Molecular modeling of human voltage-gated sodium channel complexes

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Carène Antonella Benasolo, M.Sc.

aus Nosy-Be Hell-Ville, Madagaskar

Berichter: Jun.-Prof. Dr. Jan-Philipp Machtens
Univ.-Prof. Dr. Marc Spehr

Tag der mündlichen Prüfung: 10.04.2024

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

I hereby declare that I have created this work completely on my own and used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Carène Antonella Benasolo
Aachen, August 2023

Contents

List of Publications vii

List of Figures ix

List of Tables xi

Abstract xiii

Zusammenfassung xv

1 Introduction 1

1.1 Voltage-gated Na^+channels . . . 1

1.2 Voltage-gated Na^+channel biography . . . 2

1.3 Human voltage-gated Na^+channel topology and structure . . . 3

1.4 Gating mechanisms of voltage-gated Na^+channels . . . 6

1.5 Navs and toxins . . . 9

1.6 Nav1.7 as pain target . . . 11

1.7 Aims of the thesis . . . 15

2 Theory and tools 17

2.1 Prediction tools of protein structures . . . 18

2.2 Molecular dynamics simulation . . . 19

2.3 Protein-protein docking . . . 22

2.4 Protein structure clustering . . . 23

2.5 Markov state modeling . . . 26

3 Methods 31

3.1 All-atom models . . . 31

3.2 All-atom simulation . . . 32

3.3 VAMPnets 34
3.4 Markov modeling 37
3.5 Coarse-grained models and DAFT simulation 37
3.6 Backmapping of coarse-grained models into all-atom 40
3.7 Protein-protein docking parameters 41
3.8 Prediction of hNav complexes with AlphaFold2 44

4 Prediction of hNav1.7α-hNav1.7α dimerization **47**
4.1 Dimer configurations predicted with the DAFT approach 47
4.2 Dimer configurations predicted with the docking method 50
4.3 Dimer configuration predicted with the AlphaFold2 50
4.4 Interaction interface properties 50
4.5 Interaction mechanisms 58
4.6 Scenarios of functional coupling in hNav1.7α dimer 58
4.7 Resume: investigation on the prediction of hNav1.7α in homo-dimer configurations 59

5 Prediction of hNavβ2 and β4 interaction sites on hNav1.7α **61**
5.1 All-atomistic molecular dynamics prediction 62
5.2 Docking prediction 64
5.3 Overview of hNavβ4 and Protx-II interaction sites 68
5.4 Modulation of hNav1.7α induced by hNavβ4 69
5.5 Resume: investigation of hNavβ4 interaction sites 76

6 Conclusions and perspectives **79**

A Summary of predicted hNav1.7α complexes **83**

Bibliography **85**

Acknowledgements **93**

List of Publications

In accordance with §5(4) of the doctoral regulations, I state that extracts of the thesis have been published or in progress to be published in the following papers:

[A] Benasolo C.A. and Machtens J.-P. "The auxiliary hNavβ4 as pharmacological modulator" **2023** (manuscript in progress)

[B] Ruehlmann A.H., Koerner J., Hausmann R., Bebrivenski N., Neuhof C., Detro-Dassen S., Hautvast P., Benasolo C.A., Meents J., Machtens J.-P., Schmalzing G., Lampert A. "Uncoupling sodium channel dimers restores the phenotype of a pain-linked $Na_v1.7$ channel mutation" *British Pharmacological Society* **2020** *177*:4481–4496

Chapters 5 corresponds to publication [A]. Quotations from these publications are indicated throughout the text. In accordance with §5(4) of the doctoral regulations, I state my contribution to these publications:

[A] I designed the simulations, produced and analyzed the data, and wrote the paper with contributions from all coauthors. The research was done under the supervision of Prof. Jan-Philipp Machtens.

[B] I conceived, performed, analyzed and wrote the section of the Molecular dynamics simulation with contributions from all coauthors. The research was done under the supervision of Prof. Jan-Philipp Machtens.

List of Figures

1.1 Structure and topology of voltage-gated Na^+channel 5
1.2 Gating charge residues and membrane potential 7
1.3 Illustration of Navs in closed, open and inactivated states. 8
1.4 Single mutation in hNav1.7 related to IEM and PEPD sydromes 13
1.5 Characterization of the action potential and the voltage dependency of activation in IEM patients. 14

3.1 VAMPnet models evaluation based on the scoring paramaters score3 and score2. 36
3.2 Clustering parametrization for the generated coarse-grained conformations. 39

4.1 Distribution analysis of the clustered configurations predicted from DAFT approach 48
4.2 Structural representation of the eighteen clustered dimer configurations predicted with the DAFT approach. 49
4.3 K-means clusterization of dimer configurations predicted with HADDOCK. 51
4.4 Representation of the dimer configurations predicted with the docking method. 52
4.5 Predicted models of hNav1.7α homo-dimer with Alphafold. 53
4.6 Stability analysis of dimer configurations 55
4.7 Reproducibility matrix of configurations generated from DAFT, HADDOCK and AF2 56
4.8 Mapping of the hNav1.7α interfacial residues. 57
4.9 Characterization of hNav1.7α dimer modulation. 60

5.1 Pharmacological modulation of Protx-II on hNav1.7α, in the presence and absence of hNavβ4 subunit 61
5.2 Interaction sites prediction for the Ig domain of hNavβ2 (a) and β4 (b) on hNav1.7α 62
5.3 Markov modeling analysis 63
5.4 Metastable states decomposition of the simulated trajectories of hNav1.7α/β4 complex, performed with Pyemma 65
5.5 Interaction sites prediction of hNavβ4 on hNav1.4α and hNav1.7α using docking methods. 66

5.6 VSDII interface comparison between hNav1.4α and hNav1.7α. 67
5.7 Interface comparison between hNav1.7α and hNavβ4 isoforms 68
5.8 Complex hNav1.7α/β4 prediction with Alphafold. 70
5.9 The predicted configurations of TM-hNavβ4, and the induced conformational changes at the IFM receptor and the pore of hNav1.7α 73
5.10 Conformational modulation at hNav1.7α VSDs 75
5.11 hNavβ4 might have distinct functions 78

List of Tables

1.1 Selected resolved three-dimensional structures of voltage-gated Na^+channels. Their experimental information were used to investigate structure-function properties for the prediction of hNavα1.7 configurations. 3
1.2 Highlight on some function of Navβ subunits in gating, voltage dependency and kinetics modulator 6

A.1 Summary of simulated systems 83
A.2 Summary of protein-protein docking prediction 83

Abstract

The human voltage-gated sodium channel isoform 1.7 (hNav1.7α) is defined as the main mediators of pain signal, characterized by cellular hyperexcitablity. Investigation on hNav1.7α revealed an implication of extensive interactions, influencing its mechanical-function, and responses to pharmacological treatments. In patients with gain-of-function mutations in hNav1.7α, these interactions amplify cellular hyperexcitability, and additionally impede drug efficacy. Our understanding of the interactomics of sodium channels are lacking, which limit the comprehensive insights into hNav1.7α mechanisms. This study present the first molecular insight of hNav1.7α-hNav1.7α in a dimer conformation, and characterize the interaction sites of the auxiliary hNavβ4 subunit on hNav1.7α. To accomplish this, we combined several computational techniques such as molecular dynamics simulation, protein-protein docking and AlphaFold2 predictions.
Until recently, voltage-gated sodium channels were considered as a monomer entity. Recent experimental studies have suggested that cellular hyperexcitability might be enhanced through a coupled-gating mechanism, where hNav1.7αs form dimer configuration. The prediction of hNav1.7α-hNav1.7α complex suggested the possibility of multiple conformations, impeding either directly the fast inactivation or promoting allosterically the activation of voltage sensor domains (VSDs).
The function of hNavβ4 subunit is poorly understood and its interaction sites have not yet been defined. However, hNavβ4 is suggested to potentially compete with drug interaction sites, affecting the drug efficacy. The prediction in the presence of hNavβ4 reveals conformational changes at the fast inactivation particles, the intracellular pore, and the voltage sensor domains of hNav1.7α. These changes seemingly indicate that hNavβ4 promotes the activation state of hNav1.7α. Thus, this project has provided more understanding on the mechanisms of hNav1.7α in complex quarternary structures, which will greatly contribute to the development of therapeutical strategies.

Zusammenfassung

Die spannungsgesteuerte Natriumkanalisoform 1.7 (hNav1.7α) spielt eine zentrale Rolle in der Weiterleitung von Schmerzsignalen beim Menschen. Untersuchungen an hNav1.7α haben eine Beteiligung umfangreicher Interaktionen aufgedeckt, die die Funktion des Kanals und Wirksamkeit pharmakologischer Behandlungen beeinflussen. Bei Patienten mit Gain-of-Function-Mutationen in hNav1.7α kann es zu einer zellulären Übererregbarkeit und neuropathische Schmerzsyndromen kommen, wobei diese Interaktionen diese Übererregbarkeit zusätzlich verstärken und die Wirksamkeit von Medikamenten beeinflussen können. Unser Verständnis des Interaktoms von Natriumkanälen ist bislang sehr begrenzt, was umfassende Einblicke in diese Mechanismen erschwert. Diese Studie präsentiert erste molekulare Einblicke in hNav1.7α in einer Dimerkonformation und charakterisiert die Interaktionsstellen der akzessorischen Untereinheit hNavβ4 an hNav1.7α. Um dies zu erreichen, habe ich mehrere computergestützte Techniken wie Molekulardynamiksimulationen, Protein–Protein-Docking und Proteinstrukturvorhersagen kombiniert.

Bis vor kurzem wurden spannungsgesteuerte Natriumkanäle als monomerische Entitäten betrachtet. Aktuelle experimentelle Studien weisen jedoch darauf hin, dass die zelluläre Überregbarkeit bei neuropathischen Schmerzen durch ein gekoppeltes Schaltverhalten von hNav1.7α-Dimeren zusätzlich verstärkt wird. Meine Vorhersagen des hNav1.7α–hNav1.7α-Komplexes führen zu neuen mechanistischen Hypothesen, wie die Dimerisierung zu einer funktionellen Kopplung zwischen zwei Kanälen führen kann. Die Funktion der hNavβ4-Untereinheit ist bislang unzureichend verstanden und ihre Interaktionsstellen sind noch nicht definiert. Es wird jedoch vermutet, dass hNavβ4 Wirkstoffinteraktionen modulieren und damit die Wirksamkeit von Medikamenten beeinflussen könnte. Die Strukturvorhersagen in Anwesenheit von hNavβ4 zeigen Konformationsänderungen an den schnellen Inaktivierungpartikel, dem intrazellulären Poreneingang und den Spannungssensordomänen von hNav1.7α. Diese Änderungen illustrieren, wie hNavβ4 den Aktivierungszustand von Nav1.7α beinflussen könnte. Dieses Projekt hat somit ein besseres Verständnis der Mechanismen hNav1.7α-Interaktoms ermöglicht und soll damit die Entwicklung neuer pharmakotherapeutischer Strategien unterstützen.

Chapter 1

Introduction

The treatment of pain is monopolizing medical attention. Pain is often the main symptom of many diseases, it can also be caused after injuries or surgery, which can lead to a discomfort or paralysis of the patients life. The health condition of some individuals are predominantly characterized by chronic, severe and inappropriate pain called chronic inherited pain syndromes, such as inherited Erythromelalgia (IEM) or paroxysmal extreme pain disorder (PEPD). Some individuals, however, have pain perception dysfunction manifested by an absolute pain insensitivity called chronic insensitivity to pain (CIP). Lifetime of patients with CIP are often shortened, as unconscious dangerous behaviors are common. Pain perception is considered as a protective mechanism (Cox et al. 2006).

In 2004, the mutation of the gene SCN9A encoding the protein voltage-gated Na^+channel Nav1.7 was linked with the pain disorder IEM (Yang et al. 2004), which suggests the implication of Nav1.7 in nociception or pain perception. This has been supported with the investigation led by Cox et *al.*, where the patients with CIP condition are identified with loss of function of their SCN9A. Since then, knowledge on the voltage-gated sodium channel family has increased with the advances in genetics, biochemical and biophysical experimental techniques such as polymerase chain reaction (PCR), patch-clamp and cryogenic electron microscopy (cryo-EM). This has confirmed the Nav1.7 channel isoform as the main component of pain propagation, thus the most promising drug target for pain disorders. Its experimental structure was finally resolved in 2019 (Shen et al. 2019). A lot remains to be known about Nav1.7 mechanism and its modulation.

1.1 Voltage-gated Na^+channels

The cell membrane is a semipermeable barrier constituted by lipid bilayers and crowded with proteins. Some embedded proteins play an important role as channels or transporters, diffusing specific substances across the membrane. Substances such as ions are

essential in the regulation of cells osmotic pressure, necessary in muscle and nerve cells functions, and responsible for most biological reaction in our body.
Ionic diffusion through the cells are determined by two main factors : the ion's concentration and the charge difference between the inner and the outer sides of the cellular membrane, this charge imbalance gives rise to the so-called membrane potential (Hodgkin and Keynes 1955, Clark, Choi, and Douglas 2018, chapter 35, Wright 2004).
A resting membrane potential is the electrical charge difference across a cell membrane at rest. It is influenced by the diffusion potential resulting from uneven ion distribution. This leads to an extracellular environment with a higher concentration of positively charged ions, causing the interior of the cell (cytoplasm) to carry a relatively negative charge.
Excitable cells such as neurons and skeletal muscle cells can reveal a brief reversal of membrane potential called an action potential. The action potentials mediate signal along cells, through the family of voltage-gated ions channels (VGICs) (Hodgkin and Huxley 1952). VGICs are selectively permeable to Na^+, K^+or Ca^{2+}. In response to stimulus or changes in the membrane voltage (i.e, the potential difference), these VGICs undergo conformational change into open, closed or inactivated states. The configurational states of voltage-gated Na^+channels (Navs) are detailed in Section 1.4.

1.2 Voltage-gated Na^+channel biography

In 1952, Hodgkin and Huxley discovered voltage-gated sodium currents when studying the electric conductance in axons of giant squids (Hodgkin and Huxley 1952, Lera Ruiz and Kraus 2015). The voltage-gated K^+channel was the first elucidated voltage-gated ion channel structure, using a X-ray crystallography (X-ray) (Doyle et al. 1998). This structure gave more comprehensive insight to the conformation of the voltage-gated ion channels family members. Indeed, before the availability of a Nav experimental structure, molecular modeling investigations had already defined its folding pattern (Guy and Seetharamulu 1986, Catterall 2000), which is constituted of a pore-forming domain surrounded by four voltage sensor domains (VSDs).
Nearly 60 years after Hodgkin and Huxley's research, emerged several ion channels structures with the advance of molecular structural determination methods such as X-ray crystallography (X-ray) and cryogenic electron microscopy (cryo-EM). The first Nav structure has been crystallized from the *Arcobacter butzleri* bacteria, entitled NavAb (Payandeh et al. 2011). This was followed by the elucidation of the first eukaryotic Nav structure by cryo-EM, extracted from the *American cockroach*, namely NavPas (Shen et al. 2017). NavPas shares 36 to 43 percent sequence identity with human Navs (hNavs), above the threshold to assume similar tertiary structure. In 2018, the elusive human voltage-gated Na^+channel was resolved by cryo-EM (Pan et al. 2018). Listed in table 1.1 are the resolved structures of voltage-gated Na^+channels. This list is restricted to resolved structures used to investigate on the configurations of hNavα1.7 in this project.

From the year 2020, the listing in the table 1.1 is not complete.

Table 1.1: Selected resolved three-dimensional structures of voltage-gated Na^+channels. Their experimental information were used to investigate structure-function properties for the prediction of hNavα1.7 configurations.

PDBID	Date	Organism	Resolution (Å)	Description
3RVZ[a]	2011	Arcobacter butzler	2.8	NavAb
5EK0[b]	2015	Arcobacter butzler Homo sapiens	3.5	Chimera-hNav1.7-VSD4/NavAb
5X0M[c]	2017	American cockroach	3.8	NavPas
5XSY[d]	2017	Electrophorus electricus	4	Nav1.4/β1
6AGF[e]	2018	Homo sapiens	3.2	Nav1.4α/β1
6A90[f]	2018	American cockroach	2.6	NavPas/Dc1a
6A95				NavPas/Dc1a/TTX
6A91				NavPas-Dc1a-STX
6J8G[g]	2019	Homo sapiens	3.2	Nav1.7/HWTX-IV/STX
6J8H				Nav1.7/HWTX-IV/STX
6J8I				Nav1.7/Prtx-II/TTX
6J8J				Nav1.7/Prtx-II/TTX
7DTD[h]	2021	Homo sapiens	3.3	Nav1.1α/β4
7XVE[i]	2022	Homo sapiens	2.7	mutant
7XVF				Nav1.7α/β1/β2

[a] Payandeh et al. 2011.
[b] Ahuja et al. 2015.
[c] Shen et al. 2017.
[d] Wang et al. 2017.
[e] Pan et al. 2018.
[f] Shen et al. 2018.
[g] Shen et al. 2019.
[h] Pan et al. 2021.
[i] Huang et al. 2022.

1.3 Human voltage-gated Na^+channel topology and structure

Eukaryotic Navs consist of pore-forming α subunit interacting with one or several auxiliary Navβ subunits (Pan et al. 2018).

1.3.1 Molecular architecture of the human Navα subunit

In human, there are nine isoforms of Navα (hNav1.1–hNav1.9). hNav1.1, hNav1.2, hNav1.3 and hNav1.6 are mainly expressed in the central nervous system; hNav1.4 in

skeletal muscle; hNav1.5 in heart muscle; and hNav1.7, hNav1.8 and hNav1.9 in the peripheral nervous system.
hNavα configuration is constituted of a large single polypeptide chain of 2000 amino acids, folding into four homologous repeats (DI–DIV) (Fig. 1.1). In contrast, NavAbα (a bacterial isoform) has a homotetrameric conformation (Payandeh et al. 2011). Each repeat contains six helical transmembrane segments (S1–S6). The first four segments of each repeat form the voltage-sensor domain (VSD) that are arranged laterally to the S5 and S6 pore-forming domain. The VSD domain and the pore of each repeat are connected with an intracellular linker between S4-S5. Located on the S4 segments are four to five repeated positive gating charges (GC) residues (Fig. 1.2). In response to membrane potential changes, the GC residues are driven upward or inward of the intracellular side, leading to conformational rearrangements of the channels. These conformational changes define three main states of Navs : closed, opened and inactived (developed in section 1.4).
The Na^+selectivity of the sodium channels is determined by four specific residues Asp/Glu/Lys/Ala (DEKA) located at the selectivity filter (SF) (Fig. 1.1c) (Ahern et al. 2015). The SF is rearranged in a sieve-like shape and is formed by the S5–S6 extracellular linker of each repeat.

1.3.2 Molecular architecture of Navβ subunit

The expression of Navα subunit alone demonstrated a functional channel (Lera Ruiz and Kraus 2015). However, the conformation and the mechanism of Navα can be modulated by one or multiple Navβ subunit(s) (O'Malley and Isom 2015, Brackenbury and Isom 2011). Five Navβ isoforms have been identified, β1 and its splice variant β1B, β2, β3 and β4. They have similar topology, composed by an extracellular immunoglobulin domain (Ig), a single α-helical transmembrane segment and an intracellular C-terminus (Fig. 1.1b). Except, β1B lacks a transmembrane segment and an intracellular domain.
The Navα (in the exception of Nav1.5) are known to establish a covalent disulfide interaction with the Navβ2 and β4 subunits, involving the Cys56 and Cys58, respectively (Das et al. 2016). Whereas the Navβ1 and β3 subunits are non-covalently bound to the Navsα.
The Navβ subunits are expressed in a large range of tissues and cell types, their expressions can decrease or increase over the different life cycle (embryonic, postnatal, early life and adult) (Brackenbury and Isom 2011).

The Navβ subunits play an important role in cell adhesion, which promote Navα genes expression at the plasma membrane (Wong et al. 2005, O'Malley and Isom 2015). These functions are attributed to the Ig domain of the Navβ subunits, which display structural homology to the Ig superfamily of the cell adhesion molecules (CAM). The adhesive function of Navβ has been clearly characterized in the Navβ1 isoform. Navβ2, β3 and

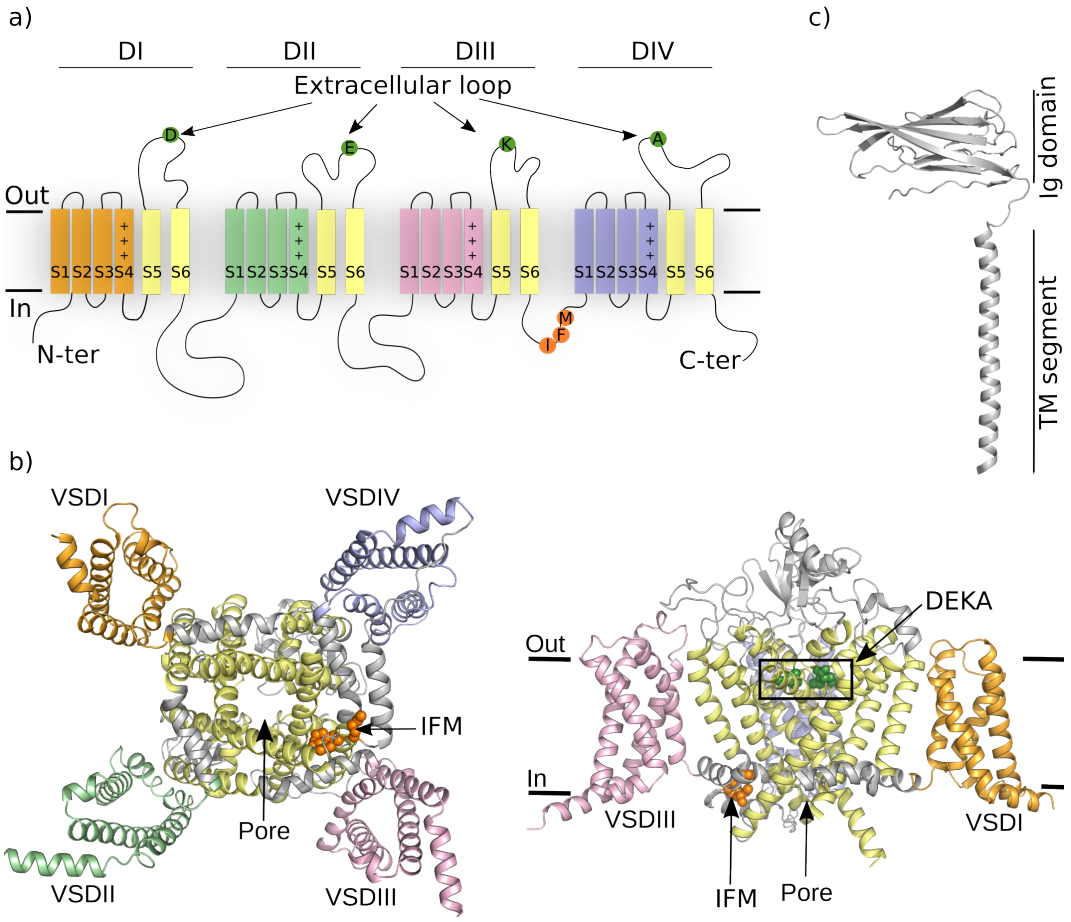


Figure 1.1: Structure and topology of voltage-gated Na^+channel. a) hNav topology consists of four homologous domains folded from a single polypeptide chain. For each repeat, the VSD formed by the S1 to S4 are respectively highlighted in light orange, green, pink and blue, for the DI, DII, DIII and DIV. The pore (S5–S6) in orange. To gain the location perspective of the DEKA and IFM motifs, they are drawn in dark green and orange circle respectively. b) The three-dimensional model of hNav1.7 based on the PDB-ID 6J8I are represented in cartoon, from the intracellular view (left) and the side view (rigth). The DEKA and the IFM motif are shown in sphere. c) Cartoon representation of β1 subunit structure formed by an Ig-domain and one transmembrane helical segment (PDB-ID: 6J8I). The Navβ subunits are well known to share similar structure. The intracellular loop (of around 29 to 36 residues) and a short N-terminal sequence (around 20 to 30 residues) are missing in this presentation, due to the unavailability of templates to model the full structure of Navβ subunits.

β4 in CAM interaction are still unclear (Davis, Chen, and Isom 2004; McEwen et al. 2009; O'Malley and Isom 2015; Yereddi et al. 2012). Moreover, they demonstrate weaker CAMs interaction compared to Navβ1.

Navβ subunits are conferred to further functions as gating, voltage dependency and kinetics modulator on Navα subunits; thus as cellular excitability regulator (O'Malley and Isom 2015, Brackenbury and Isom 2011). These functions are resumed in table 1.2.

Additionally, several studies have established pharmacological modulation functions of

Table 1.2: Highlight on some function of Navβ subunits in gating, voltage dependency and kinetics modulator

β subunits	Functions
$\beta 1^a$	- regulate the expression and subcellular distribution of α subunit proteins - increased peak sodium current density, precisely for Navα1.2 - shifted the voltage dependence of activation and inactivation to more negative membrane potentials - accelerate inactivation
$\beta 2^b$	- increased peak sodium current density - shifted the voltage dependence of inactivation to more negative membrane potentials - accelerate inactivation - stabilizing channel expression at the cell surface, which maintain normal action potential threshold -specifically regulate tetrodotoxin-sensitive channels
$\beta 3^c$	- shifts the voltage dependence of inactivation to more negative membrane potentials
$\beta 4^d$	- increase cellular excitability - slows inactivation - negative shifts in the voltage dependence of activation - promotes resurgent sodium current - promotes persistent sodium current

[a] Brackenbury et al. 2010; Chioni et al. 2009; Isom 1994; Isom et al. 1992; Isom et al. 1995; Moran, Conti, and Tammaro 2003.
[b] Isom 1994; Isom et al. 1992; Isom et al. 1995; Lopez-Santiago et al. 2006; Schmidt and Catterall 1986.
[c] Brackenbury and Isom 2011.
[d] Aman et al. 2009; Bant and Raman 2010; Grieco et al. 2005.

Navβ subunits on Navα subunits. It has been demonstrated that Navβ1 alters the pharmacological response of persistent Na^+ current to carbamazepine and lidocaine (Brackenbury and Isom 2011). The presence of Navβ4 decrease the Navα sensitivity to Protx-II (Gilchrist et al. 2013, Hampl *et al.* manuscript in progress). Understanding these pharmacological modulation by the Navβ subunits could help to develop novel drugs or therapeutical strategies. Moreover, the mutation of Navβ subunits are involved in several diseases (Brackenbury and Isom 2011).

1.4 Gating mechanisms of voltage-gated Na^+channels

In response to membrane voltage changes, the gating charge (GC) residues at the S4 segments (Fig. 1.2) are attracted or repulsed into/of the cytoplasmic side of the cell. Thus, the VSDs undergo inward–outward translocation through the membrane, which drag the adjacent segments and induce conformational changes at the pore into open or closed conformation. In mammals, three functional states of Navs have been defined : open,

closed and inactivated. The VSDs of each domain-repeat play distinct role in the Nav function (Clark, Choi, and Douglas 2018, Ahern et al. 2015, Armstrong and Bezanilla 1977, Armstrong and Bezanilla 1974). The different configurations of Nav are illustrated in Fig. 1.3.

Navs are closed at a resting membrane potential (Fig. 1.3a) (Wisedchaisri et al. 2019). At a resting membrane potential, the concentration of positive charges in the extracellular environment are higher than in the intracellular side of cell, resulting in a negatively charged potential. The GC residues are pulled toward the cytoplasm, inducing downward movements of the VSDs of all four repeats and consequently close the pore – blocking Na^+ current entirely.

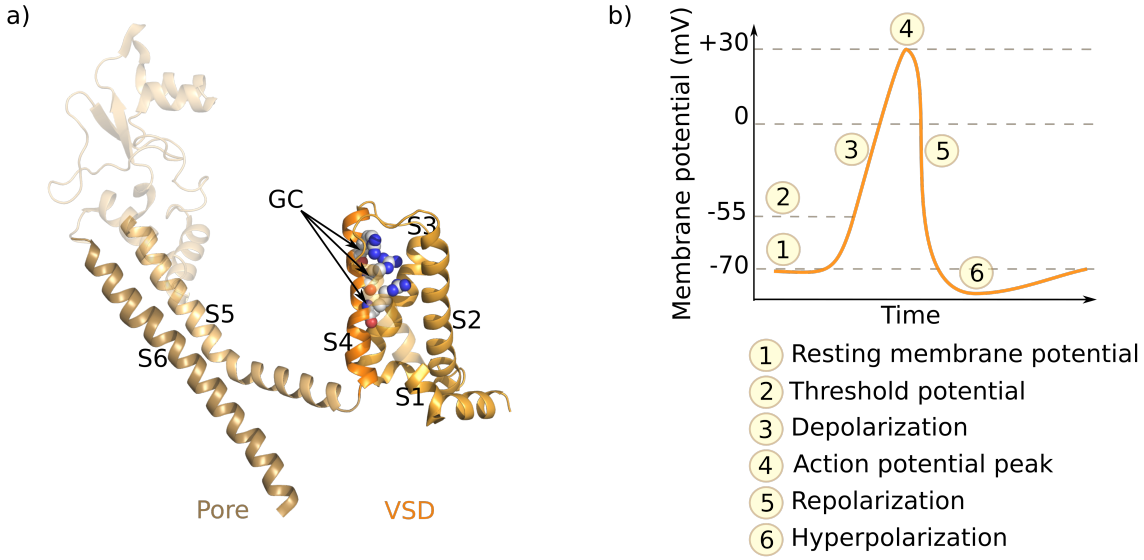


Figure 1.2: a) Cartoon representation of Navα-DI configurational structure, composed by the VSD and the pore segments. The structure model is based on PDBID 6J8I. The gating charges residues (GC) located on the S4 are highlighted in spheres. b) Illustration of the different membrane potential changes to drive Navs states.

Navs open upon membrane depolarization (Fig. 1.3b) (Ahern et al. 2015). The depolarization is initiated when the threshold potential is reached, allowing the upward translocation of the VSDI, VSDII and VSDIII, which lead to the channels opening (Armstrong and Bezanilla 1977). A large flow of Na^+ enters into the cell which decreases the charge difference between the inner and the outer cell. The membrane voltage becomes positively charged and generates the upstroke of the action potential.

Once the depolarization is complete, Navs transit to an inactivated state. There are two mechanisms of inactivation, fast and slow, which operate independently (Ulbricht 2005). Fast inactivation leads to the repolarization of the membrane, by closing of the pore (Fig. 1.3c). The rapid closure of the channel enables the restoration of the membrane

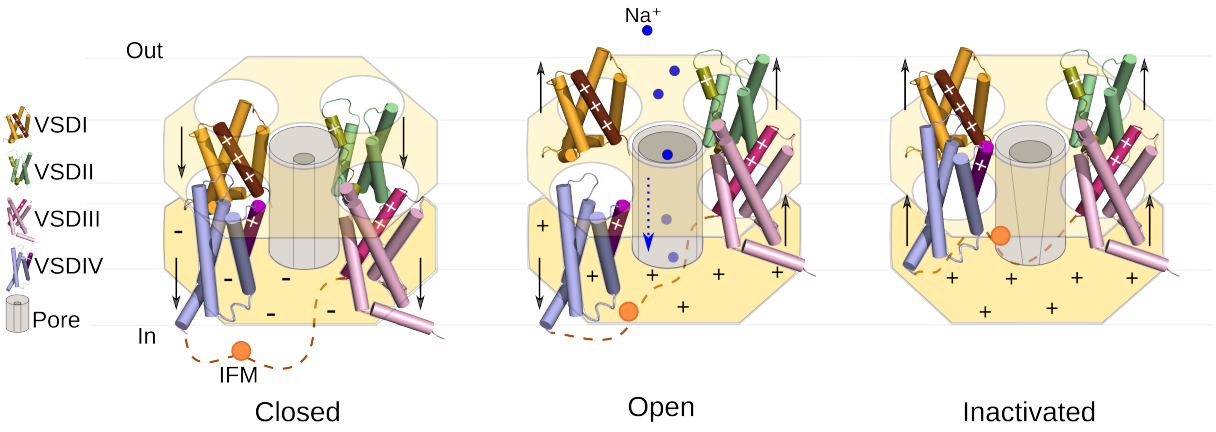


Figure 1.3: Illustration of Navs in close, open and fast-inactivated states. The VSDI, VSDII, VSDIII and VSDIV are in cylinder representation, respectively highlighted in orange, green, pink and blue. The gating charges located on the S4 segments of each domain are respectively highlighted in darker color. The VSDs motions are indicated with arrows. Schematics of the pore configurations are drawn in gray. These representations of sodium channels are embedded in a simplified lipid bilayers, in yellow; where the membrane potential of the intracellular are annotated. The IFM motif is drawn in orange spheres and the Na^+ ions in blue spheres. a) Nav in closed state: all of the four VSDs have inward movements toward the intracellular, as the potential membrane is negatively charged. b) Nav in open state: when the threshold potential is reached, the pore is enabled to open which allow a Na^+ flow into the cell. The membrane potential become positively charged and induce outward motion of the VSDI, VSDII and VSDIII. c) Nav in fast-inactivated state: the VSDIV undergo an outward motion. This is followed by the IFM motif recruitment into its receptor. The IFM receptor is surrounded by the pore segments S5–S6 of DIII and DIV. The binding of IFM to its receptor induce conformational changes around the segments forming the pore and lead to an allosteric pore occlusion.

potential. The fast inactivation is mediated by the residue IFM motif (Isoleucine-Phenilalanine-Methionine) located at the DIII–DIV linker. Recently, the fast inactivation mechanisms has been better clarified with the first resolved eukaryotic structure (Shen et al. 2017). The up translocation of the VSDI–VSDIII during the depolarization are rapidly followed by the upward motion of VSDIV. This exposes the hydrophobic receptor of IFM to the aqueous environment (Fig. 1.4, investigation performed on the PDB-IB 6J8I). A rapid recruitment of the IFM to its receptor induces an allosteric closure of the pore (Rühlmann et al. 2020). The fast inactivation occurs within millisecond timescale.

Slow inactivation is induced by prolonged or repetitive depolarization (Ulbricht 2005, Groome, Lehmann-Horn, and Holzherr 2011). It regulates the availability of excitable Nav channels for activation. The slow inactivation mechanism is unclear. However studies have demonstrated implication of certain residues located around the extracellular side of the the pore, specifically in the selectivity filter, in reference to voltage-gated K^+channel inactivation (Gawali and Todt 2016, Payandeh 2018). Slow inactivation takes seconds to minutes.

1.5 Navs and toxins

Certain animals and plants have developed defense or attack tactics, by producing chemicals substances – that might repulse the predators or lead to the paralysis of preys. From this concept has emerged the interest on several animals and plants toxins for therapeutic treatments. Some of these toxins have been demonstrated to modulate the functions of Navs. They were widely used to investigate and to characterize the different conformations of Navs (section 1.4). According to the interaction sites of the toxins and their modulation mechanism on Navs, the toxins are classified as pore-blockers or gating-modifiers. Pore-blocker toxins occlude directly the ion conduction. They bind on the intracellular or extracellular side of the pore. By contrast, the gating-modifier toxins have an allosteric effect on the ion conduction. They target the VSD domains, and maintain the VSDs in particular configurations or impair the VSDs translocation across the membrane (Gilchrist, Olivera, and Bosmans 2014, Ahern et al. 2015).

1.5.1 Pore-blocker molecules

The pore-blockers toxins often concern small molecules, that bind either in the core or at the periphery of the pore of Navs.
Tetrodotoxin (TTX) and saxitoxin (STX) toxins extracted from the Japanese puffer fish and the *Dinoflagellate algae* respectively, were widely used to study the structure–function of Navs, to characterize the Navs pore diameter, to identify the selective filter residues and in Nav purification (Gilchrist, Olivera, and Bosmans 2014). TTX and STX have non-competitive interaction sites located on S5–S6 extracellular residues and around the pore region. They demonstrated strong potential toxicity. Navs are divided in two groups according to their sensitivity to TTX. The TTX-sensitive channels include Nav1.1–1.4 and Nav1.6–1.7. The TTX-resistant channels are Nav1.5, Nav1.8 and Nav1.9 (Hille 1975, Gilchrist, Olivera, and Bosmans 2014).

The small molucules Brevetoxin and Ciguatoxin (CTX) are extracted from marine *Microalgae dinoflagellates*. They have been demonstrated to promote sodium channels activation and to alter the Na^+ permeability (Lombet, Bidard, and Lazdunski 1987). Brevetoxin interaction sites were identified on the pore forming region S6-DI and S5-DIV (Trainer, Baden, and Catterall 1994). CTX interaction sites are suggested to locate on the DI and DII (Yamaoka et al. 2009).

Batrachotoxin (BTX) can be isolated from the skin secretion of frogs of the genus *Phyllobates*. BTX plays roles in the inhibition of the fast and slow inactivation (Ahern et al. 2015). It is suggested to partially altered the pore, letting some fraction of Na^+ permeation and other ions. BTX interaction site involved multiple residues in the S6 pore segments (Gilchrist, Olivera, and Bosmans 2014).

The μ-conotoxin is a peptide toxin of 22 residues extracted from cone snail venom. The peptide μ-conotoxin has demonstrated a competitive interaction site with the toxin TTX, its interaction in the pore blocks the Na^+ current. This toxin also allowed to characterize the clockwise arrangement around the pore (Ahern et al. 2015; Dudley et al. 2000; Li et al. 2001). Additionally, it was used to identify a correlation between the GC segments motion and the pore opening (Ahern et al. 2015; French et al. 1996).

1.5.2 Gating-modifier toxins

Panel of scorpion, spider, sea anemone and cone snail venom peptide toxins have been defined as gating-modifiers (GMs) on Navs. They mostly interact on the VSDs domain, specifically on a motif pattern located on the linker S3–S4 segments (Catterall et al. 2007). Thus, these toxins were extensively used to characterize gating mechanisms of Navs.
Scorpion gating-modifier toxins are divided in two groups, α-scorpion toxins and β-scorpion toxins. The α-scorpion toxins interaction sites are identified on the Navs-VSDIV (Jover et al. 1978). The AaH2 belongs to the α-scorpion toxin, it is extracted from the scorpion *Androctonus australis*. Its structure has been resolved in complex with the chimera Nav-Pas/hNav1.7-VSDIV-DIS5 (Clairfeuille et al. 2019), where AaH2 is trapping the VSDIV in the downward configuration. The interaction of AaH2 on VSDIV has provided insight on the fast inactivation mechanism. AaH2 can also established an interaction with upward configuration of the VSDI, however with lower affinity.
In contrast to α-scorpion toxins, β-scorpion toxins enhance the channel activation by trapping the VSDII in the upward configuration (Cestèle et al. 2006).

The spider toxins can interact with multiple VSDs domain of Navs. The modulation effects of the spider toxin vary with the VSD domain that they are interacting, the VSD configurations, and the Navs isoform. Spider toxins might inhibit the channel activation, or *in contrario* promote the activation or impair the fast inactivation.
For instance, the HWTX-IV toxin extracted from the tarantula *Haplopelma schmidti* was captured trapping the VSDII in inward (Wisedchaisri et al. 2021) and outward (Shen et al. 2019) configurations. Wisedchaisri *et al.* succeeded to experimentally resolve the VSD domains in downward configuration. They linked the maltose-binding protein (MBP) with the chimera NavAb/Nav1.7-VSDII, and HWTX-IV had undergone triple mutations to improve its affinity. Successfully, they have screened the toxin HWTX-IV trapping the VSDII in the inward configuration. In this latest, the channel activation is inhibited. In the presence of HWTX-IV, extreme depolarization was required to activate the VSD *in vitro*, which is not biologically natural *in vivo*. So, the presence HWTX-IV occurs a hyperpolarization shift during the voltage dependence of activation (Xiao et al. 2008).
HWTX-IV has also been resolved trapping the hNav1.7-VSDII in outward configuration.

However, here VSDII was argued as only partially activated (i.e, in incomplete upward configuration), justifying the slight interaction of HWTX-IV.

The toxin Protx-II extracted from the tarantula *Thrixopelma pruriens* is one of the most studied with Navs. Protx-II toxin is Nav isoform specific. Protx-II has a higher affinity with the channel isoform hNav1.7, where its modulation mechanisms have been largely investigated. It can establish interaction with multiple VSDs (VSDI, VSDII and VSDIV). Protx-II has been demonstrated to inhibit the activation while interacting with the VSDI or VSDII. It also occurs the inhibition of the inactivation while interacting with the VSDIV (Xiao et al. 2010, Bosmans, Martin-Eauclaire, and Swartz 2008, Schmalhofer et al. 2008, Hampl *et al.* manuscript in progress).
Schmalhofer and co-workers have defined the residue F813 located at the VSDII, as the main anchor interaction of the Protx-II in order to inhibit the activation. The inhibition of the activation exhibit a reduction of Na^+ current in hNav1.7 (Schmalhofer et al. 2008). In Xiao *et al.* investigation, the co-expression of hNav1.7 with high concentration of Protx-II (1μM) likely occurred inhibition of inactivation. This suggests the VSDIV as an interaction site of the Protx-II. In both investigations, mutations were carried at the VSDII and the VSDIV in order to reduce the affinity of Protx-II to bind. Consequently, the effects of Protx-II were altered.
In corroboration with the experimental studies, the first three-dimensional structure of hNav1.7 has been resolved in complex with Protx-II (Shen et al. 2019). Protx-II were captured trapping the VSDII and the VSDIV in upward configuration. An additional experimental structure has also screened Protx-II interacting with the VSDII in up- and downward configurations, in the chimera NavAb/Nav1.7-VSDII (Xu et al. 2019).
The co-expression of HWTX-IV or Protx-II with Navs demonstrated divergent modulation effects. HWTX-IV seems to target only TTX-sensitive Navs (Xiao et al. 2008), whereas Protx-II can target almost the whole Navs isoform. However HWTX-IV and Protx-II exhibit a higher selectivity to hNav1.7, arousing even more interests on hNav1.7 isoform.

1.6 Nav1.7 as pain target

Mutations on Navs and its auxiliary β-subunits can lead to diverse mechanical dysfunctions and dysregulation of Navs expression. Several mutations on Navs have been associated with diverse pathology affecting the heart, muscles and brain such as arrhythmia, muscle paralysis, epilepsy, and chronic pain.

1.6.1 Nav1.7α Related-pain syndrom

According to statistical studies, one-third of all adults in the world suffer from chronic pain (Yang et al. 2018). Thanks to the genetic and functional investigations, several pain-related diseases have been linked to mutated genes encoding the Nav1.7, Nav1.8

and Nav1.9 isoforms (Faber et al. 2012, Dib-Hajj et al. 2010). Most documentations in nociception highlight the gene SCN9A encoding the sodium channel Nav1.7 as the main axis of pain transmission. Nav1.7 indeed exhibits significant depolarization of the voltage dependence of activation (Meents et al. 2019). In normal condition, Nav1.7 demonstrates a slower inactivation mechanism compared with the other isoforms, a feature allowing a late Na^+ conductance, and permitting to Nav1.7 to respond to small, slow depolarizations (Dib-Hajj et al. 2007; Herzog et al. 2003). Consequently, the mechanism of Nav1.7 can be easily perturbed when inserting mutation(s) at a single or several key regions of Nav1.7, as observed in pathological conditions.

Several gain-of-function mutations in Nav1.7 have been identified in spectrum of chronic inherited pain syndromes, such inherited Erythromelalgia (IEM) or paroxysmal extreme pain disorder (PEPD).
IEMs are characterized by an episodic burning pain localized at the distal extremities (feet and hands), accompanied with skin redness (erythema) and mild swelling. Electrophysiological studies have coupled IEM mutations with the channel activation sensitivity (Lampert, Eberhardt, and Waxman 2014). IEM mutations involve the substitution of hydrophobic residues into polar or charged residues, at critical regions regrouped around the gating domains and the S4-S5 linkers. Thus, the VSDs are promoted to the upward movement, in activated conformation. Then, only low voltage is required to open the pore; the channel activation is shifted to more hyperpolarized potential. This lead to cellular hyperexcitability. Around 22 mutations affecting the channel activation are linked to IEM pathology (Lampert, Eberhardt, and Waxman 2014).

PEPD patients suffer from prolonged pain attacks localized at the proximal parts of the body (face, torso, anal and genital region), sustained by skin flushing (Lampert, Eberhardt, and Waxman 2014). Most PEPD mutations are located within the pore domain and the intracellular side of the channel. These mutations impair inactivation mechanism of Nav1.7, by directly hampering the pore closure or allosterically hindering the fast-inactivation mechanism. These result to a prolonged channel activation, conducting to persistent current which occurs cellular hyperexcitability. During the persistent current, a small particle may enter and block the pore. It is suggested that the resurgent current results when the pore blocker is released. The resurgent current is defined as a small current of Na^+ observed upon the repolarization and contribute to cellular hyperexcitability. Few investigations involved the implication of Navβ4 subunit in the resurgent current (Chen et al. 2008; Grieco et al. 2005; Lampert, Eberhardt, and Waxman 2014) Around 10 mutations were associated to PEPD (Lampert, Eberhardt, and Waxman 2014).
It is the rise of hyperexcitability that causes the pain symptoms. In patients subjected to neuropathic and inflammatory pain-related syndromes, pain attacks can be triggered for instance by gentle mechanical stimulation or temperature change (Lampert, Eberhardt,

and Waxman 2014). In the Fig. 1.4 are illustrated a single mutation in hNav1.7 related to IEM and PEPD sydromes.

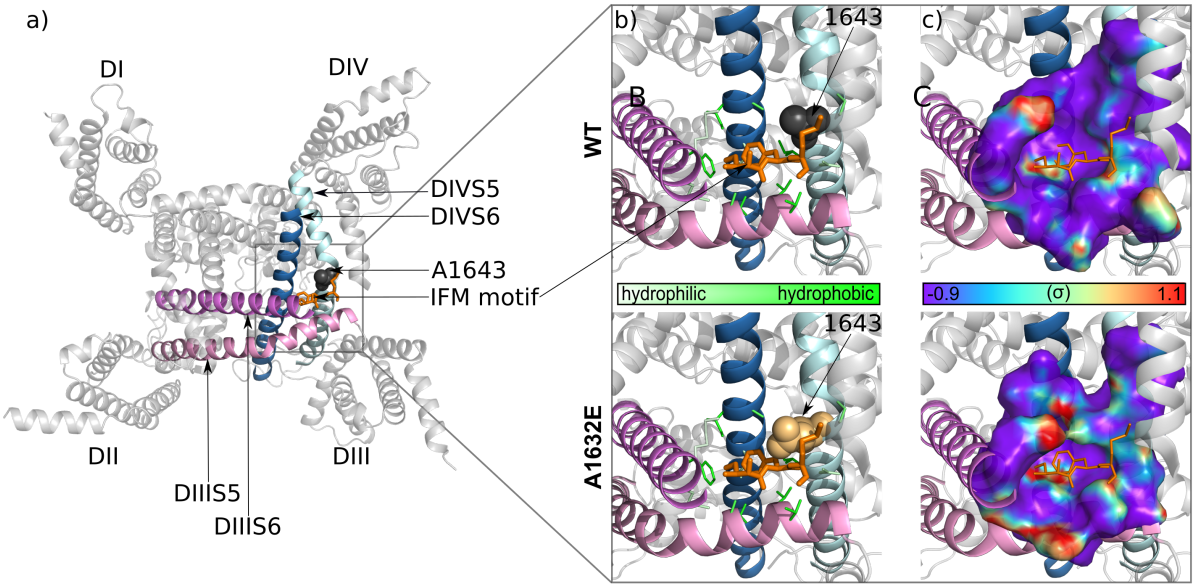


Figure 1.4: The A1643E mutation impedes binding of the IFM motif via steric repulsion and hydrophobic mismatch. a) Overview of the cryo-EM structure of the hNav 1.7α-subunit from the intracellular side (PDB ID: 6J8I). Spheres, A1643 of DIV S5; orange sticks, the IFM motif. b) Close-up view of the IFM-binding pocket, a hydrophobic cavity formed by residues from the DIII S4–S5 linker, DIII S5, DIII S6, DIV S5, and DIV S6 in both WT (upper) and A1643E Nav 1.7 (lower). Thin sticks, pocket-forming residues; thick sticks, IFM-motif residues; spheres, residue 1643. Intensity of green colour indicates the hydrophilicity/hydrophobicity. c) Solvent-accessible surface of the IFM-binding pocket from the same perspective as shown in b). Water densities from unguided 200-ns all-atom MD simulations of WT and A1643E Nav 1.7 (with the DIII–DIV linker removed) were mapped onto the protein surface, with mapping strength indicated by the colour-scale bar. The IFM motif was not present in these simulations, but its position in the IFM-bound structure is indicated by orange lines (figure extracted from the paper Rühlmann et al. 2020).

In 2006 has been established for the first time the relation between Nav1.7 loss-of-function recessive mutations and the disorder congenital insensitivity to pain (CIP) (Cox et al. 2006). Patients with CIP have insensitivity pain dysfunction. However, their somatic sense (pressure or warmth/cold) is normal. They carry non-functional Nav1.7, which is caused by non-sense mutation(s) or deletion in the gene SCN9A. These might induce an early codon-stop in the gene that shortened the protein or can occur splicing defects. This study has greatly supported the implication of Nav1.7 in pain propagation.

1.6.2 Strategies to relieve pain

To restore or regulate Nav1.7 dysfunctions, considerable efforts were invested in the development of analgesics such as antibodies, toxins and small molecules. The car-

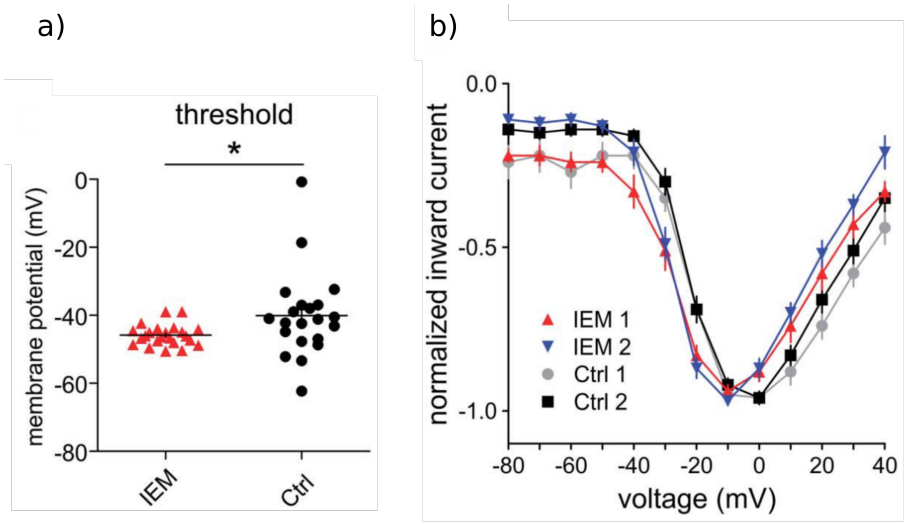


Figure 1.5: Characterization of the action potential and the voltage dependency of activation in IEM patients. a) IEM-Nav1.7 display a lower action potential threshold compared to the WT-Nav1.7 (black spheres). b) In black and gray represent two healthy control patients In contrast to the WT-Nav1.7, left-shit of the TTX-sensitive current are observed toward more negative action potential threshold. Figure extracted from the paper Meents et al. 2019

bamazepine molecule has been reported to successfully relieve pain in PEPD patients (Dib-Hajj et al. 2007). Moreover, carbamazepine shows a selectivity to inactivation-deficient channels, by enhancing its inactivation mechanism and stabilizing the inactivated state.
However, carbamazepine exhibits less effectiveness on most IEM variant mutations. IEM mutations are localized at regions affecting the channel activation. So far, existing treatments for IEM patients are reported unsatisfying (Yang et al. 2018) and the development of effective treatment is challenging. In IEM-Nav1.7, the activation is shifted towards negative potential, Meents *et al.* suggested that this lead to synchronized activation with the others sodium channels isoforms (Meents et al. 2019).

What if the synchronization of Nav activation is describing a coupled gating mechanism, where Navs form dimers? Until recently, voltage-gated sodium channels have always been represented as monomer entities. In patients with pain-related syndromes, the amplification of cellular excitability were suggested be enhanced by coupled-gating mechanisms in the channels activation. The insight emerged from investigation on cardiac diseases Brugada Syndrome (BrS), caused by mutations in the gene SCN5A encoding the cardiac sodium channel Nav1.5 isoform. Nav1.5 Brs-mutations can lead to spectrum of pathology, that are characterized with a reduction of Na^+ current.
Some BrS mutations yield an "atypical" Nav1.5 phenotype, no dysfunction are detected when they are expressed alone. However when these atypical Brs-Nav1.5 and WT-Nav1.5 are co-expressed, Hoshi *et al.* observed that the mutants displays dominant neg-

ative effect, characterized by a significant reduction of the current amplitudes (Hoshi et al. 2014). This study indicates obviously an interaction between the Navs to form a coupled-gating. This is supported by the investigation of Clatot *et al.*, who have firstly demonstrated voltage-gated sodium channels as dimer (Clatot et al. 2017). The 14-3-3 protein has demsonstrated to mediate the dimerization of Nav1.5 (Clatot et al. 2017). By uncoupling the complex atypical Brs-Nav1.5 and WT-Nav1.5, with the inhibitor of 14-3-3 protein namely Difopein, the dominant negative modulation induced between the two α-subunits are abolished and dimer conformations are not longer detected.
Thanks to all previous studies, our experimental team has demonstrated that hNav1.7 isoform also forms a dimer configuration (Rühlmann et al. 2020).

1.7 Aims of the thesis

Since Huxley's discovery, myriads of researchers have invested significant efforts to understand the voltage-gated sodium channels mechanisms and functions. The physiological function of the Nav channels are modulated and mediated by a large network of peripheral interactions (sections 1.3, 1.4). The insertion of single mutation (or multiple mutations) in Navs can disrupt the equilibrium of cells, leading to severe chronic inherited pain syndromes, such IEM and PEPD (section 1.6.1). The mutation of Navs were demonstrated to promote channel activation, inducing cellular hyperexcitability. Increased hyperexcitability observed in patients may be amplified through a coupled-gating mechanism, where Navs are suggested to form dimer conformations.
Dysregulation of Navα are also observed in the absence of some Navβ subunits or in the presence of defect Navβ subunits (section 1.3). With all that, interaction of Navs with endogenous entities (example animals and plants toxins) can modulate Navs mechanisms (section 1.5).
Investigations with several animals and plants toxins on Navs might have provided insights in the development new analgesics. Unfortunately, the existing treatments remain ineffective and exhibit poor selectivity according to the syndromes. However, Protx-II toxin demonstrated a high selectivity and affinity to hNav1.7α. Even though therapeutic application of this peptide toxin remains unfeasible, knowledge on the Protx-II toxin effect on hNav1.7α is used to drive investigation on the development of new therapeutical strategy.
In 2019, the advances in experimental structures techniques gave us access to hNav1.7 structure. Experimental knowledge on the voltage-gated hNav1.7α have provided great insights into its crucial role in pain-related diseases. However, the lack of interactomics knowledge in sodium channels limits comprehensive insights into hNav1.7α mechanisms, preventing the development of an effective drug treatment.

Thus, this research has elucidated several mechanical function of hNavα1.7 at molecular scale. With the help of experimental literatures, different computational biology tech-

niques were performed in order to:
(i) Describe atomistic mechanical behaviors of WT-hNavα1.7
(ii) Characterize the modulation induced in A1643E-hNavα1.7 mutation related to IEM and PEPD syndromes
(iii) Characterize the interaction sites of Navβ4 subunit on hNav1.7α. This aims to understand its pharmacological modulation function when expressed with Protx-II.
In order to highlight critical interaction particles in the complex hNav1.7α/β4, we also predict the complex hNav1.7α/β2 and hNav1.4α/β4 as control sets.
(iv) Providing the first dimer model of hNav1.7α isoform.

Chapter 2

Theory and tools

Despite the advances of experimental techniques, the experimental research is still facing many limits. (i) Investigation with biochemical and molecular biology methods are highly expensive and time-consuming. Not to mention the human manipulation errors and the biased essays due to environmental conditions. (ii) It omits detailed interaction at atomic level to fully understand the proteins mechanisms. Cell-based technologies often provide superficial description of macromolecules mechanisms, interpreted beyond its function. Several years might needed with the experimental research to access to detailed processes visited by the macromolecules. (iii) Although the regulations on animals ethics have improved over the last decades, animals usage remain consequent in research; for instance due to the wide marge of uncertainty of pharmacological effects.

However, experimental investigations are indispensable, complementary and often initiate biomolecular investigation with computational approach. Indeed, computational biology investigation often starts from a three-dimensional (3D) representation of a protein structure. From which is described interaction at all-atom (AA) or coarse-grained (CG) scales in order to illustrate the mechanisms. Protein structures are issued from experimental structures determination techniques such as nuclear magnetic resonance (NMR), X-ray crystallography or the cryogenic electron microscopic (cryo-EM).
These experimental structure techniques have improved rapidly this last decades, we are gaining more knowledge on biological processes and our interests are driven to more complex and larger protein structures. Consequently, the growth of resolved structures is increased exponentially in the Protein Data Bank (PDB) (Bernstein et al. 1977). However, structure resolution of large macromolecules is still challenging, either they are lacking or its structural representations are incomplete.

2.1 Prediction tools of protein structures

In order to describe at AA scale a protein and its function, several automated and supervised computational techniques have been developed to model and/or to predict protein 3D structure.

2.1.1 Modeller

Modeller is the most used 3D-structures modelisation tool (Webb and Sali 2016). Its approach is based on homology or comparative modeling, in which users provide the sequence of the protein of interest, and a single or multiple sequence alignment that is used as template(s) along with their corresponding PDB structures. Templates sharing high identity sequence with the protein of interest, should share structures similarities.

2.1.2 AlphaFold2

Recently in 2021, Jumper *et al.* have developed AlphaFold2 (AF2) which is a neural network tool to predict protein 3D-structures. AF2 requires as input only the amino acid sequence of the protein of interest. The structure prediction is trained based on multiple sequence alignments with related proteins' 3D structures from the PDB. The deep neural networks in AlphaFold2 learn patterns from the alignments and structural data, enabling the model to predict the spatial arrangement of amino acids in the target protein, to generate accurate 3D structural model (Jumper et al. 2021; Senior et al. 2020). The training are coded in an iterative process, as following:
(i) With the input sequence, AF2 builds homologous multiple sequence alignments (MSAs), mining in an implemented database such as Uniref90. (ii) The columns of the MSA can highlight co-evolving residues (which allow to preserve the structure and thus refer to residues in contact); from which are extracted residues pairwise correlation. (iii) These aim to generate residues pairwise representation, and used to infer geometrical proximity for all residues in the input sequence.
(iv) These processes are refined in the Evoformer architecture that is constituted with repeated 48 layers of a building network block. The refinements in the Evoformer are mediated by two transformers modules, the MSA transformer and the pairwise transformer. The MSA transformer drives the attention on the most correlated pairs of amino acids. These information are stored in the pair representation. The pairwise transformer can identify another possibity of pairs residues in interaction, which is passed to the MSA transformer and so on. (v) Thus, is inferred graphs of residues in interaction that are processed into the structure module block. (vi) The structure module block applies rotation and translation of each residue to assemble a full atomic structure. These steps are repeated and (vii) the predicted structures are assessed with the internal confidence measurements such as the pLDDT (predicted lDDT-Cα or Cα local-distance difference test)

and PAE (Predicted Aligned Error). pLDDT describes the per-residue local confidence in the structure. PAE describes the confidence of pairwise residues, giving an evaluation for the relative position of the predicted quaternary structures. For that, PAE assesses the expected positional error for residue x, where the predicted and actual structures are aligned (Varadi et al. 2021).
(vii) In parallel, pairs of residues inferred from the MSA can be trained on PDB structures, in order to predict spatial distances between their Cβ atoms (Senior et al. 2020).

2.1.3 AlphaFold-multimer

The development of AlphaFold2 has been extended to predict structure of protein-complexes, so called Alphafold-multimer (AM) (Evans et al. 2021). Several settings were udpated in the AM, aiming to identify co-evolving residues between two distinct macromolecules in heteromeric complexes. To prevent incorrect interfaces, the frame-aligned point error (FAPE) loss between inter-chain of complexes is defined. The FAPE measures the distance between the predicted atom positions to the true positions of different alignments.

2.2 Molecular dynamics simulation

2.2.1 All-atomistic simulation

The first interest in molecular dynamic (MD) simulation is reported in 1957 (Alder and Wainwright 1957). MD is used to study the dynamics of a macromolecule system of n atoms delimited in a predefined box that is applied to periodic boundary conditions. This aims to infer mechanical behaviors and properties of the macromolecule of interest. The MD simulations compute the atoms motion orchestrated by their interaction, according to Newton's law motion. These atoms interaction are driven by forces, that are set in a force field (FF) equation. This equation is giving by a combination of the derivative of potential energy parameters, describing the bonded and nonbonded interactions. Multiple FFs have been developed such as CHARMM (Klauda et al. 2010); they are typically expressed as following (Lindahl et al. 2020) :

$$
\begin{aligned}
U = & \sum_{bonds} k_b(b - b_0)^2 + \\
& \sum_{angles} k_\theta(\theta - \theta_0)^2 + \\
& \sum_{dihedrals} k_\phi[1 + cos(n\phi - \delta)] + \\
& \sum_{improper} k_\omega(\omega - \omega_0)^2 + \\
& \sum_{nonbond} \left[\varepsilon_{ij} \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{6} \right] + \\
& \frac{q_i q_j}{\varepsilon_r r_{ij}}
\end{aligned}
\tag{2.1}
$$

With respect to the atoms position, the three first terms regroup the bonded interaction expressed in harmonic function. b, θ, ϕ, ω are the current bond length, bond angle, dihedral angle and improper dihedral angle, respectively. $b_0, \theta_0, \phi_0, \omega_0$ are the reference (equilibrium) bond-length, angle, dihedral angle and improper torsion angle. $k_b, k_\theta, k_\phi, k_\omega$ are the stretching-bond, angle, dihedral angle and improper torsion angle force constants. Thus, the first term defines the bond-streching between atoms pair. The second term describes the valence angle energy defined for every triplet of adjacent atoms. The third term corresponds to the rotation bond of the torsion angle, formed by four consecutive bonded atoms. The cosinus function describes the rotation involving periodic, where n is the periodicity and δ is the phase. Then, the improper angles describe angle formed by four bonded atoms linked with an atom center.

The two last terms define the nonbonded interaction that are involved between atoms pair of neighboring molecules. The nonbonded interaction are defined with the Lennard-Jones (LJ) equation (a term for van-der-waals) and the Coulomb equation (a term for electrostatic interaction). The Lennard-Jones (LJ) equation modeled the attractive dispersion $(1/r)^6$ of atoms pair at long distance or repulsion $(1/r)^{12}$ at short distance, with r_{ij} corresponding to the atoms pair distance. $R_{min_{ij}}$ is the atoms distance at which the LJ energy is a minimum. The Coulomb term modeled electrostatic interaction between charged atoms. r_{ij} is the interatomic distance, q_i and q_j are the effective charges on atom$_i$ and atom$_j$.

Thus, the molecules (atoms) positions are given by the molecules degree of freedom, i.e the rotation and the translation of the molecules, with respect to the bond stretching, the bending of valence angles, and torsional degrees of freedom. The equations describing the particles motion are solved at discrete time intervals known as integration steps. These steps progress the simulation and update new positions and velocities for particles. In another words, MD simulations generate trajectories that capture snapshots of atom positions, velocities, energy, etc.

In AA-MD, each atom constituting the system is represented. The computation time of AA-MD increases with the system size, which implies large degrees of freedom. As a

result, simulations at long time-scales generate a significant amount of data. Thus, time and memory consuming underline AA-MD limitation. Several experimental structural data have not yet been addressed to AA-MD due to these limitations. For some large macromolecules, only partial configurational landscapes are explored during dynamic simulations. All of this to emphasize that, AA-MD investigation often need the combination with other computational techniques to fully understand the macromolecules mechanisms.

2.2.2 Coarse-grained molecular dynamics simulation

An approach to efficiently overcome the computational cost of AA-MD simulation is to reduce the system resolution. In coarse-grained (CG) models, groups of atoms are represented in simplified beads. One of the most accurate CG models are designed by the Marrink group, denoted Martini CG models (Marrink et al. 2007). Where four atoms are mapped into one bead (4-to-1), or in finer size bead (3-to-1 or 2-to-1) to close to more realistic mechanical properties.
To reproduce chemical interactions, these beads are described in different level of polarity according with the molecule charge. A specific FF has been developed to describe the beads interactions; the last version is denoted Martini3 FF (Souza et al. 2021). Martini FF is often combined with an elastic network model. This latest defines the structural scaffold of the macromolecule, it controls the flexibility and the overall conformation (Periole et al. 2009).
The parametrization of CG models allow to set larger values of integration steps (up to 20 fs, instead of 2 fs in AA-MD). Thus, CG-MD can be applied on more complex and larger system, such the prediction of protein-protein interaction. It can also reach a longer time-scale simulation that is yet unfeasible with AA-MD.
The CG models have greatly impacted MD investigation. However, the reduction in macromolecules resolution limit its structural and chemical description.

2.2.2.1 Docking Assay For Transmembrane components (DAFT) approach

A protein can undergo several biological processes to achieve its functions. These biological processes are characterized by conformational changes in the protein. Unbiased MD simulation generate stochastic events; as in biological environment, where the macromolecule can adopt different conformational structures to effectuate its function(s). With a single MD, we might not observed all the mechanical conformations. In addition, some conformations are occurred after longer time-scale simulation. Hence, to sample all possible configurations of a protein, it is necessary to run several MD simulations. The configurations of Navs (section 1.4) can be track by performing multiple MDs simulations.

CG-MD is implemented in Docking Assay For Transmembrane component (DAFT) approach (Wassenaar et al. 2015b), a generic framework to run a large number of simulation. DAFT is developed to perform high-throughput simulations to investigate on transmembrane protein-protein interaction, lipidomics screening or protein lipid interaction. The framework starts from CG model that is converted from an AA-model, using the function $martinize$. For protein-protein interaction prediction, the simulations start with unbound proteins, separated at d predefined distance, where each protein is oriented randomly on the xy-plane. This step is repeated several time to generate different starting configurations, in order to track interaction tendency. The starting configuration is inserted into a lipid bilayer, using $insane$. The composition of the lipid membrane can be set by the user.

2.3 Protein-protein docking

The advances in experimental techniques have made feasible the resolution of larger and complexes structures. However, the resolution of these structures requires a significant amount of effort. In MD simulation, protein-protein interaction events might occur only after long time-scale simulations, and often require several runs of MD to sample all possible complexes. This leads to the generation of large data.
Docking is an efficient method to predict protein-protein interaction. This method yields faster prediction and consumes less memory than molecular dynamics simulation. Docking approaches and workflows are often similar as follow (Dominguez, Boelens, and Bonvin 2003):
(i) Complex prediction starts from two unbound static conformations, called rigid-bodies.
(ii) One of the structure is restrained at fixed position. The second structure is rotating and translating around the first.
(iii) The previous process is iterated several times to identify matching interfaces between the two structures. Matching interfaces are based either on energetic complementarity, shape complementarity or experimental knowledges.
(iv) Each predicted conformation is evaluated and assigned to a scoring value. The scoring can be a combination of different parameters, calculated from the desolvation energy, the electrostatic interactions, the van der Waals repulsion, the ambiguous interaction restraints energy (AIR) and the buried surface area (BSA).
(v) The scoring might serves as a threshold to select the best configurations for further optimization and analysis.
HADDOCK (High Ambiguity Driven protein-protein Docking) is one of the most promising docking tools. The latest version HADDOCK2.4 was realised in 2020 (Ambrosetti, Jandova, and Bonvin 2020). With HADDOCK, the prediction of proteins interaction is driven by the ambiguous interaction restraints (AIRs). AIRs are residues listed from experimental biophysical/biochemical interaction knowledges. The residues that have experimentally been demonstrated to be involved in protein-protein interac-

tion (for instance by applying mutation) are labeled as ”active” AIRs residues. Residues surrounding the active AIRs or located at a potential interaction interfaces are labeled as ”inactive” AIRs. Then, distances restraints between AIRs are applied to ”constrain” or drive the interaction between the structures, without imposing a particular orientation.
With HADDOCK, the docking protocol consists of three steps:
(i) Rigid-body energy minimization (it0 stage): all possibilities of complexes conformation are predicted with two degrees of freedom, defined by the rotation and translation of the unbound structures. This process is repeated as much as the number of desired complexes samples. A first energy minimization are processed to remove clashes and to improve the involved energies in the interaction. Then, the rigid-body complexes are ranked according to HADDOCK scoring.
(ii) Semi-flexible simulated annealing (it1 stage): flexibilities are introduced on the residues at the interface interaction of the predicted complexes. The flexibility are firstly applied along the torsion angle of the residues side chains and followed by the flexibility on the backbones. Subsequently, HADDOCK performs a scoring and ranks the top complexes.
(iii) Refinement by energy minimization or short MD of the top ranking complexes.

Proteins are in constant motion, its interaction with other entities might induce a conformation change, and the protein might require a specific conformation in prior of an interaction. Even though HADDOCK2.4 made significant improvement in flexibility, docking methods present some limitations with induced conformation change. To improve the docking prediction, this flexibility can be addressed by applying multi-body or ensemble docking approach. Ensemble docking approach uses as input an ensemble of conformations sampled from MD simulations. The extraction of these conformations can be based on Root Mean Square Deviation (RMSD) (Bonvin 2006).

2.4 Protein structure clustering

The functions of a macromolecule emanate from its dynamic and involve interactions with other entities. Investigation with molecular computational techniques generally generate a large number of the macromolecule conformations. Often, relevant conformations are redundantly visited, defining structural similarities over large timescale. Thus, on the projection of explored configuration space, relevant conformations are regrouped in a sink of free energy. However, some relevant conformations are sampled in rare conformational events or fast biological process. In this project, clusterization algorithms (such as k-means) and Markov state modeling (MSM) are used to identify relevant conformations generated with the MD simulations and docking methods.

2.4.1 Data preparation for clustering

2.4.1.1 Similarity definition

The similarity between configurations can be calculated with pairwise Root Mean Square deviation (RMSD) and computed in a distance matrix so called all-to-all RMSD. The RMSD measures the mean distance between two proteins structures based on their atomic positions in a three-dimensional space. For large macromolecules, atoms degrees of freedom are represented in higher dimension. Consequently, the time calculation of the all-to-all RMSD is increased exponentially, handling less efficiently the data analysis. For that, we can describe the similarity between proteins conformations by a set of internal features, independent to the atoms positions; such as the distances between pairs of residues, the translation angles formed by three particles in the space, or the number of residues in contact between structures (Berg and Peter 2019). Respectively, these features might inform the orientation, the interaction interface, the relative positions or the stability of the structure(s).

2.4.1.2 Dimensionality reduction

To fully describe the explored configurations, a high number of features is often required, this raising the computation complexity for the analysis of the macromolecule. To facilitate interpretation of these features, Principal Component Analysis (PCA) is often performed in order to reduce the dimensionality of the data. PCA aims to reduce the number of variables, by combining correlated features in the original data into a new set of linearly variables (denoted principal components). The new variables are mutually orthogonal (uncorrelated) and summarized the directions of maximum variances in the data. PCA evaluates interralated variables in the original data by building a covariance matrix. The diagonalization of the covariance matrix yields to a set of eigenvectors and eigenvalues. By ordering the eigenvalues in descending order, the first eigenvector with the largest eigenvalue represents the first principal component and so on. The first component captures the largest variance in the transformed dimension (Rebala, Ravi, and Churiwala 2019).
The time-lagged independent component analysis (TICA) is an another efficient algorithm that also performed dimensionality reduction (more details are described in the section 2.5).

2.4.1.3 Clustering algorithms

To highlight macromolecules conformational tendency, clustering methods are applied on the all-to-all RMSD metric or the CVs principal components. The choice of clustering techniques and the related parameters might rely on the sampled data, the required output details. Several cluster algorithms have been developed. In this project, I used unsupervised clustering techniques such as k-means and DBSCAN.

k-means clustering algorithm aims to divide a data into a predefined number of k similarity groups. From k clusters, (i) the algorithm randomly chooses k-centroids. (ii) The squared eucledian distance of the data points from the centroids are computed, and data points are assigned to the nearest centroids. In order to refine the clustering, the mean of the data points assign to each cluster (each centroid) are calculated and consequently (iii) the centroids positions are moved. The step number (ii) is repeated. The data points are assigned to the new nearest centroids. These two last steps are reiterate until the algorithm converged, i.e when the position of the centroids do not change anymore (Rebala, Ravi, and Churiwala 2019). k-means method defines spherical-shaped clusters.

Density-based spatial clustering of applications with noise (DBSCAN) groups points based on local densities according on two hyperparameters : the radius ϵ and the minimum number of neighboring points (Minpts). DBSCAN assigns each point in the dataset to a cluster membership, (i) if within the given radius ϵ of each point contains at least Minpts of neighbours. The points within these conditions are denoted *core points*. (ii) If a point belonging to a *core point* does not fulfill the hyperparameters conditions, this point remains in the corresponding *core point* and is called *border points*. In contrast to k-means, DBSCAN does not necessarily assign points to a cluster (they are defined as a *noise points*). In addition, DBSCAN does not require to define *a priori* number of clusters (Raschka and Mirjalili 2017).

The choice of optimal hyper-parameters in k-means (number of k clusters) and DBSCAN (ϵ and Minpts) can be estimated with the "elbow method" or the "Silhouette score" metric. These methods evaluate the quality of the resulting clusters. The elbow approach consists of performing multiple clustering, using different values of hyperparameters. Then, for each run is evaluated the Sum of Squared Errors (SSE), to measure the variation within the cluster. The SSE are calculated by summing the squared distances of points to its cluster center. Thus, the lowest SSE indicates a good clustering. The graphical projection of the SSE against the hyperparameter values is represented with an elbow shape curve, where the best the hyperparameter is indicated at the distorted points before the SSE curve increase rapidly (Rebala, Ravi, and Churiwala 2019, Raschka and Mirjalili 2017). The silhouette score measures the (i) cluster cohesion and the (ii) clusters separation. Which respectively, calculates the average distance of points within a cluster and the

average distance of points with different clusters. Silhouette coefficient ranges between -1 and 1, where a score close to 1 interprets compact points within a cluster and far away from neighboring clusters. The score close to -1 describes overlapping clusters points.

2.5 Markov state modeling

To highlight relevant conformations of a macromolecule and track its modulation on evolving time, enhanced sampling techniques like Markov State Models (MSMs) have been developed (Pande, Beauchamp, and Bowman 2010, Bowman, Pande, and Noé 2014). The dynamics of a macromolecule structure can be described through the slow degrees of freedom (DOFs) over an extended simulation time. The MSM approach is a powerful framework that dissects the dynamics of a structure on a free energy configurational landscape, and defines the network pathway between metastable states. It summarizes the passage of one conformational state to cross another states over a discrete time lapse, referred as the lag time τ in MSMs.
Thus, MSM modeling is characterized by the number of states (n) and τ. MSM model is interpreted in a matrix nxn, denoted "transition probability matrix". It encodes the kinetics of transition by deducing the transition probabilities between the states (defined on rows of the matrix) and the stationary probabilities in each state (diagonal of the matrix).

Several steps are performed to build MSMs (Pande, Beauchamp, and Bowman 2010, Bowman, Pande, and Noé 2014, Wehmeyer et al. 2019).
(i) The explored configurational landscape generated from molecular dynamics simulation is discretized into microstates. The first step involves defining structural metrics or features that capture the process of interest. These features might encompass protein folding, chain or domain translation through the membrane, or the network pathway of a protein-protein interaction. To define these processes, a group of atoms is selected for calculating for instance torsion angles, relative positions, or distances relative to a reference structure. Based on the conformation description, clustering method such as k-means is applied. Generally, similar conformations are decomposed on a large number of small states (100-10000), and called microstates. The high number of microstates ensures that conformations within microstates have low internal free energy barriers.
(ii) The estimation of the transition matrix: subsequently, structures in the raw trajectories are assigned to microstates which result to a series of microstates over times. The number of transitions observed between a pair of states are counted and stored in the matrix, and is converted in transition probability matrix. Thus, transition matrix encodes the kinetics between the microstates.
(iii) The MSM coarse-grained: Microstates are generated at high resolution which is humanly difficult to gain insight of the relevant states. Coarse-grained MSM consists of merging microstates on a longer timescale, into macrostates. The merged microstates

have low barrier energy transition and defines as kinetically similar. These kinetic similarities in microstates are identified through the eigenvector of the transition matrix. The eigenvectors describe transition mode between states. The Coarse-grained MSM are performed with spectral clustering method such as Perron Cluster Cluster Analysis (PCCA+).
(iv) Validation of MSM models: The kinetics properties of MSM models can be evaluated from the eigenvalues of a transition matrix and with the Chapman-Kolmogorov Test (CK test). The eigenvalues of a transition matrix define the relaxation timescales or implied timescales (ITS), which describe the time interval, where transitions between states are history-dependent. The relaxation time increases with the energy barrier within the states, consequently the largest eigenvalues describe rates of the slowest processes. Thus, ITS (i) allows to choose an appropriate lag time τ to estimate the transition matrix that should be larger than the relaxation timescale. (ii) It also minimizes the discretization errors in MSMs (or to validate the kinetic relevance of the clustering) (Swope, Pitera, and Suits 2004).
CK test controls the quality of the Markov model. It computes on an evolve time-dependant the transition probability between metastable states. This allows to compare the probabilities of state i on lag-time τ to transition in state j on longer scale kτ ($k >= 1$) between the estimated MSM model and the observed trajectory data.
(v) Adaptive sampling: Once an initial MSM is build, we can improve its confidence by performing an adaptive sampling. It consists of running multiple simulations, in order to explore larger configurational landscape. The starting configurations for the subsequent MDs are sampled from the transition matrix, on states that highly contributes to the uncertainty of the eigenvalues and eigenvectors (Hinrichs and Pande 2007).

MSM approach is implemented in several open source software such as Pyemma (Wehmeyer et al. 2019). Pyemma software includes the time-lagged independent component analysis (TICA) in its pipeline. TICA performs the transformation of slowest-relaxing degrees of freedom (DOFs) of time series, into an approximation of the eigenvalues and the eigenvectors, given a linear combinations. Thus, TICA finds the linear combinations of high-autocorrelated DOFs. The DOFs are predefined in features set, describing the torsion angles, positions or distances. The autocorrelation refers to the correlation lag time. In contrast, PCA transforms the variables of an input data into linear combination that summarized the maximum variances of the data (Schwantes and Pande 2013). PCA and TICA are used for dimensionality reduction to describe data system (Noé and Clementi 2015).
Pyemma establishes similar pipeline as in MSM workflow describe above:
(i) setting the featurezation of system,
(ii) using TICA for dimensionality reduction that is performed on the defined feature(s),
(iii) performing cluster discretization of the reduced data,
(iv) defining lag time to perform an ITS in order to

(v) compute MSM estimation to describe the states kinetics
(vi) validate the MSMs
(vii) Coarse-grained the MSM to interpret the resulted Markov model

2.5.1 Adaptive sampling

Several MD simulations are often required to explore larger configurational landscape of a macromolecule. In this project, adaptive sampling is applied in order to obtain more accurate Markov models. The approach consists of starting the subsequent MD simulations from regions in configurational space that were fluctuating or not-well sampled. These configurational regions judged "uncertain" can be identified by using the transition matrix of MSM, built from existing trajectory simulations (Hinrichs and Pande 2007).
In MSM (section 2.5), the conformations along trajectories are projected onto a linear transformation (TICA) and are discretized in several microstates. Subsequently, the transition probabilities between the microstates are calculated and stored in transition matrix, from which the uncertainty of the eigenvalues and eigenvectors are estimated. The eigenvalues provide the states relaxation timescale of simulated system, representing how long conformations tend to remain within those states. The eigenvectors inform the transition that originate from a state i. By sampling several transition matrices, we can compute the variation of the eigenvalues and eigenvectors for each state (corresponding to each row on a transition matrix). Thus, the starting conformation of the subsequent MD generations are sampled from the states transitions that contribute the most to the variation, aiming to reduce uncertainties in eigenvalues and eigenvectors functions of a transition matrix.
An automated pipeline tool has been developed to build MSM models and to compute standard deviation to estimate the variation of the eigenvalues and eigenvectors contributions in the transition matrix (Longo, manuscript in progress). As the conventional Markov state construction, the adaptive sampling tool requires parameters such as the number of clusters *ncl* (for microstates discretization), lag τ, the number of transition matrix *nsamples* to be built for std evaluation and thus the number of new starting structures *nstr*.

2.5.2 Vampnet

The construction of MSM models are sequenced in several steps (section 2.5). The quality of the resulting MSMs are dependent on the choice of several parameters such as the feature characterization, state space decomposition (or dimension reduction) or MSM lag time. With Pyemma, each step requires an interactive decision of theses hyperparameters. This decision is often arise from qualitative observation or intuition, which can occur the failure of the Markov modeling or give non-optimal models.
Variational approach for Markov processes (VAMP) have been developed to select the

optimal hyper-parameters, by assigning a score to the built model. Thus, VAMP evaluation have to be tested with different values of hyper-parameters, and it requires the expertise of MSM modelling workflow.
In Vampnet, the MSM workflow is combine in a single end-to-end framework (Mardt et al. 2018). Vampnet is a deep learning approach trained by maximizing the VAMP variational score based on the following hyperparameters : the lag time τ, the number of states, the batch size and the epochs. The learning structure are build on two parallel network lobes $\chi 0$ and $\chi 1$. The lobes are tunneled and connected layers, which have a reduced dimension while progressing along network depth.
(i) In the input layer is passed the molecular coordinates found at xt and $xt + \tau$ along the MD. The number of nodes in the input layer are equal to the number of framed structures.
(ii) The network structure has at least one hidden layer to train (to learn) the feature transformations of the molecular configurations, with a reduced dimensionality.
(iii) The transformed feature are processed on Softmax output nodes in the output layer, i.e $\chi i(x) >= 0$ for all i and $\Sigma i \chi i(x) = 1$. This defines the probability of the conformation to be in any given metastable states. Yielding a fuzzy clustering of the molecular simulation. The transformed network data can be then interpreted in matrix of transition probabilities.
The validation of the model can be computed in implied timescales and the CK test.

Chapter 3

Methods

3.1 All-atom models

Different models of hNav1.7 were constructed using Modeller version 9.18 (Webb and Sali 2016). The structure templates used to model hNav1.7α, hNavβ1 and/or hNavβ2 subunits are based on the cryo-EM PDB ID 6J8I (Shen et al. 2019). Models with Navβ4 subunit are based on the X-ray structure PDB ID 5XAW (Shimizu et al. 2017). Where hNavβ4 was structurally aligned to the model hNavβ2 as in the PDB ID 6J8I, using pymol (Schrödinger and DeLano 2020). Modelisation of hNav1.7α with EFL domain and the corresponding conformation of DIII-DIV linker are based on the NavPas and hNav1.5 cryo-EM structures. Respectively, PDBID 6A95 (Shen et al. 2018) and PDBID 4DCK (Wang et al. 2012).

To investigate the interaction sites of Navβ subunits on hNav1.7α, the wild-type (WT) variant of the complexes (i) hNav1.7α/β1/β2 and (ii) hNav1.7α/β1/β4 were modeled. Additionally, to highlight occured modulation of hNav1.7α by Navβ subunits, (iii) the apo-structure hNav1.7α (apo-hNav1.7α) and (iv) the complexes hNav1.7α/β1 were modeled.
We attempted to investigate on the fast inactivation mechanisms, where (v) mutation A1643D/A1643E were inserted in the models apo-hNav1.7α and hNav1.7α/β1. The residue A1643 is located at the IFM receptor. Mutated residues were inserted with Modeller.
The fast inactivation is defined with the bound state of the IFM motif (b-IFM), located on the DIII-DIV linker. Yet, all experiment structures of the hNavs are resolved in the inactivated state, where the IFM is bound to its receptor and the four VDSs in upward configuration. To investigate the conformational changes induced by the binding of the IFM motif into its receptor, (vi) models with truncated DIII-DIV were generated. For further studies on the fast inactivation mechanisms, (vii) models with unbound IFM

motif (u-IFM) were built on the apo-hNav1.7α (apo-hNav1.7α/u-IFM) and hNav1.7α/β1 (hNav1.7α/u-IFM/β1) structures. Later, (viii) we also construct models with the presence of the EFL domain. These models were referred as hNav1.7α/EFL/b-IFM/β1 and hNav1.7α/EFL/u-IFM/β1.

For the modelisation of hNav1.7α subunit, the residue ranges 114–417, 726–972, and 1,175–1,768 were used. For models with EFL domain, residue ranges 1768–1878 were added. The structure of hNavβ1 with its Ig-domain and the transmembrane segment was modeled using the residues range of 20–192. For hNavβ2 and β4 subunits, available experimental structures only resolved the Ig-domain. Consequently the models were constructed in the residue ranges 29–148 and 30–152, respectively.

Apo-structure of hNav1.4α (apo-hNav1.4α) and hNav1.4α/β1 structures have also been modeled based on the cryo-EM structures PDB ID 6AGF (Pan et al. 2018), in the residue ranges of 119–464, 559–806, 1014–1607 for the α subunit and 20–192 for the hNavβ1 subunit. The residues at the extracellular loop 285–318 of hNav1.4α models were truncated. The unstructured sequences were pretty long and this can disrupt the stability of the protein during simulations. The truncated extracellular loop region is then modelled based on the electric eel Nav1.4α/β1 (EeNav1.4/β1) cryo-EM structure PDBID 5XSY (Yan et al. 2017).

According to PROPKA 3.1 (Olsson et al. 2011) to calculate the pKa, protonation states are assigned to the residues Glu180 and Asp1267 in hNav1.7α models, and the residues Asp185 and Asp1106 in hNav1.4α models. The remaining amino acids were modelled in their default ionization state. The N- and C-termini of each chain in the models or truncated extremities were capped with acetyl (ACE) and N-methyl (NME), respectively. The glycosylation site (Asn) covalently bounded with a sugar were preserved and modeled into new defined residues.

3.2 All-atom simulation

MD simulation of models with hNav1.7 and hNav1.4 were performed with GROMACS (Abraham et al. 2015), in version 2018, 2019, 2020 and 2021. The force field parameters CHARMM36m (Huang et al. 2016) and the TIP3P water model were applied. Protein models were embedded in a prior equilibrated lipid membrane using the g_membed function of GROMACS. The membrane is composed by pure 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC). The lipid bilayer are surrounded by 250 mM NaCl aqueous solution. The orientation of the protein across the membrane was guided by the Orientations of Proteins in Membranes database (OPM) (Lomize et al. 2011).

For all AA-MD simulations, the integration steps was set to 2 fs. The LINCS al-

gorithm restrains all hydrogen-involving bonds (Hess et al. 1998). Van der Waals interactions were evaluated with the Lennard-Jones potential within a cutoff radius of 1.2 nm; the forces are smoothly switched to zero in the range of 1.0–1.2 nm. Long-range electrostatics were calculated using the particle mesh Ewald (PME) method (Essmann et al. 1995). The nonbonded cutoffs was set to 1.2 nm. The simulation systems were operated under the isothermal-isobaric ensemble. The temperature were maintained at 310 K using the v-rescale thermostat (Bussi, Donadio, and Parrinello 2007) and a time constant of 0.5 ps. The pressure maintained at 1 bar using a Berendsen barostat (Berendsen et al. 1984) in a semi-isotropic manner with a time constant of 5 ps. The thermostat was separately applied to the protein, the lipid bilayer, and the water solution with ions.

Energy minimization is carried out using steepest descent, in double and single precision. The simulation systems were equilibrated in three steps. The first equilibration was performed for 4 ns, where the heavy atoms of the protein(s) and the lipids were restrained. The lipid heavy atoms were restrained only in the z-dimension. For the second equilibration, only the proteins heavy atoms were kept restrained for a timescale range of 200–500 ns. For the third equilibration, only the backbone of the proteins were restrained at least for 20 ns until the systems are ready for the production. Several replicas of each system were produced. The starting configuration for each system were selected accordingly to the adaptive sampling tools developed by Longo (Longo, manuscript in progress).
The comprehensive details of the performed MD simulation are presented in section 3.2.1, with a summarized overview provided in table A.1 (in the section A).

We also investigated on the electrostatic potential of hNav1.7α and hNav1.4α, using g_elpot tool (Kostritskii et al. 2021). The last structures from the equilibrated systems hNav1.7α/β1/β4 and hNav1.4α/β1 were extracted. The DIII-DIV linker were truncated. Then, equilibration of 20 ns were performed and followed by 120 ns production.

3.2.1 Adaptive sampling

The interaction site prediction of hNavβ2 and β4 subunits on hNav1.7α were performed with multiple free AA-MDs. To describe the visited interaction sites, we defined the relative position of hNavβ subunits along the trajectories, after superimposing hNav1.7α on a fixed referenced structure . The Cartesian coordinate corresponding to the carbone α (Cα) of hNavβ subunits were selected. The loop secondary structures were excluded in the coordinate selection, resulting to 150 and 249 Cα atoms positions of hNavβ2 and β4 subunits.
To efficiently sample broader configurational landscape of hNavβ subunits interaction,

adaptive sampling technique was used to select starting conformations for subsequent MDs simulation. The adaptive sampling consists of building a Markov State Modeling, estimate uncertainties in the transition matrix and sampling conformations in the regions that contribute the most in the uncertainties. In this project, an automated adaptive sampling developed by Longo were performed (Longo, manuscript in progress).

The adaptive sampling is performed on existing trajectories in order to build the MSM. For the first generation of MDs, the starting structures were firstly extracted during the equilibration MD. The aggregated time simulation for the first generation reach 18 μs and 31 μs, for hNav1.7$\alpha/\beta1/\beta2$ and hNav1.7$\alpha/\beta1/\beta4$ systems. Then, from existent MDs are applied adaptive sampling for the new generation of simulation. Prior each adaptive sampling, all previous trajectories are merged. Thus, 13 and 12 generation of adaptive sampling were performed for the systems hNav1.7$\alpha/\beta1/\beta2$ and hNav1.7$\alpha/\beta1/\beta4$. Respectively, 130 and 168 trajectories were generated. The timescale of the trajectories are ranged between 200 ns to 1.3 μs. Prior to analysis, these data were read every 0.2 ns, which results to a total of 260,000 and 340,000 conformations (frames), corresponding to 51 μs and 66 μs for hNav1.7$\alpha/\beta1/\beta2$ and hNav1.7$\alpha/\beta1/\beta4$ systems.
For each adaptive sampling , the uncertainties in the transition matrix are re-estimated. For that, the choice of MSM τ vary; timescale τ of 30 to 80 ns were used. The *ncl* and *nsamples* were kept constant for all adaptive sampling computation, respectively correspond to 500 and 400. For each adaptive sampling, 10 conformations were selected to start the new generation of MDs.

3.3 VAMPnets

VAMPnet is a function within the deeptime package (link: https://deeptime-ml.github.io). VAMPnet is a deep learning, single end-to-end framework of the Markov state model pipeline. To characterize the metastables states that defining the interactions sites of Nav$\beta2$ and $\beta4$ subunits and their kinetics, the overall generated trajectories of the hNav1.7$\alpha/\beta1/\beta2$ and hNav1.7$\alpha/\beta1/\beta4$ systems were applied to VAMPnet. Similarly to adaptive sampling analysis (section 3.2.1), the trajectories are read every 0.2 ns. This resulted to 260,000 and 340,000 conformations. A summary of the performed MD simulation are provided in the Table A.1. In order to analyze the interaction sites of Nav$\beta2$ and $\beta4$ subunits in each conformation, respectively 150 and 249 Cα atoms positions were described.
Subsequently, each VAMPnet lobe receives these coordinates in the input layer. Then, the learning transformations of the Cartesian coordinates were trained over four hidden layers. In the first hidden layer, dropout of 20% were applied and 50% dropout for the following hidden layers (Srivastava et al. 2014). The optimal hyper-parameters defining the number of output nodes (which refers to the number of states), the lag τ and the num-

ber of batches were evaluated with the scoring parameters score3 and score2 (developed in section 3.3.1).

3.3.1 Scoring parameters : score3 and score2

The optimal hyper-parameters were selected based on the scoring parameters score3 and score2. These scoring parameters illustrate the CK test error (CKer), the first implied component error (ICer) and the rescaled VAMP-2 score (rVamp2). Where $score3 = CKer + ICer + rVamp2$ and $score2 = CKer + ICer$
CKer defines the distance between the predicted trajectory data and the estimated model, tested on multiple lags (where mlags = 10). ICer informs on the convergence of the ITS. It computes the standard deviation on the first implied component in the lag range between 50 and 200 ns. VAMP-2 score is the sum of the squared estimated eigenvalues of the transfer operator, giving the kinetic variance explained by the model (Wehmeyer et al. 2019). VAMP-2 score were reversed (rVAMP-2), in order to refer the lowest score to the best model. Each of CKer, ICer and rVAMP-2 is rescaled to (-1,1) before calculating the score3 and score2. The scoring parameters with the lowest value indicate the optimal hyper-parameters.

3.3.2 Hyper-parameters choice : lag τ, number of states and batch size

The optimal hyper-parameters were chosen in two steps. Firstly, the optimum hyper-parameters were approximated by performing different values of lag τ = {4, 12, 20, 30, 40, 60}, units in ns; trying multiple values of metastable number n = {3, 4, 5, 6, 7, 8, 9}; and for different the batch sizes from 1000 to 20,000. The VAMPnets were performed by varying one of the hyper-parameters, while keeping the other parameters constants. The quality of each resulted VAMPnet model is evaluated with score3 and score2 scoring (section 3.3.1). Minimized score3 and score2 refer to the optimum hyper-parameters values. We observed that (i) increasing batch size reveal weaker VAMPnets models. And (ii) shorter τ resolve better the markovianity of higher ranges of metastable states (Fig. 3.1).
The best hyper-parameters are then optimized by running 50 Vampnets network. For the complexes hNav1.7$\alpha/\beta1/\beta2$, the models optimization were performed using batch size of 2200, with τ = 4 ns for n = 3 and n = 4, τ = 20 ns for n = 3 and n = 4. For the complexes hNav1.7$\alpha/\beta1/\beta4$, the models optimization were performed using batch size of 2200, with τ = 4ns for n = 3 and n = 8, τ = 12 ns for n = 3 and τ = 20 ns for n = 3. In each optimization run, the dataset was shuffled and randomly split into 70%/30%, respectively for training and validation.

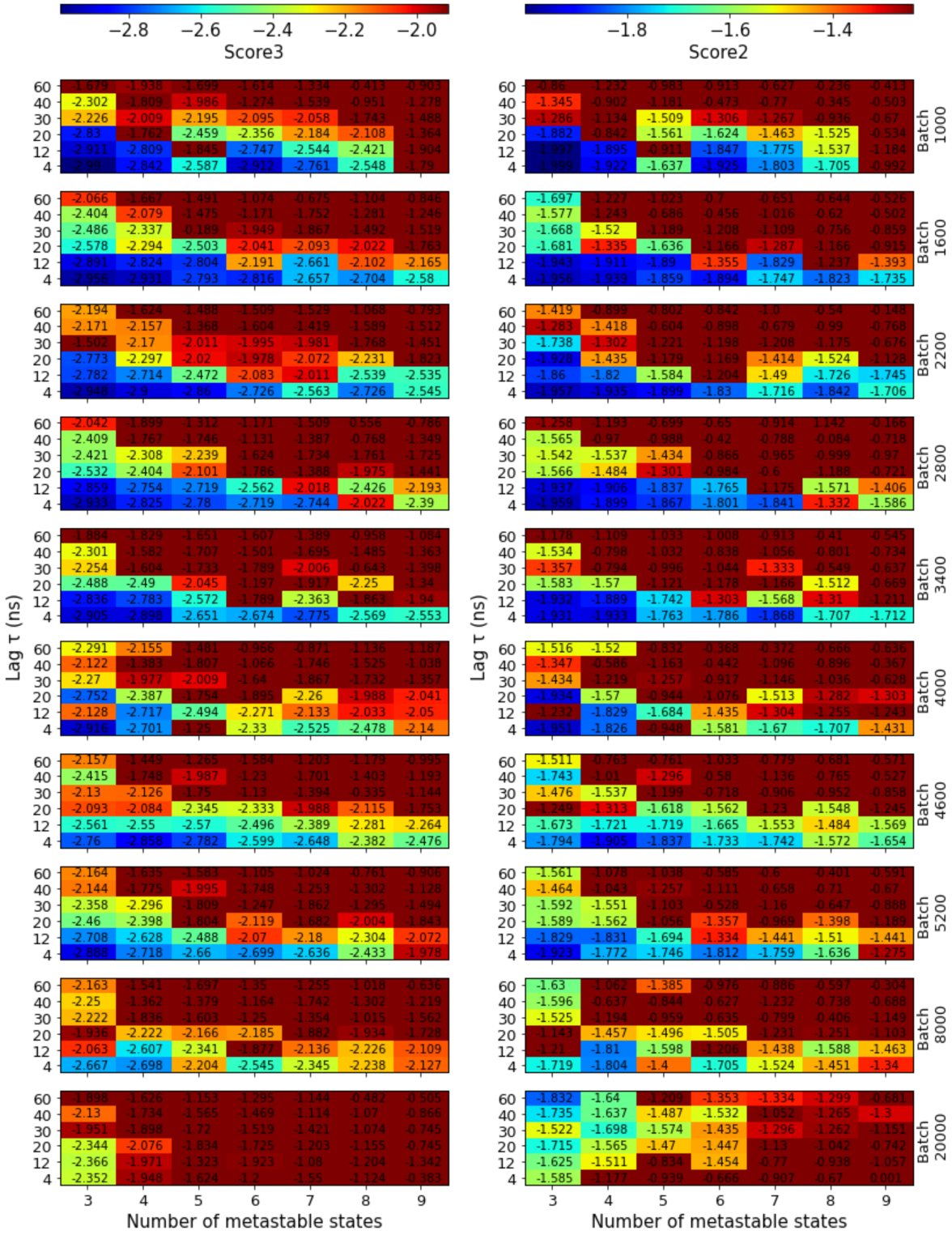


Figure 3.1: VAMPnet models evaluation based on the scoring paramaters score3 (left) and score2 (right), for the system hNav1.7$\alpha/\beta1/\beta4$. The approximation of the optimum hyper-parameters is estimated for different values of τ and n, respectively in the y-axis and x-axis. Each panel evaluate the VAMPnet model for different batch size. To distinctly highlight the scoring values for different hyper-parameters, the scaling of the colorbars are delimited from the minimum scores (corresponding to the best VAMPnet models) to mean scores values.

3.4 Markov modeling

The hyper-parameters lag time τ and number of states n chosen with VAMPnets were evaluated with the standard workflow of Markov state modeling, using Pyemma tools (Wehmeyer et al. 2019). The AA-MD simulation of the system hNav1.7α/β1/β2 and hNav1.7α/β1/β4 have generated 260,000 and 340,000 conformations. To investigate the interaction sites of hNavβ2 and hNavβ4 on hNav1.7α, the Cartesian coordinate of Navβ2 and β4 subunits were used to feature the MSM modeling. As with VAMPnets, respectively 150 and 249 Cα atoms positions were defined. These features are transformed into a lower dimension with TICA. The main variance in the data can be explained with three dimensions. Based on the features, the conformations are discretize into 500 microstates. Where the conformation similarities were computed with k-means. The kinetics between the microstates are estimated with the lag time τ defined with VAMPnets. Then, the microstates defining kinetic similarities are coarse-grained into metastable states. The VAMPnets metastable states number were applied. According to the lag time τ and n number of states, the models are forward evaluated with the implied time scale (ITS) and the CK test.

3.5 Coarse-grained models and DAFT simulation

Based on DAFT approach, several CG simulations were performed to predicted interaction interface of the hNav1.7 dimer. The CG simulations were conducted using GROMACS software package version 2019.6. Electrostatic interactions were calculated by reaction-field, with a real-space cutoff distance of 1.1 nm. Van der Waals interactions were calculated with Lennard–Jones interactions which is shifted by a constant such that it is zero at the cutoff 1.1 nm. All simulations were in isothermal-isobaric ensemble, where the temperature are maintained at 320 K using the v-rescale thermostat and a time constant of 1 ps. The pressure are maintained at 1 bar using Berendsen in a semi-isotropic manner with a time constant of 12 ps.

The settings of each simulation were implemented in an automated pipeline, consisting in four steps :
(i) The conversion of the hNav1.7 AA-model into CG model, using $martinize2$ (Kroon 2020). The AA-structure was extracted from the equilibration of apo-hNav1.7 system. The martini3 force fields of CG model were used. The elastic network of the model is turn on (-elastic) and the related elastic force (-ef) is set to 50 kJ.mol-1.nm-2, where the shortest (-el) and the longest (-eu) elastic bonds are extended to 0.5 and 2 nm respectively.
(ii) The converted CG model is duplicated into two unbound protomers, separated by a distance of 16 nm (from their respective center of mass) at xy-axis plane. The protein duplication was executed with GROMACS command $gmx\ genconf$. In addition, each protomer was randomly rotated on themselves.

(iii) The protomers were delimited in a box system with predefined dimension and embedding into bilayers lipid membrane, using $insane.py$ (Wassenaar et al. 2015a).
To control the stability and the reproducibility of the CG system, the simulations of the dimer prediction were conducted in two different membrane composition, either with pure POPC or with mixed lipids membranes. The structure of the mixed lipids membranes are asymmetric. The lipid composition in the upper leaflet is composed with 71% of POPC, 22% of POPE and 5.4% of POPS. The lower leaflet is composed with 15.7% of POPC, 52% of POPE and 32% of POPS. The lipids ratio are recalculated based on the following papers (Sharma et al. 2015, Shahane et al. 2019). Yet, the CG models of cholesterol molecules have not been designed for the martini3 FF.
Then, the system is solvated and ionized with salt concentration of 250 mM.
(iv) Next, each system is minimized using firstly double precision and then single precision, respectively for 3 ps and 45 ps. Integration time step for the minimization is set for 3 fs.
(v) This is followed with two steps of equilibration, where the proteins heavy atoms are restrained. The first equilibration (eq1) applied an integration time step of 2 fs and the simulations were generated for 200 ps. In the second equilibration (eq2) and the production run, the integration time step was increased to 20 fs. The production simulations were generated up to 20 μs.

To predict the interaction interface of the hNav1.7 dimer, a total of 90 CG simulations were generated. 74 simulations correspond to the system with pure membrane and 16 simulations for the system with mixed membrane. Each simulation has an average timescale of 18 μs. The accumulated timescale of the overall simulations reaches 1560 μs. The performed CG-MD simulation is resumed in the Table A.1 (in the appendix section).

3.5.1 Preparation of the predicted hNav1.7α dimer configurations for the analysis

To predict hNav1.7α in dimer conformation, the generated trajectories are read every 2.4 ns. This resulted to 650,337 frames for 90 CG simulations. Only dimer conformations are processed for the further analysis which refers to 290,000 frames. The CG-protomers of hNav1.7 are considered as a dimer when the shortest distance from their respective bead backbones (BB) are within 1 nm. Among the 90 generated simulations, 89 simulations occur at least one dimer snapshot frame. Along the trajectories, dimerization with weak interaction (i.e involving less than 4 residues or last less than 500 ns) are observed from 2 μs (Fig. 3.2a). To highlight visited and representative interfaces interactions of hNav1.7 dimers, conformation similarities between the frames are computed using selected features – corresponding to the BB atoms of 22 residues located in the vicinity of the membrane for each protomer. The BB pairs distances between the respective protomers were computed, resulting in 484 atoms-pair combination. These features inform on the relative

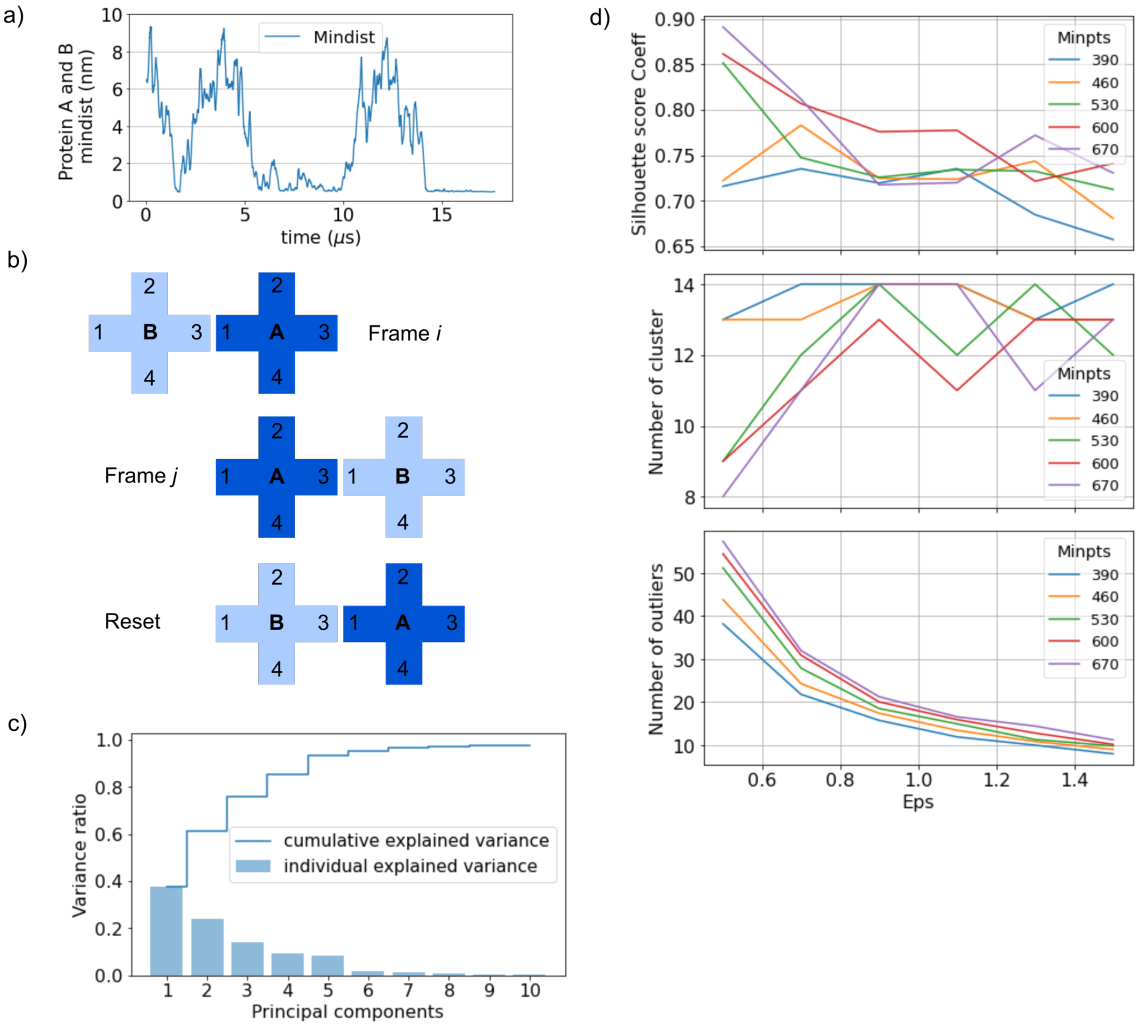


Figure 3.2: Clustering parametrization for the generated coarse-grained conformations. a) Trajectory simulation to predict hNav1.7α in dimer configuration, where the minimum distance between the protomers are tracked over the time. A dimer is defined when the protomers have a minimum distance of 1 nm. b) Identification and setting of conformations causing clustering errors. Similar conformations (frames i and j) can be assigned into two differents groups. The arrangement of interactions between frames are read in opposite direction, where same residues are involved in the interaction but not from same protomers identification (protomer A in dark blue and protomer B in light blue). To avoid the clustering errors, the protomers identification are swapped. c) Dimensionality reduction of 484 variables with PCA. d) On the up, application of silhouette score (y-axis) to chose the optimum DBSCAN hyper-parameters eps (on the x-axis) and $Minpts$ (in the legend). For each respective hyper-parameters, the number of generated clusters (middle) and the percentage of outliers which represent conformations not assigned to clusters (bottom) are represented.

orientation between the protomers and the residues involved in the interaction. Clusterisation errors might arise for protein homodimer. This is described by the attribution of similar structures in two distinct clusters – because the interaction arrangement differs between two frames (Fig. 3.2b). To avoid such errors, the chain identification of the protomers are swapped according to their position, view the Fig. 3.2b. Thus, all trajectories

are structurally aligned at the coordinates of first protomer (pA). If the second protomer (pB) is located on the right of pA, the chain identification of pA and pB are switched.
The description provided by the selected features are simplified with PCA, to transform the dataset into lower dimension. Three principal components (PCs) were able to explain 80% variance in the data (Fig. 3.2c).
The lowered dimension data was processed to DBSCAN algorithm. In order to efficiently cluster the visited conformations, the optimum hyper-parameters (eps and $Minpts$) were evaluated firstly with the 50% of the dataset. Elbow method was used to approximate the optimum values of the eps. Thus with the silhouette score coefficient was evaluated the clustering confidence with eps on the range 0.5 to 1.5, accordingly the $Minpts$ from 390 to 670. For any eps and $Minpts$, the coefficient of the silhouette score predicted a good confidence, which are ranged between 0.60 and 0.90 (Fig. 3.2d). This might illustrate good conformations sampling during the simulations and a convergence to some conformations. Reminder, the starting conformations used unbound and randomly orientated proteins. Moreover, for any hyper-parameters were generated relatively similar numbers of cluster (Fig. 3.2d). Which can conclude that DBSCAN is able to separate distinct conformations. With strict hyper-parameters i.e shorter eps and larger $Minpts$, the silhouette score describe homogeneous and well separated clusters. With this latest, large amounts of conformations are left as outliers (Fig. 3.2d). Thus, the clusterization of the whole dataset was defined with $Minpts = 1200$ within $eps = 0.9$. DBSCAN has assigned 18 clusters. In Fig. 4.1 is projected the dataset on the three first components.

Conformations similarities were also scaled with all-to-all RMSD, further analysis was quickly abandoned. The time complexity of all-to-all RMSD computation exponentially increases with the size of our data set, which handle less efficiently the analysis.

3.6 Backmapping of coarse-grained models into all-atom

In order to evaluate the stability of dimer configurations predicted from CG-MD simulation, the CG dimer models were converted in all-atom resolution. This conversion of atoms representation is called backmapping. It aims to gain finer molecular insight (i.e in all-atom resolution) of configurations, reachable at long time scale simulation as produced by coarse-grained simulation.

The backmapping tool developped by the Marrink group was used for the models conversion, called $initram - v5$ (Wassenaar et al. 2014). The backmapping requires a list of particle correspondences of the AA and CG resolution. For each residue, the transformation from CG beads to AA resolution is mapped into files. These mapping files are available in the website of Martini group.
To perfom the backmapping, the script $initram - v5$ requires a CG input structure and the corresponding atomistic topolopy. The conversion from CG to AA scale can induce

atoms clashes. To reduce eventual failing during the conversion, only the proteins were converted. The lipids membrane and the NaCl aqueous solution were excluded for the conversion.

Representative configurations highlighted from the section 3.5.1 were backmapped. After the conversion, the AA model is relaxed by processing energy minimization and positions restraints. These steps are already integrated in $initram - v5$, the parameters were kept at defaults as following :
(i) Energy minimization excluding non-bonded interactions (500 steps)
(ii) Energy minimization of the whole particles in the systems
(iii) Position restrained NVT simulation (300K), with integration step of 0.2 fs
(iii) Position restrained NVT simulation (300K), with integration step of 0.5 fs
(iii) Position restrained NVT simulation (300K), with integration step of 1.0 fs
(iii) Position restrained NVT simulation (300K), with integration step of 2.0 fs

Once the model converted in all-atom, molecular dynamics simulations were performed to evaluate the stability of dimer configurations. For the MD, two different setting of the system were (can be) used. In the first setting, the converted AA-dimer model is prepared similarly as decribed in section 3.1. The new labeled residues defining the glycosylation sites (Asn) covalently bound with a sugar are inserted. The N-terminal and C-terminal regions are capped with ACE and NME. The residues 180 and 1267 are deprotonated according to PROPKA3.1.
In the second setting, to avoid large time of protein preparation, a reference AA-structure of hNav1.7 (a priori prepared) is fitted into the converted dimer model.
All simulations were performed with GROMACS. The systems were parameterized as described in section 3.2. Each AA-dimer model is embedded into a pre-equilibrated lipids membranes. The systems were minimized for 5,000 steps using integration step of 0.1 fs. Also three steps of equilibration were applied. The first, the second and the third equilibration were carried respectively for 4 ns, 50 ns and 100 ns.

3.7 Protein-protein docking parameters

An ensemble docking consists of using multiple conformational states of the protein(s), in order to include or to mimic flexibility in docking prediction. The interaction complexes prediction (i) hNav1.4α/β4, (ii) hNav1.7α/β4 and (iii) hNav1.7α/hNav1.7α dimer were performed with the standalone HADDOCK version 2.2 and 2.4 (Ambrosetti, Jandova, and Bonvin 2020), using the ensemble docking approach.

For the complexes prediction hNav1.4α/β4 and hNav1.7α/β4, an ensemble of 10 to 20 hNav1.4α or hNav1.7α conformations, with one structure model of Navβ4 subunit were used to initiate the docking. The ensemble conformations of hNav1.4α and hNav1.7α

were extracted based on the conformational changes on the extracellular structures, simulated during the AA-MD. Mutliple simulation systems were used to extract the ensemble of conformations, apo-hNav1.4α, hNav1.4α/β1, apo-hNav1.7α, hNav1.7α/β1 or hNav1.7α/β1/β2.

To drive the docking, hNavβ4 subunit is known to establish a covalent interaction, mediated by the residue Cys58-hNavβ4/Cys895-hNav1.7α, or Cys58-hNavβ4/Cys729-hNav1.4α. These residues are defined as an "active" ambiguous distance restraints (AIRs); a distance restraints of 0.35 nm was defined. In the file parameters run.cns, the rigid body docking (it0) is set to generate a total of 3,000 to 60,000 structures. According to the implemented HADDOCK scoring, the top 10 to 30% of the structures are processed to the semi-flexible refinement (it1). This introduce flexibility on the interaction interface of the predicted complexes. Then HADDOCK re-evaluate this conformational changes, rank the best 200 to 1,000 structures and execute structures refinement with explicit water.

For the prediction of hNav1.7α/hNav1.7α dimer, the docking were initiated with an ensemble of 10 to 20 conformations for each protomer. Similarly to the setting of the previous complexes prediction, these ensemble of conformations were extracted from AA-MD; however, the choice of the starting structures were based on the conformational changes in S1–S5 transmembrane segments (TMs) of hNav1.7α. No AIRs were defined to drive the dimer prediction – any experimental investigation has highlighted critical residues involved in the interaction, precisely in the models built with the available experimental structures of sodium channels. However, we could define randomly ambiguous interaction restraints ($ranair$) of residues to drive the docking. Residues in viviniy to the lipids membrane located at the TMs were selected for $ranair$. I observed that by rescaling down to 0.01 the intermolecular interactions non-bonded energy (inter_rigid), a tighter hNav1.7αs interaction is observed. To avoid unrealistic docking, the membrane positioning restraints (z-restraining, only with HADDOCK2.4) were activated and settled on the height of the TM segments. For the prediction of larger systems, the generated structures were reduced due to the time complexity. Thus, at it0 stage, 3,000 to 10,000 structures were generated. The top 1,000 to 4,000 structures were processed for the it1 stage and the first 200 to 1,000 structures were refined with explicit water.

To sample as much larger predicted structures, the random seed (iniseed) were constantly modified for each prediction run of the complexes hNav1.4α/β4, hNav1.7α/β4 and hNav1.7α/hNav1.7α dimer. The remaining parameters were kept at the default.

3.7.1 Protein-protein docking analysis

In the table A.2 at the section A is summarized the prediction of all performed protein-protein interaction.

3.7.1.1 For the complex hNav1.4α/β4

The prediction of hNav1.4α/β4 complex was performed with 16 HADDOCK jobs. These have generated 27,000 configurations. The models of hNavβ4 structure were lacking the transmembrane segment and the C-terminal domain. Hence, some predicted configurations are unrealistic, where the Ig-domain of hNavβ4 were partially buried in the membrane or pointing vertically to the extracellular. These obvious unrealistic configurations were filtered out, 13,300 structures remain for the clustering. The clusterizations were based on the Cα Cartesian coordinate of hNavβ4, while hNav1.4α subunits were maintained at fixed position. The predicted interaction sites of hNavβ4 were described on 237 variables. These variables were transformed into a lower dimension with PCA. Two components were able to explain almost 80% variance in the dataset. To extract relevant configurations of hNav1.4α/β4 complex, the DBSCAN algorithm was used. The optimum parameters were defined with the elbow method and the silhouette score coefficient. The predicted configurations were clustered with at least $Minpts = 115$ and $eps = 1.1$. The docked configurations were regrouped into 7 clusters.

3.7.1.2 For the complex hNav1.7α/β4

To predict the interaction sites of hNavβ4 on hNav1.7α, 26 HADDOCK predictions were performed. These resulted to 23,800 hNav1.7α/β4 configurations. As in the section 3.7.1.1, unrealistic configurations were filtered out, 9,000 structures remain for further analysis. The Cα Cartesian coordinates of hNavβ4 were defined as features to cluster relevant conformations, resulting to 237 descriptors. All predicted conformations were aligned at a fixed position of hNav1.7α. A dimensionality reduction with PCA were applied to the dataset, where two components were able to explain almost 80% variance in the dataset. The clustering was performed with DBSCAN algorithm. The $Minpts$ and eps were defined for 55 and 1.4, respectively. Thus, 8 clusters were identified to define the visited interaction sites in hNav1.7α/β4.

3.7.1.3 Prediction of the hNav1.7α/hNav1.7α dimer

To predict the conformation of the hNav1.7α/hNav1.7α dimer, 107 HADDOCK prediction were performed. These have generated 79,500 dimer configurations. Despite the advance setting in HADDOCK 2.4 version for membrane proteins, unrealistic dimer structures were generated. The dimers were predicted correctly on the xy-plane, but they adopted ups and downs orientation. These structures were filtered out and 36,500 structures remained for the clustering. Similarly to the CG dimer analysis (section 3.5.1), the chains identification between the protomers were corrected if necessary, and features

were selected in order to cluster similar configurations. For each protomer, 22 atoms of Cα were selected and the pairs distance between the protomers'atoms were computed. This resulted to 484 distances of atoms-pair. These features were transformed into lower dimension with PCA. As with the CG dimer analysis, three PCs were able to explain 80% variance in the data.
The algorithm k-means was used to cluster the data. The optimal number of clusters k were defined according to the elbow method and the silhouette score (fig 4.3). The approximation of the optimum k were evaluated based on several values at the concave curve of the elbow method and validated with the silhouette coefficients. Then, the structures homogeneity within each cluster were evaluated. Thus, 35 clusters configurations were computed with k-means. The algorithm sought randomly 100 positions of the clusters centroids and the process was iterated 200 times.

DBSCAN algorithm was also used to cluster the data generated from the docking method. It has demonstrated lower efficiency to separate some configurations without excluding large number of outliers. In Fig. 4.3c and Fig. 4.3e can be compared the structures homogeneity within each cluster, respectively performed with the algorithms k-means and DBSCAN.
DBSCAN shows higher performance with larger data and more continuous data points.

3.8 Prediction of hNav complexes with AlphaFold2

AlphaFold2 version 2.0 and 2.2.0 (AF2.0 and AF2.2) (Evans et al. 2021; Jumper et al. 2021) were used to predict the structures complexes hNav1.4α/β4, hNav1.7α/β2, hNav1.7α/β4 and hNav1.7α-hNav1.7α dimer. Multiple prediction jobs were performed for each complex system, and surprisingly resulted to different conformations. For each AlphaFold2 prediction, five models were generated.
For the complexes hNav1.4α/β4, hNav1.7α/β2 and hNav1.7α/β4, prediction were conducted with AF2.0 version. With the newer version AF2.2, no interaction between the chains was observed after all attempt of parametrization.
For the input sequence corresponding to hNav1.7α, we tempted to predict the full length sequence. The structures located at the intracellular side were predicted in coily secondary structures, and presented very low plDDT confidence score. In order to improve the models confidence, the corresponding intracellular structure of the hNavα were excluded from the input sequence. The sequence of hNav1.4α and hNav1.7α were truncated similarly as the resolved in the PDB ID 6AGF and 6J8I, respectively. For hNav1.4α, the structure were predicted at the sequence range 119–463, 559–806, 1014–1607. For hNav1.7α, the structure were predicted at the sequence range 114–417, 726–972, 1175–1768. The truncated intracellular sequence were replaced with shorter Glycine linker. This linker is labeled intra-GLY linker.
The full sequence of hNavβ2 and hNavβ4 were predicted, referring to the residues range

31–216 and 30–228, respectively. These sequence ranges include the Ig domain, the transmembrane segment and the C-terminal domain. In order to assess and identify the critical residues involved in the complex interaction, some domain(s) of the hNavβ were removed in the prediction. Additionally, single or multiple mutations were inserted in the input sequence of hNavβ4. Residues were mutated according to the literature Grieco et al. 2005), or based on visual observation of the predicted structures with AlphaFold2.

The complexes hNav1.7α/hNav1.7α were predicted with the version AF2.2. In the previous version of AlphaFold2 (AF2.0), the tool was unable to predict a plausible conformation of hNav1.7α dimer. The hNav1.7α protomers were overlapping. In AF2.2 version, the clash loss and center-of-mass loss functions were redefined, allowing for complexes prediction a tighter interaction and avoiding overlapping structures. In the input sequence, different truncation of hNav1.7α sequence were tried out. Firstly, residues corresponding to the intracellular sequences of hNav1.7α were removed as in the previous prediction of hNav1.7α with β subunits. These sequences were replaced with intra-GLY linker.
Then, in order to investigate the approach in Clatot *et al.* paper, the DI-DII linkers of each respective protomer were included in the input sequence in some prediction. DI-DII linker were truncated at different position as investigated in this paper (Clatot et al. 2017).
Based on the dimer prediction with CG approach, different interaction interfaces were involved. With this latest and the Clatot *et al.* research, we would like to investigate if the DI-DII linker of Navs could establish an interaction with the intracellular structures of the other domains. Thus, in the input sequence for the dimer prediction, the first sequence of hNav1.7α included the linker DI-DII and the second sequence included either DI N-terminal, DII-DIII linker or DIV C-terminal.
For all predictions conducted with AF2.2 (this is not the case for AF2.0 version), to enable (or to force) interaction between the two hNav1.7α protomers, the input sequences were set in one single sequence. The two sequences of hNav1.7α were linked with glycine linker, such as hNav1.7α – GLY linker – hNav1.7α . This linker is labeled as extra-GLY linker. However, the prediction were conducted as multimer (-m multimer).

Chapter 4

Prediction of hNav1.7α-hNav1.7α dimerization

Multiple molecular computational methods were performed to predict the configuration(s) of hNav1.7α in a homodimer. CG simulations using the DAFT approach, molecular docking using the HADDOCK tool and ab-initio predictions with AlphaFold2 (AF2) were used. The combination of these methods can provide complementary insights and the validation for the predicted configurations of hNav1.7α in a homodimer. Moreover, employing various approaches aims to assess the robustness of the methods.

4.1 Dimer configurations predicted with the DAFT approach

The CG simulations applying the DAFT approach have generated 290,000 dimer configurations. In the section 3.5, hNav1.7α in dimer configurations were regrouped into eighteen clusters (Fig. 4.1 and Fig. 4.2). 12 of the 18 identified clusters were visited by multiple independent CG simulations – suggesting dynamic exchange of multiple dimer conformations (Fig. 4.1a). The 18 clusters could be further grouped by dimer-interface similarity, demonstrating a preferred dimer-dimer orientation (Fig. 4.2). The dominant dimer interaction mainly involves the DI–DII interface, observed in clusters 1, 2, 5, 6, 7, 9, 12, 15 (Fig. 4.2a), and clusters 3, 8 (Fig. 4.2b). The DII–DIII interface exhibits also a prevalent interaction, observed in clusters 2, 5, 12, 15 (Fig. 4.2a), clusters 11, 17 (Fig. 4.2b), and clusters 4, 16 (Fig. 4.2c). This is followed by the interaction preference at the DIV-DI interface in Fig. 4.2b. However, dimerization at the DIV–DI interface displays lower frequency (shorter lifespan).

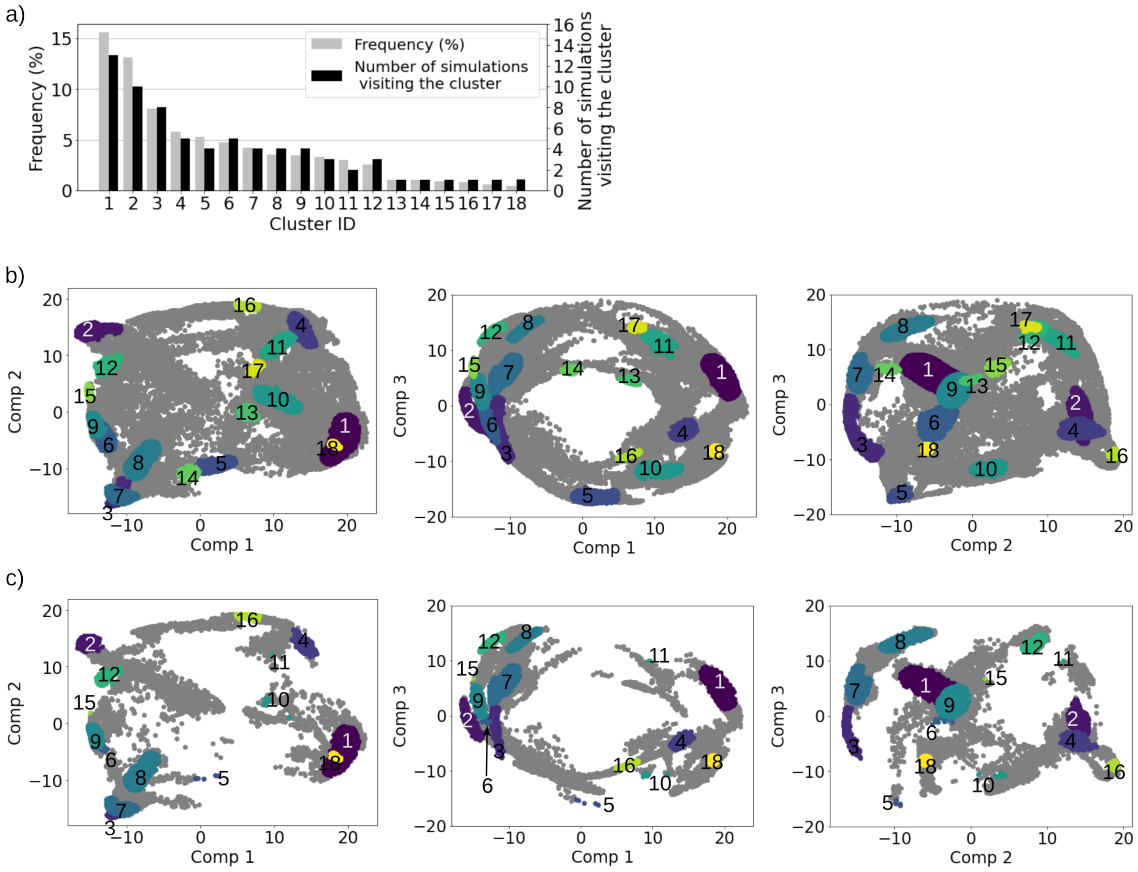


Figure 4.1: DBSCAN clustering of the dimer configurations. a) Distribution frequency of structures assigned to clusters (in grey bar) and the number of visited simulations by each cluster (in black bar). b) Data projection representing all predicted dimer frames (simulated in pure POPC and mixed membranes) onto the first three principal components. The outliers (in grey) represent 22% of the dataset. Each cluster is assigned to different colors. c) The predicted configurations simulated in mixed membrane projected onto the three first components. Most clustered configurations visited with the pure POPC membrane were also visited in the simulations with mixed membrane. However, new configurations (clusters 16 and 18) are predicted with the mixed membranes simulations, but with very low frequency (less reproducible).

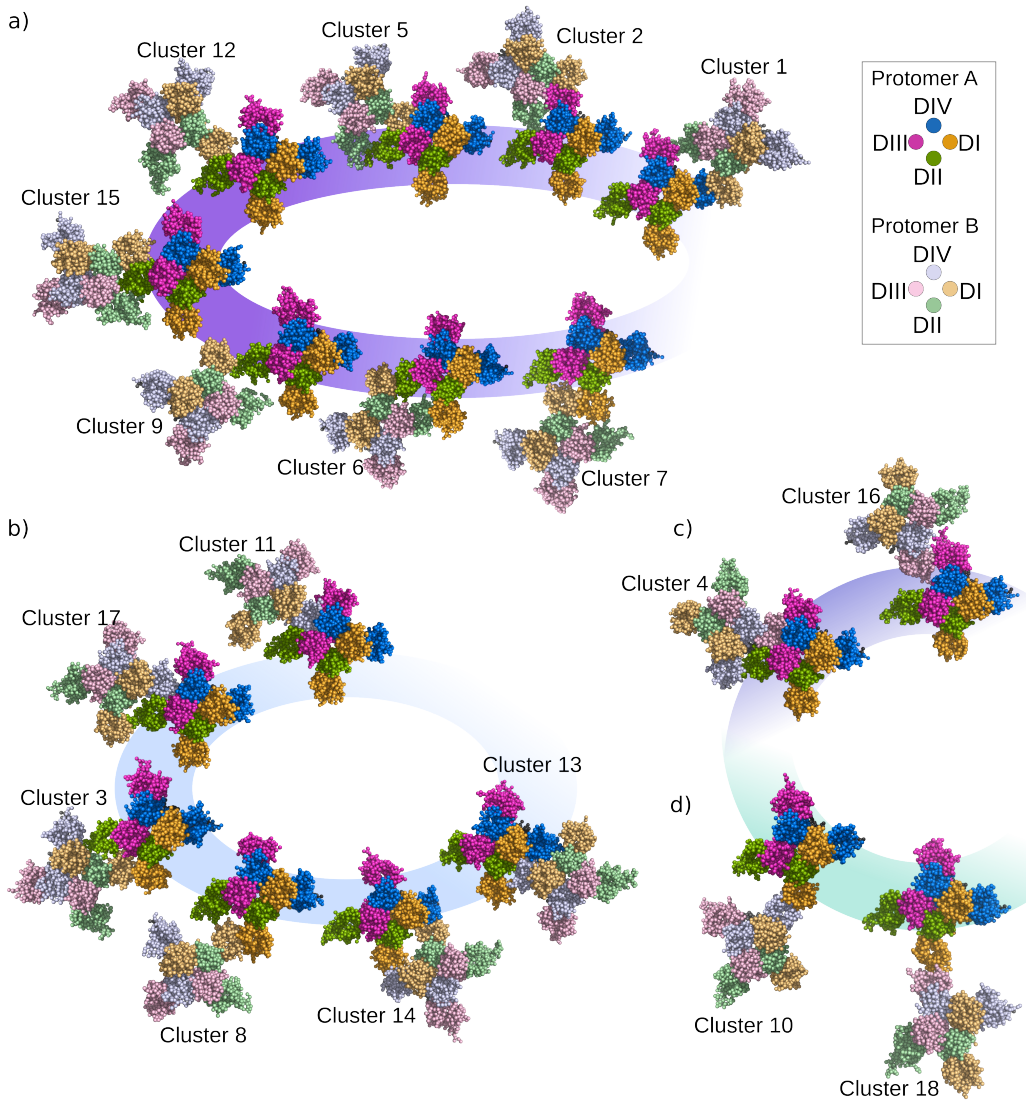


Figure 4.2: Structural representation of the eighteen clustered dimer configurations predicted with the DAFT approach. In order to highlight the visited interface of protomer B (lighter color), the configurations structures are fitted on protomer A (deeper color). DI, DII, DIII and DIV are highlight in orange, green, pink and blue, respectively. For an easy illustration, clusters sharing the same interface are regrouped into a path. In a) are the regrouped clusters involving DI–DII interface (orange-green) of protomer B in the dimerization. In b) are the clusters involving DIV–DI interface (blue-pink) of protomer B in the dimerization. In c) are the clusters involving DII–DIII interface (green-pink) of protomer A in the dimerization.. In d), the VSDI of protomer A are involved in the interactions.

4.2 Dimer configurations predicted with the docking method

The docking prediction has generated 36,500 dimer configurations (section 3.7.1.3), which were clustered into 35 clusters (Fig. 4.3). Multiple dimer configurations share the same or similar interaction interfaces. Similarly as with the DAFT structures analysis, in Fig. 4.4 are the clusters regrouped onto paths. All interfacial regions of hNav1.7α seem to be visited in the predicted dimer configurations. In Fig. 4.4a, the dimer configurations are regrouped according to interactions involving the DI–DII interface. In Fig. 4.4b, the dimer configurations are regrouped according to interactions involving the DII–DIII interface. Then in Fig. 4.4c is represented the dimer configurations involving DIII–DIV and DIV–DI interfaces.

4.3 Dimer configuration predicted with the AlphaFold2

Multiple AlphaFold2 jobs with variation at the input sequence of hNav1.7α were used to predict hNav1.7α in a dimer configuration (section 3.8). hNav1.7α sequence was truncated at different regions for each prediction; this has generated divergent conformations of hNav1.7α dimer. Thus, AlphaFold2 has predicted five dimer configurations. Most of the predicted configurations have symmetric interaction, i.e, both protomers utilize the same subunits in the interaction.

4.4 Interaction interface properties

Intermolecular interactions are predominantly driven by the physico-chemical properties at the interfacial region of the protein(s). For the homo-dimerization of hNav1.7α, all interfacial regions display similar interfacial shape – as the four homologous domains of hNav1.7α are nested in C4 symmetry.

In a complex of large macromolecules, defining prior potential features to evaluate the affinity and specificity of the interaction are less understood – in contrast to the knowledge on ligands-protein interaction. In 1984, Andrews and co-workers estimated the binding strength of functional groups for certain drugs/molecules (Andrews, Craik, and Martin 1984) – in 2004, Hunter *et al.* conclude importance of electrostatics bonding in intermolecular interaction (Hunter 2004). However, several investigations highlighted the prevalence of non-polar residues to be engaged in protein-protein interaction (Andrews, Craik, and Martin 1984, Freitas and Schapira 2017). Several investigation has also demonstrated that interaction in protein-protein complex tend at the most hydrophobic interfacial region. In 1996, Jones and Thornton observed that the interaction interface between homo-dimers involved more hydrophobicity than in hetero-dimers (Jones and Thornton 1996). Indeed, hydrophobic functional groups have wider possibilities of interactions. Eight out of twenty residues are classified as hydrophobic (Leu, Ile, Ala,

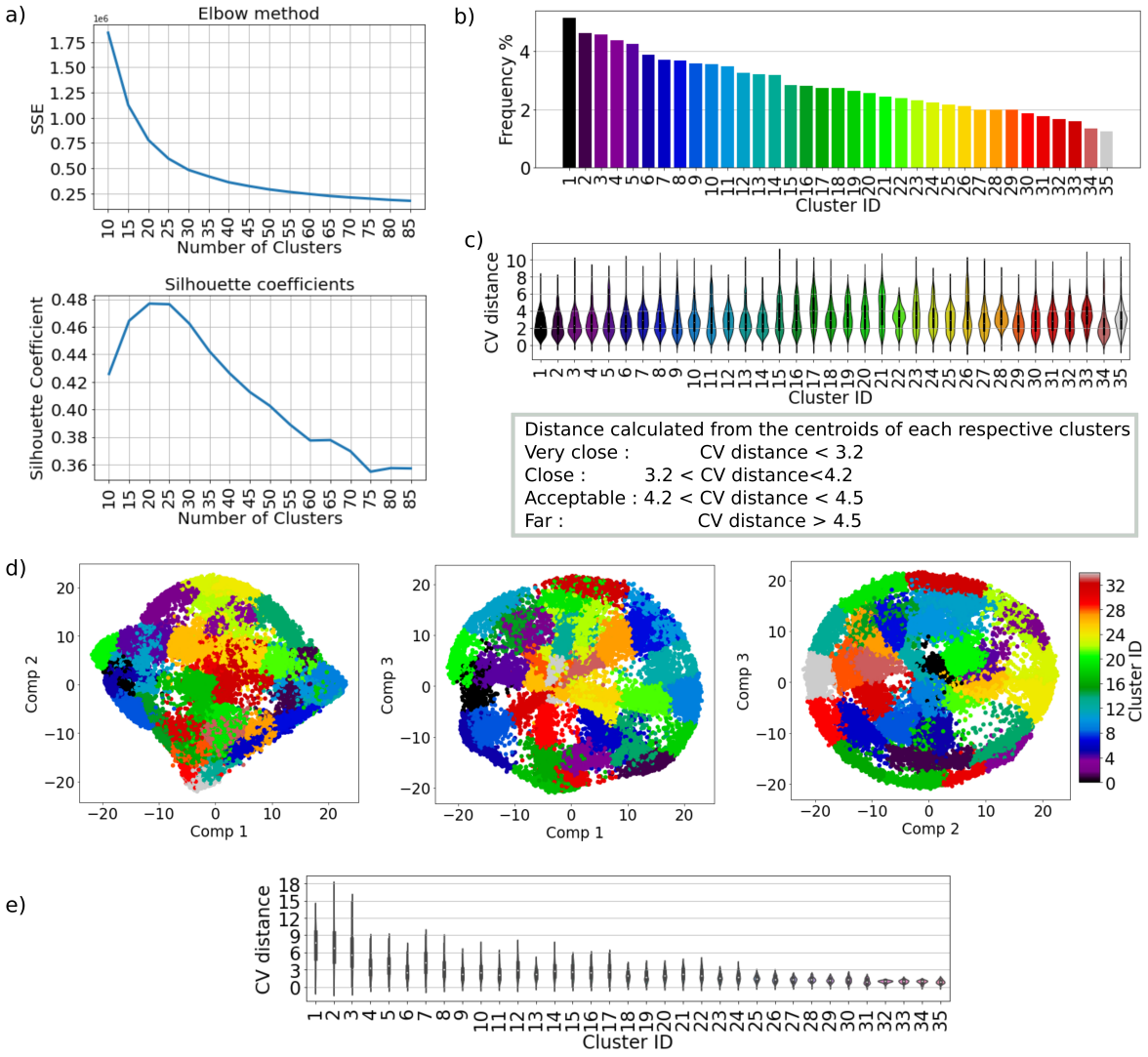


Figure 4.3: K-means clusterization of dimer configurations predicted with HADDOCK. a) The elbow method and the silhouette scores were computed for different k values. The dimer configurations were grouped in 35 clusters. In b) are represented the structures frequency within each cluster. c) The quality of the clustering were measured according to the distances between the centroid and the structures within each cluster. d) Projection of the clustered configurations on the three first components. d) Evaluation of the clusters generated from DBSCAN algorithm. The distances between the structures within each cluster and the corresponding center of cluster were computed.

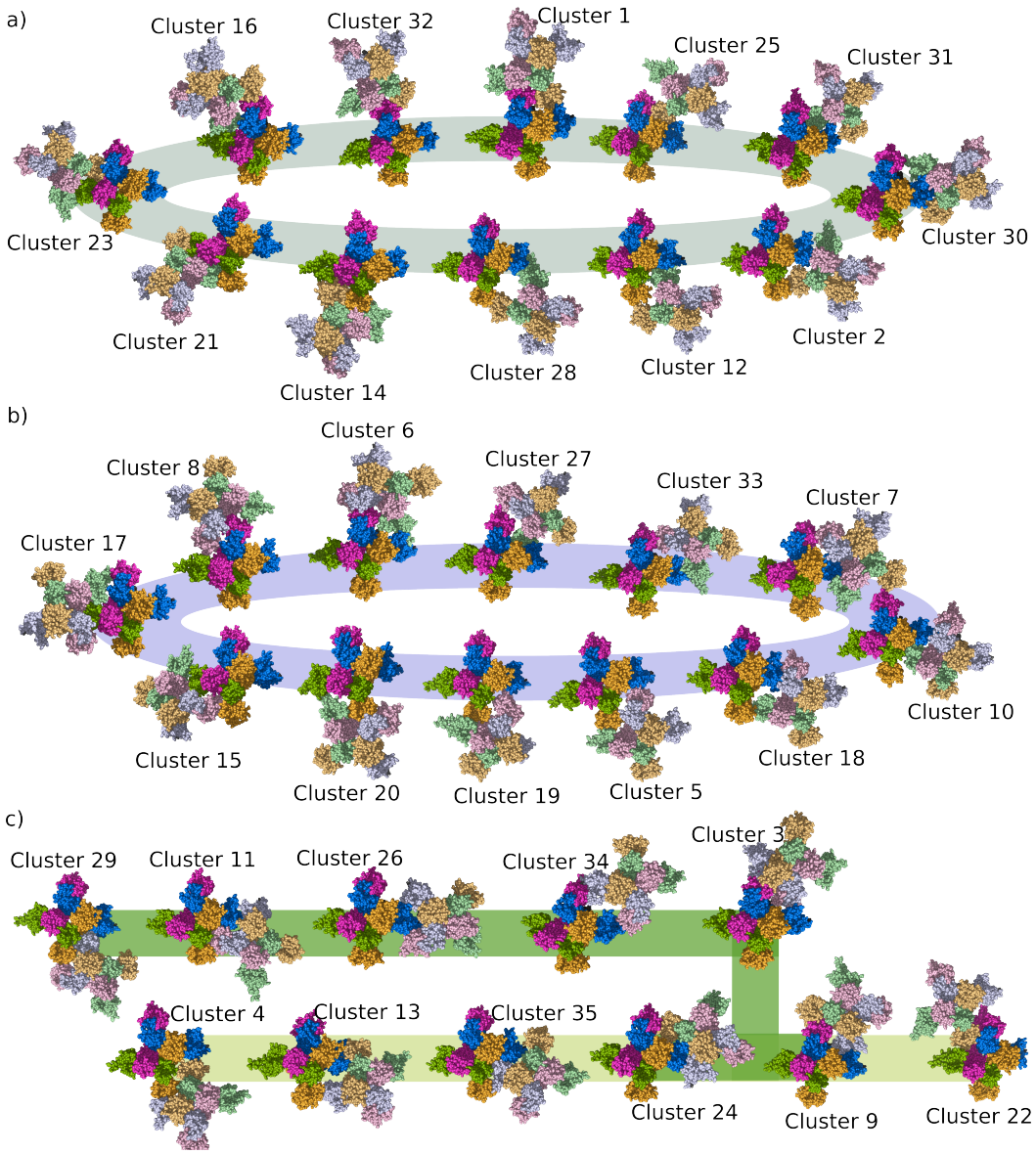


Figure 4.4: Representation of the dimer configurations predicted with the docking method. The structures are fitted on the protomer A (in deeper color) in order to highlight the visited interface of the protomer B (light color). The DI, DII, DIII and DIV are highlight in orange, green, pink and blue, respectively. For an easy illustration, clusters sharing the same interface (in protomer B) are regrouped. a) are represented structures involving the interaction interface DI–DII (orange-green) b) are the clusters involving the interaction interface DII–DIII (blue-pink). c) are represented clusters with DIII–DIV and DIV–DIV at interaction interface.

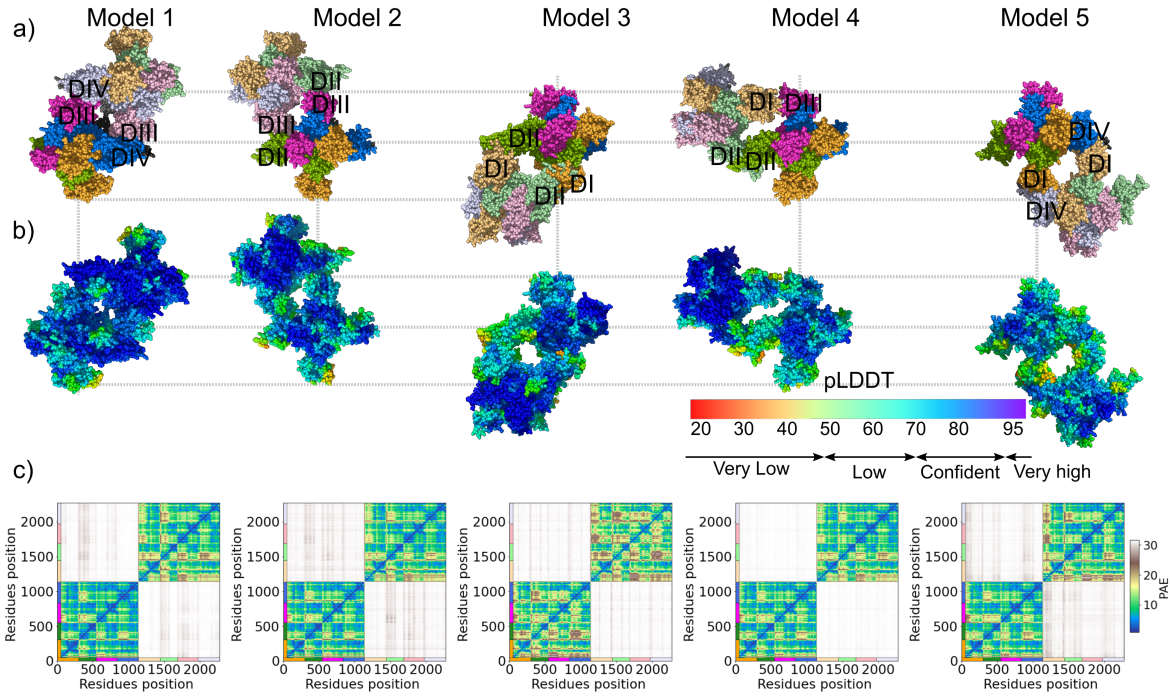


Figure 4.5: Predicted models of hNav1.7α homo-dimer with Alphafold. a) The dimer structures are fitted on the protomer A (in darker color), in order to easily visualize visited interaction of protomer B (in lighter color). DI, DII, DIII and DIV are respectively in orange, gree, pink and blue. b) The local confidence of the predicted dimer models are evaluated with pLDDT and c) the confidence of the relative residues positions are evaluated with PAE. This latest give the domains positions confidence.

Val, Met, Phe, Tyr, Trp). Thus in membrane proteins, the probabilities to establish an interaction is increased.
The aromatic hydrophobic residues Phe, Tyr and Trp, in addition with the residue His and Pro can also be involved in electrostatic interaction, called stacking bonds. The cation-π interactions also involve the aromatics residues (His, Pro, Phe, Tyr, Trp) to bond with a positively charged residues (Arg, Lys).
The salt-bridge interaction is mediated between a positively charged residues (Arg, Lys, His) and a negatively charged residues (Asp, Glu).
The amide bridge is established between two polar residues Asn and Gln.
Another most reproducible interaction type is the hydrogen bonds. They are mediated between the functional groups NH...O, OH...O and NH...N. The functional donor groups can be yielded by the polar residues (Asn, Gln, Ser, Thr), the aromatic and hydrophobic residues (Trp, Tyr) and the positive residues (Arg, Lys, His). The functional acceptor groups can also be yielded by the polar residues, the aromatic residues, the negative residues (Asp, Glu) and Gly.
Weak hydrogen bonds are mediated by CH...O, the most observed frequent acceptor are Gly, Asp, Glu and as the most frequent donor Leu (Freitas and Schapira 2017).

The strength of regular hydrogen-bonds are more steady and can be comparable to the salt brigdes, the cation-π interactions, and the amide bridges (Xie et al. 2015, link) However in aqueous solution, these interactions are more or less weakened; which could break or dissociate. In contrast, the hydrophobic residues are pulled together in aqueous solution. However, hydrophobic interaction might be transient. The electrostatic interaction between nonpolar molecules are mediated with instantaneous dipoles. Thus, the molecules composition and the electrostatic complementarity between the interfacial regions are important to maintain the stability (interaction strength) of proteins complex and drive the specificity of the interaction.

4.4.1 Mapping the interface in hNav1.7α to characterize interaction preference

For large macromolecules, the extent of hydrophobicity at a potential interfacial region (often) plays an essential driving force in proteins interaction. Obviously, in membrane protein such as the hNav1.7α, the transmembrane regions are preponderantly hydrophobic. But are all membrane proteins capable of homodimerization? Evidently, the polar and charged residues scattered at the interfacial interaction play important roles in the specificity between the binding proteins and thus their interaction stability.
In molecular computational predictions, the tendency or reproducibility of dimer configurations can decrees the stability of the interaction. The stability of hNav1.7α dimer can be quantified by its life span (frequency) occurred in a single and/or multiple prediction(s) (Fig. 4.1a, Fig. 4.6a). The prediction reproducibility of a configuration can also

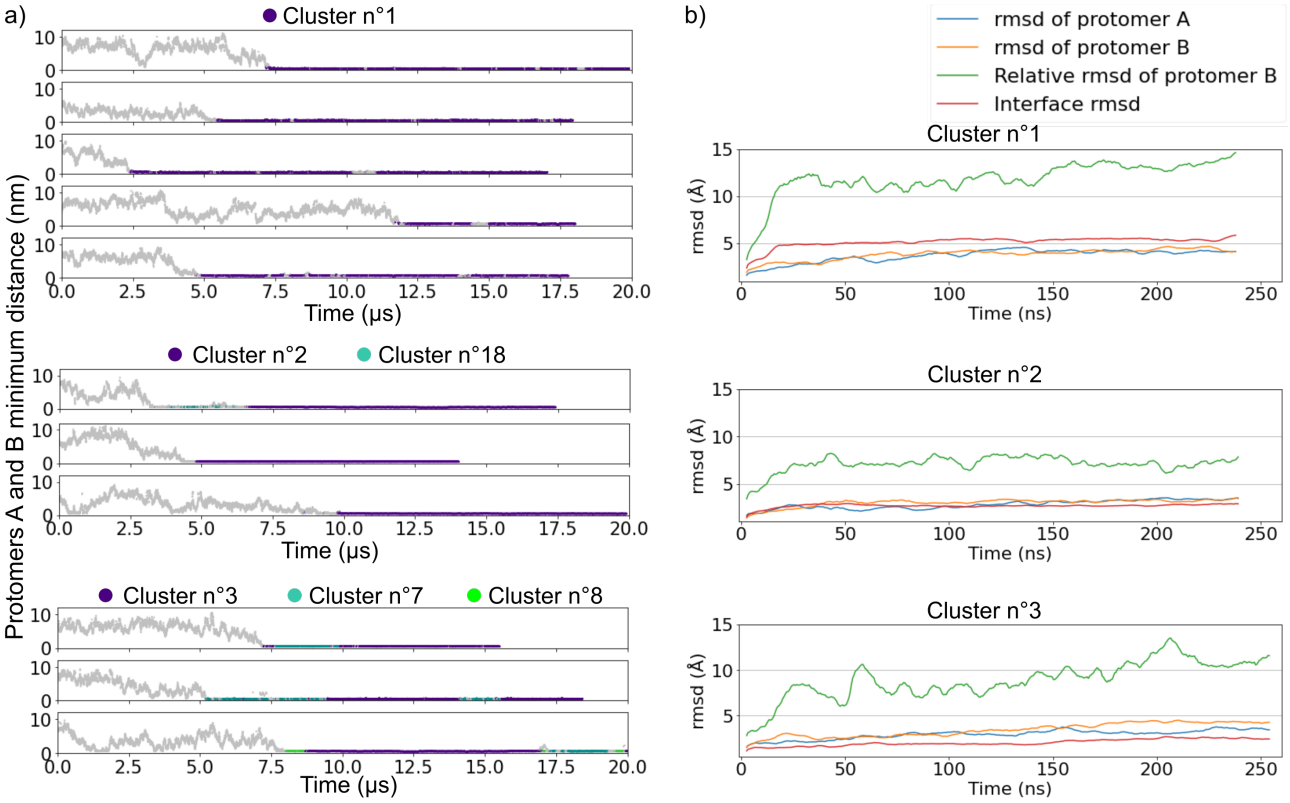


Figure 4.6: Stability analysis of the dimer configurations assigned in clusters, simulated in a) CG and b) AA scales (the backmapped dimer configurations) . In a) are represented some trajectories to predicted the hNav1.7α dimer configurations. The y-axis represent the distances between the protomers and the x-axis the simulation time. Structures regrouped into a cluster are highlighted in purple, cyan and green. Outliers structures and non-interacting protomers are in gray. b) The dimer configurations stability of the top three clusters are evaluated with an AA simulations, according to absolute rmsd in protomers A and B (blue and orange), the relative rmsd of protomer B where the coordinate of protomer A are fixed (green) and the interface rmsd (red).

be evaluated using different prediction methods (fig 4.7). In addition to all the methods performed to predict hNav1.7α in dimer configurations, the stability of these latest were also assessed in finer resolution representation, by backmapping the predicted CG configurations into AA models (Fig. 4.6b) and performing molecular dynamics simulation. Only the top four ranked configurations were backmapped. The predictions suggest that hNav1.7α can have multiple dimer configurations. In order to evaluate the preference of interfacial interaction in hNav1.7α dimerization, per-residue electrostatic calculations were mapped onto the structure in Fig. 4.8.

The DI–DII interfacial region is predominantly hydrophobic, lacking significant electrostatic interaction sites (Fig. 4.8c). The interfacial region at DII–DIII, DIII–DIV and DIV–DI have higher counts of charged residues, consequently, their interfacial region decrease in hydrophobicity compared at the interface DI–DII. The propensity of the hNav1.7α dimer to form interactions at the DI-DII interface represents the most populated configuration.

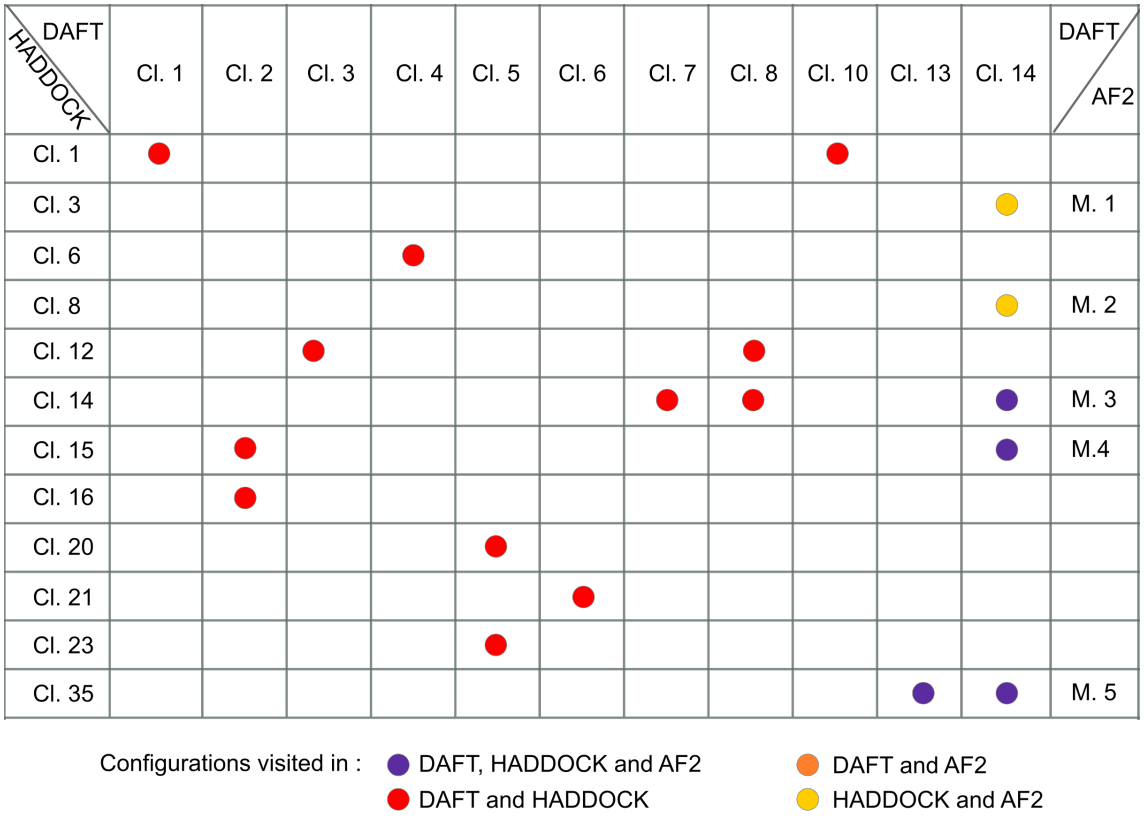


Figure 4.7: Reproducibility matrix of configurations generated from DAFT, HADDOCK and AF2. For DAFT and HADDOCK prediction, non exhaustive predicted configurations are represented.

The configuration involving the DI–DII interface are visited in the top clusters generated by the different methods. In addition, the DI–DII interface have the ability to establish interaction with the other interfaces. The extent of the hydrophobic surface at DI–DII allows the latter to interact on all hNav1.7α interfaces.

The hydrophobicity at the DII–DIII interface is weaker compared to DI–DII. However, DII–DIII interface has greater hydrophobicity in comparison to DIII–DIV and DIV–DI. This offers an explanation for the DII–DIII interaction propensity at the second position.

The DIII–DIV interface is the most prevalent interaction according to clusters of configurations generated from both DAFT and HADDOCK. The interfacial region DIII–DIV seems to match only with the DI–DII interface. DIII–DIV does not establish steady interaction with the other interfaces. The DIII–DIV interface is highly rich in positive residues.

Multiple configurations involved DIV–DI in the dimer interaction, but these configurations exhibit shorter lifespan. Interaction involving DIV–DI interface has weakened stability. Indeed, the interface of DIV–DI are highly charged (with positive and negative residues). This can reduced the electrostatic complementarity with the other interfaces, and rapidly disrupt the interaction.

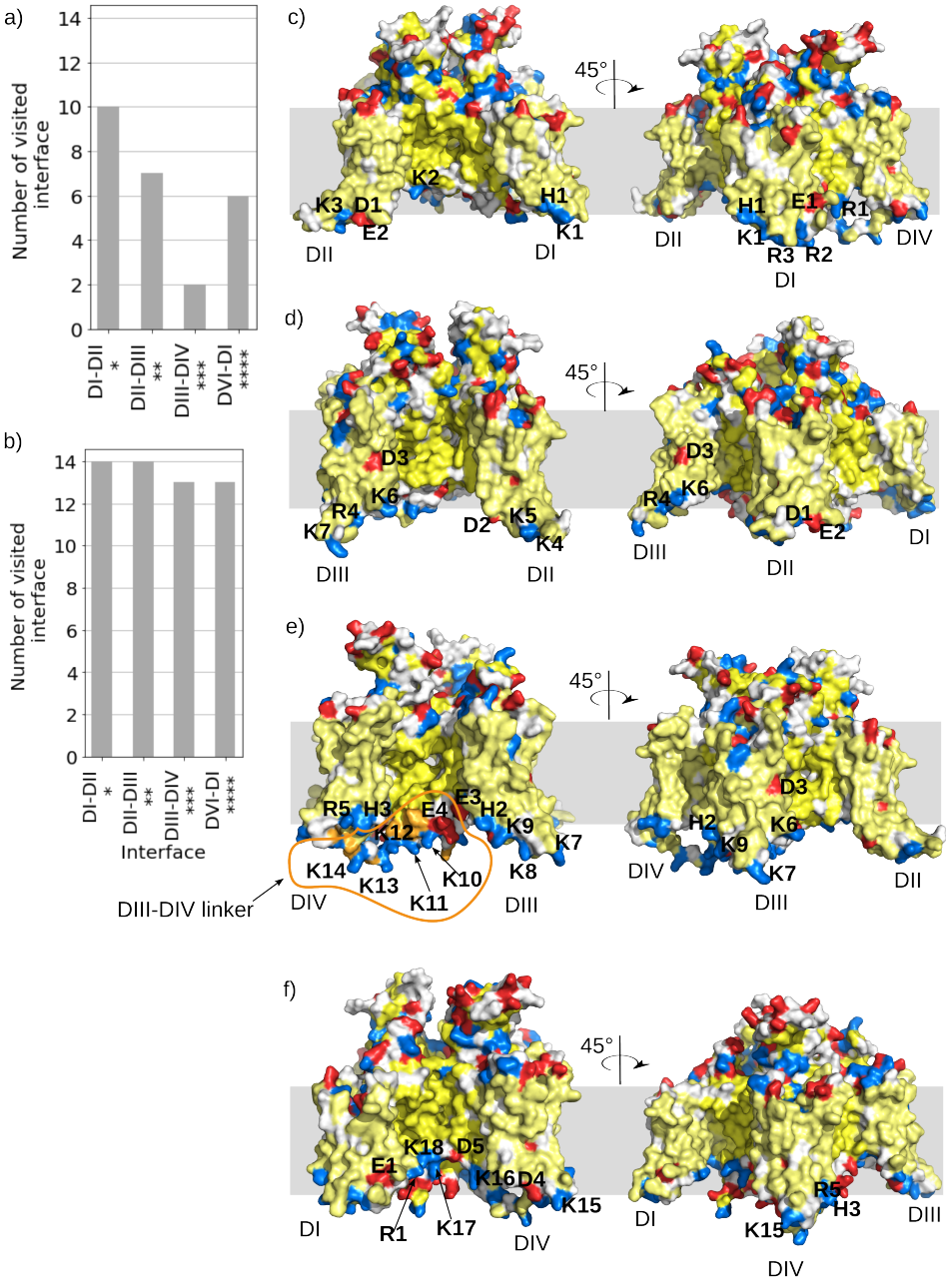


Figure 4.8: Mapping of the hNav1.7α interfacial residues. a) and b) are the number of visited interface by each cluster configuration, respectively generated with DAFT approach and HADDOCK. In c) d) e) and f) are the interfacial mapping residues at DI–DII, DII–DIII, DIII–DIV and DIV–DI. Hydrophobic residues are highlighted in yellow, polar residues are in white, positive and negative residues are in red and blue. At the DIII–DIV interface e), the DIII–DIV linker are highlighted in orange.

4.5 Interaction mechanisms

To initiate the dimerization, the interaction might firstly involve the VSDs domain, which are the most external structure of hNav1.7α. At the intracellular structure of hNav1.7α, the borders are scattered with more positively charged residues. In the most reproducible dimer configurations, involving DI-DII or DII-DIII interface, the most external structure (i.e the VSDs) are coated with few negatively charged residues (Fig. 4.8); E1 for VSDI and E2, D1, D2 for VSDII. The presence of these few negative residues is complementarily attracted to the positive properties at its protomer partner, and thus might initiate the hNav1.7α-hNav1.7α interaction. Then, the hydrophobic properties at hNav1.7α interfacial exert an attractive force to tighten hNav1.7α interaction and stabilize the dimer. For instance, as in cluster 1 generated in DAFT, Fig. 4.2a. Also, the hNav1.7α protomers can migrate into a tighter interaction, as in cluster 1 predicted with HADDOCK, Fig. 4.4a.

The other predicted configuration possibilities of hNav1.7α dimer demonstrate relatively stable interaction – multiple residues are in contact to establish tighter interaction.

4.6 Scenarios of functional coupling in hNav1.7α dimer

Until recently, voltage-gated sodium channels have always been considered as a monomer entity (Hoshi et al. 2014, Clatot et al. 2017). In patients with pain syndromes, the propagation of pain signals are characterized by an amplification of cellular hyperexcitability. This hyperexcitability might be promoted by a synchronization of the channel's pore activation (Meents et al. 2019). In few investigations, the sodium channels have been proposed to operate through a coupled-gating mechanism (Hoshi et al. 2014, Clairfeuille et al. 2019, Rühlmann et al. 2020).
In the investigations of Rühlmann and collaborators, they observed a significant reduction persistent current when uncoupling IEM-hNav1.7α variants. Additionally, when they co-expressed IEM-hNav1.7α and the WT-hNav1.7α variant, an induced dominant positive effect is observed. The persistent current characterizing IEM-hNav1.7α was significantly reduced. These experimental evidences demonstrated that hNav1.7α can indeed form a dimer configuration. In this study, multiple configurations of hNav1.7α in homo-dimer were predicted. These configurations provide insights into the modulation occurring in the coupled-gating mechanism, where the pores of the channels are suggested to be allosterically and/or directly modulated.

Three scenarios were suggested to interpret the functional coupling of hNav1.7α dimer. When the VSDIV is engaged in the dimer interaction, VSDIV might be trapped or modulated in its downward configuration, while the other VSDs are activated. This configuration describes the opened state of hNavs channel. By inspecting all the predicted dimer configurations, the DIII–DIV interface (representing high prevalence) established an in-

teraction with high specificity (only) at the DI–DII interface (Fig. 4.9A). Where VSDIV of the first protomer interacts with VSDI of the second protomer, and VSDIII interacts with VSDII. The height of VSDIV transmembrane segments are much more longer than VSDI. The VSDI could have an interaction preference with VSDIV, where the extracellular side of VSDI and VSDIV are aligned in the same plane, which means that VSDIV is at its downward configuration and VSDI in upward configuration.

The second scenario to explain the cellular hyperexcitability, hNav1.7α dimer could imply a direct modulation at the inactivation mechanism. If VSDIII-VSDIV interface is involved in the dimerization, the interaction of the IFM motifs (located at the VSDIII–VSDIV linker) to its receptor could be hindered or disrupted (Fig. 4.9B). The fast inactivation mechanism is not performed and the pores in the dimer remain open.

Finally, the dimerization involving the interface VSDI, VSDII and VSDIII can induce or trap the gating charge in the upward configuration (Fig. 4.9C). This promotes or increases the sensibility of the pore activation, which results to the cellular excitability.

4.7 Resume: investigation on the prediction of hNav1.7α in homo-dimer configurations

hNav1.7α has always been considered as a monomer entity. Few and recent investigations have suggested that the cellular hyperexcitability observed in patients with pain syndromes might be enhanced by a coupled-gating mechanims, where hNav1.7αs form a homo-dimer (Clairfeuille et al. 2019; Hoshi et al. 2014, Hampl *et al.* manuscript in progress). hNav1.7α is a very large protein. In order to efficiently predict its homo-dimer configurations, we performed the following methods: CG molecular dynamics simulation, protein-protein docking and AlphaFold2. Representative configurations were regrouped into clusters. Multiple dimer configurations were visited by the different methods cited previously. Interface preferences were observed in the dimerization of hNav1.7α, where the most hydrophobic interfaces have higher prevalence to engage dimer interaction. According to the involved interface in the interaction, we suggested that dimerization of hNav1.7α might promotes the activation state of the sodium channels, by interfering on the activation of VSDI, VSDII and VSDIII. The dimerization might also impede directly or allosterically the fast inactivation mechanism.

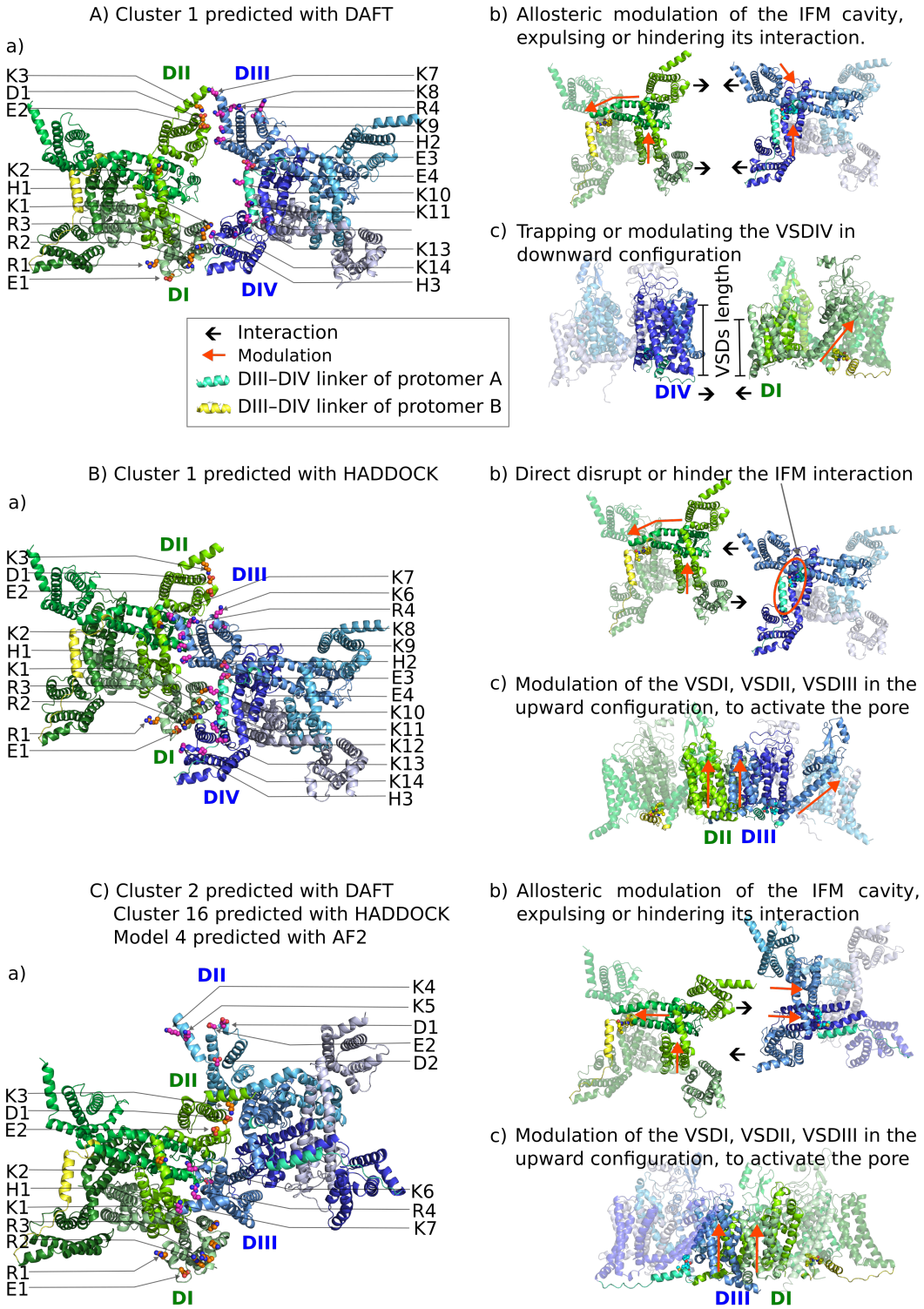


Figure 4.9: Characterization of hNav1.7α dimer modulation. On the left are highlighted the potential charged residues at the interaction interface between the protomer A (in blue) and the protomer B (in green), shown from the intracellular view. On the right are proposed modulation scenarios induced in coupled-gating mechanism, when hNav1.7αs form dimer configuration involving different interfacial interaction.

Chapter 5

Prediction of hNavβ2 and β4 interaction sites on hNav1.7α

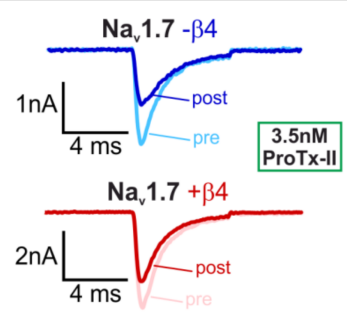


Figure 5.1: Pharmacological modulation of Protx-II on hNav1.7α, in the presence and absence of hNavβ4 subunit. (figure extract from the manuscript of Hampl *et al.*, manuscript in progress)

The mechanism of Navα can be modulated by one or multiple Navβ subunit(s), small molecules or peptides toxins (O'Malley and Isom 2015). Experimental investigations have demonstrated a high potential pharmacological effect of the tarantula Protx-II on hNav1.7. Protx-II is known to establish an interaction at three distinct sites of hNavs, located at the S3–S4 linker of the VSDI, VSDII and VSDIV (Bosmans, Martin-Eauclaire, and Swartz 2008). When the Protx-II interacts at the VSDI or the VSDII of hNavs, the mechanism of activation is inhibited, occurring the reduction of Na^+ current. In contrast, the mechanism of inactivation is inhibited when Protx-II interacts at the VSDIV of hNavs. The fast inactivation mechanism is then impeded which resulted in a persistent current. Protx-II exhibits a high affinity to the isoform hNav1.7α (Xiao et al. 2010, Bosmans, Martin-Eauclaire, and Swartz 2008).
Hampl *et al.* (manuscript in progress) is investigating the potential function of hNavβ4 as a pharmacological modulator (Fig. 5.1). When they co-expressed hNav1.7α and Protx-II

with hNavβ4, the pharmacological effect of Protx-II on hNav1.7α is weaken. The presence of hNavβ4 seems to induce a recovery of the Na^+ current in hNav1.7α/Protx-II complex. From this emerged a first suggestion, do hNavβ4 and the toxin Protx-II have competitive interaction sites? Literatures have suggested the VSDII of hNav1.7α, as the main binding site of Protx-II (Xiao et al. 2010, Hampl *et al.* manuscript in progress).
Few investigations show interests on hNavβ4. Its functions are poorly known and the experimental structure of hNavβ4 interacting with a sodium channel has not yet been characterized. In order to characterize the pharmacological functions of hNavβ4, the interaction sites of hNavβ4 are predicted using multiple computational methods. As controls and references prediction, the configuration of hNav1.7α/β2 and hNav1.4α/β4 complexes were also predicted.

5.1 All-atomistic molecular dynamics prediction

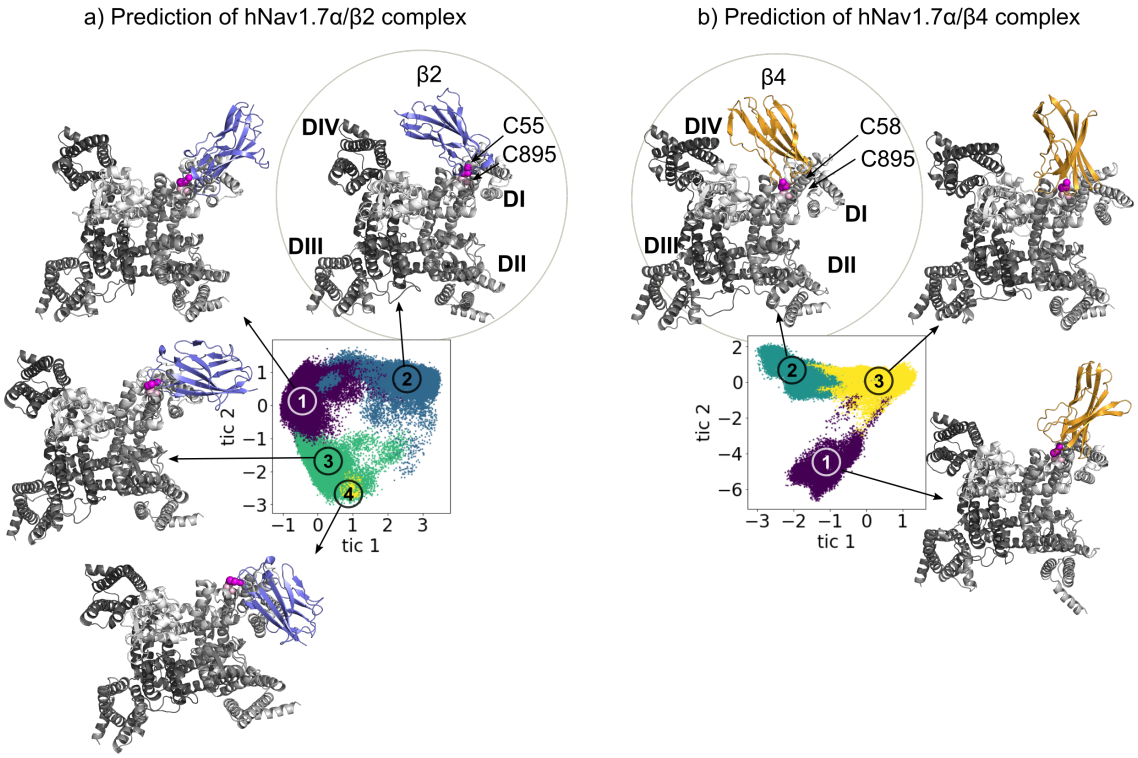


Figure 5.2: Interaction sites prediction for the Ig domain of hNavβ2 (a) and β4 (b) on hNav1.7α, performed with molecular dynamics simulation. The visited configurations by hNavβ2 and hNavβ4 are projected into the two first independent components of TICA. The regions assigned for each metastables states are annotated with different colors. The conformations representing the centroids of each metastable state are shown in cartoon; the hNav1.7α are highlighted in shades of grey according to the domains, hNavβ2 and β4 are in purple and orange, respectively. The covalent cystein bridges are in pink spheres.

The AA MDs have predicted multiple interaction sites for the hNavβ subunits on

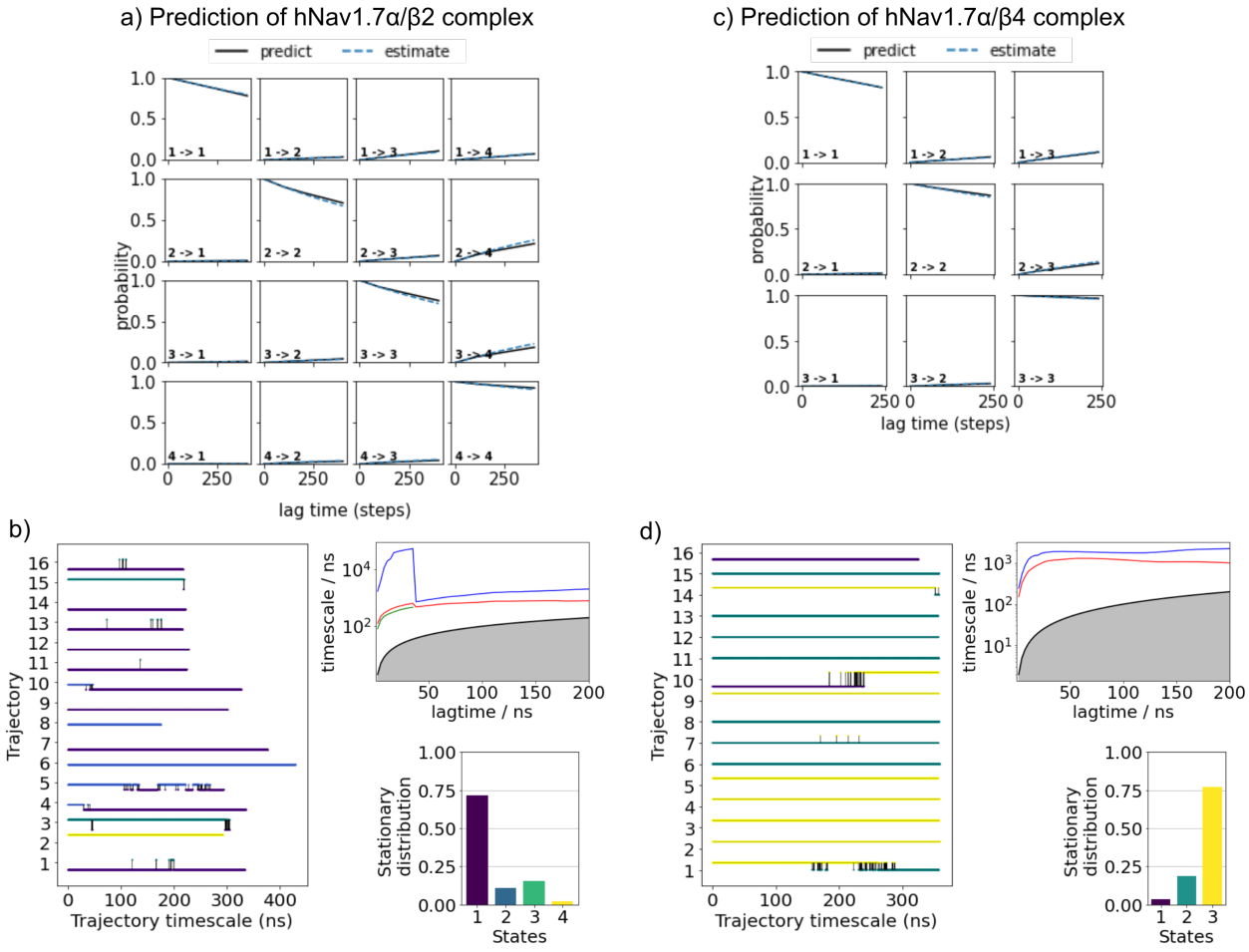


Figure 5.3: Markov modeling analysis on the interaction sites prediction of Ig domain of hNavβ2 (left) and β4 (right) on hNav1.7α. To validate the built Markov models, in a) and c) are represented the CK test. It show how process between the defined number of metastable states are resolved on longer timescale. On the left of b) and d), analysis of the transitions stability between the metastables within each trajectory. Upper right, the ITS presentation. Bottom right, the stationary distribution of each metastable state.

hNav1.7α. The relevant conformations were sampled by performing Markov state modeling, using VAMPnets and Pyemma tools. The same hyper-parameters values were applied on VAMPnets and Pyemma (n number of metastable states and lag time τ). Even though we have defined multiple "optimal" hyper-parameters to build Markov models, in Fig. 5.2 and Fig. 5.3 are only represented the Markov models performed with the hyper-parameters τ = 4 ns for n = 4 and τ = 12 ns for n = 3, respectively for the complex hNav1.7α/β2 and hNav1.7α/β4. This choice is based on the ITS plot, the CK test, the sampled metastable configurations and their transitions stability in the simulated trajectories.

For the chosen hyper-parameters values, VAMPnets and Pyemma built identical Markov models. Hence, only the VAMPnet's analysis are shown in Fig. 5.2 and Fig. 5.3. The markovianity of the defined metastable states in each complexes systems seem to be re-

solved; as evidenced by the CK test. On the left of Fig. 5.3b,d, the given lag time has resolved the transition between the metastable states, which remain stable in a series of timescale. Thus for the ITS plot (5.3b,d, upper right), the implied timescale remain constant for larger lag times.

5.1.1 Analysis of the visited interaction sites

In Fig. 5.2, we surprisingly observed that Ig-domain of hNavβ2 and hNavβ4 subunits visit similar interaction sites on hNav1.7α. Three interaction sites of hNavβ4 were predicted on hNav1.7α, where hNavβ4 interacts at the VSDI (state 1), on the VSDI-VSDIV interface (state 2), and at the VSDIV (state 3). hNavβ2 identifies the same interaction sites as hNavβ4 on hNav1.7α, where hNavβ2 interacts around VSDI and VSDIV. Additionally, hNavβ2 can interact at the VSDII.

The VSDII-hNav1.7α was defined as the main interaction site of Protx-II (Hampl *et al.* manuscript in progress, Xiao et al. 2010). Yet hNavβ4 does not identify a binding site at VSDII domain. Based on Hampl *et al.* experimental investigation, if the hNavβ4 does not exhibit same interaction site as Protx-II at the VSDII-hNav1.7α, how does hNavβ4 affect the pharmacological effect of Protx-II on hNav1.7α?
In the case of insufficient conformational sampling, the generated trajectories of hNav1.7α/β4 were discretized into higher number of metastable states ($n = 6$), Fig. 5.4. This finer sampling has defined the interaction sites of hNavβ4 around the VSDI and VSDIV of hNav1.7α, but yet not at VSDII. The states discretization were performed with Pyemma, which, with the given hyperparameters (τ = 12 ns for n = 6), successfully captured the Markovian behavior among the metastable states. The same hyper-parameters values were applied on MSM model built with Vampnets. This resulted to a failed validation test as the optimum hyper-parameters were not applied (Fig. 5.4i).

5.2 Docking prediction

The interaction sites of hNavβ4 subunit were also predicted with docking methods, in order to support the prediction performed with the molecular dynamics simulation. Similarly to the AA-MDs, the docking methods defined multiple interaction sites for hNavβ4 on hNav1.7α, precisely located at the VSDI, on the VSDI–VSDIV interface, and at the VSDIV (Fig. 5.5). The binding positions of hNavβ4 on hNav1.7α predicted with the docking methods is reproducible with AA-MDs. Accordingly, the prediction of hNav1.7α/β4 complex performed with the docking method supports that hNavβ4 does not establish interaction on VSDII-hNav1.7α.

In the prediction of hNav1.4α/β4 configuration, hNavβ4 has defined VSDII-hNav1.4α as one of the potential interaction sites. In the other hNav isoforms, hNavβ4 is able

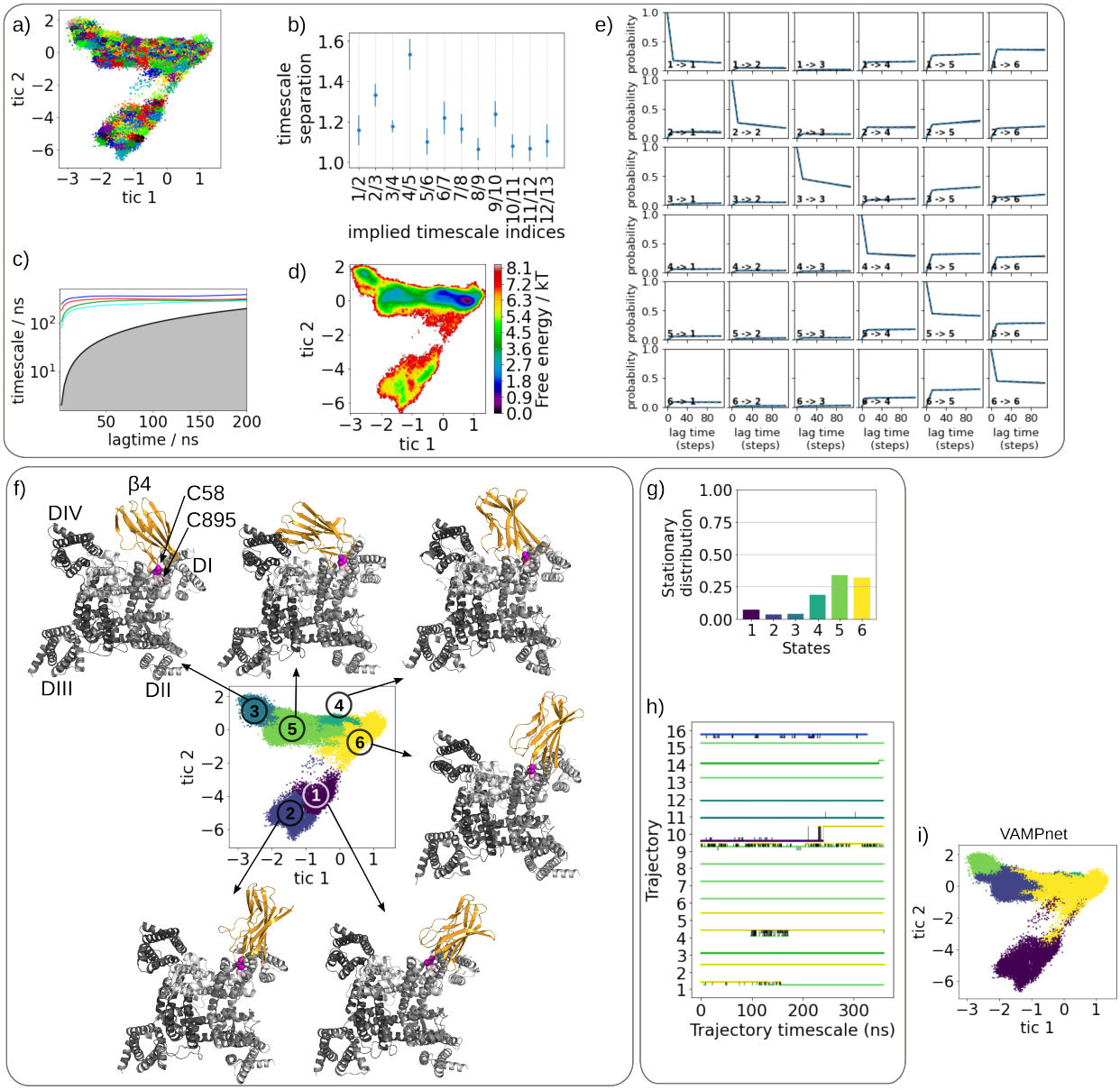


Figure 5.4: Finer metastable states decomposition of the simulated trajectories of hNav1.7α/β4 complex, performed with Pyemma. The trajectories are applied to a dimensionality reduction with TICA for a lag time τ of 12ns, based on the position of the hNavβ4. a) These latest are projected onto the two first ICs, and are firstly discretized into 500 microstates. b) In order to estimate the transition between the microstates, the suitable number of macrostates were predicted according to the time scale separation between the implied timescale. c) The transition between the microstates were evaluated with the ITS representation at multiple (increasing) lag times. At MSM τ=12ns (similar to TICA τ), the implied timescale start to converge, and remain constant on a longer timescale. d) The projection of the free energy transition from the estimation of the MSM show six minima, which can be identified as the metastable states. This correlated with b). e) The built Markov model are then validated with CK test. The process between six metastable states are resolved on longer timescale. f) The metastable states membership are well separated onto the same ICs projection. The centroids of each metastable state are rendered. The hNav1.7α are in grey.

Figure 5.4: The predicted position of hNavβ4 are highlighted in orange. The pink sphere are the covalent cystein bridge between the hNav1.7α and hNavβ4 subunits. g) The stationary distribution, describing the free energy of conformation to stay within the assigned states. h) The metastability transition between the metastables over the trajectories are in the ensemble acceptable. However, the transition between some metastables are too fast and repetitive. These metastables might be defined in one. Reminder, the aim of higher metastables decomposition were to find common interaction site between hNavβ4 and hNavβ2. i) Metastable assignment with VAMPnets.

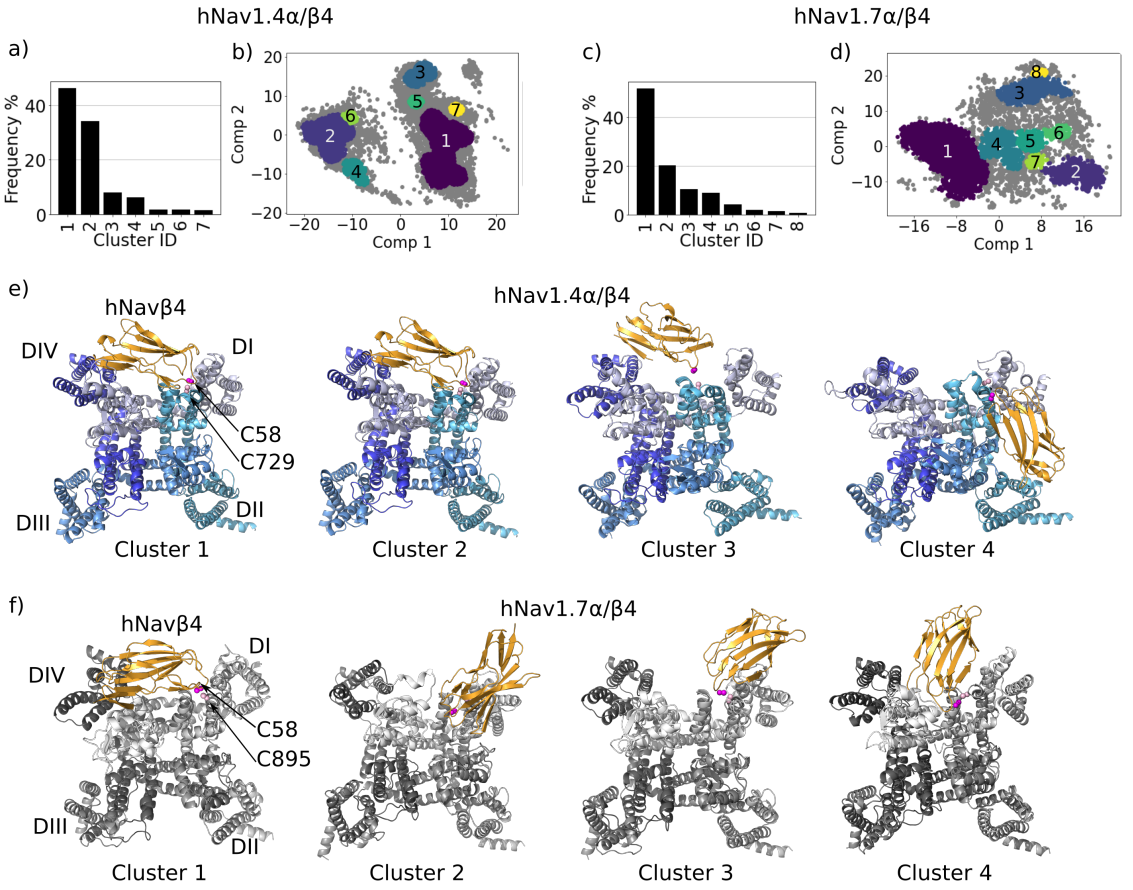


Figure 5.5: Interaction sites prediction of hNavβ4 on hNav1.4α and hNav1.7α using docking methods. a)-d) Structures distribution assigned to each cluster, represented in histogram and projected onto the two first components. e) and f) represent the structures of the first four clusters. e) The different domains in hNav1.4α are highlighted in shades of blue. f) hNav1.7α are in greys. The hNavβ4 are in orange, the cysteins involving in the covalent bonds between the hNavα and hNavβ are in pink spheres.

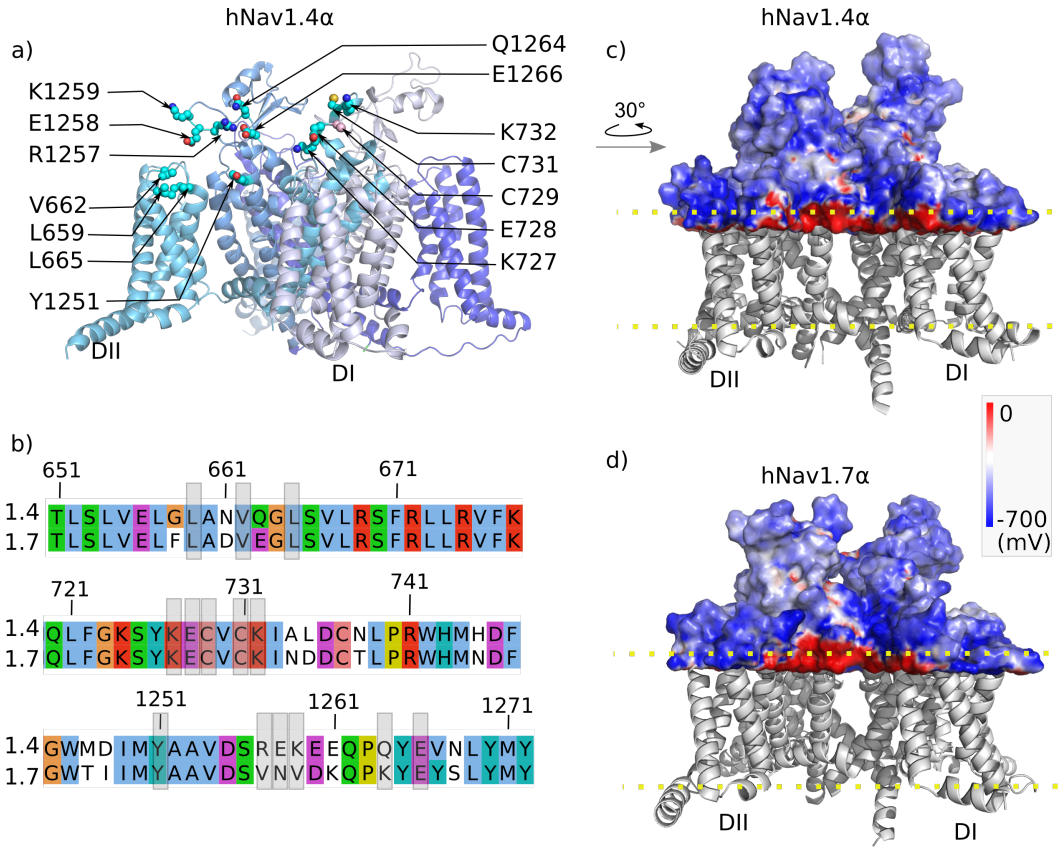


Figure 5.6: VSDII interface comparison between hNav1.4α and hNav1.7α. a) Side view of the hNav1.4α structure, highlighted in different blue shades with respect to the VSDs. Residues involved in the interaction are in spheres. b) Sequence alignment between hNav1.4α (up) and hNav1.7α (bottom), referring to the VSDII interaction interface in hNav1.4α-β4 prediction (cluster 4 in fig 5.5). The sequences were aligned according to Clustal alignment methods. Aligned residues with same chemical properties are highlithed with same color. Residues involved in hNav1.4α-β4 prediction are highlited in gray rectangle. In c) and d) are shown in surface representation, the electrostatic properties at the VSDII interface of hNav1.4α and hNav1.7α, estimated with g_elpot tool (Kostritskii et al. 2021). Region with high electrostatic potential energy illustrate regions with strong positive charged residues. Accordingly, low electrostatic potential energy illustrate the presence of more negative residues. The dashed yellow lines represent the lipid bilayer linitation.

to interact at the VSDII domain. This suggests that hNavβ subunits are VSDs specific according to the hNavα isoforms.

The interface at VSDII of hNav1.4α and hNav1.7α were compared, to evaluate hNavβ4 affinity (Fig. 5.7). Most residues in hNav1.4α implied in the interaction with hNavβ4 are comparable to the interfacial residues at VSDII-hNav1.7α (Fig. 5.7a-b). Residues in hNav1.4α/β4 complex are considered to be involved in the interaction when their respective atoms are within a range of 4 $\AA$. However, the residues neighboring the S3–S4 linker at the VSDII of hNav1.4α differs in hNav1.7α. This implies a divergence in the electrostatic properties at the VSDII interface between hNav1.4α and hNav1.7α (Fig. 5.7c-d). In hNav1.7α, the extracellular structures close to the lipid membrane are scattered with more stronger negative residues. This might reducing the electrostatic matching between hNavβ4 and hNav1.7α at VSDI–VSDII interface.

5.3 Overview of hNavβ4 and Protx-II interaction sites

Hampl and collaborators investigation suggested that Protx-II and the Ig-domain of hNavβ4 might have competitive interaction sites. Literatures have defined the VSDII domain in hNav1.7α isoform as the main interaction site of Protx-II (Xiao et al. 2010, Hampl *et al.* manuscript in progress). Experimental investigations have demonstrated that mutation of VSDI-hNav1.7α does not significantly influence the effect of Protx-II (Xiao et al. 2010). And high concentration of Protx-II is needed in order to interact at the VSDIV and

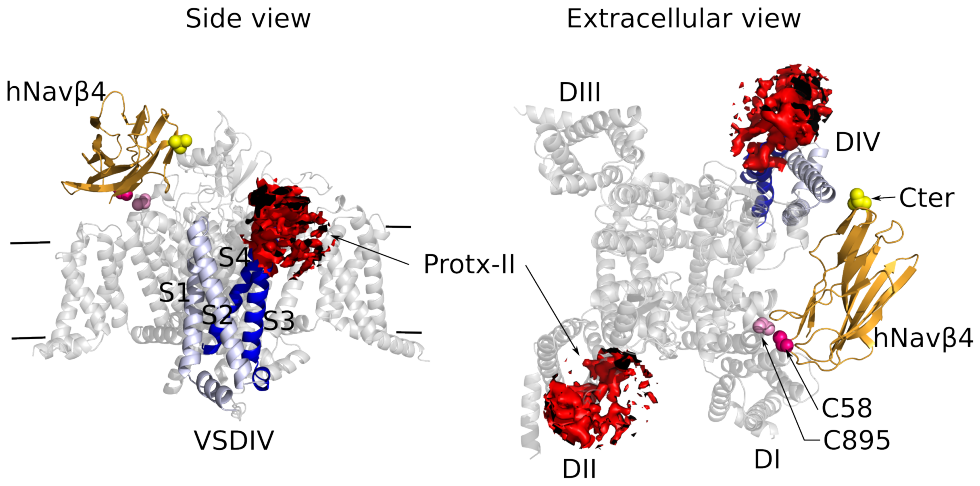


Figure 5.7: Side view and extracellular view of the hNav1.7α structure, the predicted position of hNavβ4 (in orange) and the cryo-EM density structures of Protx-II (PDB ID: 6J8I)

inducing inhibition of inactivation.

On hNav1.7α, the different prediction methods have defined only the regions around

VSDI and VSDIV as potential interaction sites of Ig-domain hNavβ4. hNavβ4 subunit was not predicted to interact at VSDII-hNav1.7α. The predicted binding site of hNavβ4 at VSDIV-hNav1.7α is located on the S1–S2 linker (Fig. 5.7). The interaction of hNavβ4 at VSDIV-hNav1.7α does not compete directly with the binding site of Protx-II, which is well known located on the S3–S4 linker. Moreover hNavβ4 does not create a steric clashing when Protx-II is bound.
If hNavβ4 and Protx-II define distinct interaction sites, how does the pharmacological effect of Protx-II in hNav1.7α is modulated in the presence of hNavβ4 (Hampl *et al.*, manuscript in progress).

5.4 Modulation of hNav1.7α induced by hNavβ4

When the full sequence of hNavβ4 including the Ig-domain, TM and Cter-hNavβ4 were predicted in complex with hNav1.7α, the AlphaFold2 prediction has highlighted interesting configurations of the C-terminal of hNavβ4 (Cter-hNavβ4) and the IFM motif of hNav1.7α. These configurations might enlighten hNavβ4 mechanical function as an open-pore blocker and thus, in Navs resurgent current (Lampert, Eberhardt, and Waxman 2014, Wang, Edrich, and Wang 2006). Additionally, these configurations might suggest an allosteric modulation of Protx-II effect on hNav1.7α, as the hypothesis on competitive interaction sites between hNavβ4 and Protx-II was rejected.

The mechanical function of Cter-hNavβ4 as an open-pore blocker emanates from the rapid recovery in high firing frequency of Purkinje neurons, where these latter require a short refractory periods (Grieco et al. 2005). Upon the depolarization, the Nav pore opens, which is suggested to give Cter-hNavβ4 access to its receptor and subsequently block the pore. Hence, the binding of Cter-hNavβ4 prevents the fast inactivation mechanism which accelerate the recovery (Grieco et al. 2005). Additionally the binding of Cter-hNavβ4 blocking the Na^+ current seems to alternate the fast inactivation in inactivation-deficient Navs (Wang, Edrich, and Wang 2006). Thus, researchers have proposed that the pore closure of Navs can be mediated by a competitive mechanism between the fast inactivation particles (IFM motifs) and Cter-hNavβ4 (Grieco et al. 2005; Wang, Edrich, and Wang 2006).
The interaction of Cter-hNavβ4 is a reversible mechanism. Upon repolarization, the electrostatic environment at the Navs intracellular side promotes the unbinding of Cter-hNavβ4. The dissociation of Cter-hNavβ4 releases the open-pore Nav, which generates the resurgent current.

5.4.1 The IFM motif is modeled unbound

Interestingly, the models predicted with Alphafold characterized competitive configurations between Cter-hNavβ4 and the IFM-hNav1.7α. When the Cter-hNavβ4 is bound to

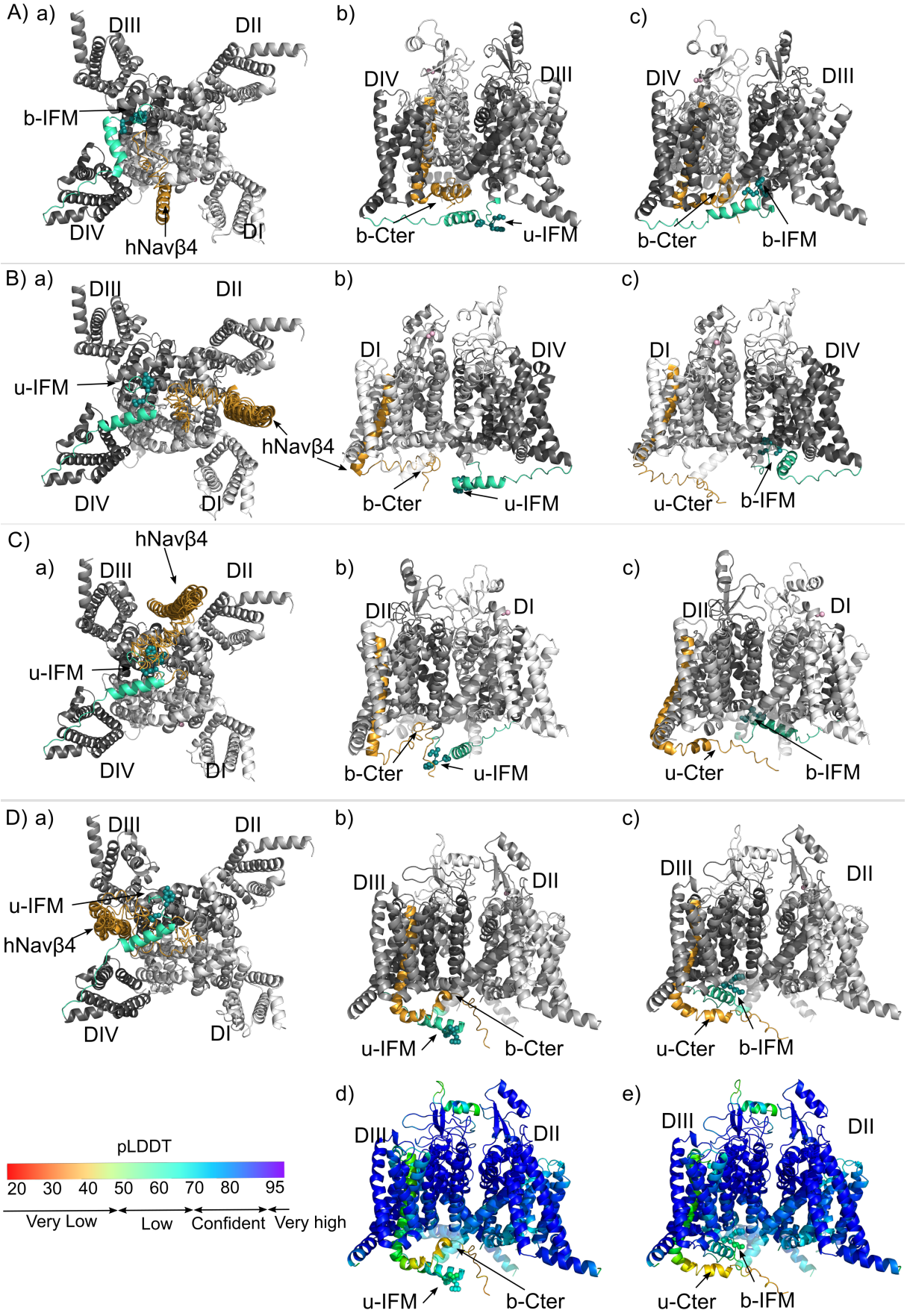


Figure 5.8: Complex hNav1.7α/β4 prediction with Alphafold.

Figure 5.8: The prediction of hNav1.7α/β4 complex is performed using the sequence of hNav1.7α truncated as the resolved cryo-EM structure PDB ID 6J8I and the full sequence of hNavβ4. The structures models of hNav1.7α are represented in cartoon with in different shades of gray color according to the domains. The DIII–DIV linker are shown in cyan, where the IFM motif are represented in spheres. For the structural representation of hNavβ4, the Ig domains were intentionally removed, due to the very low plDDT confidence and the obvious non-native configurations. Only the transmembrane segment (TM) and the C-terminal (Cter) of hNavβ4 subunit are shown here. In the panels A) to D) are represented the predicted configurations of the complex hNav1.7α/β4, when hNavβ4 interacts to the interfaces DIV–DI, DI–DII, DII–DIII and DIII–DIV, respectively. On the left of each panel are the intracellular view of the hNav1.7α/β4 structures, in order to locate the interaction site of hNavβ4. In the middle and on the right of each panel are the side view representation of the configuration complex, in order to view the IFM motif in unbound (u-IFM) and in bound (b-IFM) states, respectively. Only the configurations in panel A) has demonstrated a non-competitive mechanism between the Cter-hNavβ4 and the IFM-hNav1.7α. The bound Cter-hNavβ4 (b-Cter) at the DIV–DI interface of hNav1.7α can result to b) unbound and c) bound IFM motifs. In panels B) to D), we observed competitive interactions between Cter-hNavβ4 and the IFM-hNav1.7α, b-Cter/u-IFM and u-Cter/b-IFM. D) d-e) The confidence at the IFM motif region are similar in the bound and the unbound configurations.

hNav1.7α, the IFM motif is unbound from its receptor. Accordingly, the bound state of IFM-hNav1.7α is coupled with the unbound state Cter-hNavβ4. These competitive configurations between the IFM motif and Cter-hNavβ4 are observed when hNavβ4 interact at the interfacial region DI–DII, DII–DIII and DIII–DIV of hNav1.7α (Fig. 5.8B-D).

However, when hNavβ4 interacts at the DIV–DI interface, we observed a non-competitive interaction between the Cter-hNavβ4 and IFM-hNav1.7α (Fig. 5.8A). The bound Cter-hNavβ4 can simultaneously induce the bound and unbound states of the IFM-hNav1.7α.

As might arise from the competitive configurations, surprisingly the unbinding state of IFM-hNav1.7α is not mediated by Cter-hNavβ4. Different mutation(s) were inserted at Cter-hNavβ4 and the prediction complex hNav1.7α/β4 remain able to modeled an unbound IFM-hNav1.7α. This is supported by the truncation of Cter-hNavβ4, excluding its function in unbinding IFM-hNav1.7α. Nonetheless, the configuration of bound IFM motif in the complex prediction hNav1.7α/β4 is only modeled when the C-terminal of hNavβ4 is included. We also investigate the implication of the Ig-domain of hNavβ4 on the conformational change of IFM-hNav1.7α. Together, these data conclude that Ig-domain of hNavβ4 do not induce (allosterically) conformation change in IFM-hNav1.7α.

5.4.2 The Phe189 in hNavβ4 plays an important role in hNav1.7α modulation

The results in the previous section bring our attention to the transmembrane segment (TM) of hNavβ4. Experimental studies (Grieco et al. 2005) have investigated the se-

quence of TM-hNavβ4, corresponding to the residues range 184–197. This fragment of TM-hNavβ4 were stated to induce conformational change in hNavs, involved in resurgent current (Grieco et al. 2005). The functional specificity in the resurgent current and the binding stability of $TM_{184-197}$-hNavβ4 were demonstrated to be involved by the composition of multiple positively charged residues of Lysine, and the aromatic residue Phenylalanine.
In order to establish a correlation between these residues and the unbinding of IFM-hNav1.7α, a single or multiple mutation(s) were inserted at $TM_{184-197}$-hNavβ4. The mutation of Lysine into Alanine in $TM_{184-197}$-hNavβ4 still induces an unbound configuration of IFM-hNav1.7α. However, the mutation of Phe189Ala abolishes the modelisation of IFM-hNav1.7α in unbound state. The Phenylalanine has strong hydrophobic potential, however as observed in Fig. 5.9, Phe189 is modeled in the aqueous environment. Two configurations of TM-hNavβ4 hNav1.7α in complex with were observed, where Phe189-hNavβ4 faced toward the sodium channel or Phe189-hNavβ4 is oriented to the lipid bilayer. When Phe189-hNavβ4 faced hNav1.7α, Phe189 does not establish any short/strong interaction residues. The distance of the closest residues between Phe189-hNavβ4 and hNav1.7α is approximately at 7 $\AA$. In Fig. 5.9c is shown all residues around Phe189-hNavβ4 at a distance of 10 $\AA$.

5.4.3 The conformational change at the IFM receptor

In hNav, the fast inactivation mechanism is preceded by the activation of VSDIV in unpward configuration, which give the access of IFM motif to its receptor. The IFM motif is lodged between the S5 and S6 of the DIII and DIV domain. Conformational changes at the IFM receptor are observed when hNav1.7α is predicted with hNavβ4 (Fig. 5.9d). The S5-DIII, S6-DIII and S5-DV are moved inward, restricting the accessibility of IFM to interact to its receptor. However, when TM-hNavβ4 is mutated at the position Phe189 into Alanine, the IFM motif receptor is predicted in bound state, similar to the experimental structure PDB ID 6J8I. This excludes the implication of the Cter-hNavβ4 in the unbound state of IFM motif.
Even though Phe189 does not establish a tight interaction with hNav1.7α,the hydrophobicity of Phe189-hNavβ4 might induce a back and forth motion into a hydrophobic environment, as the lipid membrane. This suggests a similar mechanism as the IFM motif – its stability in its receptor during the fast inactivation relies on the Phenylalanine residue, and the interaction of IFM motif to its reception can also be reversible (Grieco et al. 2005). Thus, when Phe189-hNavβ4 is pushed toward the lipid membrane, it induces conformational changes at the neighboring segments of hNav1.7α. This is followed by allosterical modification of the IFM-receptor and releasing the IFM motif.

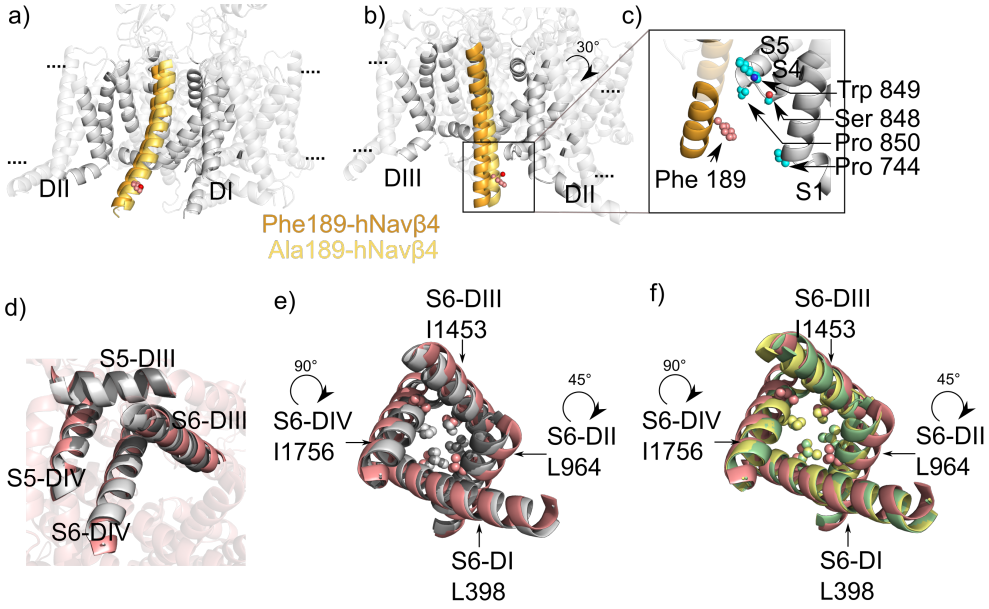


Figure 5.9: a, b) Highlight on the two predicted configurations of TM-hNavβ4. In gray are represented hNav1.7α structure. For visualization clarity purpose, structures not around the hNavβ4 are in transparent. The structures of wild-type TM-hNavβ4 are shown in orange, where Phe189 are in pink spheres. The mutated TM-hNavβ4 are shown in yellow, where Phe189Ala are in red spheres. In a), Phe189-hNavβ4 is turned facing the lipid bilayer. In b), the Phe189-hNavβ4 is oriented toward hNav1.7α. c) A closest view on the Phe189-hNavβ4 is shown, where closest residues at 10 Å in hNav1.7α are highlighted in cyan spheres. d, e, f) are highlithed conformational changes predicted at the intracellular pore and the IFM receptor of hNavA.7α models. The predicted models were compared with the experimental structure PDB ID 6J8I, represented in pink. d) Zoomed view on the IFM receptor at the intracellular view. In grays are predicted models of hNav1.7α/β4 complex. In lighter gray, Phe189-hNavβ4 was kept in wild-type. In darker gray, Phe189-hNavβ4 was mutated in Alanine. e) Close view at the intracellular pore, formed by the S6 of DI, DII, DIII and DIV. For the pore analysis, predicted models and the experimental structure were structurally aligned on the pore. The structure in gray represented the hNav1.7α models predicted with mutated Phe189Ala-hNavβ4. The reference structure is in pink. The residues forming the hydrophobic ring are represented in spheres. f) Close view at the intracellular pore of hNav1.7α/β4 complex, where in green and yellow the subunit hNavβ4 were kept in WT. The IFM-hNav1.7α were predicted in bound and unbound states, respectively.

5.4.4 Molecular insight of hNavβ4 as an open-pore blocker

The function of the IFM motif in fast inactivation mechanism has been well understood since the first eukaryotic sodium channel was resolved (Shen et al. 2017). The experimental structure displays an allosteric occlusion of the pore, induced by the IFM motif which exerts a conformational change of S6-DIV towards the ion permeation (Rühlmann et al. 2020). Consequently, one might infer that unbinding of IFM motif releases the pore open.

Unexpectedly, the predicted configurations with unbound IFM also display the channel's pore in close conformation (Fig. 5.9e-f). This suggests that hNavβ4 maintains the chan-

nel's pore in closed conformation, and only the presence of the TM domain of hNavβ4 is required. The channel's pore in the predicted configurations of hNav1.7α/β4 were compared with the experimental structure PDB ID 6J8I (Shen et al. 2019) as reference. Surprisingly, predicted models with unbound IFM have narrower pore diameter than the reference structure (Fig. 5.9e-f). This pore diameter corresponds to the positions of the hydrophobic residues ring at the intracellular side of the channel, formed by L398, L964, I1453 and I1756, respectively located at the segment S6 of DI, DII, DIII and DIV. In the experimental structure, the distance between the Cα L398–I1453 and L964–I1756 measure 12.8 $\mathring{A}$ and 14 $\mathring{A}$, respectively. In the predicted structures, L398–I1453 and L964–I1756 correspond to 12 $\mathring{A}$ and 11 $\mathring{A}$, respectively.
As illustrated in Fig. 5.9e-f, the predicted configuration shows larger conformational change at S6-DII and S6-DIV – where L964 and I1756 are rotated on 45° and 90° towards the pore which further narrows the pore diameter. In the experimental structure of hNav1.7α, intracellular pore is larger but the Na^+ current remains disrupted.

5.4.5 Conformational changes at the VSDs

As a brief reminder, the states of Navs are determined by coordination of the four VSDs (Armstrong and Bezanilla 1974, Armstrong and Bezanilla 1977). With respect to the membrane potential, the motion of the VSDs are managed by three to four repeated positively charged residues (R1 to R4) at the S4 segment, namely gating charge (GC) residues. These residues translocate across the hydrophobic constriction site (HCS). The HCS is a conserved phenylalanine or tyrosine residue. In the resolved structure of hNav1.7α (PDB ID: 6J8I), all four VSDs have structural feature of activated conformations. The three first positive residues of the gating charges are located above the HCS (Fig. 5.10). In VSDI, VSDII and VSDIV, the R1 and/or R2 are interacting with a negative charge cluster at the extracellular side, namely ENC (Payandeh et al. 2011). Except in VSDIII, where the R2 and R3 establish an interaction with the ENC. While, the R4 interacts with a negative charge cluster at the intracellular side, namely INC.

When the VSDI to VSDIII are in upward conformation, the pore opens. This is followed by the activation of the VSDIV that orchestrates the fast inactivation mechanism, characterized by the biding of IFM-hNav1.7α. The binding of the IFM motif closes the pore allosterically. In Fig. 5.10 we show the conformational changes induced in the VSDs when hNav1.7α is predicted in the presence of hNavβ4. In the predicted complex hNav1.7α/β4 with competitive configuration between the IFM and Cter-hNavβ4. For the models with bound IFM-hNav1.7α and unbound Cter-hNavβ4 (in green, Fig. 5.10a), one click downward translation on all four VSDs are observed. Where the R3 is translocated below the HCS compared to the experimental structure. This conformation is reversible when Cter-hNavβ4 is bound (in yellow, Fig. 5.10a). Where only at VSDI, VSDII and VSDIII, the R3 side chain is predicted to be located above the HCS. Whereas the R3 in VSDIV remains

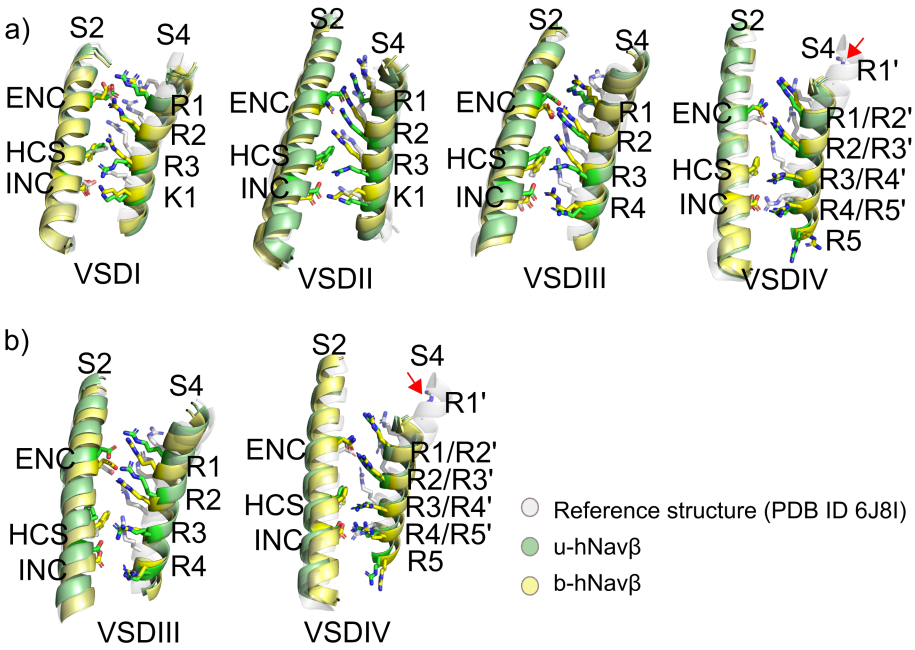


Figure 5.10: Conformational changes in the S2 and the gating charge S4 of the VSDs domain of hNav1.7α. In white are represented the experimental structure PDB ID 6J8I. In green are represented the predict structure with unbound Cter-hNavβ4 (u-hNavβ4) (in the competitive configuration, this is coupled with the bound IFM (b-IFM)). In yellow are the predicted structure with bound Cter-hNavβ4 (b-hNavβ4) (in the competitive configuration, this is coupled with the unbound IFM (u-IFM)). In sticks are highlighted the gating residues at S4, the ENC, HCS and INC at S2. a) shows the configuration of all four VSDs predicted in the competitive configuration. b) Conformational change in the non-competitive configuration are only observed in VSDIII and VSDIV, which are represented. At the VSDIV, the gating residues of the reference structure are highlighted (R') to better distinguish the conformational changes in the predicted models.

below the HCS. This latest structural feature corresponds to the mechanical conformation of VSDs when hNavs is in open state. However, the pore remains closed and narrower, even if the prediction of VSDs define the Navs open state feature (Fig. 5.10e-f)

In the non-competitive mechanism (Fig. 5.10b), hNav1.7α is also subjected to conformational changes. In hNav1.7α/hNavβ4 configuration with unbound Cter-hNavβ4, all gating charges underwent downward motions, where the R3s are located at the same height of the HCS. When the Cter-hNavβ4 is in bound state, no particular difference are observed at the VSDs compared to the configuration with bound Cter-hNavβ4.

Coming back to the fact that the presence of hNavβ4 reduces the pharmacological effect of Protx-II on hNav1.7α. The effect of Protx-II is not modulated due to competitive interaction with hNavβ4. However, hNavβ4 might play an important role in the promotion of the VSDs activation when the membrane potential is depolarized (as demonstrated in the configuration with competitive mechanism between Cter-hNavβ4 and IFM-hNav1.7α).

5.4.6 hNavβ4 and hNavβ2 have distinct function

The structural modulations of hNav1.7α mediated by hNavβ4 infer the function of hNavβ4 as an open-pore blocker and pharmacological modulator. The following investigation suggests that these functions might be proper to hNavβ4 isoform. For the prediction of hNav1.7α/β2 complex, the Ig-domain, the TM segment and C-terminal loop of hNavβ2 were also included. In constrast to hNavβ4, the presence of hNavβ2 did not predict an unbound conformation of IFM-hNav1.7α. Nonetheless, Cter-hNavβ2 is predicted binding to hNav1.7α. Firstly, TM-hNavβ2 lacks of a Phenylalanine at the region corresponding to the position Phe189 in TM-hNavβ4 (Fig. 5.11). Then, the sequence of Cter-hNavβ2 loop is formed by 30 residues which is shorter than Cter-hNavβ4 with 33 residues. Also, the electrostatic properties of hNav1.7α at the intracellular side are mainly composed with negative residues. However, Cter-hNavβ2 is contain more negative residues compared to Cter-hNavβ4. Thus, the electrostactic properties of Cter-hNavβ4 might be more favorable to interact at the intreacellular structure of hNav1.7α. Even though the Cter-hNavβ4 is not directly involved in the conformational change of hNav1.7α, the dynamics of Phe189-hNavβ4 might be enhanced by Cter-hNavβ4. For the prediction of hNav1.4α/β4, the Cter-hNavβ4 is rarely predicted to interact at hNav1.4α isoform and accordingly no structures with the IFM unbound were predicted. The intracellular side of hNav1.4α is more negatively charged (Fig. 5.11a), decreasing the electrostatic complementarity of Cter-hNavβ4 to interact to hNav1.4α. This supports the importance of Cter-hNavβ4 to enhance the dynamics of Phe189 located at the TM-hNavβ4. Moreover, Yu et al. 2003 demonstrated that hNavβ4 have greater shift in the voltage dependence of activation for hNav1.7α than in hNav1.4α. Additionally, this might support the statement in the section 5.2, that function of hNavβ4 might be Navs isoforms specific.

5.5 Resume: investigation of hNavβ4 interaction sites

In this chapter, we would like to characterize the interaction sites of hNavβ4 on hNav1.7α. Experimental investigation suggests that hNavβ4 could act as a pharmacological modulator, as demonstrated when co-expressed with hNav1.7α and the toxin Protx-II (Hampl *et al.*, manuscript in progress). Protx-II has demonstrated significant pharmacological potential on hNav1.7α. In the first years of this project, we only modeled the Ig-domain of hNavβ4 due to the availability of experimental structure. The interaction sites of Ig-hNavβ4 were predicted using molecular dynamics simulations and protein-protein docking. In parallel, we also performed the prediction of Ig-hNavβ2 interaction sites, as control set. Surprisingly, we observed that the Ig-domain of hNavβ2 and hNavβ4 visit the same interaction sites on hNav1.7α. The predicted interaction sites of hNavβ4 do not overlap or interfere with the known binding sites of Proxt-II. Despite this, hNavβ2 has not been reported to modulate the function of Protx-II.

Thanks to recent release of AlphaFold2, we are now able to predict the full complex struc-

ture of hNav1.7α and hNavβ4. The structure of hNavβ4 includes the Ig-domain, the TM segment and the C-terminal tail. In the presence of hNavβ4 and its C-terminal loop modeled bound to hNav1.7α, we observed conformational changes at the IFM motifs, the VSDs and the channel's pore. Interestingly, these conformational changes of the IFM motifs persisted even when excluding Cter-hNavβ4 from the prediction. An investigation led by Grieco and collaborators drives our attention on a Phenilalanine at the position 189 of TM-hNavβ4. By mutating only one single residue, Phe189-hNavβ4, conformational modulation at the IFM is prevented. The predictions of hNav1.7α/β2 complex did not report conformational changes as demonstrated for hNav1.7α/β4 complex . Notably, hNavβ2 lacks an equivalent residue to Phe189-hNavβ4; the C-terminal tail of hNavβ2 is shorter than hNavβ4, and it has more positively charged residues. Thus, this investigation revealed a proper function of hNavβ4 as pharmacological modulator.

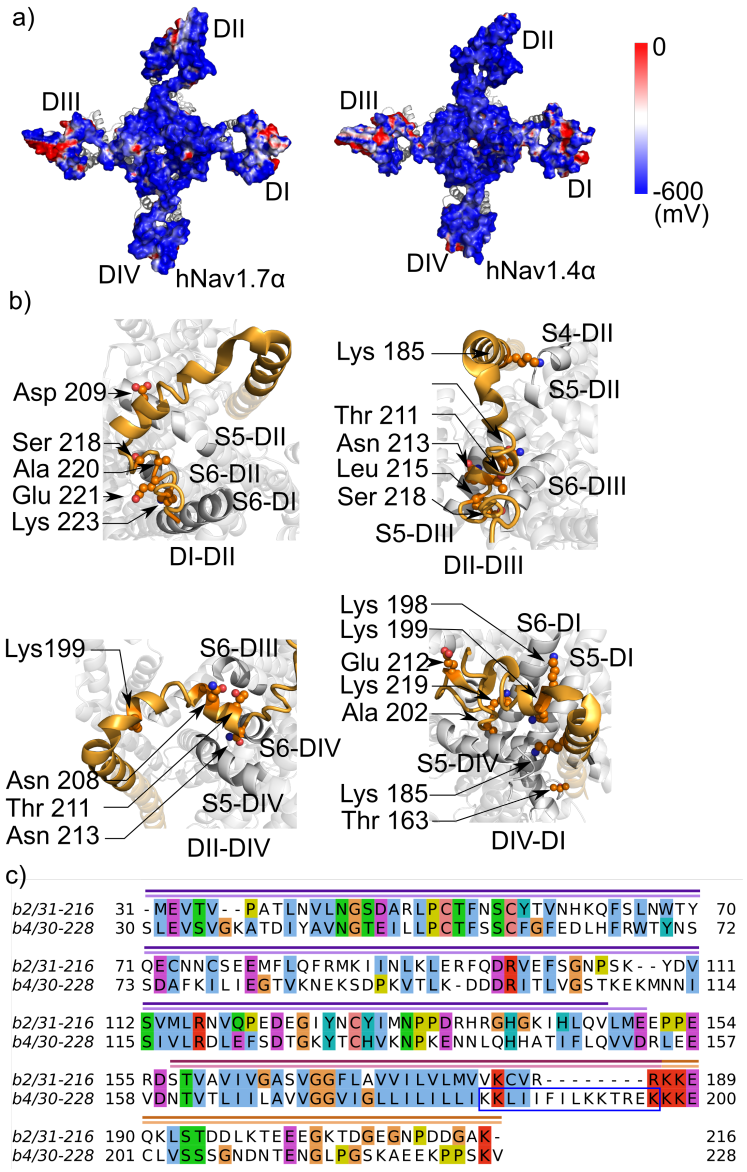


Figure 5.11: a) Electrostatic potential of hNav1.7α and hNav1.4α at the intracellular side calculated with g_elpot. b) In the prediction of hNav1.7α/β4 complex with AF2, hNavβ4 were predicted to interact at the interface DI–DII, DII–DIII, DIII–DIV and DIV–DI, viewed at the intracellular side. hNav1.7α is in gray. hNavβ4 in orange, where the interacting residues at 3 $\AA$ of hNav1.7α are highlight in spheres. c) The sequence of hNavβ2 and hNavβ4 were aligned with clustal. Residues were colored according to their electrostatic properties. The continuous lines above the residues alignment referred to Ig-, TM- and Cter-domain of the hNavβ subunits, respectively in purple, pink and orange. The sequence referring to $TM_{184-197}$-hNavβ4 investigated by Grieco team are in bleu rectangle.

Chapter 6

Conclusions and perspectives

The investigation on human voltage-gated sodium channel (hNav) is very challenging, due to its very large size and the complexity of its mechanical function. Thus, a lot remains to be unveiled about hNav mechanism and its modulation. The isoform hNav1.7α was firstly linked to pain propagation in 2004 (Yang et al. 2004). Nearly 15 years later has been resolved its experimental structures (Shen et al. 2019). In order to gain molecular insights on hNav1.7α modulation induced in complex structures, a combination of computational molecular methods, including molecular dynamics simulation, protein-protein docking and AlphaFold2, were performed. The predictions were unbiased, i.e no restraining forces were used to guide protein-protein interactions. Multiple inferences were executed to generate a diverse set of dimer configurations. Some configurations are more prevalent than others, providing a glimpse into the complex landscape of Nav1.7α in complexes structures.

Coupled-gating mechanisms of hNav1.7α are suggested to promote cellular hyperexcitability. To describe this unexplored mechanism, the dimer conformations of Nav1.7α were predicted with the above mentioned techniques. The dimerization of hNav1.7α was simulated using CG molecular dynamics, leading to the identification of 18 unique configurations. Of the 18 configurations determined from CG simulations, 12 configurations were also identified with molecular docking – where 35 dimer configurations were defined. The AlphaFold2 has predicted five distinct models, mainly involving symmetric interaction, i.e both protomers involve same interface in the interaction. All models generated with AlphaFold2 were also predicted with CG molecular dynamics and/or molecular docking methods. The convergence of these methods points towards hNav1.7α adopting multiple dimer conformations, where all interfacial regions of hNav1.7α are involved in the dimer interaction. However, preferences at certain interfaces were observed in the dimerization.
The analysis of physico-chemical properties at the interfacial regions of hNav1.7α reveal

driving features of hNav1.7α homo-dimerization. Firstly, the tertiary symmetrical shape of hNav1.7α structure can confer a complementarity-based interaction at all interfaces during the dimerization. Furthermore, the chemical properties at hNav1.7α interfacial regions are predominantly hydrophobic, which is expected for membrane protein. The interfacial region of DI–DII in hNav1.7α displays a higher propensity to establish favorable dimer interaction. The DI–DII interface has indeed more hydrophobic and non-charged residues, and is scattered with fewer charged residues compared to DII–DIII, DIII–DIV and DIV–DI. This hydrophobicity at DI–DII increases its probability to be engaged in the dimer interaction.

We suggested that when hNav1.7α forms a dimer at the interface involving DI, DII and DIII, the activation of VSDs is promoted. The synchronization of VSDI, VSDII and VSDIII activation leads to the opening of the channel's pore. The most frequently visited dimer configurations involve the interfaces DI–DII and DIII–DIV, for each respective protomer in the dimer conformation. When the dimerization involved the interfacial region of DIII–DIV, we suggest a direct modulation of the pore into an open state. The dimerization of hNav1.7α at DIII–DIV impede the fast inactivation mechanism by preventing the IFM motif interactions. Thereby, the pores are maintained continuously open. Moreover, the VSDI establish an interaction with VSDIV, where the height of the VSDI across the lipid membrane is more shorter than the VSDIV. We suggested a preferred configuration in the interaction of VSDI and VSDIV, where the VSDI is in an activated configuration and the VSDIV is trapped or modulated in the downward conformation. These configurations of the VSDI and VSDIV are proposed to promote the pore activation in coupled hNav1.7α. Thus, in patients with IEM, dimerization of hNav1.7α contributes the cellular hyperexcitability. The uncoupling of the IEM-hNav1.7α dimer has demonstrated, experimentally, a significant reduction of Na^+ conductance (Rühlmann et al. 2020).

The mechanism of hNav1.7α can also be influence by various factors, such as its interaction with hNavβ subunits, small molecules, and peptide toxins. The co-expression of hNav1.7α with the tarantula toxin Protx-II can inhibit both activation and inactivation mechanisms of hNavs. This inhibition is partially reversible when hNavβ4 is present,experimental investigation has demonstrated a partial recovery of Na^+ current in hNav1.7 (Hampl *et al.*, manuscript in progress). This suggests a potential competitive interaction between Protx-II and hNavβ4 at binding sites on hNav1.7α. Using all-atom molecular dynamics simulation and docking methods, multiple configurations of hNav1.7α/β4 complex were predicted. The models of hNavβ4 were constructed based on existing experimental structures, with the structural modeling constrained to the Ig-domain. Three interaction sites have been defined, where the hNavβ4 Ig-domain interact around the region of VSDI and VSDIV. However, the main interaction site of Protx-II on hNav1.7α has been experimentally defined at VSDII (Xiao et al. 2010, Hampl *et al.* manuscript in progress). The predicted interaction sites of hNavβ4 on hNav1.7α do

not demonstrate overlapping, competitive, or clashing interactions with Protx-II binding. This prompts the question of how hNavβ4 induces pharmacological modulation of Protx-II's effect.

The predictions with AlphaFold2 could clear up this question. With AlphaFold2, the prediction of hNavβ4 in a full complex where the Ig-domain, the TM and the C-terminal loop were included, further functions insights about hNavβ4 on hNav1.7α were provided. Remarkably, the predicted hNav1.7α/β4 complex highlight a competitive binding configuration between the Cter-hNavβ4 and IFM-hNav1.7α. When the Cter-hNavβ4 is bound, the IFM motif is unbound. Accordingly, the interaction of the IFM motifs to its receptor is associated to an unbound Cter-hNavβ4. Surprisingly, this unbound state of the IFM is not induced by Cter-hNavβ4, but rather by a specific residue Phe189 which is located at TM-hNavβ4. When the IFM motifs were modeled unbound, the hNav1.7α pores remain closed. These structural features might provide an insight to the role of hNavβ4 as an open-pore blocker. Additionally, the hNavβ4 can play an important role in sodium channels with impaired fast inactivation (Grieco et al. 2005).

The unbound state of the IFM motif are not observed in the prediction complexes hNav1.7α/β2 and hNav1.4α/β4, despite having visited similar interaction sites as hNav1.7α/β4. This might suggest that hNavβ2 and hNavβ4 have distinct function on hNav1.7α. Also, the hNavβ4 appears to be Navs specific, for instance, no conformational change in hNav1.4α isoform was observed in the presence of hNavβ4.
Conformational changes are also observed in the VSDs of the predicted complex hNav1.7α/β4; models predicted with the bound Cter-hNavβ4 are associated with an upward one-click translation of the gating charges residues on DI, DII and DIII of hNav1.7α. This suggest that hNavβ4 promotes the activation of the VSDs. This might explain the pharmacological modulation of the complex hNav1.7α/Protx-II mediated by the Phe189-hNavβ4 and the C-terminal tail.

Thus, we have characterized two interaction mechanisms of hNav1.7α. Within the densely populated cell membrane, certain proteins establish highly specific interactions with a single neighboring protein, but often, proteins interactions involve more complex networks. These interactions aim to regulate the functions of the proteins, and thus, maintaining the cellular physiology. Dysfunction of several proteins were associated to the development and the causes of many diseases. In this study, we want to better understand the pain syndromes, that have being reported and associated with dysfunctions of the membrane protein hNav1.7α. Knowledge on hNav1.7α complexes is still lacking in order to better understand the modulation of its function. To address this gap, we have characterized the first molecular insights of hNav1.7α in homo-dimer, offering insights into the coupled-gating mechanism. Additionally, we predict complex structures of hNav1.7α/β4, elucidating the roles and mechanisms of the auxiliary hNavβ4 subunits.

In this project, we have conducted computational investigations to enhance our comprehension of pain-related disease, by unveiling complexes interaction of hNav1.7α that potentially modulate its function. The insights gained from the predicted complexes of hNav1.7α at molecular scales can provided foundation for further investigations, including experimental mutagenesis. For instance, mutation can be introduced at key regions of the interaction interface, or within residues implicated in the conformation changes. Moreover, knowledge derived from these studies can also guide the development of novel therapeutic strategies. The integration of experimental research can also play a crucial roles for validating the researcher's interpretation resulting from the computational prediction, which is facing some limitation. Indeed, the use of computational techniques to predict protein's interaction might represent a simplified environment, that do not take into consideration network interactions essential for protein's function. Additionally, the setup of the prediction tools and the resolution (or condition) of the target structures might lead to biased results. However, experimental investigations are indispensable and complementary to computational investigation, to provide comprehensive understanding of the protein's mechanism.

Appendix A

Summary of predicted hNav1.7α complexes

Table A.1: Summary of simulated systems

Molecular dynamics simulation (MD)	System	Number of MD	Average simulation length	Total length of MD
All-atom	hNav1.7α/β1/β2	130	500 ns	51 μs
	hNav1.7α/β1/β4	168	500 ns	66 μs
Coarse-grained of hNav1.7α dimerization. Membrane composition with :	pure POPC	74	17 μs	1250 μs
	mixed lipids	16	19 μs	310 μs

Table A.2: Summary of protein-protein docking prediction

System	Number of performed prediction	Number of generated structures after filtering
hNav1.4α/β4	16	13,300
hNav1.7α/β4	26	9,000
hNav1.7α/hNav1.7α	107	36,500

Bibliography

Abraham, M. J. et al. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

Ahern, C. A. et al. (2015). The hitchhiker's guide to the voltage-gated sodium channel galaxy. *Journal of General Physiology* **2015**, *147*.1, 1–24.

Ahuja, S. et al. (2015). Structural basis of Nav1.7 inhibition by an isoform-selective small-molecule antagonist. *Science* **2015**, *350*.6267.

Alder, B. J. and T. E. Wainwright (1957). Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics* **1957**, *27*.5, 1208–1209.

Aman, T. K. et al. (2009). Regulation of Persistent Na Current by Interactions between β Subunits of Voltage-Gated Na Channels. *The Journal of Neuroscience* **2009**, *29*.7, 2027–2042.

Ambrosetti, F., Z. Jandova, and A. M. J. J. Bonvin (2020). *A protocol for information-driven antibody-antigen modelling with the HADDOCK2.4 webserver*. **2020**.

Andrews, P. R., D. J. Craik, and J. L. Martin (1984). Functional group contributions to drug-receptor interactions. *Journal of Medicinal Chemistry* **1984**, *27*.12, 1648–1657.

Armstrong, C. M. and F Bezanilla (1977). Inactivation of the sodium channel. II. Gating current experiments. *Journal of General Physiology* **1977**, *70*.5, 567–590. eprint: `https://rupress.org/jgp/article-pdf/70/5/567/1246582/567.pdf`.

Armstrong, C. M. and F. Bezanilla (1974). Charge Movement Associated with the Opening and Closing of the Activation Gates of the Na Channels. *Journal of General Physiology* **1974**, *63*.5, 533–552. eprint: `https://rupress.org/jgp/article-pdf/63/5/533/1245647/533.pdf`.

Bant, J. S. and I. M. Raman (2010). Control of transient, resurgent, and persistent current by open-channel block by Na channel β4 in cultured cerebellar granule neurons. *Proceedings of the National Academy of Sciences* **2010**, *107*.27, 12357–12362.

Berendsen, H. J. C. et al. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*.8, 3684–3690.

Berg, A. and C. Peter (2019). Simulating and analysing configurational landscapes of protein-protein contact formation. *Interface Focus* **2019**, *9*.3, 20180062.

Bernstein, F. C. et al. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology* **1977**, *112*.3, 535–542.

Bonvin, A. M. (2006). Flexible protein-protein docking. *Current Opinion in Structural Biology* **2006**, *16*.2, 194–200.

Bosmans, F., M.-F. Martin-Eauclaire, and K. J. Swartz (2008). Deconstructing voltage sensor function and pharmacology in sodium channels. *Nature* **2008**, *456*.7219, 202–208.

Bowman, G. R., V. S. Pande, and F. Noé, eds. (2014). *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. Springer Netherlands**, 2014**.

Brackenbury, W. J. and L. L. Isom (2011). Na+ Channel Subunits: Overachievers of the Ion Channel Family. *Frontiers in Pharmacology* **2011**, *2*.

Brackenbury, W. J. et al. (2010). Functional reciprocity between Na+ channel Nav1.6 and β1 subunits in the coordinated regulation of excitability and neurite outgrowth. *Proceedings of the National Academy of Sciences* **2010**, *107*.5, 2283–2288.

Bussi, G., D. Donadio, and M. Parrinello (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, *126*.1, 014101.

Catterall, W. A. (2000). From Ionic Currents to Molecular Mechanisms. *Neuron* **2000**, *26*.1, 13–25.

Catterall, W. A. et al. (2007). Voltage-gated ion channels and gating modifier toxins. *Toxicon* **2007**, *49*.2, 124–141.

Cestèle, S. et al. (2006). Structure and Function of the Voltage Sensor of Sodium Channels Probed by a $beta$-Scorpion Toxin. *Journal of Biological Chemistry* **2006**, *281*.30, 21332–21344.

Chen, Y. et al. (2008). Functional properties and differential neuromodulation of Nav1.6 channels. *Molecular and Cellular Neuroscience* **2008**, *38*.4, 607–615.

Chioni, A.-M. et al. (2009). A novel adhesion molecule in human breast cancer cells: Voltage-gated Na+ channel β1 subunit. *The International Journal of Biochemistry Cell Biology* **2009**, *41*.5, 1216–1227.

Clairfeuille, T. et al. (2019). Structural basis of $alpha$-scorpion toxin action on Nav channels. *Science* **2019**, *363*.6433.

Clark, M. A., J. Choi, and M. Douglas (2018). *Biology 2e*. Rice University: OpenStax**, 2018**.

Clatot, J. et al. (2017). Voltage-gated sodium channels assemble and gate as dimers. *Nature Communications* **2017**, *8*.1.

Cox, J. J. et al. (2006). An SCN9A channelopathy causes congenital inability to experience pain. *Nature* **2006**, *444*.7121, 894–898.

Das, S. et al. (2016). Binary architecture of the $Na_v1.2$-β2 signaling complex. *eLife* **2016**, *5*. Ed. by Clapham, D. E., e10960.

Davis, T. H., C. Chen, and L. L. Isom (2004). Sodium Channel β1 Subunits Promote Neurite Outgrowth in Cerebellar Granule Neurons. *Journal of Biological Chemistry* **2004**, *279*.49, 51424–51432.

Dib-Hajj, S. D. et al. (2007). From genes to pain: Nav1.7 and human pain disorders. *Trends in Neurosciences* **2007**, *30*.11, 555–563.

– (2010). Sodium Channels in Normal and Pathological Pain. *Annual Review of Neuroscience* **2010**, *33*.1, 325–347.

Dominguez, C., R. Boelens, and A. M. J. J. Bonvin (2003). HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *Journal of the American Chemical Society* **2003**, *125*.7, 1731–1737.

Doyle, D. A. et al. (1998). The Structure of the Potassium Channel: Molecular Basis of K+ Conduction and Selectivity. *Science* **1998**, *280*.5360, 69–77.

Dudley, S. C. et al. (2000). μ-Conotoxin Giiia Interactions with the Voltage-Gated Na+ Channel Predict a Clockwise Arrangement of the Domains. *Journal of General Physiology* **2000**, *116*.5, 679–690.

Essmann, U. et al. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103*.19, 8577–8593.

Evans, R. et al. (2021). Protein complex prediction with AlphaFold-Multimer**. 2021**.

Faber, C. G. et al. (2012). Gain of function NaV1.7 mutations in idiopathic small fiber neuropathy. *Proceedings of the National Academy of Sciences* **2012**, *109*.47, 19444–19449.

Freitas, R. F. de and M. Schapira (2017). A systematic analysis of atomic protein–ligand interactions in the PDB. *MedChemComm* **2017**, *8*.10, 1970–1981.

French, R. J. et al. (1996). Interactions between a Pore-Blocking Peptide and the Voltage Sensor of the Sodium Channel: An Electrostatic Approach to Channel Geometry. *Neuron* **1996**, *16*.2, 407–413.

Gawali, V. and H. Todt (2016). "Mechanism of Inactivation in Voltage-Gated Na+ Channels". *Na Channels from Phyla to Function*. Elsevier**, 2016**, 409–450.

Gilchrist, J., B. M. Olivera, and F. Bosmans (2014). "Animal Toxins Influence Voltage-Gated Sodium Channel Function". *Voltage Gated Sodium Channels*. Springer Berlin Heidelberg**, 2014**, 203–229.

Gilchrist, J. et al. (2013). Crystallographic insights into sodium-channel modulation by the 4 subunit. *Proceedings of the National Academy of Sciences* **2013**, *110*.51.

Grieco, T. M. et al. (2005). Open-Channel Block by the Cytoplasmic Tail of Sodium Channel β 4 as a Mechanism for Resurgent Sodium Current. *Neuron* **2005**, *45*.2, 233–244.

Groome, J., F. Lehmann-Horn, and B. Holzherr (2011). Open- and closed-state fast inactivation in sodium channels. *Channels* **2011**, *5*.1, 65–78.

Guy, H. R. and P Seetharamulu (1986). Molecular model of the action potential sodium channel. *Proceedings of the National Academy of Sciences* **1986**, *83*.2, 508–512.

Herzog, R. I et al. (2003). Distinct repriming and closed-state inactivation kinetics of Nav1.6 and Nav1.7 sodium channels in mouse spinal sensory neurons. *The Journal of Physiology* **2003**, *551*.3, 741–750.

Hess, B. et al. (1998). LINCS: A Linear Constraint Solver for molecular simulations. *Journal of Computational Chemistry* **1998**.

Hille, B. (1975). The receptor for tetrodotoxin and saxitoxin. A structural hypothesis. *Biophysical Journal* **1975**, *15*.6, 615–619.

Hinrichs, N. S. and V. S. Pande (2007). Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *The Journal of Chemical Physics* **2007**, *126*.24, 244101.

Hodgkin, A. L. and A. F. Huxley (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology* **1952**, *117*.4, 500–544.

Hodgkin, A. L. and R. D. Keynes (1955). Active transport of cations in giant axons from Sepia and Loligo. *The Journal of Physiology* **1955**, *128*.1, 28–60.

Hoshi, M. et al. (2014). Brugada Syndrome Disease Phenotype Explained in Apparently Benign Sodium Channel Mutations. *Circulation: Cardiovascular Genetics* **2014**, *7*.2, 123–131.

Huang, G. et al. (2022). Unwinding and spiral sliding of S4 and domain rotation of VSD during the electromechanical coupling in Nav1.7. *Proceedings of the National Academy of Sciences* **2022**, *119*.33.

Huang, J. et al. (2016). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2016**, *14*.1, 71–73.

Hunter, C. A. (2004). Quantifying Intermolecular Interactions: Guidelines for the Molecular Recognition Toolbox. *Angewandte Chemie International Edition* **2004**, *43*.40, 5310–5324.

Isom, L (1994). Auxiliary subunits of voltage-gated ion channels. *Neuron* **1994**, *12*.6, 1183–1194.

Isom, L. L. et al. (1992). Primary Structure and Functional Expression of the $beta1$ Subunit of the Rat Brain Sodium Channel. *Science* **1992**, *256*.5058, 839–842.

Isom, L. et al. (1995). Structure and function of the $\beta 2$ subunit of brain sodium channels, a transmembrane glycoprotein with a CAM motif. *Cell* **1995**, *83*.3, 433–442.

Jones, S and J. M. Thornton (1996). Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences* **1996**, *93*.1, 13–20.

Jover, E. et al. (1978). Scorpion toxin: Specific binding to rat synaptosomes. *Biochemical and Biophysical Research Communications* **1978**, *85*.1, 377–382.

Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*.7873, 583–589.

Klauda, J. B. et al. (2010). 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.

Kostritskii, A. Y. et al. (2021). g_elpot: A Tool for Quantifying Biomolecular Electrostatics from Molecular Dynamics Trajectories. *Journal of Chemical Theory and Computation* **2021**, *17*.5, 3157–3167.

Kroon, P. C. (2020). "Aggregate, automate, assemble". PhD thesis. University of Groningen**, 2020**.

Lampert, A., M. Eberhardt, and S. G. Waxman (2014). "Altered Sodium Channel Gating as Molecular Basis for Pain: Contribution of Activation, Inactivation, and Resurgent Currents". *Voltage Gated Sodium Channels*. Springer Berlin Heidelberg**, 2014**, 91–110.

Lera Ruiz, M. de and R. L. Kraus (2015). Voltage-Gated Sodium Channels: Structure, Function, Pharmacology, and Clinical Indications. *Journal of Medicinal Chemistry* **2015**, *58*.18, 7093–7118.

Li, R. A. et al. (2001). Clockwise Domain Arrangement of the Sodium Channel Revealed by -Conotoxin (GIIIA) Docking Orientation. *Journal of Biological Chemistry* **2001**, *276*.14, 11072–11077.

Lindahl et al. (2020). GROMACS 2020 Manual**. 2020**.

Lombet, A., J.-N. Bidard, and M. Lazdunski (1987). Ciguatoxin and brevetoxins share a common receptor site on the neuronal voltage-dependent Nav channel. *FEBS Letters* **1987**, *219*.2, 355–359.

Lomize, M. A. et al. (2011). OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research* **2011**, *40*.D1, D370–D376.

Lopez-Santiago, L. F. et al. (2006). Sodium Channel β2 Subunits Regulate Tetrodotoxin-Sensitive Sodium Channels in Small Dorsal Root Ganglion Neurons and Modulate the Response to Pain. *The Journal of Neuroscience* **2006**, *26*.30, 7984–7994.

Mardt, A. et al. (2018). VAMPnets for deep learning of molecular kinetics. *Nature Communications* **2018**, *9*.1.

Marrink, S. J. et al. (2007). The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B* **2007**, *111*.27, 7812–7824.

McEwen, D. P. et al. (2009). The voltage-gated Na+ channel β3 subunit does not mediate trans homophilic cell adhesion or associate with the cell adhesion molecule contactin. *Neuroscience Letters* **2009**, *462*.3, 272–275.

Meents, J. E. et al. (2019). The role of Nav1.7 in human nociceptors: insights from human induced pluripotent stem cell–derived sensory neurons of erythromelalgia patients. *Pain* **2019**, *160*.6, 1327–1341.

Moran, O., F. Conti, and P. Tammaro (2003). Sodium channel heterologous expression in mammalian cells and the role of the endogenous β1-subunits. *Neuroscience Letters* **2003**, *336*.3, 175–179.

Noé, F. and C. Clementi (2015). Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation* **2015**, *11*.10, 5002–5011.

Olsson, M. H. M. et al. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation* **2011**, *7*.2, 525–537.

O'Malley, H. A. and L. L. Isom (2015). Sodium channel subunits: emerging targets in-channelopathies. *Annu Rev Physiol 77: 481–504* **2015**.

Pan, X. et al. (2018). Structure of the human voltage-gated sodium channel Nav1.4 in complex with β1. *Science* **2018**, *362*.6412.

Pan, X. et al. (2021). Comparative structural analysis of human Nav1.1 and Nav1.5 reveals mutational hotspots for sodium channelopathies. *Proceedings of the National Academy of Sciences* **2021**, *118*.11.

Pande, V. S., K. Beauchamp, and G. R. Bowman (2010). Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010**, *52*.1, 99–105.

Payandeh, J. (2018). Progress in understanding slow inactivation speeds up. *Journal of General Physiology* **2018**, *150*.9, 1235–1238.

Payandeh, J. et al. (2011). The crystal structure of a voltage-gated sodium channel. *Nature* **2011**, *475*.7356, 353–358.

Periole, X. et al. (2009). Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *Journal of Chemical Theory and Computation* **2009**, *5*.9, 2531–2543.

Raschka, S. and V. Mirjalili (2017). *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow, 2nd Edition*. Packt Publishing**, 2017**.

Rebala, G., A. Ravi, and S. Churiwala (2019). *An Introduction to Machine Learning*. Springer International Publishing**, 2019**.

Rühlmann, A. H. et al. (2020). Uncoupling sodium channel dimers restores the phenotype of a pain-linked Nav 1.7 channel mutation. *British Journal of Pharmacology* **2020**, *177*.19, 4481–4496.

Schmalhofer, W. A. et al. (2008). ProTx-II, a Selective Inhibitor of Nav1.7 Sodium Channels, Blocks Action Potential Propagation in Nociceptors. *Molecular Pharmacology* **2008**, *74*.5, 1476–1484.

Schmidt, J. W. and W. A. Catterall (1986). Biosynthesis and processing of the α subunit of the voltage-sensitive sodium channel in rat brain neurons. *Cell* **1986**, *46*.3, 437–445.

Schrödinger, L. and W. DeLano (2020). *PyMOL*. Version 2.4.0**. 2020**.

Schwantes, C. R. and V. S. Pande (2013). Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *Journal of Chemical Theory and Computation* **2013**, *9*.4, 2000–2009.

Senior, A. W. et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*.7792, 706–710.

Shahane, G. et al. (2019). Physical properties of model biological lipid bilayers: insights from all-atom molecular dynamics simulations. *Journal of Molecular Modeling* **2019**, *25*.3.

Sharma, S. et al. (2015). A Coarse Grained Model for a Lipid Membrane with Physiological Composition and Leaflet Asymmetry. *PLOS ONE* **2015**, *10*.12. Ed. by Paci, E., e0144814.

Shen, H. et al. (2017). Structure of a eukaryotic voltage-gated sodium channel at near-atomic resolution. *Science* **2017**, *355*.6328.

Shen, H. et al. (2018). Structural basis for the modulation of voltage-gated sodium channels by animal toxins. *Science* **2018**, *362*.6412.

Shen, H. et al. (2019). Structures of human Nav1.7 channel in complex with auxiliary subunits and animal toxins. *Science* **2019**, *363*.6433, 1303–1308.

Shimizu, H. et al. (2017). Parallel homodimer structures of the extracellular domains of the voltage-gated sodium channel β4 subunit explain its role in cell–cell adhesion. *Journal of Biological Chemistry* **2017**, *292*.32, 13428–13440.

Souza, P. C. T. et al. (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature Methods* **2021**, *18*.4, 382–388.

Srivastava, N et al. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res* **2014**, *15*, 1929–1958.

Swope, W. C., J. W. Pitera, and F. Suits (2004). Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory. *The Journal of Physical Chemistry B* **2004**, *108*.21, 6571–6581.

Trainer, V., D. Baden, and W. Catterall (1994). Identification of peptide components of the brevetoxin receptor site of rat brain sodium channels. *Journal of Biological Chemistry* **1994**, *269*.31, 19904–19909.

Ulbricht, W. (2005). Sodium Channel Inactivation: Molecular Determinants and Modulation. *Physiological Reviews* **2005**, *85*.4, 1271–1301.

Varadi, M. et al. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **2021**, *50*.D1, D439–D444.

Wang, C. et al. (2012). Crystal Structure of the Ternary Complex of a NaV C-Terminal Domain, a Fibroblast Growth Factor Homologous Factor, and Calmodulin. *Structure* **2012**, *20*.7, 1167–1176.

Wang, G. K., T. Edrich, and S.-Y. Wang (2006). Time-Dependent Block and Resurgent Tail Currents Induced by Mouse β4 154–167 Peptide in Cardiac Na+ Channels. *Journal of General Physiology* **2006**, *127*.3, 277–289.

Wang, Y. et al. (2017). In Silico Identification of a Novel Hinge-Binding Scaffold for Kinase Inhibitor Discovery. *Journal of Medicinal Chemistry* **2017**, *60*.20, 8552–8564.

Wassenaar, T. A. et al. (2014). Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *Journal of Chemical Theory and Computation* **2014**, *10*.2, 676–690.

Wassenaar, T. A. et al. (2015a). Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *Journal of Chemical Theory and Computation* **2015**, *11*.5, 2144–2155.

Wassenaar, T. A. et al. (2015b). High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *Journal of Chemical Theory and Computation* **2015**, *11*.5, 2278–2291.

Webb, B. and A. Sali (2016). Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **2016**, *54*.1.

Wehmeyer, C. et al. (2019). Introduction to Markov state modeling with the PyEMMA software [Article v1.0]. *Living Journal of Computational Molecular Science* **2019**, *1*.1.

Wisedchaisri, G. et al. (2019). Resting-State Structure and Gating Mechanism of a Voltage-Gated Sodium Channel. *Cell* **2019**, *178*.4, 993–1003.e12.

Wisedchaisri, G. et al. (2021). Structural Basis for High-Affinity Trapping of the NaV1.7 Channel in Its Resting State by Tarantula Toxin. *Molecular Cell* **2021**, *81*.1, 38–48.e4.

Wong, H.-K. et al. (2005). β Subunits of Voltage-gated Sodium Channels Are Novel Substrates of β-Site Amyloid Precursor Protein-cleaving Enzyme (BACE1) and γ-Secretase. *Journal of Biological Chemistry* **2005**, *280*.24, 23009–23017.

Wright, S. H. (2004). Generation of resting membrane potential. *Advances in Physiology Education* **2004**, *28*.4, 139–142.

Xiao, Y. et al. (2008). Tarantula Huwentoxin-IV Inhibits Neuronal Sodium Channels by Binding to Receptor Site 4 and Trapping the Domain II Voltage Sensor in the Closed Configuration. *Journal of Biological Chemistry* **2008**, *283*.40, 27300–27313.

Xiao, Y. et al. (2010). The Tarantula Toxins ProTx-II and Huwentoxin-IV Differentially Interact with Human Nav1.7 Voltage Sensors to Inhibit Channel Activation and Inactivation. *Molecular Pharmacology* **2010**, *78*.6, 1124–1134.

Xie, N.-Z. et al. (2015). Exploring Strong Interactions in Proteins with Quantum Chemistry and Examples of Their Applications in Drug Design. *PLOS ONE* **2015**, *10*.9. Ed. by Obukhov, A. G., e0137113.

Xu, H. et al. (2019). Structural Basis of Nav1.7 Inhibition by a Gating-Modifier Spider Toxin. *Cell* **2019**, *176*.4, 702–715.e14.

Yamaoka, K. et al. (2009). Synthetic Ciguatoxins Selectively Activate Nav1.8-derived Chimeric Sodium Channels Expressed in HEK293 Cells. *Journal of Biological Chemistry* **2009**, *284*.12, 7597–7605.

Yan, Z. et al. (2017). Structure of the Nav1.4-β 1 Complex from Electric Eel. *Cell* **2017**, *170*.3, 470–482.e11.

Yang, Y et al. (2004). Mutations in SCN9A, encoding a sodium channel alpha subunit, in patients with primary erythermalgia. *Journal of Medical Genetics* **2004**, *41*.3, 171–174.

Yang, Y. et al. (2018). Na V 1.7 as a Pharmacogenomic Target for Pain: Moving Toward Precision Medicine. *Trends in Pharmacological Sciences* **2018**, *39*.3, 258–275.

Yereddi, N. R. et al. (2012). The immunoglobulin domain of the sodium channel β3 subunit contains a surface-localized disulfide bond that is required for homophilic binding. *The FASEB Journal* **2012**, *27*.2, 568–580.

Yu, F. H. et al. (2003). Sodium Channel β4, a New Disulfide-Linked Auxiliary Subunit with Similarity to α 2. *The Journal of Neuroscience* **2003**, *23*.20, 7577–7585.

Acknowledgments

I would like to express my deepest gratitude to my supervisor Prof. Dr. Jan-Philipp Machtens. Thank you for trusting on me to lead this huge research project, for sharing your knowledge, and for providing your support. Working under your supervision has been an honor, and I have gained immense learning from you. Thank you for your guidance and for the privilege of collaborating with you.

This research project would not have been possible without the research grant from the RWTH-AICES Graduating school. I would like to express my sincere gratitude for giving the opportunity to work on my research project.

I would also like to thank Prof. Dr. Marc Spehr for his willigness to take over the position of the primary referee for this dissertation.

Many thanks to my colleagues in the Computational Neurophysiology group for their help and support. I would also want to thank all IBI-1 members, I really appreciate the environment you have built.

Profound gratitude for my mother, for giving me the love of learning.

Malagasy proverb :
"Ny fianarana no lova tsara indrindra"
"Education is the most beautiful heritage"