.54). From Eq. (3.66), one has

$$\sigma^2(p_i^u(z)) = \left[e^{\beta[U_{wi}(z)-f_{wi}]} \right]^2 \sigma^2(p_i^b(z)) \quad (3.68)$$

The statistical error $\sigma^2(p_i^b(z))$ of the computed histogram $p_i^b(z)$ can be expressed as in reference [165]:

$$\sigma^2(p_i^b(z)) = \frac{g_i(z)}{n_i \Delta z} \overline{p_i^b(z)} \quad (3.69)$$

where $\overline{p_i^b(z)}$ is the histogram computed from an infinite length simulation, n_i is the finite length of the simulation i , *i.e.* the number of coordinate sets used to compute $p_i^{(b)}(z)$. Δz is the resolution of $p_i^{(b)}(z)$, that is the width of the bins used to calculate the histogram. Moreover,

$$g_i(z) = \left(1 + \frac{\tau_i(z)}{\delta t} \right) \quad (3.70)$$

where $\tau_i(z)$ is a correlation time and δt is the time step between two consecutive coordinate sets. Let us consider that the factors $g_i(z)$ are the same for all the windows. An estimate of $\overline{p_i^b(z)}$ is given by Eq. (3.60) where $p_i^{(b)}(z)$ is replaced by $\overline{p_i^b(z)}$ and $p_i^{(u)}(z)$ by the probability distribution $p(z)$.

$$\overline{p_i^b(z)} = e^{-\beta[U_{wi}(z)-f_i]} p(z) \quad (3.71)$$

Using Eqs. (3.60), (3.67), (3.69), (3.71), it can be shown that

$$\pi_i(z) = \frac{n_i e^{-\beta[U_{wi}(z)-f_i]}}{\sum_{j=1}^{N_w} n_j e^{-\beta[U_{wj}(z)-f_j]}} \quad (3.72)$$

For all the US simulations performed, the WHAM version from the Grossfield group, available at <http://membrane.urmc.rochester.edu/content/wham>, was used to reconstruct the global PMF.

For the statistical error evaluation, the method described in [165], which introduces Eq. (3.69), was not used. Alternatively, we used another method *i.e.* the bootstrap method [166–169], which is included in the WHAM code. The premise of bootstrapping error analysis requires that, for a time series containing N points, N are chosen randomly, allowing duplication to create sets of so-called *fake data*. Then the average is computed from this fake data sets. This procedure is repeated a number of times to compute the standard deviation of the average of the data sets. This standard deviation is an estimate for the statistical uncertainty of the average computed using the real data.

3.5.2 Milestoning MD Simulations

Milestoning is a theory and a collection of algorithms that were introduced to compute long time kinetics, in addition to thermodynamical properties. The method has been chosen for its ability to perform not only PMF calculations but also kinetic calculations. These simulations give, on one hand, a robust comparison with the previous US calculations, while, on the other hand, they produce the details of the transition rates for ionic transport. Milestoning at present is a general and rigorous theory [138] and the original version of the method was used to study relevant biological processes such as the conformational transitions of the HIV Reverse Transcriptase [170] and the impact of protonation on early translocation of the anthrax lethal factor [171].

In this Thesis we used a variant of the original method, named *Voronoi Tessellated Markovian Milestoning* (VTMM), with the soft-walls potential restraints for confinement, introduced in [139]. In VTMM [60], several MD simulations are confined within cells that span the space between metastable states, and the kinetic properties of the transition are reconstructed from the hittings on the cell boundaries, identified as milestones.

Suppose we are given a set of N points in the CV z -space (z_1, z_2, \dots, z_N) to which we associate a Voronoi tessellation of the space in N cells. It was shown [60] that the statistical properties of the long-time dynamics of the system can be reconstructed from independent simulations, each confined inside one of the N cells. A fundamental requirement of the confinement is that it must leave unperturbed the dynamical properties of the systems when it is in the interior of the cell as well as the probability flux in and out of the cells.

Soft-Walls restraining potentials. One possible strategy to perform the confinement in a cell is the use of the so-called Soft-Walls restraints, *i.e.* half-pseudoharmonic restraining potentials [139]. In the case of the 1D CV, $\theta(\vec{r}) \equiv z$, these potentials are expressed, for each Voronoi cell N_α identified by the point z_α , as:

$$\tilde{\omega}_\alpha(z) = \begin{cases} \frac{1}{2}\tilde{\kappa}(z - z_{upper}^\alpha)^2 & \text{if } z > z_{upper}^\alpha \\ 0 & \text{if } z_{lower}^\alpha \leq z \leq z_{upper}^\alpha, \text{ centered in } z_\alpha \\ \frac{1}{2}\tilde{\kappa}(z - z_{lower}^\alpha)^2 & \text{if } z < z_{lower}^\alpha \end{cases} \quad (3.73)$$

where z_{lower}^α and z_{upper}^α denote the edges of the cell and are the two midpoints between the center z_α and the adjacent ones.

Rates calculation. By considering the edges of the Voronoi cells as milestones, the dynamics of the system is reduced to that of a discrete state continuous-time Markov chain in the state space of milestones indices.

Let us define a rate matrix q_{ij} , with i and j indexes of milestones, whose elements are given by $q_{ij} = N_{ij}/R_i$, where N_{ij} is the number of transitions from i to j , and R_i is the total time that milestone i was the last crossed. They can be expressed in terms of quantities extracted from the confined simulations, *i.e.* if we indicate with α the cell index: $N_{ij} = \sum_{\alpha=1}^N \pi_\alpha(N_{ij}^\alpha/T_\alpha)$ and $R_i = \sum_{\alpha=1}^N \pi_\alpha(R_i^\alpha/T_\alpha)$. Here π_α is the equilibrium probability of finding the system in cell α , T_α is the duration of the simulation in the cell, and N_{ij}^α and R_i^α are defined as N_{ij} and R_i but for the simulation in α .

The PMF associated to the cells is obtained as $-k_B T \ln(\pi_\alpha)$. The rate matrix q_{ij} can be used to quantify the mean first passage times (MFPTs) from any milestone to any other. In fact,

the MFPTs, as the inverse of the rates, are calculated from the matrix q_{ij} by solving a linear system, as reported in [60, 139]. For instance, if τ_i^N $i = 1, \dots, N - 1$ denote the MFPTs from milestone i to milestone N ($\tau_N^N = 0$ by definition), then these MFPTs can be computed by solving:

$$\sum_{j=1}^{N-1} q_{ij} \tau_j^N = -1 \quad i = 1, \dots, N - 1 \quad (3.74)$$

In particular, in this study, τ_i^N denotes the MFPT from the starting milestone i at the mouth of the channel to milestone N at its end.

Since we are interested in the relative permeation time-scales of the ions, and because there is only one entry and one exit portal in our structure, the entry kinetics contributions related to bulk concentration and diffusivity, that are discussed in [172], are neglected.

3.5.3 Single-Sweep Simulations and Pathway Reconstruction

Finally the SSw algorithm [61–63] together with the String Method [173–176] were used to calculate the PMF surfaces in an extended three dimensional (3D) CV space, as well as the minimum free energy paths (MFEPs) and energy profiles along them. Calculations were performed for a set of cations (Li^+ , Na^+ , K^+ , Ru^+ , Cs^+) and the α -D-Glucose (ADG) molecule. The SSw method combines the Temperature Accelerated Molecular Dynamics (TAMD) [121, 177] with a variational reconstruction method using radial-basis functions for the representation of the PMF surface. In particular, TAMD is used to rapidly sweep through the important regions of the PMF landscape and compute the gradient of the free energy locally at points in these regions. TAMD alone was applied in different relevant biological studies such as [178–183]. After the selection of a set of discrete points of the landscape, the variational method is then used to reconstruct the PMF globally from the mean force at these points. SSw has already been used in different applications such as the study of the pathways of diffusion of small molecules inside myoglobin proteins [172, 184, 185], oxygen pathways in monomeric sarcosine oxidase [186] and in in [FeFe]-Hydrogenase [187], but not for the study of ionic permeation. After 3D-PMF calculation with the SSw method, a 1D-PMF profile is calculated along the MFEPs extracted from the 3D map using the string method. Studies and applications of the String Method can be found in [188–194].

For a system with n degrees of freedom whose position in configuration space $\Omega \subseteq \mathbb{R}^n$ will be denoted by \vec{r} , we consider a set of N collective variables $\vec{\theta}(\vec{r}) = (\theta_1(\vec{r}), \dots, \theta_N(\vec{r}))$, the potential energy of the system $V(\vec{r})$, $1/\beta \equiv k_B T$ where k_B is Boltzmann’s constant and T the system’s temperature.

TAMD for exploring the unknown PMF surface. In TAMD [121, 177], the system is extended to include \vec{z} variables that act as anchor points for CVs, involving the potential: $\omega_\kappa(\vec{r}, \vec{z}) = V(\vec{r}) + \frac{1}{2} \kappa |\vec{\theta}(\vec{r}) - \vec{z}|^2$, where $\kappa > 0$ is an adjustable spring-constant-like parameter. The extended system evolves according to the equations:

$$\begin{cases} m \ddot{\vec{r}} = -\nabla V(\vec{r}) - \kappa \sum_{\alpha=1}^N (\theta_\alpha(\vec{r}) - z_\alpha) \nabla \theta_\alpha(\vec{r}) + \text{thermostat at } \beta^{-1} \\ \bar{\gamma} \dot{\vec{z}} = \kappa (\vec{\theta}(\vec{r}) - \vec{z}) + \sqrt{2\bar{\gamma}\bar{\beta}^{-1}} \bar{\eta}(t) \end{cases} \quad (3.75)$$

where m is the mass, $\bar{\eta}(t)$ is a Gaussian process with mean 0 and covariance $\langle \eta_\alpha(t) \eta_{\alpha'}(t') \rangle = \delta_{\alpha\alpha'} \delta(t - t')$, $\bar{\gamma}$ is a friction coefficient and $1/\bar{\beta}$ is an artificial temperature with $1/\bar{\beta} > 1/\beta$, where $1/\beta$ is the inverse of the temperature $k_B T$.

As shown in [121] by adjusting the parameter κ so that $\vec{z}(t) \sim \theta(\vec{r}(t))$ and friction coefficient γ so that \vec{z} moves slower than \vec{r} , a trajectory $\vec{z}(t)$ can be obtained which moves at the artificial temperature $1/\bar{\beta}$ on the PMF computed at the physical $1/\beta$.

By $1/\bar{\beta} > 1/\beta$, the $\vec{z}(t)$ trajectory visits rapidly the regions where the PMF is relatively low (*i.e.* within a range of a few $1/\bar{\beta}$) even if these regions are separated by barriers which the system would take a long time to cross at the physical temperature $1/\beta$.

Mean Force Calculations. After the TAMD simulations, a set of representative points visited during the $\vec{z}(t)$ trajectory (named *centers*) is selected. To compute the mean forces \vec{f}_k , at each of the centers \vec{z}_k , a simulation of

$$m\ddot{\vec{r}} = -\nabla V(\vec{r}) - \bar{\kappa} \sum_{\alpha=1}^N (\theta_{\alpha}(\vec{r}) - z_{k,\alpha}) \nabla \theta_{\alpha}(\vec{r}) + \text{thermostat at } \beta^{-1} \quad (3.76)$$

with \vec{z}_k fixed, is performed and then the mean force is computed:

$$\vec{f}_k = \frac{1}{T} \int_0^T \bar{\kappa} (\vec{\theta}(\vec{r}(t)) - \vec{z}_k) dt \quad (3.77)$$

because the negative gradient of the PMF $G(\vec{z})$, *i.e.* the mean force $\vec{f}(\vec{z}) = -\nabla G(\vec{z})$, can be computed locally at point \vec{z} via calculation of an expectation [195,196].

In general, the choice of a large $\bar{\kappa}$ reduces the error due to the spring constant finiteness [195,196].

PMF representation. After the calculation of the mean forces $\vec{f}_1, \dots, \vec{f}_K$, the N -dimensional PMF is reconstructed using a radial basis function (RBF) representation of $G(\vec{z})$ with centers $\vec{z}_1, \dots, \vec{z}_K$:

$$G(\vec{z}) = \sum_{k=1}^K a_k \phi_{\sigma}(|\vec{z} - \vec{z}_k|) + C \quad (3.78)$$

Here C is a constant, $|\dots|$ indicates the Euclidean norm in \mathbb{R}^N , and $\phi_{\sigma}(u)$ is a RBF with length-scale parameter σ . In this work, we used Gaussian functions $\phi_{\sigma}(u) = \exp(-u^2/2\sigma^2)$. The unknown coefficients a_k in the representation and the optimal σ parameter are determined as described in the original work [61] by minimizing an objective function, which describes the difference between the negative gradients of Eq. (3.78) and the computed mean forces :

$$E(a, \sigma) = \sum_{k=1}^K |\nabla_z A(z_k) + f_k|^2 = \sum_{k=1}^K \left| \sum_{k'=1}^K a_{k'} \nabla_z \phi_{\sigma}(|\vec{z}_k - \vec{z}_{k'}|) + \vec{f}_k \right|^2 \quad (3.79)$$

The minimization of Eq. (3.79) is performed in two steps. First a minimization of $E(a_1, \dots, a_K, \sigma)$ over a_k at σ fixed is carried out. Since Eq. (3.79) is quadratic in a_k , the minimizer is the solution of the following linear algebraic system:

$$\sum_{k'=1}^K B_{k,k'}(\sigma) a_{k'}^*(\sigma) = c_k(\sigma) \quad (3.80)$$

where

$$B_{k,k'}(\sigma) = \sum_{k''=1}^K \nabla_z \phi_{\sigma}(|\vec{z}_k - \vec{z}_{k''}|) \cdot \nabla_z \phi_{\sigma}(|\vec{z}_k - \vec{z}_{k''}|) \quad (3.81)$$

$$c_k(\sigma) = - \sum_{k'=1}^K \nabla_z \phi_{\sigma}(|\vec{z}_k - \vec{z}_{k'}|) \cdot \vec{f}_{k'} \quad (3.82)$$

Therefore, Eq. (3.80) can be solved by any standard technique. Then the scalar function $\epsilon_{\sigma} \equiv E(a_1(\sigma), \dots, a_K(\sigma), \sigma)$ is introduced and minimized over σ .

Globally, compared to other methods, the SSw technique combines several advantages [61]:

- It does not require *a priori* assumptions on the PMF since it uses TAMD to find the important regions in the landscape automatically.
- The computation of the mean forces at the centers, which is the most costly step of the calculation, can be distributed on different, independent, processors. This is the main feature of the SSw method and it differentiates this approach from another TAMD-based PMF reconstruction, named Temperature Accelerated MD/On-The-Fly Parameterization (TAMD/OTFP) [2–6].

- The results can be easily monitored for convergence, and systematically improved if desired. In particular new centers can be added on top of previous ones along the same TAMM trajectory to increase the accuracy without having to repeat the previous calculation.
- The method can be used in more than 2 dimensions and its computational complexity is the same regardless of the dimension.

MFEPs as ion permeation paths. Once the PMF surface is known, pathways for diffusion inside the channel are identified as MFEPs on the surface. An MFEP is defined as the curve whose tangent is always parallel to $\mathbf{M}\nabla G(\vec{z})$, where \mathbf{M} is a metric tensor [175] which for the case of linear CVs chosen here, is constant and diagonal. Since an analytical approximation of this surface was obtained with Gaussian functions, such paths can be computed with the Zero Temperature String (ZTS) method [173, 174]. Given an initial guess for a curve on the PMF surface, the ZTS method finds the closest MFEP by moving a discrete set of points on the curve by steepest descent on the PMF landscape, at the same time keeping the points at constant distance from each other.

3.5.4 Root-Mean-Square Deviation

RMSD. The *Root-Mean-Square Deviation* (RMSD) is the measure of the average distance between the atoms (usually the backbone atoms) of superimposed proteins defined as:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$$

where δ is the distance between N pairs of equivalent atoms. Equivalently, given two sets of N points \vec{v} and \vec{w} , where each point i of the two points is described by a Cartesian tern (v_{ix}, v_{iy}, v_{iz}) and (w_{ix}, w_{iy}, w_{iz}) , the RMSD is defined as follows:

$$\text{RMSD}(\mathbf{v}, \mathbf{w}) = \sqrt{\frac{1}{N} \sum_{i=1}^N \|v_i - w_i\|^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N ((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)}.$$

In this work, $\vec{v} \equiv \vec{r}_0$ where \vec{r}_0 is the set of Cartesian coordinates of each atom of the protein in the initial structure and $\vec{w} \equiv \vec{w}_t$ is the set of Cartesian coordinates of each atom of the protein at time instant t . In general, an increase of the RMSD indicates that the protein moves to a conformation different from the initial structure and thus suggests an incomplete sampling or a conformational change. In this work, the RMSD will be used during the analysis of the MD simulations, to evaluate the stability of the protein models.

3.6 Coarse Grained Elastic Network Models

Normal Mode Analysis (NMA) provides information on the equilibrium modes accessible to a system. In the 1990s it has become a tool widely used for exploring functional motions. In this section, a brief description of two NMA methods is introduced, *i.e.* the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM).

3.6.1 Gaussian Network Models

A GNM is a representation of a biological macromolecule as an elastic mass-and-spring network to study, understand, and characterize the mechanical aspects of its long-time large-scale dynamics [16]. The GNM was proposed by Bahar et al. in 1997 [197]. Detailed description the method can be found in [16]. Here only the main features are discussed. In GNM, the structure of a protein is represented by a network of N nodes identified by the C_α carbons. The pairs of nodes within a cutoff distance R_c ($\sim 7 \text{ \AA}$) are assumed to be connected by uniform springs of force constant γ , representative of the interactions that stabilize the protein. For nodes i and j ,

we define the equilibrium position vectors, R_i^0 and R_j^0 ; the equilibrium distance vector, R_{ij}^0 ; the instantaneous fluctuation vectors, ΔR_i and ΔR_j ; and instantaneous distance vector, R_{ij} . These quantities are shown in **Fig 3.4**.

The difference between equilibrium position vector and instantaneous position vector of residue i gives the instantaneous fluctuation vector, $\Delta R_i = R_i - R_i^0$. Hence, the instantaneous fluctuation vector between nodes i and j is expressed as $\Delta R_{ij} = \Delta R_j - \Delta R_i = R_{ij} - R_{ij}^0$ (see **Fig 3.4**). The potential energy of the network in terms of ΔR_i is

$$V_{GNM} = \frac{\gamma}{2} \left[\sum_{i,j}^N (\Delta R_j - \Delta R_i)^2 \right] = \frac{\gamma}{2} \left[\sum_{i,j}^N \Delta R_i \Gamma_{ij} \Delta R_j \right] \quad (3.83)$$

where γ is a force constant uniform for all springs and Γ_{ij} is the ij^{th} element of the Kirchhoff (or connectivity) matrix of inter-residue contacts, Γ , defined by

$$\Gamma_{ij} = \begin{cases} -1, & i \neq j \text{ and } R_{ij} \leq R_c \\ 0, & i \neq j \text{ and } R_{ij} > R_c \\ -\sum_j \Gamma_{ij} & \text{for } i = j \end{cases} \quad (3.84)$$

The inverse of Γ provides information on the mean-square (MS) fluctuations $\langle (\Delta R_i)^2 \rangle$ of residues (diagonal elements) and their cross correlations (off diagonal elements). The major utility of the GNM is the decomposition of the fluctuation spectrum into a set of normal modes. The contribution of the K^{th} nonzero mode is given by [197]:

$$[(\Delta R_i)^2]_K = (3k_B T / \gamma) \lambda_K^{-1} [\mathbf{u}_K]_i^2 \quad (3.85)$$

where \mathbf{u}_K and λ_k are the K^{th} eigenvector and eigenvalue of Γ , respectively.

3.6.2 Anisotropic Network Models

To characterize the directions of motions, the Anisotropic Network Model (ANM) can be adopted. The ANM was introduced in 2000 [198,199]. It represents the biological macromolecule as an elastic mass-and-spring C_α network, as GNM, but it also accounts explicitly for directionality. In this method, the Γ matrix of the GNM is replaced by the $3N \times 3N$ Hessian matrix H , with the second derivatives of the ANM potential V :

$$H = \begin{bmatrix} \frac{\partial V}{\partial x_i x_j} & \frac{\partial V}{\partial x_i y_j} & \frac{\partial V}{\partial x_i z_j} \\ \frac{\partial V}{\partial y_i x_j} & \frac{\partial V}{\partial y_i y_j} & \frac{\partial V}{\partial y_i z_j} \\ \frac{\partial V}{\partial z_i x_j} & \frac{\partial V}{\partial z_i y_j} & \frac{\partial V}{\partial z_i z_j} \end{bmatrix} \quad (3.86)$$

where

$$V = (\gamma/2) \sum_i \sum_{j>i} H_{ij} (R_{ij} - R_{ij}^0)^2 \quad (3.87)$$

always with respect to α carbon positions (**Fig 3.4**).

3.7 Basics of Parallel Calculus for Molecular Dynamics Simulations

As already stressed in the previous sections, the time scale accessible to MD simulations is strongly restricted by the computing power available. Thus, many technologically and biochemically interesting systems which extend over time scales longer than microseconds are difficult to simulate adequately using standard MD codes running on conventional single

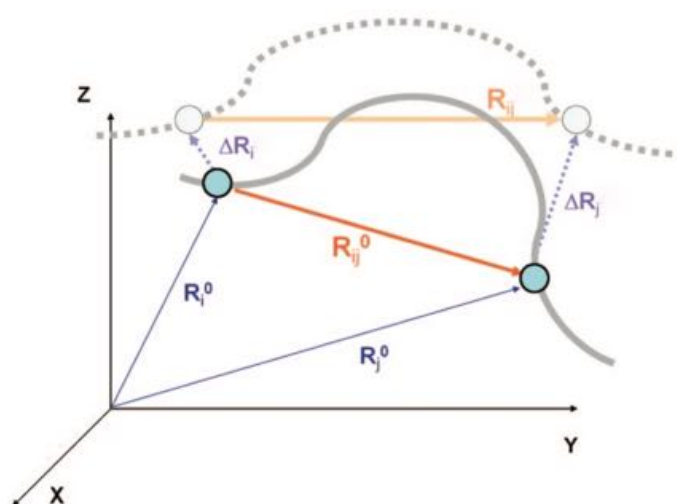


Fig 3.4. Schematic representation of nodes. Every node is connected to its spatial neighbors by uniform springs. Distance vector between two nodes, i and j , is shown by an arrow and labeled R_{ij} . Equilibrium positions of the i^{th} and j^{th} nodes, R_i^0 and R_j^0 , are shown in the Cartesian coordinate system. R_{ij}^0 is the equilibrium distance between the nodes. ΔR_i and ΔR_j , are the instantaneous fluctuation vectors. Reprinted from [16], <https://pubs.acs.org/doi/10.1021/cr900095e>. Copyright (2010), with permission from the American Chemical Society.

processor computers. Therefore, it is essential to develop new parallel algorithms capable of handling complex biochemical and technological systems [200]. To date, a number of strategies were proposed for mapping MD onto Multiple Instruction Multiple Data (MIMD) machines based on either Particle Decomposition (PD), *i.e.* assigning particles to home processors using their indices, or Domain Decomposition (DD) (also named Spatial Decomposition, SD), *i.e.* assigning particles to processors according to their position in space [201–207]. In the DD strategy, used in the NAMD code, each Processor Element (PE) is responsible for particles residing in a particular sub-volume of the MD box and each processor accumulates the total forces on the particles in its domain and integrates their equations of motion. Therefore, only a small fragment of the system has to be simulated in each processor and the most efficient division is not based in the list of particles, but in their position in the space. The most popular simulation codes have long been compatible with the Messaging Passing Interface (MPI). When a large number of computer cores can be used simultaneously, MPI can greatly reduce the computation time.

The main devices used in parallel computing can be grouped as follows:

- Accelerators such as Graphic Processing Units (GPUs). Originally designed to handle computer graphics, GPUs have evolved into general-purpose, fully programmable, high performance processors and represent a major technical improvement to perform atomistic MD. GPUs are mainly produce by two companies: the NVIDIA Corporation and the Advanced Micro Devices (AMD).
- Advanced processors such as the Xeon Phi manycore processors designed and made entirely by Intel. They are intended for use in supercomputers, servers, and high-end workstations.

NAMD uses C^{++} to implement the design DD ideas in a modular and efficient manner.

Chapter 4

Materials

Nil satis est, inquit, quia tanti quantum habeas sis.

Quintus Horatius Flaccus, *Saturae*.

4.1 Computational Protocols

In this section, all the key protocols and methodological approaches adopted are reported to allow adequate replication of the findings. In particular,

- **Subsection 4.1.1** describes the preparation of the protein structures,
- **Subsection 4.1.2** introduces the elements of the unbiased MD simulations for the Cldn15 paracellular channel models,
- **Subsection 4.1.3** shows all the details of the biased MD simulations for PMF and kinetic calculations for the Cldn15 single paracellular channel model,
- **Subsection 4.1.4** illustrates the details of the Elastic Network Models, used to study the dynamics of the putative Cld5 Suzuki-like models.

For all the protocols, as a starting point, I used the crystal structure of the single mCldn15 protomer (PDB ID: 4P79) and the PDB file provided by Suzuki et al. [1], which is pictured in **Fig 4.1**.

4.1.1 Preparation of Cldn Monomers

Preparation of Cldn15 structures. The available crystal structure of Cldn15 (PDB ID: 4P79 [40]) requires some additions in order to be used in molecular simulations. Specifically, both the monomer crystal structure and the assembled protomers in the TJ model of Suzuki et al. miss few extracellular residues: the crystal lacks eight residues (34 to 41) in ECL1, while the Cldn15 protomer in the TJ model is missing nine residues (34-42) in ECL1 and two (149 and 150) in ECL2.

Our strategy comprises the preparation of two distinct Cldn15 protomers. The first one, named *Model1*, is a fair replica of the crystallographic structure with *in silico* addition of the missing residues (34-41), modeled by the highest ranked conformation from the loop-closure modeling RCD+ software [208, 209], which uses an *ab initio* algorithm. The full protein structure was further refined with Chimera tools [210], after changing Selenomethionines to Methionines, and using the Dunbrack rotamer library to design missing or partial side chains [211].

Model1 is suitable for all the MD simulations which involves single protomer of Cldn15 and for the assembly of the single-pore system which takes into account only the second *cis* (*face-to-face*) interface between protomers.

For the double-pore simulation, an alternative configuration of the loops of Cldn15 is required. Indeed, the complete architecture of the model shows head-to-head protomers packing in the regions of missing ECL1 and ECL2 residues and, when these amino-acids are directly inserted,

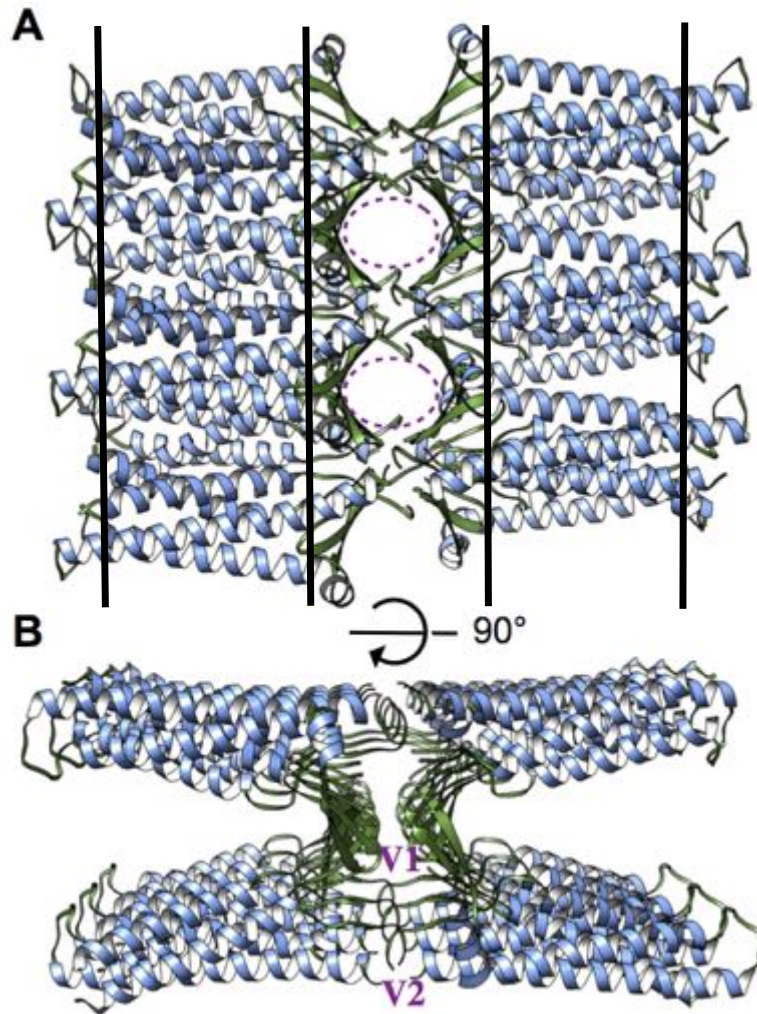


Fig 4.1. Model of paracellular TJ channels proposed by Suzuki et al. The model of Cldn-based paracellular channels proposed by Suzuki et al., represented as ribbons and viewed from the apical (**A**) and the lateral side (**B**). The pore regions are marked with purple circles, and the V1 and V2 regions (see text) are indicated. Horizontal lines in panel **A** indicate the membrane boundaries of the two cells. Adapted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

steric clashes are created. To overcome this problem, an alternative structure of Cldn15 (named *Model2*) was prepared, where, using again RCD+, new coordinates for the residues of ECL1 (34-41) and an extended segment in ECL2 (residues 148-154) were generated, with non-overlapping extracellular protein regions. The structure was again further refined using Chimera.

The difference between the two final protomer models is highlighted in **Fig 4.2**. It is important to stress that the different arrangements of ECL1 and ECL2 in the two models did not alter the crystallographic orientation of the side chains in the linear *cis* interface, since residues F146, F147 and L158 are not part of the reconstructed regions.

As discussed below, these protomers were used to build two refined versions of the Cldn15 paracellular channel, as illustrated in **Fig 4.3**.

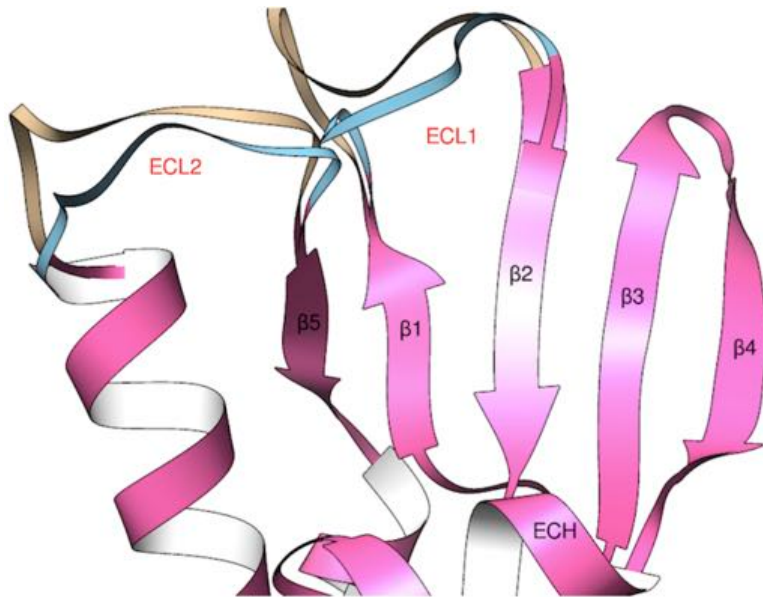


Fig 4.2. Modeling of Cldn15 loops. Superposition of the Cldn15 structures named *Model1*, used for the monomer and the single-pore, and *Model2* for the the double-pore. The identical folding is pictured in pink, while the two different conformations of the loops, ECL1 and ECL2, are colored in brown for *Model1* and cyan for *Model2*, respectively. Secondary structure elements are labeled in black and red. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

Preparation of Cldn5 structures. Multiple monomer models of Cldn5 were built using *homology modeling*.

Homology modeling, also known as *comparative modeling*, refers to constructing an atomic-resolution model of a *target* protein from its amino-acid sequence (named the *query sequence*) and an experimental three-dimensional structure of a related homologous protein, *i.e.* the *template*.

This approach is based on the identification of at least one template, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence.

Because protein structures are more conserved than DNA sequences, detectable levels of sequence similarity usually imply significant structural similarity.

Both the templates used for the modelig (mCldn15 and hCldn4) share ~ 35% sequence identity with the hCldn5 target.

In particular, two softwares were used for structural modeling:

1. The MODELLER suite [212–215]. It is based on a method, termed *satisfaction of spatial restraints*, by which a set of geometrical criteria are implemented to create a probability density function for the location of each protein atom. The UCSF Chimera software provides a Graphical User Interface (GUI) to running Modeller. Here the crystal structure of mCldn15 (PDB ID 4P79) was used and 10 Cldn5 models, diversified for the conformation of the incomplete loop ECL1 in the template, were produced. Multiple sequence alignment of Cldn proteins was performed using Clustal Omega [216] and the alignment was visualized using UCSF Chimera.
2. The Swiss Model webserver [217] <http://swissmodel.expasy.org/>. After the selection of a template and the sequence alignment, SWISS-MODEL implements a rigid fragment assembly approach: deviations in the protein structure geometry are regularized in the last modeling step by steepest descent energy minimization with the GROMOS96 force field. Here, to investigate an additional conformation of the extracellular loops, the crystal structure of Cldn4 (PDB ID: 5B2G) was used.

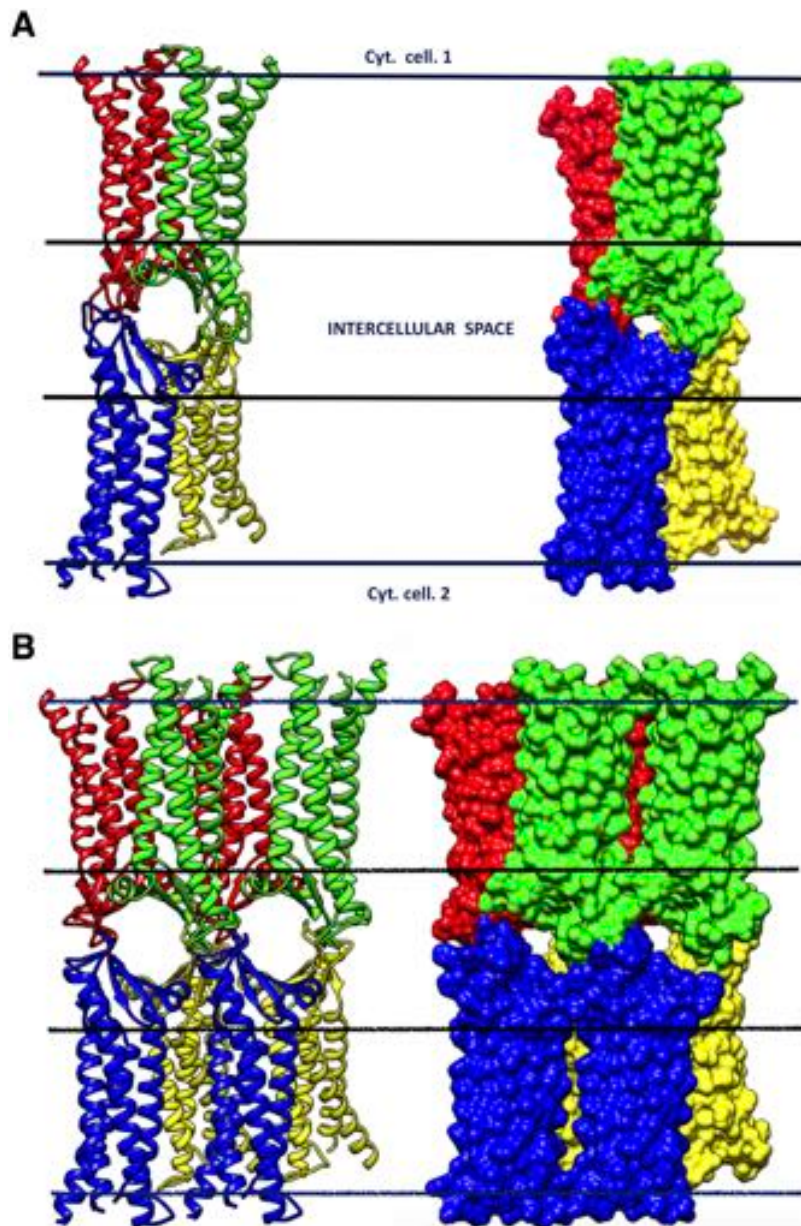


Fig 4.3. The single and double-pore channel models. **A**, structure of the single-pore system viewed from the apical side, in ribbon (left) and surface representation (right). The complex is formed by the trans interaction of two *cis*- dimers (red-green, P1-P2 dimers; and blue-yellow, P3-P4 dimers). **B**, Structure of the double-pore system represented and viewed as in panel **A**. Horizontal lines indicate the membrane boundaries. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

To improve both the global topology and the local atomic geometry of the Cldn5 models, a further refinement step was adopted. In this part of the protocol, each Cldn5 protomer was refined with the ModRefiner tool of the Zhang Lab [17], freely available at <http://zhanglab.ccmb.med.umich.edu/ModRefiner>. This software revises protein structures from C_{α} traces based on a two-step, atomic-level energy minimization. The main-chain structures are first constructed from initial C_{α} traces and the side-chain rotamers are then refined together with

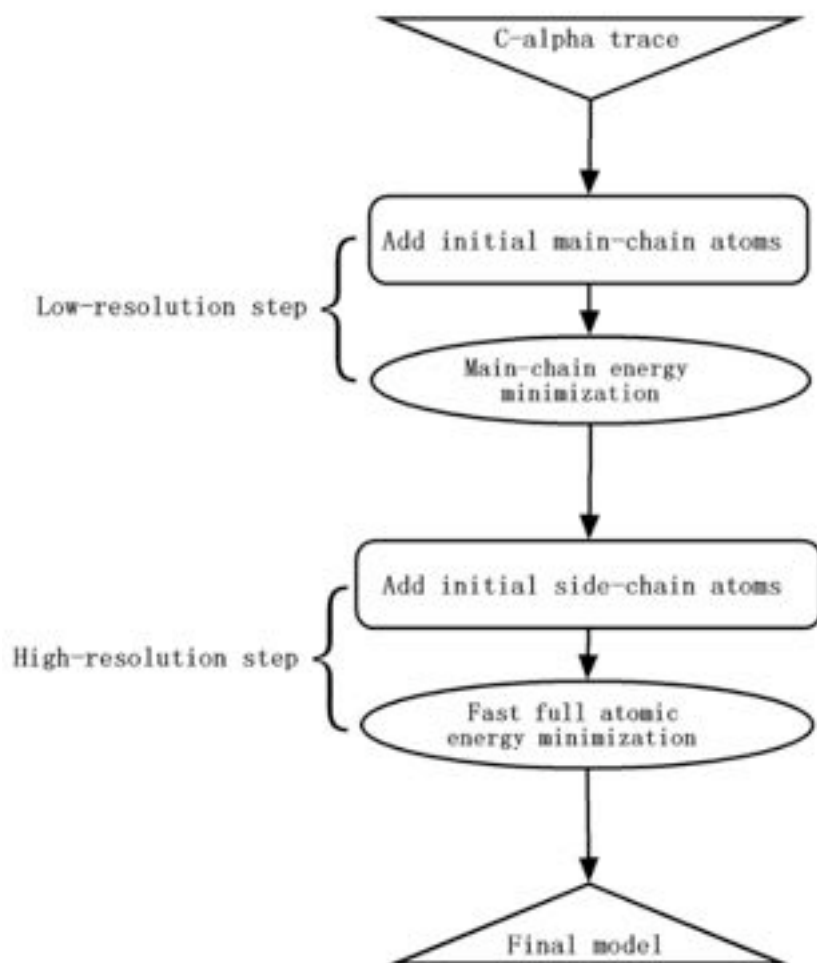


Fig 4.4. Flow chart of the ModRefiner algorithm. The two step ModRefiner procedure. If backbone and side-chain atoms are already included in the configuration of the model, the algorithm starts with energy minimization. Reprinted from [17], [https://www.cell.com/biophysj/references/S0006-3495\(11\)01245-8](https://www.cell.com/biophysj/references/S0006-3495(11)01245-8). Copyright (2011), with permission from Elsevier.

the backbone atoms with the use of Monte Carlo simulations with a composite physics- and knowledge-based force field, to obtain more accurate side-chain positions, better hydrogen-bonding networks, and fewer atomic overlaps. Then, using these protomers, three Cldn5 tetramers (with the configuration introduced in **Fig 4.3**, panel **A**) were prepared with Chimera tools and energy minimized *in vacuo*, using the GalaxyRefineComplex utility [218, 219].

4.1.2 Unbiased MD Simulations

Unbiased MD simulation of the Cldn15 monomer. The structure described above as *Model1* was oriented using the PPM Server [220] in a hexagonal 1-palmitoyl-2-oleoyl-SN-glycero-3-phosphocholine (POPC) bilayer (253 molecules) and solvated with three point (TIP3P) water molecules (17540) and 150 mM NaCl solution neutralizing the net charge of the protein. The conserved disulphide bridge in ECL1 (C52-C62) was maintained in the structure. The membrane builder application of the CHARMM-GUI server [221] was used for the preparation of all the input files. Hexagonal PBCs were used to replicate the system and remove box surface effects. After an initial energy minimization, the system was heated to 310 K and then simulated using positional restraints on the protein atoms for 2 ns in the *NPT* ensemble

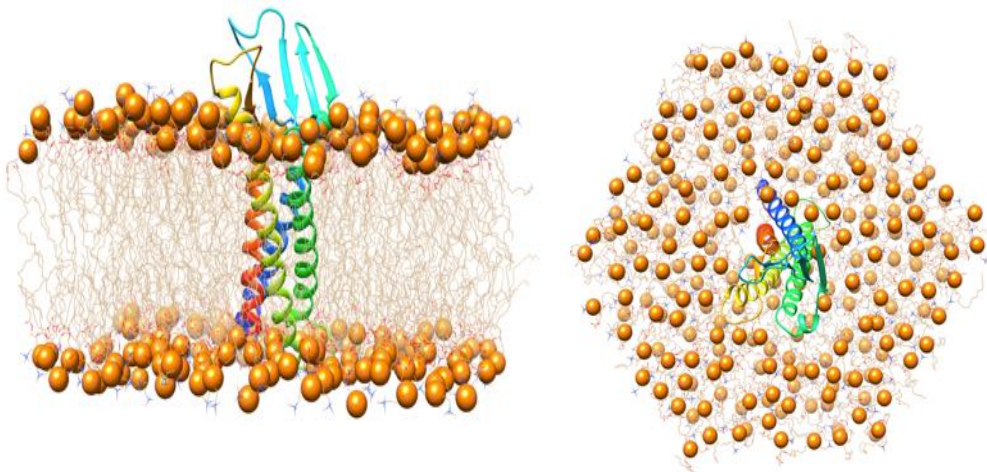


Fig 4.5. The equilibrated Cldn15 monomer. Viewed from the transmembrane domain (left) and from the extracellular environment (right). Cldn15 monomer (rainbow cartoon) is embedded in a POPC bilayer, shown as wire structures. Phosphorus atoms of the membrane are introduced as spheres. Water molecules and ions are not shown for clarity. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

at 310 K and 1 bar. Successively, a 130 ns unrestrained MD simulation was produced in the same ensemble at the same conditions maintained by a Langevin thermostat and Nosé-Hoover Langevin piston pressure control [80,81]. The NAMD 2.12 program [13] with the CHARMM36 force field [65,66] was used. During the trajectory, the length of the cell basis vector box was stationary around $99.0 \times 99.0 \times 103.0$ Å. Water molecules were described with the three point TIP3P model [68] and ions with the parameters included in the force field [222]. Electrostatic and van der Waals interactions were calculated with a cutoff of 12 Å and the application of a smoothing decay starting to take effect at 10 Å. Long range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) algorithm [64]. All covalent bonds involving hydrogen atoms (except those of water molecules) were kept fixed using SHAKE [71], and those in water molecules using SETTLE [73]. A time-step of 2 fs was employed. To ensure maximum accuracy, electrostatic and van der Waals interactions were computed at each simulation step. The MD trajectory was visualized and analyzed to validate the structural stability of the Cldn15 protein using Chimera and VMD and associated plugins [223].

Unbiased MD simulations of the single-pore model. To obtain the starting configuration of the single-pore system, four replicas of *Model1* in the same conformation provided by Suzuki et al. were assembled, by using the Matchmaker Chimera tool for structure superimposition. Then, the tetramer was energy minimized *in vacuo*, using the GalaxyRefineComplex utility [218,219], preserving the conformation of the transmembrane domains and relaxing the extracellular domains. The resulting conformation is shown in **Fig 4.3**, panel **A**.

After the described procedure, the side chains of D55 residues on ECL1 of each protein point towards the interior of the channel, consistent with their position in the crystal structure of the protomer and their known regulation of Cldn15 selectivity for cations. Subsequently, the proteins were embedded in a double lipid bilayer to reproduce the TJ paracellular environment. Water molecules were added in the intercellular space and the two cytosolic regions, and counterions were inserted to neutralize the system charge. The configurations of the two bilayers and of the solvent were produced with the CHARMM-GUI server and suitably assembled with the tetramer via Chimera tools, eliminating unfavorable protein-lipid and protein-water contacts. The advices of [224] were carefully followed to optimally combine CHARMM and NAMD input files, for example concerning the orientation of the periodic box.

The final configuration includes four Cldn15 structures (two replicas of the monomer in *face-to-*

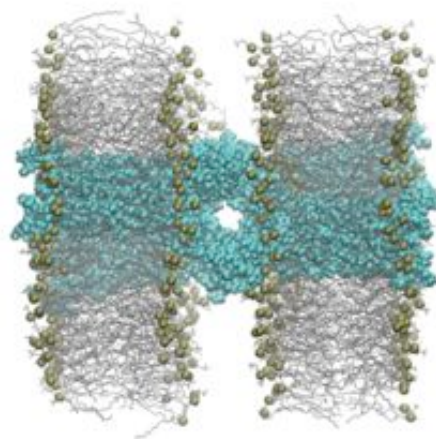


Fig 4.6. Conformation of the equilibrated single-pore system. The structure of the single pore, after the equilibration protocol. Protomers are shown as cyan spheres. Each *cis* dimer is embedded in a hexagonal POPC bilayer, shown as wire structures with phosphorus atoms as spheres. Solvent molecules are not shown for clarity.

face cis- interaction per side of the junction), ~ 280 POPC lipids in each bilayer, TIP3P water molecules and counterions, for a system of ~ 165000 atoms.

VMD tools were used to prepare the topology of the whole system with the CHARMM36 force field, always including the disulphide bridge of ECL1, for each protomer. All other simulation settings were the same of those used for the monomer.

After an initial energy minimization, the system was heated to 310 K and then simulated for 2 ns in the *NPT* ensemble ($P=1$ bar) with all protein atoms fixed. The system was further equilibrated for 1 ns with the C_{α} atoms restrained, and finally for 1 ns with only the α helices C_{α} atoms restrained. The equilibrated cell box is an hexagonal prism with the length of the cell basis vectors at $\sim 106 \times 106 \times 162.0 \text{ \AA}^3$. In **Fig 4.6**, we show a conformation of the single-pore system after the equilibration procedure.

After minimization and restrained equilibration, all restraints were removed and a free MD simulation of the system was produced for 260 ns in the *NPT* ensemble ($T=310$ K, $P=1$ bar). The final 250 ns were used for analysis. The MD trajectory was visualized and analyzed to validate the structural stability of the system using Chimera and VMD.

For the same single-pore system, we also performed a replica simulation to monitor the effect of a different equilibration stage. In this replica, after energy minimization and heating to 310 K, a longer equilibration phase was produced: 10 ns with all proteins fixed, 10 ns with only C_{α} atoms restrained, and 1 ns with only the C_{α} atoms of the transmembrane α -helices restrained, always in *NPT* at 1 bar. Then, the system was simulated with all restraints removed for 35 ns. No differences were found in the structure obtained with respect to that from the protocol described above.

MD simulation of the double-pore model. The structure of *Model2* was used for the MD simulations of the octameric double-pore model, which includes both the linear and *face-to-face cis* interfaces introduced by Suzuki et al. and described above. Eight replicas of *Model2* were superimposed to the pdb file provided by Suzuki et al. by using the Matchmaker Chimera tool.

The octamer was energy minimized *in vacuo* to relax the extracellular domains, using the GalaxyRefineComplex utility, and restraining the conformation of the transmembrane domains. The resulting conformation (see **Fig 4.3** panel **B**) was embedded in a double hexagonal bilayer and solvent was added. Globally, the system includes 8 protomers (four replica for face of the junction), ~ 455 POPC lipids for each bilayer, TIP3P water molecules and counterions, for a system of ~ 338000 atoms.

An extended equilibration was performed to relax the large structure. After energy minimization and heating to 310 K, the system was submitted to 30 ns total of equilibration in *NPT* ($P=1$ bar) with positional restraints on protein atoms: 6 ns with restraints on the whole proteins, 6

ns with restraints on backbone atoms only, 12 ns with restraints on the C α atoms, and finally 6 ns for the C α atoms of the α helices only, to allow relaxation of the paracellular region.

Evaluation of hydrogen bonds survival ratio, hydrophobic interactions and pore size analysis. Hydrogen bonds (HBs) across the complex interfaces were searched using the VMD software. An HB was considered formed when the donor-acceptor distance and bonding angle were less than 4.0 Å and 30°, respectively. The survival ratio of HBs was defined as the percentage of time the HB was found present along the trajectory. Interactions between hydrophobic residues were studied with standard TCL scripts, introducing a 4.0 Å cutoff. To study the size of the paracellular channel, we employed the HOLE program [225,226], which calculates the radius of a protein pore along a given axis by determining the maximum size for a spherical probe that can fit with the van der Waals radii of the atoms. We used a 5 Å threshold for the pore radius which is sufficient to explore the narrow part of the pore. Representative structures spaced by about 10 ns along the simulated trajectory were selected and analyzed. The nomenclature and the color legend for protomers are introduced in **Fig 4.7**.

4.1.3 General protocols for PMF calculations

The configuration of the Cldn15 channel and general MD parameters for ionic diffusion. A configuration of the full single pore system was extracted from the extended, unbiased MD simulation at equilibrium, and it was used as starting conformation for all the biased MD simulations. To optimize the use of parallel computational resources, for all the simulations, starting conformations were prepared with the ion in the proper position along the pore axis. For each configuration, the tested ion was swapped with equilibrated water molecules. The ion was then kept fixed, together with the proteins and the lipids, and ~ 1 ns of MD was performed for each configuration to allow equilibration of surrounding water molecules. Furthermore, for all the SSw mean force calculations, a preliminary run with 100 steps of energy minimization and over than 50 ps of equilibration was performed, where the test ion was held fixed to allow water to equilibrate around it. All trajectories were visualized and analyzed using UCSF Chimera [210] and VMD [223].

US simulations and Milestoning MD simulations for ionic diffusion. 1D-US simulations with harmonic biases were performed with the Collective Variable Interfaces (Colvars library) [227], using the coordinate of the ion channel along the axis as CV. We performed US simulations with the restraints of Eq. (3.54) and $\hat{\kappa} = 2$ kcal/(mol \cdot Å²) with 5 ns for each window. We employed a total of 65 umbrella windows with a uniform spacing of 1 Å for each ion. The first 0.5 ns of each window were not used for analysis.

The WHAM code from the Grossfield group, available at <http://membrane.urmc.rochester.edu/content/wham>, was used to reconstruct the global 1D PMF. For WHAM analysis, we used 600 bins, and a tolerance of 0.0001. Furthermore, the statistical uncertainty at each bin was estimated with bootstrapping, using 100 bootstrap trials for each PMF. For Na⁺ and Cl⁻ two simulations were performed for ~ 1.3 μ s, while one simulation was produced for K⁺ (~ 325 ns). Overall the US calculations are based on a cumulative production of ~ 1.6 μ s.

Milestoning MD simulations were performed to evaluate the kinetic properties of the Na⁺, K⁺, Cl⁻ ions. The Colvar plugin was also used to apply the half-harmonic restraints of Milestoning MD simulations with the same CV of the US simulations. For each ion the length of the channel was mapped with 23 cells. Pseudoharmonic restraining potentials were applied (Eq. (3.73)) with $\tilde{\kappa} = 100$ kcal/(mol \cdot Å²). The simulations lasted for 3 ns per cell for Na⁺ and K⁺ and 6 ns per cell for Cl⁻. The first 100 ps of each run was not used for analysis. Overall these calculations are based on a cumulative production of ~ 276 ns.

To avoid any rigid body rotational or translational displacement of the channel, in all these simulations, further harmonic restraints were applied to the C α atoms of the residues 11, 20, 80, 94, 122, 133, 167 and 177 (all in the TM domain).

SSw 3D PMF reconstruction for cations diffusion. TAMM Eq.s (3.75) were implemented with the NAMD tclForces module, using as CVs the three Cartesian coordinates of the center of mass (COM) of the single ion. To fill the space within the pore, ~ 60 configurations

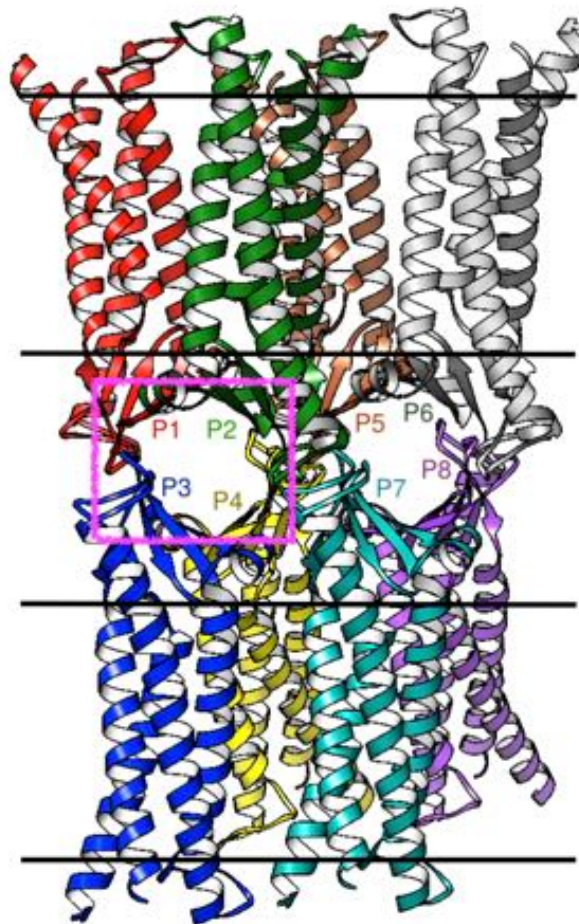


Fig 4.7. Legend of the protomers. The nomenclature and color legend for each protomer of the Suzuki model is introduced. For single pore simulations the system is restricted to the four protomers P1-P4, included in the pink rectangle. Horizontal lines indicate the membrane boundaries. Adapted and reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

were prepared with a Na^+ ion placed along the pore axis with a uniform spacing of 1 \AA from each other. These conformations were minimized and then equilibrated for more than 1 ns , with the addition of harmonic restraints to the proteins, the membranes and the Na^+ ion. Then, for each configuration, a short 1 ns TAMM simulation was produced, using the parameters $\kappa = 200 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$, $\bar{\beta}^{-1} = 5 \text{ kcal}/\text{mol}$, $\bar{\gamma} = 200 \text{ ps}^{-1}$. All TAMM trajectories were combined to obtain a single sweep of the ion COM coordinates and 177 centers in the CV space were selected. The mean forces at the centers were computed with a first battery of 600 ps for each center using Eq.s (3.76-3.77) and $\bar{\kappa} = 200 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$. The duration of 600 ps at a fixed center, with $\bar{\kappa} = 200 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$ for Eq.s (3.76-3.77), defines a protocol which we used for the other ions. Overall the SSw calculations for Na^+ are based on a cumulative production of $\sim 106 \text{ ns}$. To study the K^+ diffusion, the previous ~ 60 configurations with the Na^+ ion placed along the axis of the pore were used. We changed the Na^+ counterions of the system with K^+ ions and

then ~ 95 TAMD simulation of 1 ns were produced (some of them repeated), with parameters $\kappa = 200$ kcal/(mol \cdot \AA^2), $\bar{\beta}^{-1} = 5$ or 7 kcal/mol, $\bar{\gamma} = 200$ ps $^{-1}$. Here, 177 centers were selected. The mean forces at the centers were computed with a 600 ps simulation with $\bar{\kappa} = 200$ kcal/(mol \cdot \AA^2). Overall the SSw calculations for K^+ are based on a cumulative production of ~ 200 ns. To study the diffusion of Li^+ , Ru^+ and Cs^+ , we used the 177 centers obtained from the Na^+ TAMD simulations. For these configurations, the Na^+ counterions were replaced with the associate monovalent ions. Then, the calculation of the mean forces were produced for 600 ps with $\bar{\kappa} = 200$ kcal/(mol \cdot \AA^2). Overall the SSw calculations for each of these ions are based on a cumulative production of ~ 106 ns, for a total production of ~ 315 ns.

SSw 3D PMF reconstruction for α -D-Glucose diffusion. The atomic coordinates of ADG were obtained from the crystal structure of the α -,1,3 galactosyltransferase-ADG complex (PDB ID: 1GWW). Hydrogens were added using the Glycan Reader tool [228] in the CHARMM-GUI [229, 230]. The parameters of ADG were obtained from the CHARMM36 force field. The molecule was positioned at the mouth of the channel by replacing 12 water molecules and then pulled through the pore. To this aim, a harmonic force of 25 kcal/(mol \cdot \AA^2) was added to the coordinates of the ADG COM, slowly changing the COM reference positions z_0 , to dock the molecule inside the channel and obtain starting configurations for TAMD. During this procedure, harmonic restraints were also added to the proteins and membranes to preserve the starting configuration of the system. Finally, multiple TAMD runs were produced for a cumulative time production of 42 ns, with the parameters $\kappa = 200$ kcal/(mol \cdot \AA^2), $\bar{\beta}^{-1} = 5$ or 7 kcal/mol, $\bar{\gamma} = 100$ ps $^{-1}$. All TAMD trajectories were combined to obtain 180 centers in the CV space. Then for each center, a simulation of 1500 ps was produced for the evaluation of the mean forces (Eq. 3.77) with $\bar{\kappa} = 600$ kcal/(mol \cdot \AA^2). Overall the SSw calculations for ADG are based on a cumulative production of ~ 270 ns.

4.1.4 Protocols for Elastic Network Models

The DynOmics Portal <http://enm.pitt.edu/Tutorial.php> provides information on the equilibrium dynamics of biomolecular systems. At the core of DynOmics Portal are two elastic network models: the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM). Distinctive functionalities of DynOmics Portal include the estimation of key sites involved in collective mechanics, allostery, and mapping of the Coarse Grained (CG) conformations driven by collective modes to their full-atomic representations. The Cldn5 models were used as input files. A cutoff distances of 7 \AA was used for GNM and 15 \AA for ANM. A spring constant for all connections was taken as unity (*i.e.*, 1 kcal/(mol \cdot \AA^2)).

Here, GNM and ANM are used to predict potential functional sites (PFSs). Using the GNM slowest two modes, the algorithm, referred to as the *minima-screening algorithm* [231], calculates the PFSs *via* inter-residue contact topology exclusively. The PFSs are the hinge residues that control the slowest modes.

4.2 Computational Resources

For all the simulations, the computational resources provided by the MARCONI supercomputer of the CINECA consortium <https://www.cineca.it> were used. MARCONI is a Tier-0 system, based on the Lenovo NeXtScale platform, formed by three partitions.

- The A1 partition. Based on Intel Xeon processor E5-2600 v4 product family (Broadwell, KNC) with a computational power of ~ 1 Pflop/s. In particular, the A1 system was used in almost all the simulations, because it is able to produce good performance for NAMD simulations.
- The A2 partition. By the end of 2016 a section is added, equipped with the next-generation of the Intel Xeon Phi product family (Knights Landing, KNL), based on a many-core architecture.
- The A3 partition. In August 2017, a third partition is added, based on Intel Xeon 8160 (SkyLake).

Chapter 5

Results and Discussion

Eppure resta
che qualcosa è accaduto, forse un niente
che è tutto.

Eugenio Montale, Xenia (Satura).

5.1 Structural Features of the Refined Cldn15 Paracellular Channel

5.1.1 Analysis of the Standard MD Simulations

In this subsection the results produced with the standard MD simulations are introduced. MD simulations were analyzed, to detect the general features of the structure and its ability to fit the main available experimental results.

Analysis of the MD simulation of the Cldn monomer. To verify the stability of the Cldn15 monomer structure and to determine membrane-driven effects on the Cldn conformation, the protein was embedded in a 1-Palmitoyl-2-oleoyl-SN-glycero-3-phosphocholine (POPC) bilayer based on the calculated hydrophobic thickness (Ht) and respective orientation angle of the protein. POPC lipids represent one of the main components of different membranes and are widely used in MD simulations to represent the lipid domain.

After preliminary equilibration, as illustrated in the Methods, all restraints on protein atoms were released and a 130 ns simulation was produced. In **Fig 5.1**, panel **A**, a comparison between the monomer structures in the extracellular region at the beginning and at the end of the trajectory is reported.

Overall, the extracellular domains did not change their structure, maintaining the β strand conformations. Relevant interactions observed in the crystal structure were preserved. In particular, R79 on TM2 establishes hydrogen bonds (HBs) with L48 on ECL1, for half the duration of the simulation. These interactions, which are illustrated in **Fig 5.1**, panel **B**, contribute the anchoring of the most internal segment of ECL1 to the protein core.

To observe the stability of the protein during all the frames of the simulation, the Root-Mean-Square Deviation (RMSD) of the atomic positions belonging to the backbone atoms was calculated from the initial *Model1* conformation. Results show a plateau at about 2 Å for the full protein backbone (**Fig 5.2**, panel **A**) and at about 3 and 2.5 Å for ECL1 and ECL2, respectively (**Fig 5.2**, panel **B**), highlighting the stability of the entire protein structure with limited conformational heterogeneity in the extracellular region.

Analysis of the MD simulations of the single-pore model. It is important to remark that for the study of the structural features and of the permeation events through the paracellular environment, the use of a single paracellular channel is sufficient, at least at a first stage of the investigation, in particular to reduce the computational cost of the simulations.

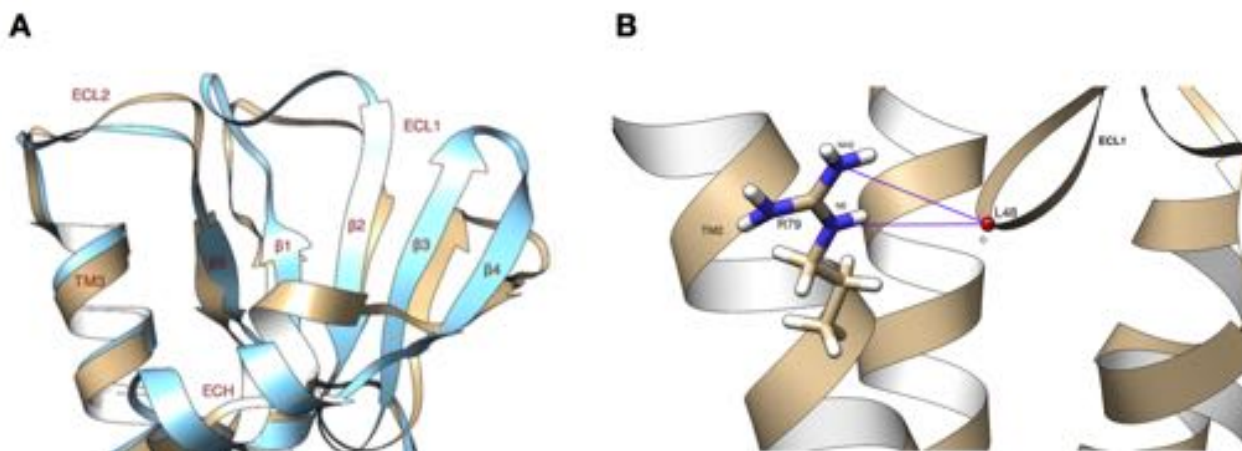


Fig 5.1. Structural features of the Cldn15 monomer. **A**, Superposition of the starting Cldn15 *Model1* monomer (cyan) and the final configuration from the MD run (brown). **B**, The side chain of R79 establishes two HBs with the main-chain carbonyl group of L48. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

To test and refine our Cldn15-based single pore model, we ran a 250 ns simulation. This MD run is another step to further refine the *in silico* loops, that are missing in the original version of the Suzuki model. During the MD simulation, no restraints were applied on either the protein or the membrane atoms.

The single-pore model remained extremely stable, with an average RMSD of 2 Å from the equilibrated initial structure for each protomer (**Fig 5.3**). The four chains are indicated with names P1 to P4 as reported in **Fig 4.7**. Similarly to what observed for the single protein, RMSD profiles of each protomer show good convergence (**Fig 5.3**, panel **A**), with the exception of P3 where the larger fluctuations are due to movement of ECL1 (see **Fig 5.4**, below).

During the simulation, a slight change was observed in the relative orientation of the protomers. This rearrangement can be observed in **Fig 5.3** panels **B** and **C**, where the structure of the tetramer at the beginning (pink ribbon) and at the end (orange ribbon) of the trajectory are compared. However, these limited readjustments do not alter the geometry of the channel, as evidenced below from the analysis of inter-protomer distances and interactions (see in particular the results illustrated in **Fig 5.7** and **Table 5.1**).

RMSD profiles of the extracellular loop regions (**Fig 5.4**) reveal the flexibility of these domains, which is generally similar to what observed for the monomer. One of the four protomers (P1) shows an increase of RMSD values for ECL1 between 75 and 100 ns, and a similar, although less marked, peak can also be observed for P4. However, all these deformations appear to be reversible and both protomers return rapidly to a conformation similar to the initial one.

The only permanent deviation from the starting configuration, is observed for the P3 ECL1,

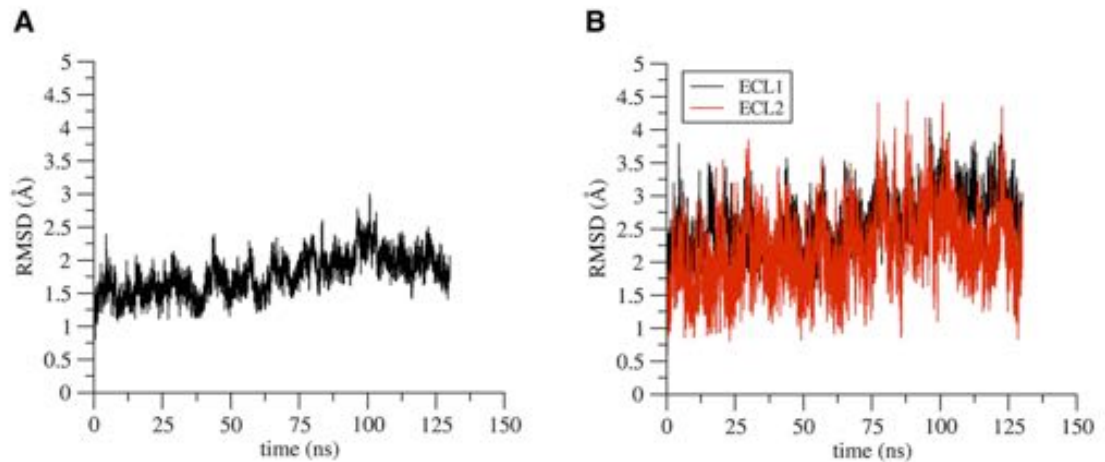


Fig 5.2. RMSD of Cldn15 monomer. RMSD values of the protein backbone (**A**), and of the backbone of the two extracellular loops, ECL1 and ECL2 (**B**) along the simulation of the Cldn15 monomer. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

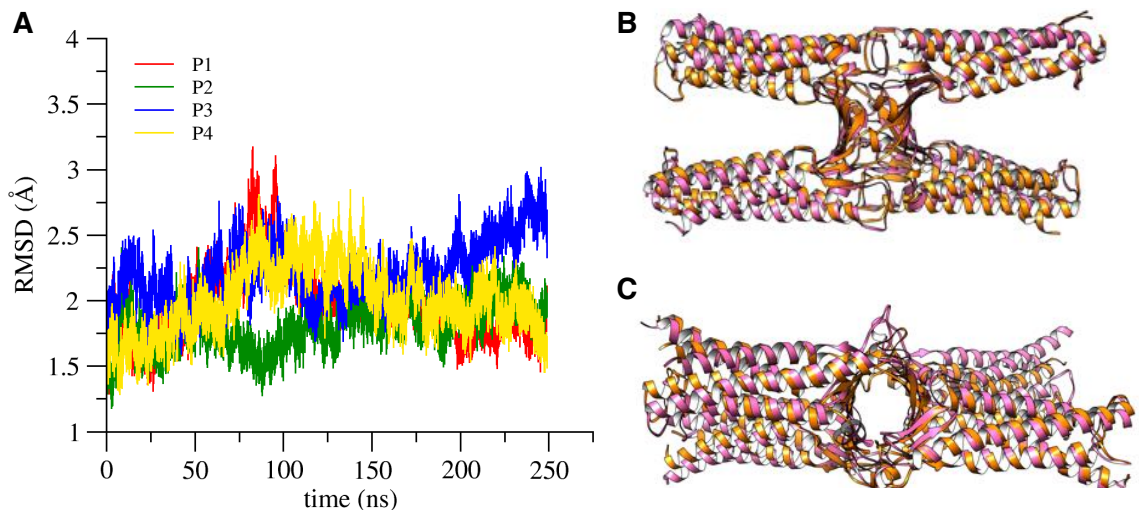


Fig 5.3. RMSD of the single-pore. **A**, RMSD values of the backbone atoms of each protomer in the single-pore simulation with respect to the initial model. **B** and **C**, superposition of the initial (pink ribbon) and final (orange ribbon) configurations of the channel, viewed from the lateral and apical sides, respectively. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

which adopts a conformation that is slightly different from the other ECLs, with its tip pointing towards the solvent, but also this modification does not affect the global structure of the pore, as evidenced below (see analysis of **Fig 5.5** and **Fig 5.6**).

Then we calculated the time evolution of the channel pore size. In the original paper of the Suzuki model [1], the determination of the pore size could not be performed accurately, because of the lack of residues in the intercellular space, and only an upper bound of 10 Å is reported. As a consequence, the paracellular pore should permeate ions and eventually restrict their

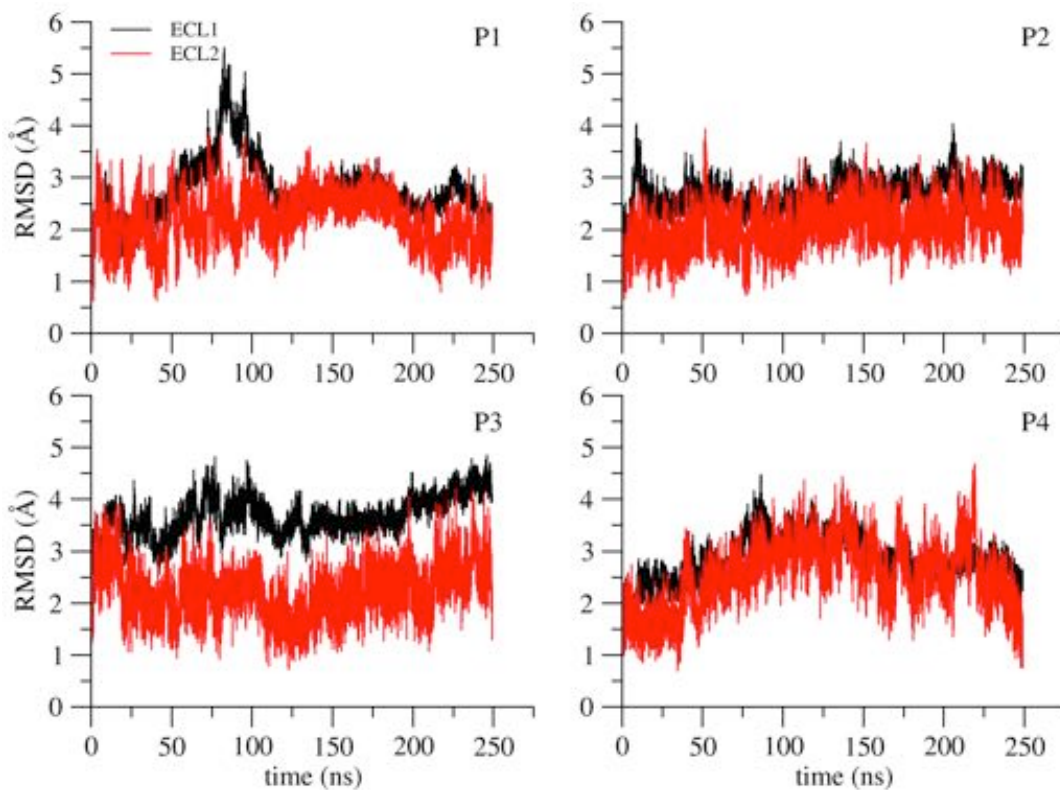


Fig 5.4. RMSD of the extracellular domains of the single-pore. RMSD values of the backbone of the extracellular loops ECL1 (black) and ECL2 (red) in the four protomers. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

diffusion depending on the composition of the internal cavity. This conclusion is consistent with previous experimental observations [232], where Claudin 2 (Cldn2) paracellular channels are described as narrow cavities with a diameter of 6.5 Å. These dimensions also fit the topology of many other biological channels, and the ion selectivity of the paracellular channels would depend on the specific Cldn subtypes that are considered. For our analysis, we focused on the region in the middle of the β -barrel structure, indicated as surface in **Fig 5.5**.

The minimal value of the pore radius was calculated using snapshots taken from the trajectory at approximately 10 ns from each other. The evolution of the minimal pore radius and a representative profile calculated at 130 ns (*i.e.* in the portion of the trajectory that shows less fluctuations of the minimal radius) are shown in **Fig 5.6, A and B**, respectively.

In **Fig 5.6 B**, the positions of the C_{α} atoms of the selectivity-determining residues D55 (pore forming, red bars) and D64 (pore lining, blue bars) in the four protomers (P1-P4) are indicated. As can be observed, they surround the region of maximal constriction of the paracellular pore.

The minimal radius is 2.0 Å in the assembled structure before equilibration, and it oscillates around 2.5-3 Å during the dynamics. Within the first 100 ns, a large fluctuation is observed between 4 and 1 Å, revealing a breathing motion of the pore. To represent these pore fluctuations visually, in **Fig 5.6** the initial tetramer conformation (**C**) is compared with three structures taken from the simulated trajectory at 30 ns (**D**), 80 ns (**E**) and 130 ns (**F**).

Despite the described fluctuations in pore width, the channel scaffold is quite stable, as can be deduced from visual inspection and from the time evolution of the cross-distances between the C_{α} atoms of the four pore-lining C52 residues, located in the middle of the β 3 strands, which are stationary around 21 Å, and show limited fluctuations (see **Fig 5.7**).

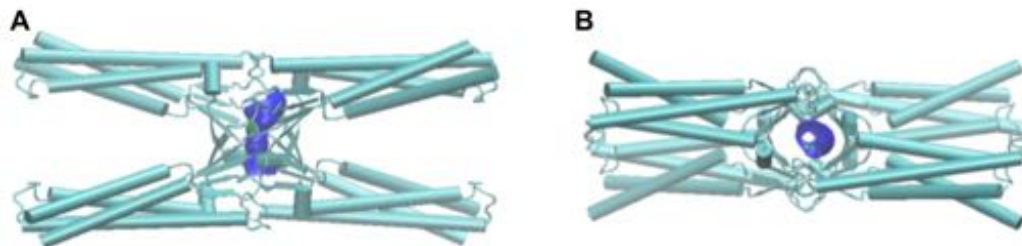


Fig 5.5. Pore cavity region of the single-pore. Representation of the pore cavity region (blue surface) of the paracellular channel (represented with the Tube style of the VMD software). Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

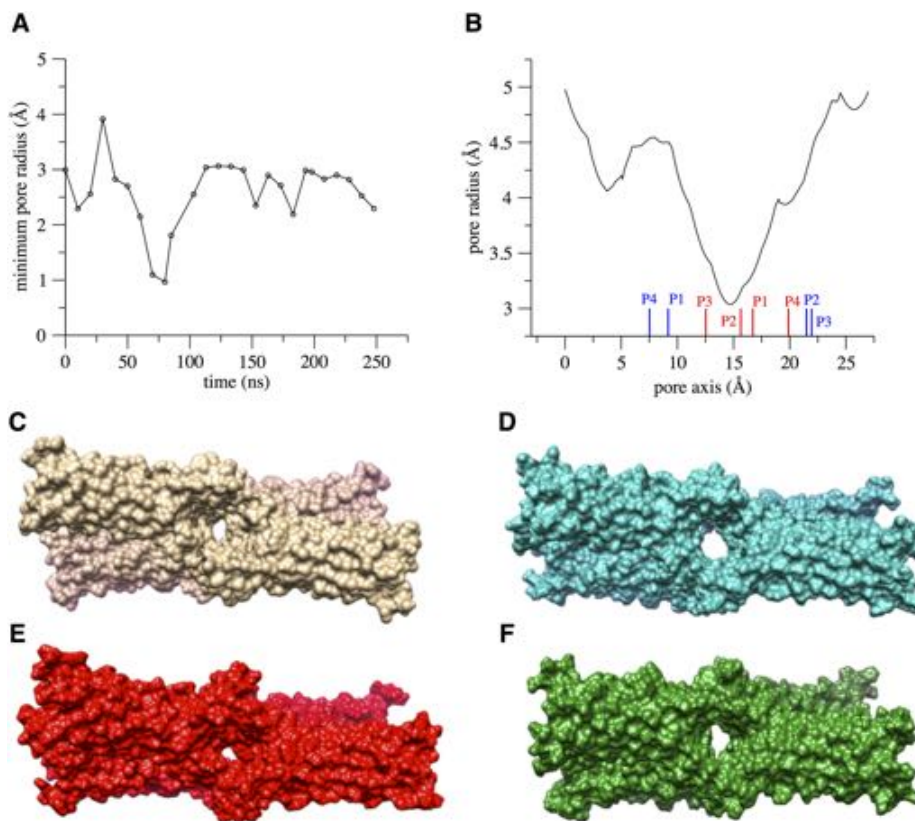


Fig 5.6. The pore size and structure in the single-pore. **A**, time evolution of the minimal pore radius along the single-pore simulation. **B**, the pore profile along the channel axis for the configuration taken at 130 ns, with red and blue bars indicating the C_{α} positions of D55 and D64 residues, respectively. **C**, The starting conformation of the single-pore and three snapshots extracted at 30 ns (**D**), 80 ns (**E**), and 130 ns (**F**). Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

In addition, from our simulations, we can also analyze accurately the inter-monomer interactions that characterize the model. In particular the following network of interactions, can be introduced:

1. the *face-to-face cis*- interactions for the dimers P1-P2, P3-P4;
2. different *trans*- interactions.

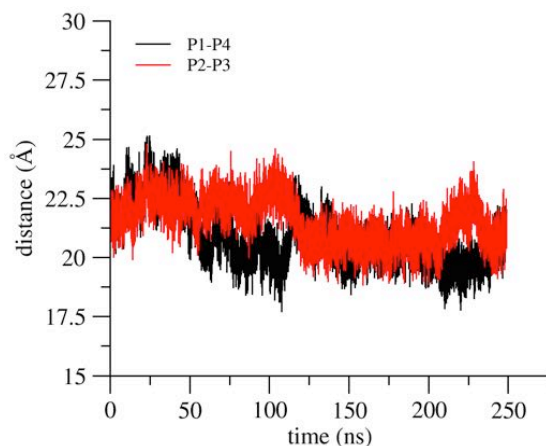


Fig 5.7. Cross-distances between facing C52 C α atoms in the single-pore. Measure of cross-distances between facing C52 C α atoms. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

Along the 250 ns MD simulation, each *cis*- dimer retains the *face-to-face* interaction between β 4 strands, which are at the basis of the pore model by Suzuki et al. The presence of stable HBs at this interface is essential to preserve the global stability of the model and to maintain the negatively charged surface of the strand pointing towards the interior of the channel. In **Table 5.1** a list of the most persistent *cis*- HBs found during the simulation are reported. Stable HBs are formed by β 4 residues of the two dimers such as S60, N61 and C62, consistent with the enhanced *face-to-face* dimerization observed in cross-linking experiments where N61 is mutated in Cysteine [1]. Remarkably, the network of HBs is essentially symmetric in the two *cis*- interfaces where two HBs formed between the backbone nitrogen of C62 and the backbone oxygen of C62 of the associate *cis*- protomer are the most stable in both dimers. It is important to stress the interesting role of the conserved C62 residue, which contributes to stabilizing each extracellular region via a disulfide bridge with the other conserved cysteine C52, and also forms a persisting HB with the C62 of the *cis*- partner is highlighted.

The single-pore dynamics shows how the two *cis* dimers engage stable *trans*- HBs. The existence of these favorable interactions had only been hypothesized when the model was first described [1], since various residues of the extracellular zone were missing. In **Table 5.2** a list of the most persistent *trans*- HBs found along the trajectory is reported.

The *trans*- interface also comprises contacts between hydrophobic residues that seal together facing protomers. In particular, each opposing pair (P1-P4 and P2-P3) establishes ECL2-ECH interactions by bringing the conserved hydrophobic residue A152 close to the facing cluster of conserved residues M68, L69, A70, L71 as illustrated in **Fig 5.8**, panel **A** an arrangement that persists for most of the simulation: the distance between the C α of A152 and the backbone center of mass of ECH residues 68 to 71 in the facing protomer is stationary around about 7 Å during the whole trajectory for all pairs except P4-P1, where it is so for 75% of the trajectory.

Stable interactions are also formed by ECL1 hydrophobic residues of the opposed pairs P1-P3 and P2-P4, as introduced by a representative snapshot in **Fig 5.8**, panel **B**. Specifically, L57 in the loop connecting β 3 and β 4 is very frequently at a distance smaller than 4 Å from a group of ECL1 residues which includes the conserved V38, I39 and I44, with percentages that vary among protomer pairs and specific residues, but always fall between 50% and 95%, except for the P4-P2 pair, where L57 is in contact with A152 of ECL2 for 57% of the trajectory.

Analysis of the MD simulation of the double-pore model. In this paragraph, we validate the entire model published by Suzuki et al., by performing MD simulations of a double-pore configuration. Studying a double-pore structure is necessary to obtain a more extensive validation of the entire model published by Suzuki et al, because this is the minimal configuration showing the *cis-linear* arrangement between monomers in the same membrane (as introduced in **Fig 1.8** and **Fig 4.7**).

Table 5.1. HBs in the *face-to-face cis-* interactions of the single pore. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

Dimer 1			Dimer 2		
%	P1 residue (atom)	P2 residue (atom)	%	P3 residue (atom)	P4 residue (atom)
80	C62 (N)	C62 (O)	86	C62 (O)	C62 (N)
74	C62 (O)	C62 (N)	80	C62 (N)	C62 (O)
68	S60 (O)	D64 (N)	74	D64 (N)	S60 (O)
32	N61 (ND2)	N61 (OD1)	66	S60 (O)	D64 (N)
			38	D64 (OD2)	S60 (OG)
			36	N61 (ND2)	N61 (OD1)
			35	N61 (ND2)	C62 (O)
			34	D64 (OD1)	S60 (OG)
			30	D64 (OD2)	S60 (N)

Table 5.2. HBs in the *trans-* interactions of the single pore. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

P1-P3			P2-P4		
%	P1 residue (atom)	P3 residue (atom)	%	P2 residue (atom)	P4 residue (atom)
50	T41 (OG1)	D55 (OD2)	65	N42 (OD1)	N42 (ND2)
44	T41 (N)	D55 (OD2)	62	N42 (ND2)	N42 (OD1)
			33	T41 (OG1)	T41 (OG1)

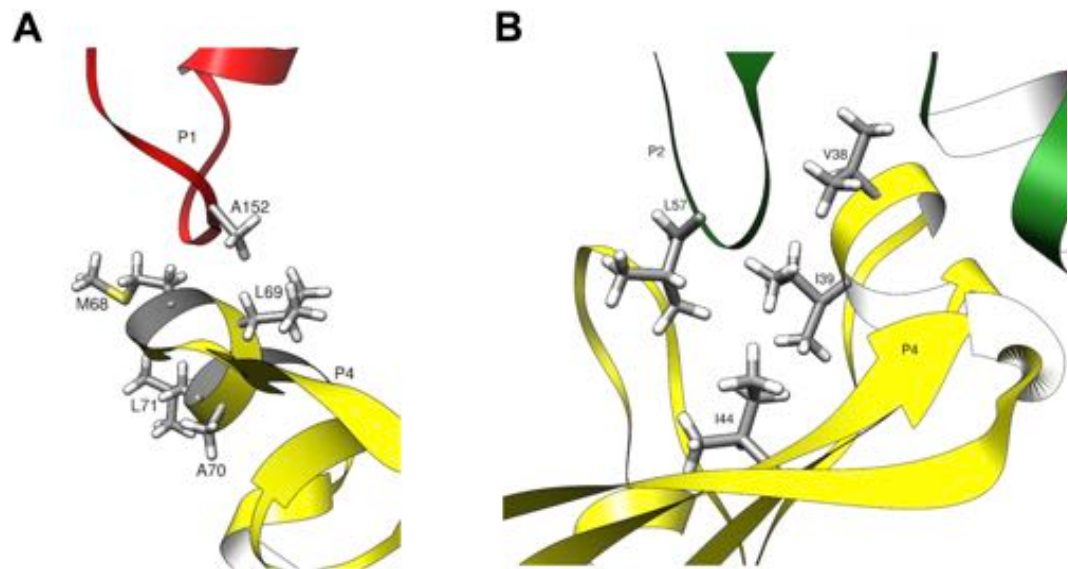


Fig 5.8. Hydrophobic contacts in the single-pore. **A** Contacts between the P1 A152 residue and the conserved residues of the P4 ECH region. **B** Hydrophobic interactions between the ECL1 segments of diagonally opposed protomers. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

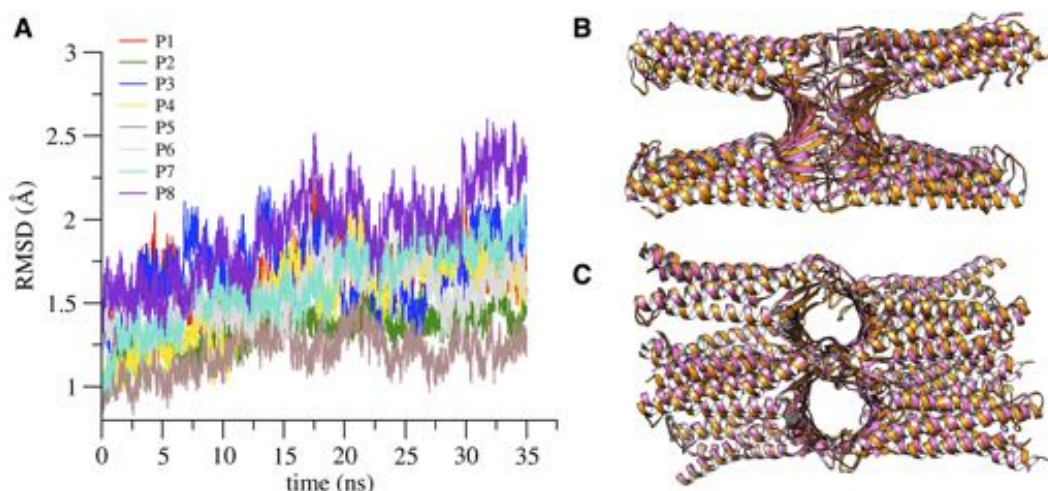


Fig 5.9. RMSD of the double-pore structure. **A**, RMSD values of the backbone atoms of each protomer along the double-pore simulation. **B** and **C**, superposition of the initial (pink ribbon) and the configuration at 30 ns, viewed from the lateral and apical side, respectively. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

As done for the previous model, the backbone atoms RMSD with respect to the starting conformation was calculated after optimal alignment of the structures (**Fig 5.9**, panel **A**). The different chains are indicated with names P1 to P8 following again the legend illustrated in **Fig 4.7**. The time evolution of each protomer shows values similar to those in the single-pore simulations.

The architecture of the double channel is well preserved along the trajectory, as can be observed in **Fig 5.9** panels **B** and **C**, where the double-pore conformation at the beginning of the simulation (pink ribbon) is compared with a snapshot taken at 30 ns (orange ribbon). The scaffold of the double channel is very stable, as suggested also by the time evolution of the cross-distances between the C_{α} atoms of the four pore-lining C52 residues in both the channels (see **Fig 5.10**, panel **A**).

Moreover, all the main *cis*- and *trans*- inter-monomer surfaces were analyzed. Here, the arrangement of the different protomers lead to a more complex network of interactions that can be summarized as follows:

1. the *face-to-face cis*- interactions for the dimers P1-P2, P3-P4, P5-P6, P7-P8;
2. the *linear cis*- interaction for the dimers P1-P5, P3-P7, P6-P2, P8-P4; where the first protomer name in the pair is that of M68;
3. different *trans*- interactions.

The analysis starts with the *face-to-face* interactions, now occurring between the dimers formed by P1-P2, P3-P4, P5-P6 and P7-P8, and listed in **Table 5.3**. Similarly to the tetramer case, each pair shows HBs between residues of the $\beta 4$ barrel such as S60, N61 and C62. The interaction between C62 residues of opposing protomers is also detected here for three of the four dimers, while in the missing case (P3-P4) C62 of P3 is in interaction with the C62 preceding residue N61 in P4.

Most importantly, the presence of the *cis*-linear interaction surface, described in the model of Suzuki et al, is the distinctive feature of the double pore. This motif comprises interactions between the ECH residue M68 of one protomer with residues F146, F147, and L158 in the TM3 and ECL2 of the adjacent chain. In particular residues M68, F146, and F147 of mCldn15 were reported to be involved in a hydrophobic interacting network which is essential for TJ strand

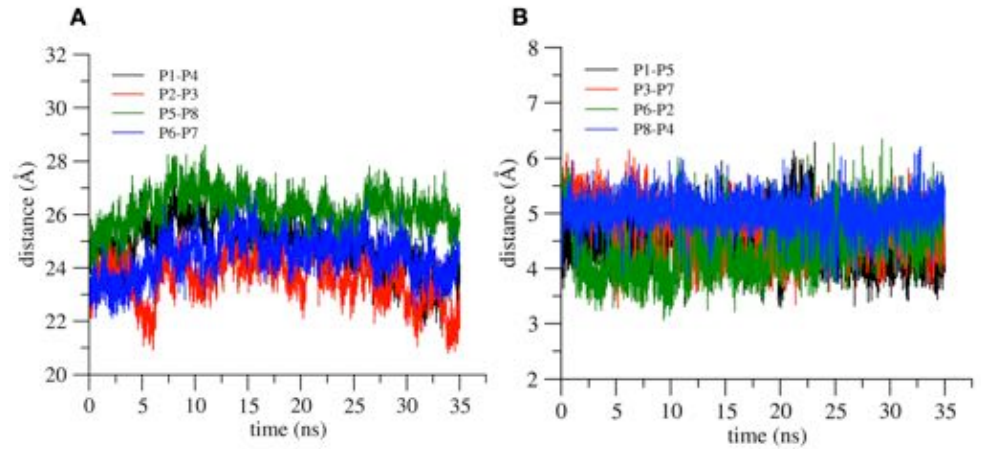


Fig 5.10. Cross-distances between facing C52 C α atoms and *Cis*- linear arrangement in the double-pore. A Cross-distances between facing C52 C α atoms. **B** *Cis*- linear arrangement in the double-pore. Distances between the M68 sulfur atom and the center of mass of F146 benzene ring. In the legend of **B**, the first protomer name in the pair is that of M68. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

formation [40]. Therefore, to verify the persistence of this interaction, the distance between the M68 sulfur atom and the center of mass of the F146 benzene ring, for each of the four linear *cis*- interacting dimer, was calculated. These results are shown in **Fig 5.10**, panel **B**, and reveal values that nicely correspond to the main peak of the probability density of Met sulfur - Aromatic distances (~ 5.5 Å), calculated from *in vitro* experiments, multi-scale simulations, and a bioinformatics screen of the PDB, as reported by [233].

Additionally, the persistence of HBs for *trans*- interactions between protomers is studied as reported in **Table 5.4**. Although many involved residues are the same of the single-pore *trans*-interactions (**Table 5.2**), the specific patterns show differences. This is a consequence of the different modeling performed for the missing segments in the two systems. However, it is important to remark that these interactions were only hypothesized in the architecture by Suzuki et al., and hence both the patterns fit with the original model structure.

Table 5.3. HBs of the *cis*- face-to-face interactions of the double pore. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

%	P* residue (atom)	P* residue (atom)
74	P6 D64 (O)	P5 S60 (O)
70	P2 C62 (N)	P1 C62 (O)
69	P6 C62 (N)	P5 C62 (O)
67	P5 C62 (N)	P6 C62 (O)
64	P3 C62 (N)	P4 N61 (OD1)
48	P8 C62 (N)	P7 C62 (O)
47	P1 C62 (N)	P2 C62 (O)
40	P2 N61 (ND2)	P1 N61 (OD1)
38	P4 N61 (ND2)	P3 C62 (O)

Finally, in the double-pore model additional *trans*- interactions can be observed between ECL2 segments of facing protomers. Indeed, in Suzuki et al., it is hypothesized that the missing region V2, part of ECL2, plays a primary role in the formation of multi-pore strands between Cldns from adjacent cells and it is stated that ECL2 has an anchoring function for *trans*- interactions in TJ strands [51]. In the double-pore model, we observed a dense packing of hydrophobic

Table 5.4. HBs of the *trans*- interactions of the double pore. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

%	P* residue (atom)	P* residue (atom)
84	P5 T41 (OG1)	P4 T40 (O)
83	P5 T40 (N)	P4 V34 (O)
65	P5 N37 (N)	P4 N37 (O)
60	P4 H35 (ND1)	P5 V38 (O)
53	P5 N37 (ND2)	P4 T40 (OG1)
49	P6 T41 (N)	P8 S56 (O)
46	P7 T40 (OG1)	P5 D55 (OD1)
45	P5 T40 (OG1)	P7 S56 (O)
40	P7 H35 (ND1)	P2 V38 (O)
36	P7 N37 (N)	P2 V34 (O)

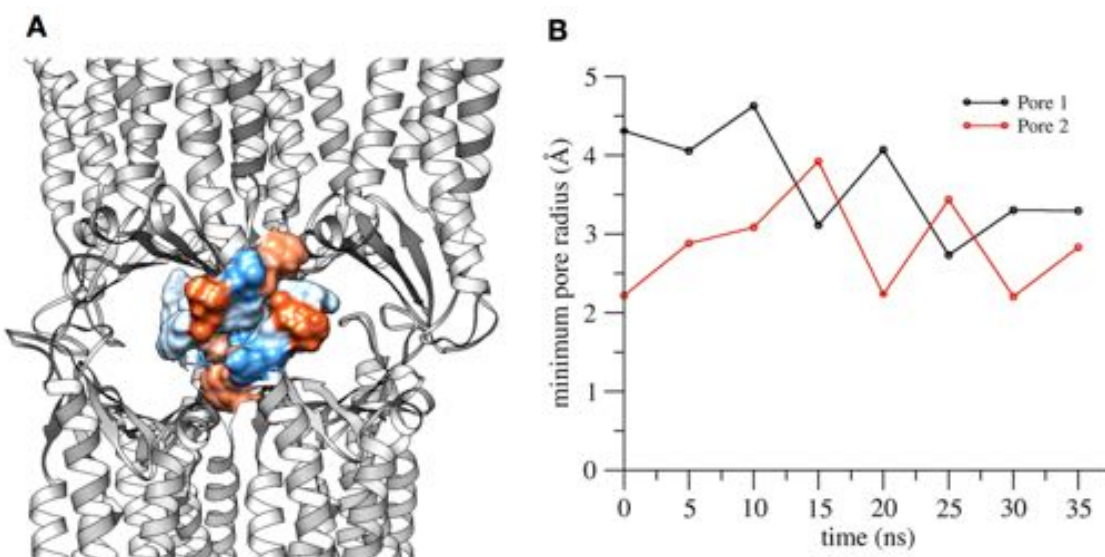


Fig 5.11. Hydrophobic *trans*- interactions and pore minimal radius of the double pore. **A**, Representation of the *trans*- ECL2-ECL2 interactions as surfaces, and colored following the Kyte-Doolittle scale: from blue (most hydrophilic), to orange (most hydrophobic). **B**, Evolution of the minimal radii of the two pores. Reprinted from [7], <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0184190>. Copyright (2017), with permission under the terms of the Creative Commons Attribution License.

residues from symmetrically facing ECL2s, including F146, F147, L150, A152, G153. This cluster is illustrated in **Fig 5.11**, panel **A**.

Last, the analysis of the dimension of the pore radius was performed for both the channels. The values of the minimal radius of both pores are shown in **Fig 5.11**, panel **B**, where the evolution of the constriction is similar in the two channels and it converges to values close to that obtained for the single pore.

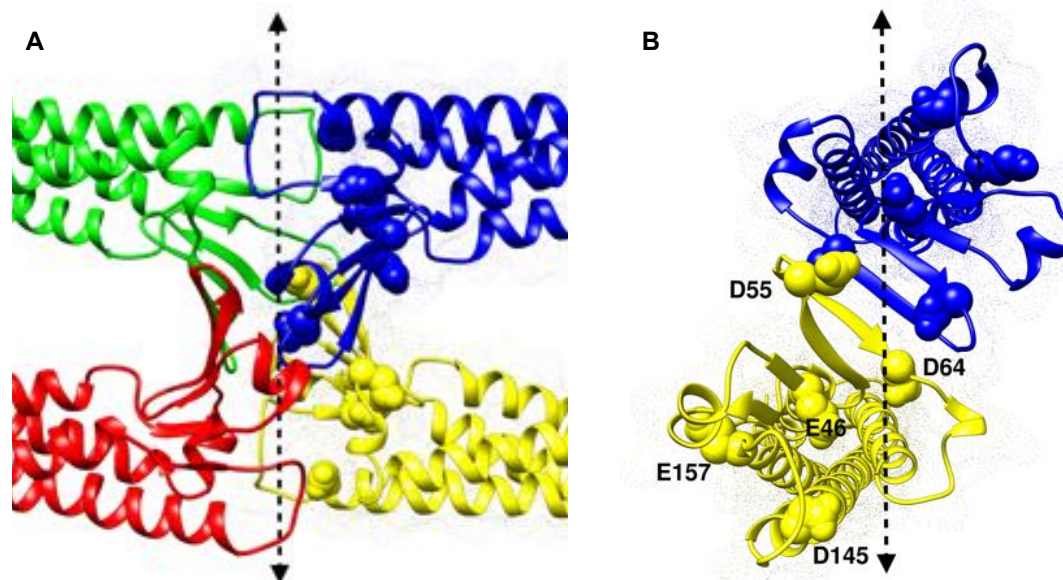


Fig 5.12. Amino-acid composition of the paracellular pore surface. Representative structure of the single pore model, used for thermodynamic and kinetic calculations, using the same legend of Fig 4.3. **A**, Lateral, zoomed view of the Cldn15 TJ pore, with the surface representation used for protomers 3 and 4, only. The side chains of acidic residues in protomers 3 and 4 are shown as spheres; the dashed line represents the axis used in the simulations of ion permeation. **B**, Cut-away view of the channel, obtained with a 90° rotation of the figure in panel A. The section shows only protomers 3 and 4 for clarity and with acidic residues numbering indicated for protomer 4. Note that the β -strands are tilted with respect to the channel axis, as in β -barrel structures.

5.1.2 PMF Calculations and Ion Permeation Rates with US and VTMM Simulations

A wealth of experimental data confirms that Cldn15 paracellular channels are highly selective for cations, particularly Na^+ . Therefore, to further validate the structural configuration of the channel introduced by Suzuki et al., it is mandatory to test the ionic selectivity through the cavity of the pore. To this aim, we investigate the selectivity features of the model, with a single CV, using both US and VTMM MD simulations.

Results from US simulations. To verify the effect of channel structure and amino-acid composition on ion permeation, PMF calculations were performed for Na^+ , K^+ and Cl^- , using the position of the ion along the pore axis as CV (Fig 5.12).

As already stressed in Section 3.4, the main feature of an accelerated simulation for PMF calculations is that it allows investigation of processes that are not easily accessible by direct MD simulations. Among such schemes, Umbrella Sampling (US) is one of the most used methods and it was used here for preliminary evaluation of the PMF.

The resulting US PMF profiles for the ionic permeation are reported in Fig 5.13, where the error is estimated using bootstrap analysis. All these results show one main feature of the model, and they are symmetric with respect to the center of the channel. The profiles of Na^+ and K^+ reveal a minimum, in both cases about 4 kcal/mol deep, while that of Cl^- shows a large barrier of about 8 kcal/mol.

Remarkably, the shape of the profiles correlates with the position of the acidic residues and the narrowing of the channel towards its center. Indeed, by observing the channel structure, several negatively charged residues might drive the flux of Na^+ ions through the pore, and hinder that of Cl^- . Moving from the mouth of the channel towards the narrow center, these residues are, in each protomer (see Fig 5.12, for a visual inspection): D145 (whose side chains

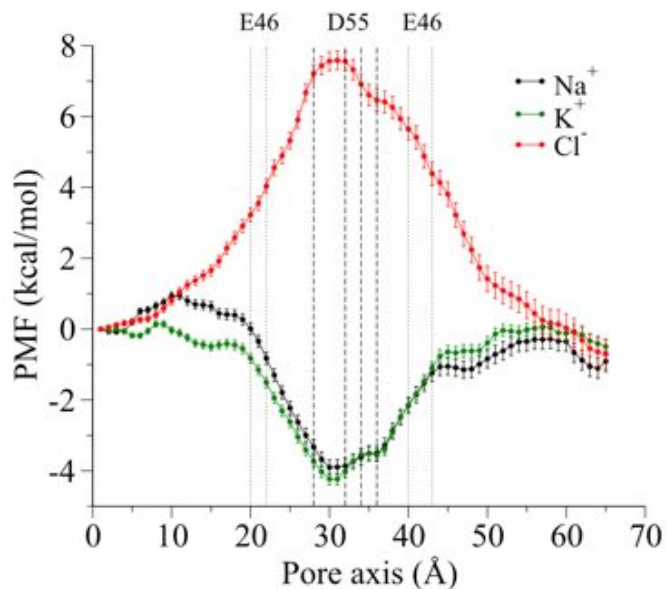


Fig 5.13. Energetics of ion permeation (US-WHAM simulations). PMF profiles for the permeation of Na^+ (black), K^+ (green) and of Cl^- (red), obtained from US-WHAM calculations. The error estimation for each window, obtained with bootstrap analysis, is represented with bars. The position of the C_α atoms of D55 and E46 residues are represented as dashed and dotted lines, respectively. The C_α atoms of D55 residues are located at the narrow region of the pore (windows at 28, 32, 34, 36 Å, respectively, see text). Reprinted from [8], <https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484>. Copyright (2018), with permission from the American Chemical Society.

point towards the channel interior), E157 (side chains towards the exterior), E46 (internal side chains), D64 (external side chains) and D55 (internal side chains). Conversely, there are only two basic amino-acids per protomer in the paracellular space, R30 and K155, both located at the mouth of the channel, where the pore is wider. The side chain of K155 points towards the exterior of the channel, while that of R30 points toward a facing transmembrane helix, and it is tangential to the channel walls. The positions of all these residues are symmetrical with respect to the channel axis and the channel center.

With respect to the PMF profiles in **Fig 5.13**, the C_α atoms of D64 are at the level of windows at 25 Å and 38 Å, and those of E46 at 20, 22 Å and 40, 43 Å. The positions of D55, corresponding to the narrow region of the pore, are at 28, 32, 34 and 36 Å, and coincide with the minimum in the cation PMF and the maximum in anion PMF.

Results from VTMM MD simulations. After the determination of the PMF along the channel axis using US, the results were verified with another method, the soft-walls Voronoi Tessellated Markovian Milestoning (VTMM). These additional PMF calculations return very similar profiles (**Fig 5.14**, panel **A**) and globally confirm all the main features of the US simulations.

Moreover, in addition to thermodynamic calculations, Milestoning permits to estimate the time-scale of the ionic translocation through the pore. In particular, the Mean First Passage Times (MFPTs) of ion permeation through the channel were obtained from the Milestoning MD simulations. The MFPTs of ion permeation (cumulative from the first milestone to all the successive ones) are shown in **Fig 5.14** panel **B** as a function of the coordinate along the channel axis. Global MFPTs are reported in **Table 5.5**, and they refer to the transit of an ion from a milestone at one mouth of the channel to a milestone at the opposite mouth. As can be observed, while the rates of cation permeation are quite similar, that of Cl^- is reduced by a

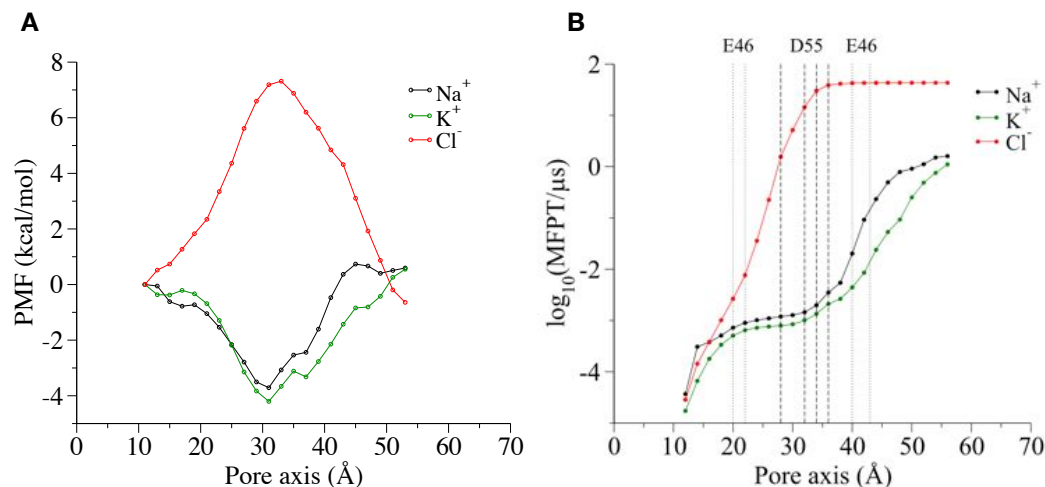


Fig 5.14. Energetics and kinetic of ion permeation (VTMM MD simulations). **A**, PMF profiles for the permeation of Na⁺ (black), K⁺ (green) and of Cl⁻ (red). **B**, MFPTs along the pore axis. Reprinted from [8], <https://pubs.acs.org/doi/10.1021/acs.jpcb.8b06484>. Copyright (2018), with permission from the American Chemical Society.

factor 10, further confirming the strong channel selectivity for cations.

In summary, the combination of MD, US, and Milestoning tools provide, for the first time, both thermodynamic and kinetic estimates of ionic translocation through this model of Cldn15 paracellular channels.

Table 5.5. Kinetic measures of ion permeation. MFPTs and relative rates of single ion permeation through the Cldn15 paracellular channel. Reprinted from [8], <https://pubs.acs.org/doi/10.1021/acs.jpcb.8b06484>. Copyright (2018), with permission from the American Chemical Society.

Ion	MFPT (μ s)	Relative rate
Na ⁺	1.60	1.00
K ⁺	1.10	1.40
Cl ⁻	43.10	0.04

5.1.3 Pore Size and Hydration of Na⁺ during Permeation

From the previous analysis, it is clear that the D55 residues provide a binding site for cations. In contrast to the D55 site for Na⁺, other pore-forming negatively charged residues, such as E46, have minor interactions with the Na⁺ ion. Thanks to its electrostatic affinity, Na⁺ ions interact with these pore-forming residues. In addition, the dimension of the paracellular channel allows the flux of water that passes through the pore. These results agree with the current experimental results for Cldn15 paracellular complexes, as efficient channels for the transport of Na⁺ ions.

To further assess the reliability of the model, the correlation between the pore size and the hydration mechanisms of the Na⁺ ion was investigated. Interestingly, the MD simulations maintain a stationary minimum pore size with a diameter of \sim 5-6 Å (Fig 5.6, panel A), in agreement with available experimental results for intestinal channel forming, cation selective, Cldns (Cldn2 [232] and Cldn15 [234]). Remarkably, the experimental work [234] was published after the results described in this Thesis.

In the single pore Cldn15 model, the pore has a minimal size which is comparable with the dimension of the first hydration shell for Na⁺. Therefore it should critically affect the hydration

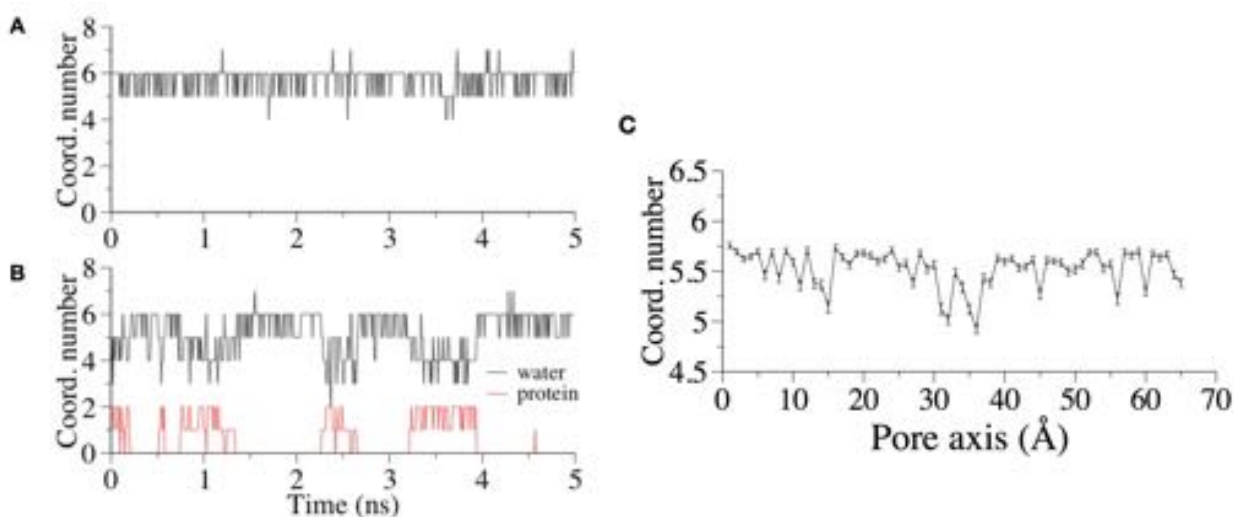


Fig 5.15. Number of oxygen atoms coordinating the Na^+ ion in US windows. **A**, Window at 2 Å. **B**, Window at 32 Å: the black line indicates the coordinating oxygen atoms from water molecules, while the red line indicates those from the D55 sidechains. **C**, Average of the number of oxygen atoms coordinating the Na^+ , as a function of the reaction coordinate, from all the US simulations. Reprinted from [8], <https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484>. Copyright (2018), with permission from the American Chemical Society.

state of the ion during permeation. Moreover, the D55 ring is located at the narrowest region of the pore, so that size and electrostatic effects combine to determine selectivity.

Furthermore, the hydration state of the Na^+ ion over the permeation pathway from the US calculations was analyzed and the number of oxygen atoms from water molecules and protein residues coordinating the Na^+ ion in the first hydration shell (using 3.0 Å as threshold) was computed. In **Fig 5.15**, this analysis is performed in the US windows at 2 Å, panel **A**, and at 32 Å, panel **B**, near the mouth of the channel and at the central D55 ring site, respectively. In the most external window, the ion is fully solvated (the average number of coordinating water molecules is 5.7 ± 0.5). Conversely, when in the narrow channel region, the Na^+ ion occasionally loses one or more hydrating water molecules, replaced by oxygens from the sidechains of D55. This is remarked by the calculation of the average of the number of oxygen atoms coordinating the ion as a function of the reaction coordinate (**Fig 5.15**, panel **C**) where the main dehydration events of the Na^+ ion are located at the middle of the channel.

To confirm this observation, the entire unbiased MD trajectory was analyzed and a time window was observed where a Na^+ ion in the paracellular space spontaneously enters the channel and binds to the aspartic acids of the ring in a partially dehydrated state. In **Fig 5.16**, two snapshots representing such events are reported.

5.1.4 PMF Calculations with SSw Simulations

Permeation of cations. To verify the effect of the channel structure on permeating cations (Li^+ , Na^+ , K^+ , Rb^+ , Cs^+), further PMF calculations for ionic permeation with new CVs were performed. The SSw method was used to compute the 3D-PMF landscape of ionic Cartesian coordinates. This approach was already successfully applied to map irregular landscapes (see [172, 184–187]). The resulting PMF landscape for Na^+ inside the channel is reported in **Fig 5.17**, panel **A**, where the isosurfaces reveal a deep minimum at the constriction region of the pore, in correspondence with the four D55 residues. The permeation pathway identified by the Minimum Free Energy Path (MFEP) on the landscape passes through that minimum, and

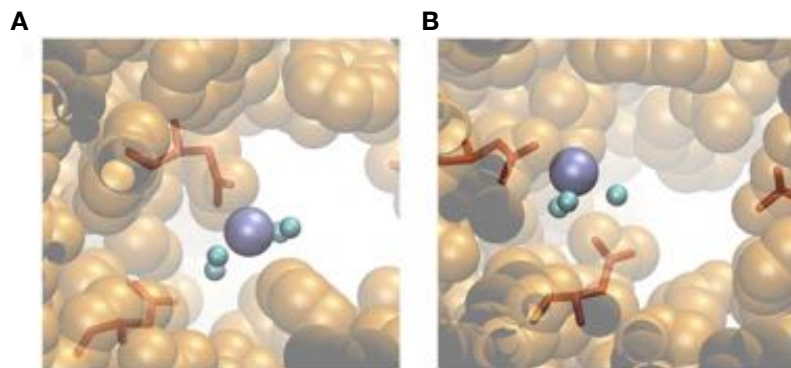


Fig 5.16. Snapshots of partial dehydration events of a sodium ion in the pore. Partial dehydration events of a Na^+ ion in the pore region. **A**, Snapshot extracted from the US simulation at 32 Å. **B**, Snapshot extracted from the unbiased MD simulation. The ice-blue sphere is the Na^+ ion, small cyan spheres represents water molecules in the ion hydration shell, while all other water molecules in the system are not shown. The Na^+ ion is hydrated with only 4 and 3 water molecules, respectively. The side chains of the D55 residues are represented as red sticks, and the protein as orange VdW spheres. Reprinted from [8], <https://pubs.acs.org/doi/10.1021/acs.jpcc.8b06484>. Copyright (2018), with permission from the American Chemical Society.

the corresponding 1D-PMF profile reveal a well ~ 5 kcal/mol deep (**Fig 5.17**, panel **B**). To further evaluate the charge selectivity of the paracellular channel model, 3D PMF calculations were performed for the series of cations Li^+ , K^+ , Rb^+ , Cs^+ . PMF profiles along the MFEPs on the landscapes, shown in **Fig 5.18**, reveal that all cations have an energy minimum in correspondence of the D55 ring with a range between 5-9 kcal/mol. Interestingly, the Na^+ ion has a less deep PMF minimum, which should provide a more facilitated passage through the channel. These simulations indicate that the channel allows the passage of the different cations, without presenting evident mechanisms of selectivity, especially between K^+ and Na^+ , in accordance with the previous calculations with the single CV.

Permeation of ADG. To fully validate the single pore configuration, a final investigation was performed to study the permeation of small molecules through the cavity of the model. As already stressed, in particular in **Subsection 1.4.1**, all Cldns TJ architectures have a typical attitude to prevent the flux of almost all the molecules.

In this Thesis, the analysis of molecule permeation is limited to the representative α -D-Glucose (AGD), which is a neutral molecule with a relevant physiological role in organisms. Inspired by many results of PMF calculations with multiple CVs *via* SSw calculations [172,184–187], this method was preferred to US simulations which is often limited to 1 or 2 CVs.

The 1D-PMF profile calculated along the MFEP on the 3D-PMF with SSw simulations shows a ~ 7 kcal/mol barrier in the narrow region of the the Cldn15 pore, revealing that the channel does not allow ADG passage. The action of a barrier is mainly due to the charged and polar residues in the third and fourth β -sheets of each protomer, with the formation of HBs between the different residues and the ADG hydroxyl groups.

This result demonstrates that our pore model does not allow the flux of small molecules and act as a tight regulator of the solute traffic, in accordance with the main role of Cldns in TJs.

5.2 Discussion

5.2.1 Structural Features of the Models

Understanding the molecular mechanisms at the basis of biological barriers is a target of pivotal importance in molecular biology and pharmacology.

TJs are molecular architectures that link adjacent epithelial cells by forming a network of strands on each cell wall, and regulate paracellular diffusion of solutes across the intercellular

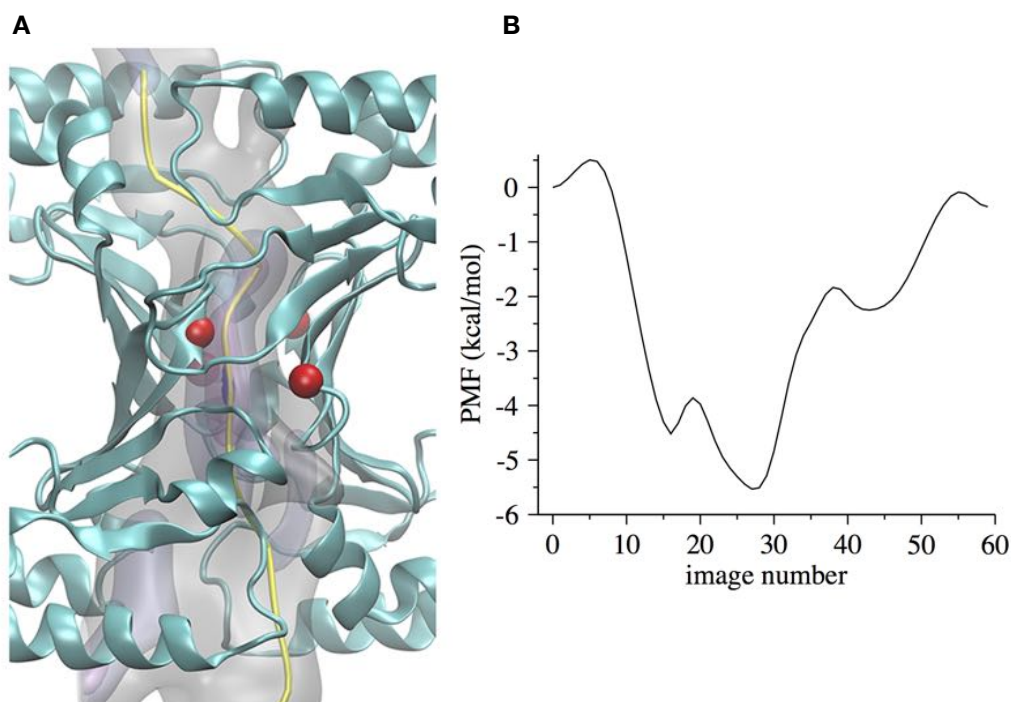


Fig 5.17. 3D PMF profile of Na^+ permeation through the channel from SSW simulations. **A**, Isosurfaces for the 3D PMF map for the coordinates of Na^+ inside the channel (blue: 0 kcal/mol; mauve: 1 kcal/mol; azure: 3 kcal/mol; grey: 8 kcal/mol). The D55 C_α atoms are shown as red spheres. The yellow line represents the MFEP. **B**, The PMF profile along from the MFEP.

space. The structural and functional properties of TJs are mainly ascribed to members of the Claudin (Cldn) family.

Structural studies of Cldns have only recently started to help dissecting the molecular architecture of TJs and the mechanism of their function. At the moment, there are only four available crystal structures of Cldns: Cldn15 alone [40], and, on the other hand, Cldn4, Cldn19, and Cldn3 bound to the same fragment toxin. In [1] the Suzuki model of TJ channels was proposed by assembling multiple Cldn15 monomers to form a paracellular architecture. The detailed atomic model fits multiple experimental data such as crosslinking, mutational experiments and EM images, but misses residues in Cldn extracellular loops that, once added, could generate steric overlaps. For these reasons, it is so far considered a valuable starting hypothesis that needs further testing. Intrigued by the growing interest in this paracellular model, the first part of my PhD project was devoted to the generation of a refined all-atom structural model of a Cldn15 channel [7], based on the work of Suzuki and collaborators.

Using structural modeling and all-atom MD simulations, the assembly of Suzuki et al. was refined to provide high resolution, integral chain structures of Cldn-based paracellular channels in explicit double membrane bilayer and solvent environment. In MD models, the architecture of the channel was well maintained, as revealed by overall stability of the protomers and preservation of the β -barrel-like pore scaffold by residue-residue *cis*- and *trans*- interactions between monomers known from experiments to be critical for TJ strand formation.

Moreover, these simulations showed a stationary minimal pore diameter of 5-6 Å. In the more extended unrestrained trajectory of the single-pore model, we observed large fluctuations in minimal pore radius within 0-100 ns, followed by stabilization around 2.5 Å up to 250 ns (**Fig 5.6**), while the double pore system showed rapid convergence around 3 Å for the minimum

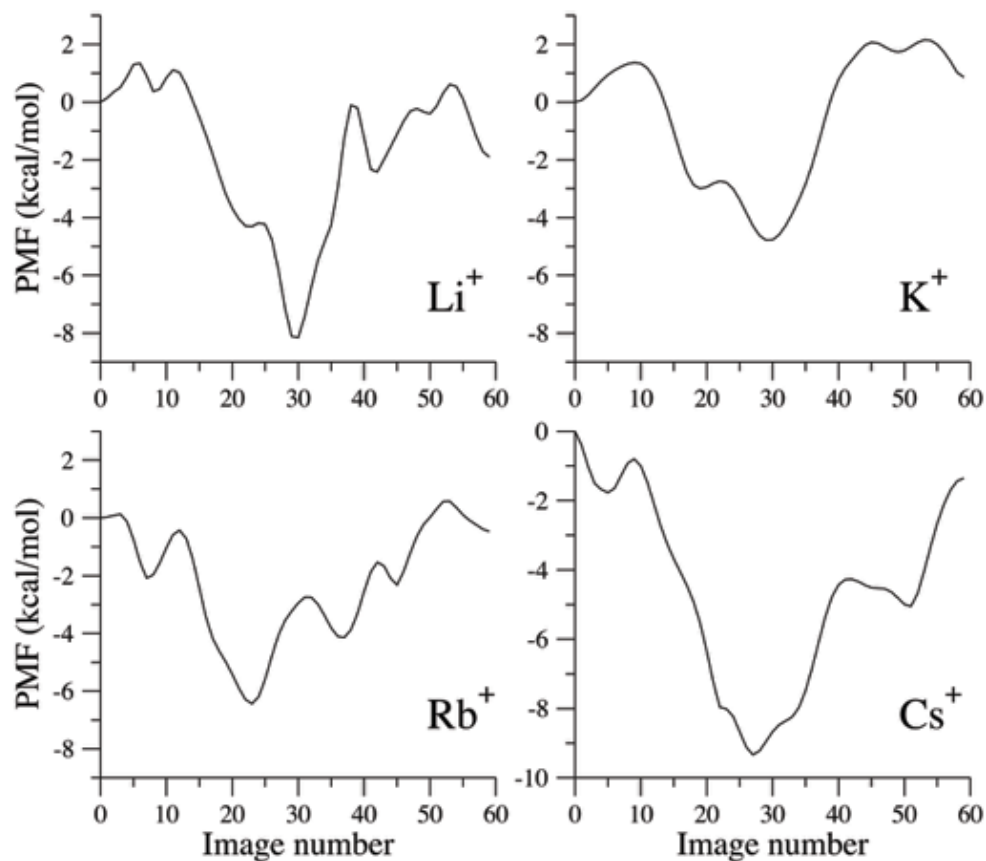


Fig 5.18. PMF profile of additional cations from SSw simulations along their MFEP. PMF profiles obtained from SSw simulations for Li^+ , K^+ , Rb^+ , Cs^+ .

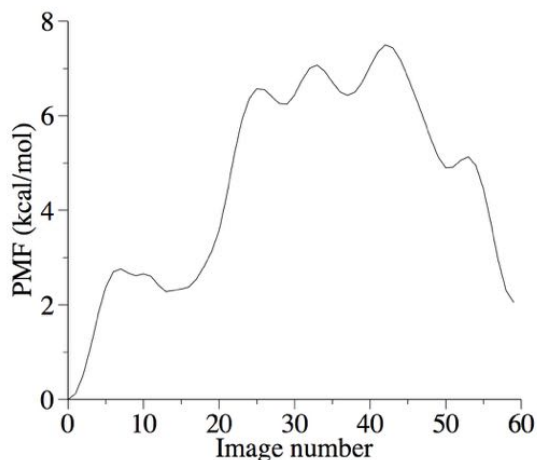


Fig 5.19. PMF of α -D-Glucose permeation through the channel. The PMF barrier of α -D-Glucose (ADG) permeation through the channel, calculated via SSw simulations.

radius of both pores (**Fig 5.11**). These results are in agreement with electrophysiological studies performed on another pore-forming Claudin, *Cldn2*, where the formation of tight pores of 6.5 Å diameter was suggested [232].

By analysis of HB lifetime and other interactions patterns, we confirmed the relevant role of specific residues in establishing protomer-protomer interacting interfaces. For the *cis face-to-face* interface, the $\beta 4$ strands engage persistent HBs which include S60, N61 and C62 residues.

The double pore model simulation also allowed observing the linear *cis*- interaction surface. This motif is based on the Cldn15 crystal packing [40], and contains ECL2 residues whose corresponding ones in Cldn5 and in Cldn3 have been shown to regulate TJ strand formation [51]. For Cldn15, the role of M68, F146 and F147 has been observed by freeze fracture EM, where mutations to smaller or charged residues at these sites hampered strand formation [40]. In our double-pore trajectory, a very stable interaction was observed among these residues, as revealed by distances between the M68 sulfur atom and the center of mass of F146 benzene ring that are stationary around 5 Å.

Most importantly, the MD simulations allow to investigate for the first time the formation of *trans*- interactions between dimers that involve the missing protein segments in the model by Suzuki et al. This analysis reveals that ECL1 and ECL2 residues engage persistent HBs and hydrophobic interactions with residues of facing protomers.

As a final remark, while the single-pore structure provides a reduction in system size that is advantageous for MD simulations, it is still of general interest since it is known that the main molecular determinants of TJ selectivity are at the individual pore level. On the other hand, the double-pore system comprises structural features that complement those of the single-pore and that are responsible for TJ strand formation.

It is important to remark that, after the publication of our first work [7], two other papers which investigate the Suzuki Cldn15 model, were published by two different groups [235,236]. In particular, in [235], atomistic MD simulations showed partial loss of structure in a Cldn15 strand of 16 protomers forming four pores, assembled as in the Suzuki model, while stable double-pore configurations was observed, as in our study [7]. Together with ours, these results point to the validity of the Suzuki model at least for single- or double-pore conformations. Very recently, another all-atom MD refined version of the Cldn15 Suzuki model has become available [236]. In this last work, the authors rely on the periodic replicas of the simulation box to build a virtually continuous strand of channels from a unit cell of three, and observe that stable pore architectures are maintained by the same *cis*- and *trans*- interactions between monomers we identified in [7]. Overall, it is remarkable that, regardless of minor differences in the protocols followed to complete the missing protein segments and to assemble the inter-membrane architecture, several independent studies demonstrated the plausibility of the paracellular channel putative structure of the Suzuki model.

Furthermore, our observations for the additional Cldn5 Suzuki model are in agreement with the recently published papers [50,237], where Coarse Grained (CG) and Molecular Docking simulations are used to investigate the Cldn *cis*- self-assembly and the formation of a pore, as described by the Suzuki model, for other classic Cldns such as Cldn-1, -2, -4, -5, -15, and -19. All these Cldns share a strong sequence similarity, while being diverse in their ion selectivity and physiological significance. In our simulations, we observed for the Cldn15 pore model a distribution of the charged residues able to produce a clear selectivity, with the action of a negatively charged cage in the middle of the channel, with the D55 and D64 residues.

5.2.2 Solute Permeation Features

The concept of charge selectivity is widely used to illustrate the ability of channel proteins to regulate the flux of different ions *via* the action of charged residues at specific sites (for an introduction to Cldn charge selectivity see [238]). Since experimental atom-detail structures have become available, many computational studies have investigated the selectivity properties of conventional transcellular ion channels such as voltage gated potassium KV channels [144–148,239–242] and sodium NaV channels [149–156,243–246]. On the contrary, due to lack of structural information, calculations of selectivity in paracellular channels are only starting to appear. Because a reliable model must recapitulate the physiological properties of the system it represents, the single-pore model introduced in this work was further validated by verifying that it correctly reproduces the experimentally known selectivity of Cldn15 pores for cations [46].

At a first stage, the US method was applied to calculate the PMF for single-ion permeation through the channel, using Na⁺, K⁺ and Cl⁻ ions, and the VTMM technique to determine further independent PMF calculations and the kinetic rates for the same process. In all these

cases, the position of the ion along the main axis of the channel was used as CV. The PMF curves obtained from US are symmetrical with respect to the channel center, consistent with the symmetric structure of the tetrameric pore (**Fig 5.13**). From a functional point of view, symmetric conductance was observed in Cldn2 cation channels, indicating ion fluxes electrochemically driven by a concentration gradient between the opposite ends [247]. PMF profiles show a single minimum, about 4 kcal/mol deep, for Na^+ and K^+ and a barrier of about 8 kcal/mol for Cl^- , centered at the narrowest region of the channel. These features well correlate with the distribution of amino-acids along the pore. Indeed, the structure displays several negatively charged residues from the different protomers lining the pore surface in symmetrical positions with respect to the main axis, capable of generating an electric potential to drive the flux of cations and hinder that of anions. These results show that the single pore structural model of the Cldn15 channel preserves the required selectivity for cations, and substantiate the experimental observation that selectivity is mostly regulated by the D55 residues [52]. The permeation time-scales for Na^+ , K^+ and Cl^- ions, calculated using soft-walls VTMM in 1D, yielded ratios of permeation rates that again confirm that our Cldn15 channel model replicates the marked charge discrimination of cation-selective paracellular channels.

It is important to stress that we performed calculations for single-ion permeation only, while the actual transport properties might result from the crossing of multiple ions. However, single-ion calculations provide the fundamental features of the process, as for example the location of binding sites and repulsive barriers, or the magnitude of the ion-protein interaction. Moreover, they determine conduction at relatively low ion concentration [248]. The results of our calculations, published in [8], fit well with those of other independent contemporary studies such as [237] and the already cited [236]. In [237], the authors performed a 300 ns-long all-atom MD simulation of a Cldn2 pore (also cation selective) similar to the Suzuki model, and observed permeation of 13 Na^+ and only 1 Cl^- . Even more recently, Samanta et al. [236], using all-atom MD, investigated the selectivity of the transport properties of a Cldn15 paracellular pores model by calculating ionic currents from simulated trajectories of Na^+ , other small positively charged molecules and Cl^- , under the application of a constant external electric field. In addition to these results, the thermodynamics and kinetics calculations introduced here provide a quantitative link between the ion permeation mechanism and the structure of the channel, in terms of an attractive energy well at the cation binding site and a repulsive energy barrier for the anion. In all-atom MD simulations, both ionic current calculations under external electric fields and enhanced PMF calculations are widely used to investigate the conduction and selectivity mechanisms of transmembrane channels. Remarkable examples include studies of potassium [240, 249, 250] and, more recently, sodium [149, 243, 245] channels. While the current provides a direct measure of conduction, it requires statistically sufficient sampling of ion crossing events, which implies long simulated trajectories or high applied voltages that can affect the channel structure and/or its mechanism. Ion fluxes, however, are ultimately determined by the energetically favorable/unfavorable sites encountered by the particles inside the protein, and a detailed map of these features is provided by PMF calculations. Hence, PMF maps and profiles allow a quantitative interpretation of the experimentally determined conduction in terms of the protein's amino-acid composition and conformation [248, 251]. In addition, for the specific case of paracellular channels, ionic current calculations require sealing the paracellular space with multiple pores to avoid passage of ions in the aqueous environment between bilayers, thus considerably increasing the size of the simulated system. PMF and rates calculations, conversely, while permitting the use of a smaller, single-pore structure, also yield the thermodynamic and kinetic properties of the individual conducting unit that are required to interpret TJ electrophysiology experiments using single-channel models [232, 247].

Furthermore, the observation that the minimal pore size of our Cldn15 channel is stationary around 5-6 Å, close to the size of a hydrated Na^+ atom, pushed us to investigate the details of ion hydration during permeation. Indeed, electrostatic and steric effects couple to determine ion transport in most channels, and pore width-related ion dehydration effects are known to be relevant in narrow channels such as gramicidin [252], or in the selectivity filter of potassium [241, 242] and sodium [149] channels. By inspection of the restrained simulations for PMF calculations and of the unrestrained MD trajectory, it was observed that, when in close proximity of the acidic amino-acids, the Na^+ ion can shed one or two water molecules from its hydration shell by substituting them with oxygens from the sidechains of the residues

(Fig 5.16). A first mechanism of *partial* dehydration was hypothesized to characterize Na^+ permeation through Cldn2 channels in Ref. [232], in agreement with the fact that Cldn2 form both cations and water channels. Furthermore, after the publication of the computational simulations introduced in this Thesis [7,8], a paper focused on experimental Cldn15 permeation was published [234]. Results of this work demonstrate that Cldn15, similar to Cldn2, forms a paracellular cation and water channel, but, in functional contrast to Cldn2, water and Na^+ fluxes through Cldn15 inhibit each other. Therefore, Cldn15 allows Na^+ to retain part of its hydration shell within the pore, reducing the simultaneous passage of additional water through the pore.

Our calculations of PMF and time-scales of ion permeation through the Cldn15 pore reveal interesting details on the mechanism of ion transport in paracellular channels. While these share some properties with other widely studied transmembrane channels, they also bear relevant differences. For example, ionotropic glutamate receptor channels have a pore diameter of 5.5-7.0 Å, similar to what we observe for the Cldn15 pore, and are able to permeate hydrated and partially hydrated cations across the membrane [253]. However, while in most channels ion permeation occurs through fairly rigid regions of the proteins embedded in a membrane environment [254,255], paracellular channels regulate transport parallel to the membrane planes, in the hydrated space between adjacent epithelial or endothelial cells.

The Suzuki model, our refined version and the only few others currently available [50,237], all predict Cldn channels with a pore region where side chains of extracellular loops form the narrow filter and modulate selectivity. This arrangement results in a pore structure that is overall less rigid than in transmembrane channels, and whose size is affected by fluctuations of the extracellular loops, particularly ECL1 [7].

The permeation of a family of solutes was further investigated with an alternative method named Single Sweep (SSw) for PMF calculations. This computational approach allows us to study the permeation problem with a different protocol, where the single CV of the ion position along the channel axis is replaced by a set of three CVs such as the Cartesian coordinates of the Center of Mass of the molecule. At this stage, this method is used to confirm that the choice of the single CV for US and VTMM calculations is appropriate to map the cavity of the model which has a *hourglass* shape. The algorithm is used to reconstruct the 3D PMF map for monovalent cations *i.e.* Li^+ , Na^+ , K^+ , Rb^+ , and Cs^+ , and to study the behavior of the D55 binding site for these different cations. Indeed, the reconstruction of the full 3D landscapes allows an exhaustive map of all possible ion binding sites. In our case, SSw calculations confirm that the four D55 at the narrowest region of the channel are the major regulators of the ionic flux. The isosurfaces of the Na^+ PMF landscape are characterized by a minimum centered at the constriction region of the pore, in correspondence with the four D55 residues. The 1D PMF profile of the MFEP extracted from the 3D landscape with the String Method quantifies the minimum for Na^+ as 5 kcal/mol deep. SSw calculations were also performed to reconstruct the 1D PMF profile of other cations such as Li^+ , K^+ , Rb^+ , and Cs^+ . All computed landscapes show an attractive energy minimum in correspondence of the D55 ring. The different depths reveal that the channel model is sensitive to differences among cation types.

Finally, in order to test the selectivity of the model for molecular size, the MFEP was obtained for ADG transport through the channel, using SSw. ADG is an interesting test molecule, because of its small size, charge neutrality and biological relevance. Recently, the authors of [50] investigated *via* US calculations the permeation process of ADG through a different model of a Cldn5 pore. Their model was obtained from spontaneous association of Cldn5 monomers in CG representation, and it is quite different from the Suzuki model we started from. Specifically, it displays a distinct *cis*- arrangement of monomers in the same membrane, resulting in a different paracellular pore architecture. Results for the Cldn5 pore showed that a ~ 6.5 kcal/mol barrier hinders the passage of ADG [50]. In our calculations for the Cldn15 pore, we also observe a repulsive barrier at the centre of the channel, measuring ~ 7 kcal/mol at its maximum. This is consistent from a physiological point of view since nutrients such as glucose and amino-acids are absorbed in the small intestine through the transcellular route [256,257], and although it is not excluded that there exists paracellular absorption of ADG, its contribution is likely to be negligible [258].

As a final consideration, it is important to highlight how the calculations based on the SSw method can be further refined by revising the protocol used. In fact, during our simulations, we

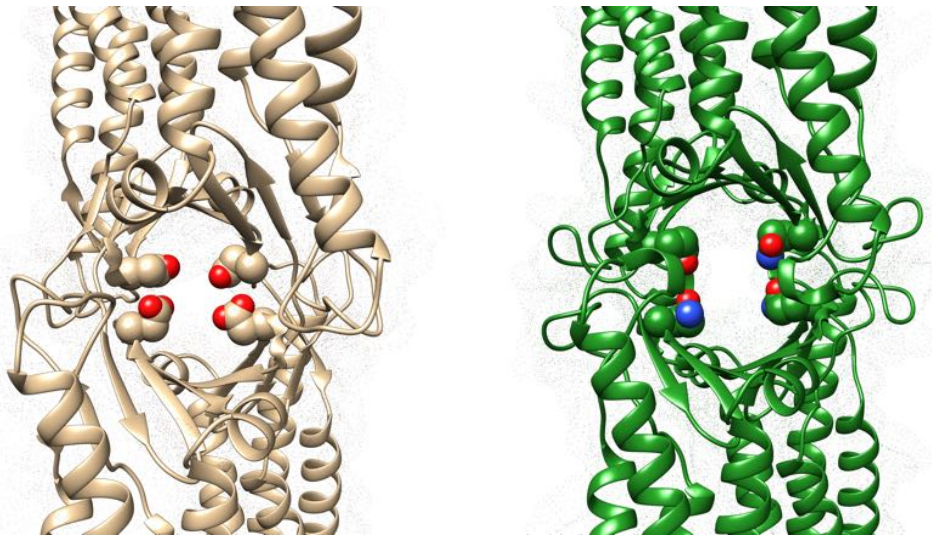


Fig 5.20. Pore characterization in Cldn15 and Cldn5 complexes. View of the Cldn15 (brown model) and Cldn5 (green model) Suzuki pore structures, perpendicular to the pore axis. The models are described with the ribbon style and their surface with a dot-mesh. Pore-lining residues located in the narrow region of the filter (Cldn15 D55 and Cldn5 Q57) are highlighted with spheres.

noticed that mean forces calculations for centers that are closer to the protein (with a distance from the channel below 3.0 Å) could result in very high value, giving problems with the PMF reconstruction. A possible solution we will explore is to repeat the mean forces calculation for these centers to remove harmful fluctuations of the system. Furthermore, our method includes a number of centers that could be further increased, using a larger and more finely resolved irregular mesh of centers. In previous application of the SSw method, it has already been shown how the increase in the number of centers can improve the quality of the results [61, 172, 185]. An additional strategy includes the performance of longer simulations for each center to obtain a better convergence of all the mean forces.

5.2.3 Structural Features of the Cldn5 Suzuki Model

The investigation of the TJ structural features in the Blood Brain Barrier (BBB) is an important topic in Neuroscience. In particular, Cldn5 complexes in the BBB are important gatekeeper of the paracellular environment. In this subsection, we discuss the possibility of describing paracellular Cldn5 pores with the topology of the Cldn15 Suzuki model, in agreement with the current experimental knowledge. Only very recently, in [12], Krystofiak et al. observed *directly* a double-row arrangement for Cldn11 strands, using amorphous carbon replicas and Phase-contrast Electron Microscopy. Dimensions of these Cldn11 fibrils are consistent with the original indirect observations that led to the introduction of the Cldn15 strand model proposed by Suzuki et al. [1]. This result considerably supports the hypothesis that Cldn complexes belonging to different tissues follow a common pattern while the difference in selectivity among different tissues are produced by the different amino-acid composition of the Cldn ECL domain in the paracellular environment. Here, we investigated this idea, performing Molecular Modeling to build refined Cldn5 Suzuki-like models.

In **Fig 5.20**, the filters of the Cldn15 and Cldn5 models are shown. Interestingly, the structure of the Cldn5 model presents a filter with four uncharged Glutamine residues (Q57) which could dramatically change the transport properties through the cavity, with respect to the Cldn15 D55 cage. Remarkably, this Glutamine is conserved among the Cldns responsible for the formation of barrier-like TJs (see **Fig 1.10** from [1]), and not among other members that form paracellular channels. Furthermore, the additional pore lining Cldn15 D64 residues that increase the repulsion among equally charged residues are replaced by the hydrophobic Valine (Cldn5 V66) residues. The absence in the Cldn5 pore space of charged residues could induce an

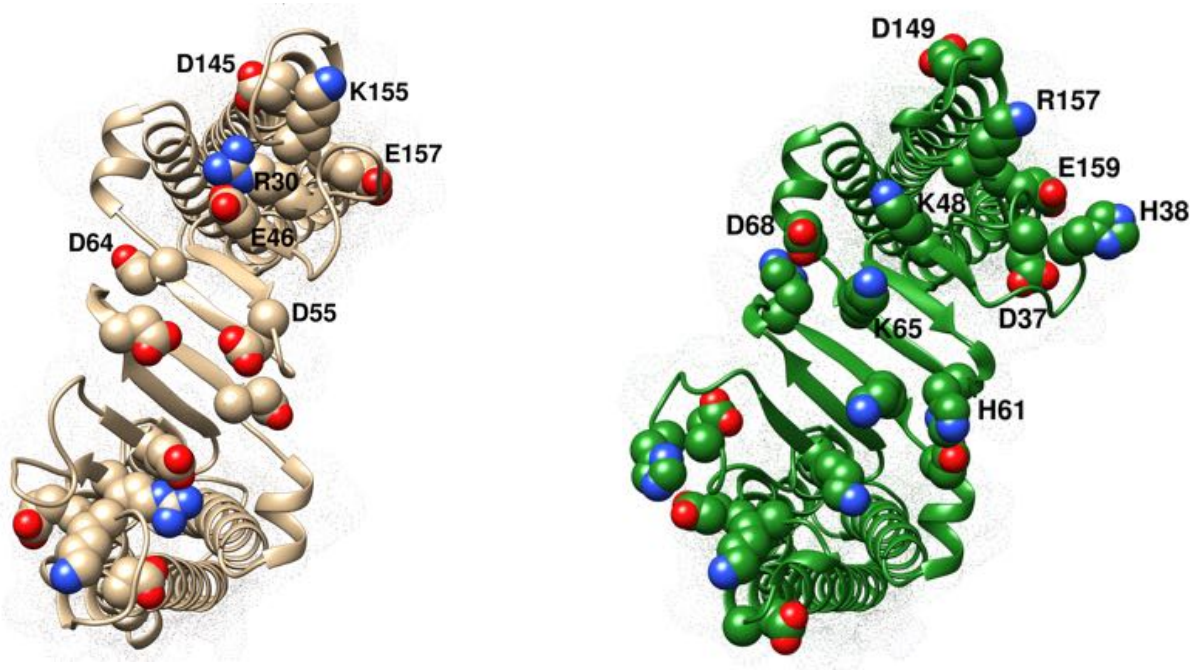


Fig 5.21. Surface characterization of the *cis face-to-face* interaction for Cldn15 and Cldn5 dimers. Cut-away view of the Cldn-based complex for Cldn15 (brown model) and Cldn5 (green model) proteins. Only two proteins forming the *cis face-to-face* interaction are shown. Charged residues located in the cavity are shown with VdW spheres.

occlusion of the intercellular space because of the absence of the electrostatic repulsion. This observation will be tested with future explicit MD simulations in double bilayer, as described in the future perspectives, in **Chapter 6**.

These preliminary topological results agree with the observation that the BBB environment does not contain paracellular channels [10], but, at most, small pore cavities, named *tightnesses*, due to the encounter of uncharged residues belonging to the ECL1 domain [34], as already discussed in **Subsection 1.4.1**. Therefore in our Cldn5 model, the tight configuration of the pore is supported by a different distribution of uncharged residues located in the selective filter. The occluded space of the BBB paracellular region is also known to act as a barrier for cations [35]. This physiological behavior might generate from the specific distribution of positively charged residues (K48 and K65) in the most external part of the Cldn5 cavity (**Fig 5.21**).

To further validate the Cldn5 model, additional Gaussian Network Modeling (GNM) and structural analysis were performed, using the DynOmics server <http://enm.pitt.edu/Tutorial.php>. The GNM approach is a minimalist, coarse-grained approach to study biological molecules. In the GNM model, proteins are represented by nodes corresponding to α -carbons of the amino-acid residues. This representation makes the calculations computationally inexpensive and, therefore, faster than the all-atom MD strategy. Using the two slowest GNM modes, it is possible to calculate the Potential Functional Sites (PFSs) of the protein complex. This analysis for the Cldn15 model recapitulates the importance of the C_{α} position of the D55 residue (**Fig 5.22**, brown model), detected for all the four protomers, together with two additional uncharged residues (V59 and W63). When applied to Cldn5, the algorithm highlights a rich family of residues in the cavity (**Fig 5.22**, green model), such as:

- the polar residues (T42, Q44, Q57, S58) that could facilitate the formation of HBs with external molecules; and
- the positively charged (H61 and K65) residues, that could be responsible for the barrier

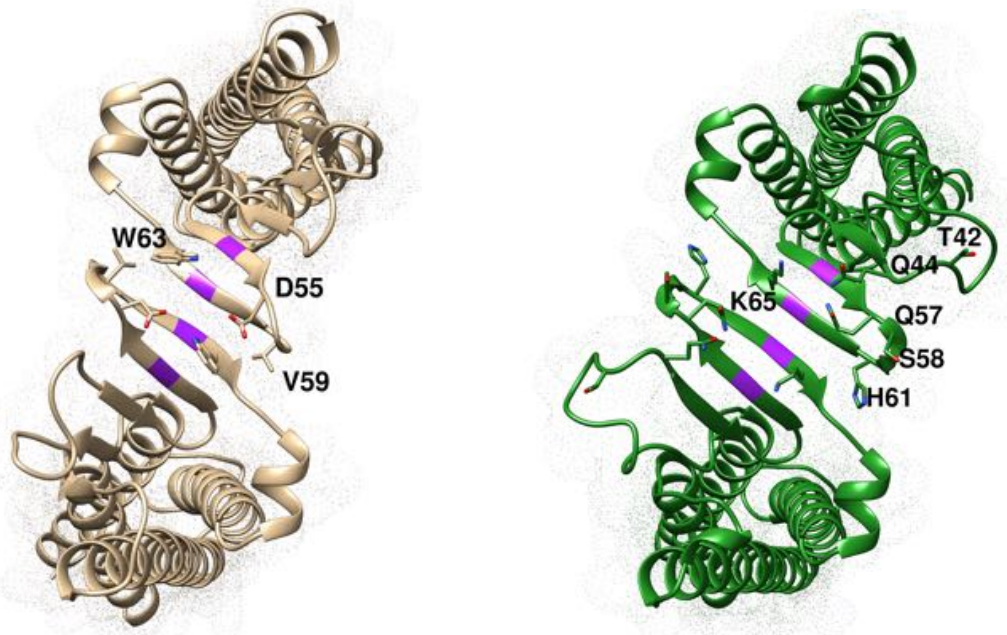


Fig 5.22. Potential Functional Sites for *cis- face-to-face* Cldn dimers. Potential Functional Sites for *cis- face-to-face* for the Cldn15 brown dimer (D55, V59, W63) and for the Cldn5 green dimer (T42, Q44, Q57, S58, H61, K65). Residues are shown in stick style for clarity. The potential conserved cysteine bridge is highlighted in purple.

action against cations.

Globally, these preliminary observations point to a validity of the Suzuki-based model also for Cldn5. Interestingly, other authors have recently drawn similar conclusions. In [237] CG simulations together with Molecular Docking calculations proved the transferability of the Suzuki model to a group of Classic Cldns.

Chapter 6

Conclusions and Future Perspectives

**Più veloci di aquile i miei sogni
Attraversano il mare.**

Francesco Battiato / Manlio Sgalambro, *La Cura*.

6.1 Further Computational Validations

This dissertation examines the dynamics of a paracellular channel model [1] *via* the application of Molecular Modeling and Molecular Dynamics techniques [7, 8]. Missing segments were successfully added to the original version of the structure, avoiding steric clashes and distortions. Further MD simulations relaxed the configuration in a double bilayer environment. Moreover, thermodynamic and kinetic calculations demonstrate the validity of the charge selectivity of the system.

The main model introduced in this Thesis describes a paracellular channel made by four Cldn15 protomers. In particular, the charge selectivity for the single Cldn15 paracellular channel is formed by a cage of four pivotal D55 residues (one for each monomer), placed in the narrow region of the filter. Globally, this all-atom model is stable in long MD simulations. Furthermore, our results reproduce the correct selectivity of the system. However, future efforts are still required to further investigate the general validity of this model. They include:

1. the study of the effects of multi-ionic permeations;
2. the introduction of other proteins belonging to the junction in the model.

Almost all of these efforts will depend on an ambitious investment of computational resources. By extension, our procedures can be applied to other Cldn based systems, such as the Cldn5 complexes belonging to the Blood Brain Barrier (BBB). In the Cldn5 case, the amino-acid composition of the pore cavity is different, as suggested by the comparison with the Cldn15 model in **Fig 5.21**. Consistently with the experimental literature, where the Cldn5 complexes are described as biological barriers, the Suzuki model for Cldn5 proteins is characterized by an uncharged cage in the pore region, where the Cldn15 D55 residue is replaced by the Glutamine Q57. The study of this Cldn5 structural complex could be useful to provide a solid model for the paracellular BBB environment and to guide the rational design of new molecules able to permeate the barrier in a safe and reversible manner.

To this aim, the next step of this project will be the simulation of ionic permeation through the Suzuki-like assembly of Cldn5 monomers. These results will be pivotal to understand the feasibility of this model for Cldn-based complexes belonging to other tissues.

Furthermore, the results of these new simulations will be compared with those of other models, such as an alternative configuration of a Cldn5 pore, introduced in [49, 50], which is based on *cis*- interactions between TM helices and not by the *cis*- *face-to-face* interactions.

6.2 Future Computational Work and Experiments

In addition to further computational validations, the results illustrated in this Thesis could be helpful to guide experiments for the evaluation of molecular permeation through biological tissues.

The selectivity properties of the paracellular complexes, which are critical in protecting the internal environment, could be extremely promising in being modulated by artificial molecules, in a non-disruptive manner. So far, however, this has been hampered by lack of a detailed molecular description of the TJs.

Indeed, the work presented in this Thesis is part of a long-term project in collaboration with experimentalists, whose final goal is the rational design of drugs and peptides able to cross the paracellular environment. Several efforts have already been made to obtain *in vitro* effective vectors capable of passing paracellular barriers, but the structural details of the permeation mechanisms remain elusive [259–263]. Our efforts led to the introduction of a structural model that describes all the main features of intestinal paracellular channels and that will be exploited to investigate the paracellular environment in the Blood Brain Barrier.

On the experimental side, an *in-vitro* model of the biological barriers has been developed at the IIT-NSYN laboratory. This model is based on a transwell system and allows studying transport of various substances or nanoparticles, as well as live measurements of the tissue integrity and functionality by imaging techniques.

A library of peptides bound to optogenetic probes [264] will be developed to tune the paracellular permeability by perturbing Cldn-Cldn interactions. To this aim, computational studies will include protein-peptide docking simulations. In recent years, a variety of approaches have been developed for protein-peptide docking [265, 266], that is, predicting the structure of a protein-peptide complex, including variable degrees of information about the peptide binding site.

6.3 Thesis Impact

In this Thesis, the structure, dynamics and function of a member of the unconventional paracellular channel class were investigated. These systems perform vital functions and they are a family of potential targets for drug delivery. The main achievements of the work are the following:

1. Molecular Modeling and MD simulations techniques allowed the refinement of the starting configuration of the Cldn15 Suzuki model and the evolution of the pore size. The refined model is consistent with available experimental results, in terms of pore size, charge selectivity and water transport.
2. The thermodynamic and kinetic features of ionic transport by the Cldn15 channel were efficiently quantified, using a single CV. The PMF and MFPT of ions are sensitive to a selective filter made by four D55 residues. In particular, the model allows the flux of single cations, without appreciable selectivity mechanisms between Na^+ and K^+ . On the contrary, the passage of anions is prevented. Further multi-CVs TAMD-SSw simulations were performed and they confirm the main results of the single CV calculations.
3. Finally, the use of Molecular Modeling and low resolution Coarse-Grained (CG) simulations provided a first description of other Cldn5 complexes based on the same 3D-model proposed by Suzuki et al. This should help to discuss the possible application of this structural model to other Cldn complexes. In particular, using Homology Modeling, a family of BBB Cldn5 complexes were introduced.

Examination of the distribution of the charged residues in the internal cavity and analysis of the residues of the selective filter are in agreement with the features of the BBB paracellular environment, which include the absence of paracellular channels made by charged selectivity filters and a distribution of positively charges at the mouth of the cavities able to form cation barriers.

Chapter 7

Further Projects

Homo sum, humani nihil a me alienum puto.

Publius Terentius Afer, *Heautontimorumenos*.

In the context of a series of projects of the IIT-NSYN laboratory, we produced a database of 3D protein structures belonging to the family of Voltage Gated Sodium (NaV) Channels, a group of proteins that are of central importance for the generation and transmission of action potentials in excitable cells.

As an illustration, genetic studies on patients with epilepsy have identified more than 700 mutations belonging to NaV proteins, attesting to their role in the pathogenesis [267].

The final aim of this work is the introduction of suitable models able to improve the comprehension of the structural features of NaV channels.

In particular these models are used in the following projects:

1. structural modeling of the interactions between NaV channels and the Proline-Rich Transmembrane Protein 2 (PRRT2) [268]. Our lab described for the first time how PRRT2 modulates NaV channels activity [269] and on-going efforts include the introduction of 3D NaV-PRRT2 models.
2. elucidation of the Na⁺ permeation features of NaV channels, performing advanced MD simulations. These efforts are currently underway in collaboration with Prof. Cameron F. Abrams (Drexel University, Philadelphia, USA) and Dr. Alexis Paz (Universidad Nacional de Córdoba, Córdoba, Argentina).

Our databases includes structures of the human NaV (hNaV) proteins which are mostly expressed in the Central Nervous System (CNS) such as NaV1.1, NaV1.2, NaV1.6. The development of these models has benefited from the recent publication of a series of eukaryotic NaV structures through the use of Single Particle Cryo Electron Microscopy (see **Table 7.1** for a list of these proteins).

In the following,

- **Section 7.1** summarizes the main functional and structural properties of NaV channels and their role mediating Na⁺ influx in excitable cells;
- **Section 7.2** introduces all the methods and protocols followed for the production of the models.
- **Section 7.3** introduces the first results of the ongoing collaboration with Prof. Cameron F. Abrams and Dr. Alexis Paz, to perform MD simulations of Na⁺ permeation, through a NaV1.2 model.
To this aim, a TAMD-based PMF reconstruction, named Temperature Accelerated MD/On-The-Fly Parameterization (TAMD/OTFP) [2–6] is used.

7.1 An Introduction to the Structure-Function Features of Voltage Gated Sodium Channel

Table 7.1. List of the published experimental structures of eukaryotic Voltage Gated Sodium Channels. List of experimental structures of eukaryotic Voltage Gated Sodium Channels, obtained with Single Particle Cryo Electron Microscopy. The atomic resolution spans from 2.6 to 4.2 Å.

PDB ID [Reference]	Organism and NaV type	Configuration
5X0M [270]	Periplaneta Americana NaVPas	<i>apo</i> (closed pore)
5XSY [271]	Electrophorus Electricus NaV1.4	<i>apo</i> (open pore)
6A90, 6A91, 6A95 [272]	Periplaneta Americana NaVPas	Ligand (Toxin)-bound
6AGF [273]	Human NaV1.4	<i>apo</i> (open pore?)
6N4R, 6N4Q, 6N4I [274]	Human NaV1.7	Ligand-bound
6NT3, 6NT4 [275]	Human-cockroach Hybrid NaV channel	Ligand-bound
6J8J, 6J8I, 6J8H, 6J8G [276]	Human NaV1.7	Ligand-bound
6J8E [277]	Human NaV1.2	Ligand-bound
6UZ3, 6UZ0 [278]	Rat NaV1.5	Ligand-bound

Ion permeation in NaV channels. NaV channels belong to a wide family of proteins capable of spanning the cell membrane, and responding to changes in the electrical properties of the membrane to generate a transmembrane potential.

To perform their function, these proteins undergo complex conformational transitions in response to changes of membrane potentials [279].

Normally, the inner portion of the membrane is at a negative voltage (resting phase) and, when the voltage increases (depolarization phase) because of the production of an *action potential*, NaV channels begin to allow Na⁺ ions to cross the membrane [279].

These proteins were first characterized by Hodgkin and Huxley in a series of seminal papers published in 1952, where the authors described different mechanisms such as voltage-dependent activation, fast inactivation, and selectivity over other ions [280–285].

Subsequently, before the determination of ion channel sequences or structures, Hille [280, 286] was able to construct a structural model of the narrowest constriction in NaV channels, the selectivity filter (SF), predicting a rectangular SF of dimensions 3 × 5 Å.

After many years of research, scientists discovered that different Na⁺ channels have a different aminoacidic composition of the SFs, with rings of facing residues that can be EEEE, EKEE, EEKE, DKEA, or DEKA [279].

The last motif, in particular, is responsible for the selectivity filter in Mammalian NaV channels. However, the fine details of the conduction mechanisms produced by these signature sequences are still unclear.

Structure and gating mechanisms of NaV channels. The eukaryotic NaV channels consist of a core α subunit, which is sufficient for voltage-gated ion permeation, and auxiliary β subunits [287].

There are nine subtypes of α subunit in human, encoded by genes SCNXA ($X = 1-9$ for NaV1.1-1.9, respectively) [288, 289]. NaV1.1-1.3 and NaV1.6 are the primary subtypes in the central nervous system (CNS), NaV1.7-1.9 mainly function in the peripheral nervous system, NaV1.5 is predominantly expressed in the heart, and NaV1.4 is the prevailing subtype in the skeletal muscle.

Remarkably, all NaV α subunits share identical topology formed by a single polypeptide chain folds to four repeats (repeats I-IV), each containing six transmembrane segments (TMs) (S1-S6) (see **Fig 7.2** for a schematic representation).

As all the other Voltage-gated cation ion channels, each repeat of a NaV channel consists of

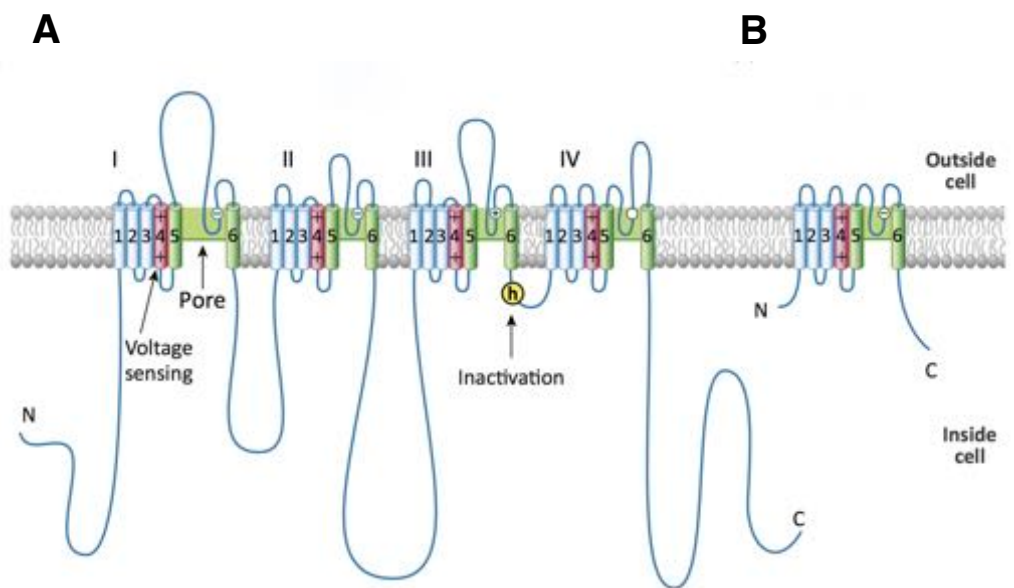


Fig 7.1. 2D schematic map of the NaV α subunit. **A** The typical α subunit of a Mammalian NaV channel is illustrated as a transmembrane folding diagram in which cylinders represent transmembrane α helices and lines represent loop regions, in proportion to their length. The Roman numerals indicate the four homologous repeats and the Arabic numerals are used to label the six transmembrane helices of each repeat. The S4 helices are colored in red with + signs indicating gating charges. The S5-S6 helices are colored in green and the small white circles indicate key residues in the selectivity filter with + and - signs indicating their charge states. The yellow circle with an h indicates the inactivation gate. **B** Schematic map of a bacterial channel, which contains all the minimal functional elements of a single mammalian NaV domain. In each panel, the N and C letters indicate the N- terminus and the C- terminus, respectively. Adapted and reprinted from [18], [https://www.cell.com/trends/biochemical-sciences/fulltext/S0968-0004\(15\)00121-8](https://www.cell.com/trends/biochemical-sciences/fulltext/S0968-0004(15)00121-8). Copyright (2015), with permission from Elsevier.

two structurally and functionally distinct modules, namely, the voltage-sensing domains (VSDs) formed by S1-S4 in each repeat and the ion-conducting pore domain enclosed by S5 and S6. The S5-S6 domain of each repeat arranges in a pore forming conformation which is opened by specific changes of membrane potentials.

The voltage sensors are featured by repetitively occurring K or R residues at every third position in S4. Voltage-dependent activation is understood in terms of the outward movement of these S4 gating charges, that are stabilized in their transmembrane position by interactions with negative charges and hydrophilic moieties in neighbouring transmembrane segments of the sensor. This voltage sensor activation is coupled to pore opening in the central cavity [290–292]. After the pore opening and before the return of the channel to the starting resting configuration, NaV channels pass to a so called *inactive* configuration within 1-2 ms. This *fast inactivation* process is required for repetitive firing of action potentials in neural circuits and for control of excitability in nerve and muscle cells. The inactivation gate is located in the intracellular loop linker between Repeat III and IV.

7.2 Modeling of NaV Channels

Modeling of NaV channels in a Closed State. The American cockroach NaVPaS channel (PDB ID: 5X0M), which is the first experimental structure of a eukaryotic NaV channel, has been solved in a closed state [270].

NavPas shares 36-43% sequence identities with human NaV1.1-1.9, contains the signature DEKA residues of the selectivity filter and highly conserved transmembrane segments. There are missing segments of the structure at the intracellular side, including the N-terminal region, the I-II linker and II-III linker, and the C-terminal region.

The sequences of the human NaV 1.1, 1.2 and 1.6 ion channels were obtained from the website <http://www.uniprot.org>. The four transmembrane (TM) repeats of each channel were selected for the structural modeling, neglecting all the disordered long cytoplasmic regions that connect the different repeats.

The model of each channel was constructed in two steps, following the approach of [293]. First, each of the four repeats was modeled using the protein prediction algorithm I-TASSER (Iterative Threading ASSEMBly Refinement) [294–296], a software with a hierarchical approach for the prediction of protein structures. In this case, I-TASSER was preferred to other softwares that performs pure homology modeling, since the sequence homology between the template and the target is good but not excellent (< 50%). In all the modeling procedures I-TASSER uses the NaVPaS channel as the main template. Accuracy of individual domain models can be reliably estimated by the confidence score (C-score), which has a high correlation to the actual TM-score, a widely used measure of similarity between two protein structures with values in the [0,1] range. Unlike RMSD (root-mean-square-deviation) score, TM-score is insensitive to local variation and protein size. In general, a TM-score > 0.5 indicates high similarity between two protein structures. Estimated TM-scores based on C-scores for our models of (domains I; II, III, IV) were (0.78; 0.87; 0.90; 0.93) for Nav1.1, (0.74; 0.88; 0.90; 0.92) for Nav1.2, (0.93; 0.90; 0.92; 0.92) for Nav1.6, indicating high similar folding structures to those of the template selected. Then, the whole model for each channel was built by structurally aligning the four individual repeats models to the global structural template. Each single domain was aligned to the corresponding domain of the template by the structural alignment algorithm TM-align [297]. Finally, the model was refined with the FG-MD software [298]. FG-MD is a molecular dynamics (MD) based algorithm for atomic-level protein structure refinement. Given an initial protein structure, FG-MD first identifies analogous fragments from the PDB by the structural alignment program TM-align. Spatial restraints extracted from the fragments are then used to re-shape the funnel of the MD energy landscape and guide the MD conformational sampling. FG-MD aims to refine the initial models closer to the native structure. It can also improve the local geometry of the structures by removing the steric clashes and improving the torsion angle and the hydrogen-binding networks.

Modeling of human NaV channels in an open state. The structure of the human NaV1.4 (hNaV1.4) channel (PDB ID: 6AGF), which is the first experimental structure of a human NaV channel, has been solved in an open state with activated (also named "up") configurations for all the four VSDs [273]. hNaV1.4 shares 80-90% sequence identities with each repeat of the other human NaV proteins.

In addition to the signature DEKA residues of the selectivity filter and highly conserved transmembrane segments, it also contains the important linker between the third and fourth domain (III-IV), responsible for the fast inactivation process.

The transmembrane (TM) sequences of the repeats of the human NaV1.1, 1.2 and 1.6 channels were selected for the structural modeling, including the III-IV linker.

The first, the second and the third linked to the fourth repeat were modeled using the protein prediction algorithm SWISS [217], which implements a rigid fragment assembly approach. In this case, SWISS was preferred to I-TASSER, since the sequence homology between the template and the target is more than 80%. Finally, each channel model was built by structurally aligning the four individual repeats models to the global structural template with the MATCHMAKER tool of the UCSF Chimera suite and, finally, refined with FG-MD code.

A representative NaV1.2 model is shown in **Fig 7.2**.

7.3 MD Simulations of the NaV1.2 Channel in an Open State

MD simulations of the human NaV1.2 channel in an open state. The model of the hNav1.2 channel, which is one of the main NaV subtypes in the brain, was selected to perform MD simulations to elucidate the mechanisms of Na⁺ permeation.

The final aim is the elucidation of the thermodynamics of the process, studying multi-dimensional PMFs, also named *Free Energy Surfaces*, FESs.



Fig 7.2. Extracellular view of a human NaV structural model. Ribbon style representation of the Nav1.2 structural model. Repeat I is pictured in gray; repeat II, in yellow; repeat III in green; the linker between repeat III and IV in orange; repeat IV, in blue. The four repeats are assembled in a clockwise order. The pore fenestration is shown with the red ellipse.

To this aim, one- and two- dimensional FES profiles will be reconstructed using a TAMD-based method, named TAMD/on-the-fly parameterization (TAMD/OTFP) [2–6].

To calculate the FESs, TAMD/OTFP uses the same variational approach of the SSw method [61], minimizing the same error function (Eq. 3.79). The main difference between the two methods is the strategy for the calculation of the Mean Forces. In fact, the SSw method calculates the Mean Forces in different points of the Cartesian space obtained *previously* from a family of TAMD simulations, while the OTFP approach quantifies the FES *directly* during the TAMD simulation.

Calculation of these profiles for permeation of Na^+ through the Nav1.2 pore can improve the comprehension of this mechanism and, in particular, the roles of both the DEKA and vestibular EEDD rings of the selectivity filter.

After the beginning of the production of the both the MD and TAMD/OTFP simulations, the experimental structure of the hNav1.2 structure was solved using cryoEM (PDB ID: 6J8E, [277]). To fit the Cryo-EM map of hNav1.2, the authors used the atomic coordinates of hNav1.4 (PDB ID: 6AGF), obtaining a very similar conformation.

In the words of the authors [277]: *The conformation of Nav1.2 is nearly identical to that of Nav1.4 with a root-mean-square deviation (RMSD) of 0.696 Å, over 981 C_α atoms when the two structures are superimposed.* Therefore, because of the sharing of the same template, it is not surprising that our hNav1.2 model overlaps nicely with the Cryo hNav1.2.

Since we used the 6AGF structure as template for homology modeling, the experimentally derived conformation of hNav1.2 is highly similar to our model. Specifically, the RMSD for the backbone atoms of our Nav1.2 model and Cryo Nav1.2 structure is 1.19 Å for all residues of the whole system; and 0.82 Å for all residues of the pore domain only (excluding VSDs and intracellular loops).

The alignment between the structures at the selectivity filter region is quite remarkable, even for sidechains. **Fig 7.3** and **Fig 7.4** show the similarity of the outer EEDD and the inner DEKA filter motives, respectively, between the model and the Cryo structure.

Therefore, globally, the two structures can be considered equivalent. However, since small differences are known to propagate effects over the whole structure, we will investigate both of them in the project to identify possible differences.

MD simulations of the human Nav1.2 channel in an open state. The relaxed FG-MD model of the human Nav1.2 (hNav1.2) channel includes:

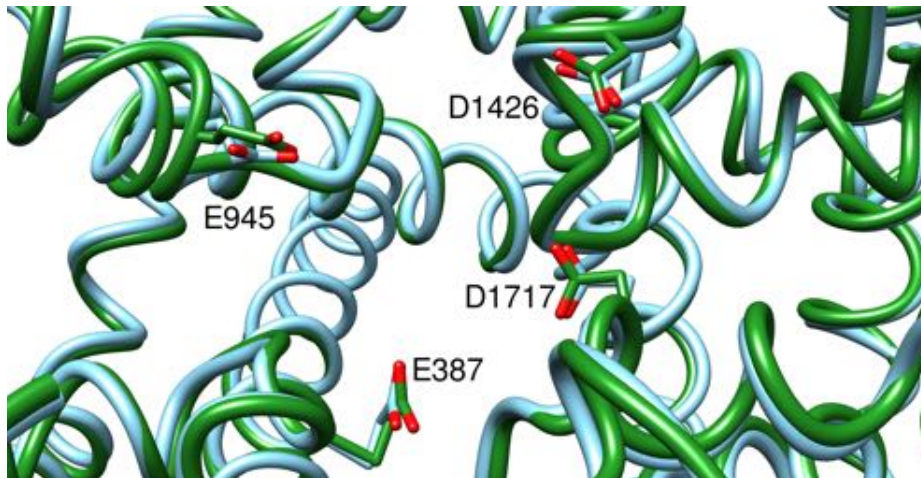


Fig 7.3. Superposition of the hNaV1.2 model and the experimental hNaV1.2 structure. The EEDD motif. Top view of the comparison between the hNaV1.2 model (template hNaV1.4, PDB ID: 6AGF, green ribbons) and the experimental hNaV1.2 structure (PDB ID: 6J8E, cyan ribbons). The conformation of the outer vestibular EEDD motif is highlighted. Each residue is labelled and side chains are shown in sticks.

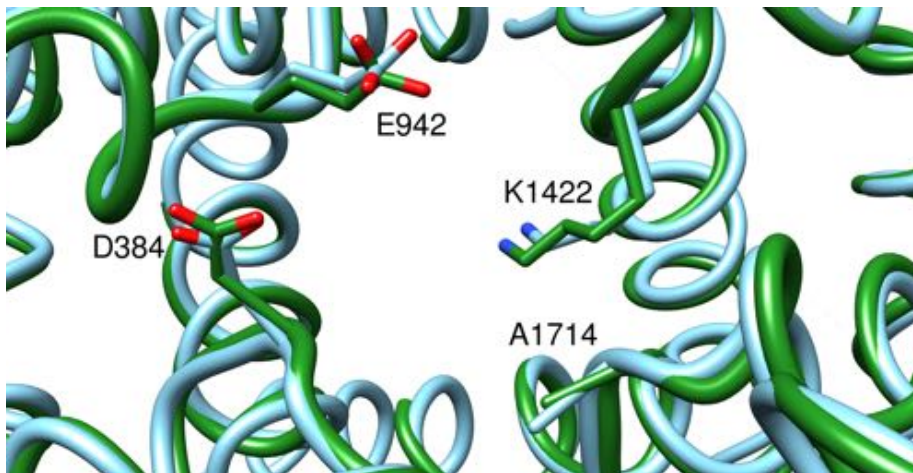


Fig 7.4. Superposition of the hNaV1.2 model and the experimental hNaV1.2 structure. The DEKA motif. Top view of the comparison between the hNaV1.2 model (template hNaV1.4, PDB ID: 6AGF, green ribbons) and the experimental hNaV1.2 structure (PDB ID: 6J8E, cyan ribbons). The conformation of the inner DEKA motif is highlighted. Each residue is labelled and side chains are shown in sticks.

1. residues belonging to Repeat I: from 125 to 284, and (linked to) from 306 to 427;
2. residues belonging to Repeat II: from 755 to 983;
3. residues belonging to Repeat III, the loop responsible for the Fast Inactivation and Repeat IV: from 1210 to 1775.

The structure was oriented using the PPM Server [220] in a 1-palmitoyl-2-oleoyl-SN-glycero-3-

phosphocholine (POPC) bilayer and solvated with three point (TIP3P) water molecules and 150 mM NaCl solution neutralizing the net charge of the system.

The membrane builder application of the CHARMM-GUI server [221] was used for the preparation of all the input files. Tetragonal PBCs were used to replicate the system. The system was pre-equilibrated following the preliminary CHARMM-GUI protocol. Then the system was equilibrated with a 35 ns long procedure in the *NPT* ensemble:

1. 5 ns with harmonic restraints for the heavy atoms of the protein,
2. 5 ns with restraints for the backbone of the protein,
3. 15 ns with restraints for the C_α atoms,
4. 10 ns with restraints for the C_α atoms of the α -helices.

The system was maintained by a Langevin thermostat and Nosé-Hoover Langevin piston pressure control [80,81]. The NAMD 2.12 program [13] with the modified CHARMM36 force field (CHARMM36m) [299] was used, including parameters of the ions [222].

Electrostatic and van der Waals interactions were calculated with a cutoff of 12 Å and the application of a smoothing decay starting to take effect at 10 Å. Long range electrostatic interactions were calculated using the Particle Mesh Ewald (PME) algorithm [64]. To ensure maximum accuracy, electrostatic and van der Waals interactions were computed at each simulation step. All covalent bonds involving hydrogen atoms (except those of water molecules) were kept fixed using SHAKE [71], and those in water molecules using SETTLE [73]. A time-step of 2 fs was employed.

A stable configuration of the channel, with two Na^+ ions located in the selectivity filter was selected as a starting configuration for FES calculations (**Fig 7.5**). The double ion occupancy is the most common configuration in the selectivity filter, as suggested by multi- μs fully atomistic simulations published in [300]. In this work, the authors studied the conduction mechanisms of a model of the hNaV1.2 channel, which was constructed by grafting residues of the mammalian NaV selectivity filter onto the high-resolution crystal structure of a NaV bacterial channel.

MD simulations of the human NaV1.2 channel in an open state (pore only). To investigate the main features of the ionic permeation, an equilibrated configuration of the pore region (which includes the S5-S6 segments of each repeat) was selected from the previous model after the first 30 ns of restrained equilibration.

Solvent molecules inside the cavity of the channel were kept, to preserve the sampling of the previous simulations. The benefit of a pore-only model, for MD simulations of Voltage Gated channels, is a widely used strategy. In this way, the system is less computationally expensive, while maintaining the main features that drive conduction in a specific conformation [3, 150, 240, 243, 246].

The pore-only model includes:

1. residues belonging to Repeat I, from 234 to 284 - and (linked to) - from 306 to 427;
2. residues belonging to Repeat II, from 866 to 983;
3. residues belonging to Repeat III, from 1320 to 1476;
4. residues belonging to Repeat IV, from 1641 to 1775.

Here the loop responsible for the fast inactivation events is not included. The selected configuration of the pore region was embedded in a POPC bilayer with the CHARMM-GUI server.

The membrane and the protein were solvated with explicit TIP3P water and the electrostatic charge of the full system was neutralized using a 150 mM NaCl solution.

The system contains 217291 atoms, for a volume of $135 \times 135 \times 128$ Å, in a rectangular box (**Fig 7.6**). As done for the full system, the modified CHARMM36 force field (CHARMM36m) was used together with the NAMD 2.12 code to perform MD simulations in the *NPT* ensemble, maintaining the pressure at 1 atm and the temperature at 310 K.

After the multi-step equilibration with gradually decreasing harmonic restraints on lipid

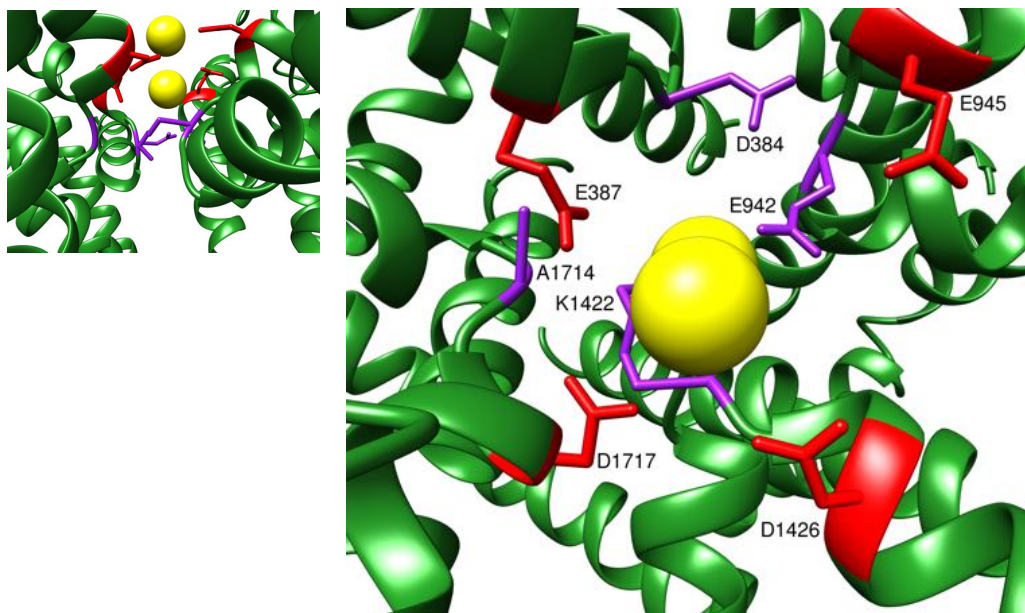


Fig 7.5. The representative configuration with two sodium ions in the selectivity filter. Top view of the representative pore structure, used as a starting configuration for thermodynamic calculations. NaV1.2 is described by green ribbons. Residues belonging to the EEDD motif are illustrated with red sticks and those belonging to the DEKA motif with purple sticks. Na⁺ ions are represented as yellow spheres. The small panel on the left shows the TM view of the configuration.

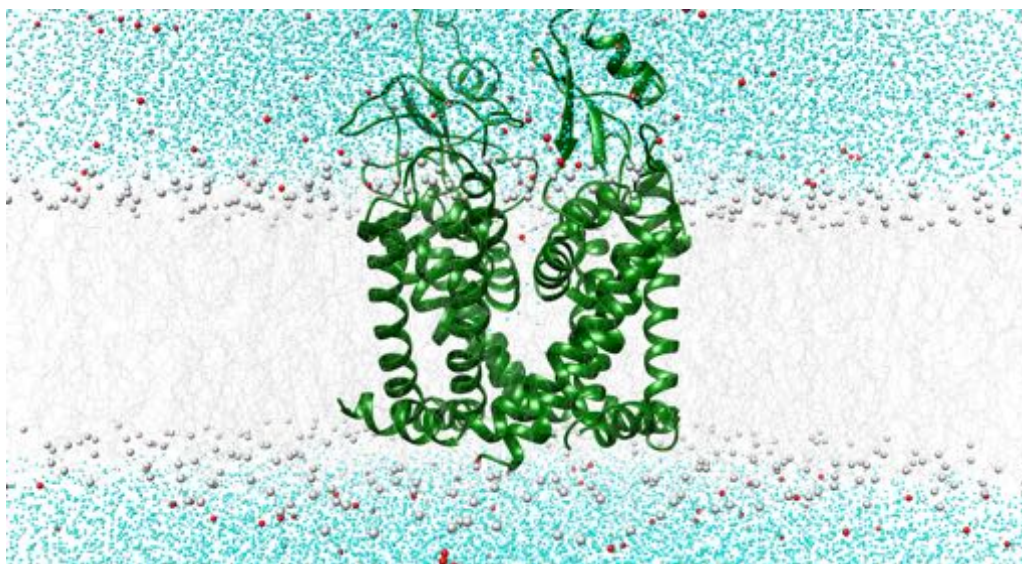


Fig 7.6. Transmembrane view of the hNaV1.2 pore. Ribbon style representation of the hNaV1.2 pore model (green). The POPC molecules are represented using lines and white spheres for the phosphorus atoms. Water molecules and ions are represented as cyan and red spheres, respectively.

molecules and protein atoms, following the CHARMM-GUI membrane builder equilibration protocol, and additional 25 ns restrained equilibration was performed to allow equilibration of the pore channel in the new membrane environment. The details of the protocol are identical

to those used for the simulation of the whole NaV1.2 channel. The equilibration procedure includes:

1. a preliminary step of ~ 12 ns with harmonic restraints for the heavy atoms of the protein;
2. further 5 ns with harmonic restraints for the protein C_α atoms;
3. 10 ns with harmonic restraints for the protein C_α atoms of the protein α helices, such as
 - residues 234-271 and 399-427 in Repeat 1,
 - residues 866-907 and 956-983 in Repeat 2,
 - residues 1321-1359 and 1447-1476 in Repeat 3,
 - residues 1643-1682 and 1749-1775 in Repeat 4.

While the comprehension of the molecular mechanisms of ionic permeation has been greatly improved in the last years thanks to MD Simulations for both mammalian KV channels [146,239–242] and Bacterial NaV channels [149–155,243–246], the study of the same process for mammalian NaV channels has been hampered because of the lack of experimental structures. In the following, we report the outcomes of simulations performed by Dr. Alexis Paz and Prof Cameron F. Abrams. Since they employ the homology model and system above described and generated by myself, and since they allow a first inspection of the ion permeation process that shows promising results, they are presented here.

Preliminary Results and Discussion. Our preliminary results, based on the use of the recently solved Cryo-electron structures of hNaV channels, are devoted to the search of an exhaustive set of CVs.

A first set of calculations is based on a single CV, *i.e.* the position of a single Na^+ ion along the z pore axis, with only one of the two Na^+ ions maintained inside the selectivity filter. Three different 45 ns simulations were produced. In **Fig 7.7**, we report, for each simulation, the FE profile of the Na^+ permeation through the pore. All the simulations shows a minimum in the inner cavity where the DEKA selectivity filter is located ($z = 4\text{-}10$ Å). However a net difference among the three profiles can be observed for the most external region ($z = 10\text{-}16$ Å). This demonstrates the presence of a clear sampling issue in the 1D simulations. The analysis of the CV trace, together with the the fluctuations of the center of MASS (COM) of the two motifs located in the filter (the EEDD and DEKA rings) show a clear correlation of the ion motion with the behavior of the K1422 belonging to the inner DEKA ring **Fig 7.8**. In particular, the K1422 residue seems to oscillate between two distinct positions, the *down configuration* at $z = 8$ and the *up configuration* at $z = 10$ Å, respectively.

In the simulation that returns the FE profile with lower barriers (~ 2 kcal/mol), named 1Da, **Fig 7.7**, K1422 is located almost all the time of the trajectory in $z = 8$ Å. In simulation 1Db, K1422 jumps to $z = 10$ in the middle of the trajectory and the CV ion changes its behavior, getting stuck in a lower position for a while, returning the larger FES profile for larger z value. This is consistent with the idea that K1422 is blocking the Na^+ ion to go up. In the last simulation, 1Dc, K1422 has a more dynamic behavior. The residue jumps repeatedly times between $z = 8$ and $z = 10$ Å, and as a consequence, the ion visit a lower position only 2 times (at ~ 15 and 30 ns of the simulation). In this case the FE profile shows and intermediate behavior.

These simulations demonstrate that the K residue acts in two distinct configurations, the first is a non-barrier configuration (the *down* state), while the second (the *up* state) performs a barrier function against the conduction of the cation. Therefore, the vertical position of the K1422 side-chain strongly affects the permeation of the cation, and hence an additional CV able to describe its dynamics must be included in further calculations.

Since the Na^+ permeation seems to be dominated by a two-ion process, further simulations were performed using, as CVs, the coordinate z of two Na^+ ions, located in the selective filter.

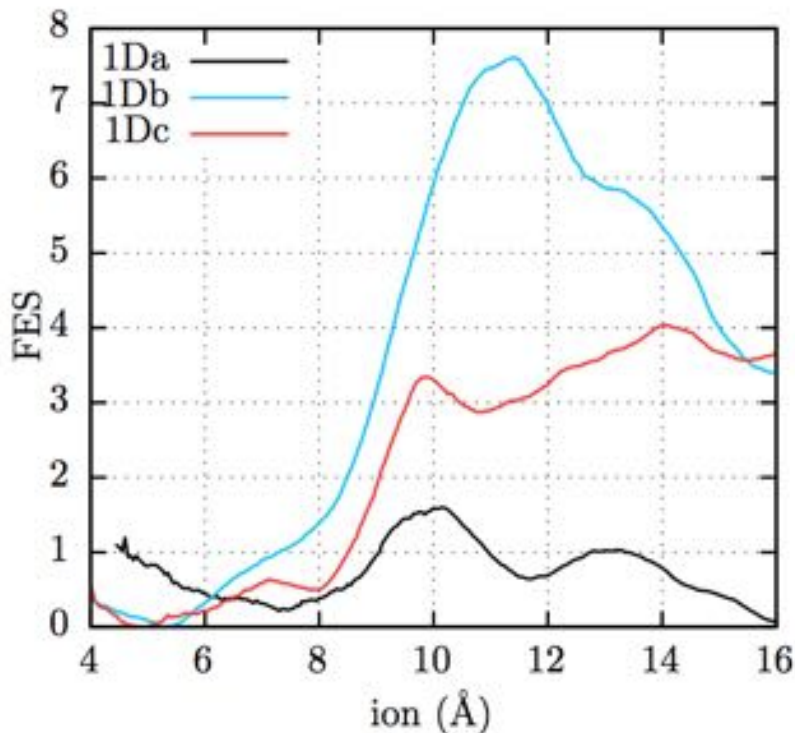


Fig 7.7. FESs from 1D-OTFP simulations. FESs (kcal/mol) obtained from three 1D-OTFP simulations (replicas 1Da, 1Db and 1Dc, respectively). Courtesy of Dr. Alexis Paz (Universidad Nacional de Córdoba, Córdoba, Argentina) and Prof. Cameron F. Abrams (Drexel University, Philadelphia, USA), ©.

The analysis of this two-ion configuration is important to understand, even in this case, how the presence of the K1422 can complicate the conduction mechanism.

For this set-up, three independent TAMD/OTFP were produced and the related 2D-FESs are shown in **Fig 7.9**.

For 2D-OTFP, the plots in **Fig 7.9** are presented in two different ways, for each independent run. One (on the left) is the direct way, which describes the original sampling of the simulation. The second, on the right, is a projection obtained with a symmetrization of the statistics collected by the simulations, to improve the quality of the plot. This procedure is justified by the fact that the original (partial) FESs are almost symmetric.

Although the three FESs are globally similar, some differences can be observed and explained by the behavior of the K1422 residue, as represented by the traces of **Fig 7.10**.

In particular, these FES projections reveal the relationship between the two cations and the z position of the K1422 ammonium group, demonstrating the conduction for 2-ion occupancies in the selectivity filter. In simulation 2Da, K1422 is always around the *down* position at $z = 8$ Å and it seems to prevent that both ions go below this value. In simulation 2Db, K1422 visit almost all the time the *up* position at $z = 12$ Å, allowing the passage of both the ions in the lower position. This explains the low free energy in the square below (8, 8). In the last simulation, 2Dc, despite K1422 remains in the down configuration, the two cations are able to explore the square, giving a high free energy when both the ions and the K1422 are all together below 10 Å.

To further justify the operation of symmetrization of the FESs, the statistics of the three 2D-OTFP was combined to obtain a unique FES, as illustrate in **Fig 7.11**, with the left panel. The associated symmetric plot (right panel of the same figure) is very similar and altogether this observation justifies the use of a symmetric operation to improve the quality of the FESs.

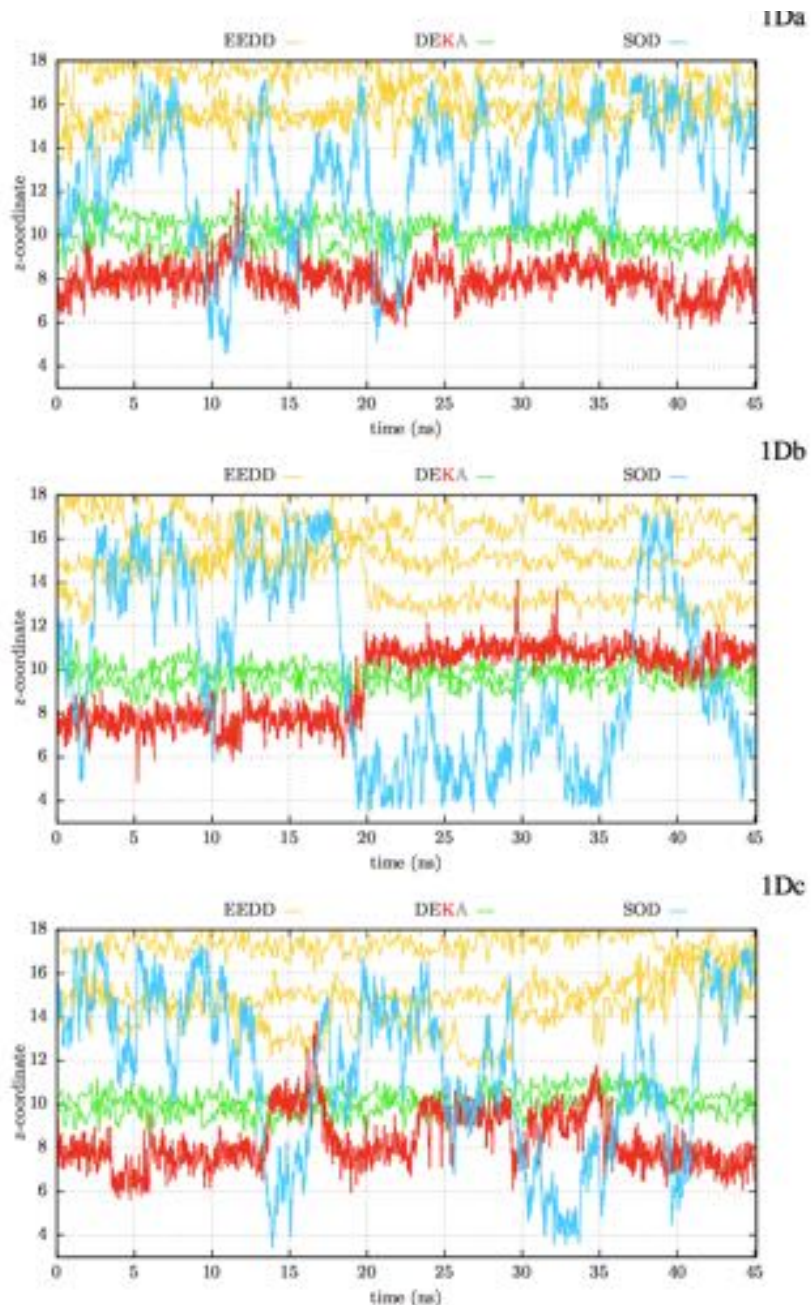


Fig 7.8. CV traces of the 1D-OTFP simulations. The CV traces for three different simulations of the previous figure, together with the z coordinate of the nitrogen atom of the K1422 side chain. The fluctuations of the center of mass of the EEDD and DEKA motifs are also included in the plot. The z -coordinate unit is in Å. Courtesy of Dr. Alexis Paz and Prof. Cameron F. Abrams, ©.

Conclusion and future perspectives. Our understanding of the structural features of mammalian NaV channels remained limited in the state of the art, because of the lack of experimental structures.

The recently solved Cryo-electron Microscopy NaV configurations (listed in **Table 7.1**) offer new invaluable data to study the complex mechanisms responsible for ion conduction in NaV channels.

Ionic permeation is a dynamic process that is likely to involve several many components such as

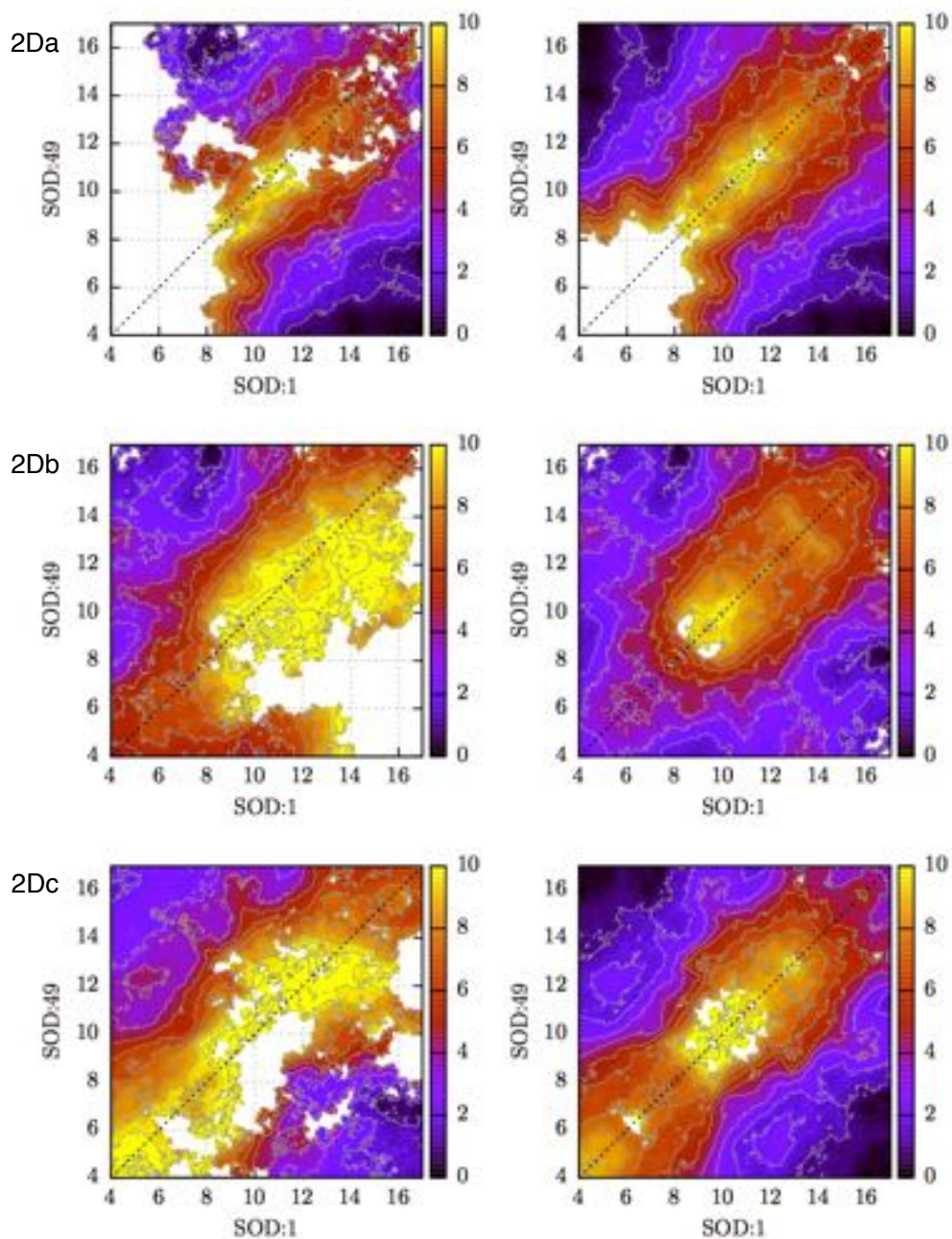


Fig 7.9. FESs from 2D-OTFP simulations. The FESs obtained from three different 2D-OTFP simulations (replicas 2Da, 2Db and 2Dc, respectively). The coordinate z of two Na^+ ions (labeled as SOD:1 and SOD:49) are used as CVs. The plots at the left correspond to the FESs as obtained from the simulation. The plots at the right are obtained after processing the collected statistics to obtain a symmetric CV space. Courtesy of Dr. Alexis Paz and Prof. Cameron F. Abrams, ©.

1. isomerization of the residues belonging to the SF, such as the K1422 residue belonging to the DEKA filter;
2. cooperation between different structural motifs such as the DEKA and the EEDD rings;
3. participation of multiple ions;
4. (partial) dehydration of the ionic shell.

A fair comprehension of this process requires a molecular description at the atomic resolution. Biased MD simulations could help the description of the ionic permeation, providing thermody-

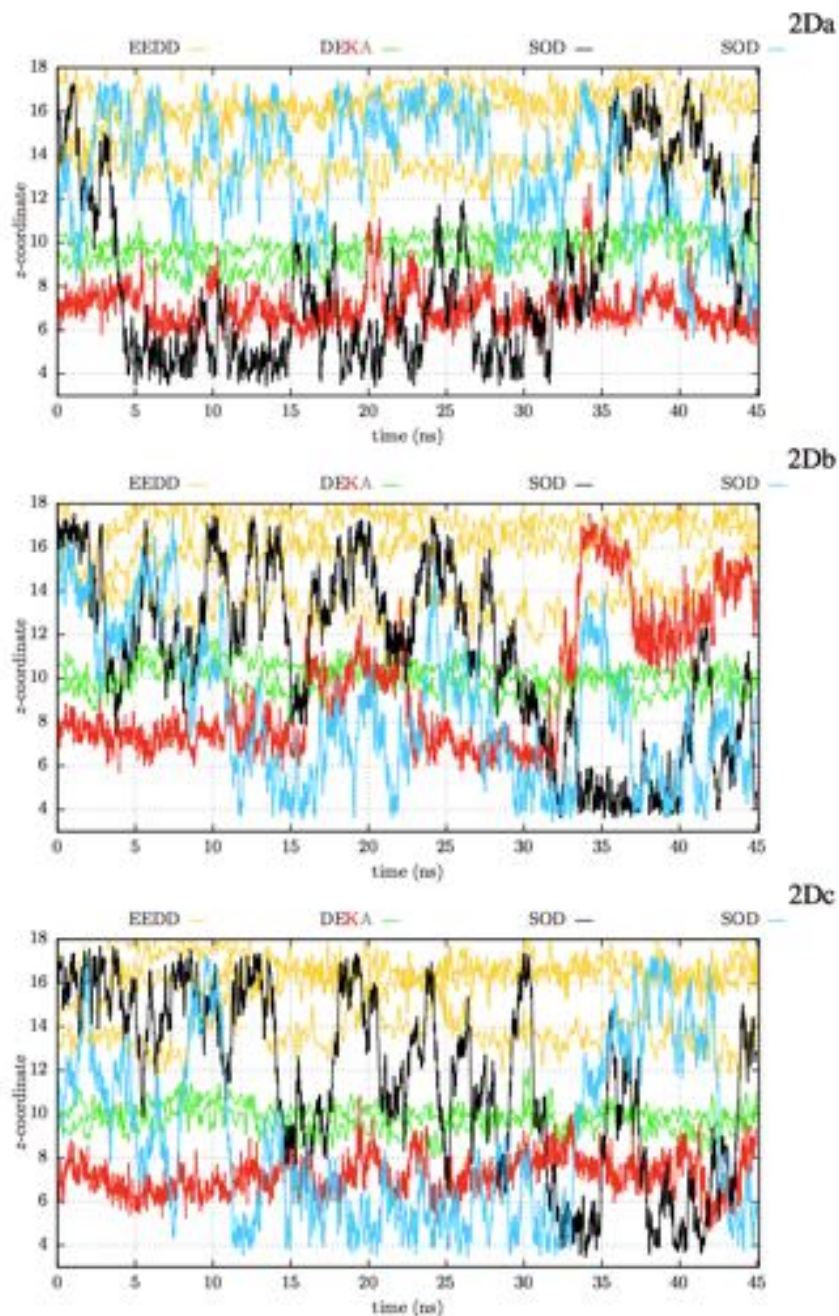


Fig 7.10. The CV traces of the 2D-OTFP simulations. The CV traces for the three different 2D-OTFP simulations of the previous figure, together with the z coordinate of the nitrogen atom of the K1422 side chain. The fluctuations of the center of mass of the EEDD and DEKA motifs are included in the plot. The z -coordinate unit is in Å. Courtesy of Dr. Alexis Paz and Prof. Cameron F. Abrams, ©.

namics calculations of the process.

In this project, structural modeling and TAMD based FES calculations, with on-the-fly parametrization (OTFP) were performed to elucidate the mechanism. Our starting efforts demonstrate that these structures, even if characterized by a low atomic resolution, are able to reproduce the correct features of the process, such as the main 2 ion mechanism and the pivotal regulating role of the Lysine belonging to the DEKA filter, as suggested by previous

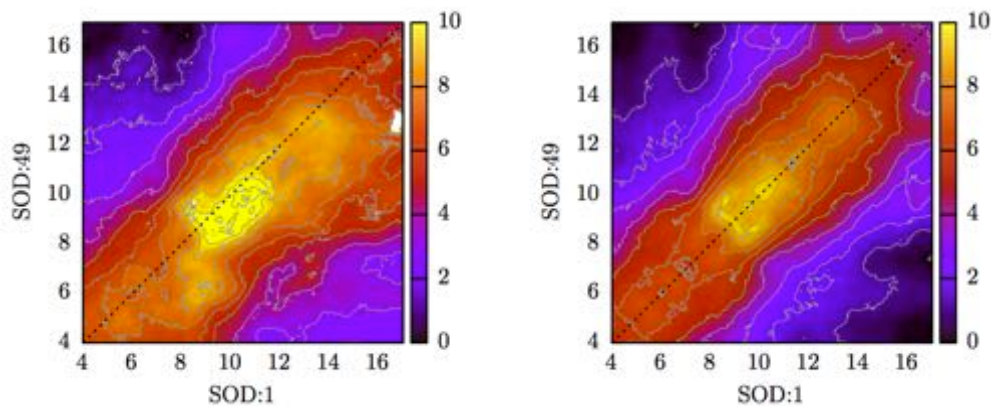


Fig 7.11. The cumulative CV traces of the three 2D-OTFP simulations. The left panel illustrates the FESs obtained from averaging together the statistics of the 3 different 2D-OTFP simulations used in **Fig 7.10**. The right plot takes into account the CV-space symmetry. The z -coordinate unit is in Å. Courtesy of Dr. Alexis Paz and Prof. Cameron F. Abrams, ©.

long unbiased MD simulations [300] performed on a chimera model based on a high resolution crystal structure of a bacterial NaV channel with a modified mammalian selectivity filter. Ongoing efforts are focused on MD simulations where the coordinate of the nitrogen atom of the K1422 side chain is introduced as an additional CV. Furthermore, since it is difficult to predict the exact pK_a shift of the Lysine residue in the microenvironment of the selectivity filter, the configuration with the neutral K1422 will be studied to understand the role of this residue in ion conduction. Finally, we will also study the effect on Na^+ permeation of mutating K1422 into E1422. This defect has been shown to induce loss of channel selectivity for ions and it is associated to a pathological state [301].

Appendices

Appendix A

Academic Transcript

Ph.D. students at the Neuroscience School of the University of Genoa are expected to obtain 40 credits in the ECTS (European Credit Transfer System) format during the 3-years long PhD course. I earned 42 credits within the second year, 58 in total. Most of the courses were taught in English.

12 ECTS - Courses of the CINECA Consortium for Automatic Computing.

- Introduction to Marconi KNL Cluster for Users and Developers, 27/03/2017. **1 Credit**
- Introduction to Scientific and Technical Computing in C, 05/05/2017. **3 Credits**
- 26th Summer School on Parallel Computing, 21/07/2017. **5 Credits**
- High Performance Bioinformatics, 15/12/2017. **3 Credits**

7 ECTS - Courses of the Center for Synaptic Neuroscience and Technology (IIT-NSYN Center).

- NSYN@UNIGE Journal Club based on [302], 23/11/2016. **1 Credit**
- Molecular Structure and Function at the Synapse, 08/11/2016. **3 Credits**
- Molecular Signaling and Adhesion at the Synapse 09/11/2017. **3 Credits**

12 ECTS - Courses of the Neuroscience and Brain Technologies Center (NBT Center).

- Fundamentals of Analog Processing from Basic Circuit Theory to Signal Manipulation, 29/11/2016. **4 Credits**
- Behavioral Genetics in Mice, 22/11/2017. **4 Credits**
- Molecular and Cellular Determinants and Neocortical Development, 24/11/2017. **2 Credits**
- Neuronal-Nanotechnology: a Systematic Review of the Bio-nano Interactions with the CNS, 23/11/2017. **2 Credits**

2 ECTS - Courses of the Nanochemistry Center (NACH Center).

- Basics of Crystallography and Diffraction by Crystals, 24/03/2017. **1 Credit**
- Introduction to Polymer Physical Chemistry and Characterization Techniques, 28/02/2019. **1 Credit**

25 ECTS - Courses of the University of Genoa.

- Synaptic Basis of Brain Pathology, 23/11/2016. **2 Credits**
- Corso Avanzato di Dinamica Molecolare di Proteine, 26/04/2017. **3 Credits**
- Introduction to Computer Programming to Researchers (together with the Robotics Brain and Cognitive Sciences IIT Center), 01/06/2017. **5 Credits**
- Nanophotonic Devices: from Fabrication to Application (together with the Robotics Brain and Cognitive Sciences IIT Center), 01/07/2019. **Settore scientifico disciplinare: FIS/07. 4 Credits**
- The Optical Microscope: Optics, Image Formation and Resolution (together with the Robotics Brain and Cognitive Sciences IIT Center), 05/07/2019. **Settore scientifico disciplinare: FIS/07. 3 Credits**
- Fluorescence Super-Resolution Microscopy: Basis, Applications and Perspectives (together with the Robotics Brain and Cognitive Sciences IIT Center), 27/06/2019. **Settore scientifico disciplinare: FIS/07. 3 Credits**
- Summer School of the School of Neuroscience of the University of Genova, from 10/09/19 to 12/09/19. **5 Credits**

Bibliography

1. Suzuki H, Tani K, Tamura A, Tsukita S, Fujiyoshi Y. Model for the Architecture of Claudin-Based Paracellular Ion Channels through Tight Junctions. *Journal of Molecular Biology*. 2015;427(2):291–297.
2. Abrams CF, Vanden-Eijnden E. On-the-fly Free Energy Parameterization via Temperature Accelerated Molecular Dynamics. *Chemical Physics Letters*. 2012;547:114–119.
3. Paz SA, Abrams CF. Free Energy and Hidden Barriers of the β -Sheet Structure of Prion Protein. *Journal of Chemical Theory and Computation*. 2015;11(10):5024–5034.
4. Paz SA, Vanden-Eijnden E, Abrams CF. Polymorphism at 129 Dictates Metastable Conformations of the Human Prion Protein N-terminal \hat{I}^2 -sheet. *Chemical Science*. 2017;8:1225–1232.
5. Paz S, Abrams C. Testing Convergence of Different Free-Energy Methods in a Simple Analytical System with Hidden Barriers. *Computation*. 2018;6:27.
6. Paz SA, Maragliano L, Abrams CF. Effect of Intercalated Water on Potassium Ion Transport through Kv1.2 Channels Studied via On-the-fly Free-energy Parameterization. *Journal of Chemical Theory and Computation*. 2018;14(5):2743–2750.
7. Alberini G, Benfenati F, Maragliano L. A Refined Model of Claudin-15 Tight Junction Paracellular Architecture by Molecular Dynamics Simulations. *PLoS One*. 2017;12(9):e0184190.
8. Alberini G, Benfenati F, Maragliano L. Molecular Dynamics Simulations of Ion Selectivity in a Claudin-15 Paracellular Channel. *The Journal of Physical Chemistry B*. 2018;122(48):10783–10792.
9. Fukuoka A, Yoshimoto T. Barrier Dysfunction in the Nasal Allergy. *Allergology International*. 2018;67:18–23.
10. Günzel D, Yu ASL. Claudins and the Modulation of Tight Junction Permeability. *Physiological Reviews*. 2013;93(2):525–569.
11. Tsukita S, Tanaka H, Tamura A. The Claudins: From Tight Junctions to Biological Systems. *Trends in Biochemical Sciences*. 2018;44(2):141–152.
12. Krystofiak E, Heymann J, Kachar B. Carbon Replicas Reveal Double Stranded Structure of Tight Junctions in Phase-contrast Electron Microscopy. *Communications Biology*. 2019;2(1):98.
13. Philips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. 2005;26(16):1781–1802.
14. Kasahara K, Sakuraba S, Fukuda I. Enhanced Sampling of Molecular Dynamics Simulations of a Polyalanine Octapeptide: Effects of the Periodic Boundary Conditions on Peptide Conformation. *The Journal of Physical Chemistry B*. 2018;122(9):2495–2503.
15. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-grained Protein Models and Their Applications. *Chemical Reviews*. 2016;116(14):7898–7936.
16. Bahar I, Lezon TR, Bakan A, Shrivastava IH. Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. *Chemical Reviews*. 2010;110(3):1463–1497.

17. Xu D, Zhang Y. Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-Step Atomic-Level Energy Minimization. *Biophysical Journal*. 2011;101:2525–2534.
18. Catterall WA, Zheng N. Deciphering Voltage-gated Na⁺ and Ca²⁺ Channels by Studying Prokaryotic Ancestors. *Trends in Biochemical Sciences*. 2015;40(9):526–534.
19. Blanco E, Shen H, Ferrari M. Principles of Nanoparticle Design for Overcoming Biological Barriers to Drug Delivery. *Nature Biotechnology*. 2015;33(9):941–951.
20. Furuse M, Fujita K, Hiiiragi T, Fujimoto K, Tsukita S. Claudin-1 and -2: Novel Integral Membrane Proteins Localizing at Tight Junctions with No Sequence Similarity to Occludin. *The Journal of Cell Biology*. 1998;141(7):1539–1550.
21. Anderson JM. Molecular Structure of Tight Junctions and their Role in Epithelial Transport. *Physiology*. 2001;16(3):126–130.
22. Farquhar MG, Palade GE. Junctional Complexes in Various Epithelia. *The Journal of Cell Biology*. 1963;17(2):375–412.
23. Staehelin LA. Further Observations on the Fine Structure of Freeze-Cleaved Tight Junctions. *Journal of Cell Science*. 1973;13(3):763–786.
24. Gumbiner B. Structure, Biochemistry, and Assembly of Epithelial Tight Junctions. *American Journal of Physiology–Cell Physiology*. 1987;253(6):C749–C758.
25. Powell DW. Barrier Function of Epithelia. *American Journal of Physiology–Gastrointestinal and Liver Physiology*. 1981;241(4):G275–G288.
26. Tang VW, Goodenough DA. Paracellular Ion Channel at the Tight Junction. *Biophysical Journal*. 2003;84(3):1660–1673.
27. Garcia V, Quiros M, Nusrat A. Intestinal Epithelial Claudins: Expression and Regulation in Homeostasis and Inflammation. *Annals of the New York Academy of Sciences*. 2017;1397:66–79.
28. Lal-Nag M, Morin PJ. The Claudins. *Genome Biology*. 2009;10(8):235.
29. Angelow S, Ahlstrom R, Yu ASL. Biology of Claudins. *American Journal of Physiology–Renal Physiology*. 2008;295(4):F867–F876.
30. Morita K, Sasaki H, Furuse M, Tsukita S. Endothelial Claudin. *The Journal of Cell Biology*. 1999;147(1):185–194.
31. Krause G, Winkler L, Mueller SL, Haseloff RF, Piontek J, Blasig IE. Structure and Function of Claudins. *Biochimica et Biophysica Acta*. 2008;1778(3):631–645.
32. Furuse M, Sasaki H, Fujimoto K, Tsukita S. A Single Gene Product, Claudin-1 or -2, Reconstitutes Tight Junction Strands and Recruits Occludin in Fibroblasts. *The Journal of Cell Biology*. 1998;143(2):391–401.
33. Rosenthal R, Milatz S, Krug SM, Oelrich B, Schulzke JD, Amasheh S, et al. Claudin-2, a Component of the Tight Junction, Forms a Paracellular Water Channel. *Journal of Cell Science*. 2010;123(11):1913–1921.
34. Krause G, Winkler L, Piehl C, Blasig I, Piontek J, L Müller S. Structure and Function of Extracellular Claudin Domains. *Annals of the New York Academy of Sciences*. 2009;1165:34–43.
35. Wen H, Watryand DD, Marcondes MCG, Fox HS. Selective Decrease in Paracellular Conductance of Tight Junctions: Role of the First Extracellular Domain of Claudin-5. *Molecular and Cellular Biology*. 2004;24(19):8408–8417.
36. Milatz S, Krug SM, Rosenthal R, Günzel D, Müller D, Schulzke JD, et al. Claudin-3 Acts as a Sealing Component of the Tight Junction for Ions of Either Charge and Uncharged Solutes. *Biochimica et Biophysica Acta*. 2010;1798(11):2048–2057.
37. Amasheh S, Schmidt T, Mahn M, Florian P, Mankertz J, Tavalali S, et al. Contribution of Claudin-5 to Barrier Properties in Tight Junctions of Epithelial Cells. *Cell and Tissue Research*. 2005;321(1):89–96.

38. Veshnyakova A, Krug SM, Mueller SL, Piontek J, Protze J, Fromm M, et al. Determinants Contributing to Claudin Ion Channel Formation. *Annals of the New York Academy of Sciences*. 2012;1257:45–53.
39. Hou J, Goodenough DA. Claudin-16 and Claudin-19 Function in the Thick Ascending Limb. *Current Opinion in Nephrology and Hypertension*. 2010;19(5):483–488.
40. Suzuki H, Nishizawa T, Tani K, Yamazaki Y, Tamura A, Ishitani R, et al. Crystal Structure of a Claudin Provides Insight into the Architecture of Tight Junctions. *Science*. 2014;344(6181):304–307.
41. Saitoh Y, Suzuki H, Tani K, Nishikawa K, Irie K, Ogura Y, et al. Structural Insight into Tight Junction Disassembly by Clostridium Perfringens Enterotoxin. *Science*. 2015;347(6223):775–778.
42. Shinoda T, Shinya N, Ito K, Ohsawa N, Terada T, Hirata K, et al. Structural Basis for Disruption of Claudin Assembly in Tight Junctions by an Enterotoxin. *Scientific Reports*. 2016;6(33632).
43. Nakamura S, Irie K, Tanaka H, Nishikawa K, Suzuki H, Saitoh Y, et al. Morphologic Determinant of Tight Junctions Revealed by Claudin-3 Structures. *Nature Communications*. 2019;10(1):816.
44. Vecchio AJ, Stroud RM. Claudin-9 Structures Reveal Mechanism for Toxin-induced Gut Barrier Breakdown. *Proceedings of the National Academy of Sciences*. 2019;PNAS Latest Articles(PNAS Latest Articles):1–8.
45. Colegio O, Van Itallie C, McCreah H, Rahner C, JM Anderson. Claudins Create Charge-selective Channels in the Paracellular Pathway between Epithelial Cells. *American Journal of Physiology - Cell Physiology*. 2002;283(1):C142–7.
46. Hou J, Rajagopal M, Yu ASL. Claudins and the Kidney. *Annual Review of Physiology*. 2013;75(1):479–501.
47. Piontek J, Winkler L, Wolburg H, Müller SL, Zuleger N, Piehl C, et al. Formation of Tight Junction: Determinants of Homophilic Interaction between Classic Claudins. *The FASEB Journal*. 2008;22(1):146–158.
48. Goliaei A, Adhikari U, Berkowitz ML. Opening of the Blood-Brain Barrier Tight Junction Due to Shock Wave Induced Bubble Collapse: A Molecular Dynamics Simulation Study. *ACS Chemical Neuroscience*. 2015;6(8):1296–1301.
49. Irudayanathan FJ, Trasatti JP, Karande P, Nangia S. Molecular Architecture of the Blood Brain Barrier Tight Junction Proteins—A Synergistic Computational and In Vitro Approach. *The Journal of Physical Chemistry B*. 2016;120(1):77–88.
50. Irudayanathan FJ, Wang N, Wang X, Nangia S. Architecture of the Paracellular Channels Formed by Claudins of the Blood–brain Barrier Tight Junctions. *Annals Of The New York Academy of Sciences*. 2017;1405:131–146.
51. Krause G, Protze J, Piontek J. Assembly and Function of Claudins: Structure–function Relationships Based on Homology Models and Crystal Structures. *Seminars in Cell and Developmental Biology*. 2015;42:3–12.
52. Colegio OR, Van Itallie CM, Rahner C, Anderson JSM. Claudin Extracellular Domains Determine Paracellular Charge Selectivity and Resistance but not Tight Junction Fibril Architecture. *American Journal of Physiology - Cell Physiology*. 2003;284 6:C1346–54.
53. Müller T, Gross H, Winkler H, Moor H. High Resolution Shadowing with Pure Carbon. *Ultramicroscopy*. 1985;16(3):340–348.
54. Van Itallie CM, Mitic LL, Anderson JM. Claudin-2 Forms Homodimers and is a Component of a High Molecular Weight Protein Complex. *Journal of Biological Chemistry*. 2011;286(5):3442–3450.
55. Rossa J, Ploeger C, Vorreiter F, Saleh T, Protze J, Günzel D, et al. Claudin-3 and Claudin-5 Protein Folding and Assembly into the Tight Junction are Controlled by Non-conserved Residues in the Transmembrane 3 (TM3) and Extracellular loop 2 (ECL2) segments. *Journal of Biological Chemistry*. 2014;289(11):7641–7653.

56. Abrams C, Bussi G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy*. 2014;16(1):163–199.
57. Torrie GM, Valleau JP. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-critical Lennard-Jones Fluid. *Chemical Physics Letters*. 1974;28(4):578–581.
58. Torrie GM, Valleau JP. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *Journal of Computational Physics*. 1977;23(2):187–199.
59. Kästner J. Umbrella Sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2011;1(6):932–942.
60. Vanden-Eijnden E, Venturoli M. Markovian Milestoning with Voronoi Tessellations. *The Journal of Chemical Physics*. 2009;130(19):194101.
61. Maragliano L, Vanden-Eijnden E. Single-sweep Methods for Free Energy Calculations. *The Journal of Chemical Physics*. 2008;128(18):184110.
62. Vanden-Eijnden E. Some Recent Techniques for Free Energy Calculations. *The Journal of Computational Chemistry*. 2009;30(11):1737–1747.
63. Fujisaki H, Moritsugu K, Matsunaga Y, Morishita T, Maragliano L. Extended Phase-Space Methods for Enhanced Sampling in Molecular Simulations: A Review. *Frontiers in Bioengineering and Biotechnology*. 2015;3(125).
64. Darden T, York D, Pedersen L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics*. 1993;98(12):10089–10092.
65. Best R, Zhu X, Shim J, Lopes P, Mittal J, Feig M, et al. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain $\chi(1)$ and $\chi(2)$ Dihedral Angles. *Journal of Chemical Theory and Computation*. 2012;8(9):3257–3273.
66. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al. Update of the CHARMM All-atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B*. 2010;114(23):7830–7843.
67. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. *Interaction Models for Water in Relation to Protein Hydration*. Intermolecular Forces; 1981.
68. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics*. 1983;79(2):926–935.
69. Berendsen HJC, Grigera JR, Straatsma TP. The missing Term in Effective Pair Potentials. *The Journal of Physical Chemistry*. 1987;91(24):6269–6271.
70. Li J, Yip S. *Handbook of Materials Modeling*. Springer Netherlands; 2005.
71. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes. *Journal of Computational Physics*. 1977;23(3):327–341.
72. Frenkel D, Smit B. *Understanding Molecular Simulations: from Algorithms to Applications*. Academic Press, San Diego; 2002.
73. Miyamoto S, Kollman PA. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *Journal of Computational Chemistry*. 1992;13(8):952–962.
74. Andersen HC. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *The Journal of Chemical Physics*. 1980;72(4):2384–2393.
75. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular Dynamics with Coupling to an External Bath. *The Journal of Chemical Physics*. 1984;81(8):3684–3690.
76. Bussi G, Donadio D, Parrinello M. Canonical Sampling through Velocity Rescaling. *The Journal of Chemical Physics*. 2007;126(1):014101.

77. Nosé S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Molecular Physics*. 1984;52(2):255–268.
78. Hoover WG. Canonical Dynamics: Equilibrium Phase-space Distributions. *Physical Review A*. 1985;31:1695–1697.
79. Parrinello M, Rahman A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *Journal of Applied Physics*. 1981;52(12):7182–7190.
80. Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant Pressure Molecular Dynamics Simulation: the Langevin Piston Method. *The Journal of Chemical Physics*. 1995;103(11):4613–4621.
81. Martyna GJ, Tobias DJ, Klein ML. Constant Pressure Molecular Dynamics Algorithms. *The Journal of Chemical Physics*. 1994;101(5):4177–4189.
82. Quigley D, Probert MIJ. Langevin Dynamics in Constant Pressure Extended Systems. *The Journal of Chemical Physics*. 2004;120(24):11432–11441.
83. Leach AR. *Molecular Modelling: Principles and Applications*. Pearson Education Limited, England; 2001.
84. Schlick T. *Molecular Modeling and Simulation: an Interdisciplinary Guide*. Springer, New York; 2002.
85. Huang K. *Statistical Mechanics*. J. Wiley & Sons; 1963.
86. Balescu R. *Equilibrium and Non-equilibrium Statistical Mechanics*. John Wiley & Sons; 1975.
87. McQuarrie DA. *Statistical Mechanics*. Harper & Row, New York; 1976.
88. Chandler D. *Introduction to Modern Statistical Mechanics*. Oxford University Press; 1987.
89. Pathria RK. *Statistical Mechanics*. Elsevier Science; 1996.
90. Wilde RE, Singh S. *Statistical Mechanics, Fundamentals and Modern Applications*. John Wiley & Sons, New York; 1998.
91. Zwanzig R. *Non Equilibrium Statistical Mechanics*. Oxford University Press; 2001.
92. Dellago C, Hummer G. Computing Equilibrium Free Energies Using Non-equilibrium Molecular Dynamics. *Entropy*. 2013;16:41–61.
93. Frenkel D. Free-Energy Computation and First-Order Phase Transitions. *Proceedings of the Enrico Fermi Summer School, Varenna, 1985*; Ciccotti, G., Hoover, W.G., Eds.; North-Holland Elsevier Science Publisher: Amsterdam, The Netherlands; 1987.
94. Jarzynski C. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*. 1997;78:2690–2693.
95. Jarzynski C. Equilibrium Free-energy Differences From Nonequilibrium Measurements: A Master-equation Approach. *Physical Review E*. 1997;56:5018–5035.
96. Crooks GE. Path-ensemble Averages in Systems Driven Far from Equilibrium. *Physical Review E*. 2000;61:2361–2366.
97. Crooks GE. Nonequilibrium Measurements of Free Energy Differences for Microscopically Reversible Markovian Systems. *Journal of Statistical Physics*. 1998;90:1481–1487.
98. Crooks GE. Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences. *Physical Review E*. 1999;60:2721–2726.
99. Elber R. A new paradigm for atomically detailed simulations of kinetics in biophysical systems. *Quarterly Reviews of Biophysics*. 2017;50:e8.
100. Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation*. 2012;8(5):1542–1555.
101. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation*. 2013;9(9):3878–3888.

102. Stone JE, Hardy DJ, Ufimtsev IS, Schulten K. GPU-accelerated Molecular Modeling Coming of Age. *Journal of Molecular Graphics and Modelling*. 2010;29(2):116–125.
103. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. Best Bang for Your Buck: GPU Nodes for GROMACS Biomolecular Simulations. *Journal of Computational Chemistry*. 2015;36(26):1990–2008.
104. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High Performance Molecular Simulations through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX*. 2015;1-2:19–25.
105. Bernardi RC, Melo MCR, Schulten K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochimica et Biophysica Acta*. 2015;1850(5):872–877.
106. Doshi U, Hamelberg D. Towards Fast, Rigorous and Efficient Conformational Sampling of Biomolecules: Advances in Accelerated Molecular Dynamics. *Biochimica et Biophysica Acta*. 2015;1850(5):878–888.
107. Spiwok V, Sucer Z, Hosek P. Enhanced Sampling Techniques in Biomolecular Simulations. *Biotechnology Advances*. 2015;33(6, Part 2):1130–1140.
108. Pietrucci F. Strategies for the Exploration of Free Energy Landscapes: Unity in Diversity and Challenges Ahead. *Reviews in Physics*. 2017;2:32–45.
109. Grubmüller H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Physical Review E*. 1995;52:2893–2906.
110. Rahman JA, Tully JC. Puddle-jumping: A Flexible Sampling Algorithm for Rare Event Systems. *Chemical Physics*. 2002;285:277–287.
111. Rahman JA, Tully JC. Puddle-skimming: An Efficient Sampling of Multidimensional Configuration Space. *The Journal of Chemical Physics*. 2002;116(20):8750–8760.
112. Voter AF. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Physical Review Letters*. 1997;78(20):3908–3911.
113. Hamelberg D, Mongan J, McCammon JA. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *The Journal of Chemical Physics*. 2004;120(24):11919–11929.
114. Laio A, Parrinello M. Escaping Free-energy Minima. *Proceedings of the National Academy of Sciences*. 2002;99(20):12562–12566.
115. Barducci A, Bussi G, Parrinello M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Chemistry Chemical Physics*. 2008;100(2):020603.
116. Limongelli V, Bonomi M, Parrinello M. Funnel Metadynamics as Accurate Binding Free-energy Method. *Proceedings of the National Academy of Sciences*. 2013;110(16):6358–6363.
117. Barducci A, Bonomi M, Parrinello M. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2011;1(5):826–843.
118. Hénin J, Chipot C. Overcoming Free Energy Barriers Using Unconstrained Molecular Dynamics Simulations. *The Journal of Chemical Physics*. 2004;121(7):2904–2914.
119. Darve E, Rodríguez-Gómez D, Pohorille A. Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. *The Journal of Chemical Physics*. 2008;128(14):144120.
120. Comer J, Gumbart JC, Hénin J, Lelièvre T, Pohorille A, Chipot C. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *The Journal of Physical Chemistry B*. 2015;119(3):1129–1151.
121. Maragliano L, Vanden-Eijnden E. A Temperature Accelerated Method for Sampling Free Energy and Determining Reaction Pathways in Rare Events Simulations. *Chemical Physics Letters*. 2006;426(1):168 – 175.
122. Earl DJ, Deem MW. Parallel Tempering: Theory, Applications, and New Perspectives. *Physical Chemistry Chemical Physics*. 2005;7:3910–3916.

123. Sugita Y, Okamoto Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters*. 1999;314:141–151.
124. Liu P, Kim B, Friesner RA, Berne BJ. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proceedings of the National Academy of Sciences*. 2005;102(39):13749–13754.
125. Gao YQ. An Integrate-over-temperature Approach for Enhanced Sampling. *The Journal of Chemical Physics*. 2008;128(6).
126. Roe DR, Bergonzo C, Cheatham TE. Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. *The Journal of Physical Chemistry B*. 2014;118(13):3543–3552.
127. Ostermeir K, Zacharias M. Advanced Replica-exchange Sampling to Study the Flexibility and Plasticity of Peptides and Proteins. *Biochimica et Biophysica Acta*. 2013;1834(5):847–853.
128. Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K. Molecular Dynamics Study of Unbinding of the Avidin-biotin Complex. *Biophysical Journal*. 1997;72(4):1568–1581.
129. Lu H, Schulten K. Steered Molecular Dynamics Simulations of Force-induced Protein Domain Unfolding. *Proteins: Structure, Function, and Bioinformatics*. 1999;35(4):453–463.
130. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, et al. *Steered Molecular Dynamics*. Springer Berlin Heidelberg Berlin, Heidelberg; 1999.
131. Isralewitz B, Baudry J, Gullingsrud J, Kosztin D, Schulten K. Steered Molecular Dynamics Investigations of Protein Function. *Journal of Molecular Graphics and Modelling*. 2001;19(1):13–25.
132. Isralewitz B, Gao M, Schulten K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Current Opinion in Structural Biology*. 2001;11(2):224–230.
133. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. Free Energy Calculation from Steered Molecular Dynamics Simulations Using Jarzynski's Equality. *The Journal of Chemical Physics*. 2003;119(6):3559–3566.
134. Eyring H. The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics*. 1935;3(2):107–115.
135. Chandler D. Statistical Mechanics of Isomerization Dynamics in Liquids and the Transition State Approximation. *The Journal of Chemical Physics*. 1978;68(6):2959–2970.
136. Dellago C, Bolhuis PG, Chandler D. On the Calculation of Reaction Rate Constants in the Transition Path Ensemble. *The Journal of Chemical Physics*. 1999;110(14):6617–6625.
137. Dellago C, Bolhuis PG, Geissler PL. *Transition Path Sampling*. John Wiley & Sons, Ltd; 2003.
138. Faradjian AK, Elber R. Computing Time Scales from Reaction Coordinates by Milestoning. *The Journal of Chemical Physics*. 2004;120(23):10880–10889.
139. Maragliano L, Vanden-Eijnden E, Roux B. Free Energy and Kinetics of Conformational Transitions from Voronoi Tessellated Milestoning with Restraining Potentials. *Journal of Chemical Theory and Computation*. 2009;5(10):2589–2594.
140. Voter AF, Montalenti F, Germann TC. Extending the Time Scale in Atomistic Simulation of Materials. *Annual Review of Materials Research*. 2002;32(1):321–346.
141. Hummer G, Kevrekidis IG. Coarse Molecular Dynamics of a Peptide Fragment: Free Energy, Kinetics, and Long-time Dynamics Computations. *The Journal of Chemical Physics*. 2003;118(23):10762–10773.
142. Ren W, Vanden-Eijnden E, Maragakis P, E W. Transition Pathways in Complex Systems: Application of the Finite-Temperature String Method to the Alanine Dipeptide. *Journal of Chemical Physics*. 2005;123:134109.

143. Moroni D, Bolhuis PG, van Erp TS. Rate Constants for Diffusive Processes by Partial Path Sampling. *The Journal of Chemical Physics*. 2004;120(9):4055–4065.
144. Thompson AN, Kim I, Panosian TD, Iverson TM, Allen TW, Nimigean CM. Mechanism of Potassium-channel Selectivity Revealed by Na⁺ and Li⁺ Binding Sites Within the KcsA Pore. *Nature Structural & Molecular Biology*. 2009;16(12):1317–1324.
145. Berneche S, Roux B. Energetics of Ion Conduction Through the K⁺ Channel. *Nature*. 2001;414:73–7.
146. Piccinini E, Ceccarelli M, Affinito F, Brunetti R, Jacoboni C. Biased Molecular Simulations for Free-Energy Mapping: A Comparison on the KcsA Channel as a Test Case. *Journal of Chemical Theory and Computation*. 2008;4(1):L72–L74.
147. Domene C, Klein ML, Branduardi D, Gervasio FL, Parrinello M. Conformational Changes and Gating at the Selectivity Filter of Potassium Channels. *Journal of the American Chemical Society*. 2008;130(29):9474–9480.
148. Baştuğ T, Kuyucak S. Comparative Study of the Energetics of Ion Permeation in Kv1.2 and KcsA Potassium Channels. *Biophysical Journal*. 2011;100:629–36.
149. Corry B, Thomas M. Mechanism of Ion Permeation and Selectivity in a Voltage Gated Sodium Channel. *Journal of the American Chemical Society*. 2012;134(3):1840–1846.
150. Furini S, Domene C. On Conduction in a Bacterial Sodium Channel. *PLoS Computational Biology*. 2012;8(4): e1002476.
151. Ke S, Zangerl EM, Stary-Weinzinger A. Distinct Interactions of Na⁺ and Ca²⁺ Ions with the Selectivity Filter of the Bacterial Sodium Channel NavAb. *Biochemical and Biophysical Research Communications*. 2013;430(4):1272–1276.
152. Qiu H, Shen R, Guo W. Ion Solvation and Structural Stability in a Sodium Channel Investigated by Molecular Dynamics Calculations. *Biochimica et Biophysica Acta*. 2012;1818(11):2529–2535.
153. Corry B. Na⁽⁺⁾/Ca⁽²⁺⁾ Selectivity in the Bacterial Voltage-gated Sodium Channel NavAb. *PeerJ*. 2013;1:e16.
154. Lenaeus MJ, Gamal El-Din TM, Ing C, Ramanadane K, Pomès R, Zheng N, et al. Structures of Closed and Open States of a Voltage-gated Sodium Channel. *Proceedings of the National Academy of Sciences*. 2017;114(15):3051–3060.
155. Callahan KM, Roux B. Molecular Dynamics of Ion Conduction through the Selectivity Filter of the NavAb Sodium Channel. *The Journal of Physical Chemistry B*. 2018;122(44):10126–10142.
156. Jiang D, Gamal El-Din T, Ing C, Lu P, Pomès R, Zheng N, et al. Structural Basis for Gating Pore Current in Periodic Paralysis. *Nature*. 2018;557:590–594.
157. Provasi D, Johnston JM, Filizola M. Lessons from Free Energy Simulations of δ -Opioid Receptor Homodimers Involving the Fourth Transmembrane Helix. *Biochemistry*. 2010;49(31):6771–6776.
158. Lau AY, Roux B. The Free Energy Landscapes Governing Conformational Changes in a Glutamate Receptor Ligand-Binding Domain. *Structure*. 2007;15(10):1203–1214.
159. Wied T, Chin A, Lau A. High Conformational Variability in the GluK2 Kainate Receptor Ligand-Binding Domain. *Structure*. 2018;27(1):189–195.e2.
160. Baştuğ T, Chen PC, Patra SM, Kuyucak S. Potential of Mean Force Calculations of Ligand Binding to Ion Channels from Jarzynski's Equality and Umbrella Sampling. *The Journal of Chemical Physics*. 2008;128(15):155104.
161. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The method. *Journal of Computational Chemistry*. 1992;13(8):1011–1021.
162. Souaille M, Roux B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Computer Physics Communications*. 2001;135(1):40 – 57.

163. Roux B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Computer Physics Communications*. 1995;91(1):275–282.
164. Ferrenberg AM, Swendsen RH. New Monte Carlo Technique for Studying Phase Transitions. *Physical Review Letters*. 1988;61:2635–2638.
165. Müller-Krumbhaar H, Binder K. Dynamic Properties of the Monte Carlo Method in Statistical Mechanics. *Journal of Statistical Physics*. 1973;8(1):1–24.
166. Bradley Efron R. *An Introduction to the Bootstrap*. Chapman and Hall/CRC; 1994.
167. Lin D, Grossfield A. Thermodynamics of Antimicrobial Lipopeptide Binding to Membranes: Origins of Affinity and Selectivity. *Biophysical Journal*. 2014;107:1862–1872.
168. Bell DR, Cheng SY, Salazar H, Ren P. Capturing RNA Folding Free Energy with Coarse-Grained Molecular Dynamics Simulations. *Scientific Reports*. 2017;7(45812).
169. Udayabhaskararao T, Altantzis T, Houben L, Coronado-Puchau M, Langer J, Popovitz-Biro R, et al. Tunable Porous Nanoallotropes Prepared by Post-assembly Etching of Binary Nanoparticle Superlattices. *Science*. 2017;358(6362):514–518.
170. Kirmizialtin S, Johnson KA, Elber R. Enzyme Selectivity of HIV Reverse Transcriptase: Conformations, Ligands, and Free Energy Partition. *The Journal of Physical Chemistry B*. 2015;119(35):11513–11526.
171. Ma P, Cardenas AE, Chaudhari MI, Elber R, Rempe SB. The Impact of Protonation on Early Translocation of Anthrax Lethal Factor: Kinetics from Molecular Dynamics Simulations and Milestoning Theory. *Journal of the American Chemical Society*. 2017;139(42):14837–14840.
172. Yu TQ, Lapelosa M, Vanden-Eijnden E, Abrams CF. Full Kinetics of CO Entry, Internal Diffusion, and Exit in Myoglobin from Transition-Path Theory Simulations. *Journal of the American Chemical Society*. 2015;137(8):3041–3050.
173. E W, Ren W, Vanden-Eijnden E. String Method for the Study of Rare Events. *Physical Review B*. 2002;66(052301).
174. E W, Ren W, Vanden-Eijnden E. Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-crossing Events. *The Journal of Chemical Physics*. 2007;126(164103).
175. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. String Method in Collective Variables: Minimum Free Energy Paths and Isocommittor Surfaces. *The Journal of Chemical Physics*. 2006;125(024106).
176. Maragliano L, Vanden-Eijnden E. On-the-fly String Method for Minimum Free Energy Paths Calculation. *Chemical Physics Letters*. 2007;446:182–190.
177. Stoltz G, Vanden-Eijnden E. Longtime convergence of the Temperature-Accelerated Molecular Dynamics Method. *Nonlinearity (ArXivPreprint Available arXiv:170808800)*. 2017;31(8).
178. Abrams CF, Vanden-Eijnden E. Large-scale Conformational Sampling of Proteins using Temperature-accelerated Molecular Dynamics. *Proceedings of the National Academy of Sciences*. 2010;107(11):4961–4966.
179. Vashisth H, Skiniotis G, Brooks CL. Using Enhanced Sampling and Structural Restraints to Refine Atomic Structures into Low-Resolution Electron Microscopy Maps. *Structure*. 2012;20(9):1453–1462.
180. Vashisth H, Abrams CF. All-atom Structural Models of Insulin Binding to the Insulin Receptor in the Presence of a Tandem hormone-binding Element. *Proteins: Structure, Function, and Bioinformatics*. 2013;81(6):1017–1030.
181. Vashisth H, Storaska AJ, Neubig RR, Brooks CL. Conformational Dynamics of a Regulator of G-Protein Signaling Protein Reveals a Mechanism of Allosteric Inhibition by a Small Molecule. *ACS Chemical Biology*. 2013;8(12):2778–2784.
182. Hu Y, Liu H. Case Study on Temperature-Accelerated Molecular Dynamics Simulation of Ligand Dissociation: Inducer Dissociation from the Lac Repressor Protein. *The Journal of Physical Chemistry A*. 2014;118(39):9272–9279.

183. Cortes-Ciriano I, Bouvier G, Nilges M, Maragliano L, Malliavin TE. Temperature Accelerated Molecular Dynamics with Soft-Ratcheting Criterion Orients Enhanced Sampling by Low-Resolution Information. *Journal of Chemical Theory and Computation*. 2015;11(7):3446–3454.
184. Maragliano L, Cottone G, Ciccotti G, Vanden-Eijnden E. Mapping the Network of Pathways of CO Diffusion in Myoglobin. *Journal of the American Chemical Society*. 2010;132(3):1010–1017.
185. Lapelosa M, Abrams CF. A Computational Study of Water and CO Migration Sites and Channels Inside Myoglobin. *Journal of Chemical Theory and Computation*. 2013;9(2):1265–1271.
186. Bucci A, Abrams CF. Oxygen Pathways and Allostery in Monomeric Sarcosine Oxidase via Single-sweep Free-Energy Reconstruction. *Journal of Chemical Theory and Computation*. 2014;10(7):2668–2676.
187. Mohammadi M, Vashisth H. Pathways and Thermodynamics of Oxygen Diffusion in [FeFe]-Hydrogenase. *The Journal of Physical Chemistry B*. 2017;121(43):10007–10017.
188. Pan AC, Sezer D, Roux B. Finding Transition Pathways Using the String Method with Swarms of Trajectories. *The Journal of Physical Chemistry B*. 2008;112(11):3432–3440.
189. Johnson ME, Hummer G. Characterization of a Dynamic String Method for the Construction of Transition Pathways in Molecular Reactions. *The Journal of Physical Chemistry B*. 2012;116(29):8573–8583.
190. Branduardi D, Faraldo-Gómez JD. String Method for Calculation of Minimum Free-Energy Paths in Cartesian Space in Freely Tumbling Systems. *Journal of Chemical Theory and Computation*. 2013;9(9):4140–4154.
191. Stober ST, Abrams CF. Energetics and Mechanism of the Normal-to-Amyloidogenic Isomerization of β 2-Microglobulin: On-the-Fly String Method Calculations. *The Journal of Physical Chemistry B*. 2012;116(31):9371–9375.
192. Pan AC, Weinreich TM, Shan Y, Scarpazza DP, Shaw DE. Assessing the Accuracy of Two Enhanced Sampling Methods Using EGFR Kinase Transition Pathways: The Influence of Collective Variable Choice. *Journal of Chemical Theory and Computation*. 2014;10(7):2860–2865.
193. Maragliano L, Roux B, Vanden-Eijnden E. Comparison between Mean Forces and Swarms-of-Trajectories String Methods. *Journal of Chemical Theory and Computation*. 2014;10(2):524–533.
194. Lev B, Murail S, Poitevin F, Cromer BA, Baaden M, Delarue M, et al. String Method Solution of the Gating Pathways for a Pentameric Ligand-gated Ion Channel. *Proceedings of the National Academy of Sciences*. 2017;114(21):4158–4167.
195. Carter EA, Ciccotti G, Hynes JT, Kapral R. Constrained Reaction Coordinate Dynamics for the Simulation of Rare Events. *Chemical Physics Letters*. 1989;156(5):472 – 477.
196. Ciccotti G, Kapral R, Vanden-Eijnden E. Blue Moon Sampling, Vectorial Reaction Coordinates, and Unbiased Constrained Dynamics. *ChemPhysChem*. 2005;6(9):1809 – 1814.
197. Bahar I, Atilgan A, Erman B. Direct Evaluation of Thermal Fluctuations in Protein Using a Single Parameter Harmonic Potential. *Folding & design*. 1997;2:173–81.
198. Doruker P, Atilgan AR, Bahar I. Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to α -amylase inhibitor. *Proteins: Structure, Function, and Bioinformatics*. 2000;40(3):512–524.
199. Atilgan A, Durell SR, Jernigan R, Demirel M, Keskin O, Bahar I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *Biophysical Journal*. 2001;80(1):505–515.
200. Brown D, Maigret B. Large Scale Molecular Dynamics Simulations using the Domain Decomposition Approach. *Speedup*. 1999;12(2):33–41.

201. Fincham D. Parallel Computers and Molecular Simulation. *Molecular Simulation*. 1987;1:1–45.
202. Petersen HG, Perram JW. Molecular Dynamics on Transputer Arrays. *Molecular Physics*. 1989;67(4):849–860.
203. Pinches MRS, Tildesley DJ, Smith W. Large Scale Molecular Dynamics on Parallel Computers using the Link-cell Algorithm. *Molecular Simulation*. 1991;6(1-3):51–87.
204. Raine ARC, Fincham D, Smith W. Systolic Loop Methods for Molecular Dynamics Simulation using Multiple Transputers. *Computer Physics Communications*. 1989;55(1):13–30.
205. Rapaport DC. Microscale Hydrodynamics: Discrete-particle Simulation of Evolving Flow Patterns. *Physical Review A*. 1987;36:3288–3299.
206. Smith W. Molecular Dynamics on Distributed Memory (MIMD) Parallel Computers. *Theoretica Chimica Acta*. 1993;84:385–398.
207. Smith W. Molecular Dynamics on Hypercube Parallel Computers. *Computer Physics Communications*. 1991;62(2):229–248.
208. Chys P, Chacón P. Random Coordinate Descent with Spinor-matrices and Geometric Filters for Efficient Loop Closure. *Journal of Chemical Theory and Computation*. 2013;9(3):1821–1829.
209. López-Blanco JR, Canosa-Valls AJ, Li Y, Chacón P. RCD+: Fast Loop Modeling Server. *Nucleic Acids Research*. 2016;44(W1):W395–W400.
210. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera — A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry*. 2004;25(13):1605–1612.
211. Shapovalov M, Dunbrack RJ. A Smoothed Backbone-dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure*. 2011;19(6):844–858.
212. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*. 2014;47(1):5.6.1–5.6.32.
213. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A. Comparative Protein Structure Modeling of Genes and Genomes. *Annual Review of Biophysics and Biomolecular Structure*. 2000;29(1):291–325.
214. Sali A. Comparative Protein Modeling by Satisfaction of Spatial Restraints. *Molecular Medicine Today*. 1995;1(6):270–277.
215. Fiser A, Kinh Gian Do R, Sali A. Modeling of Loops in Protein Structures. *Protein science : a publication of the Protein Society*. 2000;9:1753–1773.
216. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, Scalable Generation of High-quality Protein Multiple Sequence Alignments using Clustal Omega. *Molecular Systems Biology*. 2011;7(1).
217. Waterhouse A, Rempfer C, Heer FT, Studer G, Tauriello G, Bordoli L, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*. 2018;46(W1):W296–W303.
218. Ko J, Park H, Heo L, Seok C. GalaxyWEB Server for Protein Structure Prediction and Refinement. *Nucleic Acids Research*. 2012;40(W1):W294–W297.
219. Heo L, Lee H, Seok C. GalaxyRefineComplex: Refinement of Protein-protein Complex Model Structures Driven by Interface Repacking. *Scientific Reports*. 2016;6(32153).
220. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM Database and PPM Web Server: Resources for Positioning of Proteins in Membranes. *Nucleic Acids Research*. 2012;40(D1):D370–D376.
221. Sunhwan J, Taehoon K, Vidyashankara GI, Im W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. *Journal of Computational Chemistry*. 2008;29(11):1859–1865.

222. Beglov D, Roux B. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations. *The Journal of Chemical Physics*. 1994;100(12):9050–9063.
223. Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. *Journal of Molecular Graphics*. 1996;14(1):33–38.
224. Langham A, Kaznessis YN. Molecular Simulations of Antimicrobial Peptides. *Methods in Molecular Biology*-Wiley. 2010;618:267–285.
225. Smart OS, Neduvilil JG, Wang X, Wallace BA, Sansom MSP. HOLE: A Program for the Analysis of the Pore Dimensions of Ion Channel Structural models. *Journal of Molecular Graphics*. 1996;14(6):354 – 360.
226. Smart OS, Breed J, Smith GR, P SMS. A Novel Method for Structure-based Prediction of Ion Channel Conductance Properties. *Biophysical Journal*. 1997;72(3):1109–1126.
227. Fiorin G, Klein ML, Hémin J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Molecular Physics*. 2013;111(22-23):3345–3362.
228. Jo S, C Song K, Desaire H, D MacKerell A, Im W. Glycan Reader: Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins. *Journal of Computational Chemistry*. 2011;32:3135–3141.
229. Park SJ, Lee J, Patel D, Ma H, Sun Lee H, Jo S, et al. Glycan Reader is Improved to Recognize Most Sugar Types and Chemical Modifications in the Protein Data Bank. *Bioinformatics*. 2017;33:3051–3057.
230. Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*. 2016;12(1):405–413.
231. Yang LW, Eyal E, Bahar I, Kitao A. Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. *Bioinformatics*. 2009;25(5):606–614.
232. Yu ASL, Cheng MH, Angelow S, Günzel D, Kanzawa SA, Schneeberger EE, et al. Molecular Basis for Cation Selectivity in Claudin-2–based Paracellular Pores: Identification of an Electrostatic Interaction Site. *The Journal of General Physiology*. 2009;133(1):111–127.
233. Valley CC, Cembran A, Perlmutter JD, Lewis AK, Labello NP, Gao J, et al. The Methionine-aromatic Motif Plays a Unique Role in Stabilizing Protein Structure. *Journal of Biological Chemistry*. 2012;287(42):34979–34991.
234. Rosenthal R, Günzel D, Piontek J, Krug SM, Ayala-Torres C, Hempel C, et al. Claudin-15 Forms a Water Channel through the Tight Junction with Distinct Function Compared to Claudin-2. *Acta Physiologica*. 2019;0(ja):e13334.
235. Zhao J, Krystofiak ES, Ballesteros A, Cui R, Van Itallie CM, Anderson JM, et al. Multiple Claudin–claudin Cis Interfaces are Required for Tight Junction Strand Formation and Inherent Flexibility. *Communications Biology*. 2018;1(1):50.
236. Samanta P, Wang Y, Fuladi S, Zou J, Li Y, Shen L, et al. Molecular Determination of Claudin-15 Organization and Channel Selectivity. *The Journal of General Physiology*. 2018;150(7):949–968.
237. Irudayanathan FJ, Wang X, Wang N, Willsey SR, Seddon IA, Nangia S. Self-Assembly Simulations of Classic Claudins—Insights into the Pore Structure, Selectivity, and Higher Order Complexes. *The Journal of Physical Chemistry B*. 2018;122(30):7463–7474.
238. Krug SM, Günzel D, Conrad MP, Lee IM, Amasheh S, Fromm M, et al. Charge Selective Claudin Channels. *Annals of the New York Academy of Sciences*. 2012;1257:20–28.
239. Allen TW, Kuyucak S, Chung SH. Molecular Dynamics Study of the KcsA Potassium Channel. *Biophysical Journal*. 1999;77(5):2502–2516.
240. Khalili-Araghi F, Tajkhorshid E, Schulten K. Dynamics of K⁺ ion conduction through Kv1. 2. *Biophysical Journal*. 2006;91(6):L72–L74.

241. Noskov SY, Roux B. Importance of Hydration and Dynamics on the Selectivity of the KcsA and NaK Channels. *The Journal of General Physiology*. 2007;129(2):135–143.
242. Bostick DL, Brooks CL. Selectivity in K⁺ Channels is Due to Topological Control of the Permeant ion's Coordinated State. *Proceedings of the National Academy of Sciences*. 2007;104(22):9260–9265.
243. Ulmschneider MB, Bagn eris C, McCusker EC, DeCaen PG, Delling M, Clapham DE, et al. Molecular Dynamics of Ion Transport through the Open Conformation of a Bacterial Voltage-gated Sodium Channel. *Proceedings of the National Academy of Sciences*. 2013;110(16):6364–6369.
244. Boiteux C, Vorobyov I, Allen TW. Ion Conduction and Conformational Flexibility of a Bacterial Voltage-gated Sodium Channel. *Proceedings of the National Academy of Sciences*. 2014;111(9):3454–3459.
245. Domene C, Barbini P, Furini S. Bias-Exchange Metadynamics Simulations: An Efficient Strategy for the Analysis of Conduction and Selectivity in Ion Channels. *Journal of Chemical Theory and Computation*. 2015;11(4):1896–1906.
246. Guardiani C, Rodger PM, Fedorenko OA, Roberts SK, Khovanov IA. Sodium Binding Sites and Permeation Mechanism in the NaChBac Channel: A Molecular Dynamics Study. *Journal of Chemical Theory and Computation*. 2017;13(3):1389–1400.
247. Weber CR, Liang GH, Wang Y, Das S, Shen L, Yu ASL, et al. Claudin-2-dependent Paracellular Channels are Dynamically Gated. *Elife*. 2015;4(e09906).
248. Zhu F, Hummer G. Pore Opening and Closing of a Pentameric Ligand-gated Ion Channel. *Proceedings of the National Academy of Sciences*. 2010;107(46):19814–19819.
249. Berneche S, Roux B. Energetics of Ion Conduction through the K⁺ Channel. *Nature*. 2001;414:73–77.
250. Jensen M , Borhani DW, Lindorff-Larsen K, Maragakis P, Jogini V, Eastwood MP, et al. Principles of Conduction and Hydrophobic Gating in K⁺ Channels. *Proceedings of the National Academy of Sciences*. 2010;107(13):5833–5838.
251. Allen TW, Andersen OS, Roux B. Molecular Dynamics — Potential of Mean Force Calculations as a Tool for Understanding Ion Permeation and Selectivity in Narrow Channels. *Biophysical Chemistry*. 2006;124(3):251–267.
252. Allen TW, Andersen OS, Roux B. Energetics of Ion Conduction through the Gramicidin Channel. *Proceedings of the National Academy of Sciences*. 2004;101(1):117–122.
253. Villarroel A, Burnashev N, Sakmann B. Dimensions of the Narrow Portion of a Recombinant NMDA Receptor Channel. *Biophysical Journal*. 1995;68(3):866–875.
254. Taly A, H nin J, Changeux JP, Cecchini M. Allosteric Regulation of Pentameric Ligand-gated Ion Channels. An Emerging Mechanistic Perspective. *Channels (Austin)*. 2014;8(4):350–360.
255. Roux B. Ion Channels and Ion Selectivity. *Essays In Biochemistry*. 2017;61(2):201–209.
256. Weinman SA, Carruth MW, Dawson PA. Bile Acid Uptake via the Human Apical Sodium-Bile Acid Cotransporter is Electrogenic. *Journal of Biological Chemistry*. 1998;273(52):34691–34695.
257. Schultz SG, Curran PF. Coupled Transport of Sodium and Organic Solutes. *Physiological Reviews*. 1970;50(4):637–718.
258. Lane JS, Whang EE, Rigberg DA, Hines OJ, Kwan D, Zinner MJ, et al. Paracellular Glucose Transport Plays a Minor Role in the Unanesthetized Dog. *American Journal of Physiology-Gastrointestinal and Liver Physiology*. 1999;276(3):G789–G794.
259. Staat C, Coisne C, Pfeil S, M Stamatovic S, Andjelkovic A, Wolburg H, et al. Mode of Action of Claudin Peptidomimetics in the Transient Opening of Cellular Tight Junction Barriers. *Biomaterials*. 2015;54(06):9–20.

260. Bocsik A, Walter FR, Gyebrovski A, Fülöp L, Blasig I, Dabrowski S, et al. Reversible Opening of Intercellular Junctions of Intestinal Epithelial and Brain Endothelial Cells With Tight Junction Modulator Peptides. *Journal of Pharmaceutical Sciences*. 2016;105(2):754–765.
261. Dithmer S, Staat C, Müller C, Ku MC, Pohlmann A, Niendorf T, et al. Claudin Peptidomimetics Modulate Tissue Barriers for Enhanced Drug Delivery. *Annals of the New York Academy of Sciences*. 2017;1397:169–184.
262. Hashimoto Y, Shirakura K, Okada Y, Takeda H, Endo K, Tamura M, et al. Claudin-5-Binders Enhance Permeation of Solutes across the Blood-Brain Barrier in a Mammalian Model. *Journal of Pharmacology and Experimental Therapeutics*. 2017;363(2):275–283.
263. Neuhaus W, Piontek (Veshnyakova) A, Protze J, Eichner M, Mahringer A, Subileau ea, et al. Reversible Opening of the Blood-brain Barrier by Claudin-5-binding Variants of Clostridium Perfringens Enterotoxin's Claudin-binding Domain. *Biomaterials*. 2018;161(02):129–143.
264. Kim C, Adhikari A, Deisseroth K. Integration of Optogenetics with Complementary Methodologies in Systems Neuroscience. *Nature Reviews Neuroscience*. 2017;18(03):222–235.
265. London N, Raveh B, Schueler-Furman O. Peptide Docking and Structure-based Characterization of Peptide Binding: from Knowledge to Know-how. *Current Opinion in Structural Biology*. 2013;23(6):894–902.
266. Ciemny M, Kurcinski M, Kamel K, Kolinski A, Alam N, Schueler-Furman O, et al. Protein-peptide Docking: Opportunities and Challenges. *Drug Discovery Today*. 2018;23(8):1530–1537.
267. Kaplan DI, Isom LL, Petrou S. Role of Sodium Channels in Epilepsy. *Cold Spring Harbor Perspectives in Medicine*. 2016;6(6).
268. Rossi P, Sterlini B, Castroflorio E, Marte A, Onofri F, Valtorta F, et al. Novel Topology of Proline-Rich Transmembrane Protein 2 (PRRT2): Hints for an Intracellular Function at the Synapse. *Journal of Biological Chemistry*. 2016;291(12):6111–6123.
269. Fruscione F, Valente P, Sterlini B, Romei A, Baldassari S, Fadda M, et al. PRRT2 controls neuronal excitability by negatively modulating Na⁺ channel 1.2/1.6 activity. *Brain*. 2018;141(4):1000–1016.
270. Shen H, Zhou Q, Pan X, Li Z, Wu J, Yan N. Structure of a Eukaryotic Voltage-gated Sodium Channel at Near-atomic Resolution. *Science*. 2017;355(6328).
271. Yan Z, Zhou Q, Wang L, Wu J, Zhao Y, Huang G, et al. Structure of the Nav1.4- β 1 Complex from Electric Eel. *Cell*. 2017;170(3):470–482.
272. Shen H, Li Z, Jiang Y, Pan X, Wu J, Cristofori-Armstrong B, et al. Structural Basis for the Modulation of Voltage-gated Sodium Channels by Animal Toxins. *Science*. 2018;362(6412).
273. Pan X, Li Z, Zhou Q, Shen H, Wu K, Huang X, et al. Structure of the Human Voltage-gated Sodium channel Nav1.4 in complex with β 1. *Science*. 2018;362(6412).
274. Xu H, Li T, Rohou A, P Arthur C, Tzakoniati F, Wong E, et al. Structural Basis of Nav1.7 Inhibition by a Gating-Modifier Spider Toxin. *Cell*. 2019;176:702–715.
275. Clairfeuille T, Cloake A, Infield DT, Llongueras JP, Arthur CP, Li ZR, et al. Structural Basis of α -scorpion Toxin Action on Nav Channels. *Science*. 2019;363(6433).
276. Shen H, Liu D, Wu K, Lei J, Yan N. Structures of Human Nav1.7 Channel in Complex with Auxiliary Subunits and Animal Toxins. *Science*. 2019;363(6433):1303–1308.
277. Pan X, Li Z, Huang X, Huang G, Gao S, Shen H, et al. Molecular Basis for Pore Blockade of Human Na⁺ Channel Nav1.2 by the μ -conotoxin KIIIA. *Science*. 2019;363(6433):1309–1313.
278. Jiang D, Shi H, Tonggu L, Gamal El-Din T, Lenaeus M, Zhao Y, et al. Structure of the Cardiac Sodium Channel. *Cell*. 2019;180(1):122–134.e10.

279. Ing C, Pomès R. Chapter Eight - Simulation Studies of Ion Permeation and Selectivity in Voltage-Gated Sodium Channels. In: Na Channels from Phyla to Function. Academic Press; 2016. p. 215–260.
280. Hille B. The Permeability of the Sodium Channel to Organic Cations in Myelinated Nerve. *The Journal of General Physiology*. 1971;58(6):599–619.
281. Hodgkin ALH, Huxley AH. A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve. *Journal of Physiology*. 1952;117(4):500–544.
282. Hodgkin ALH, Huxley AH. Currents Carried by Sodium and Potassium Ions through the Membrane of the Giant Axon of Loligo. *Journal of Physiology*. 1952;116(4):449–472.
283. Hodgkin ALH, Huxley AH. The Components of Membrane Conductance in the Giant Axon of Loligo. *Journal of Physiology*. 1952;116(4):473–496.
284. Hodgkin ALH, Huxley AH. The Dual Effect of Membrane Potential on Sodium Conductance in the Giant Axon of Loligo. *Journal of Physiology*. 1952;116(4):497–506.
285. Hodgkin ALH, Huxley AH. The Potassium Permeability of a Giant Nerve Fibre. *Journal of Physiology*. 1955;128(1):61–88.
286. Hille B. The Permeability of the Sodium Channel to Metal Cations in Myelinated Nerve. *The Journal of General Physiology*. 1972;59(6):637–658.
287. Catterall W. The Molecular Basis of Neuronal Excitability. *Science*. 1984;223(4637):653–661.
288. Goldin AL, Barchi RL, Caldwell JH, Hofmann F, Howe JR, Hunter JC, et al. Nomenclature of Voltage-Gated Sodium Channels. *Neuron*. 2000;28(2):365–368.
289. Plummer NW, Meisler MH. Evolution and Diversity of Mammalian Sodium Channel Genes. *Genomics*. 1999;57(2):323–331.
290. Aggarwal SK, MacKinnon R. Contribution of the S4 Segment to Gating Charge in the Shaker K⁺ Channel. *Neuron*. 1996;16(6):1169–1177.
291. Schoppa N, McCormack K, Tanouye M, Sigworth F. The Size of Gating Charge in Wild-type and Mutant Shaker Potassium Channels. *Science*. 1992;255(5052):1712–1715.
292. Yang N, Horn R. Evidence for Voltage-dependent S4 Movement in Sodium Channels. *Neuron*. 1995;15(1):213–218.
293. Yang Y, D Dib-Hajj S, Zhang J, Zhang Y, Tyrrell L, Estacion M, et al. Structural Modelling and Mutant Cycle Analysis Predict Pharmacoresponsiveness of a Na(V)1.7 Mutant Channel. *Nature Communications*. 2012;3:1186.
294. Zhang Y. I-TASSER Server for Protein 3D Structure Predictions. *BMC Bioinformatics*. 2008;9:40.
295. Roy A, Kucukural A, Zhang Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nature protocols*. 2010;5:725–738.
296. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function Prediction. *Nature methods*. 2014;12:7–8.
297. Zhang Y, Skolnick J. TM-align: a Protein Structure Alignment Algorithm Based on the TM-score. *Nucleic Acids Research*. 2005;33:2302–2309.
298. Zhang J, Liang Y, Zhang Y. Atomic-level Protein Structure Refinement Using Fragment-Guided Molecular Dynamics Conformation Sampling. *Structure*. 2011;19(12):1784–1795.
299. Huang J, Rauscher S, Nawrocki G, Ting R, Feig M, de Groot B, et al. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Biophysical Journal*. 2017;112(1):175a–176a.
300. Flood E, Boiteux C, Allen T. Selective Ion Permeation Involves Complexation with Carboxylates and Lysine in a Model Human Sodium Channel. *PLoS Computational Biology*. 2018;14.

301. Sanders SJ, Campbell AJ, Cottrell JR, Moller RS, Wagner FF, Auldridge AL, et al. Progress in Understanding and Treating SCN2A-Mediated Disorders. *Trends in Neurosciences*. 2018;41(7):442–456. doi:doi.org/10.1016/j.tins.2018.03.011.
302. Stöbberg T, Mctague A, J Ruiz A, Hirata H, Zhen J, Long P, et al. Mutations in SLC12A5 in epilepsy of infancy with migrating focal seizures. *Nature communications*. 2015;6:8038.