Structural and functional characterization of the protein PDZD7 as part of the Usher2 complex

Baptiste Colcombet-Cazenave

Sorbonne Université

ED515 Complexité du Vivant

Unité Récepteurs-Canaux, CNRS UMR3571 / Institut Pasteur

Structural and functional characterization of the protein PDZD7 as part of the Usher2 complex

Par Baptiste Colcombet-Cazenave

Thèse de doctorat de Biochimie et Biologie Structurale

Dirigée par Nicolas Wolff

Présentée et soutenue publiquement le 26 septembre 2022

Devant un jury composé de :

Dr. Françoise Ochsenbein	Rapporteur
Dr. Matthieu Sainlos	Rapporteur
Dr. Bruno Collinet	Examinateur
Dr. Sébastien Granier	Examinateur
Dr. Julien Henri	Examinateur
Dr. Xavier Morelli	Examinateur

Acknowledgments

First and foremost, I want to thank my supervisor Nicolas Wolff and our unit director Pierre-Jean Corringer for giving me the opportunity to perform my PhD, as well as all members of the jury who have kindly accepted to evaluate my work. I also thank the members of my thesis comity for following my project, providing ideas and support during the four years.

I am grateful to every member of the Channel-Receptor Unit, for their friendship and support. In particular, Nicolas, Célia, Mariano and Yanlei, with whom I spent the most part of my life these last four years. More than colleagues, you have been inspiring mentors and true friends. Thank you to the new comers, Yasmine and Flavio, for your energy and joy. I do not doubt for a second you will perform outstanding PhDs. Talking of energy, this thank extends to all the *file*, see you soon in Greece! I am grateful to the three interns who worked with me, Estelle, Eleana and Louise, I learned a lot working with you. Eleana, I wish you the best luck for your upcoming PhD, I am sure I will hear from you! I could not have done this work without the help of Marie, Nathalie, Stéphanie and, of course, Vincent. Marie and Nathalie, thank you for helping me on a near daily basis. Stéphanie, I apologize for disturbing you so much, I am glad you were here to help in times of need. Vincent, I loved working with you on all animal-related projects, I wish we could continue these projects together.

My project made use of a varied array of techniques, which was rendered possible only by our precious collaborators and platforms of the Institut Pasteur. I will start by a special thanks to Florence. You dealt with my messy organization with patience and put up with our last-minute-emergency-spectra-analysis with kindness. Another thanks go to Julien, you taught me a set of skills that I now apply to all projects I am involved in. I am also obliged towards the members of the Structural Studies of Macromolecular Machines in Cellulo and Structural Imaging units, with whom I started the tomography project and who were always available to help. I am grateful to the whole platform of biophysics, in particular Bertrand and Sébastien, for the cheerful working atmosphere you convey, your dedication to users' assistance and your flexibility, always finding solution to ease my work. I also want to thank the nanoimaging core facility, in particular Stéphane and Jean-Maris, for their help on the electron microscopes. Double thanks to Stéphane for the time spent on the Aquilos microscope, despite the failures associated to testing and developing new approaches. Finally, I thank Gérard and Olivier from the Ultrastructural BioImaging platform, who introduced me to electron microscopy and high pressure freezing.

Above all, my gratitude goes to my family and friends, for their patience and care throughout my PhD. I have neglected you these past years, and for your understanding I am forever in debt.

Summary / Résumé

General audience

In mammals, the cochlea is the organ of hearing, and the retina the organ of vision. Usher syndrome is the first disease associating hearing loss and blindness in human, accounting for half the cases. This disease is caused by mutations in proteins expressed both in the cochlea and retina. In the cochlea, they are responsible for the morphogenesis of sensory cells and the detection of sounds. In the retina, they would promote the recycling of photosensitive molecules. In Usher patients, one or multiple mutations alter these functions. However, the way these proteins perform their function is not yet understood. This PhD project therefore aims at deciphering their interactions at the molecular scale, and at characterizing the complex assembly at the cellular level, comparing the healthy proteins to their mutated counterparts to understand the mechanisms underlying the physiopathogenesis of the disease.

Résumé grand public

Chez les mammifères, la cochlée est l'organe sensoriel de l'audition et la rétine celui de la vision. Le syndrome Usher est la première cause de surdité combinée à de la cécité chez l'Homme, représentant la moitié des cas recensés. Cette maladie est provoquée par des mutations génétiques de protéines dites Usher, trouvée à la fois dans la cochlée et la rétine. Dans la cochlée, elles sont responsables du développement des cellules sensorielles et de la détection des sons. Dans la rétine, ces protéines permettraient le recyclage des molécules photosensibles. Chez les patients Usher, une ou plusieurs mutations altèrent ces fonctions. Cependant, la façon dont ces protéines interagissent pour assurer leurs différentes fonctions n'est pas encore comprise. Ce projet de thèse vise donc à décrire ces interactions depuis l'échelle moléculaire jusqu'à l'échelle cellulaire, comparant les protéines normales aux protéines mutées, afin de comprendre les mécanismes conduisant à la maladie.

Scientific audience

Hearing relies on the capacity of specialized sensory hair cells in the cochlea to transduce sound-induced vibrations into electrical signals that are transmissible to the brain. Hair cells possess actin-filled stereocilia structured into staircase-shaped bundles deflected by sound-waves. Large protein complexes are found at the anchoring sites of extracellular links that interconnect stereocilia. Mutations of these proteins are responsible for hereditary sensory diseases, notably the Usher syndromes. The Usher syndrome of type 2 (Usher 2) is the most common form genetic cause of combined congenital deafness and progressive blindness. The Usher 2 protein complex, involved in the morphogenesis of the hair bundles, encompasses two large transmembrane proteins, usherin and the G protein-coupled receptor (GPCR) ADGRV1, with very large extracellular domains forming fibrous links between the stereocilia. These two proteins possess a cytoplasmic region in interaction with the scaffolding proteins whirlin and PDZD7, which in turn associate to actin-binding proteins. Usher proteins contain numerous protein-protein interaction domains necessary to the intricacy of the network, but the network's assembly remains elusive, thus leaving the effect of mutations detected in patients to speculation. In this project, I studied the Usher 2 complex components from the molecules (*in silico* and *in vitro*) to the complex assembly at the cellular level.

First, I deeply analyzed the sequences of the orphan domain family HHD, found in few neuronal proteins, using bioinformatics tools to predict functional surfaces involved in protein-protein interactions and the effect of human pathogenic mutations. Then, I characterized the determinants of interaction between the adhesion GPCR ADGRV1 and its PDZ-domain containing regulator PDZD7. I showed that the two N-terminal PDZ domains of PDZD7 are able to interact with ADGRV1 PDZ binding motif, with a higher affinity for the second PDZ. This interaction requires C-terminal extension of the PDZ domains which likely adopts a beta strand conformation in solution. I showed that two human pathological mutations of PDZD7 PDZ domains trigger a drastic decrease of affinity for ADGRV1, potentially disrupting the Usher2 complex. To further understand the activation mechanism of ADGRV1, I started its structural characterization in complex with its associated G protein by cryo-electron microscopy. At the cellular level, I used high resolution STED microscopy to decipher the accurate localization of the Usher 2 complex and its anisotropic distribution between hair cell stereocilia rows. Finally, I initiated an ambitious cryo electron tomography project to solve the general organization of the Usher 2 complex in situ. To this end, I optimized the cochlea preparation for cryo-Correlative Light and Electron Microscopy Focused Ion Beam milling (cryo-CLEM FIB milling) and performed the first lift-out procedure on mouse dissected cochlea. Altogether, the obtained results will help to understand the physiopathology of mutations associated to the Usher syndrome of type 2.

Résumé scientifique

L'audition repose sur la transduction, par les cellules sensorielles de la cochlée, des ondes sonores en signaux électriques alors transmis au cerveau. Ces cellules, dites ciliées, possèdent des touffes de stéréocils d'actine organisés en rangées de taille croissante. De larges complexes protéiques sont trouvés aux sites d'ancrage de liens extracellulaires connectant ces stéréocils entre eux. Des mutations dans ces protéines induisent des maladies sensorielles héréditaires, notamment le syndrome Usher. Le syndrome Usher de type 2 (Usher 2) est la première cause de surdité congénitale associée à une cécité chez l'Homme. Le complexe protéique Usher 2, impliqué dans le développement des touffes ciliaires, est composé de deux protéines transmembranaires, l'usherine et ADGRV1, qui possèdent de larges domaines extracellulaires formant des liens fibreux à la base des stéréocils. Ces deux protéines ont un domaine intracellulaire en interaction avec les protéines d'échafaudage whirline et PDZD7, qui interagissent à leur tour avec des protéines de liaison à l'actine. Les protéines Usher possèdent de multiples domaines d'interaction protéine-protéine nécessaires à la formation de complexes, mais l'organisation générale du réseau reste largement inconnue, ce qui ne permet pas d'interpréter l'effet de mutations Usher. Dans ce projet de thèse, j'ai étudié le complexe Usher 2 depuis l'échelle moléculaire (*in silico* et *in vitro*) jusqu'à l'échelle cellulaire.

J'ai en premier lieu analysé les séquences protéiques de la famille de domaines HHD, exprimés seulement dans quelques protéines neuronales, en utilisant des outils bioinformatiques afin de prédire la fonction de différentes surfaces dans des interactions protéine-protéine ainsi que l'effet de mutations pathogènes humaines. Par la suite, je me suis intéressé aux déterminants d'interaction entre le récepteur d'adhésion couplé aux protéines G (RCPG) ADGRV1 et sa protéine régulatrice à domaine PDZ PDZD7. J'ai ainsi montré que les deux domaines N-terminaux de PDZD7 interagissent avec le motif d'interaction aux PDZ (PBM) d'ADGRV1, avec une meilleure affinité pour le second PDZ. Cette interaction requiert la présence d'extensions C-terminales dans les deux domaines PDZ. Ces extensions adoptent une structure en brins beta. J'ai montré que des mutations de PDZD7 provoquent une perte drastique d'affinité pour ADGRV1, dissociant potentiellement le complexe Usher 2. Afin de mieux comprendre les mécanismes d'activation d'ADGRV1, j'ai entamé l'étude structurale du complexe RCPG/protéine G par cryo-microscopie électronique. A l'échelle cellulaire, j'ai utilisé la microscopie de fluorescence à haute résolution STED afin de décrire la distribution du complexe Usher 2 entre les différents stéréocils des cellules ciliées. Enfin, j'ai entamé un projet ambitieux de cryo-microscopie électronique en tomographie visant à déterminer l'organisation générale du complexe Usher 2 *in situ*, dans l'organe de l'audition. J'ai ainsi optimisé la préparation d'échantillon pour « cryo-Correlative Light and Electron Microscopy Focused Ion Beam milling (cryo-CLEM FIB milling) » et réalisé le premier prélèvement d'une lamelle de cochlée de souris disséquée. Les résultats obtenus permettront de mieux comprendre la physiopathologie des mutations associées au syndrome Usher de type 2.

Table of contents

Introduction

Hearing is the capability to perceive sounds, which is one of the five physiological senses through which an organism can interface with its environment. Largely attributed to the vertebrates[1], it is a key survival trait aiding in avoiding predators and pursuing preys. Hearing also became a social tool to engage communication with other individuals, within and across species. In particular, humans have developed elaborate languages unparalleled by other animals[2] and made complex sounds the base of most social interactions. As a result, hearing defects in human correlate with lower subjective well-being[3,4] (loneliness, low self-esteem), social isolation[3], increased psychological distress[4–7] (anxiety, depression) and increased susceptibility to dementia[7,8]. According to the World Health Organization, over 430 million people suffer from disabling hearing loss worldwide in 2021, accounting for more than 5% of the population[9]. Hearing loss is clinically heterogenous, with varying severity, frequency coverage or age of onset. It can be progressive or sudden and affect one or both ears. This heterogeneity reflects the diverse causes underlying the hearing loss, including genetic factors, infections, trauma, other conditions or risk factors (medication, smoking, chronic diseases…), aging or recreative exposure to loud noise[9]. These can affect any step of the hearing process, from the entry of sound waves by the outer ear up until the signal interpretation by the brain.

A sound is a mechanical and longitudinal wave travelling through a non-vacuum medium. Molecules of a plane perpendicular to the propagation of the wave undergo in phase harmonic oscillations around a position of equilibrium. Progressive phase desynchronisation between the oscillations of molecules from different planes induces local high-pressure regions (compressions) and low-pressure regions (rarefactions) that appear to travel the medium. Sound waves are thus effectively periodic pressure waves, capable of inducing oscillations and resonance effects. The human ear is capable of detecting sounds of amplitudes corresponding to atom displacements in the tenths of picometers, barely above brownian motion[10]. It processes sound amplitudes in a logarithmic manner, with a ten-fold increase in pressure resulting in a single increment of perceived intensity (decibel, dB). In addition to a sensitivity over a broad range of amplitudes, from 0dB to about 120dB, the human ear is able to process sounds over a large range of frequencies, as low as 20Hz and up to 20kHz.

Here I will describe how the ear performs such a feat by presenting its individual components, following the travel of sound-induced vibrations throughout the auditory system. I will focus on the mammalian ear, using human nomenclature for proteins length and organization.

I – Functioning and physiology of the ear

I.1 – The outer and middle ears

The outer ear consists of the auricle and the ear canal (Figure 1). Its main function is to focus sound waves to the eardrum. Collection of sound waves is anisotropic, favouring the direction facing the auricle. In human, the auricle has very reduced motility, focusing sounds coming from the direction faced by the individual. In many animals, the auricle can be reoriented by the three auricular muscles to track sounds coming from several directions without tilt of the head. Due to its shape, the auricle is also capable to discriminate the height of a sound's source relative to the ears. The outer ear is separated from the middle ear by the eardrum. It is a membrane that oscillates in response to the sound-induced variations of pressure in the ear canal, effectively acting as a reversed drum instrument. It translates local vibrations of air molecules in a mechanical oscillation that is further transmitted to the middle ear. The middle ear is normally filled with air, ventilated by the Eustachian tube. The ossicles are middle ear's mechanical intermediates in the transmission of pressure variations from the outer ear to the liquid medium of the inner ear. They consist of the three smallest human bones, namely the malleus, incus and stapes.

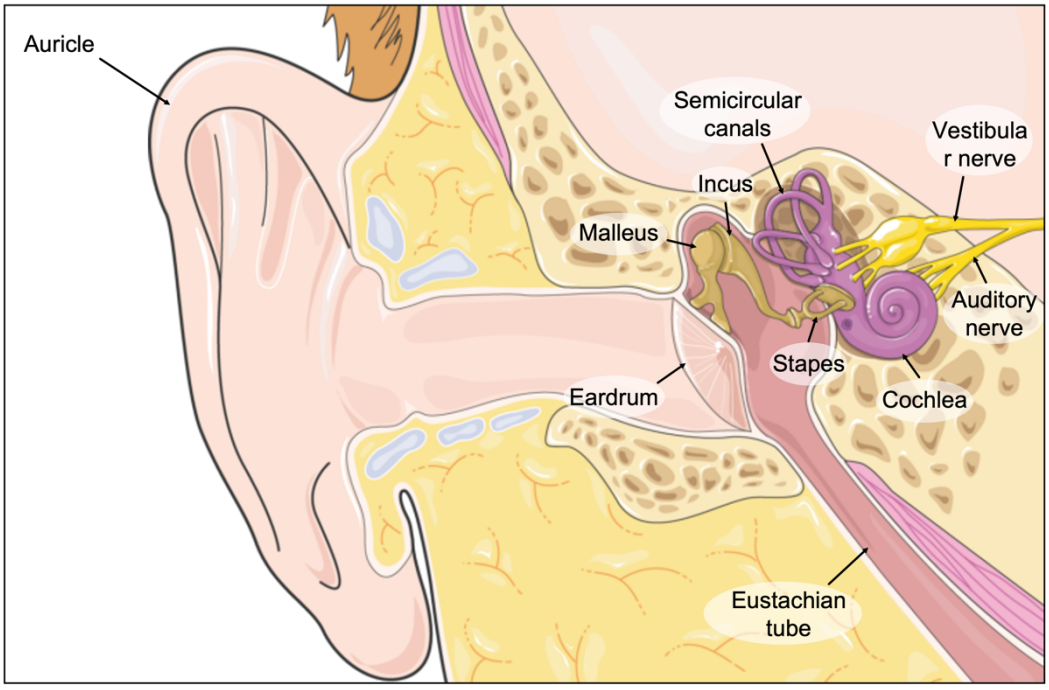


Figure 1 - Anatomy of the human ear. Made with Biorender.

The need for a mechanical intermediate between the airborne outer ear and the aqueous inner ear arises from the large difference in impedance of the two media. Acoustic impedance is calculated as the ratio between the acoustic pressure and the celerity of associated particles in a medium. Therefore, it depends on the density of the medium and reflects the energy of a given sound wave travelling it, or to the energy necessary to induce such sound wave. A sound wave travelling in a low impedance medium (air) carries too little energy to induce a similar wave in a high impedance medium (aqueous). This results not only on a low amplitude transmitted wave, but also in a reflected wave that carries most of the incident energy. To overcome this issue, ossicles focus the acoustic pressure received on the eardrum to a much smaller surface area of the inner ear. The malleus and incus complex motion also levers the oscillations of the eardrum. These two mechanisms increase the pressure applied by the stapes at the interface with the liquid medium of the inner ear, performing impedance adaptation between the two media[11].

I.2 – The inner ear

I.2.a – General morphology

In contrast with the airborne outer and middle ears, the inner ear is filled with aqueous medium. It contains two main structures, the vestibule and the cochlea (Figure 2A). The vestibule consists of the semi-circular canals, which detect angular acceleration of the head in the (x,y,z) axes, and of the otolithic organs, which detect planal acceleration of the head in the (x,y) plane (performed by the utricule) and vertical acceleration in the (z) axis (performed by the saccule). The information is relayed to the brain through the vestibular nerve, allowing to maintain balance. The perception of sounds is achieved by the cochlea. The mammalian cochlea is a bony structure with a snail-like shape. Encased in the so-called bony labyrinth is found the cochlear duct, forming a coiled tube. In human, the cochlear duct makes about two and a half turns for a total length of about 30 millimeters[12]. Its walls are covered by cells that form the membranous labyrinth. The cochlear duct is split in three compartments that run all along the axis of the cochlea. These three canals are the *scala vestibuli*, the *scala media* and the *scala tympani* (Figure 2B). The *scala vestibuli* and *scala tympani* are filled with perilymph (high Na^+ concentration, low K^+ and Ca^{2+} concentrations)[13]. They communicate at the helicotrema[14], located at the apex of the cochlea. The *scala media* is closed on both ends, separated from the *scala vestibuli* by the Reissner's membrane, and from the *scala tympani* by the basilar membrane. Seclusion of the *scala media* is necessary as it is filled with endolymph

(low Na^+ concentration, high K^+ and Ca^{2+} concentrations). The difference between ionic concentrations of the perilymph and endolymph in the cochlea is mainly maintained by the *stria vascularis*[15]. The *stria vascularis* is located along the lateral wall of the cochlear duct, facing the *scala media*. Cells of the *stria vascularis* notably express $Na^+/2Cl^-/K^+$ cotransporters and Na^+/K^+ ATPases and cycle potassium ions from the *scala vestibuli* and *scala tympani* to the *scala media*, while maintaining endolymph's low Na^+ concentration[16,17].

The organ of Corti, sensory epithelium for hearing, lies on the basilar membrane and emerges in the *scala media*, facing the tectorial membrane. It encompasses the sensory hair cells, as well as several types of supporting cells (Figure 2C).

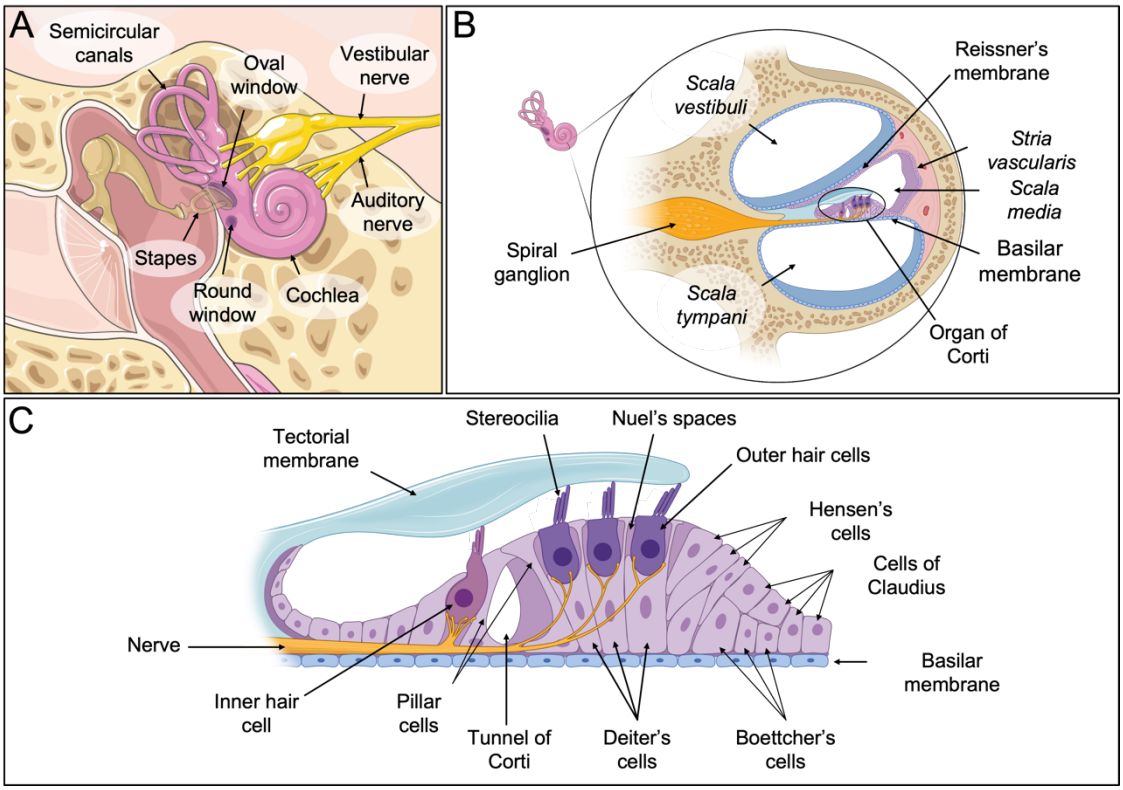


Figure 2 - A) The human inner ear. B) View of a cut perpendicular to the running axis of the cochlea. C) The architecture of the organ of Corti. Made with Biorender.

Hair cells possess actin-based stereocilia at their apical surface, referred to as hair bundle, bathed in the endolymph of the *scala media*. There are two types of hair cells, inner hair cells (IHCs) and outer hair cells (OHCs). There is a single row of IHCs and three rows of OHCs, running all along the axis of the cochlea. The different functions of the IHCs and OHCs in sound detection will be described in the following section. Amongst supporting cells are found Deiter's cells. These cells emerge from the basilar membrane, holding the base of the OHCs. They also project a branch-like extension, the phalangeal process, towards the surface of the organ of Corti. The extremity of the phalangeal process (phalangeal apex) fills the gaps

between OHCs[18]. The gap between the apex of OHCs and IHCs is filled by pillar cells, largely contributing to the structural cohesion of the organ of Corti[15]. Hensen's cells are found next to the third row of OHCs[19]. The extremities of the organ of Corti are sealed by the cells of Claudius, as well as Boettcher cells found in the lower turn of the cochlea. Altogether, these supporting cells and hair cells form a continuous surface in the *scala media*, also covered by the collagen III-rich reticular lamina, preventing ion flux towards the inner part of the Organ of Corti[15]. Several cavities are found inside the organ of Corti, the tunnel of Corti formed between pillar cells and the spaces of Nuel formed between hair cells. These cavities are filled with cortilymph, of similar composition to the perilymph of the *scala vestibuli* and *scala tympani*[20]. Therefore, the apical surface of the hair cells is exposed to a different environment than the basal bodies, surrounded by endolymph and perilymph, respectively.

I.2.b – Propagation of sound waves

Upon perception of a complex sound, stimulation of the eardrum induces oscillation of the ossicles that relay the sound wave to the cochlea. The stapes is in contact with the *scala vestibuli* at the base of the cochlea, through a hole in the bony labyrinth called the oval window. The sound waves propagate in the *scala vestibuli* up to the apex of the cochlea, echoes at the helicotrema and travels back to the base through the *scala tympani*. A second opening, the round window, is located at the base of the *scala tympani* and oscillates in response to the displacement of liquid induced by the stimulation of the stapes and allows the diffusion of energy carried by the sound wave. Propagation of the sound wave along the cochlear duct results in vibrations of the basilar and tectorial membranes. These two membranes decrease in thickness and stiffness all along the axis of the cochlea[21,22]. Their elasticity is therefore higher at the apex. As a result, the specific frequency at which stimulation induces maximum oscillation of these membranes differs all along the cochlea, with the stiffer base resonating with high frequencies and the apex with low frequencies. Therefore, the cochlea mechanically achieves the Fourier transform of a complex incoming sound in its frequency components. Each frequency induces local vibrations of the basilar and tectorial membranes around a specific location of the cochlea, allowing their independent detection by few hair cells at that position. Furthermore, the tectorial and basilar membranes are anisotropic tissues. Notably, the tectorial membrane is mainly composed of collagen fibbers (types II, V, IX and XI), extending perpendicular to the axis of the cochlea. As a result, vibrations of the tectorial membrane travel further radially (along the collagen fibbers) than longitudinally (along the length of the cochlea)[23]. These fibbers vary in length along the axis of the cochlea, further

tuning the specific frequency of radial oscillations of the tectorial membrane at different locations of the cochlea[24]. This further restricts the area stimulated for each pure component of a complex sound.

The shape of the bony labyrinth is critical to the propagation of sound waves up to the apex of the cochlea. Its curvature allows the propagation of sound waves with minimal energy loss. Vibrations are focused on the outer wall of the cochlear duct, bouncing further down the cochlea. This phenomenon is similar as in whispering galleries, capable of carrying even low amplitude sounds over long distances. This effect is stronger towards the apex of the cochlea, as the curvature of the cochlear duct increases[23]. As presented in the previous paragraph, low frequencies are analysed at the apex of the cochlea. These frequencies could thus not be detected without the coiled structure of the cochlear duct. The low frequency hearing limit across mammalian species correlates with the coil radius and length of the cochlea[25].

I.3 – Hair cells

Hair cells possess actin based stereocilia forming a hair bundle at their apical pole, organized in rows of increasing height (Figure 3). Both types of hair cells, inner hair cells (IHCs) and outer hair cells (OHCs), can detect local vibrations through deflection of their hair bundle.

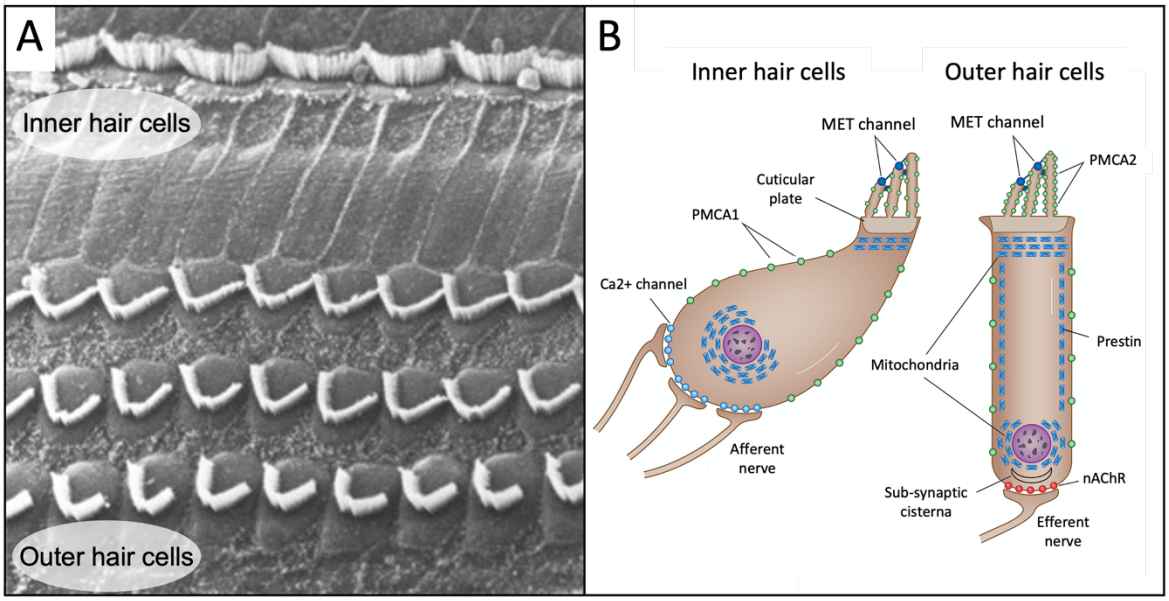


Figure 3 - A) Scanning electron microscopy image of dissected mouse organ of corti surface. From Maria, P. L., & Soghalai, J. (2015)[373]. B) Morphology of cochlear hair cells. Modified from Fettiplace, R., & Nam, J. H. (2019)[374].

IHCs are largely responsible for the transduction of the local vibrations into an electrical signal, referred to as mechanotransduction. Their basal body possesses multiple synapses directly interfacing with efferent nerves connected to the spiral ganglion. Outer hair

cells act as amplifiers and frequency filters. They are connected to the nervous system by afferent nerves and are actively stimulated by the brain. In total, a mature human organ of Corti houses about 15000 hair cells (one row of IHCs for 3000 cells and three rows of OHCs for 12000 cells)[26].

I.3.a – General morphology

There are three rows of OHCs in the organ of Corti. The hair bundle of each OHC consists of three rows of actin-based stereocilia of increasing height, adopting a V or U shape. The tip of the tallest row is embedded in the facing tectorial membrane, attached to it by the otogelin, otogelin-like and stereocilin proteins. These proteins also contribute to the formation of the top connectors, fibrous links tethering together stereocilia. These links emerge from the lateral face of stereocilia tip, connecting to the tip of neighbouring stereocilia from the same row (intra-row) or to the side of neighbouring stereocilia from the next taller row (inter-row)[27,28]. In addition, diverse protein links interconnect stereocilia, such as lateral links or shaft connectors, aiding the cohesive deflection of the hair bundle[29].

There is one row of IHCs in the organ of Corti. The hair bundle of each IHC consists of two main rows of actin-based stereocilia of increasing height, with the last row being significantly taller. The IHCs' hair bundle has little curvature of its stereocilia rows compared to OHCs, and overall longer stereocilia. It was recently suggested that the tallest row of stereocilia of IHCs is embedded within the tectorial membrane in mature cochlea[30], while it was thought to be the case only during early stages of development[26,31]. The hypothesis of fully anchored IHCs in the tectorial membrane would indicate direct mechanical coupling between IHCs and OHCs. Like OHCs, stereocilia of IHCs are connected by diverse lateral links.

The hair bundles of both cell types share similar features. At their base, stereocilia width decreases towards the surface of the cell body, forming the taper region (Figure 4). The resulting dense actin network is inserted in the cell surface, forming stereocilia rootlets. Insertion of the stereocilia as a dense structure with a small surface area is necessary to allow deflection of the stereocilia upon stimulation, while rootlets ensure resistance to damage. Tightening of the actin filaments is promoted by large protein complexes. It notably involves

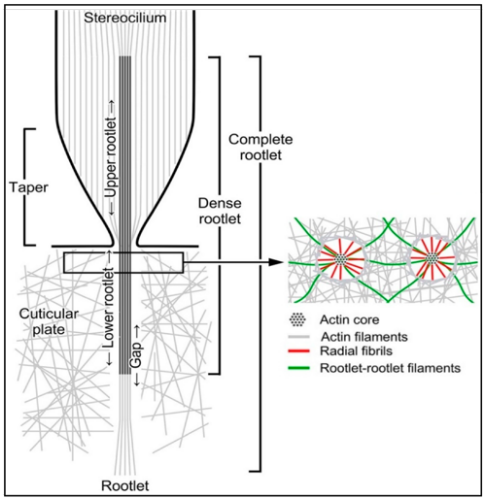


Figure 4 - The taper region. From Pacentine et al. (2020)[32].

the actin-binding proteins TRIOBP-4 and -5, as well as taperin[32]. These proteins appear to participate in the protein assembly by interaction with other partners (GRXRC2[33], whirlin[34]). Other proteins, notably spectrin and RIPOR2, form ring-like structures that sheath stereocilia rootlets. The apical surface of hair cell soma also has peculiar properties. The submembrane space encompasses a dense network of actin filaments called the cuticular plate. Actin filaments are organized by the LMO7 protein and cross-linked by actin-binding proteins such as spectrin[35] and supervillin[36], further strengthening stereocilia rootlets anchoring. The cuticular plate also contributes to the overall surface structure of the organ of Corti.

I.3.b – OHCs as amplifiers and energy filters

The cellular body of OHCs is cylindrical, increases in length from the base (~10µm) to the apex (~80µm) of the cochlea[37,38], and rests on Deiter's cells. The basal body of OHCs is mostly interfacing with efferent neurons (signal coming from the brain)[39]. Therefore, OHCs can be activated by direct mechanotransduction at the hair bundle in response to local vibrations, as well as by active stimulation from the brain.

OHCs are contractile cells, a property required to perform their function as frequency filters and amplifiers. The cellular body contracts during depolarization and elongates with hyperpolarization[40], for a change in length of about 5%[38]. Upon stimulation, K^+ and Ca^{2+} influx from the *scala media* induces depolarization and shortening of the cell[40]. Electromotility of the OHCs is independent from ATP, microtubules or actin[41]. Instead, it is promoted by the protein prestin. Prestin has evolved from an anion-transporter ancestor[42]. It is abundantly expressed in OHCs and forms oligomers at their basolateral membrane[43]. Its hydrophobic core is inserted in the membrane but has lost its transport capability in mammals[44]. Prestin still undergoes structural rearrangement depending directly on membrane polarization[45]. This rearrangement would result from partial translocation of Cl^- anions in hyperpolarized cells[46]. The anion-bound form of prestin occupies a larger surface area in the membrane. Upon depolarization of the cell, anion binding is no longer favoured, resulting in a compaction of prestin structure and the macroscopic contraction of the outer hair cell. Then, the membrane potential is restored quickly by Ca^{2+}-activated K^+ (BK) channels, releasing potassium ions in the K^+ poor cortilymph, resulting in cellular elongation. Successive hyper- and depolarizations of OHCs result in oscillations transmitted to the tectorial membrane, therefore amplifying local vibrations. Notably, the rate of hyperpolarization after stimulation is tuned by differential expression of the BK potassium channel α and β subunits. The α subunit forms the pore of

the channel. Splicing variants of the α subunit exhibit differential current amplitudes and gating kinetics[47]. Co-expression of the auxiliary β subunit further decreases gating kinetics of the channel[48]. Higher densities of BK channels α subunits are found at the base of the cochlea, associated with lower expression of accessory β subunits. Towards the apex, BK channel density on the hair cell lateral walls decreases, with an increased expression of accessory β subunits[49]. This induces a faster hyperpolarization of hair cells at the base of the cochlea and slower at the apex, resulting in different maximum frequency stimulations along the cochlea. This mechanism is called electrical tuning, further establishing the frequency specificity of hair cells[50].

I.3.c – IHCs as primary mechanosensors

Both IHCs and OHCs can detect sound-induced local vibrations, a mechanism referred to as mechanotransduction. However, IHCs are largely responsible for reporting sound detection to the brain (Figure 5). Mechanotransduction is triggered by deflection of the hair bundle in the direction of the taller stereocilia row (positive), which increases the tension on a specific protein link, the tip link, which bridges the top of each lower row stereocilia to the side of their taller neighbour. The force is relayed through an unclear mechanism to a cation channel, increasing its opening probability and leading to an influx of K^+ and Ca^{2+} from the *scala media* that depolarizes the cell. The depolarization ultimately results in the release of neurotransmitter at the base of the cell, which is mostly innervated by afferent neurons[15]. Deflection of the hair bundle in the opposite direction (negative) closes the MET channels, dampening cations influx and allowing repolarization of the cell. Depolarization and hyperpolarization rates must match the frequency of the stimulation. To this end, physical properties of the MET channels at the stereocilia apex and ion channels of the cell body differ, depending on the cell characteristic frequency. The next chapter is dedicated to the mechanotransduction and its key molecular components.

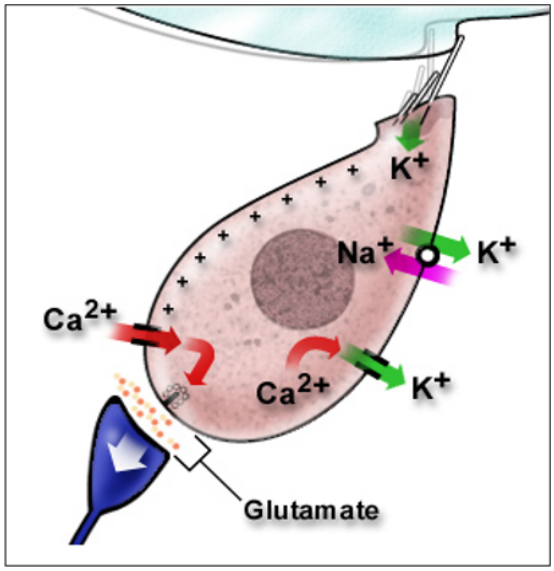


Figure 5 - Excitation scheme of an inner hair cell upon positive deflection of the hair bundle.

I.3.d – Morphogenesis of the hair bundle

During embryogenesis, the apical pole of cochlear hair cell is covered by brush-like microvilli. A unique microtubule-based cilium, the kinocilium, grows at the center of the cell surface and then migrates towards the outer coil of the maturing cochlea, defining the polarization of the hair cell apical pole. The kinocilium behaves as a prop and promotes the growth of a first row of microvilli, which will become the taller stereocilia row of the hair bundle. Formation of a stereocilia row in turn induces the growth of microvilli close to it, resulting in the staircase organization of the hair bundle. During this process, multiple protein links form between nascent stereocilia, thought to stabilize the fragile hair bundle and regulate its architecture. Later, remaining microvilli will disappear and grown stereocilia will thicken by addition of actin filaments and anchor in the cuticular plate of the cell. The protein links then regroup at more defined locations along the stereocilia, reflecting the diversity of composing proteins and functions (Figure 6)[29]. In most mammals, the kinocilium then disappears before birth. Some of these links will be maintained in the mature hair bundle (tip-link, top connectors), while others disappear (kinocilial links, shaft connectors, ankle links). Depending on the organism, cochlear maturation steps might happen before or after birth and some molecular components might be maintained in the mature hair bundle or disappear in later stages of its morphogenesis. In addition, maturation of hair cells along the axis of the cochlea is not concomitant, with about a two-day delay for the cells of the apex compared to those of the base. For simplicity, most studies report morphological and protein expression changes of cells located at the middle turn of the cochlea.

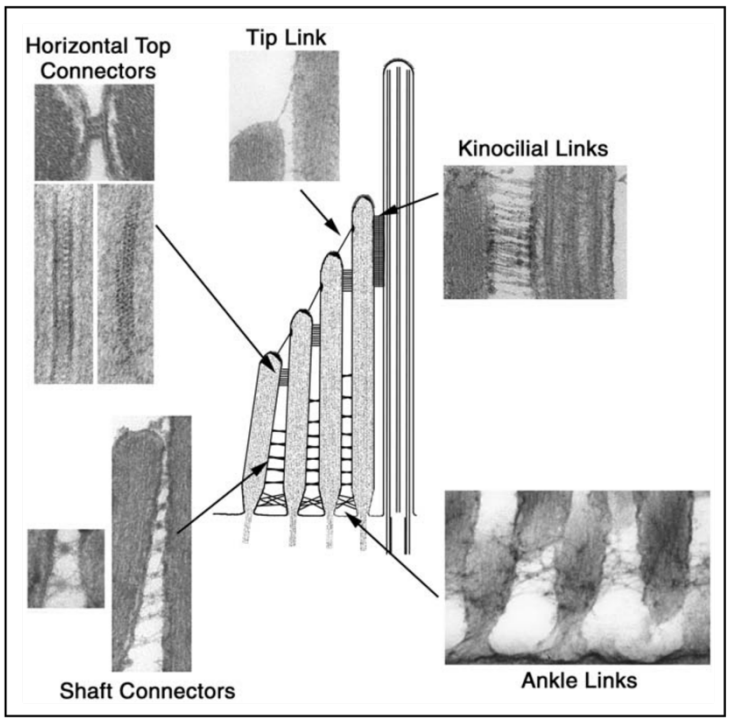


Figure 6 - Transmission electron microscopy imaging of the various links found between hair cell stereocilia during development. From Nayak et al. (2007)[375].

Morphogenesis of the hair bundle is critical for the good functioning of hair cells. The mature hair bundle must meet several criteria, including stereocilia of the correct height, with correct pattern of organization (bundle shape and stereocilia alignment) and capable of tilting in response to local vibrations in a cohesive manner. These characteristics are required for the concerted opening of all MET channels upon stimulation of the hair bundle and the correct excitability of hair cells.

I.4 – Molecular basis of the mechanotransduction

In this chapter, I will present the key molecular components necessary to hair cells for sound detection, including the coupling of the tip-link to cytoplasmic protein networks and the release of neurotransmitter at the cell synapses. If several of the presented interactions have been studied using structural and biophysical methods, most of the network interactions has been described in cellular models using pull-down or co-immunoprecipitation assays.

I.4.a – The tip-link

The tip link is composed of two proteins, the protocadherin15 (PCDH15, bottom part) and the cadherin23 (CDH23, top part), each forming a homodimer. These two proteins have a large N-terminal extracellular region composed of 11 and 27 extracellular cadherin domains (ECD), respectively. These domains promote protein-protein interactions, such as the formation of PCDH15/PCDH15 and CDH23/CDH23 helical dimers, as well as the hetero-tetramer handshake structure bridging the two homodimers together[51,52] (Figure 7). Calcium ions coordinate the linkers between ECDs of the two proteins and are crucial for tip-link mechanical properties[53]. A bend between the ECD 9 and 10 of PCDH15 allows for an elastic response of the tip-link to the applied tension (Figure 8). Both PCDH15 and CDH23 possess a single transmembrane segment and a C-terminal cytoplasmic domain of about 550 and 250 residues, respectively. The transmembrane segment and cytoplasmic domain of the two proteins mediate the interaction with protein networks at the lower and upper anchoring sites of the link. Several isoforms of PCDH15 and CDH23 are expressed differentially depending on cell maturity[54], with sequence differences notably located in the cytoplasmic tail of the proteins.

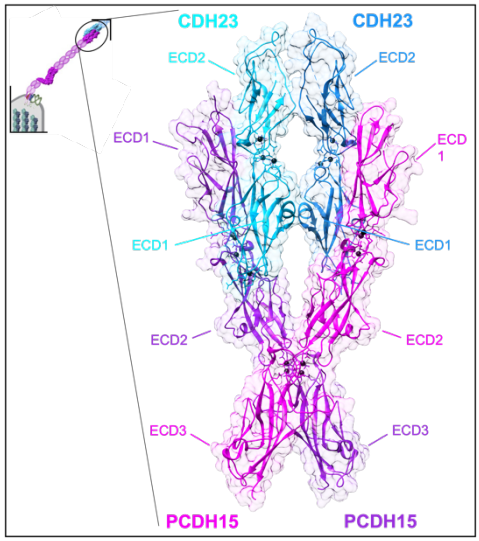


Figure 7 - X-ray structure of the protocadherin-15 and cadherin23 dimers handshake structure. Calcium ions are colored in black. PDB entry: 6N2E.

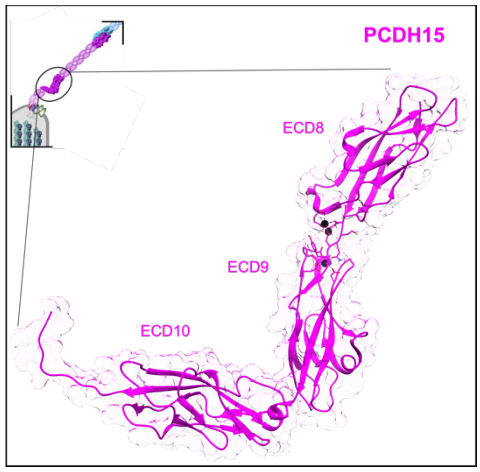


Figure 8 - X-ray structure of protocadherin-15 ECD9-10 bend, highlighting the role of calcium ions (in black) in conformation switches of ECD domains pairs. PDB entry: 4XHZ.

In mature cells, the isoform CD2 of PCDH15 is mostly expressed as the lower tip-link component.

I.4.b – The Lower Tip-Link Density (LTLD)

At the lower anchoring site of the tip-link, PCDH15 interacts with several transmembrane and cytoplasmic proteins, forming the lower tip-link density. These proteins notably form the mechanotransduction (MET) channel. The lower tip-link density is therefore responsible for coupling variation of tension in the tip-link to the opening of the MET channel, inducing K^+ and Ca^{2+} influx into the cell.

At the lower anchoring site, PCDH15 interacts with the scaffolding protein whirlin[55]. Whirlin is a 907 residues protein containing three PDZ (PSD95; Dlg1; Zo-1) domains and two HHDs (Harmonin Homology Domain). Both domain types promote protein-protein interactions and will be further documented in their dedicated sections. Whirlin in turn associates to actin-binding proteins such as myosin15a[34], anchoring the link to the actin filaments of stereocilia. PCDH15 also interacts with several transmembrane proteins forming the MET channel or associated to it (Figure 9). In mature cells, PCDH15 interacts with the channel-like protein 1 (TMC1)[56] forming the pore of the MET channel[57]. TMC1 possesses six transmembrane segments and assembles as a homo-dimeric channel. TMC1 trafficking to the lower tip-link anchoring site is regulated by the TOMT (Transmembrane O-methyltransferase) protein[58]. TOMT does not localize in the stereocilia and is rather enriched in the golgi apparatus of hair cells. The mechanism through which TMC trafficking is regulated is still unclear, but the identification of deafness-associated proteins such as TOMT in the secretory pathway of the cell highlights the need for a specialized transporting machinery to address MET complex components to the secluded environment of the stereocilia tip.

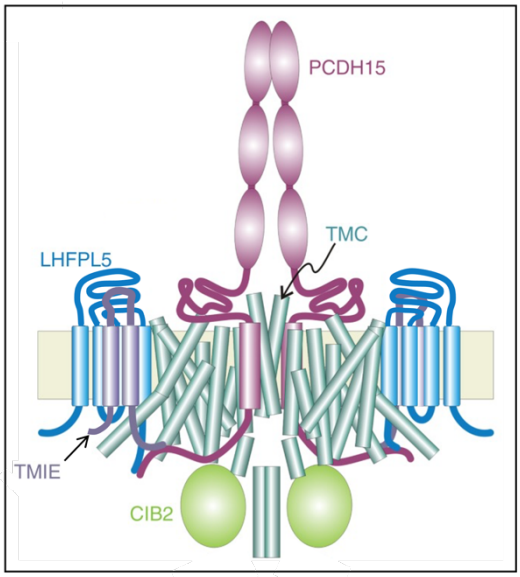


Figure 9 - Model of the molecular assembly associated to the MET channel. Modified from Corey et al. (2019)[376].

Further maintenance of TMC1 at the lower tip-link anchoring site requires co-expression of the essential channel subunits TMIE (TransMembrane Inner Ear) protein[59,60]

and CIB2 (Ca^{2+} and Integrin Binding 2) protein[61–63]. CIB2 is a 187 residues protein containing three EF-hand Ca^{2+} binding domains. CIB2 interacts with TMC1 at the tip-link lower anchoring site, as well as other scaffolding proteins such as whirlin and integrins (protein promoting the interaction with the extracellular matrix) in diverse location of the stereocilia[62]. Knock-out of CIB2 abolishes MET currents in cochlear hair cells and induces late defects in hair bundle morphogenesis and ultimately hair bundle regression[62,64]. This reflects a role both in mechanotransduction and hair cell survival. Interestingly, CIB2 is also a negative regulator of the apoptosis signal-regulating kinase-1 (ASK1)[65]. CIB proteins stability relies on Ca^{2+} binding[66], while MET currents are the main source of hair cell intracellular calcium ions[67]. CIB2 might therefore act as a control of MET current generation and promote the degradation of non-functioning hair cells[68], so that new hair cells can be generated[69–72]. The ability of hair cells to regenerate has, however, been lost in mammals, while other vertebrates can replace hair cells damaged by injuries, diseases or natural aging.

The other protein involved in TMC1 maintenance, TMIE, is a 156 residues two-pass transmembrane domain. Its N-terminal region and first transmembrane helix are involved in MET channel gating, suggesting a role in force transmission onto the channel[60], but more data is required to validate this hypothesis. The second transmembrane helix and the cytoplasmic region of TMIE are involved in TMC1 binding, as well as channel unitary conductance and ion selectivity[60]. Its C-terminal cytoplasmic tail (78 residues in human) is heavily charged, encompassing 41 cationic or anionic residues (53%). TMIE cytoplasmic tail also mediates Phosphatidylinositol-4,5-bisphosphate (PIP_2) binding, which are enriched in the membrane of hair cell stereocilia and are required for mechanotransduction[60,73]. Lastly, TMIE can interact directly with PCDH15-CD2 cytoplasmic domain by its cytoplasmic tail, as well as indirectly through the transmembrane protein LHFPL5. LHFPL5 is a 219 residues protein with 4 transmembrane segments that interacts with PCDH15[74], TMC1[75] and TMIE[74]. The function of this protein is still unclear, but the multiple interactions reported thus far suggest a role in the stabilization of the transmembrane machinery associated to the MET channel.

Altogether, these proteins form the MET channel machinery responsible for the entry of cations inside the cell during positive tilt of the hair bundle. As previously mentioned, hair cells must be able to carry hyperpolarization at their characteristic frequency. To this end, physical properties associated to the MET channels vary along the cochlea. Notably, MET channels conductance is higher at the base of the cochlea and decreases towards the apex[76,77]. This allows for the faster depolarization of high characteristic frequency cells in response to positive tilt of the hair bundle.

I.4.c – The Upper Tip-Link Density (UTLD)

At the upper anchoring site of the tip-link, CDH23 interacts with multiple cytoplasmic partners, forming the upper tip-link density. The upper tip-link density has been described as responsible for tuning the resting tension of the tip-link[78]. Adequate tip-link tension is critical for tuning hair cell sensitivity to sound-induced vibrations. A loose tip-link does not relay small increases in tension during stimulation of the hair bundle to the MET channel, effectively rendering the hair cell insensitive to low amplitude stimulation. On the other hand, a tip-link with excessive resting tension maintains the MET channel in an overly opened conformation, resulting in easily saturated ion influx and an incapacity to distinguish low and high amplitude stimulations.

At the upper anchoring site, CDH23 interacts with the scaffolding protein harmonin[79-82], myosin7a and membrane phospholipids (PIP_2)[83]. Harmonin is homologous to the cytoplasmic protein whirlin and has an overall similar architecture. The harmonin isoform b, expressed in the sensory hair cells of the cochlea[84], also contains three PDZ domains, one HHD domain, but lacks the second HHD domain found in whirlin. At the upper anchoring site of the tip-link, it mediates the assembly of a protein network involving CDH23, SANS and myosin7a[85,86]. SANS is also a scaffolding protein, with three ankyrin repeats of unknown function and a SAM domain. Together, harmonin, SANS and myosin7a form a complex thought to anchor CDH23 to the actin core of the stereocilia. The network intricacy, however, remains elusive. The complex assembles by multiple transient and redundant interactions (Figure 10). CDH23 interacts with both harmonin HHD domain and two N-terminal PDZ domains[79], as well as directly to myosin7a tail[83]. Harmonin in turn associates with the SANS protein by its HHD and first PDZ domain[87], and also directly with myosin7a by its third PDZ domain[86]. Again, SANS interacts with myosin7a[86], suggesting a direct interaction of the motor protein with all three other partners. Finally, SANS has been reported to form homomeric structures[85] and harmonin is also able to auto-associate, potentially allowing oligomerization[86].

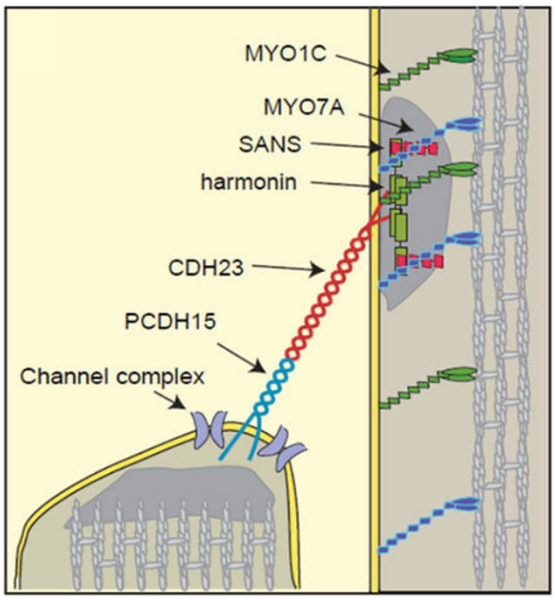


Figure 10 - Scheme of the upper tip-link anchoring site. From Dionne et al. (2018)[377].

The multiplicity of interactions induces the formation of condensates by phase-separation with high local protein concentrations[88]. Despite the micromolar affinities displayed by most of the above-mentioned protein-protein interactions, this suggests a stable anchoring of PCDH23 to the actin filaments of the stereocilia.

The use of multiple and redundant interactions of different affinities for tip-link anchoring, compared to stable high-affinity ones, is unclear. Possibly, this allows for a more dynamic protein assembly, with physical properties that can be tuned by differential protein expression. This hypothesis is supported by the expression of several isoforms of CDH23, harmonin and myosin7a in the cochlea[78–80].

I.4.d – Adaptation

Hair cells response to local vibrations is not constant over time. MET channels conductance decays during sustained stimulation, a phenomenon called adaptation[89,90]. Adaptation is thought to continuously maintain MET channels in a sensitive operating point and prevent MET saturation[90]. Two adaptation rates have been observed in hair cells, referred to as fast and slow adaptations and likely resulting of different molecular mechanisms (Figure 11).

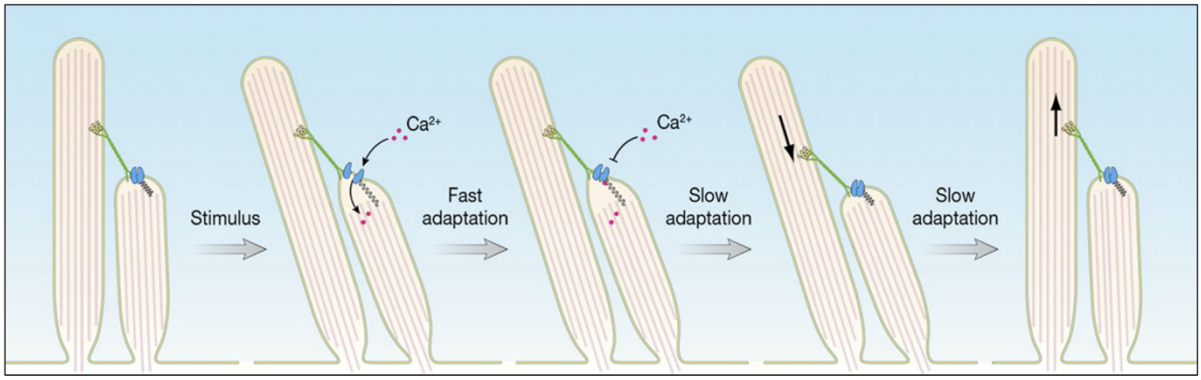


Figure 11 - Step by step adaptation of hair cells mechanotransduction machinery, regulating tip-link tension in response to a maintained stimulation. Modified from Schwander et al. (2010)[91].

Fast adaptation occurs in a millisecond time frame after hair bundle deflection[92]. Such adaptation rates are likely due to a direct effect of entering Ca^{2+} on proteins of the MET channels[93], resulting in an increased resistance to the force-driven opening. Therefore, MET channel closure rates should directly depend on conductance. As a consequence, cells of high specific frequencies, with larger MET channels conductance, should perform fast adaptation at higher rates than low specific frequency cells. To my knowledge, such difference has not been reported thus far. Interestingly, resting tension of the tip-link varies along the axis of the cochlea, from a higher tension at the base to a lower tension at the apex[53]. This gradient is

proportional to the decrease in the characteristic frequency of the hair cells along the organ[53]. It is hypothesized[53] that the increased tension applied on the tip-link at rest for cells of high characteristic frequency compensates for their high conductance[93] and thus faster-closing[76,77] MET channels. This would result in fast-depolarizing cells capable of responding to high-frequency stimulations, while maintaining MET channels in the adequate response range for optimal sensitivity.

Slow adaptation occurs in the tenths of millisecond time frame. It is associated with the downward motion of the tip-link upper anchoring site, effectively reducing tension during extensive hair bundle stimulation. Myosin motors have been suggested to promote the slip of the tip-link upper anchoring site along the actin filaments of the stereocilia. Specifically, the inhibition of myosin1c results in a loss of slow adaptation in hair cells[94]. Myosin1c has been identified in several locations of the stereocilia, notably at the upper tip-link anchoring site (Figure 10). This localization requires the expression of CDH23 in the stereocilia, suggesting coupling between the two proteins[95]. Myosin1c is composed of an N-terminal motor head, three IQ domains, a post-IQ region and a C-terminal PH (Pleckstrin Homology) domain (Figures 12A). It adopts an elongated conformation, with the motor region promoting force generation on the actin filaments on one end and the PH domain mediating membrane interaction on the other end (Figures 12B). This conformation is favoured by a rigid organization of its IQ domains in interaction with three Ca^{2+}-free CaM (Calcium-Modulated) proteins[95,96]. The rigid CaM-bound state of myosin1c is associated to a high duty-ratio of the motor head[96]. At higher intracellular concentrations of calcium, CaM proteins adopt their Ca^{2+}-bound state and detach from myosin1c. This results in an increased flexibility of the protein and a decrease of its motor head load and duty ratio[96]. Therefore, slow adaptation would arise from a decrease in myosin1c activity upon entry of Ca^{2+} in the cell, allowing the tip-link upper anchoring site to slide down

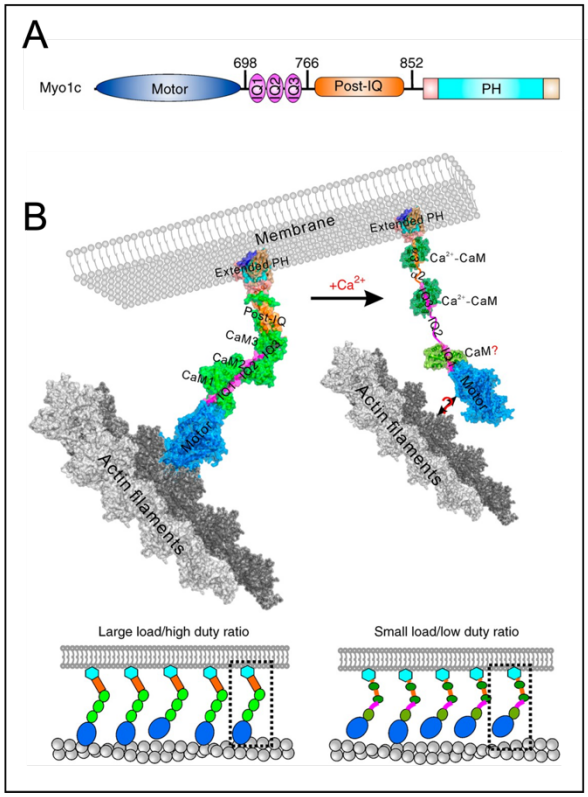


Figure 12 - A) Schematic organization of the myosin1c protein. B) Model for the calcium dependent actin decoupling of myosin1c underlying slow adaptation. Modified from Lu et al. (2015)[96].

the actin filaments of the stereocilia. In this model, myosin1c continuously promotes tensioning of the tip-link by climbing up the stereocilia and would also contribute to tuning the resting tension of the tip-link. Interestingly, several isoforms of myosin1c with different pre-motor N-terminal sequences exhibit diverse motor properties[97] and could tune resting tension and slow adaptation rates along the cochlea. The expression pattern of these isoforms, however, has not yet been documented.

I.4.e – Ion regulation

Mechanoelectrical transduction relies on the controlled ion balance between the extracellular medium of the hair bundle, the endolymph, and the intracellular medium of the hair cell. The MET mechanisms presented thus far induce an influx of K^+ and Ca^{2+} in the intracellular medium. Several channels and pumps then correct the cell membrane polarization by redistributing these ions in the extracellular medium.

Potassium ions are largely redistributed by channels located at the basolateral membrane of hair cells. Passive transport of K^+ towards the extracellular medium is rendered possible by the composition of the cortilymph, which itself contains little potassium (similar to the perilymph, few millimolar)[98,99]. It is mediated by several types of potassium channels, notably the Kir2.1 family member[100], the voltage dependant KCNQ4 channel[101] and the previously mentioned Ca^{2+}-activated K^+ (BK) channels. The Kir2.1 channels contribute to maintain the basal intracellular K^+ concentration and membrane potential. KCNQ4 and BK channels also contribute to establishing resting membrane potential, but achieve highest conductance in depolarized cells and high intracellular Ca^{2+} concentrations, respectively. Therefore, they mainly promote the return to negative membrane potentials after activation of the MET channels, subsequently maintained by Kir2.1 channels.

The Ca^{2+} influx in the soma of hair cells induced by opening of the MET channels is rather limited compared to potassium. This is due to the composition of the *scala media* in which hair bundle are bathed. The *scala media* contains endolymph, which has a much higher K^+ concentration (about 150mM) than Ca^{2+} concentration (about 0.02mM)[67]. Despite MET channels producing the main source of Ca^{2+} by passive influx at rest, they are not solely responsible for the larger increase in intracellular calcium observed during stimulation of the hair cells. Instead, the entry of large K^+ concentration and membrane depolarization resulting from mechanotransduction induce the opening of Cav1.3 Ca^{2+} channels at the basolateral membrane of hair cells[102]. This triggers calcium ions to flow from the surrounding perilymph,

with a Ca^{2+} concentration of about 1.2 mM. This is mostly occurring in IHCs, where calcium channels cluster at the presynaptic densities[103]. The regulation of intracellular Ca^{2+} concentration then depends on three mechanisms. First, Cav1.3 Ca^{2+} voltage dependent channels are inhibited by calcium[104]. Calcium influx from the cortilymph is therefore intrinsically limited. Secondly, intracellular structures, such as mitochondria, are capable to quickly intake large quantities of calcium and provide a calcium buffering property to the cell soma[67,92]. Thirdly, PMCA (Plasma Membrane Ca^{2+} ATPase) pumps use ATP to translocate calcium ions back to the cortilymph[105,106]. Active transport is necessary to pump calcium out of the cell, as Ca^{2+} concentrations are higher in all surrounding extracellular compartments. Several PMCA isoforms are expressed in hair cells, notably the PMCA2 pump in the stereocilia[105,106] and PCMA1 pump in the basolateral membrane[105].

I.4.f – The ribbon synapse

Inner hair cells are capable of quickly releasing neurotransmitters at afferent neuron synapses during excitation and continuously responding to sustained stimuli[107]. To this end, IHCs possess specialized presynaptic structures, the ribbon synapses (Figure 13). Ribbon synapses are electron-dense structures which, in cochlear hair cells, appear to sit directly on the presynaptic membrane. The core of this structure is thought to primarily be composed of the ribeye protein[108] in interaction with cytomatrix proteins such as piccolino[109] and bassoon[110], but its molecular organization is yet to be characterized. Vesicles are found tethered around the ribbon synapse, with a population docked at the membrane ready for release[110]. These vesicles are typically filled with glutamate, thanks to the VGLUT3 transporter[111]. Positioning of the vesicles would rely on the vesicular protein otoferlin, interacting with myosin6 at the ribbon synapse surface[112–114].

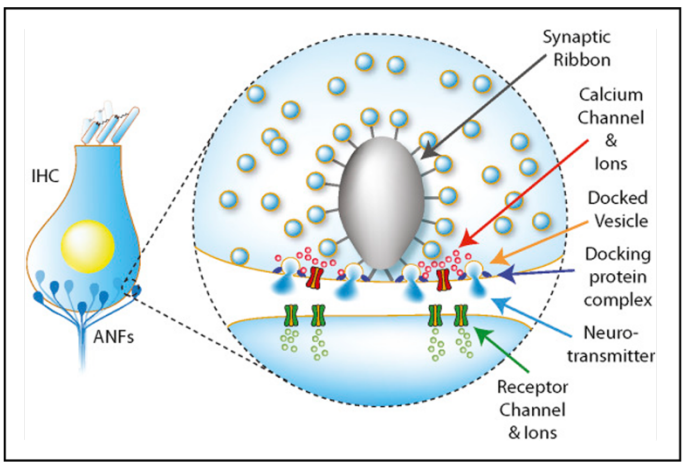


Figure 13 - Inner hair cell ribbon synapse. Glutamate-filled vesicles are docked at the presynaptic membrane. Fusion is triggered by an influx of calcium, resulting in release of neurostransmitter in the synaptic cleft. From Bruce et al. (2018)[378].

Fusion of these vesicles with the plasma membrane is linearly related to the intracellular calcium concentration[114,115]. It is triggered during depolarization of the hair cell membrane in the time of sound-evoked mechanotransduction, notably by the opening of the voltage-dependent Cav1.3 channels[116] found clustered at the ribbon synapse by interaction with otoferlin[112]. Physical properties of the calcium channels would participate in hair cell phase-locking, which refers to vesicular fusion triggered in phase with hair bundle deflection, critical for transducing the stimuli with temporal fidelity[116]. Calcium dependent vesicle fusion is a rather common mechanism in neurotransmitter release (Figure 14). In this model, vesicles are docked near the plasma membrane by the interaction between complexin and SNARE proteins. Calcium-bound synaptotagmin then displaces complexin from the SNARE complex and promotes vesicle fusion[117]. In IHCs, however, involvement of SNARE proteins in vesicle fusion is debated[118]. Moreover, synaptotagmin IV, expressed in mature IHCs, does not have calcium binding sites, but is still required for linear release of neurotransmitter depending on intracellular Ca^{2+} concentration[119]. Otoferlin was also suggested to contribute to calcium dependent vesicle fusion[120], but the underlying mechanisms remain elusive. More data is necessary to unveil the molecular processes regulating vesicle fusion at the ribbon synapse of auditory hair cells.

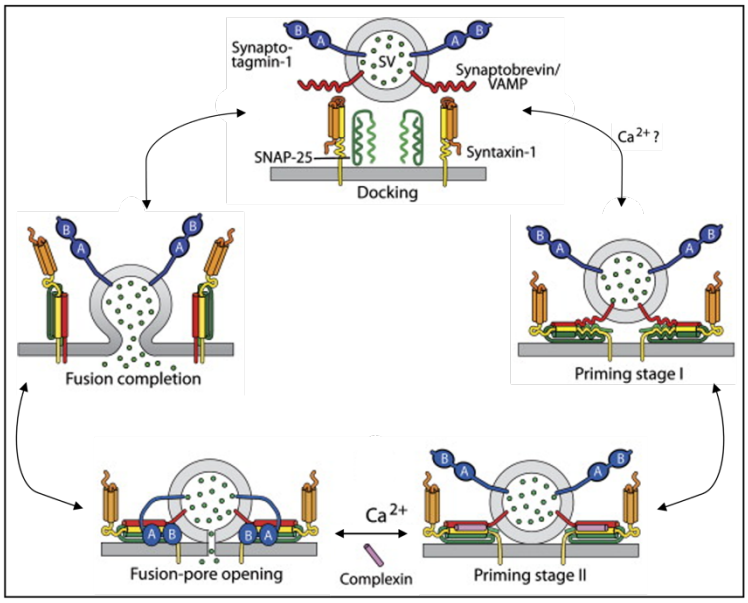


Figure 14 - Model for calcium-regulated vesicle fusion mediated by synaptobrevin and complexin proteins. Modified from Tang et al. (2006)[117].

After vesicle fusion, glutamate is released at the synaptic cleft and activates AMPA-type receptors (AMPARs) at the nerve afferent buttons[121,122]. Opening of AMPARs induces depolarization of the nearby heminode, opening of voltage-gated Na+ channels and generation of a spike propagated to the spiral ganglion neuron (SGN). Each hair cell connects to multiple afferent nerve fibbers, from one (or few) SGN, by one ribbon synapse each. These ribbon synapses have different Ca^{2+} dependence[123]. In high characteristic frequency cells, it allows to desynchronise spikes transmitted to the SGN by different nerve fibbers. This results in a better representation of the stimulus by triggering a response at almost each cycle of a sound wave

of shorter period than individual nerve refractory period (usually in the millisecond time frame). For low characteristic frequency cells (<1kHz), nerve fibber stimulation can occur at almost every cycle of the sound wave and a higher degree of synchrony is observed between several nerve fibbers[124].

I.5 – Deafness in human

In the previous chapters, I have presented the main steps of the hearing process, from the capture of sound-evoked vibrations by the outer ear to the release of neurotransmitter at the ribbon synapse. Any disruption of this process, or of the processing of the produced signal by the brain, may lead to deafness. Deafness corresponds to a decrease in the perceived intensity of sounds. It is usually defined as a hearing loss greater than 20dB at the worst hearing frequency. The severity is then classified depending on the loss of the perceived intensity, referred to as mild (20-35dB loss), moderate (35-50dB loss), moderately severe (50-65dB), severe (65-80dB loss), profound (80-95dB loss) or complete (>95dB loss). Deafness is considered debilitating for a hearing loss above 30dB, where the patient may have difficulties hearing conversational speech. For these people, hearing aid is usually needed, with unaddressed cases being associated with a lower quality of life and increased risks of isolation, depression and dementia[6–8]. In addition to the medical and psychological consequences posed at the individual level, unaddressed hearing loss costs about 980 billion dollars annually worldwide[9]. These costs come from health care fees (314 billion dollars), specialized educational support (27 billion dollars), unemployment and premature retirement (182.5 billion dollars) and an additional 456.5 billion dollars resulting from social friction, including communication difficulties and stigma. These estimates highlight the economic necessity to address hearing loss.

I.5.a – Types and prevalence

Deafness is the most common sensory deficit in human, affecting about 20% of the total population (Figure 15). If most people suffer from mild hearing loss, about 430 million people (>5.5% of the population) have moderate to complete hearing loss[9].

The variability in hearing loss severity and onset reflects its multifactorial origins. For good prognosis and adequate treatment options, it is important to identify the type of loss (conductive, sensorineural or mixed), time of onset (congenital or acquired), stability (stable or progressive) and causality (hereditary or environmental).

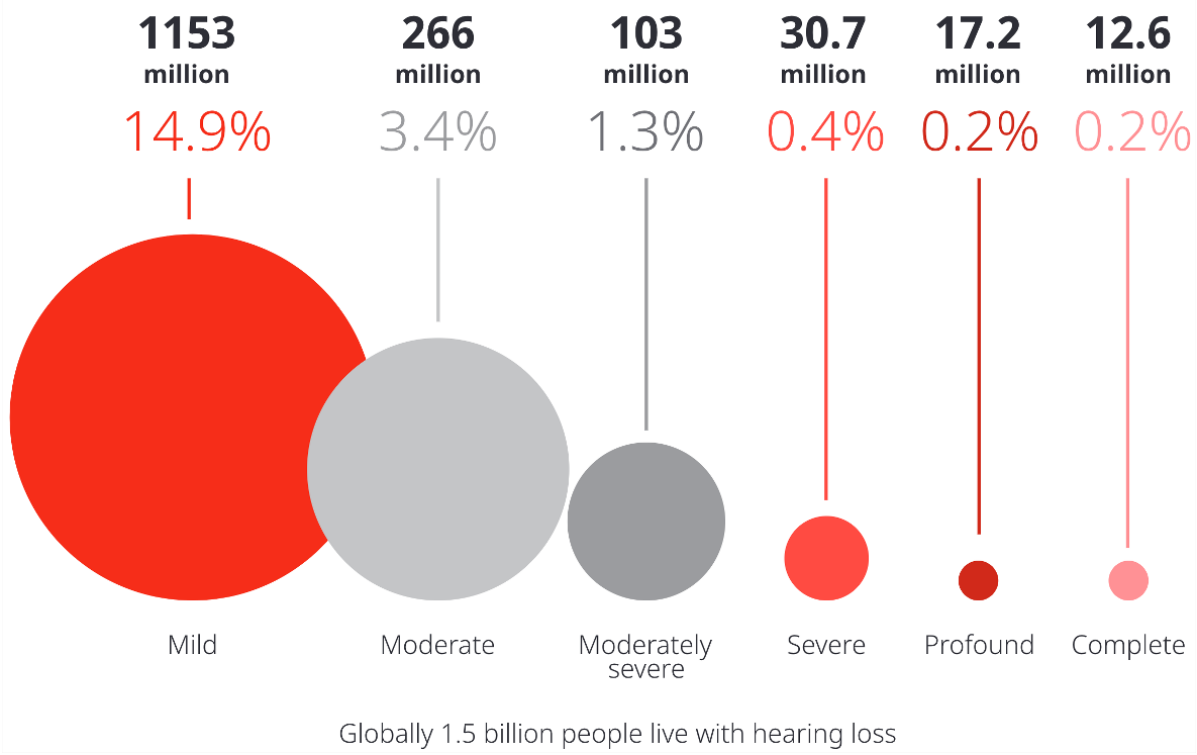


Figure 15 – Prevalence of hearing loss. From World Health Organization (WHO) World Report on Hearing (2021)[9].

Conductive deafness refers to defects of the outer and middle ears and is often the consequence of a malformation, a trauma or an infection. It consists in an impairment of the mechanical relay of the sound to the cochlea. The induced hearing loss is usually mild to moderate and primarily affects low frequencies. Common manifestations include cerumen impaction (blocking the ear canal), tympanic perforation (mechanical or due to an infection) or bone dislocation (usually trauma).

Sensorineural deafness refers to defects in the detection of the sound wave by the inner ear, an incapacity to relay the signal to the brain or the inability of the auditory cortex to process the information. As such, it results from damages of the cochlear structure, hair cell and/or hair bundle, auditory nerve or auditory cortex. Causes are again very varied, including genetic factors, infections, trauma, environmental factors (smoking, nutritional deficiencies, ototoxic medicines, exposure to loud noises…) and aging. The variability in the underlying causes also results in a wide diversity in severity, from mild to complete hearing loss, as well as age of onset, from congenital to age-related.

Determination of the age of onset and progressive or stable nature of the hearing loss is critical. Unattended deafness before speech and language acquisition is strongly associated to lower communication skills and impaired cognitive and psychosocial development[125], as well as later isolation. In developed countries, 3 newborns in 4000 (1.33‰) are estimated to suffer from a hearing loss greater than 40dB, an estimate that doubles (2.7‰) at five years

old[126,127]. Amongst neonates with sensorineural hearing loss, at least half of cases are inherited while the other half results from environmental causes[128] (Figure 16). Non-genetic hearing loss often results from prenatal infections, the most common being by cytomegaloviruses, lymphocytic choriomeningitis virus, rubella virus, *Toxoplasma gondii* or *Treponema pallidum*[128]. For hearing impairment of genetical origin, 35% result in non-syndromic hearing loss. The transmission mode varies, it can be recessive (DFNB, 28%), dominant (DFNA, 7%), X-linked or mitochondrial (<1%). More than 110 genes have been associated to non-syndromic hearing loss[129]. Surprisingly, a single gene (GJB2) is responsible for about 50% of all cases[130]. It encodes for the connexin-26 protein, expressed at the tight-junctions of inner ear cells where it forms connexons, channels notably involved in the regulation of the K^+ homeostasis of the cochlea[131].

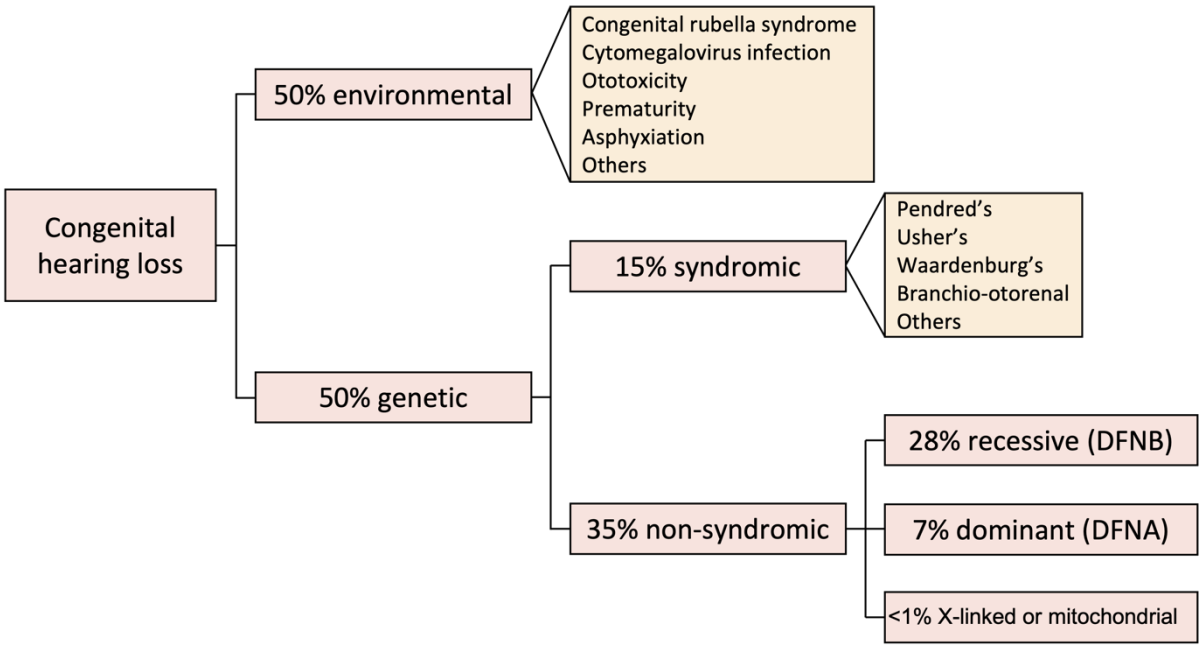


Figure 16 – Underlying causes for congenital hearing loss. Adapted from Smith et al. (2005)[128].

Syndromic hearing loss accounts for 15% of congenital deafness and associates other symptoms to the hearing impairment. These other symptoms vary widely depending on the affected genes, usually expressed in other tissues than those of the ear. Common types of sensorineural hearing loss syndromes include Pendred's, Usher's, Waardenber and Branchio-otorenal syndromes (Figure 17). A single gene can be associated to several syndromic or non-syndromic hearing loss, depending on the position of the mutation(s) and the type of substitution(s).

	Gene	Phenotype
Dominant		
Waardenberg (WS1)	*PAX3*	Major diagnostic criteria include dystopia canthorum; congenital hearing loss; heterochromic irises; white forelock; and an affected first-degree relative. About 60% of affected children have congenital hearing loss; in 90%, the loss is bilateral.
Waardenberg (WS2)	*MITF*, others	Major diagnostic criteria are as for WS1 but without dystopia canthorum. About 80% of affected children have congenital hearing loss; in 90%, the loss is bilateral.
Branchio-otorenal	*EYA1*	Diagnostic criteria include hearing loss (98%), preauricular pits (85%), and branchial (70%), renal (40%), and external-ear (30%) abnormalities. The hearing loss can be conductive, sensorineural, or mixed, and mild to profound in degree.
Recessive		
Pendred's	*SLC26A4*	Diagnostic criteria include sensorineural hearing loss that is congenital, non-progressive, and severe to profound in many cases, but can be late-onset and progressive; bilateral dilatation of the vestibular aqueduct with or without cochlear hypoplasia; and an abnormal perchlorate discharge test or goitre.
Usher (USH1)	*USH1A*, *MYO7A*, *USH1C*, *CDH23*, *USH1E*, *PCDH15*, *USH1G*	Diagnostic criteria include congenital, bilateral, and profound hearing loss, vestibular areflexia, and retinitis pigmentosa (commonly not diagnosed until tunnel vision and nyctalopia become severe enough to be noticeable).
Usher (USH2)	*USH2A*, *USH2B*, *USH2C*, others	Diagnostic criteria include mild to severe, congenital, bilateral hearing loss and retinitis pigmentosa; hearing loss may be perceived as progressing over time because speech perception decreases as diminishing vision interferes with subconscious lip reading.
Usher (USH3)	*USH3*	Diagnostic criteria include postlingual, progressive sensorineural hearing loss, late-onset retinitis pigmentosa, and variable impairment of vestibular function.

Figure 17 – The most common forms of syndromic hearing loss and their symptoms. From Smith et al. (2005)[128].

I.5.b – Palliative medicines and therapeutic approaches

Prevention, therapeutics and hearing restoration methods differ depending on the origin and severity of the hearing loss.

For few infection-related hearing impairment, the extent of the loss can be reduced by early and adequate treatment. For example, prenatal or early (6 first months after birth) antiviral therapy, as well as immunoglobulin therapy, can improve neurological development and hearing in children with symptomatic congenital cytomegalovirus infection[130,132,133]. However, limited data is available on treatment efficacy for most infections, such as *Toxoplasma gondii*, *Treponema pallidum,* or zika virus. When applicable, prophylactic vaccination is the main way to prevent congenital infection[134].

Once acquired, several devices exist to restore, even partially, the hearing loss[127]. The most commonly used is the conventional hearing aid (Figure 18A). It is suitable for patients with no or limited damages to the outer and middle ear, meaning that an air-relayed sound can be perceived, and with a hearing loss between 35 and 65dB. Conventional hearing aids usually detect and amplify sounds, reemitting them in the ear canal. These devices are usually flexible in use, with a tuneable intensity gain and the possibility to be easily removed[135].

For patients with disruption of the outer and middle ear sound relay (narrowing or absence of auditory canal, non-functioning ossicles…), bone-anchored hearing aid has proved efficient for rehabilitation of sensorineural hearing loss (Figure 18B)[136]. Instead of emitting air-borne vibrations to the ear canal, they bypass the outer and middle ear by relaying vibrations through bone conduction. To this end, an implant must be surgically inserted in the skull close to the ear. Sounds detected by the device are reemitted through the bone to directly stimulate the inner ear. This approach, however, relies on the insertion of the implant in the bone (osseo-integration) and cannot usually be performed on infants under six years old.

Conventional and bone-anchored hearing aid both require a functioning inner ear cochlea to function. For patients with defects of the cochlea or sensory organ of corti, a cochlear implant should be used (Figure 18C). Instead of reproducing sound vibrations, cochlear implants encode sounds into electrical impulses that directly stimulates the auditory nerve through an electrode array. Cochlear implants are also suited for patients suffering of hearing loss restricted to certain frequencies, resulting in a hybrid electro-acoustic stimulation[137]. The U.S. Food and Drug Administration (FDA) has recently approved the use of cochlear implants in children as early as 9 months of age.

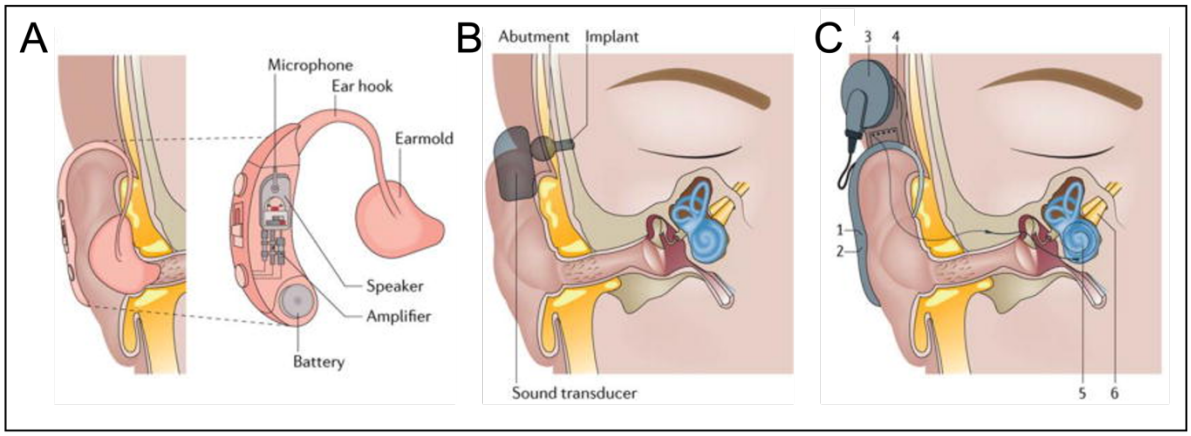


Figure 18 – Available devices for hearing loss restauration. A) The conventional hearing aid. B) The bone-anchored hearing aid. C) The cochlear implant. Modified from Korver et al. (2017)[127].

Therapeutic approaches to restore hearing loss are the focus of multiple investigations, but none is currently available to my knowledge. Amongst studied strategies, gene therapy appears to be a promising method for rescuing hearing loss of genetic origin. One of the underlying ideas is to correct the DNA sequence of a mutated gene or supplement the cells with the corresponding wild-type sequence. The first approach usually relies on the use of the molecular editing tool Crispr/CAS9 and has been successfully used in mice carrying a pathogenic variant of cadherin23[138]. One notable downside of this technique is the need to repair the DNA of all precursor cells expressing the gene, including hair cells but also (ideally) other corresponding cell lines in the case of syndromic hearing loss. Thus far, this is achieved by injecting the mouse zygote, which poses direct practical issues when treating organisms that were not purposefully generated with the defect. Furthermore, off-site mutations in the rest of the genome might occur and potentially cause other issues, while raising ethical questions for an application in human. The second approach often makes use of adeno-associated viruses (AAV) engineered to infect various cell-lines of an organ and insert a DNA fragment of choice in the host genome instead of a replicative provirus. Examples of successful

studies in *vivo* include mice defective for the genes encoding VGLUT3[139], GJB2[140], TMC1/2[141], whirlin[142] and CLRN1[143]. In these cases, AAV are injected in the cochlea of neonatal mice (~P0-P2). This allows to more easily time the procedure as well as restrict the genome modification to the neighbouring tissues (no modifications in other cell lines, including germinal). However, this solution is also limited in time by the expression patterns of the targeted gene and the permanent cochlear damages induced by several mutated genes. Later treatment is therefore rarely an option if damages have already occurred[140,144]. This approach might still be used for progressive symptoms, including hearing loss but also other associated syndromic manifestations[145].

Another research avenue to rescue numerous types of hearing loss is hair cell regeneration, by promoting differentiation of new and healthy hair cells. Again, two routes are currently being explored. Stem cell therapy is the first strategy and consists in delivering undifferentiated cells to the cochlea, which would integrate the tissue and make new hair or supporting cells. If stem cells are used in the field of regenerative medicine, their application in cochlear regeneration seems arduous thus far[146]. This would be due to the peculiar composition of the cochlear medium, the cohesive nature of the tissue limiting stem cells insertion and the overall difficulty to differentiate into hair cells. Moreover, limited data are available *in vivo*, differentiation from stem cells into hair cells being mostly studied in *in vitro* 3D cultures (see Lee M. *et al.*, 2018 for review[146]). The second strategy would be to stimulate the differentiation of cochlear supporting cells into hair cells. Such mechanism has been lost in mammals, but spontaneously occurs in several other vertebrates[147,148]. These approaches rely on the differential gene expression between supporting cells and hair cells. Notably, hair cells express the Math1 (or Atoh1) gene, required for hair cell differentiation[72]. Supporting cells, however, inhibit Math1 expression and express the cyclin-dependent kinase inhibitor $p27^{Kip1}$ (or Cdkn1b), reducing their proliferative capacity[69]. Multiple approaches have thus been considered to promote supporting cell differentiation into hair cells, such as the overexpression of Math1[70,71], the expression of transcription factors specific to hair cells[149] or the suppression of supporting cells Notch effectors, in turn promoting (likely amongst other) expression of Math1[150]. Several examples of these strategies induced the differentiation of new hair cells both *in vitro* and *in vivo*, but little to no improvement in the hearing threshold of deaf animals was observed[71,151]. Even if induced hair cells may display similar transcriptome to juvenile hair cells and produce voltage-dependent currents[149], poor ribbon synapse formation and innervation by afferent nerves is observed in treated animals[151]. Ribbon synapse formation and connection to developing nerve fibbers are several-steps processes

starting around birth in mouse and are promoted by multiple signalling events[152]. The lack of synaptic connections observed in hair cells differentiated by gene therapy suggest that they cannot recapitulate synaptic development, either because expression (or repression) of other genes should be induced in these hair cells or because additional regulating factors are carried out on the nerve end. More data is required to identify additional targets triggering nerve growth and hair cell interfacing. Lastly, hair cell regeneration would be suited for restoring hearing loss resulting from environmental causes or aging. For hearing loss induced by pathogenic variants of key auditory genes, later differentiation of new hair cells should generally prove inefficient if their genetic background remains altered. A combined use of new hair cell differentiation and gene replacement or editing, as presented in the previous paragraph, should therefore be used. However, the complexity of the approach would dramatically increase.

I.5.c – The Usher syndrome

Amongst genetic disorders inducing hearing loss, I will here provide more details about the Usher syndrome, which will bring the focus on the main topic of my thesis.

This syndrome takes the name of Charles Usher, Scottish ophthalmologist who described it for the first time, based on a cohort of patients, in the early 20th century. The Usher syndrome then refers to a combination of congenital bilateral hearing loss, balance defects and progressive blindness due to *retinitis pigmentosa*. *Retinitis pigmentosa* corresponds to a progressive degeneration of rods and/or cones of the retina, resulting first in night blindness, a progressive constriction of the visual field and ultimately total blindness[153].

This rare disease is the most common genetic cause of combined deafness and blindness in human. The Usher syndrome is responsible for 3 to 6% of congenital hearing loss, for about 400,000 people worldwide[154]. It regroups an ensemble of autosomal recessive disorders, further categorized within one of three subtypes[155] (Figure 19). These subtypes (Usher type 1, 2 or 3) are defined clinically based on the severity of the symptoms and the age of onset. Usher type 1 is characterized by a profound deafness at birth, vestibular dysfunctions and the onset of progressive *retinitis pigmentosa* before puberty. Usher type 2 patients exhibit mild to moderate congenital hearing impairment, delayed onset of *retinitis pigmentosa* compared to type 1 (during puberty) and usually no vestibular dysfunction. Finally, patients with Usher type 3 have progressive hearing loss during childhood, further delayed *retinitis pigmentosa* starting during adulthood and mild or no vestibular defects. The clinical

heterogeneity of the three Usher subtypes has a molecular basis, with mutated genes encoding for different proteins. These genes have been identified by genetics on patients suffering from Usher syndrome and helped define the proteome associated to hearing.

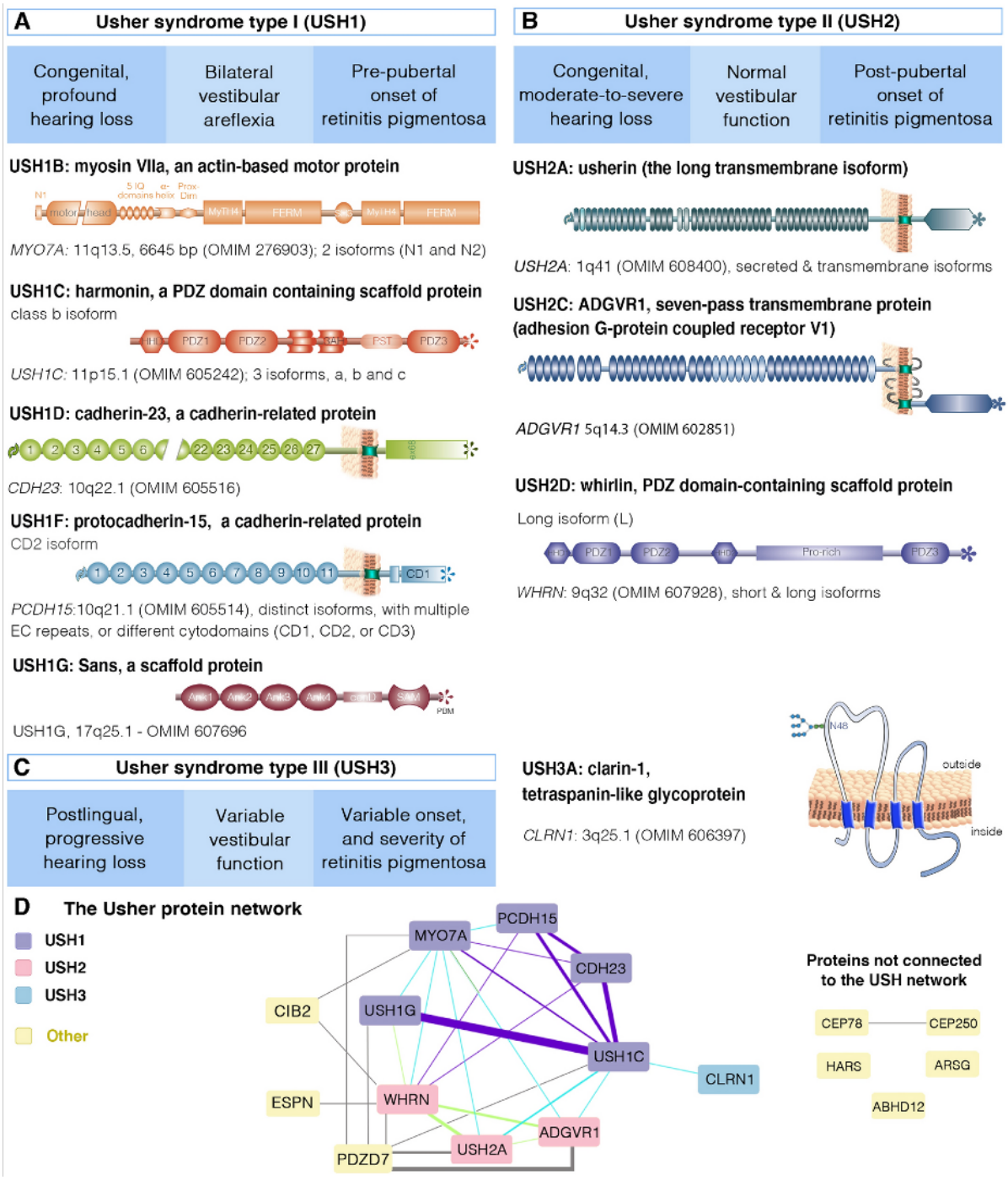


Figure 19 – The Usher syndrome subtypes, clinical manifestations and associated proteins. From Delmaghani et al. (2022)[155].

The Usher syndrome of type 1 (Usher 1) is the most severe subtype, mostly involving proteins directly responsible for the detection of local vibrations at all stages of the development. Affected genes often encode for proteins associated to the MET channel machinery, such as the scaffold protein harmonin and associated partners (SANS), the molecular motor myosin7a, as well as proteins of the tip-link (protocadherin-15, cadherin-23)[156]. The molecular assembly and variant's physiopathogenesis of these proteins have been the focus of extensive studies, from the genetics to the molecular level, even though key information is still missing (channel assembly, mechanisms for mechanical coupling, transport to the correct localization, …).

The Usher syndrome type 2 (Usher 2) is the most prevalent Usher subtype, accounting for more than 50% of all patients in Europe and North America[154]. Affected genes transiently encode for proteins thought critical for hair bundle differentiation and morphogenesis during the development of the cochlea. The associated proteins, usherin, ADGRV1 (Adhesion G protein-coupled receptor V1) and whirlin[156], take part in a same molecular complex at the base of hair cell stereocilia, the ankle link complex. This complex also involves PDZD7 (PDZ-containing protein 7), an Autosomal Recessive Non-Syndromic Hearing Loss (ARNSHL) causative also described as modifier for the Usher 2 phenotype. Despite being the most prevalent subtype, limited information is available concerning these four Usher 2 associated proteins in terms of function and protein-protein interactions. The study of the Usher 2 complex is the main topic of my PhD, with a focus on the interaction between PDZD7 and ADGRV1.

The Usher syndrome type 3 (Usher 3) is the least prevalent (2-4% of all patients), with the least severe yet most variable symptoms across the three subtypes. Two proteins have been associated to Usher 3, the translation factor HARS[157] (Histidyl-tRNA Synthetase) and the synaptic protein Clarin-1[158].

The molecular mechanisms underlying the physiopathogenesis of Usher syndrome appear complex, with different variants of a same protein being associated to syndromic or non-syndromic hearing loss. In the syndromic Usher-type hearing loss, variants impair the function of the affected proteins in the ear (deafness), but also in the vestibule (balance defects) and retina (blindness). The semi-circular canals and otolithic organs of the vestibule both possess hair cells with similar mechanisms of activation than the IHCs of cochlea. The stimuli, however, differ. Each semi-circular canals, sensing angular acceleration of the head, has an ampullae at its base, containing the sensory cells and the cupula (Figure 20A). When tilting the head, the gel-like cupula is displaced, inducing the tilt of hair cell bundles. In the otolithic

organs, sensing planar and vertical acceleration of the head, hair cells are covered by a gel on which lays calcium carbonate structures called otoliths. Linear acceleration of the head induces the relative movement of the otoliths, deformation of the gel-like environment and deflection of the cells hair bundles. Proteins expressed in the hair bundles of vestibular hair cells are partially the same as those expressed in the cochlea, resulting in a combined sensory defect when pathologically mutated. In the retina, isoforms of Usher 1 and Usher 2-associated proteins are notably expressed in the tubulin-based connecting cilium of rods and cones (Figure 20B)[159,160]. Expression of defective genes encoding for proteins of the connecting cilium would prevent the renewal of photosensitive molecules contained in the discs of the outer segment, resulting in the accumulation of cytotoxic products and eventually cellular death. These photoreceptor cells also signal to the brain through ribbon synapses, albeit different in morphology than those of the cochlea, and mutation of Usher 3-associated proteins might also affect their functioning.

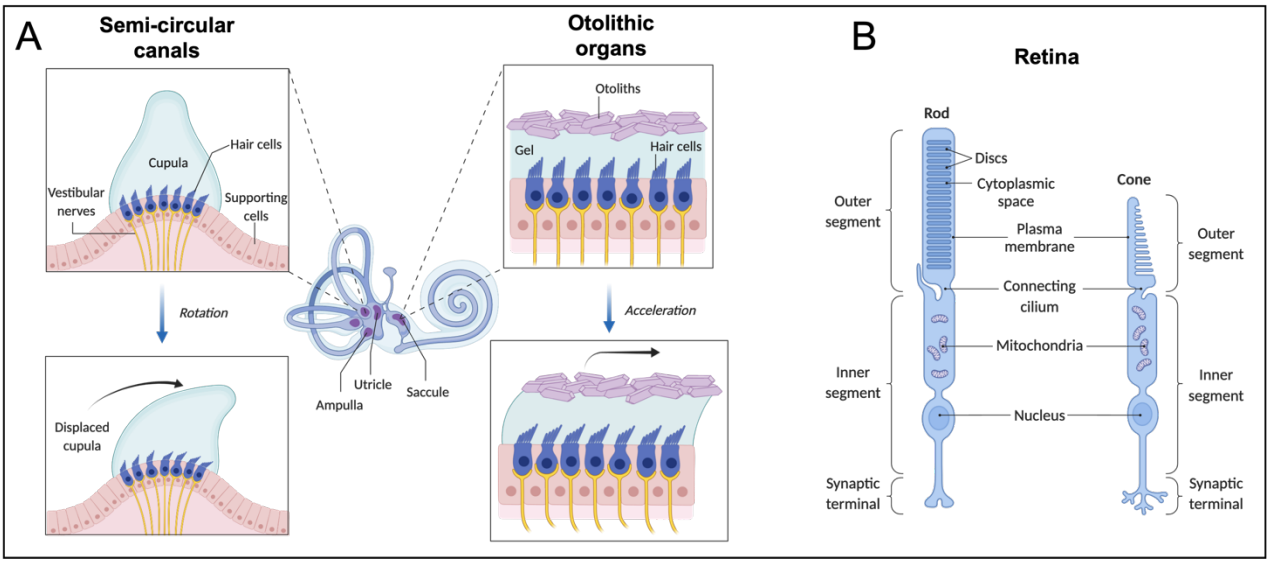


Figure 20 - Other sensory cells expressing Usher genes. A) Hair cells of the vestibule have hair bundles resembling those of cochlear hair cells and express Usher proteins in similar patterns. B) Photoreceptor cells of the retina express Usher proteins at the connecting cilium region. The general organization of the complexes and the function of Usher proteins in these cells is unclear. Made with BioRender.

Overall, the syndromic phenotype of Usher mutations is a consequence of the expression of the same proteins in cochlea, vestibule and retina. The variability in symptoms exhibited by the patients thus results from the location of the variant(s) in the related genes, corresponding to more or less severe mutations and potentially affecting only some of the isoforms, which are expressed differentially depending on the cell types.

II – The Usher 2 complex

Three proteins are associated to the syndrome Usher of type 2, namely usherin, ADGRV1 and whirlin. Together with the deafness protein and usher modifier PDZD7, they form a quaternary complex at the base of hair cell stereocilia[161,162] (Figure 21). Usherin and ADGRV1 are two transmembrane proteins with very large extracellular domains, ADGRV1 belonging to the adhesion GPCR protein family. These two proteins have cytoplasmic domains with a same C-terminal PDZ Binding Motif (PBM, -DTHL$_{COOH}$), interacting with the two homologous cytoplasmic proteins whirlin and PDZD7. These cytoplasmic proteins are thought to act as scaffolds and promote the interaction with other partners, notably actin-binding proteins such as myosins[34,163].

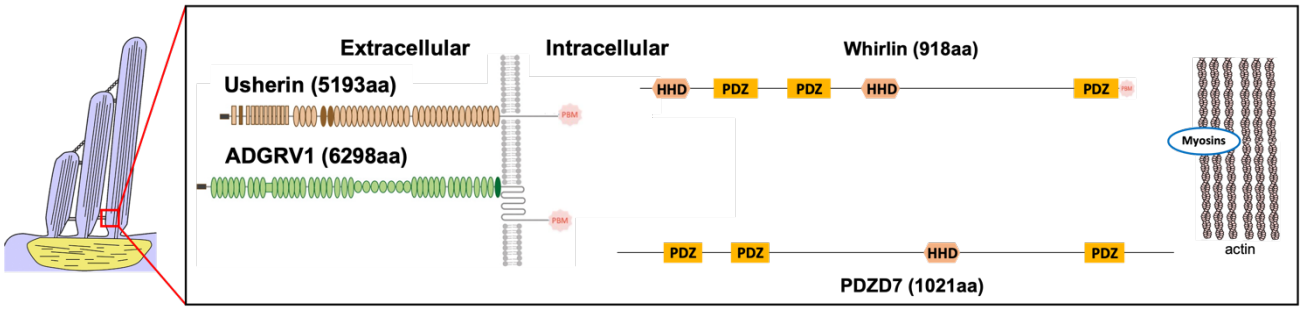


Figure 21 – Schematic organization of the Usher 2 complex. Indicated residue numbers correspond to the mouse proteins.

The large extracellular domains of usherin and ADGRV1 are thought to form the fibrous links observed transiently at the base of hair cell stereocilia (Figure 22A), reported in mice during the postnatal days (~P2 to P12)[164]. These links later disappear and Usher 2 proteins are thought to no longer localize at the ankle link region. Knock-out of any of the four proteins results in aberrant hair bundle morphology (Figure 22B)[165], linking this complex to late hair bundle morphogenesis. However, hair cells of neonatal mice (P0-P2) already exhibit hair bundles with staircase organization of the stereocilia[164]. The function of the Usher 2-associated proteins is therefore unclear, possibly contributing to maintain bundle architecture (scaffold) and/or mediate signalling for gene expression switches in the maturing cell (GPCR activity).

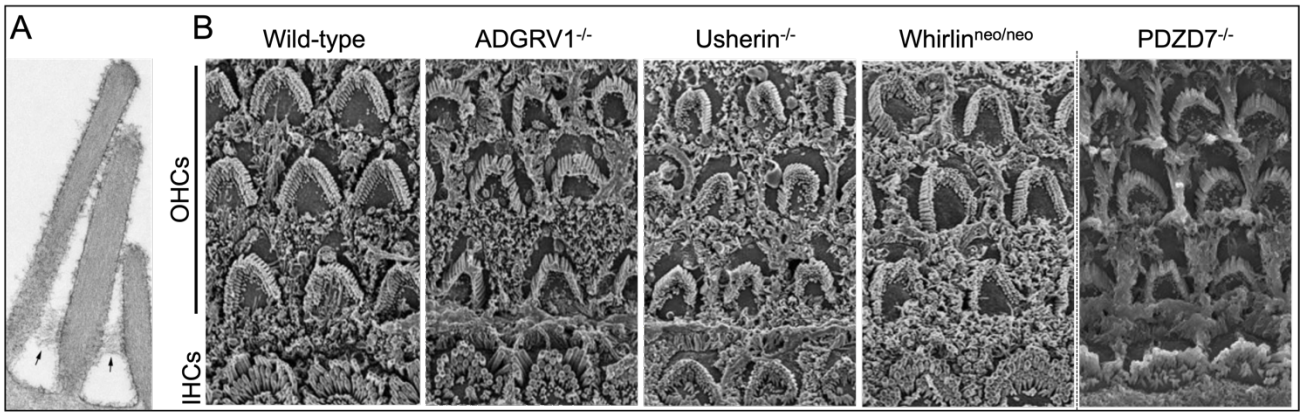


Figure 22 – A) Transmission electron microscopy image of three rows of hair cells stereocilia. The fibrous ankle-links are indicated by black arrows. B) Morphology of healthy wild-type hair bundles compared to hearing impaired mouse with Usher 2 proteins knock-outs. Modified from Goodyear et al. (2005)[164], Zou et al. (2015)[165] and Zou et al. (2014)[161].

II.1 – The extracellular ankle-links

The individual contributions of usherin and ADGRV1 extracellular regions to the fibrous ankle links observed by scanning electron microscopy are unclear. Specifically, the proportion of each protein in the links and the directionality of each molecule with respect to the inter-stereocilia space remain to be described. Thanks to the length of their extracellular regions, the two proteins may theoretically cross the entire space between adjacent stereocilia (about 100nm). However, there is currently no evidence in the literature that this is the case. It is thus also possible that multiple facing molecules form handshake-like structures, as observed for the tip-link. In such case, the links could consist of ADGRV1/ADGRV1 or usherin/usherin homomeric structures, as well as ADGRV1/usherin heteromers. The study of the ankle-links composition and organization is relevant for hair bundle organization, as it may harbour cellular mechanisms underlying bundle polarity. As an example, heteromeric ankle links could help discriminate stereocilia rows by asymmetric expression of the two proteins.

II.1.c – Usherin

Usherin is a single-pass transmembrane protein with two described isoforms[166]. The full-length protein has a large extracellular region (5011 residues) and a cytoplasmic domain (139 residues) exhibiting a C-terminal PBM. Usherin is expressed in the basement membrane of several tissues, including the ear, retina, testis, ovary, intestine and colon[167]. Basement membranes are extracellular matrices responsible for the attachment of epithelial tissues to the

supporting connective tissues. Its extracellular region encompasses 13 laminin-types domains and 33 fibronectin type-III domains. (Figure 23).

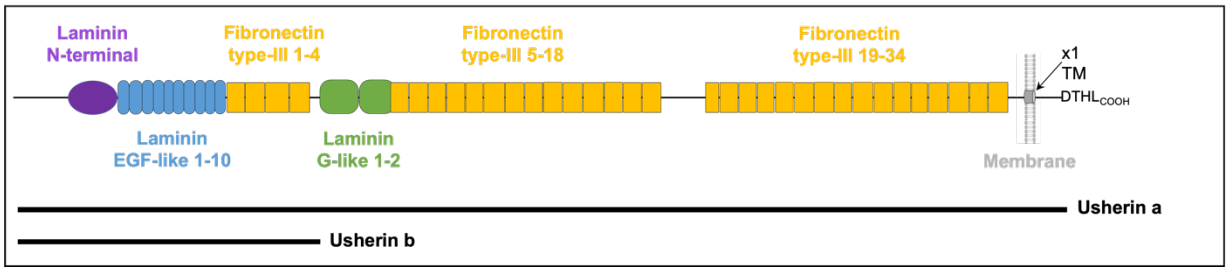


Figure 23 – Schematic representation of the usherin protein from N- to C-terminal. Domain delimitations are based on the UniprotKB entry O75445. Isoforms are indicated in black.

Fibronectin type-III domains consist of about 100 residues folded as a 7 strands β-sandwich. These domains were first identified in the fibronectin protein, an extracellular matrix component involved in cell adhesion, growth, migration and differentiation. Deletion of fibronectin type-III domains from the fibronectin protein can affect cell adhesion and migration during development[168]. The function of these domains in usherin are to be determined, but the large number of fibronectin type-III repeats in its extracellular region suggests a role in extracellular scaffolding required for hair cell development. Usherin extracellular region also contains one laminin domain VI (or N-terminal laminin domain or LN domain), 10 laminin EGF-like domains and two laminin G-like domains. The types of laminin domains are defined by homology to the various domains of the laminin proteins. All laminin domains mediate protein-protein interactions, but differ in their fold and binding partners. The laminin domain VI, first folded domain predicted at the N-terminus of the protein, consists of about 247 residues with a mixed α/β fold. In laminin proteins, the N-terminal domain (LN) can promote homo- and hetero-oligomerization with other LN domains[169]. The laminin N-terminal domain of usherin could therefore participate in the cohesive assembly of the fibrous ankle links. Laminin EGF-like domains are short β motifs of about 60 residues sharing structural homology with part of the extracellular region of the Epidermal Growth Factor Receptor (EGFR)[170]. EGF-like motifs have been shown to bind to the extracellular matrix component nidogen 1/2[171,172], recently identified as gene associated with self-reported adult hearing difficulties[173]. This reflects the expression of usherin in basement membranes, of which nidogen is a constituent[167]. In a similar fashion, the laminin G-like domains of usherin are about 180 residues long and adopt a β fold and promote the interaction with dystroglycan[174], another component of basement membranes. Interestingly,

both nidogen and dystroglycan are Ca^{2+} dependent cell-matrix adhesion proteins, involved in signalling events of basement membranes.

Usherin having laminin-like protein-protein interaction domains is consistent with its expression pattern in the basement membrane of various tissues. However, the function of these domains at the ankle-links of hair cell stereocilia is unclear. Either some basement membrane components or homologous proteins are expressed in the reticular lamina found at the surface of the organ of Corti, or laminin domains contribute in another way to the function of the ankle-links.

The function of the cytoplasmic domain of usherin is unknown. Several stretches of conserved residues are found when aligning orthologous usherin sequences, some corresponding to predicted secondary structures on the human sequence (Figure 24). The last 6 residues are almost perfectly conserved across compared sequences, the last three residues consisting in a PDZ binding motif (-THL_{COOH}). I will further comment the PDZ/PBM binding modes in a dedicated section. Lastly, no tertiary structure is predicted by the software Alphafold[175] for this cytoplasmic domain (not shown).

Of the four proteins of the Usher 2 complex, usherin has probably been the less studied structurally. Yet, it appears to be the most mutated gene of the four Usher 2 associated proteins, accounting for more than 50% of patients suffering from Usher syndrome of type 2[176]. As declared in the Leiden Open Variation Database (LOVD; www.lovd.nl), 1652 different pathogenic or likely pathogenic variants of the protein have been reported.

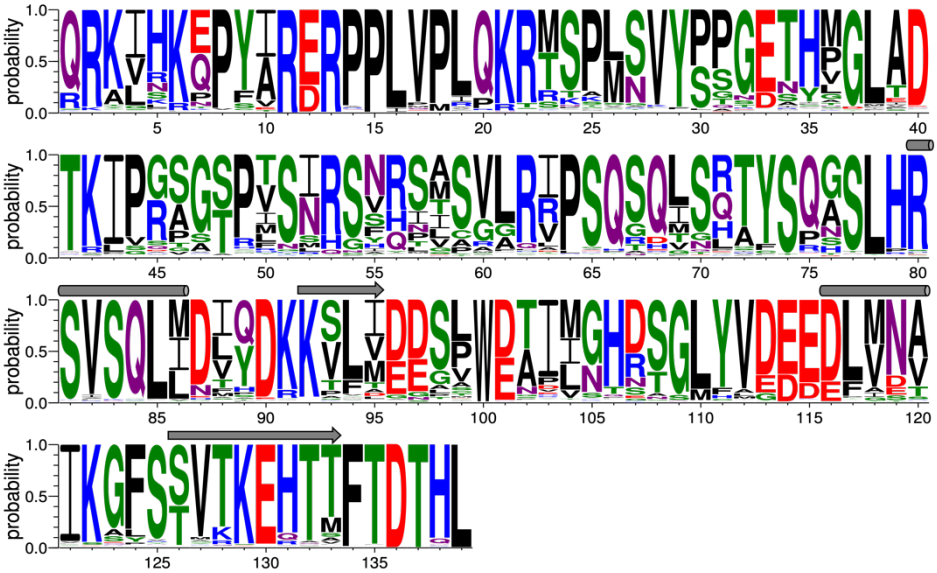


Figure 24 – Logo representation of the cytoplasmic domain of 276 usherin orthologs generated using Weblogo3.Sequences were retrieved from the UniprotKB database and aligned using the ginsi algorithm from the Mafft package. Only positions ungapped in the human sequence (entry O75445) were retained. Secondary structures predicted using Jpred4 (Jnetpred) are depicted by arrows (strands) and cylinders (helices).

II.1.d – ADGRV1 (Adhesion G-protein coupled receptor V1)

ADGRV1 is a seven-pass transmembrane protein of the adhesion GPCR[177] (aGPCR) family B2[178], referred to by several names over the years, including GPR98, MASS1 and VLGR1. Multiple isoforms have been described (up to 8[179]), but the full-length protein (ADGRV1b) is mainly expressed in the cochlea and retina[166]. Its extracellular region consists of 5879 residues with predicted 35 CalX-β domains, 6 EAR domains and one GAIN domain. Using Alphafold prediction, we can instead identify 39 CalX-β domains and 7 EAR domains. Furthermore, CalX-β 11 is split in two parts with a PTX domain inserted in its sequence, same for CalX-β 18 that has an unidentified fold inserted in the sequence, as well as CalX-β 24 which has the 7 EAR repeats inserted in its sequence (Figure 25).

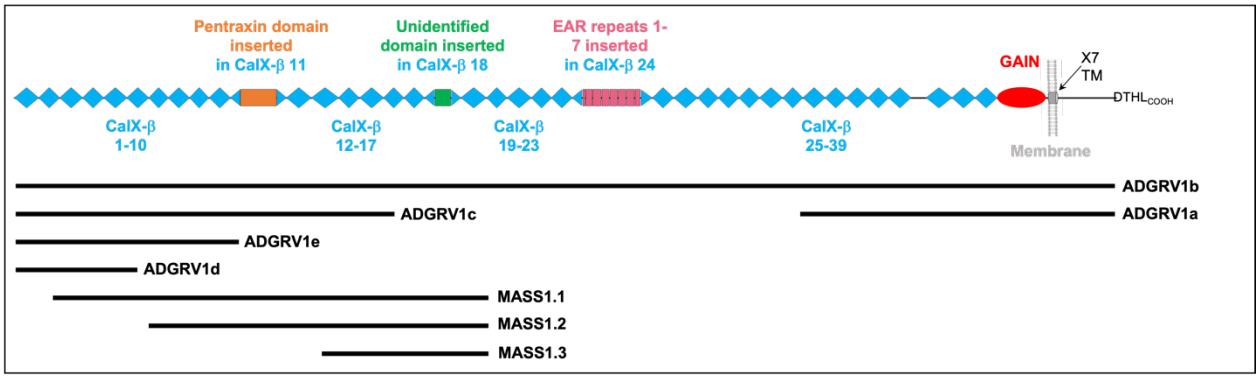


Figure 25 – Schematic representation of the ADGRV1 protein from N- to C-terminal. Domain delimitations are based on Alphafold predictions. Isoforms are indicated in black.

Domains of the extracellular region

PTX (pentraxin) domains are about 200 residues long, forming a large β sandwich. In ADGRV1, the PTX domain is found inserted in the CalX-β 11 of the extracellular region (Figure 26). PTX domains are described by homology to the domains of the pentraxin proteins. They mediate protein-protein and ligand interactions in a calcium-dependent manner[180,181]. PTX domains can be found in ring-like homopentameric structures, in complex with various small molecules (PDB entries 1B09, 1D2S, 2W08, 3L2Y, 4AVS, 3PVN). The role of the PTX domain of ADGRV1 remains to be determined.

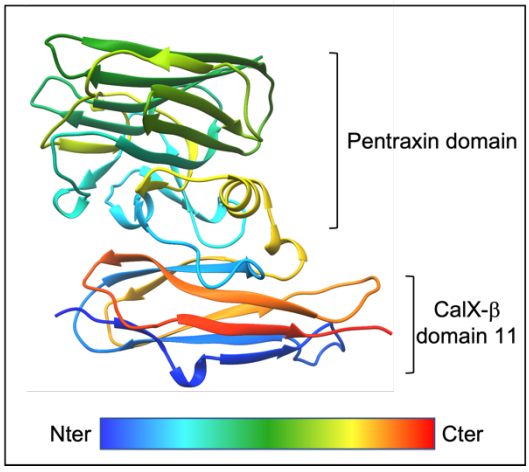


Figure 26 – Alphafold prediction of a pentraxin domain inserted in ADGRV1 CalX-β domain 11.

Alphafold predicts a folded insertion in the CalX-β 18 of ADGRV1 (Figure 27). Blastp searches on the corresponding sequence returns hits from ADGRV1 orthologous sequences only and fold prediction by homology modelling using Phyre2 does not return any hit with a satisfying threshold. I could not identify this fold, so it is either a new module of unknown function, a conserved fold with highly divergent sequence compared to the rest of the family, or a prediction mistake by Alphafold despite the high confidence level.

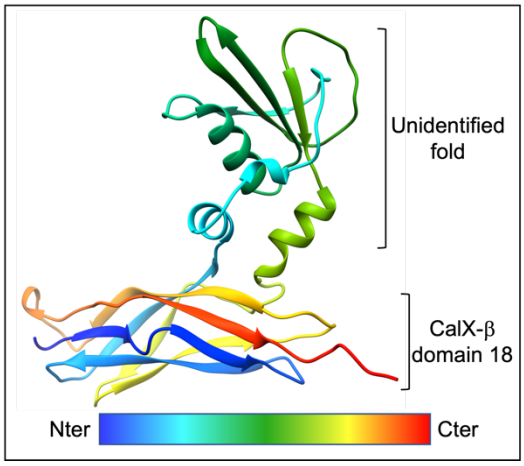


Figure 27 – Alphafold prediction of an unidentified fold in ADGRV1 CalX-β domain 18.

EAR (Epilepsy associated repeat) or EPTP repeats are small motifs of about 50 residues, adopting a 4 strands conformation, assembling as β-propeller structures. In ADGRV1, there are 7 of these repeats, inserted in CalX-β 24 domain (Figure 28). The function of these repeats has not been extensively characterized, with few available structures. The only example comes from the LGI1 protein, where it mediates the interaction with the protein ADAM22 as well as calcium binding (PDB entry 5Y2Z)[182]. In terms of function, EAR/EPTP-containing proteins are involved in some forms of epilepsy[183], but the molecular mechanism associating the disease to the repeat itself is unclear. For ADGRV1, the link to epilepsy likely relates to its expression in the central nervous system[184], where it regulates cell migration of astrocytes and GABAergic interneurons in the auditory cortex[185,186]. The functional relevance of these repeats in cochlear hair cells and retinal photoreceptor cells is unknown.

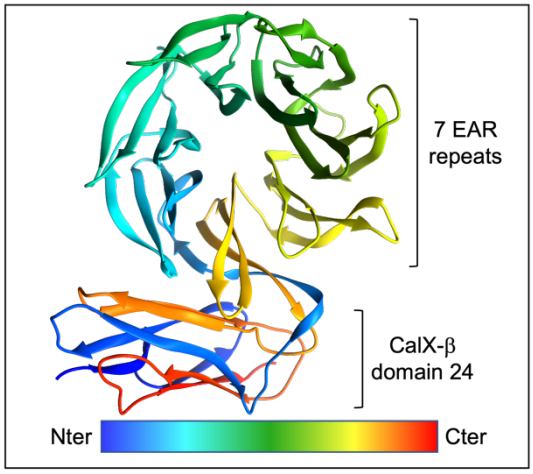


Figure 28 – Alphafold prediction of 7 EAR repeats β-propeller structure inserted in ADGRV1 CalX-β domain 24.

The C-terminal extracellular domain of ADGRV1 is a GAIN (G protein-coupled receptor autoproteolysis-inducing domain) located upstream of the 7 transmembrane segments. It consists of about 277 residues in human, folding as a helical bundle of 5 α helices (~residues 5623 to 5728) and a 12 strands β sandwich (residues 5729 to 5899). The GAIN domain contains a G protein-coupled receptor Proteolytic Site (GPS) mediating the autoproteolysis of a loop upstream of the last β strand of the domain (Figure 29). By analogy with proposed mechanisms for other adhesion GPCRs[187,188], the histidine 5889 would act as a

general base, deprotonating the side chain of serine 5891 that would in turn make a nucleophilic attack on the carbonyl carbon of the methionine 5890. This induces the hydrolysis of the peptide bond between methionine 5890 and serine 5891, resulting in two peptide chains corresponding to the adhesion subunit α (extracellular region) and GPCR subunit β (7TM and cytoplasmic domain). The last strand of the GAIN domain then becomes the new N-terminal end of the β subunit, called stachel peptide, which could activate the receptor. The stachel peptide may remain non-covalently bound to the GAIN domain, dissociating from it under adequate stimuli, such as a mechanical stress. This hydrolysis has been shown to occur in cellular models for ADGRV1 and would result in a switch[189] between $G_{\alpha s}/G_{\alpha q}$[190] and $G_{\alpha i}$[191] coupling with the full length and cleaved ADGRV1, respectively. The connection of ADGRV1 to several cellular signalling pathways may reflect tissue-dependent functions and/or a response to a cellular event.

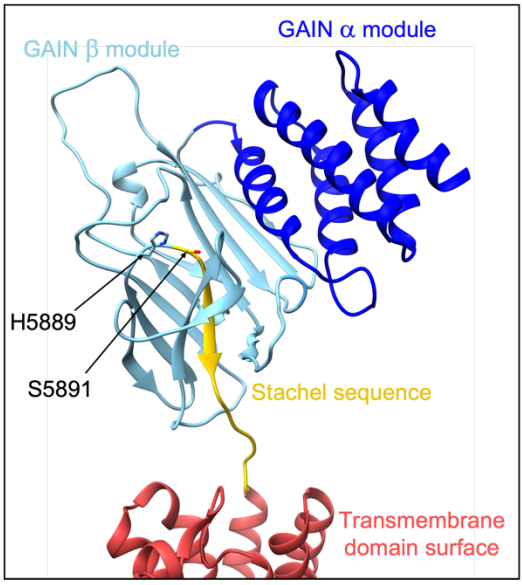


Figure 29 – Alphafold prediction of ADGRV1 GAIN domain before autoproteolysis at the GPS.

Generalities on G protein-coupled receptors (GPCRs)

GPCRs are the largest known family of cell-surface receptors, with more than 800 identified proteins in human[192]. They generally mediate cell signalling in response to extracellular ligands. As such, they are primary targets for therapeutic approaches, with ~30% of FDA approved drugs targeting GPCRs[193]. This large protein family is diverse in terms of sequence, ligand and function. Several classifications of GPCRs have been proposed over the years, based on sequence/structure homology and types of ligands. A recent classification proposes 7 main groups: class A (Rhodopsin), class B1 (secretin), class B2 (adhesion), class C (glutamate) and class TAS2 (or classes F and T; Frizzled/taste) receptors[194]. The GPCRdb[195] (GPCR database) tracks receptor classification, with many additional information such as phylogenetic analysis, sequence clustering and consensus.

The canonical fold of GPCRs consists of 7 transmembrane helices (TM). Generally, binding of an extracellular ligand induces the activation of the receptor, characterized by a rearrangement of the 7TM favouring the interaction with the partner heterotrimeric G protein. This interaction promotes the substitution of the bound G_{α} subunit GTP to GDP, triggering dissociation of G_{α} and $G_{\beta\gamma}$ subunits, which separately mediate signalling[192]. Note that GPCRs

have dynamic conformations with or without ligand, both potentially compatible with G-protein coupling. Each GPCR therefore exhibits a different basal activity, which may be enhanced by ligand binding (GPCR activation).

A second signaling pathway is mediated by arrestin proteins, which bind to the cytoplasmic tail of GPCRs in a phosphorylation dependent manner. β-arrestins (four sub-types in human) are versatile proteins, mediating GPCR desensitization, endocytosis and signalling. The function of β-arrestin binding depends on its bond-state conformation[196,197], which is determined by the phosphorylation pattern of the GPCR intracellular domain. It has been proposed that different combinations of phosphorylation sites on the GPCR cytoplasmic domain would act as a barcode resulting in a specific β-arrestin binding mode[198–200]. This results in a differential regulation of the GPCR activity depending on tissue-specific factors, such as ligands, types of GPCR kinases (GRKs) and types of β-arrestin.

Activation mechanisms by stachel peptides

Adhesion GPCRs (aGPCRs, class B2) often possess a large extracellular region containing a GAIN domain promoting the autoproteolysis of the protein. The new N-terminal end of the receptor consists in a short (~10-15 residues) and rather conserved sequence called stachel peptide, thought to act as an agonist and mediate the activation of the GPCR (Figure 30A, F). This activation mechanism of aGPCRs by their N-terminal stachel sequence has been described recently in four articles[201–204], with structures of the activated β subunits of 7 different receptors in complex with several G or miniG proteins. For simplicity, I will number residues of each receptor by their position in their sequence alignment with human ADGRV1 β subunit (Figure 30A), starting at position 1 for the S/T residues initiating the stachel sequence. In all structures, the N-terminal of the stachel sequence forms a helix turn parallel to the membrane, buried in a groove formed between the tips of TM 1, 6, 7 and 2, 3, 5, in place of the orthosteric ligand[205] (Figure 30B). A first set of residues, F^3, L^6 and M/L^7, have their side chains oriented towards the hydrophobic core of the receptor. Residues in positions 3 and 7 interact with W^{263} on TM6 (Figure 30C), similar to the "toggle-switch" tryptophan responsible for ligand-induced opening of the TM6 cytoplasmic end in family A GPCRs[204,206], enabling G protein binding. The interaction network of the downwards facing position 9 of the stachel peptide is more variable between the different structures[202], reflected by a lower degree of conservation from one receptor to the other. In order to accommodate the N-terminal stachel sequence, TM6 and TM7 extracellular tips kink away from the TMD (Figure 30B, D). For TM6, the kink occurs at a conserved glycine (G^{260}) position, three

residues upstream the toggle-switch tryptophan (W^{263}) (Figure 30D). For TM7, the kink is mediated by disruption of the canonical α helix backbone hydrogen bond network between the conserved G^{291} and L^{289} residues of the rather conserved "SLQG" motif (Figure 30D). The TM7 glutamine (Q^{290}) residue engages in electrostatic interaction with the indole nitrogen of the TM6 toggle-switch tryptophan (W^{263}), stabilizing the concerted opening of TM6 and TM7[204] (Figure 30D). The TM6/TM7 kink is a feature of the active state of class B1 GPCRs[201,204]. This results in the opening of the rather large groove where the N-terminal of the stachel sequence is inserted. Accompanying the movement of the TM7 tip, a clockwise motion (seen from the extracellular side) of the TM1 is observed, consisting in a tilt of the helix rather than a kink, further opening the surface of the TMD. The C-terminal part of the stachel peptide and extracellular loop 2 (ECL2) cover the TMD, isolating the hydrophobic core of the receptor (Figure 30B). The position of the ECL2 on top of the TM 2, 3, 4 and 5 side of the groove is stabilized by a disulphide bridge between loop residue C^{190} and TM3 residue C^{97}, as well as interaction with residues of the N-terminal stachel positions pointing outwards[201,202]. Furthermore, ELC2 residue W^{191} is inserted inside the TMD and interacts with the conserved stachel L^{6} pointing downwards[204]. The position of the C-terminal stachel sequence on top of TM 1, 6 and 7 is stabilized by a variable set of interactions with the tip of these helices.

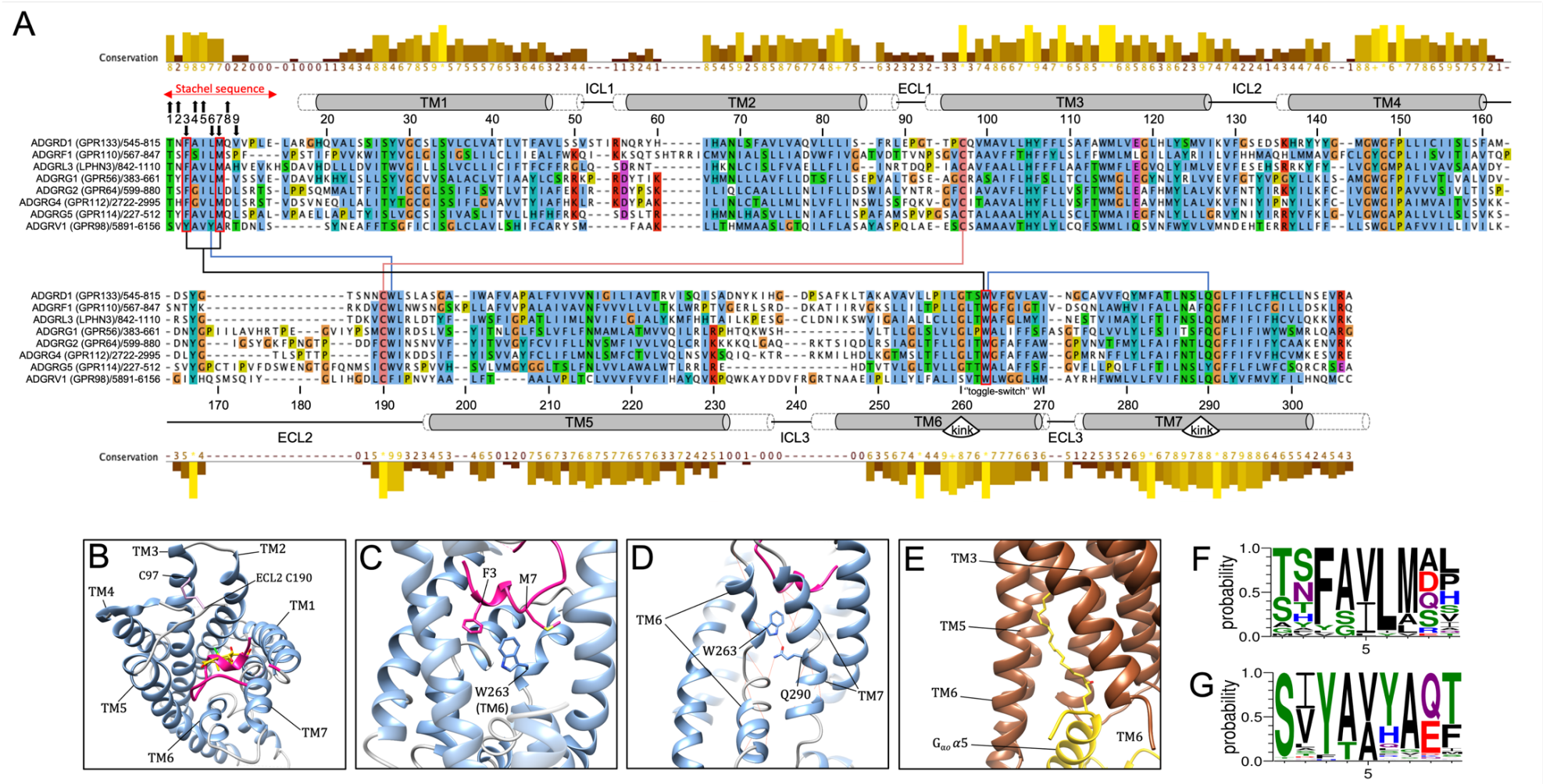
In terms of G protein binding, the main common feature of the stachel-activated aGPCRs results from the kinked TM6 and TM7, resulting in a wide cavity in the intracellular surface of the TMD[201,202], where is inserted the α5 helix of the G protein α subunit. However, the interaction networks between the different G proteins (G_s, mini-G_s, mini-G_{i-1}) and the 7 stachel-activated aGPCRs display a higher degree of variability. A noteworthy feature of the TM6 kink is the bulge of the pivot region towards the middle of the TMD, which would impair binding to post-translationally modified G protein $\alpha5^{201}$, as opposed to the ligand-induced GPR97-G_o complex where a palmitoylation of $G_{\alpha o}$ α5 is inserted between TM3, TM5 and TM6 (Figure 30E). Consistently, no density corresponding to post-translational modification of the G protein was observed in the TMD of stachel-activated aGPCRs[203].

ADGRV1 β subunit was shown to constitutively signal through $G_{\alpha i}$[191] coupling in cellular models. However, the stachel sequence of ADGRV1 differs from those of other adhesion GPCRs. The consensus for stachel sequences is very high for positions 3, 6 and 7, responsible for the interaction with W^{263}. The phenylalanine position 3 is substituted to a conserved tyrosine in ADGRV1, which may recapitulate the interaction network of the phenylalanine residue as well as engage in electrostatic interactions. The residue in position

7, often a methionine, is replaced by an alanine in ADGRV1, which otherwise occurs only in ADGRB1 to 3 aGPCRs. It has been shown that ADGRB1 and ADGRB2 have a cleaved β subunit with higher G protein coupling compared to the uncleaved proteins[207,208], suggesting stachel-driven activation. Despite not having a structural insight, these results suggest that the alanine substitution at position 7 of the stachel sequence is compatible with a constitutive activity of the receptor. ADGRV1 also has a tyrosine residue instead of the leucine found almost exclusively in other a GPCRs. The only other human aGPCR without a leucine at that position is ADGRE1, which does not have the catalytic residues upstream of the stachel peptide and may therefore not undergo stachel-driven activation. ADGRV1 being cleavable[191], the function of its tyrosine in position 6 is therefore unclear. The concomitant substitution W191F on the ECL2 loop, normally interacting with L7, may participate in adapting another interaction mode. Finally, ADGRV1 does not have the conserved G^{260} residue described as pivot for TM6 kink, but has a conserved TM7 kink region. The observed G260S substitution may suggest a decrease in flexibility of the TM6, but the effect on a potential helix kink has to be documented.

Overall, ADGRV1 has been shown to undergo autoproteolysis at the GPS and has a β subunit capable of constitutively coupling to G proteins in *vitro*[191,209]. The divergent stachel sequence of ADGRV1, however, suggests that the interaction network leading to constitutive activation would differ from other aGPCRs. Yet, the autoproteolysis of ADGRV1 and potential function of its constitutive activation remains to be demonstrated *in vivo*.

Figure 30 - A) Alignment of protein sequences from the 6 structurally solved stachel-activated adhesion GPCRs, as well as human ADGRV1. Alignment was performed using the einsi algorithm from the Mafft package. Conservation was estimated using Jalview. Shared secondary structures are indicated by grey cylinders, dashed cylinders indicate the extreme delimitations observed in the structures. ICL: Intracellular loop. ECL: Extracellular loop. Colored solid lines indicate key interactions found in all structures. B) Top view of the stachel-activated structure of ADGRG4 (PDB entry 7WUJ), with the stachel sequence in pink. In yellow is the ligand of the activated ADGRG3 (PDB entry 7D76) after structural alignment with ADGRG4. C) Side view of the stachel-activated structure of ADGRG4, with residues position 3 and 7 of the stachel sequence on both sides of TM6 ''toggle-switch'' tryptophan. D) Side view of the stachel-activated structure of ADGRG4, highlighting TM6 and TM7 kink, as well as Q290 mediated hydrogen bond network. E) Side view of the ligand-activated adhesiion GPCR ADGRG3, with the $G_{\alpha o}$ α5 helix palmitoylation inserting in a cavity formed by the receptor's TM3/4/6 helices. F) Logo representation of stachel sequences from A (positions 1 to 9) generated using Weblogo3. Letter height indicates the probability of finding the corresponding residue at a given alignment position. G) Logo representation of the stachel sequences from 97 ADGRV1 orthologs. Orthologs were retrieved from the UniprotKB database using Hidden Markov Profile (see result section 1), with subsequent sequence length and PBM filtering.


A
Conservation
Stachel sequence
TM1
ICL1
TM2
ECL1
TM3
ICL2
TM4
ADGRD1 (GPR133)/545-815
ADGRF1 (GPR110)/567-847
ADGRL3 (LPHN3)/842-1110
ADGRG1 (GPR56)/383-661
ADGRG2 (GPR64)/599-880
ADGRG4 (GPR112)/2722-2995
ADGRG5 (GPR114)/227-512
ADGRV1 (GPR98)/5891-6156
"toggle-switch" W
TM5
TM6
kink
TM7
ECL2
ICL3
ECL3
B
TM3
TM2
C97
ECL2 C190
TM4
TM1
TM5
TM7
TM6
C
F3
M7
W263
(TM6)
D
TM6
W263
Q290
TM7
E
TM3
TM5
TM6
$G_{\alpha o}$ α5
TM6
F
probability
1.0
0.5
0.0
5
G
probability
1.0
0.5
0.0
5

ADGRV1 cytoplasmic domain

Downstream its TMD, human ADGRV1 has a 152 residues long cytoplasmic domain of unknown function. Three stretches of conserved residues are found when aligning ADGRV1 protein sequences from various orthologs, some corresponding to predicted secondary structures (Figure 31). These conserved stretches of residues, encompassing hydrophobic and aromatic residues, likely mediate protein-protein interactions relevant for complex assembly and/or receptor function[189].

Conserved positions of the cytoplasmic domain include multiple serine and threonine residues. The cytoplasmic tails of GPCR are known to be phosphorylated on serine and threonine residues, often in multiple positions, mediating the recruitment of β-arrestin proteins[195,196,197–199]. It has also been shown that some constitutively active GPCRs have increased β-arrestin binding[207,210], likely resulting from an enhanced phosphorylation rate of their cytoplasmic domain. The contribution of conserved residue stretches and S/T residues in ADGRV1 cytoplasmic domain function, in particular in the coupling to β-arrestins, remains to be determined in hair cells.

Finally, the last 4 residues are the same as in usherin, forming a conserved type 1 PBM (-DTHL$_{COOH}$), allowing the interaction with PDZ-containing cytoplasmic proteins.

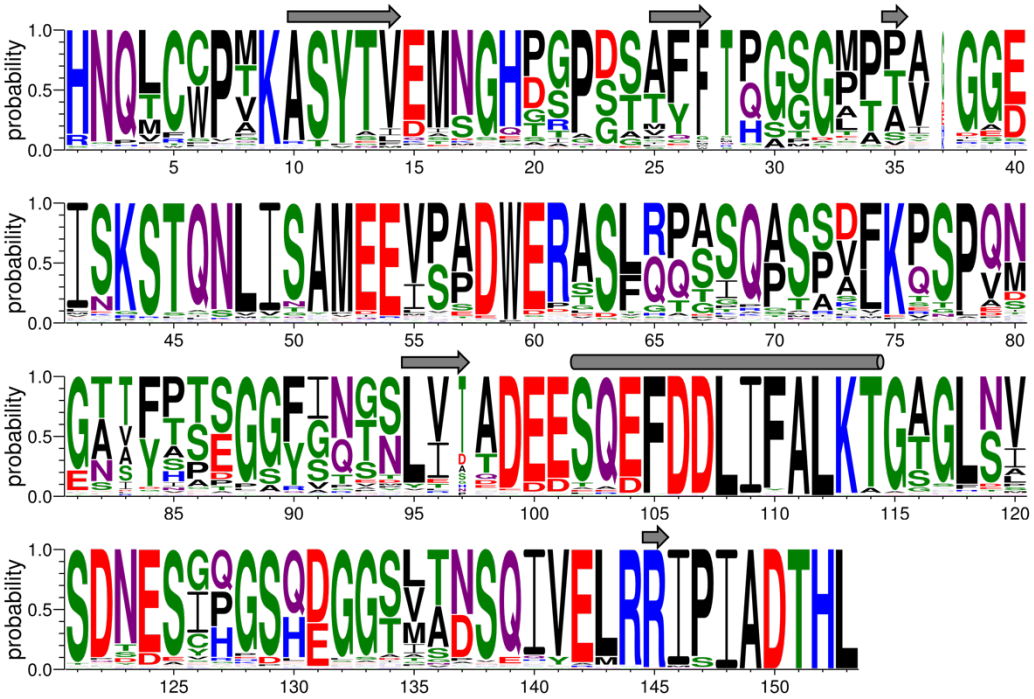


Figure 31 – Logo representation of the cytoplasmic domain of 97 ADGRV1 orthologs generated using Weblogo3. Orthologs were retrieved from the UniprotKB database using Hidden Markov Profile (see result section 1), with subsequent sequence length and PBM filtering, and aligned using the ginsi algorithm from the Mafft package. Letter height indicates the probability of finding the corresponding residue at a given alignment position. Secondary structures predicted by JPred4 (Jnetpred) are depicted by arrows (strands) and cylinders (helices).

II.2 – Usher 2 cytoplasmic proteins

The cytoplasmic domains of usherin and ADGRV1 associate to two paralogous proteins, whirlin and PDZD7. These two cytoplasmic proteins are thought to promote the assembly of a large cytoplasmic network through protein-protein interactions mainly mediated by PDZ domains and HHDs. Before I summarize the known function of the two proteins, I will describe the general binding mechanisms of their protein-protein interaction domains.

II.2.a – PDZ domains

PDZ domains take their name from the three first proteins in which they were identified, PSD-95 (PostSynaptic Density 95), DLG1 (Discs Large tumor suppressor 1) and ZO-1 (Zonula Occludens 1). This acronym saw its first use in 1995[211], replacing its previous names DHR (DLG Homology Domain) or "GLGF repeat"[212]. They compose a large family of protein-protein interaction modules, with 273 PDZ domains found in 152 proteins in human[213]. Interestingly, multiple PDZ domains are commonly found in PDZ-containing proteins, with up to 13 PDZ domains in the Multiple PDZ Domain Protein (MPDP). PDZ-containing proteins mediate both scaffolding and signalling in various cellular environments[214–217].

Fold and canonical binding mode

PDZ domains are usually about 85 residues long, with a rather high sequence variability and few conserved positions (Figure 32). The canonical fold of PDZ domains consist of five β-strands forming an incomplete β-sandwich closed on the sides by two α-helices (Figure 33). Only four pairs of residues located on each end of the α1 helix ("GG" and "GL") and β4 strand ("GD" and "NG") have a higher degree of conservation despite not directly being part of the hydrophobic core or participating in the ligand binding (Figure 32).

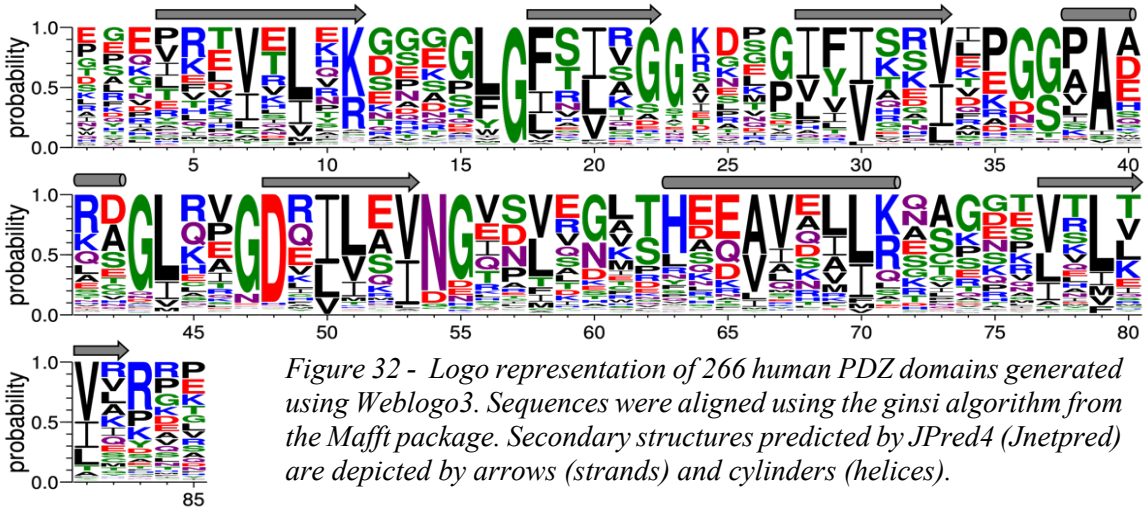


Figure 32 - Logo representation of 266 human PDZ domains generated using Weblogo3. Sequences were aligned using the ginsi algorithm from the Mafft package. Secondary structures predicted by JPred4 (Jnetpred) are depicted by arrows (strands) and cylinders (helices).

PDZ domains canonically interact with short motifs at the C-terminus PBM of their partner. For simplicity in describing PDZ/PBM complexes, PBM residues are often numbered from the last C-terminal residue p(0) towards the N-terminal of the protein (p(-1), p(-2)…, p(-N)). In the bound state, the last three residues of the partner insert in a groove formed between the PDZ β2 strand and α2 helix, forming an antiparallel intermolecular β-sheet with the PDZ β2 and β3 strands (Figure 33). For all PDZ/PBM complexes, three determinants of interaction are nearly always met. The first two involve the conserved "GLGF" loop between the β1 and β2 strands of the PDZ, at the top of the binding groove. The consensus for this loop is closer to [G/ϕ]-ϕ-G-ϕ where ϕ is a hydrophobic residue. The C-terminal carboxylic acid of the PBM forms hydrogen bonds with the backbone amide groups of this loop. In addition, the side chain of the PBM C-terminal residue p(0), facing the hydrophobic core of the PDZ, establishes hydrophobic contacts with the side chains of the "GLGF" hydrophobic residues. This usually constrains the last residue to have a medium-sized hydrophobic side chain, allowing to contribute to the hydrophobic core, without clashing with the crowded environment. Lastly, the side chain of the residue p(-2) of the PBM interacts with the side chain of the first residue of the PDZ α2 helix. The type of interaction made at that position defines the canonical classes of PBM. Class 1 PBM have a serine or threonine residue in position p(-2), forming a hydrogen bond with a histidine on the first position of the PDZ α2 helix. PBM of class 1 are thus defined by the consensus -[S/T]-X-ϕ_{COOH} (where X can be any residue). Class 2 PBM have a hydrophobic residue in position p(-2), forming hydrophobic contacts with a hydrophobic or aliphatic side chain on the first position of the PDZ α2 helix. PBM of class 2 are thus defined by the consensus -ϕ-X-ϕ_{COOH}. Class 3 PBM have an anionic residue in position p(-2), often forming a hydrogen bond with a polar residue on the first position of the PDZ α2 helix. Class 3 PBM are thus defined by the consensus -[D/E]-X-ϕ_{COOH}. To my knowledge, no interaction with a PBM having a cationic residue at position p(-2) has been reported thus far.

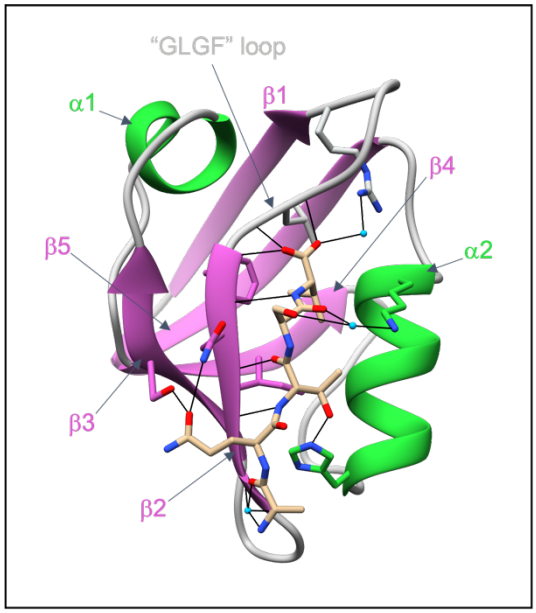


Figure 33 - Structure of PSD-95 third PDZ domain in complex with the C-terminal PBM of the protein CRIPT (PDB entry ABE9). The PBM is inserted with its C-terminal carboxylic acid directed towards the ''GLGF'' loop.

Based on these three consensuses alone, PDZ/PBM interactions would not seem to be highly specific. Endogenous PDZ-mediated interactions are usually of medium affinity, with dissociation constants (Kd) between few micromolar and few tenths of micromolar. However, if a certain degree of promiscuity exists between different PDZ domains, additional determinants of interaction tune the specificity and affinity of one PDZ for different PBM with similar p(0) and p(-2) residues. Residues in position p(-1) and p(-3), exposed to the solvent near the binding groove, can also participate in the interaction, resulting in a strengthened or weakened binding. Additional contacts can also occur between residues upstream of the canonical PBM and various parts of the PDZ. The furthest reported contact occurs between the PDZ domain of the protein MAST2 in complex with the PBMs of the PIP kinase PTEN (PDB entry 2KYL) or of the rabies virus glycoprotein (PDB entry 2KQF). In both structures, position p(-11) of the PBM interacts with residues of unusually extended β2/β3 strands of the PDZ domain.

Finally, PDZ/PBM binding can also be tuned by post-translational modifications of either partner. These modifications can be varied in nature (methylation[218], acetylation[219]…), but phosphorylation is the most documented modification affecting PDZ/PBM interactions[220-226]. In particular, the S/T residue in position p(-2) of type 1 PBMs is a prime candidate for regulatory phosphorylations. Yet post-translational modifications may also occur few residues upstream the canonical PBM and can also affect PDZ/PBM interactions.

Overall, each PDZ evolves a network of residues allowing to discriminate between different PBM beyond the canonical p(0) and p(-2) positions. Individual PDZ domains are thus expected to have different specificity profiles, binding strongly a limited variety of PBM sequences or in the contrary have a rather large array of possible partners. However, drawing the binding profile of a PDZ domain against all possible PBM sequences, including to some extent the upstream residues, is challenging and few examples exist[227]. The concept of unique specificity profile also applies to PBM sequences, each interacting with multiple PDZ domains with varied promiscuity. High-throughput techniques have been developed to describe the binding profile of individual PDZ or PBM[228,229], allowing to probe the intricate PDZ/PBM interactome.

Internal motifs

The large majority of reported PDZ domain interactions involve the C-terminal end of the partner protein, described as canonical PBM. However, at least some PDZ domains are able to specifically recognize motifs other than a partner's C-terminus. These motifs are referred to as internal PBM. The first internal PBM was described in 1999 between PDZ domains of syntrophin and PSD95 with a folded β-hairpin downstream of the PDZ domain of the protein nNOS[230,231]. Since, several examples of PDZ domains bound to internal PBMs have been described, including PDZ domains from Par6[232], NOS[233], Tip-1[234], PICK1[235], whirlin[34], Scribble[236] and Shank1[237] proteins. Internal motifs can be found both in disordered regions and secondary structures from the partner. Note that some internal PBMs consist of a shift of two residues from the C-terminus of the protein, which can also consist of a canonical PBM. The interaction is thus still mediated by the C-terminal region of the partner, which exhibits potentially competing PBM sequences with different specificities[34,235]. Internal PBM/PDZ complexes exhibit specific structural features allowing to mimic the interaction network of canonical PBMs while preventing steric clashes with the core of the PDZ (Figure 34).

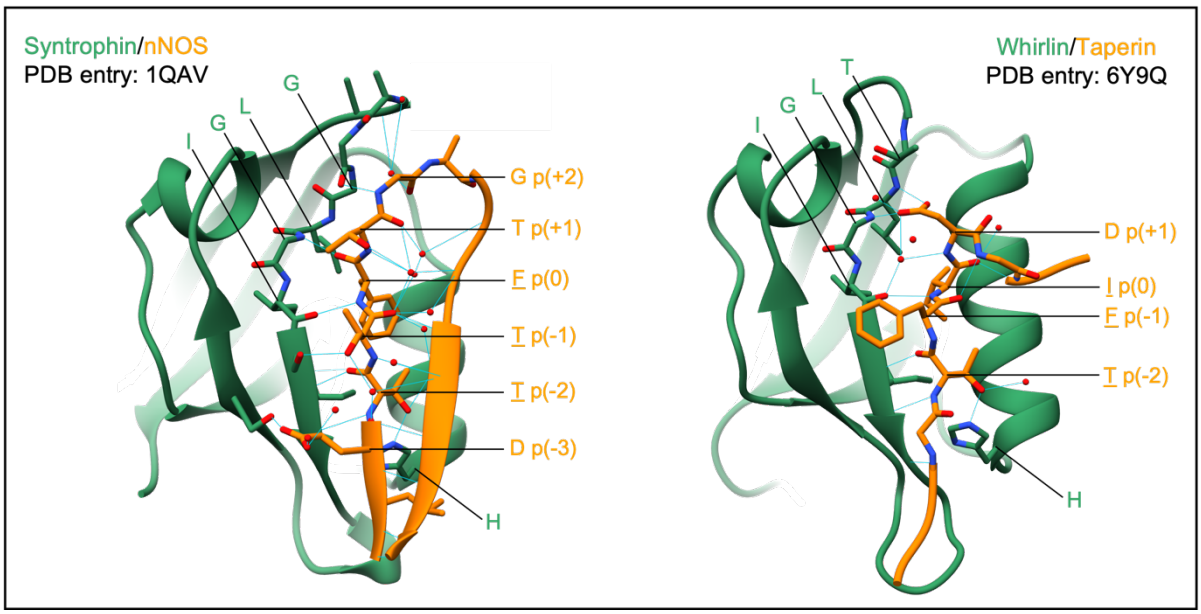


Figure 34 - Structures of syntrophin PDZ domain and whirlin third PDZ domain in complex with the internal PBMs of the proteins nNOS and taperin respectively. PDB entries 1QAV and 6Y9Q.

The syntrophin/nNOS complex (PDB entry 1QAV) consists of the interaction between the PDZ domain of the alpha-1 syntrophin protein and a β-hairpin downstream from the nNOS protein, highlighting the capability of PDZ domains to interact with secondary structures and not exclusively disordered regions. In this complex, the PBM does not insert deeply in the binding groove, with limited contacts between the α2 helix and the loop that links the two strands of the hairpin, allowing to minimize steric clashes. Only the threonine at pseudo-position p(-2) of the PBM engages the canonical hydrogen bonding with the first histidine residue of the PDZ α2 helix. The phenylalanine residue at PBM position p(0) initiating the turn after hairpin first strand has its side chain buried canonically in the contact with the hydrophobic core of the PDZ. The interaction between canonical C-terminal carboxyl however is not perfectly recapitulated by the phenylalanine carbonyl group, with a single hydrogen bond formed with the amide group of the second glycine residue of the PDZ "GLGF" loop. A second hydrogen bond is instead formed between the p(+2) residue's amide group and the carbonyl group of the "GLGF" loop first glycine residue. The interaction is further stabilized by PBM residues p(-2) and p(-3) forming hydrogen bonds with side chains of the PDZ β2 and β3 strands.

The whirlin/taperin complex (PDB entry 6Y9Q) consists of the interaction between the third PDZ domain of whirlin and a flexible internal PBM of the taperin protein. In this complex, the threonine residue at p(-2) and the isoleucine residue at p(0) of the PBM engage in the classical type 1 PBM/PDZ interaction mode, respectively forming a hydrogen bond with the histidine residue at the first position of the PDZ α2 helix and making hydrophobic contacts with the core of the PDZ domain. However, the backbone of the PBM does not recapitulate any of the hydrogen bonds canonically formed between the terminal carboxyl and the "GLGF" loop. Instead, the peptide turns away from the PDZ, allowing the side chain of an aspartate residue in position p(+1) to fold back towards the GLGF loop and mimic this interaction network. In both presented complexes, several structural water molecules participate in the stabilization of both the PBM conformation and the interaction with the PDZ. Multiple water molecules participating in the hydrogen bonding network of internal PBM/PDZ interactions seem to be a recurring feature to adapt unique conformations (see PDB entries 1X8S, 6Y9O, 6MYE and 7A9B).

The short length of internal PBMs consensus renders their prediction by bioinformatic tools impossible thus far, with thousands potential hits in the human proteome. Several approaches based on high-throughput methods have been proposed to overcome this issue[228,238,239]. However, many PDZ domains are known to have a promiscuous interaction

profile with a specificity of interaction tied by cellular protein localization. The results yielded by these studies should therefore be confirmed case by case *in vitro*. Nonetheless, the high number of hits generated by these studies suggest that functional internal PBMs may be abundant and that their propensity in cellular context is likely underestimated.

Interaction with lipids

Despite being primarily described as protein-protein interaction modules, multiple PDZ domains binding to phospholipids (often phosphatidylinositol phosphates or PIPs) have been characterized over the years[240–244]. This recurring feature appears to be of functional importance for the cellular localization of multiple PDZ-containing proteins, as well as to tune the affinity of the PDZ domains for protein partners. Since, several high-throughput and predictive studies aimed at more generally describe the propensity of PDZ binding to lipids, resulting in an estimated ~20%[243] to 40%[245] of mammalian PDZ domains engaging in such interaction. Interestingly, PDZ domains reported to interact with lipids are mostly part of proteins involved in submembrane densities of transmembrane proteins. Lipid binding may therefore have evolved over time as an additional tool for regulating the function of membrane-associated PDZ-containing proteins. A molecular phylogenetic analysis of the PDZ domain family may unveil evolutionary-linked clusters of domains and improve our understanding of the sequence-function interplay in PDZ domain lipid binding.

From *in vitro* experiments, it appears that PDZ domains identified with lipid binding capability usually have a high isoelectric point (pI; above 9) compared to the family average (pI ≈ 7 for the human PDZome)[246]. This results from an enrichment of cationic residues (K, R) exposed at the surface of the domain. This is consistent with PDZ domains interacting with the polar heads of membrane phospholipids, often negatively charged in the intracellular leaflet. However, the pI value is not of sufficient predictive power and the position of positive side chains on the PDZ domain surface locates the binding site. Clusters of lysine and arginine residues on the tertiary structure of the domain are usually responsible for binding to phospholipid heads (Figure 35).

Chen *et al.*[245] have proposed a model for predicting the effect of lipid interactions on PDZ domains function depending on the position of the cationic patch. According to their model, PDZ domains may be partitioned into three classes with respect to their lipid binding mode. Class A PDZ domains have their lipid binding site distinct from the peptide binding groove, resulting in a largely independent binding of the two types of molecules. Class B1 PDZ domains have their cationic patch in the vicinity of the peptide binding groove, affecting

the affinity of the domain for protein partners while allowing tertiary complex formation. Class B2 PDZ domains have overlapped lipid and peptide binding sites, resulting in mutually exclusive interaction.

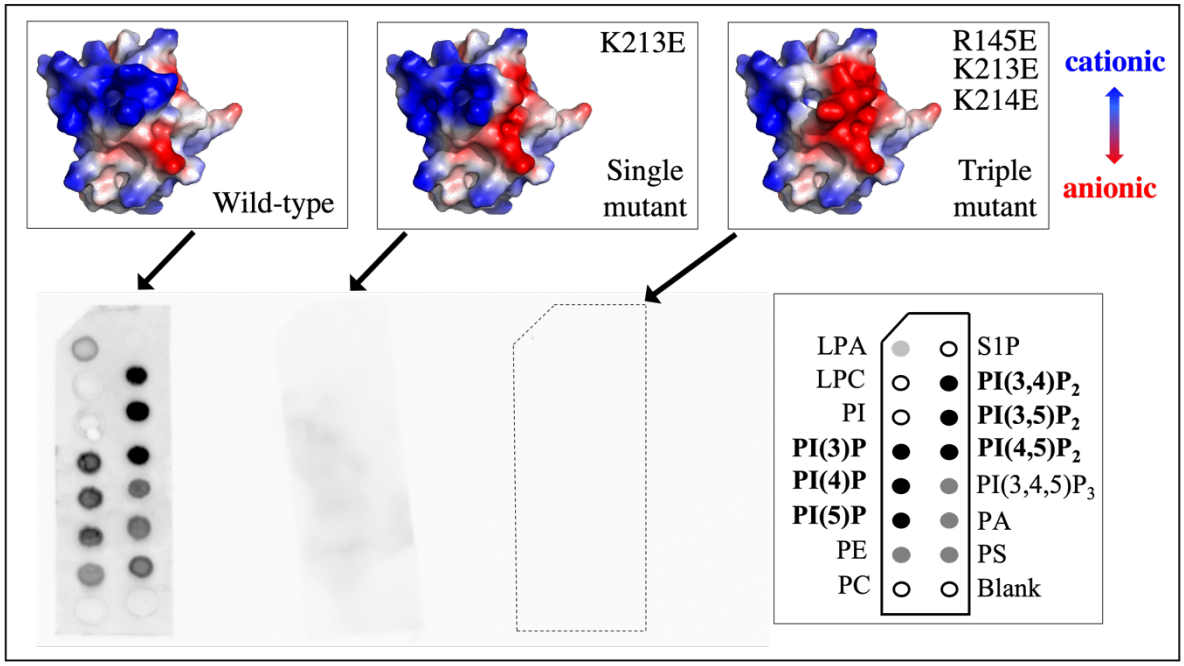


Figure 35 - Lipid-binding assays of the first PDZ domain of whirlin, as well as single and triple charge reverted mutants. The surface of each PDZ construct is colored according to its electrostatic potential. The protein is incubated with a strip exhibiting spots of bound lipids. After washing, the bound protein is detected using a chemiluminescent probe. The map of lipid spots on the strips is illustrated in the bottom right. Wild-type whirlin first PDZ interacts with several phospholipids, especially PIPs. Disruption of the cationic patch abrogates the interaction entirely. Unpublished data.

Homotypic supramodules

PDZ domains are often found in multiple copies in a single protein[214] and exhibit a strong tendency to function as PDZ supramodules. These supramodules may consist of adjacent PDZ domains of a same protein, forming a PDZ tandem, or result from homo- and hetero-dimers between different molecules. Chang *et al.*[247] estimated that about 30% of mammalian PDZ domains engage in homotypic interactions.

The term of PDZ tandem usually refers to two sequential PDZ domains from a same protein with a linker of 25 residues or less. About 35 human proteins have PDZ domains meeting this criterion, yet little structural information is available about PDZ tandems. From available data, each PDZ tandem seems to have unique features, such as the relative positioning of the two domains or the exchange between different conformational states. I will here describe some examples highlighting different aspects of tandems in PDZ domains function.

The association of two PDZ in a tandem may serve to synergistically bind to partner proteins or lipids. In syntenin (Figure 36A), the PDZ domains are tethered by a short linker of 5 residues, therefore constrained in close proximity, yet with diverse conformations in solution[248]. In this tandem, both PDZ domains are functional[249] and can interact with the C-terminal PBM of syndecan. Syndecan assembling as dimers, the interaction with syntenin PDZ1-2 tandem results in an increased affinity compared to isolated domains[250,251]. In PSD-95 (Figure 36B), the two N-terminal PDZ domains are also tethered by a short linker of 4 residues containing 2 prolines. In that case, the orientation of the two PDZ domains is constrained by the short and rigid linker, favouring a unique conformation with both PBM binding sites facing the same direction. This would allow a synergistic binding of the tandem to oligomeric partners, notably transmembrane glutamate receptors. Interestingly, this constrained orientation of the PDZ tandem of PSD-95 is favourable for the design of small molecules binding simultaneously to the two domains[252].

PDZ tandems may also serve as folding regulators. A good example is the protein GRIP1, which possesses two PDZ tandems (Figure 36C and 36D). In the first PDZ tandem (PDZ1-2), the peptide binding site of the second PDZ domain seems buried in the interface with the first PDZ and therefore unable to canonically bind to protein ligands[253]. In the second PDZ tandem (PDZ4-5), the first PDZ does not bind to canonical ligands, as its PBM binding groove is distorted[254]. Each tandem is thus composed of one supposedly non-functional PDZ domain regarding direct PBM binding. However, they appear critical for the folding of their respective adjacent PDZ[253,255], potentially further tuning their binding specificity or affinity.
PDZ domains can also serve as negative regulatory units, as illustrated in PDZ tandem of the Inad protein, expressed in drosophila photoreceptor cells (Figure 36E). In this protein, the closed conformation of the PDZ4-5 tandem is maintained in a pH-dependent manner by an extension of the PDZ5, wrapping around the PDZ4. In this conformation, the binding groove of the PDZ5 can accept a partner's PBM. Upon pH variation, the tandem would open, modifying the redox potential of a cysteine pair across the PDZ5 PBM binding groove, leading to disulfide bridge formation and closure of the binding pocket[256].

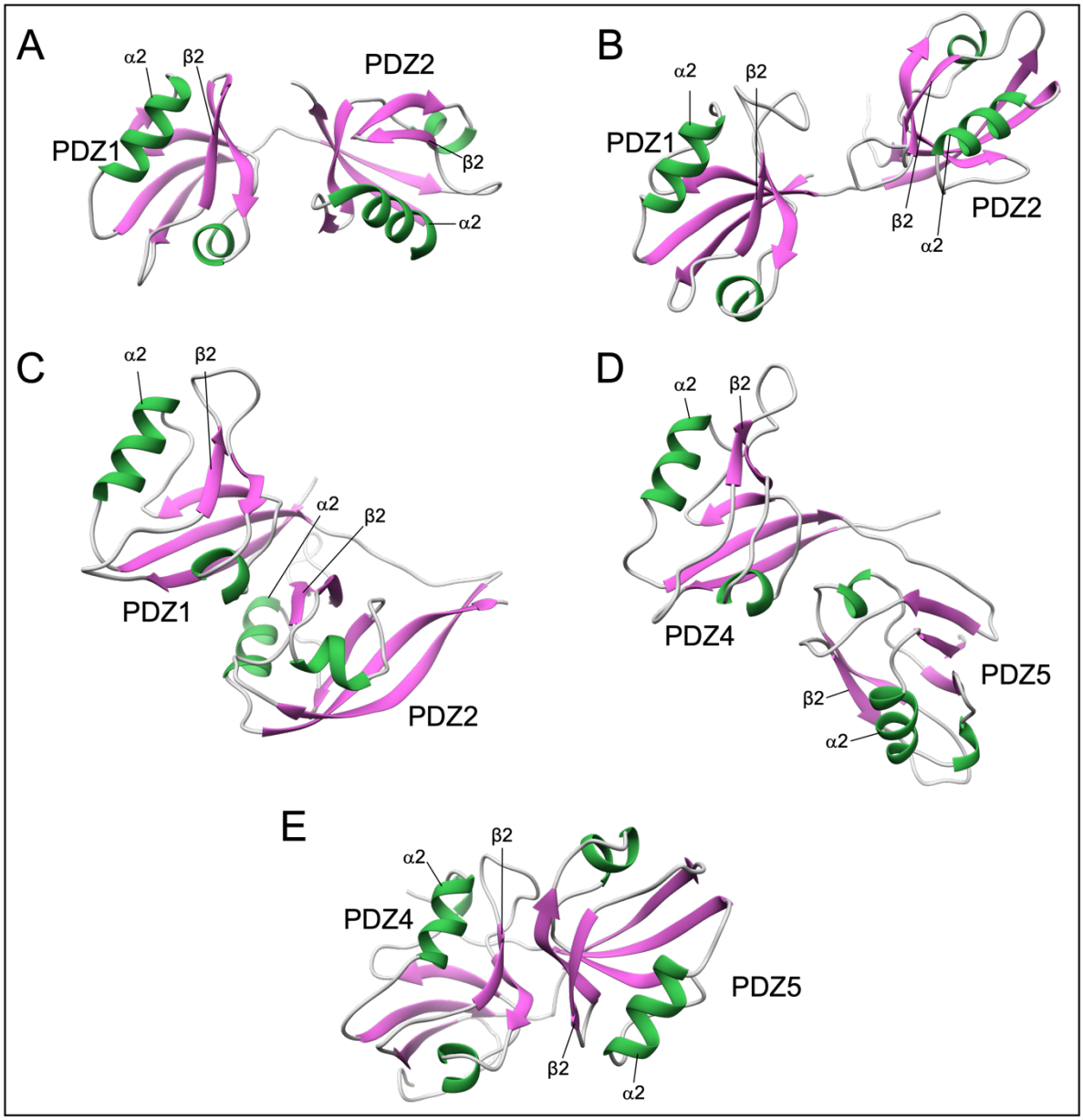


Figure 36 - Different conformations in PDZ tandems. In all examples, the orientation of the N-terminal PDZ is kept relatively similar, highlighting the conformational variability of the C-terminal PDZ positioning. A) Syntenin PDZ1-PDZ2 tandem (PDB entry 1N99). B) PSD-95 PDZ1-PDZ2 tandem (PDB entry 3GSL). C) GRIP PDZ1-PDZ2 tandem (PDB entry 2QT5). D) GRIP PDZ4-PDZ5 tandem (PDB entry 1P1D). E) INAD PDZ4-PDZ5 tandem (PDB entry 3R0H).

Heterotypic supramodules

PDZ domains can also engage in the formation of supramodules with other types of domains. This is well illustrated by the proteins PTPN3 and PTPN4, which both possess a PDZ domain making transient contacts with a phosphatase domain. These contacts inhibit the catalytic activity of the protein. However, when the PDZ domain is in complex with a PBM, the supramodule opens and the phosphatase domain resumes its catalytic activity[257–259].

Another example of catalytic activity regulated by PDZ domains occurs in the bacterial DegS protein, from the Htra protease family, containing adjacent PDZ and proteolytic domains. DegS proteins assemble in oligomers by their catalytic domains[260], with PDZ domains directed towards the solvent. In this resting conformation, the activity of the proteolytic domain is inhibited by the adjacent PDZ. During cellular stress, the PDZ domains can interact with PBMs of misfolded targets, inducing a conformational switch and the activation of proteolytic domains[261,262].

An example of PDZ domain taking part in non-catalytic heterotopic supramodule would be the MAGUK protein family. Proteins of this family have a conserved architecture mediating highly specific partner binding through a triple recognition pattern. This pattern is referred to as PSG, for PDZ, SH3 and GuK domains. In this interaction mode, the PBM binds to the PDZ and the preceding residues to a hydrophobic pocket formed between the PDZ and SH3 domains[263]. The upstream residues further interact with the non-catalytic Guanylate kinase (GuK) domain, usually recognizing phosphorylated motifs.

PDZ domain extensions

As illustrated by the aforementioned Inad PDZ4-5 tandem, PDZ domains function sometimes rely on conserved extension outside of their canonical fold. These extensions can be located upstream (shank3[264]) or downstream (PSD95 PDZ3[265], MAGI1 PDZ1[266,267], NHERF1 PDZ2[268], harmonin PDZ1[87], whirlin PDZ1 and PDZ2[269]) of the domain. In whirlin and harmonin, PDZ extensions also mediate the formation of supramodules. In harmonin, the first PDZ possesses a C-terminal hairpin extension mediating an interface with the preceding HHD domain as well as direct contacts with the partner[87] (Figure 37A). This supramodule displays a ~10 fold increased affinity for protocadherin-15 compared to the isolated PDZ[55]. In whirlin, the N-terminal PDZ tandem is in equilibrium between open states and a close conformation. In this compact supramodule, each PDZ extension folds as a β hairpin packed onto the inner leaflet of its adjacent PDZ. The two extensions then form the interface between the two PDZ domains (Figure 37B). This results in an increase in affinity of about 3 fold for partners compared to the isolated PDZ1. Mutations targeting conserved positions of each PDZ β hairpin results in the opening of the supramodule[269].

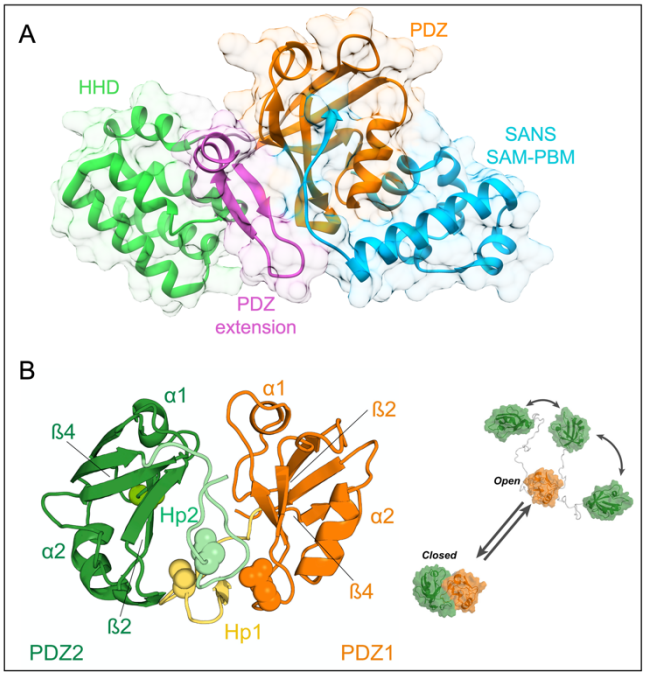


Figure 37 - PDZ extensions mediate the assembly of hetero- and homo-typiques supramodules. A) Harmonin HHD (Harmonin Homology Domain) and first PDZ domain form a supramodule exclusively mediated by the PDZ extension. The extension also contacts residues upstream of the PDZ partner's PBM (PDB entry 3K1R). B) The PDZ1 and PDZ2 domains of whirlin both possess C-terminal hairpin extensions (Hp) mediating the interface of a transient supramodule, in exchange between opened and closed conformations.

In shank3, the PDZ domain also has an extension folding as a β hairpin at the N-terminal end of the domain. In the shank3 dimer, this extension makes direct contacts with residues upstream of the partners PBM bound to the other protomer. The resulting crossed dimer have an affinity 100 fold higher for the partner[264]. In NHERF1, the second PDZ has a C-terminal extension folding as two α helices, stabilizing the folded PDZ domain and decreasing entropic cost upon PBM binding[268]. In PSD-95, the third PDZ domain possesses a C-terminal extension folding as a α helix, increasing the affinity of the domain by 25 fold for the PBM of the CRIPT protein. This dramatic increase is thought to arise from both a modulation of PDZ side chains dynamic[265] (decreased entropic cost upon PBM binding), as well as direct contacts with the ligand[270].

These multiple examples highlight the diversity of PDZ domains extensions both in fold (disordered, helix or strand) and function (supramodule mediation, allosteric modulation or additional anchoring point for the partner). The number of examples available does not however reflect the predicted prevalence of extensions in PDZ functioning. About 40% of human PDZ domains have a predicted secondary structure outside of their canonical fold[271], which may consist in extension critical of their function. Furthermore, this estimate does not consider unstructured extensions that may be involved in binding and regulatory mechanisms.

The prevalent role of tandems and extensions in PDZ domains function is detrimental to high-throughput approaches using delimitations encompassing the canonical fold of the domain. In auditory proteins, both extensions and PDZ supramodules have been characterized and affect the binding affinity and specificity towards partners. These includes homotypic

(whirlin) and heterotypic supramodules, as illustrated by the PDZ1 of harmonin forming extension-mediated contacts with its preceding HHD.

II.2.b – Harmonin Homology Domains

Harmonin Homology Domains were first identified in harmonin. They compose a small family of protein-protein interaction domains, with only 9 domains in 6 proteins, namely harmonin, whirlin, PDZD7, delphilin, CCM2 and RTEL. Interestingly, 4 of these proteins also contain PDZ domains and exhibit a similar organization (Figure 38). This includes the three deafness and Usher-associated proteins whirlin, harmonin and PDZD7, as well as the post-synaptic scaffolding protein delphilin. The last two HHD-containing proteins are CCM2 (cerebral cavernous malformations 2 protein) and RTEL (regulator of telomere elongation helicase).

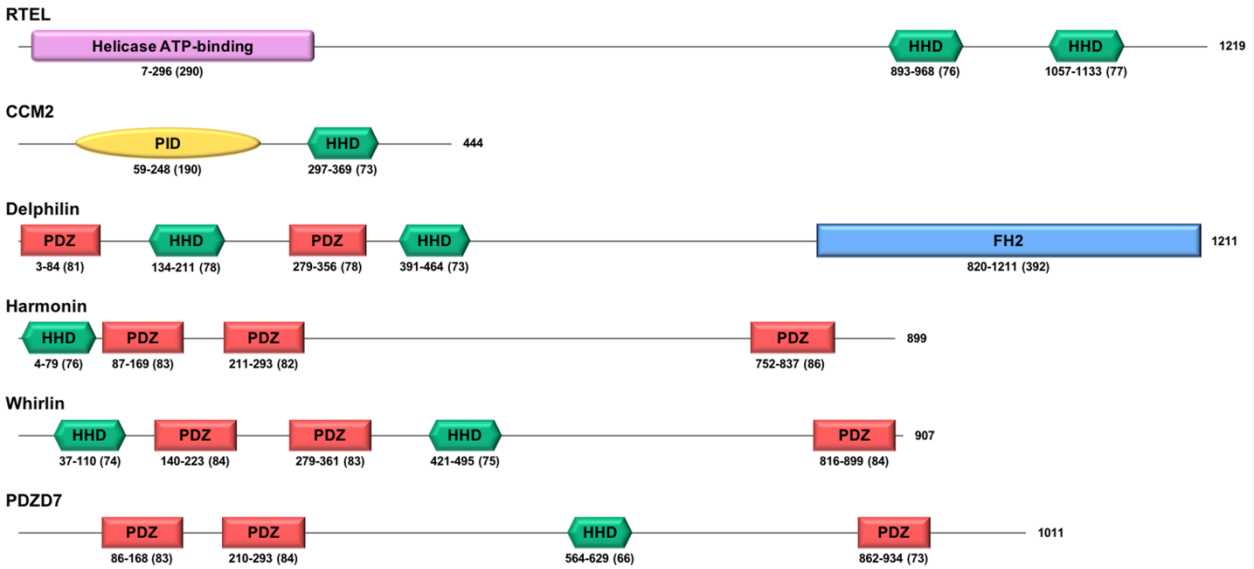


Figure 38 - Schematic representation of HHD-containing proteins known to date.

Canonical fold

HHD structures of CCM2 (PDB entries 4YKD and 4FQN), harmonin (PDB entry 2KBQ) and whirlin (6FDD and 6FDE) consist of about 80 residues folded as a bundle of 5 α helices. The structure of these domains appears very similar, with a globular fold largely maintained by hydrophobic contacts of its core and assisted by electrostatic interactions between inter-helix loops. The main difference consists in a length variation of the first helix of the bundle, from 10 (whirlin second HHD) to 14 residues (HHD from CCM2). Despite sharing a similar tertiary structure, HHD domains share low sequence identities, with

conserved residues largely tied to the hydrophobic core (Figure 39A). The identification of HHD domains by conventional approaches based on sequence similarity has therefore proven difficult, leaving this domain family mostly unexplored.

HHDs share a common fold with PAH (Paired amphipathic helices) domains, with the 4 first helices adopting a largely similar tertiary structure with conserved positions in the hydrophobic core[272] (Figure 39B). PAH domains are found in the human sin3a, sin3b and Gon4-like proteins, three transcriptional repressors. The striking sequence and structure similarities between the HHD and PAH domains suggest that they may have diverged from a common ancestor.

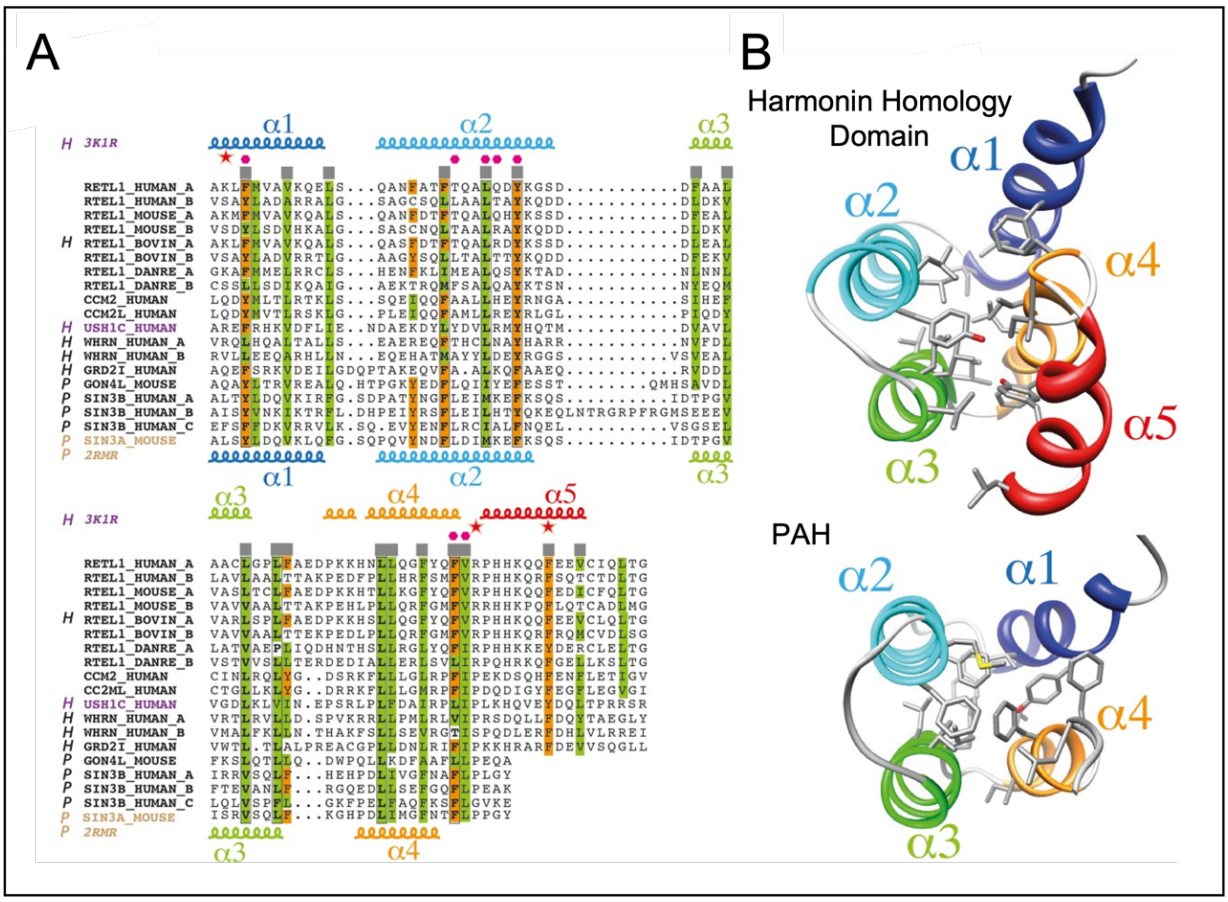


Figure 39 - A) Sequence alignment of HHD and PAH domains. B) Structural comparison between HHDs and PAH domains. PDB entries 3K1R and 2RMR. Adapted from Faure et al. (2014)[272].

A second protein-protein interaction mode is shown by the structure of harmonin HHD (Figure 40B). In this structure, the HHD forms a supramodule with the adjacent PDZ domain through its C-terminal α3-α5 surface[87], mediated by the PDZ's β hairpin extension[87]. The conformation of the PDZ and its extension is stabilized by the HHD, resulting in a high affinity inter-molecular interaction with SANS despite the HHD not directly contacting the partner.

This result suggests that HHDs may act as modulators tuning the affinity and possibly specificity of adjacent domains. This is confirmed in whirlin, with constructs encompassing both first HHD and first PDZ having 10 fold higher affinity for protocadherin-15 PBM compared with the isolated PDZ[55], though no structure is available.

A third protein-protein interaction can be observed on the structure of the complex between CCM2 HHD and MEKK3[273] (Figure 40C). This interaction involves the canonical amphipathic helix binding described above, as well as an additional contact area between CCM2 HHD helices α2-α3 and MEKK3 PB1 domain. The resulting complex has a large contact area between the two proteins. However, no affinity values were measured by the authors of the study and point mutations made to test the physiological relevance of the complex only target the canonical interaction surface.

Lastly, a more recent study has shown that the HHD domain of PDZD7 is able to bind to lipids by a cationic patch located its N-terminal end, although no structure of the complex is available[274]. This interaction appears necessary and sufficient to promote the submembrane localization of the protein in HEK cells. The obtained structure (PDB entry 4Y5O) also shows that the upper part of the HHD is kinked and folds back on top of the canonical binding groove. Authors of the study suggest that PDZD7 HHD might therefore not promote protein-protein interactions through this interface.

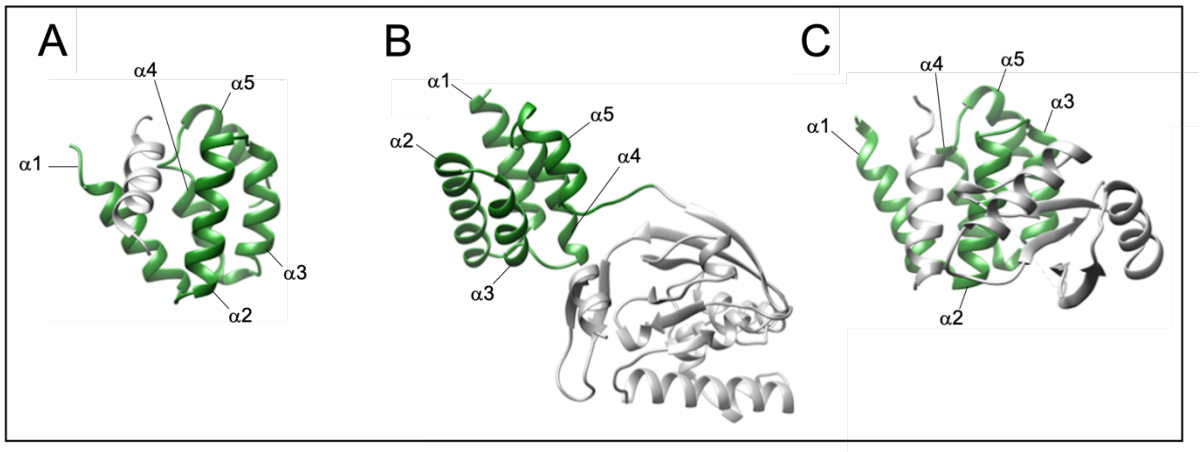


Figure 40 - Structures of HHDs highlighting three different binding interfaces. A) Harmonin HHD in complex with a cytoplasmic helix from cadherin-23. This binding mode is common with PAH domains. PDB entry 2KBR. B) Harmonin HHD forming a supramodule with the adjacent PDZ domain by interaction with its C-terminal extension. PDB entry 3K1R. CCM2 HHD in complex with the NPB1 domain of MEKK3, highlighting a second interface in addition to the canonical amphipathic helix binding groove. PDB entry 4Y5O.

Altogether, these results highlight 4 different interaction surfaces located throughout the bundle surface, allowing for direct protein and lipid binding as well as multidomain supramodular organization. HHDs thus appear to be flexible domains for tuning protein interactions in various cellular environments, which contrasts with the limited domain family size. The first part of my PhD results will present a bioinformatics study of the HHD family, aiming at identifying all HHD-containing protein sequences and propose a phylogenetic tree describing the differentiation between various HHD and PAH-containing proteins.

II.2.c – Whirlin

Whirlin (also named DFNB31 or CIP98) is a cytoplasmic protein of 907 residues in human, encompassing 2 PDZ flanked by 2 HHDs in its N-terminal part. After the second HHD is found a large, disordered and flexible[34] region, rich in prolines. The C-terminal end of the protein consists in a PDZ domain followed by a PBM of type 2. The multiple protein-protein interaction domains and lack of catalytic activity suggest that whirlin acts as a scaffold to promote the assembly of large protein complexes. It is expressed as several isoforms in various cellular locations[275] (Figure 41).

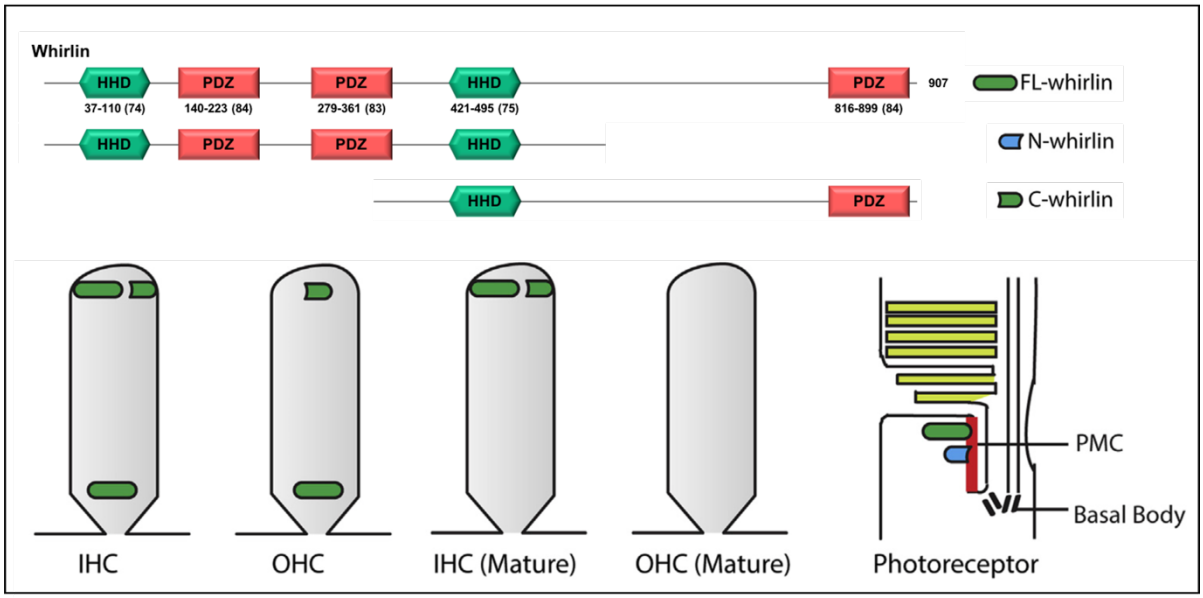


Figure 41 - Schematic representation of whirlin isoforms and their expression patterns in cochlear hair cells and photoreceptor cells. Modified from Mathur, P., & Yang J. (2019)[275].

In the cochlea, the long isoform is expressed at ankle-links of developing IHCs and OHCs, as well as the tip of developing and mature IHCs. The full-length whirlin is also expressed in photoreceptor cells of retina. The C-terminal isoform encompasses the second

HHD domain, the extended proline-rich region and the C-terminal third PDZ domain. It is expressed at the tip of both developing IHCs and OHCs, as well as mature IHCs. No whirlin isoform is expressed in mature OHCs. The N-terminal whirlin isoform contains the first HHD and the N-terminal PDZ tandem. This isoform is specific to photoreceptor cells.

Whirlin is therefore exposed to different cellular environments and available data suggest that it promotes the assembly of distinct protein complexes. The formidable plasticity of whirlin third PDZ domain reflects the functional diversity of the protein, interacting with multiple cellular partners with unique binding networks[34].

At the tip of stereocilia, whirlin coordinates a dense complex[276] primarily involving myosin15a[277,278], eps8[279,280], CASK[281], p55[282], p4.1[283] and GPSM2[284,285] proteins. The molecular motor myosin15a appears necessary for whirlin localization at the stereocilia tip[278], suggesting a main role in protein transport. Myosin15a and whirlin knockouts both lead to a similar shortening of hair cell stereocilia. Eps8 is an actin-capping protein and p55 and p4.1 actin cytoskeleton stabilizing proteins. These three proteins have therefore been suggested to promote the actual growth of hair cells stereocilia. Notably, overexpression of Eps8 is associated to stereocilia lengthening and larger densities of Eps8 labelling is observed in stereocilia forming the larger rows of the hair bundles[280]. CASK is a protein kinase of the MAGUK family, shown to interact with whirlin in the central nervous system[281] and to regulate gene expression after translocation in the nucleus of neurons[286,287]. The function of CASK at the stereocilia tip is unclear, where it may act as a scaffold for protein-protein interactions (L27, PDZ and SH3 domains), notably binding to p4.1[287]. GPSM2 is a member of the G protein signalling modulator family, promoting GPCR independent activation of the G protein $G_{\alpha i\text{-}3}$[284]. It is required in the hair cell stereocilia to stabilize higher concentrations of whirlin, myo15a and eps8 at the tip of the longer stereocilia. Knock-out of GPSM2 results in shortened stereocilia[285], but the underlying signalling process is unknown. At the tip of the shorter rows of stereocilia, whirlin also participates in the assembly of the lower tip-link density, notably interacting with protocadherin-15 and CIB2 proteins.

At the ankle links, whirlin coordinates, together with PDZD7, the assembly of the Usher 2 protein complex[288]. It interacts with both transmembrane proteins usherin and ADGRV1 identical PBMs through its N-terminal PDZ domains, as well as its paralogous protein PDZD7. Using pull-down assays, Chen *et al*.[288] suggested that all domains of whirlin can interact with all domains of PDZD7, while its third PDZ would also promote homodimerization of the protein, but such interaction network seems unlikely. Furthermore, the localization of the other three Usher 2-associated proteins is largely unaffected[165] in mouse

strains which do not have detectable whirlin expression at the base of the stereocilia[163]. Lastly, the protein CIB2 has also been suggested to interact with whirlin and myosin7a[289], a protein motor participating in molecule transport in photoreceptor cells[290] also expressed in cochlear hair cells stereocilia. The function of whirlin at the ankle link cytoplasmic complex is overall unclear, potentially including scaffolding properties and a role in molecule transport throughout the stereocilia.

In the retina, whirlin full-length and N-terminal isoform also colocalize with usherin and ADGRV1, together with the protein SANS, at the photoreceptor periciliary membrane complex[160,291,292] (Figure 42). SANS is a scaffold protein interacting with tubulin-associated proteins such as myomegalin[293] and spectrin βV[294], respectively involved in microtubules polymerization and kinesin binding. The correct transport of photosensitive molecules, notably opsin, along the connecting cilium between inner and outer segments of photoreceptor cells seem to rely on both actin-associated (myosin7a) and tubulin-associated (kinesins) molecular motors. However, the molecular mechanisms underlying Usher proteins function in photoreceptor cells are unclear. With pathogenic variants leading to the bleaching of the retina over time[295], Usher proteins have been suggested to promote the trafficking of photosensitive molecules from the inner segment to the outer segment of photoreceptor cells. This hypothesis is supported by the association of Usher proteins to both actin and tubulin binding proteins, but the variability of photoreceptor cells' morphology across species complicates the general understanding of their function.

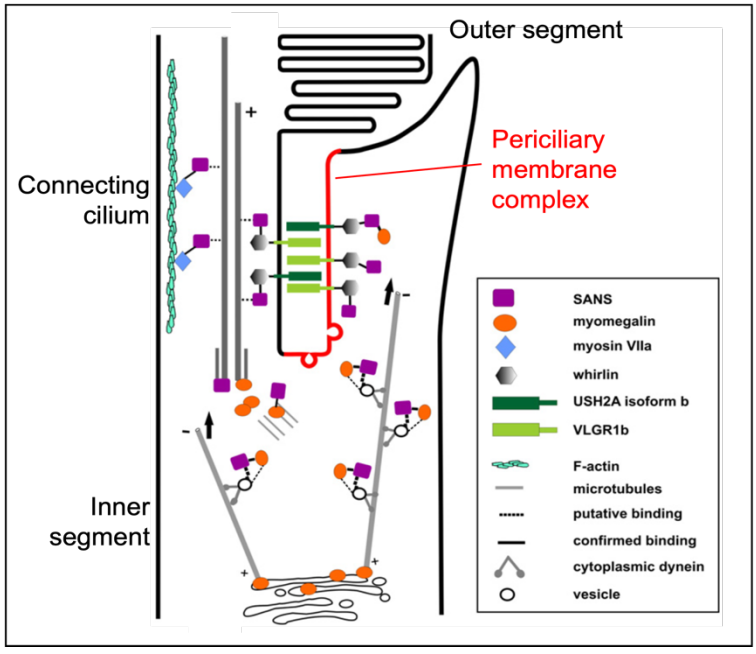


Figure 42 - Suggested complex assembly for Usher proteins at the connecting cilium region of photoreceptor cells. Modified from Overlack et al. (2011)[293].

II.2.c – PDZD7

PDZD7 is a paralogous protein of harmonin and whirlin, but has been the focus of fewer studies compared to its homologues. It consists of 1033 residues in human, encompassing a N-terminal PDZ tandem, a single HHD in its central region and a C-terminal PDZ domain (Figure 43). It is supposedly expressed as different isoforms, including the full-length protein[296,297], potentially 2 N-terminal isoforms[297] and 2 C-terminal isoforms[296]. It is also unclear if PDZD7 is solely expressed transiently during the development, or if some isoforms are retained in mature hair and photoreceptor cells[161,296].

Figure 43 - Schematic representation of the PDZD7 protein organization. Domain delimitations correspond to the human protein (UniprotKB entry Q9H5W4).

Knock-out of PDZD7 has been characterized as inducing hearing loss[161,297–299]. The type of hearing loss, however, seems to depend on the type of variant expressed by the patients[298,300–304], with down slopping[298,300,304] (more severe impairment at high frequencies) and possibly progressive[298,303] hearing loss as recurring features. PDZD7 has also been described as an Usher phenotype modifier[299,305], with more acute symptoms including both hearing loss and *retinitis pigmentosa*. Inclusion of PDZD7 as an Usher causative gene is debated, with knock-outs leading to Usher-like *retinitis pigmentosa* in zebrafish[305], but patients with PDZD7 variants having no visual impairment. This apparent discrepancy might arise from different retinal structures between humans and zebrafish or from a differential physiopathogenesis process between knock-out and punctual/truncated pathogenic variants, suggesting a partial overlap of function between Usher 2 associated proteins, especially in the retina.

In hair cell stereocilia, PDZD7 is mainly detected at the ankle-links[162], a localization that would depend on its third PDZ domain[297] and on the interaction with myosin7a [163]. Myosin7a does not exhibit a C-terminal PBM and the determinants of interaction with PDZD7 remains to be identified. PDZD7, however, possesses an 80 residues extension after its PDZ3 domain, predicted to be disordered, which may mediate a similar interaction mode (independent of a PBM) as observed in the paralogous protein harmonin in complex with myosin7 C-terminal MyTH4-FERM domain[86,306]. Inversely, the localization of myosin7a does not appear to depend on PDZD7 expression[161]. Knock-out of PDZD7 in mouse results in a loss of whirlin labelling at the ankle-links and a diffused localization of ADGRV1 and

usherin[161], suggesting that it acts as a primary scaffold for the complex assembly of the three Usher 2 proteins. This function appears to be carried out through direct interaction with the cytoplasmic domains of both ADGRV1 and usherin, in addition to their other scaffolding partner whirlin, as monitored by pull-down experiments[288]. This hypothesis is strengthened by PDZD7, ADGRV1 and usherin localizations being largely unaffected in mice strains not expressing whirlin at the ankle links[161,165]. PDZD7 also interacts with the Usher protein SANS[296], but the localization of the two proteins is independent[161]. Other partners of PDZD7 include PIP5k1c[297] (phosphatidylinositol-4-phosphate 5-kinase), an enzyme expressed in hair cell stereocilia participating in $PI_{[4,5]}P2$ biosynthesis.

In the retina, PDZD7 is also found at the periciliary membrane complex, where it is necessary to localize ADGRV1 to the connecting cilium[288,305]. It interacts there with the Usher protein harmonin[296]. PDZD7 does not appear to be responsible for usherin localization in photoreceptor cells[305], a function carried out by whirlin[307]. Overall, it appears that whirlin would preferentially bind to usherin and PDZD7 to ADGRV1, despite these interactions being partially redundant[288]. The molecular mechanisms responsible for the specificity of such interaction are unknown. Interestingly, PDZD7, but not whirlin, has been shown to negatively regulate the GPCR activity of ADGRV1. This result in restored cellular cAMP levels in cellular models, reduced when the receptor is active through $G_{\alpha i}$ coupling[191]. ADGRV1 and PDZD7 are therefore prime candidates for mediating signalling associated to the ankle link complex.

III – Objectives of the PhD project

The Usher 2 complex is composed of two large transmembrane proteins, ADGRV1 and usherin, and two cytoplasmic proteins, whirlin and PDZD7. These proteins form the fibrous ankle links at the base of hair cell stereocilia and assemble as a dense cytoplasmic network through means of multiple protein-protein interaction domains, including HHD and PDZ domains. This complex potentially serves three functions in hair cells, including mechanical scaffolding of the hair bundle, signalling and regulation of gene expression, as well as transport of molecules along the stereocilia.

The interactions between all four proteins are mediated by multiple binding sites and should be partially redundant both *in vitro* and *in cell*. This is reflected by apparent discrepancies observed in animal models and their modified strains. In order to better understand the physiopathogenesis associated with the Usher 2 complex, there is a need to decipher the network assembly. This goal relies both on continuing the characterization of the protein-protein interactions and the identification of their structural determinants, as well as the study of the general complex architecture *in situ*, which remains elusive.

My project consists in an integrative structural biology approach, making used of biophysical and imaging techniques to characterize the Usher 2 complex at different scales (Figure 44).

Harmonin Homology Domains (HHDs) form a poorly characterized orphan domain family, abundant in scaffolding deafness proteins, but which function is unclear. The first objective of the project is to propose an evolutionary scheme for the various HHDs using bioinformatics tools, associated to a thorough sequence analysis aiming at identifying conserved interaction surfaces between different branches. To complement this study, I present *in vitro* experiments performed on the first HHD of whirlin and the sole HHD of PDZD7.

The second objective of my project is to characterize the molecular mechanisms underlying the negative regulation of ADGRV1 GPCR activity by PDZD7. The first part consists in determining the determinants of interaction between the two proteins, comparing interactions between wild-type proteins and variants carrying human pathogenic mutations. The second part consists in solving the structure of the activated GPCR with primary effectors (G protein, …) to later compare with ternary complexes including PDZD7.

The third objective is to provide a cellular context to the molecular events studied *in vitro*. This includes determining the fine localization of Usher 2 protein components using

high resolution fluorescence microscopy, as well as monitoring changes such as cleavage of ADGRV1 extracellular region associated to the receptor activation in dissected organs of Corti. Lastly, I aimed at setting up the sample preparation procedure for *in situ* structural characterization of the Usher 2 complex in the cochlea using Soft X-ray and cryo-EM tomographies.

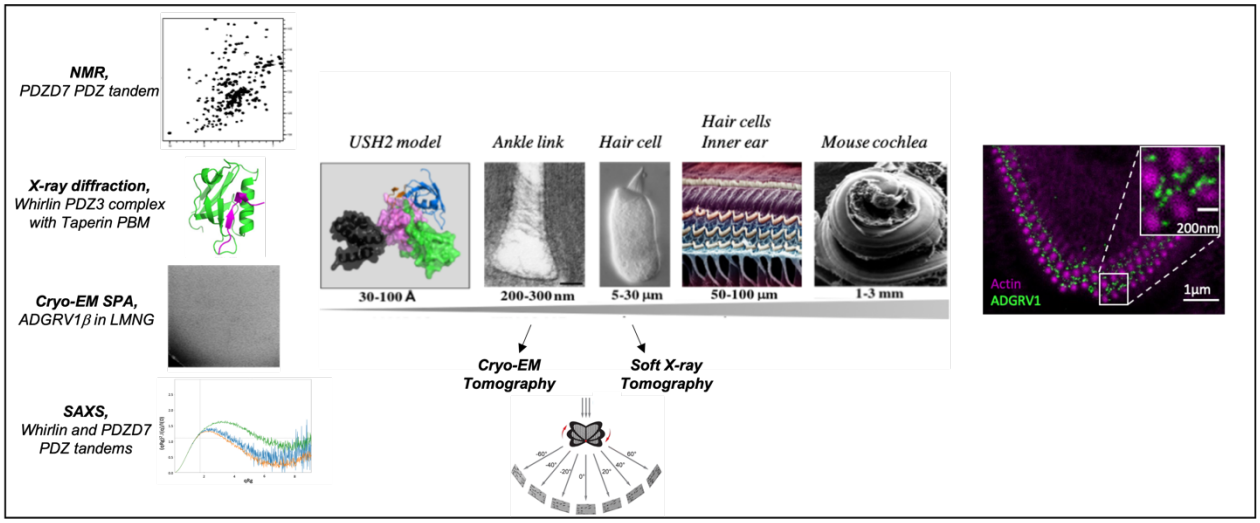


Figure 44 – Scheme of the implemented integrative biology approach.

Results

I – Harmonin Homology Domains

Harmonin Homology Domains (HHDs) are found in the three paralogous scaffold proteins harmonin, whirlin and PDZD7. Through their numerous protein-protein interaction domains (mainly PDZ and HHD), these proteins mediate the assembly of dense protein networks in diverse locations of cochlear hair cell stereocilia and retinal photoreceptor cells connecting cilium. HHDs may contribute in a unique fashion to the assembly of such protein densities. The origin and function of the HHDs, however, is poorly described. As discussed in the introduction, these domains mediate protein-protein and protein-lipid interactions through various surfaces of their helix bundle. As a first step in my project, I aimed at proposing a thorough description of the domain family by means of bioinformatic tools. This includes the identification of potential HHD-containing proteins not previously described, drawing an evolutionary scheme of the domain family and trying to infer the functional interfaces present for each branch of the family. This work was done in collaboration with Julien Guglielmini from the HUB of Bioinformatics of the Institut Pasteur (Paris, France). For the article presented in the next section, supplementary information (SI) is reported in the Annex 1.

I.1 – Article: Phylogenetic analysis of Harmonin Homology Domains

Baptiste Colcombet-Cazenave[1,2], Karen Druart[3], Crystel Bonnet[4,5], Christine Petit[4,5], Olivier Spérandio[3], Julien Guglielmini[6], Nicolas Wolff[1]

1. Unité Récepteurs-Canaux, Institut Pasteur, 75015 Paris, France
2. Sorbonne Université, Collège Doctoral, F-75005 Paris, France
3. Unité de Bio-informatique Structurale, Institut Pasteur, 75015 Paris
4. Unité de Génétique et Physiologie de l'Audition, Institut Pasteur, 75015 Paris, France
5. Institut de l'Audition, Institut Pasteur, INSERM, F-75012 Paris, France.
6. Hub de Bioinformatique et Biostatistique – Département Biologie Computationnelle, Institut Pasteur, USR 3756 CNRS, Paris, France

I.1.a – Abstract

Background – Harmonin Homogy Domains (HHD) are recently identified orphan domains of about 70 residues folded in a compact five alpha-helix bundle that proved to be versatile in terms of function, allowing for direct binding to a partner as well as regulating the affinity and specificity of adjacent domains for their own targets. Adding their small size and rather simple fold, HHDs appear as convenient modules to regulate protein-protein interactions in various biological contexts. Surprisingly, only nine HHDs have been detected in six proteins, mainly expressed in sensory neurons.

Results – Here, we built a profile Hidden Markov Model to screen the entire UniProtKB for new HHD-containing proteins. Every hit was manually annotated, using a clustering approach, confirming that only a few proteins contain HHDs. We report the phylogenetic coverage of each protein and build a phylogenetic tree to trace the evolution of HHDs. We suggest that a HHD ancestor is shared with Paired Amphipathic Helices (PAH) domains, a four-helix bundle partially sharing fold and functional properties. We characterized amino-acid sequences of the various HHDs using pairwise BLASTP scoring coupled with community clustering and manually assessed sequence features among each individual family. These sequence features were analyzed using reported structures as well as homology models to highlight structural motifs underlying HHDs fold. We show that functional divergence is carried out by subtle differences in sequences that automatized approaches failed to detect.

Conclusions – We provide the first HHD databases, including sequences and conservation, phylogenic trees and a list of HHD variants found in the auditory system, which are available for the community. This case study highlights surprising phylogenetic properties found in orphan domains and will assist further studies of HHDs. We unveil the implication of HHDs in their various binding interfaces using conservation across families and a new protein-protein surface predictor. Finally, we discussed the functional consequences of three identified pathogenic HHD variants involved in Hoyeraal-Hreidarsson syndrome and of three newly reported pathogenic variants identified in patients suffering from Usher Syndrome.

Keywords

HHD domains; Profile HMM; Screening; Phylogeny; Usher syndrome; Sequence analysis

I.1.b – Background

Hearing relies on the capacity of highly specialized sensory hair cells to transduce sound-induced cochlear vibrations into electrical signals that are transmissible to the brain. Hair cells possess an ensemble of actin-filled stereocilia tightly structured into staircase-shaped bundles deflected by sound-waves. Multiple extracellular links interconnect stereocilia allowing their concerted deflections and thus the simultaneous opening of mechano-transduction channels. Large protein complexes are found at the anchoring sites of these extracellular links. Mutations in the genes encoding these proteins are responsible for hereditary sensory diseases, notably the Usher syndrome[308–311]. This rare disease, affecting the sensory cells of the inner ear and the retina, is the most common genetic cause of combined congenital deafness and progressive blindness, associated in some cases with balance defects[312].

Proteins encoded by usher genes contain numerous scaffolding domains and protein-protein interaction motifs necessary to the intricacy of the network. The molecular interactions mediated by protein domains underlie the cohesive assembly of the complex, but the network's organization remains elusive. All except one Usher proteins contain short C-terminal motifs that interact with PDZ (PSD95 Dlg1 Zo-1) domains, known as PDZ Binding Motifs or PBM (Figure S1). In contrast, only three proteins contain PDZ domains and have thus been suggested to act as scaffolds for the assembly of the complexes[313–315], each interacting with several partners. These three scaffolding proteins are Whirlin (DFNB31)[292,316,317], Harmonin (Ush1C)[318–320] and the PDZ domain containing protein 7 (PDZD7)[321–323], a deafness protein associated with the Usher syndrome. Interestingly, the three proteins also systematically contain protein-protein interaction domains called Harmonin Homology Domain (HHD). This globular domain was first identified in Harmonin and consists of about 70 residues arranged as a bundle of five helices.

Only nine HHDs have been identified in six proteins so far (Figure 1), including the three Usher-associated and PDZ-containing proteins mentioned before. The post-synaptic scaffolding protein Delphilin (GRID2I) also contains both HHD and PDZ domains. The other two proteins with HHDs are the cerebral cavernous malformations 2 protein (CCM2) and the regulator of telomere elongation helicase (RTEL). The structures of the HHDs from Harmonin and CCM2 proteins have been solved in five hetero-complexes. They exhibit three different binding modes, including inter-molecular[318,324,325] and intra-molecular[87,326] interactions. These structures show that HHDs can bind to isolated amphipathic helices of their partners, but are

also prone to interact with larger domains *via* other surfaces of the alpha-helix bundle, highlighting its plasticity of interaction for binding partners. We previously showed that the intrinsic structural plasticity of the second HHD of Whirlin may play a role in its function, by regulating the interaction with binding partners and promoting its dimerization[269].

Altogether, these results suggest that HHDs are efficient tools to tune protein-protein interactions in rather different biological contexts, through direct binding to the partner and intramolecular modulations of other protein-protein interacting domains, such as PDZ domains, in large modular proteins. In contrast, the variety of proteins in which HHDs are found is surprisingly narrow. Using bioinformatic tools, we have characterized the HHD family diversity and phylogeny from a whole UniProtKB databank screening. We have also reported HHD mutations identified in Usher patients and discussed their functional consequences.

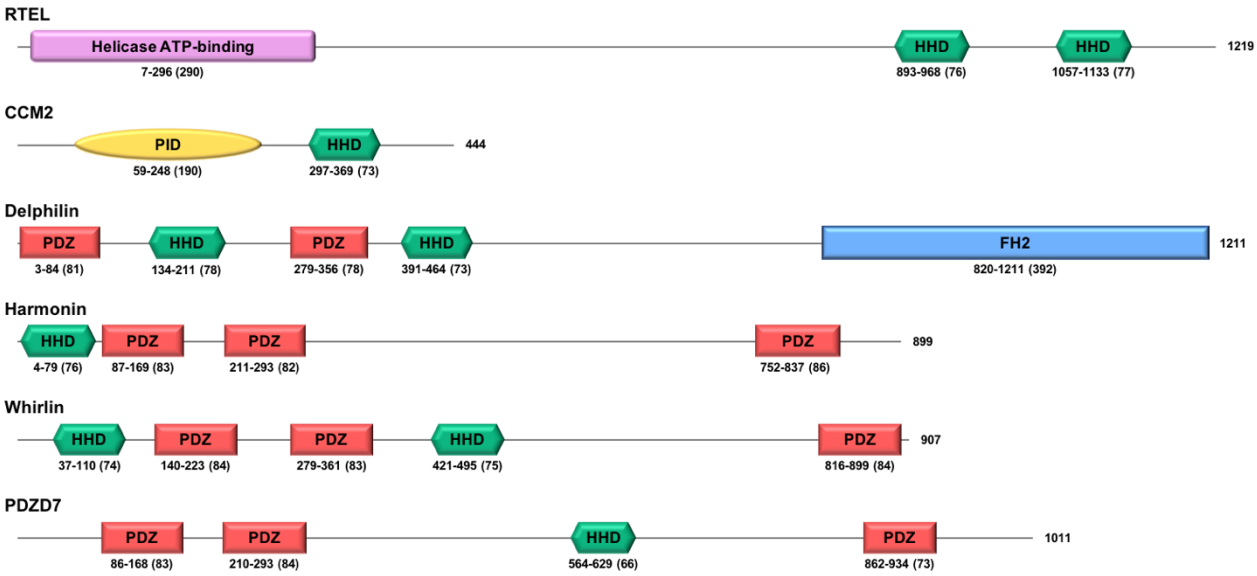


Figure 1- Schematic representation of the six human HHD-containing proteins. Folded domains are highlighted by colored frames. Delimitations of HHDs are derived from our study, delimitations for other domains correspond to the matching UniProtKB entries.

I.1.c – Results and discussion

Structure of Harmonin Homology Domains

Twelve structures of HHDs have been reported so far, covering three out of the nine identified domains and providing information on both free conformations (Figure 2A, E, F, J, K) and complexes (Figure 2B, C, D, G, H, I).

As previously described, the free HHDs of CCM2 (PDB entries: 4YKD & 4FQN), Harmonin (PDB entry: 2KBQ) and Whirlin (second HHD, PDB entries: 6FDD & 6FDE) adopt a similar five-helix fold. The main difference concerns the size of the first helix, CCM2 exhibiting the longest (14 amino acids) and the second HHD of Whirlin being the shortest of the three (10 amino acids). The globular fold of the helical bundle is maintained by the extended hydrophobic core of the domain, assisted by essential electrostatic interactions around the interhelix loops.

The most documented binding mode of HHDs, which can be considered canonical, is reported in Figure 2B, C, G, H and I. It involves an amphipathic helix from the partner that positions itself between helices α1 and α2 of the HHD. Comparing free (Figure 2A) and complexed HHDs of Harmonin (Figure 2B and 2C) and of CCM2 (Figure 2E & F against 2G & H), the helix α1 slightly opens upon binding (RMSD values on Ca and N atoms ranging from 1.3Å to 3.5Å), but the rest of the bundle remains mostly identical (RMSD on Ca and N atoms ranging from 0.8Å to 1.3Å). The interaction is primarily driven by hydrophobic effects involving the HHD's hydrophobic core, but also by hydrogen or ionic bonds between the hydrophilic faces of the helices. As an example, R10 of α1 helix and R31 of α2 helix from Harmonin form ionic bonds with D113 and E111 side chains of Cadherin23, respectively. This mode of binding leads to the interaction in *trans* with the partners (Figure 2B and 2C). The binding in *cis* of a helix can also be observed, as illustrated by a structure from CCM2 (Figure 2I) where a downstream helix from CCM2 folds back to interact with the HHD, competing with the helix from the partner in a regulatory manner[326].

Two other surfaces on the side of the bundle were also observed, providing *cis* (Figure 2D) and *trans* (Figure 2I) noncanonical protein-protein interactions, involved in the scaffolding function of HHDs together with the formation of intramolecular supramodules, as illustrated by the HHD-PDZ complex in Harmonin (Figure 2D) that tunes the binding affinity of the PDZ towards its partners[327,328].

Finally, HHDs can be found in crystals as oligomers (Figure 2F and 2J), or swapped dimer (Figure 2K) with the merged α2/α3 helices of one monomer interacting with the merged α4/α5 helices of the second monomer[329].

Despite a high structural conservation, available structures suggest that HHDs can accommodate various partners *via* different binding sites exposed at the surface of the five-helix bundle.

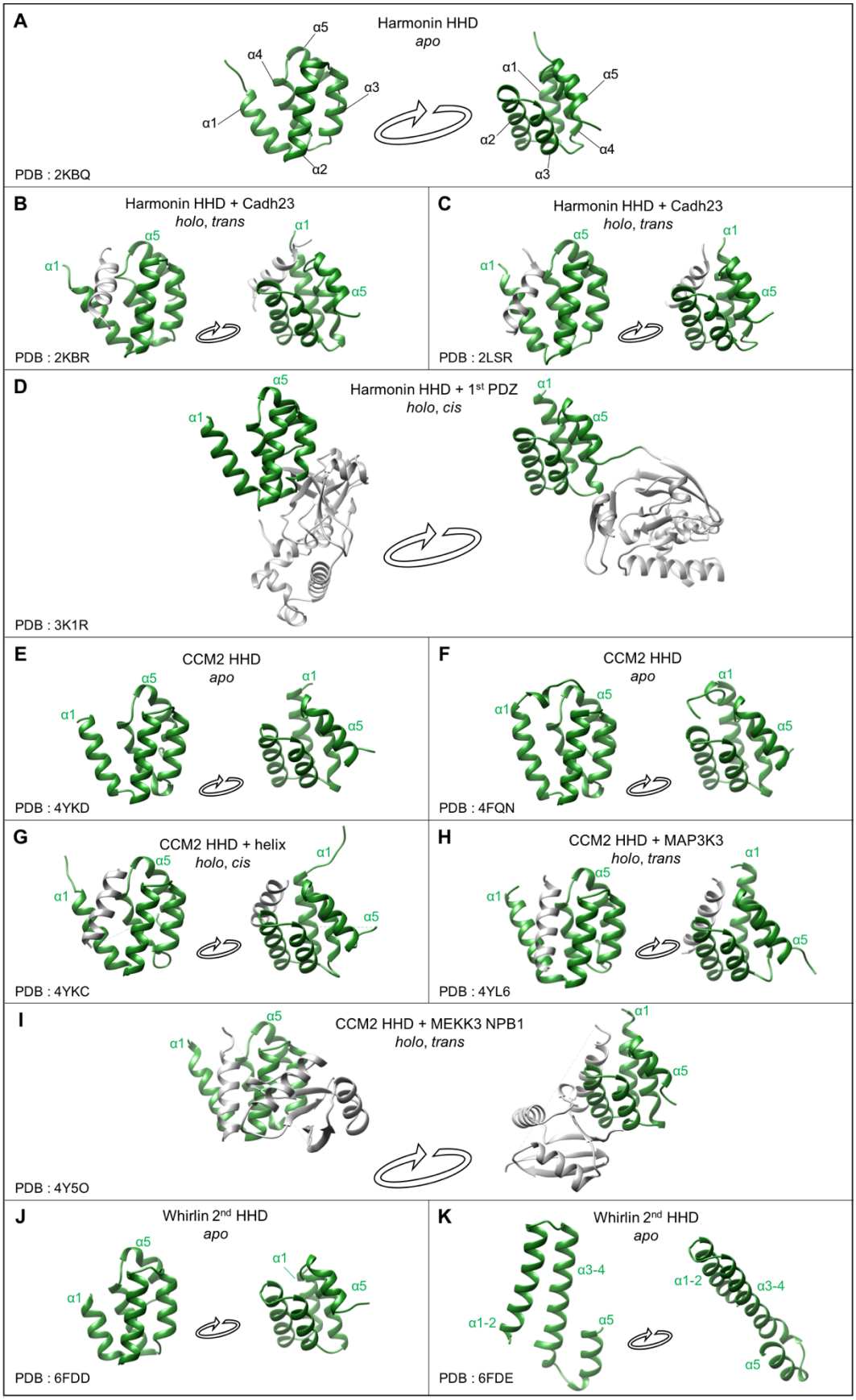


Figure 2 - All HHD structures available for Harmonin, CCM2 and Whirlin in the Protein Data Bank (PDB). Structures of free (2A, E, F, J, K) and complexed (2B, C, D, G, H, I) HHDs. HHDs are colored in green and helical partners or associated domains in gray. The PDB codes are indicated for each structure.

Identification of HHD-containing sequences

As a first task, we aimed at identifying new HHD-containing proteins. The Harmonin-N-like superfamily from the National Center for Biotechnology Information (NCBI; accession cd07347) was used as a starting set of sequences. It consists of the 94 known sequences of HHDs among the six HHD-containing proteins from various organisms. These domain sequences were used as a seed on which we built a profile HMM (Hidden Markov Model) to screen the UniProtKB database[330,331]. Profiles HMM are probability-based models allowing to predict true probabilities of finding each residue at every position of an alignment. These are derived from observed frequencies using a statistical approach, instead of directly using these frequencies for scoring. This provides a model that is less biased towards the starting alignment than frequency-based approaches and thus allows to detect more distant homologues. HMM profiles are adequate tools when using a rather short alignment of sequences as a starting point for a large dataset screening.

Screening of the entire UniProtKB database using our generated profile HMM yielded 2939 domain hits, with E-values lower than 10^{-5} as inclusion threshold. Profile HMM and all sequence alignments are accessible online on http://doi.org/10.5281/zenodo.4255847.
Our second task was to identify full-length proteins corresponding to each domain hit, given that most associated Uniprot entries were not annotated. We ran the BLASTP algorithm on each domain sequence and corresponding full-length protein, in addition to manually checking each individual UniProtKB entry. This process was supported by a clustering approach (Figure 3) based on sequence identity of the domains. For this analysis, we started with the whole set of domain sequences (2939) and ran the clustering algorithm with an increasing identity inclusion threshold.

The following procedure was implemented: 1) We increased the inclusion threshold until a significant cluster is isolated, representing more than 5% of the initial full set of sequences. 2) If annotated sequences appear homogeneous, this annotation is used to name the cluster and we proceed to step 3, otherwise we go back to step 1. 3) We further increased the inclusion threshold until another split happened, with hits in subclusters accounting for more than 10% of the total sequences of the parental cluster. The incremental approach allows to isolate all clusters at their appearing threshold. An alternative display of the clustering is presented Figure 3C. It is the result of a calculation at a single threshold applied to all sequences, thus fragmenting clusters of lower homogeneity (RTEL from insects and PAH).

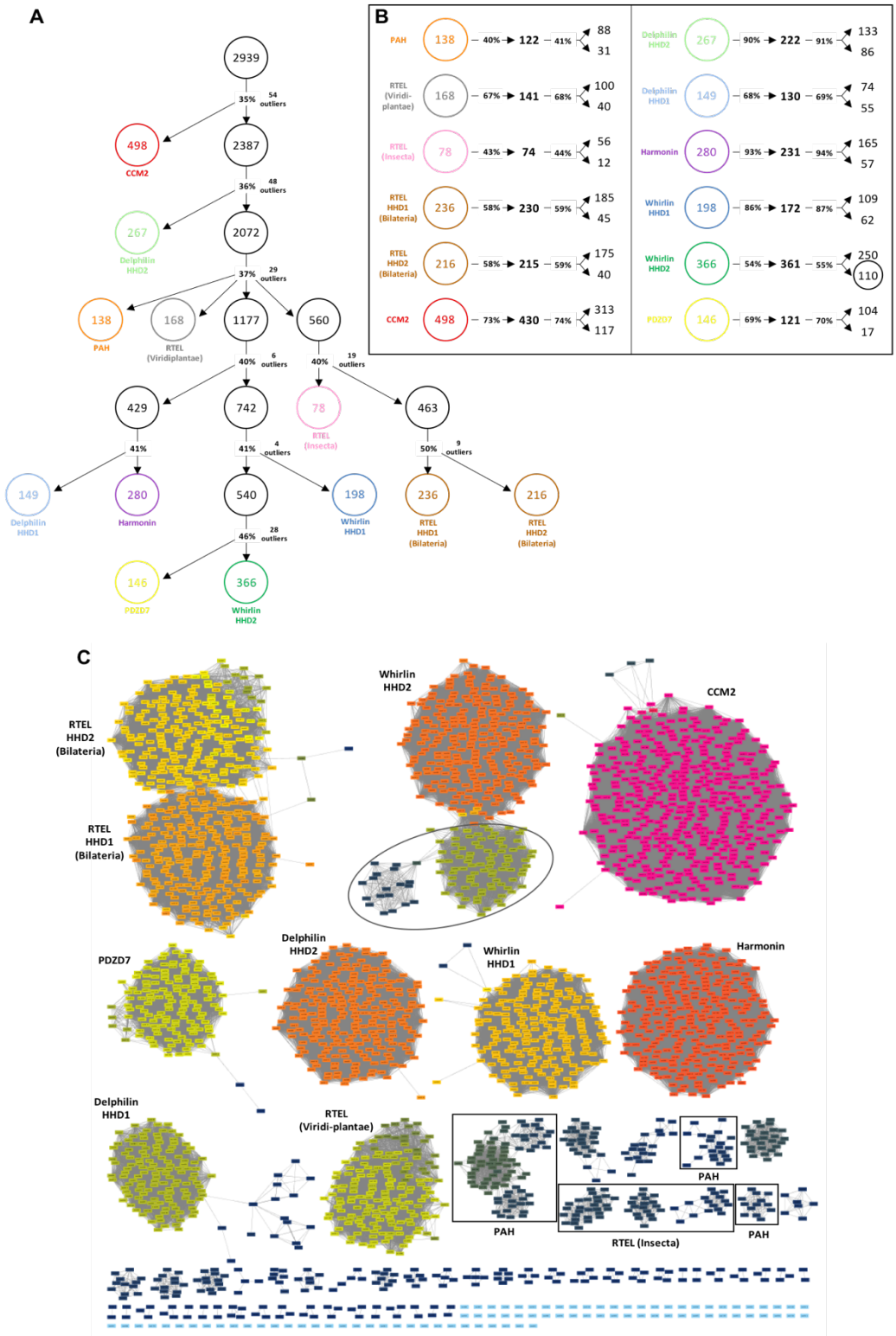


Figure 3 - Clustering analysis of HHD-containing sequences. A: The inclusion threshold (%identity) is incremented from top to bottom. Indicated percentages refer to inclusion thresholds where subclusters emerged, indicated by arrows. Outliers correspond to sequences that did not cluster successfully. The inclusion threshold is further incremented for subclusters with inhomogeneous sequence annotations. B: Each resulting cluster is again submitted to an incrementation of the inclusion threshold. The first percentage value and arrow correspond to the limit identity where only one main cluster remains and point toward the number of sequences in that cluster. The second percentage value and arrows indicate the identity value where multiple clusters accounting for more than 10% of starting sequences emerge, pointing to the sizes of the two main clusters at the given identity threshold. C: EFI – Enzyme Similarity Tool representation using a filter-value of 18. This threshold is determined empirically to differentiate and display all clusters in a single snapshot. The circled cluster highlights sequences from insects, corresponding to the Dyschronic sequences also circled in Figure 3B.

Our procedure allowed to estimate the conservation within each cluster and to identify subclasses of proteins. From our dataset, only 79 domain sequences (out of 2939) arise from proteins without known HHD or that could not be identified (2.7% of the dataset). These sequences did not cluster together and formed a collection of isolated hits that do not make up for a new HHD-containing family.

Interestingly, 5% of the hits (146 over 2939) corresponded to PAH (Paired Amphipathic Helices) domains that satisfied the inclusion threshold and clustered together. PAH domains are protein-protein interaction domains involved in eukaryotic transcription, found in proteins such as Mad1 and Sin3[332]. It consists of four helices instead of five as found in HHDs. However, the similarity between PAH and HHDs, considering the four first helices, is striking both in fold and sequence conservation (Faure, et al., Proteins (2014) (https://doi.org/10.1002/prot.24438))[272], with an overall identity of 12% ± 8%, ranging from 8% between Delphinin-HHD2 and PAH to 17 % between HHD from RTEL of plants and PHA (Figure S2, Table S1). Importantly, PAH also can interact with amphipathic helices through the interface formed by its α1 and α2 helices (ex PDB : 1pd7, 2rms), a binding mode conserved in HHDs. The overall conservation of fold and function suggests the two domains are evolutionary related.

The 2939 sequences are clustered in eight groups shown in Table 1.

Table 1

Proteins associated with the 2939 domain hits

Protein name	Number of domain hits	Phylogenetic coverage
PAH-containing	146	Eukaryota
RTEL	787	Eukaryota
CCM2	490	Metazoa
Whirlin	592	Opisthokonta
Delphilin	413	Metazoa
Harmonin	277	Metazoa
PDZD7	155	Metazoa
Unidentified	79	Eukaryota

HHDs in RTEL

We found 787 hits corresponding to HHDs in RTEL proteins from eukaryotic organisms. The HHD is located at the C-terminal of the protein. In contrast with other HHD-containing protein families, HHDs in RTEL form three clusters corresponding to viridiplantae, bilaterian and specifically insects. Based on our analysis, only vertebrates possess two HHDs in their RTEL proteins, as previously suggested[272], whereas a single HHD is found in other bilaterian, insects or plants.

HHDs in CCM2

We identified 490 sequences corresponding to HHDs in CCM2 proteins, only found in metazoans. For this family, exclusively one domain hit per protein sequence was found, as previously reported[333]. The HHD domain in CCM2 is identified in the C-terminal 200 amino acid region involved in the induction of cell death.

HHDs in Whirlin

We pointed out 592 hits that can be identified as belonging to Whirlin proteins. Two clusters were obtained, one with 198 sequences including the first HHD, and the other with 366 hits including the second HHD of known sequences. However, at an inclusion threshold over 55% identity, this second larger cluster leads to an additional subcluster of 110 hits. Interestingly, these sequences come exclusively from insects, and 75 of them can be identified as a Dyschronic protein, a homolog of Whirlin involved in the synaptic development of flies expressed in major neuronal tracts[334–336]. We can detect only one HHD in these sequences. The N-terminal region of 300 amino acids is predicted as disordered upstream its first PDZ. By contrast, Whirlin has a first HHD in the N-terminal 140 amino acids region upstream its PDZ tandem. The unique HHD of Dyschronic is located between a PDZ tandem, but its HHD is tethered to the third PDZ by a shorter disordered region (about 230 residues instead of 550 in Whirlin). Finally, the Dyschronic protein has 50 additional residues following its third and last PDZ. The two proteins end with a type-I PBM. We hereby predict that the protein referred to as Dyschronic possesses a HHD similar to the second HHD of Whirlin. However, given the confusingly high similarity of Dyschronic to Whirlin, it will be considered as a Whirlin subclass in the rest of the study.

HHDs in Delphilin

Among a pool of 413 hits, two clusters were obtained for Delphilin, one comprising the first HHDs and the other the second HHDs of annotated proteins. Protein sequences only originate from metazoans.

HHDs in Harmonin

A total of 277 hits correspond to HHDs in Harmonin proteins from metazoans. This cluster has the highest degree of conservation, up to 93% identity for cluster core. The highly conserved residues cover almost all the surface of the HHD. As illustrated in Figure 2B, 2C and 2D, it has been shown that the HHD of human Harmonin exhibits several surfaces of binding providing both intramolecular (PDZ domain) and intermolecular (Cadherin23) interactions. We suggest that the highest degree of conservation of HHDs in Harmonin might underlie multiple potential binding interfaces, resulting in an overall well conserved surface.

HHDs in PDZD7

Coming from metazoans only, 155 hits have been identified as PDZD7 proteins. The vast majority of these sequences encode for one HHD domain, however two HHDs are detected in 11 sequences from 6 organisms (*Branchiostoma floridae*, *Crassostrea gigas*, *Lingula unguis*, *Stylophora pistillata*, *Capitella teleta*, *Strongylocentrotus purpuratus*), with an additional hit before the first PDZ, leading to a Whirlin-like organization of domains. This suggests that PDZD7 might have diverged from Whirlin and lost its first HHD. This observation is consistent with the phylogenetic tree (see below) that highlights a low evolutionary distance between the second HHD of the two proteins.

Overall, only the Dyschronic homologue of Whirlin appears as a new HHD-containing protein. All domain hits identified in this study falls into known HHD-containing protein families, with no significant new architecture. We provide an ensemble of sequences to support HHD studies, with an improved classification of HHD families, highlighting 11 main clusters.

Phylogeny

Given the similarity between PAH and HHDs, in terms of conserved residues, fold and canonical binding mode, we hypothesized that PAH and HHDs are evolutionarily related and that the loss or gain of the fifth helix is the evolutionary event to transition from one domain to the other. Using this hypothesis, we decided to root our phylogenetic trees between PAH and HHDs. The various trees we generated based on the full set of sequences presented some variability. This is due to the high number of sequences we compared relatively to their length, resulting in a low signal to anchor each sequence. To overcome this issue, we subsampled each cluster for representative sequences using CD-HIT[337] and successfully obtained a stable tree with statistically supported branching (Figure 4).

The branching between HHD families in the tree is consistent with the phylogenetic groups in which the proteins are found. First, PAH-containing proteins are found in all eukaryotic organisms and consist in a branch opposed to all HHDs, this is a direct consequence of our working hypothesis. Then, are found HHDs from RTEL in Viridiplantae against all HHDs in Opisthokonta. Among those, a branch containing HHDs from RTEL in Bilateria, with a sub-branching for HHDs in Insecta, is opposed to other HHD containing proteins that are not in plants. Finally, the last branch of the tree exclusively supports neuronal proteins and splits CCM2 from PDZ-containing sequences (Delphilin, Harmonin, Whirlin and PDZD7). The fact that the six HHDs arise closely in the tree suggests that they evolved from one ancestral HHD sequence, that was retained only in these neuronal proteins. We note that, at least, the HHD1 domains of Harmonin and Whirlin can form a supramodule organization together with the adjacent PDZ domain that can regulate the inter- and intra-molecular interactions of Usher proteins[325] (also supported by unpublished data). The co-occurrence of HHD and PDZ may indicate a complementarity of function of the domains and suggests that supramodular interactions have been evolutionarily selected in this branch of paralog proteins.

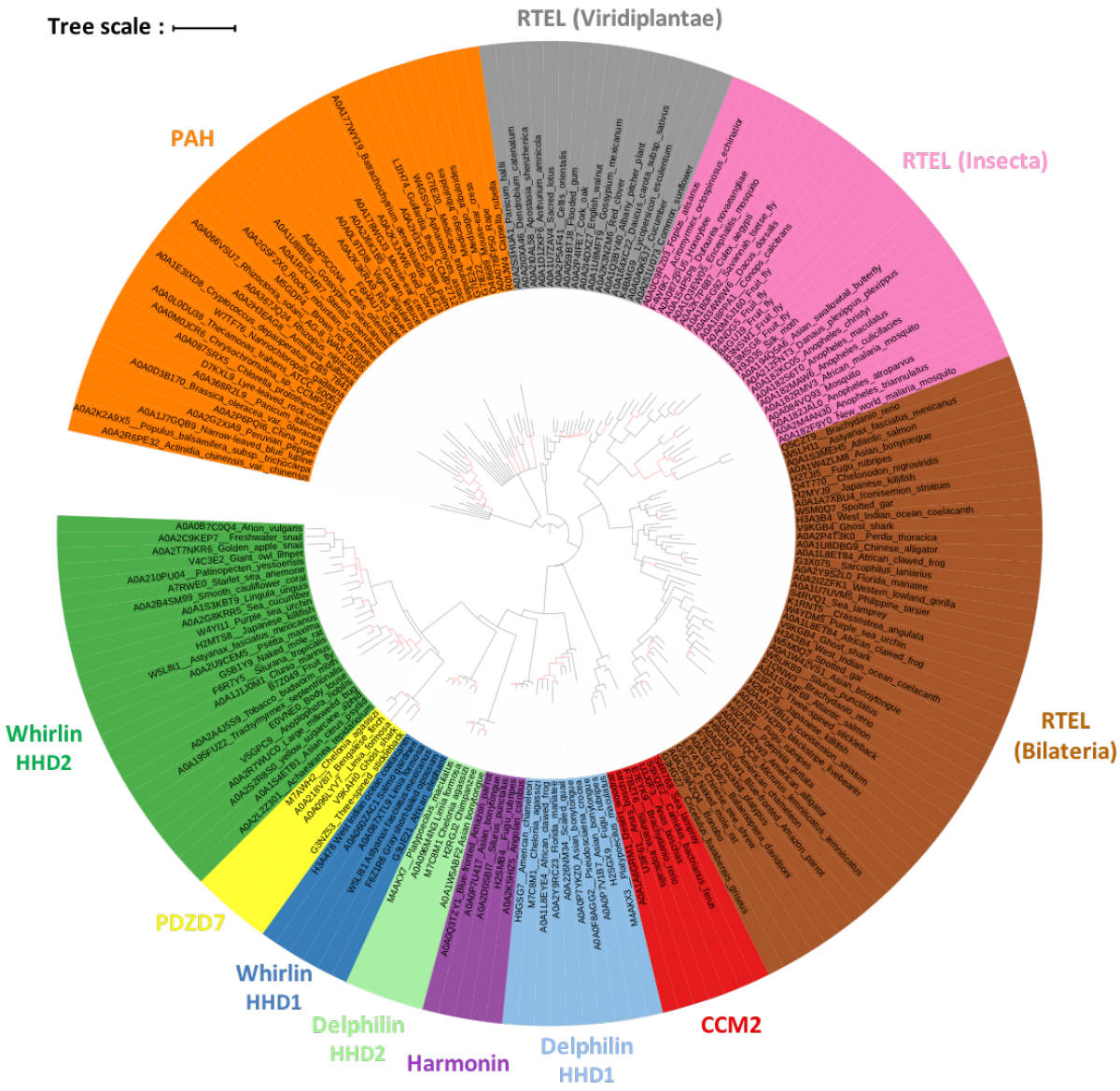


Figure 4 - Phylogenetic tree obtained using subsampled sequences from each previously identified cluster. Branches with Transfer Bootstrap Expectation lower than 70% are depicted in red. Phylogenetic tree can be visualized on the iTOL website (https://itol.embl.de/shared/2PuqbCYlUfiSH).

HHD-family analysis

Heatmap

Phylogenetic analysis based on isolated domains helps to understand the processes promoting protein differentiation, notably through partial (domain) or whole gene (protein) duplication. However, it is not straightforward to conclude on which ensembles of sequences, or branches, are significantly different from one another. Here, we used the BLASTP tool to

score each HHD sequence against every other sequence detected using our HMM approach. Converting the obtained data to an adjacency matrix allows us to cluster sequences into communities. Sequences within one community are then not significantly different. Using this approach, we could group all sequences of Whirlin HHD2 and PDZD7 HHD, Delphilin HHD1 and Harmonin HHD, RTEL HHD1 and HHD2 in Metazoa and the specific cluster from insects, as well as the sequences from PAH and HHDs from plant RTEL. CCM2 HHD, Delphilin HHD2 and Whirlin HHD1 are significantly different from any other HHD cluster (Figure 5).

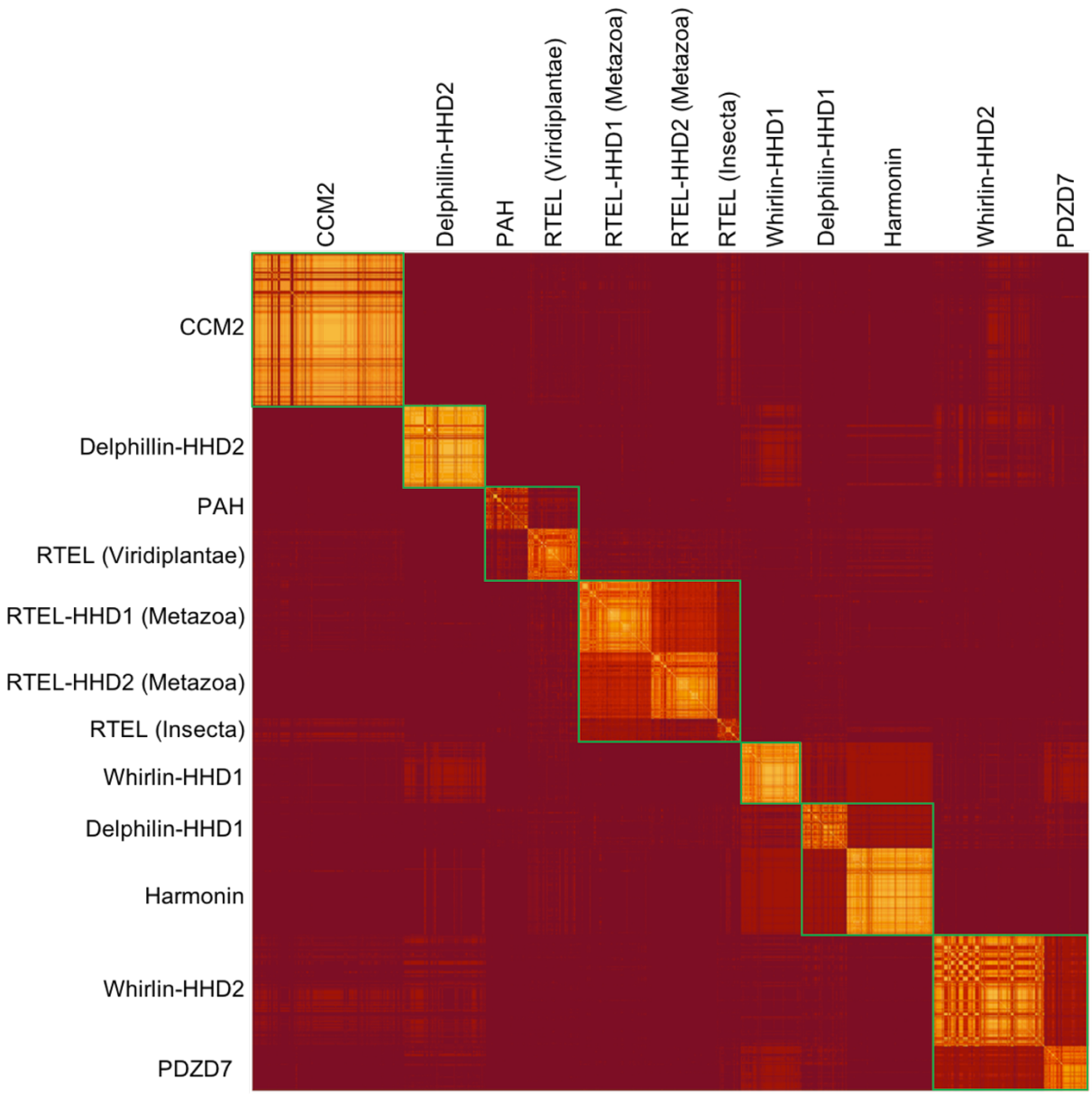


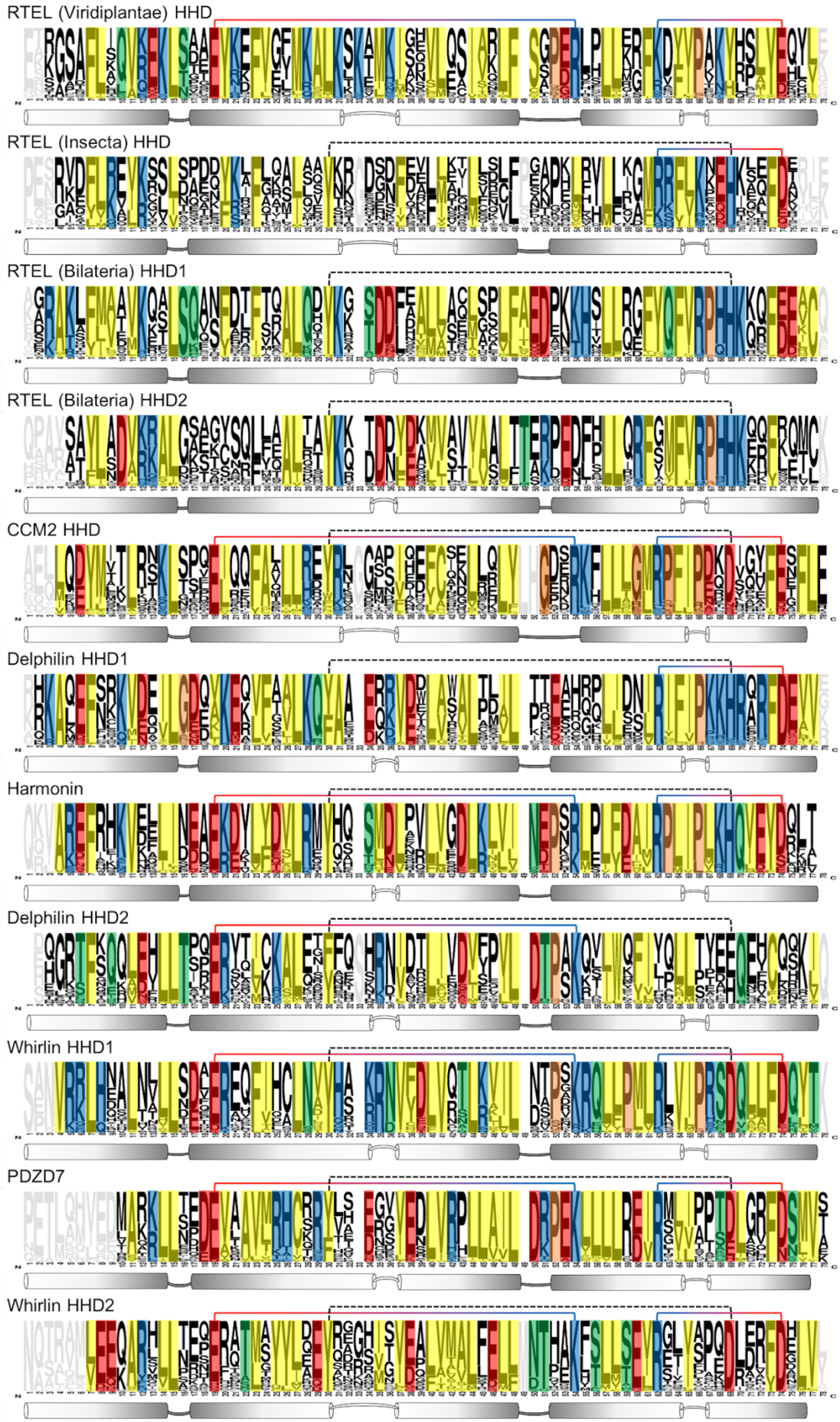
Figure 5 - Heatmap based on BLASTp scores - lighter pixels correspond to pairs with a lower e-value, i.e. sequences of higher similarity. Communities are indicated by green squares.

These results are consistent with the phylogenetic tree (Figure 4) and branches can be regrouped as follow: PAH and HHDs from RTEL-Viridiplantae; HHD1 and HHD2 in RTEL-Metazoa and in RTEL-Insecta; CCM2 HHD; Delphilin HHD1 and Harmonin HHD; PDZD7 HHD and Whirlin HHD2; Delphilin HHD2; Whirlin HHD1.

Motifs analysis

Whole domain comparison, as performed with phylogenetic analysis and blast, can fail at differentiating overall similar domains based on short motifs in the sequences. Therefore, the functional features that are specific to each domain, or rather conserved among them, remain elusive. Unfortunately, the use of automatized motif discovery approaches to identify stretches of residues specific to a family, or conserved from one domain to another, did not lead to results we could decipher. This is most likely due to the high conservation within as well as between clusters, leading to a low signal to noise ratio when searching for local stretches of conserved residues with associated functions. Therefore, we manually calculated the conservation for each position of individual HHDs, mapped on a logo representation of the corresponding clusters. Sequences of each cluster were filtered to reduce global redundancy and increase signal for the detection of position-specific conservation. To discuss the conservation of given positions across all HHDs, we will further use the numbering from Figure 6.

Figure 6 - Logo representation of sequences for each individual HHD, extracted from the overall alignment of all HHDs to maintain column correspondence. Clusters corresponding to each HHD were filtered using CD-hit at 95% identity threshold to remove redundant sequences and increase diversity. Number of remaining sequences: 68 RTEL (Viridiplantae), 54 RTEL (Insecta), 83 RTEL HHD1 (Bilateria), 73 RTEL HHD2 (Bilateria), 78 CCM2, 65 Delphilin HHD1, 23 Harmonin, 42 Delphilin HHD2, 33 Whirlin HHD1, 48 PDZD7, 91 Whirlin HHD2. Grayed positions are represented in less than 80% of the sequence hits, non-colored positions are represented in most sequences (>80%), but are not conserved (<75%) in the final set. Colored positions are conserved in more than 75% of sequences, yellow for hydrophobic, blue for cationic, red for anionic, green for polar and non-charged residues and orange for prolines and glycines. Residues were grouped as follow for conservation calculation: [AVILM], [FY], [TS], [QN], [ED], [RKH], [W], [C], [P], [G]. Recurrent charge-charge and polar interactions are indicated by colored and dashed lines, respectively.


RTEL (Viridiplantae) HHD
RTEL (Insecta) HHD
RTEL (Bilateria) HHD1
RTEL (Bilateria) HHD2
CCM2 HHD
Delphilin HHD1
Harmonin
Delphilin HHD2
Whirlin HHD1
PDZD7
Whirlin HHD2

We identified three stabilizing pairs of positions shared among most HHDs (Figure 7). The residues position 30 and 69, forming a hydrogen bond (Y to charged) or cation-pi interaction (F to cationic), stabilizing the bundle between the top of helices α2 and α5. These two positions are found conserved in all clusters except RTEL from Viridiplantae and the bound is indicated as a dashed line Figure 6. A second pair of residues, positions 19 (anionic) and 54 (cationic) facing each with charge to charge interaction, stabilizes the bundle between the bottom of helices α2 and α4. The two positions are found conserved in 7 of the 12 clusters as indicated Figure 6 by a plain colored line. Similarly, a third pair of residues, positions 62 (cationic) and 74 (anionic), links the α4 to the shorter α5, further stabilizing the helix bundle. This last stabilizing pair is found in 8 of the 12 clusters, indicating by a plain colored line Figure 6.

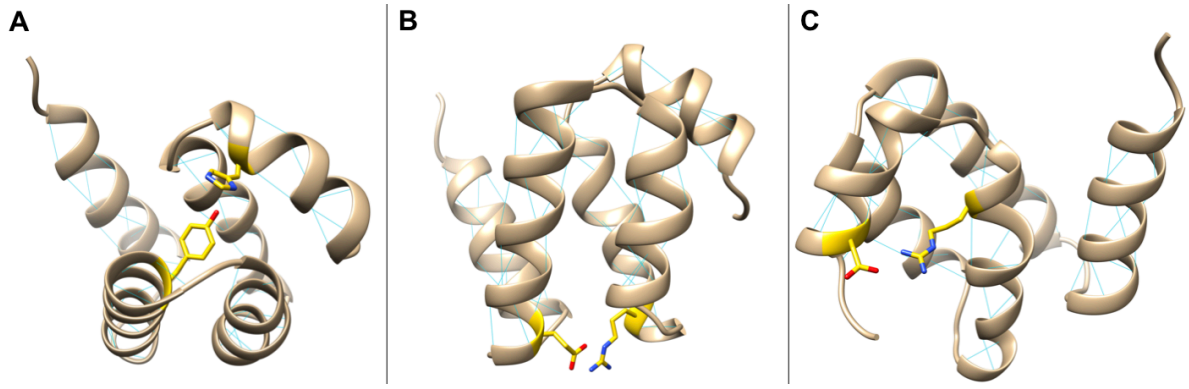


Figure 7 - Pair of residues shared among HHDs to promote bundle stability; highlighted on our model of the Harmonin HHD. A: Positions 30 and 69. B: Positions 19 and 54. C: Positions 62 and 74.

Based on available experimental models, we identified a list of positions in close proximity to the canonical helix binding site that are or could be involved in the interaction with the partner. They can be sorted into two sets, some with side chain participating in the hydrophobic core of the HHD (positions 7, 11, 27 and 61) forming the background of the binding groove and the others located on the α1 and α2 helices and lining the binding groove (positions 4, 5, 8, 12, 17, 20, 21, 24, 25, 28 and 31). The first set cannot be used to discriminate between HHDs, as the four positions are conserved across all clusters with hydrophobic residues, as taking part in the hydrophobic core of the domain. Looking at the conservation of the second set of positions: 7/11 are conserved in the HHD of RTEL from Viridiplantae, 4/11 for RTEL Insecta, 8/11 for RTEL HHD1 Bilateria, 3/11 for RTEL HHD2 Bilateria, 6/11 for CCM2, 8/11 for Delphilin HHD1, 9/11 for Harmonin, 3/11 for Delphilin HHD2, 8/11 for Whirlin HHD1, 3/11 for PDZD7 and 3/11 for Whirlin HHD2. These observations suggest that some HHDs did not evolve to bind to a helical partner. However, using a protein-protein

interaction surface predictor, we could not distinguish the various HHDs in terms of size and shape of the binding site (Figure 8, method submitted for publication). The binding pocket might be too shallow for accurate prediction, with hydrophobic contacts contributing too much to the interaction compared to subtle variations of residues along the binding site determining their specificity and promiscuity. However, the interface was detected unambiguously for every HHD suggesting that all HHDs could accommodate an amphipathic helix as a binder. More experimental data are required to support any further conclusion.

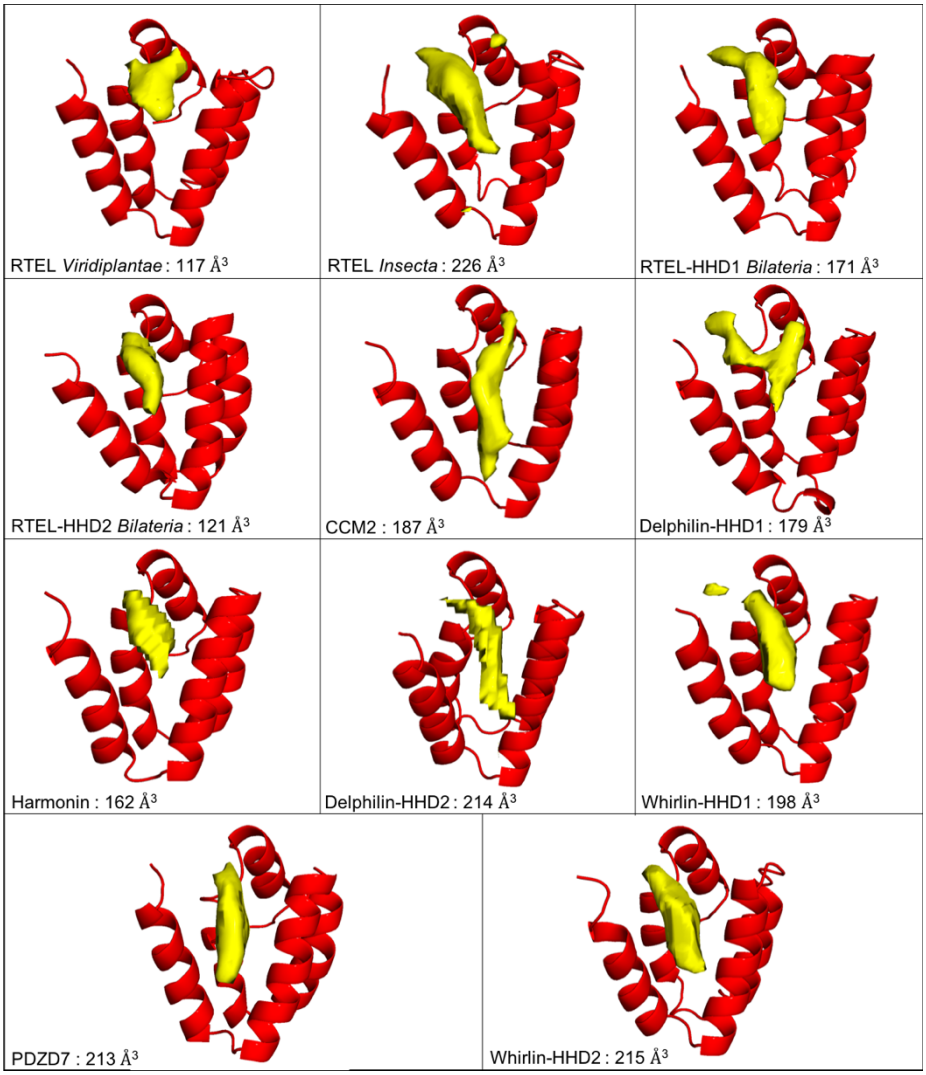


Figure 8 - Prediction of protein-protein interaction volumes for each HHD using DeepPPI-Pocket. The predicted volume is depicted at a probability threshold of 37% and its value is indicated for each individual HHD.

Two other interfaces for protein-protein interactions are observed in crystal structures. The first one is found on the α2-α3 face of the CCM2 HHD[324] (PDB entry 4Y5O). In the crystal structure, the interaction is mediated by an alpha helix from the MEKK3 protein, as discussed so far, as well as its NPB1 domain, which contacts the HHD on this α2-α3 surface. Picking residues from these two helices facing the NPB1 domain rises the following list of

positions: 17, 18, 21, 22, 25, 28, 29, 38, 39, 42, 43 and 46. However, only position 28 is conserved using our criteria, and it is also involved in helix binding. This observation suggests that this interface was not conserved and results from crystal packing, leaving the interaction mainly mediated by the MEKK3 N-terminal helix. This is supported by authors indicating that mutations in the N-terminal helix of MEKK3 severely reduces its ability to pull-down CCM2, while mutations in the NPB1 domain only moderately affect the interaction. The second one consists of the lower segments of α3, α4 and α5 of the Harmonin HHD[328] (PDB entry 3K1R). In the crystal structure, this interface is located at the C-terminus of the HHD and promotes the formation of a supramodule with the following PDZ. Picking positions taking part in this surface of interaction are: 45, 50, 51, 55, 58, 59, 74 and 75. From our study, 7 of the 8 positions are conserved, which is consistent with a functional supramodule affecting the binding of the PDZ to its partners.

Mutations in HHD domains

According to the low number of HHD-containing proteins and their recent discovery, only few mutations associated with human diseases have been described.

The first HHD of the human RTEL protein is affected by three mutations responsible for the Hoyeraal-Hreidarsson syndrome (HHS), a severe form of *dyskeratosis congenita*[338]. The first mutation, K897E (UniProtKB entry Q9NZ71), is positioned in the beginning of the first helix (position 5 of the alignment) and is fully exposed to the solvent. This position is conserved in more than 75% of Bilateria RTEL HHD1 sequences we identified in this study. It is located in close proximity to the canonical binding groove of HHDs and the charge switch induced by the mutation could affect the binding of the partner. However, this cannot be ascertained as no helix binding to the RTEL HHD1 have been identified thus far. The second mutation, R957W, is located in the short loop between helices α4 and α5, corresponding to the position 66 of the alignment. This position is conserved and fully exposed and the function of the arginine or effect of the tryptophan is unclear. The last mutation, F964L is located in the middle of the α5, corresponding to the conserved position 73 of the alignment. This phenylalanine takes part in the hydrophobic core of the HHD, occupying a rather large pocket. This mutation might destabilize the core of the domain by introducing a shorter aliphatic side-chain.

Harmonin, Whirlin and PDZD7 are deafness proteins sometimes associated with the Usher syndrome. We have identified 99 variants in HHD domains of these proteins on the

largest cohort of USH patients studied to date[339]; 37 in Harmonin HHD, 25 in Whirlin HHD1, 22 in Whirlin HHD2 and 12 in PDZD7 HHD. Three of them have been identified as pathogenic, while the 96 others are classified as Variants of Unknown Significance (VUS, complete list per domain Table S2). The first pathogenic variant is the R63W mutant in the Harmonin HHD. It corresponds to the conserved arginine position 62 of the alignment, involved in the stability pair stabilizing helices α4 and α5, as previously described. This mutation may thus impair the stability of Harmonin HHD. The second pathogenic variant is the D447H mutant in Whirlin HHD2. This mutation corresponds to position 28 of the alignment and is not conserved. However, this position is located right next to the binding groove and is directly involved in the binding to the helix in Harmonin and CCM2. This mutation may thus affect the binding to a partner. The last pathogenic variant is the R490H mutant in Whirlin HHD2. It corresponds to the non-conserved position 72 of the alignment, completely exposed to the solvent, and its effect is unclear.

I.1.d - Conclusion

HHDs are found across the whole Eukaryota phylogenetic domain and evolved from a single ancestor, in common with PAH domains. However, only six proteins could be detected in our screening, resulting in a surprisingly narrow variety of proteins encompassing. Adding to these six proteins, we identified a subcluster containing sequences referred to as Dyschronic, the closest homolog to Whirlin in Drosophila. The Dyschronic protein is a component of the circadian output pathway present in the fly central brain. This study led us to identify one HHD domain in this protein similar to the second HHD of Whirlin. Among HHD-containing proteins, the human RTEL protein was reported to possess two HHDs[272]. Here, we highlight that only bilaterians have a two-HHDs RTEL protein, except insects, while RTEL in other organisms only possess one C-terminal HHD. Interestingly, the HHDs of the various RTEL proteins constitute multiple clusters, reflecting consistent differences in sequence that could be associated with the variability of telomeric properties between taxons.

HHDs are involved in intermolecular interactions, both in homotypic (dimer) and heterotopic (HHD-helix, HHD-PDZ) complexes. The small number of domains identified in the UniProtKB databank, and the diversity of molecular recognition characterized up to now prevents us from predicting potential partners. The main function of HHD is likely to promote the formation of large oligomeric complexes with the recruitment of multi-domain scaffold proteins. The reasons are still unclear to explain why the number of identified HHD is limited

despite our extensive search. HHDs are small and well-folded domains and would be good candidates for duplication in proteins to serve as molecular scaffolds for partner interactions. We are currently investigating the role of HHDs in Whirlin and PDZD7 proteins to shed light on the roles of these domains in hearing-associated protein networks. These results will provide clues to decipher the molecular basis of their interactions.

I.1.e - Methods

Sequence alignments

All sequence alignments in this paper were performed using the G-INS-i algorithm from the MAFFT package[340], suitable for sequences containing single domains. Position filtering depending on gap percentage was carried out, when indicated, with Goalign "clean sites" option (https://github.com/evolbioinfo/goalign). This allows to remove positions that are not representative of the overall alignment, but rather of unique sequence features. Alignment quality was assessed using TCS from the T-COFFEE package (doi.org/10.1093/molbev/msu117).

Profile HMM and hits search

Hmmbuild (default parameters) from the package HMMER[341] was used for profile Hidden-Markov Model (profile HMM) building based on sequence alignment. This model consists of probabilities to find a residue at a given position derived from observed counts for each amino acid at this given position, and counts at this position depending on the preceding residue. Hmmsearch from the same package was then used to scan all UniProtKB entries (2018/12/10) for hits with e-values under 10^{-5}, regardless of query coverage (--incdomE 1e-5).

Clustering analysis and detection of communities

Pairwise all against all blast were performed among the 2939 hits using BLASTp from the BLAST+ package[342] (default parameters). Clustering was performed tuning identity cutoff for cluster inclusion using silix[343] (default parameters, only inclusion threshold -i is changed). From the matrix of Blastp scores we extracted a matrix of significant scores by removing all values greater than 1e-4. We took the log of these scores and calculated an adjacency graph using the R package igraph[344]. We then detected the communities in this graph using the

Louvain algorithm[345] from the same package. Resulting communities are depicted by green boxes on the heatmap.

Phylogeny

Prior to tree calculation, we reduced redundancy within the dataset using CD-HIT with a 70% identity threshold on clusters presented Figure 3B, after identity threshold incrementation (-c 0.7 -b 10 -T 6 -n 4 -d 500 -g 1 -M 2000). The Harmonin cluster (231 sequences) was processed separately with 92% identity threshold to obtain more than one representative sequence. The tree in Figure 4 was generated using IQ-TREE v1.6.7.2[346] with automatic extended model selection[347] followed by tree inference. The selected model was LG+R5. We used the option -allnni, set the -pers option to 0.2 and the -nstop criterion to 500 for more thorough searches. The alignment was processed 50 times and only the best tree was retained. Using the same software, model and search parameters, we performed 1,000 bootstrap inferences resulting in 1,000 bootstrap trees. We used these trees to calculate Transfer Bootstrap Expectation branch support values[348] using Goalign.

Motifs

The overall alignment of all HHDs detected in our study was filtered to remove columns containing more than 90% gaps. Sequences gapped for more than 20% of the remaining positions were removed. Subsets of sequences corresponding to each individual HHD were then extracted from the overall sequence alignment of all detected HHDs, conserving position numbering from one subset to the other. We finally used CD-hit[337] with a 95% identity threshold to reduce redundancy within each subset. Logos were finally generated using the online tool WebLogo (logo@combio.berkeley.edu) to represent amino acids frequencies at each position of the alignment for the individual HHDs (*i.e.* subset).

HHD Modeling

HMM-detected sequences to model are aligned to the HHD sequence of the human Harmonin (UniProtKB: Q9Y6N9), using delimitations from a published *holo* structure (PDB: 2KBR). In order to account for possibly shorter predicted sequences due to an increased variability at domain extremities, flanking regions to the HMM prediction are added as necessary to match template length. Models are calculated using MODELLER[349] using the

automodel() function. The best out of 10 generated models (lowest MODELLER objective function value) is retained. For the DeepPPI-Pocket prediction, side chains orientation within the binding groove is optimized using HADDOCK[350] in the presence of the ligand from the template Harmonin structure (PDB: 2KBR).

Protein-protein interaction surface predictor

In order to make the protein-protein interaction surface prediction between HHD and its partners (Figure 8), we have used DeepPPI-Pocket (manuscript under preparation), a deep-learning model based on a double-task FCN (Fully Convolutional Network) capable of predicting the interaction sites of other protein partners or the binding sites of small molecules ligands, from the three-dimensional structure of a protein. This artificial neural network has been trained from known structures of protein complexes and small molecule protein-ligand complexes. Thus, the 3D structure of the protein, here HHD, is placed within a 3D grid and probabilities of binding another protein are predicted for each grid point. The combination of the probabilities allows us to measure a volume for the binding site that matches a chosen probability threshold. Once a specific threshold is fixed (here 0.36) the comparison of the volumes between different systems can be used to gauge the propensity of a protein region to bind another protein.

I.1.f – Additional comments

Using the approach presented in this article, we failed to detect new protein families containing HHDs. This result contrasts with the diversity of function (direct binding to proteins and lipids, modulation of other domain specificity) exhibited by this domain family. This is reflected by a rather high degree of conservation within each branch of the HHD family. For some domains (notably harmonin HHD, whirlin HHD1 and delphilin HHD2), this results in a bundle surface almost entirely conserved, potentially due to a selective pressure induced by several functional interfaces. Predicting individual HHD function without additional information thus proves challenging.

To go further, the study of co-evoluted stretches of residues between HHDs and other protein complex components may provide new insights in the interaction promoted by individual domains. In particular, the enrichment of HHDs in PDZ-containing proteins is striking. Characterizing the co-evolution of the two types of domains may therefore be a first step to continue this study.

The bioinformatics study of HHDs would also benefit from having more experimental data. In the next section, I will describe some experiments performed on the HHDs of whirlin and PDZD7.

I.2 – HHDs in deafness proteins

I.2.a – Whirlin HHD1

The HHD1 of whirlin induces a concentration-dependent oligomerization of all constructs encompassing it, leading to progressive aggregation of the protein. A former PhD student, Yanlei Zhu, has shown that this behavior can be controlled by addition of short amphipathic polymers named amphipols (A8-35), on a construct encompassing both the HHD1 and PDZ1 (N-PDZ1) (Figure 45A). Amphipols have been designed to stabilize membrane proteins in aqueous solution[351]. The obtained sample is monodispersed in solution, as monitored by gel filtration and dynamic light scattering. Subsequent small angle X-ray scattering (SAXS) experiments yielded data compatible with a 114.6kDa molecule, which would correspond to a tetrameric form of the protein construct. In collaboration with another PhD student, Dania Zuniga (Catherine Venien-Bryan's lab, Sorbonne University, Paris), we calculated a model by negative staining electron microscopy that would be consistent with a tetrameric organization of N-PDZ1 (Figure 45B).

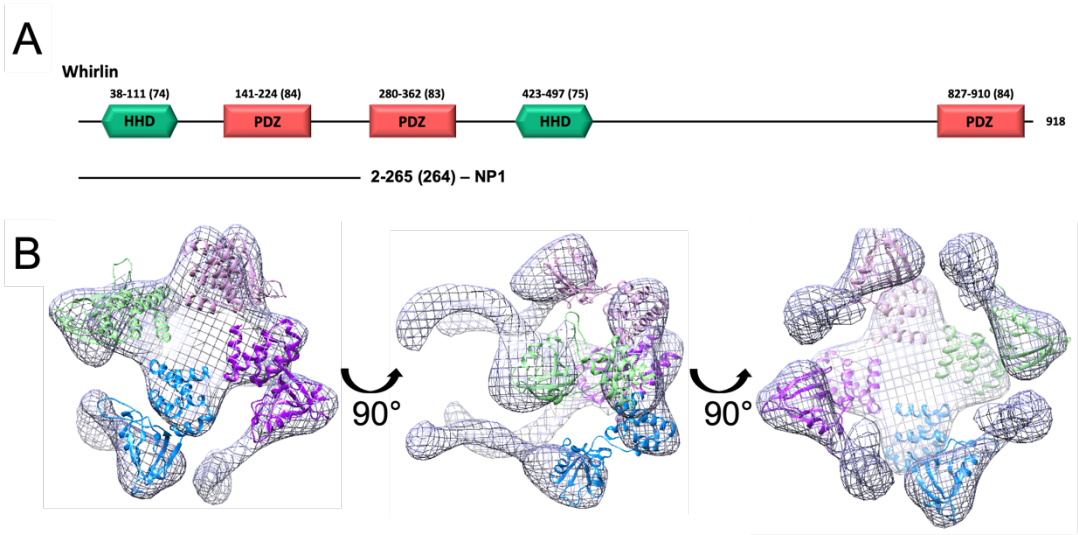


Figure 45 – A) Schematic organization of whirlin. The delimitation of the NP1 construct is indicated by a black line. B) Electron density obtained by negative staining electron microscopy for whirlin NP1 with amphipol. The density would be compatible with the tetramer organization measured with biophysical methods.

Finally, Yanlei Zhu showed that this tetramer retains both the PBM binding and lipid-binding properties (see Figure 35 in the introduction) of the PDZ1. Altogether, these results suggested that whirlin is able to form functional oligomers in a HHD1 dependent manner. The model at low resolution produced for the tetramer, however, did not allow to decipher the detailed molecular organization of the protein, the function of each domain in the oligomer assembly and the role of A8-35 in tetramer stabilization.

In order to answer those questions, we decided to use cryo-electron microscopy to solve the tetramer organization at higher resolution than by negative stain electron microscopy. The sample was prepared at 0.7mg/mL in Tris-HCl pH7.5 50mM, NaCl 150mM, TCEP 0.5mM, with a 1(A8-35):0.5(protein) (w/w) ratio of amphipols. The best grids were obtained using Quantifoil Cu R2/2/, glow discharged for 25 seconds at 15mA. After loading 2.5µL of sample onto the grid, blotting was performed using a Vitrobot (Thermofisher) with 30 seconds wait time and 4 seconds blot time (force 0) at 15°C and 100% humidity. Data collection was performed on a Glacios 200keV microscope equipped with a Falcon3 camera (NanoImaging platform, Institut Pasteur, Paris) in linear mode, yielding 1235 images with maximum resolution between 3.3 and 6.5Å. An objective aperture of 100µm was used. Defocus values ranged from -2µm to -3.5µm. Magnification was of 150k, for a final pixel size of 0.97Å. Total dose used for imaging was $58e^{-}/Å^{2}$. A typical image obtained using these parameters is presented Figure 46.

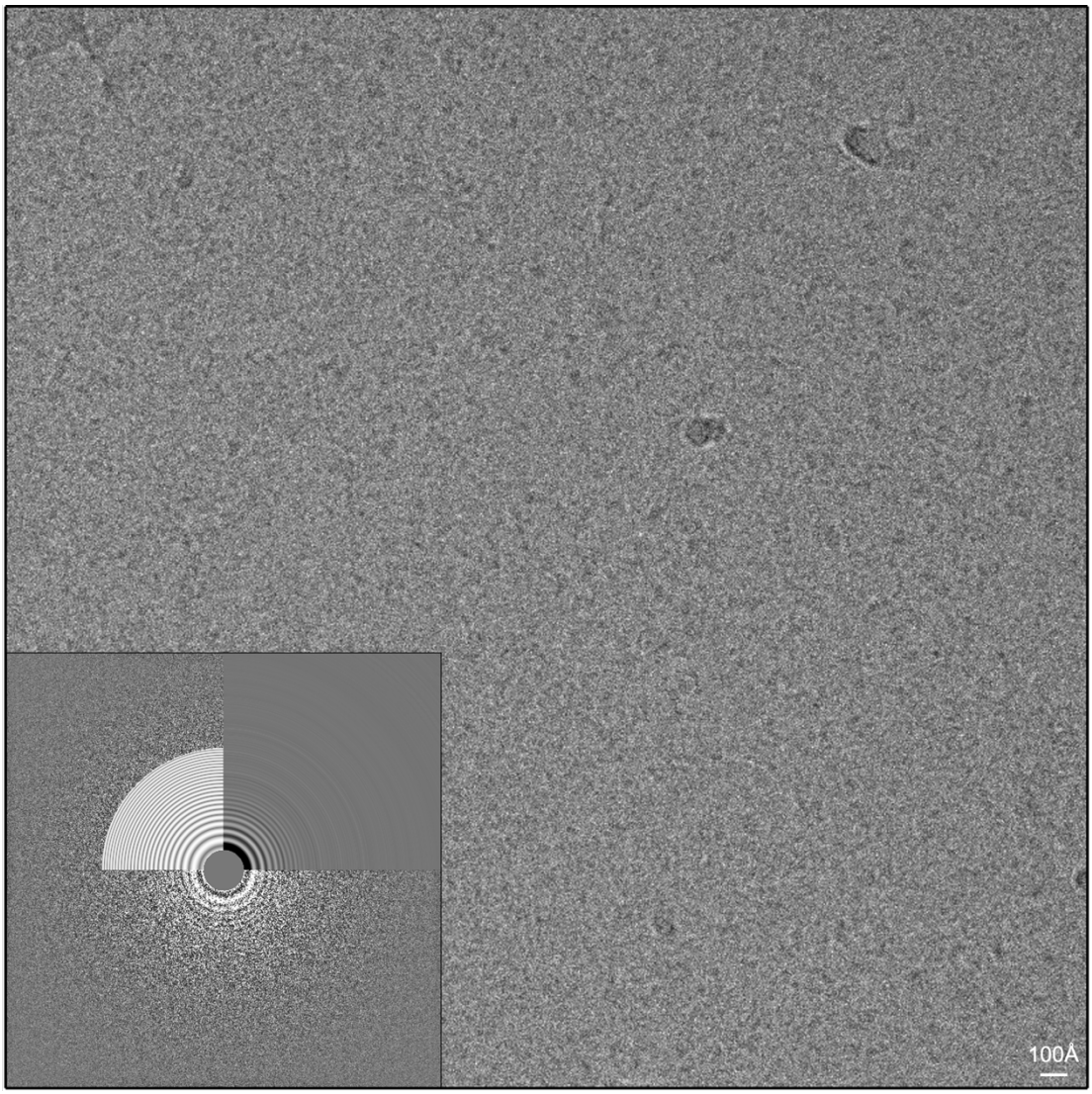


Figure 46 – Typical image obtained for whirlin NP1 on a glacios microscope equipped with a falcon3 camera operating in linear mode. Defocus for this image is of -2.4 microns.

In this image, particles of 80-100Å can be observed, but already appear visually heterogenous. After particle picking (about 1.2 million particles), boxes of 152 pixels (~147.4Å) were extracted and resized to 76 pixels (binning by 2) for 2D classification. Unfortunately, 2D classification did not converge after multiple iterations (Figure 47). For each iteration, the total particle number decreases while the quality of the classification does not improve. Eventually, the number of particles in each class dwindles to a point where signal to noise ratio is too low. An explanation for this result is a high degree of conformational variability, with different objects on the grids. I tried using 3D classification to cluster global conformations before further cleaning orientations by 2D and 3D classification. This approach did not prove successful either. Examples of structured obtained by cryo-EM on a subset of particles, few percents of the total dataset, exist in the literature, though for larger proteins.

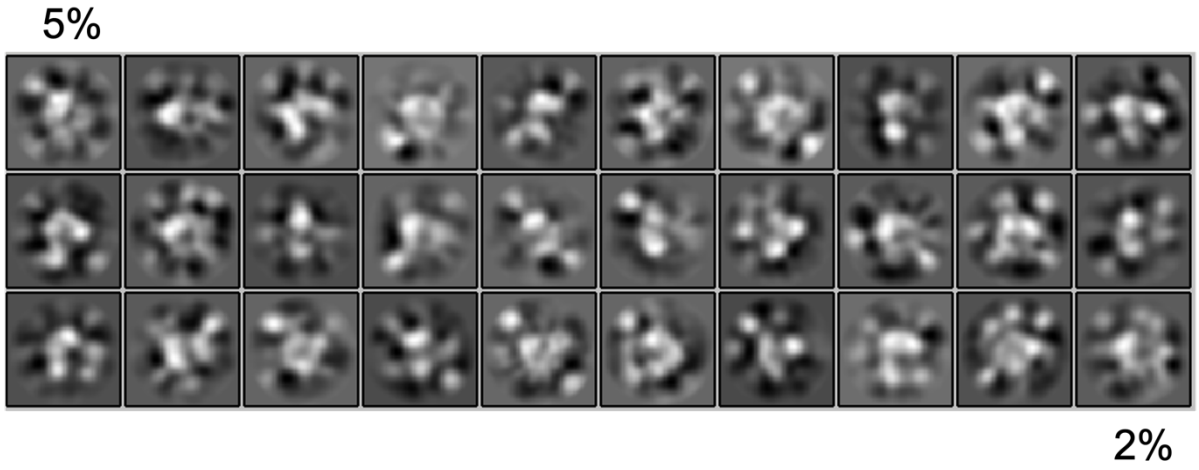


Figure 47 – 2D classification resulting from the particle picking on images such as Figure 46. About 200000 particles are distributed across 30 classes, ranked from the most (5%) to the least (2%) populated.

Our N-PDZ1 construct of whirlin forms particles with too weak contrast in comparison with the potential conformational variability and lack of asymmetry of the tetramer. This may arise from a maintained degree of dynamics of the tetramer, with potential transient tri- and penta- oligomers, in opposition to a stable assembly promoted by amphipols. Additional constraints, such as freezing of the sample, also likely alter the final state of the oligomer on the grids in a heterogeneous fashion.

The concentration-dependent oligomerization property of whirlin HHD1 may play a role in assembling large complexes. This is compatible with the unique environment of hair cells stereocilia, where proteins are secluded in a confined space between a dense actin network and the membrane. Such property may also support the demonstrated role of whirlin in the formation of phase-separated protein condensates at the tip of stereocilia[276]. The second HHD of whirlin may also participate in the protein oligomerization. HHD2 was found to form a swapped dimer in crystallographic conditions[352]. Further experiments suggested that the domain may be in equilibrium between the monomeric and swapped dimer conformations. While *in vitro* experimental conditions largely fail to render the stereocilia's environment, the high local concentrations found in protein condensate may favor otherwise unfavored conformations.

Overall, whirlin appears to be a polarized protein. Its N-terminal region mediates protein oligomerization through both HHDs, as well as protein-protein and protein-lipids interactions by its PDZ tandem, notably with usherin and ADGRV1 in the submembrane compartment. Downstream the HHD2, a large region of the protein (about 300 residues) is flexible and mainly unfolded in solution, exposing the third PDZ domain at its C-terminus.

The PDZ3 mediates interactions with multiple partners, including myosin-binding proteins, reflecting the diverse function of whirlin in cochlear hair cells[34].

I.2.b – PDZD7 HHD

PDZD7 has a single HHD isolated in a predicted unstructured region of the protein. According to the bioinformatics study presented in the preceding section, this HHD encompasses residues 553 to 636 in the mouse sequence (UniprotKB entry E9Q9W7), located about 300 residues downstream the second PDZ domain and 200 residues upstream the third PDZ domain. I showed that based on sequence analysis of the HHD family, PDZD7 HHD is most similar to the two HHDs of whirlin. As discussed in the previous section, whirlin HHDs promote the protein oligomerization *in vitro*. Therefore, we hypothesized that PDZD7 HHD may also promote the protein auto-association.

In order to experimentally test this hypothesis, I expressed a construct corresponding to residues 553 to 636 of mouse PDZD7, encompassing its HHD, fused to a GST (Glutathione-S-transferase) tag at its N-terminal. Expression was carried out at 20°C overnight in *E. coli* BL21 DE3 after induction using 0.2mM IPTG and in M9 minimum media. The composition of the medium is chemically controlled. This allows to exclusively provide isotopically labelled sources of various elements (^{15}N, ^{13}C…) which will be incorporated by the bacteria in translated proteins. Protein labelling is a critical step for Nuclear Magnetic Resonance (NMR), as the active isotopes of protein's atoms are naturally poorly represented (except for ^{1}H). Notably, the natural abundance of ^{13}C and ^{15}N is of about 1% and 0.1% respectively. The counterpart, however, is a reduced biomass compared to rich medium such as LB (lysogeny broth), which may reduce the protein yield. In our case, we used ^{15}N labelled ammonium chloride as unique source of nitrogen. After expression, the protein was purified by a first step of affinity chromatography (GSTrap column, *Cytiva*), followed by cleavage of the tag (TEV protease) and gel filtration (HiPrep 16/60 Sephacryl S-100 HR, *Cytiva*; Figure 48). The final buffer is Tris-HCl pH7.5 50mM, NaCl 150mM, TCEP 0.5mM and anti-protease tablets. The cleaved construct has a "GAMGS" exogenous sequence at its N-terminus.

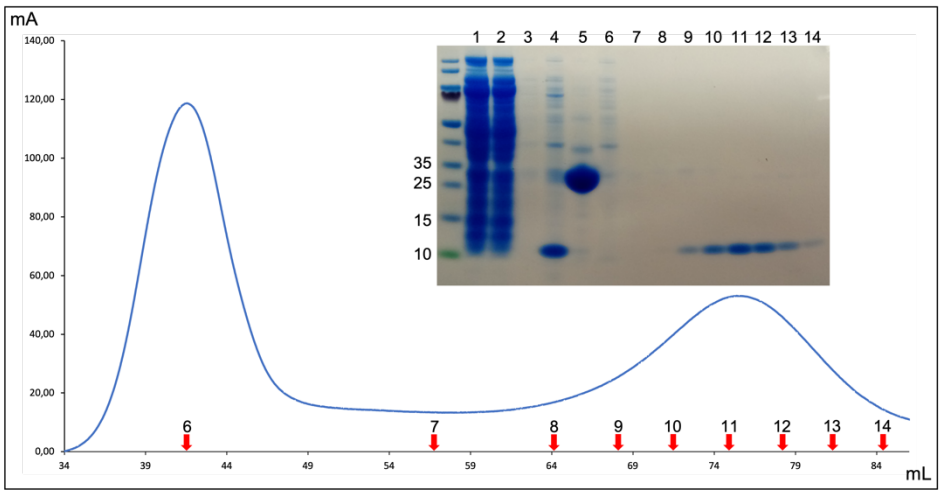


Figure 48 – Gel filtration elution of PDZD7 HHD construct. Relevant positions on the corresponding polyacrylamide gel are indicated on the chromatogram.

According to the specifications of our gel filtration column, a globular protein of 12kDa (cytochrome C) elutes at a volume of 70mL. PDZD7 HHD (10kDa) has an elution volume of about 75mL, strongly suggesting that the protein is mainly monomeric in solution, in such experimental conditions (in presence of TCEP). The final yield for PDZD7 HHD after expression in M9 and purification is of ~0.6mg per liter of culture. Purification of a 2L culture allowed us to obtain an NMR sample at 580µM and record a ^{1}H-^{15}N HSQC spectrum of the protein (Figure 49) on an 800MHz Bruker spectrometer. The obtained spectrum displays scattered peaks characteristic of a folded protein. Further estimation of the protein correlation time by performing NMR relaxation experiments also suggested that PDZD7 HHD mainly behaves as a monomer (rotational correlation time Tc~6.4ns). The function of this domain is therefore unlikely related to the protein oligomerization. This domain most likely promotes interactions with protein partners, yet unknown.

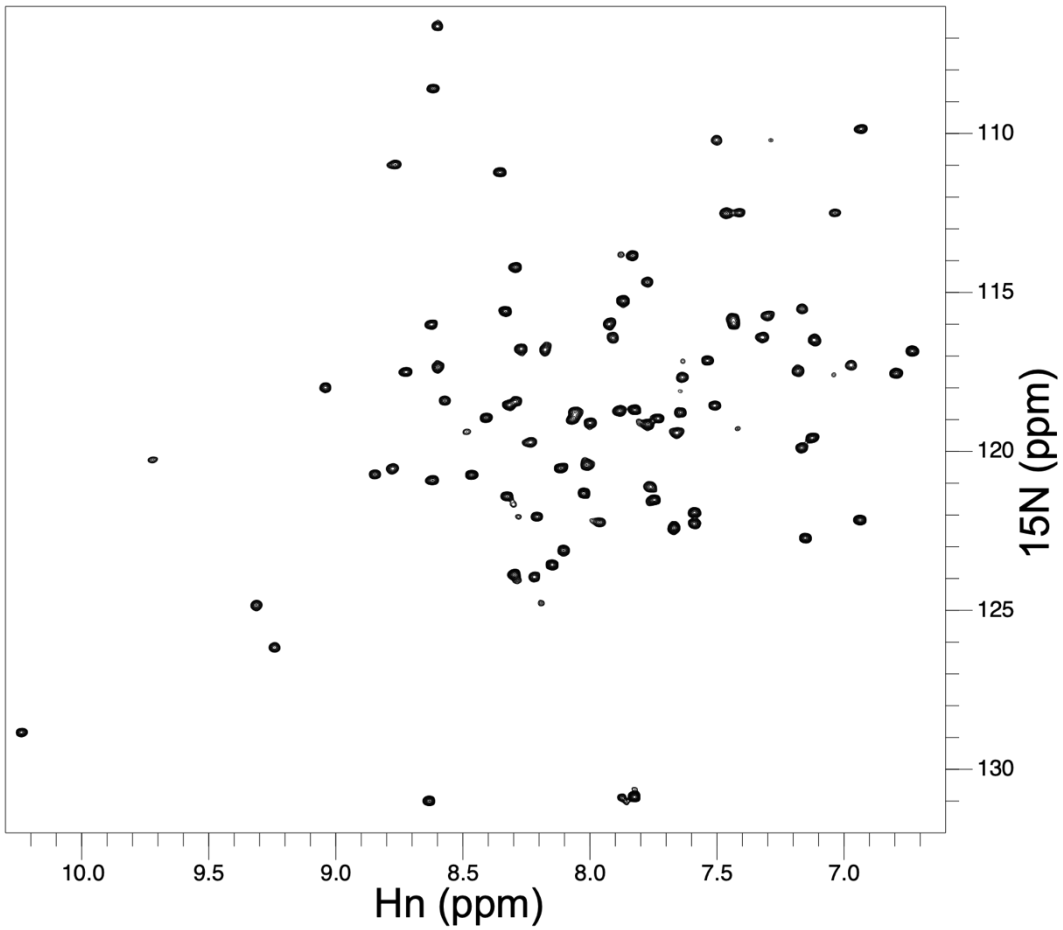


Figure 49 – HSCQ spectrum of PDZD7 HHD recorded at 25°C on an 800MHz Bruker spectrometer.

In the meantime, Lin *et al.*[353] showed last year that PDZD7 HHD is capable to interact with lipids and promotes the membrane targeting of the protein in HEK cells. The association of PDZD7 with plasma membrane is coherent with its interaction with transmembrane proteins, notably usherin and ADGRV1. The next section will present my work on the interaction between PDZD7 N-terminal region and ADGRV1.

II – PDZD7 interacts with ADGRV1β subunit

The second objective of my PhD project was to identify the structural determinants for interaction between ADGRV1 and PDZD7. According to the literature, the interaction requires ADGRV1 PBM and would involve the N-terminal region of PDZD7 without more details. Naturally, PDZ1 and PDZ2 of PDZD7 appear as primary candidates for mediating PBM binding. The first step was therefore to design constructs allowing to decipher which domains participate in the binding.

II.1 – Constructs delimitation

For PDZD7, our constructs were focused on the N-terminal region of the protein. A conserved stretch of residues is observed downstream of the first (PDZ1) and second (PDZ2) PDZ domains of the protein when comparing the sequences of 117 PDZD7 orthologs, using an HMM approach similar to the one described in the previous result section, (Figure 50). As described in the introduction, some PDZ extensions have been characterized as playing an important role in the domain function, with multiple examples in hearing proteins.

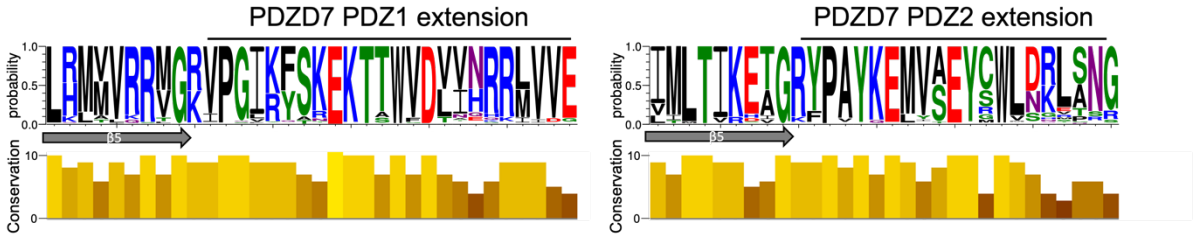


Figure 50 – Consensus and conservation for the extensions of PDZD7 PDZ1 and PDZ2 domains.

Therefore, we designed multiple constructs that would allow us to decipher the function of the N-terminal PDZ domains of PDZD7, isolated, with their extension and as a PDZ tandem (Figure 51A). We also designed constructs including the first N-terminal 80 residues found upstream of PDZD7 PDZ tandem. This region is predicted as flexible and unfolded, but may yet contribute to the interaction between PDZD7 and ADGRV1. For expression in bacterial system, constructs were cloned in pHTP11 vector (NZYtech) with a N-terminal fused GST-tag, followed by a linker containing a TEV protease cleavage site. For experiments in cellular models, constructs corresponding to the full-length protein were cloned in the mammalian expression vector pMT3, with a N-terminal fused strep-tag and a C-terminal fused GFP. For consistency with experiments planned in animal, amino acid sequences

corresponding to the mouse PDZD7 (UniprotKB entry E9Q9W7) have been used. The PDZ tandem of PDZD7 from mouse and human have 97% identity. The few substitutions (7 in 233 residues) mainly maintain the residue properties (size, charge, hydrophobicity) and are not located near the PDZ binding grooves nor PDZ extensions. Comparison with the human protein is therefore relevant.

Interestingly, four human pathogenic variants of PDZD7 have single-point mutations in the PDZ tandem region of PDZD7 (R66L, G103R, G228R and M285R). In particular, two variants have the same glycine to arginine substitution, one located on the β2 strand of the PDZ1 (G103R) and the other on the β2 strand of the PDZ2 (G228R). Due to their position lining the binding groove and type of substitution, we hypothesized that they likely affect the binding properties of the PDZ domains and potentially affect the interaction with ADGRV1. The article recently published *in Frontiers Mol. Sci.* presented in the next section will describe the role of each PDZD7 N-terminal PDZ domains in binding to ADGRV1, the importance of their C-terminal extensions and the effect of the two pathogenic mutations located in their respective binding groove.

For *in vitro* binding assays involving ADGRV1, we used a peptide corresponding to the last 13 residues of the receptor cytoplasmic domain (VELRRIPIADTHL$_{COOH}$), as well as the whole cytoplasmic domain of 152 residues (Figure 51B). As mentioned in the introduction, ADGRV1 cytoplasmic domain is largely unfolded and flexible as monitored by NMR, with stretches of conserved residues of unknown function. Comparing the last 13 residues and the entire cytoplasmic domain should allow to identify PBM and non-PBM contributions to the interaction with PDZD7. The peptide was synthesized by the *ProteoGenix* company. The cytoplasmic domain was cloned in the prokaryotic expression vector pHTP1 (NZYtech company) for expression in fusion with a N-terminal His-tag (6xH). For cellular experiments, we used a construct corresponding to ADGRV1β subunit, from the stachel peptide sequence ($_{NH2}$SVYAVYA-) to the C-terminus of the protein (-IPIADTHL$_{COOH}$). This construct was cloned in the mammalian expression vector pRK5, without tag or with a C-terminal FLAG tag. Again, for consistency with experiments planned in animal, amino acid sequences corresponding to the mouse ADGRV1 (UniprotKB entry B8JJE0) have been used. The ADGRV1β subunit and cytoplasmic domain from mouse and human have 90.4% and 92.3% identity, respectively. No substitution is found in the C-terminal 13 residues.

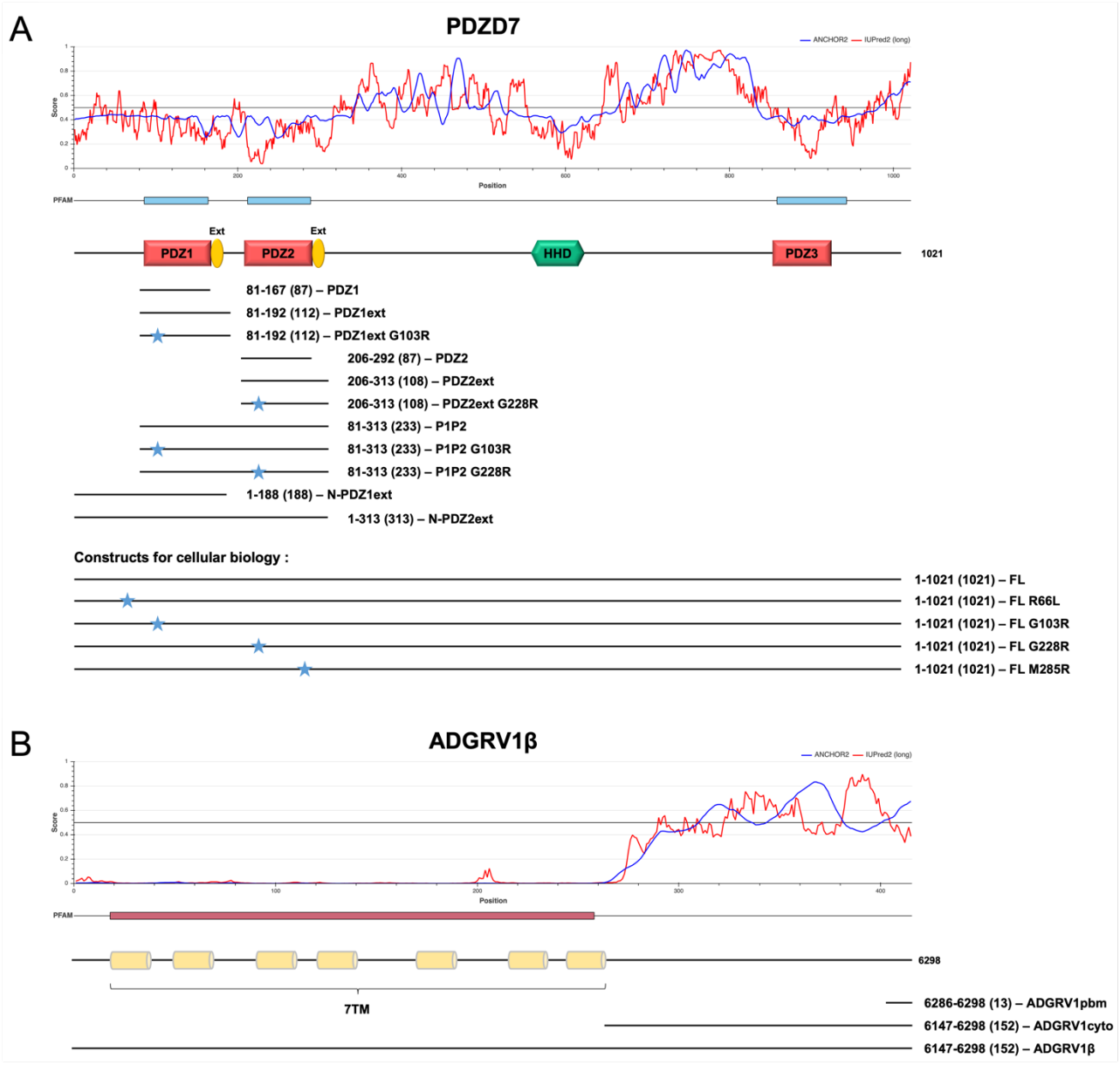


Figure 51 – A) Iupred disorder predictor mapped on a schematic representation of PDZD7 modular organization. A score above 0.5 is associated to favored disordered conformations, while a score below 0.5 is associated to more ordered globular folds. The delimitations of all constructs used for PDZD7, including pathogenic variants, are indicated by black lines. B) Iupred disorder prediction on ADGRV1 7TM and cytoplasmic domain. Constructs used in this study are indicated by black lines.

II.2 – Article: Deciphering the molecular interaction between the adhesion G protein-coupled receptor ADGRV1 and its PDZ-containing regulator PDZD7

Baptiste Colcombet-Cazenave[1,4], Florence Cordier[2,5], Yanlei Zhu[1], Guillaume Bouvier[2], Eleni Litsardaki[1], Louise Laserre[1], Marie S. Prevost[1], Bertrand Raynal[3], Célia Caillet-Saguy[1], Nicolas Wolff[1]

[1]Channel Receptors Unit, UMR CNRS 3571, Institut Pasteur, Université de Paris, 75015 Paris, France.
[2]Structural Bioinformatics Unit, CNRS, UMR 3528, Institut Pasteur, Université de Paris, Paris, France.
[3]Molecular Biophysics Platform-C2RT, CNRS, UMR 3528, Institut Pasteur, Université de Paris, Paris, France.
[4]Complexité du Vivant, Sorbonne Université, 75005 Paris, France.
[5]Biological NMR and HDX-MS Technological Platform, CNRS, UMR 3528, Institut Pasteur, Université de Paris, Paris, France.

1. Channel Receptors Unit, UMR CNRS 3571, Institut Pasteur, Université de Paris, 75015 Paris, France Sorbonne Université, Collège Doctoral, F-75005 Paris, France
2. Structural Bioinformatics Unit, UMR CNRS 3528, Institut Pasteur, Université de Paris, Paris, France Unité de Génétique et Physiologie de l'Audition, Institut Pasteur, 75015 Paris, France
3. Molecular Biophysics Platform-C2RT, UMR CNRS 3528, Institut Pasteur, Université de Paris, Paris, France
4. Complexité du Vivant, Sorbonne Université, 75005 Paris, France
5. Biological NMR and HDX-MS Technological Platform, UMR CNRS 3528, Institut Pasteur, Université de Paris, Paris, France

Keywords

Deafness; G protein-coupled receptor; PDZ-containing protein; ADGRV1; PDZD7

II.2.a – Abstract

Hearing relies on the transduction of sound-evoked vibrations into electrical signals, occurring in the stereocilia bundle of inner ear hair cells. The G protein-coupled receptor (GPCR) ADGRV1 and the multi-PDZ protein PDZD7 play a critical role in the formation and function of stereocilia through their scaffolding and signaling properties. During hair cell development, the GPCR activity of ADGRV1 is specifically inhibited by PDZD7 through an unknown mechanism. Here, we describe the key interactions mediated by the two N-terminal PDZ domains of PDZD7 and the cytoplasmic domain of ADGRV1. Both PDZ domains can bind to the C-terminal PDZ binding motif (PBM) of ADGRV1 with the critical contribution of atypical C-terminal β extensions. The two PDZ domains form a supramodule in solution, stabilized upon PBM binding. Interestingly, we showed that the stability and binding properties of the PDZ tandem are affected by two deafness-causing mutations located in the binding grooves of PDZD7 PDZ domains.

II.2.c – Introduction

Cochlear hair cells are the primary receptors for the detection of sound-induced vibrations. Each cell possesses a bundle of actin-based stereocilia organized in rows of increasing height. The shape of the bundle is critical for the cohesive deflection of the stereocilia necessary for the cell function. Morphogenesis of the hair bundle is regulated during cochlear development, but the underlying signaling pathways remain elusive. The ankle link protein complex, transiently expressed at the base of the stereocilia, is the main candidate for regulating bundle morphogenesis, with pathogenic variants of its composing proteins leading to defects in stereocilia organization and impaired auditory function. The genes encoding these proteins are affected by mutations responsible for hereditary sensory diseases, notably the Usher syndrome type 2 (Usher2) that combined congenital deafness and progressive blindness. The protein complex is thus referred to as Usher2 complex. It encompasses two transmembrane proteins, usherin and ADGRV1, exposing very large N-terminal extracellular regions (5011 and 5879 residues in human, respectively) thought to form fibrous links in between stereocilia. Usherin is a single-pass transmembrane protein of largely unknown function. It contains multiple laminin and fibronectin type domains that are likely promoting extracellular matrix adhesion and local scaffolding. ADGRV1 (Adhesion G-protein Receptor V1) is a seven-pass transmembrane protein of the adhesion G protein-coupled receptor (GPCR)[177] family B[178]. Its extracellular region contains multiple protein-protein and

calcium interaction domains, supporting potential calcium-dependent scaffolding properties. Additionally, the extracellular region of ADGRV1 contains a GAIN domain located upstream of the first transmembrane segment, comprising a G protein receptor Proteolytic Site (GPS) triggering autoproteolysis of the protein. Self-catalyzed hydrolysis of the adhesion GPCR at the GPS cleaves the large extracellular domain from the GPCR β subunit. Both fragments might remain non-covalently bound until another trigger, such as a rearrangement of the extracellular region, disrupts the interaction[188]. The new N-terminal end of the protein then consists of an extracellular Stachel peptide[354] thought to act as an agonist and constitutively activate the GPCR[355], leading to a decrease of cellular cAMP through the $G_{\alpha i}$ signaling pathway[191]. The ADGRV1 cleaved β subunit possesses a 152 residues long cytoplasmic domain with a C-terminal PDZ Binding Motif (PBM) promoting the interaction with the PDZ-containing proteins whirlin and PDZD7[191,356]. These two cytoplasmic proteins from the Usher 2 complex in turn associate with various partners such as actin binding proteins[34,306,322]. The formation of the PBM dependent PDZD7-ADGRV1 complex downregulates the GPCR activity and restores cAMP levels in cellular models[191]. However, the structural determinants mediating PDZD7-ADGRV1 complex formation remained undetermined. Furthermore, multiple pathogenic variants of PDZD7 have been identified in patients suffering from Autosomal Recessive Non-Syndromic Hearing Loss (ARNSHL)[296,298–300,302–304,357]. Interestingly, two human pathogenic missense variants, consisting of the same amino acid substitution, are located at the same position in the binding groove of the first (G103R) and second (G228R) PDZ domains of PDZD7[299]. Here we deciphered the contribution of the N-terminal PDZ region of PDZD7 and the cytoplasmic domain of ADGRV1 in the PDZD7-ADGRV1 interaction using complementary biophysical approaches. We highlighted the role of proximal PDZ extensions that tune the affinity towards ADGRV1. The two PDZ domains and their extensions form a supramodule in solution, able to simultaneously bind ADGRV1 PBM. Finally, we characterized the effect of the two deafness associated variants on the protein stability and on complex formation, disrupting the interaction between full-length PDZD7 and ADGRV1 β subunit, underlying potential defects in hair bundle morphogenesis.

II.2.d – Material and Methods

Protein expression and purification

All PDZD7 constructs are based on the murine long isoform (Uniprot entry E9Q9W7-1); PDZ1 (81-167), PDZ1ext (81-192), PDZ2 (206-292), PDZ2ext (206-313), P1P2 (81-313),

N-PDZ1ext (1-188). All were expressed in fusion with a N-terminal GST tag followed by a Tobacco Etch Virus (TEV) protease site. Protein production is performed in *E. coli* BL21 DE3 in LB or M9 minimal medium supplemented with $^{15}NH_4Cl$ and/or ^{13}C-glucose, induced with 0.2mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and performed overnight at 18°C. Lysis is performed in 550mM NaCl, 2mM β-mercaptoethanol, 250mM EDTA and complemented with anti-proteases (Roche Diagnostics) (buffer and pH as indicated for the respective experiment) using a Constant Systems Ltd cell disruptor at 1.3kbar. After 1h centrifugation at 19000G, tagged proteins are purified from the supernatant by affinity chromatography using a GSTrap (Cytiva) column followed by overnight (on-column) TEV cleavage. The cleaved proteins are eluted in 300mM NaCl, 0.5mM tris(2-carboxyethyl)phosphine (TCEP), 250mM EDTA complemented with anti-proteases (buffer and pH identical to the lysis step). A last step of purification is performed by size-exclusion chromatography (SEC) using a Sephacryl S-100 HP 16/60 (Cytiva).

ADGRV1 cytoplasmic domain is based on the murine long isoform (CCDS36737; 6149-6298) and is expressed in fusion with a N-terminal Histidine tag followed by a TEV protease site. Protein production is performed in *E. coli* BL21 DE3 in LB, induced with 1mM IPTG for 3 hours at 30°C. Lysis is done in Tris HCl pH 7.5 50mM, NaCl 300mM, 2mM β-mercaptoethanol and complemented with anti-proteases using a Constant Systems Ltd cell disruptor at 1.3kbar. After 1h centrifugation at 19000G, the protein is purified from inclusion bodies retrieved in the insoluble fraction. Solubilization is performed using 6M guanidine hydrochloride and the resulting fraction is loaded on a HIStrap (Cytiva) column. Guanidine hydrochloride is gradually washed using Tris HCl pH 7.5 50mM, NaCl 150mM, 0.5mM TCEP complemented with anti-proteases. The His-tag is removed by overnight (on-column) TEV cleavage. The cleaved protein is subsequently eluted and a last step of purification is performed by SEC using a Sephacryl S-100 HP 16/60 (Cytiva).

Nuclear Magnetic Resonance (NMR)

All experiments were recorded at 25°C on a Bruker 800MHz spectrometer equipped with a triple resonance $^{1}H\{^{13}C/^{15}N\}$ cryoprobe. The ^{1}H, ^{15}N and ^{13}C backbone resonances assignment of PDZ1ext was carried out using 3D HNCA, HNCACB, HNCO, HNCOCA experiments (Tris HCl buffer adjusted to pH 6). The 2D ^{1}H-^{15}N HSQC spectra used to compare PDZ1, PDZ1ext, PDZ2 and PDZ2ext were recorded in Tris HCl pH 7.5 (50mM). The 2D ^{1}H-^{15}N HSQC spectra used to compare PDZ1ext, PDZ2ext and P1P2 were recorded in Phosphate Buffer pH 6.5 (50mM). Titration experiments of PDZ1 (WT), PDZ1ext (WT) and PDZ1ext

(G103R) against ADGRV1pbm peptide were performed in Tris HCl pH 7.5 (50mM). Averaged Chemical Shifts Perturbations (CSP) were calculated as: $\Delta\delta_{avg} = ((\Delta\delta(^{1}H))^{2}+(\Delta\delta(^{15}N)*0.159)^{2})^{½}$. For the determination of dissociation constants (Kd) of each peptide/PDZ interaction, averaged CSP δ_{avg} was plotted as a function of the molar ratio (peptide/PDZ) for several peaks and fitted using nonlinear regression and assuming a simple complex formation model. The average and standard deviation on minimum 10 significantly shifting peaks were calculated for each complex.

Small-Angle X-ray Scattering (SAXS)

X-ray scattering data were collected at the SWING beamline at Soleil Synchrotron (Saclay, France). 50µl of sample are injected on a Superdex-75 5/150GL size exclusion column in-line with the SAXS measuring cell. PDZD7 samples are prepared in the same buffer (Phosphate Buffer 50mM pH 6.5, NaCl 300mM, TCEP 0.5mM, anti-proteases). The experiment is performed at initial concentrations of 119µM (3.1mg/ml) for P1P2 WT and 108µM (2.8mg/ml) for P1P2 G103R. For complexes, 2µl of pH-adjusted ADGRV1pbm peptide (VELRRIPIADTHL$_{-COOH}$; 13,7mM) are added to the 50µl from the sample one hour before recording. Initial data processing was performed using sxCuBE, FOXTROT and PRIMUS. The radii of gyration were evaluated using Guinier approximation. The Dmax was determined from the distance distribution function P(r) obtained with the program GNOM.

Analytical UltraCentrifugation (AUC)

AUC experiments were performed at 10°C using an analytical Beckman Coulter Optima ultracentrifuge equipped with a AN50-Ti rotor. The P1P2 sample was prepared at 0.4mg/ml in Phosphate Buffer pH 6.5 (50mM). For the complex, 10µl of pH-adjusted ADGRV1pbm peptide (13,7mM) were added to 300µl of P1P2 sample. Interference profiles were analyzed with SEDFIT.

Metadynamics

The metadynamics simulation ran for 100 ns with 4 replicas for each experiment. The triple-strand conformation predicted by AlphaFold was used as starting point for PDZ1ext, and a homology model[358] based on whirlin second PDZ for PDZ2ext. The Beta Collective Variable has been applied on the β strand extension for residues 92-117 and 92-113 for PDZ1ext and PDZ2ext respectively. We used Gaussians of height 1.28 kJ/mol with a width of 0.02, adding a Gaussian every 5 ps. A lower wall bias at a β propensity of 1 was used to

compensate for the entropically favored unfolded states. The free energy associated with each conformation was calculated[359].

Fluorescence titrations

Fluorescence titrations were performed on a Jasco V-730 spectrometer using quartz cell 109.004F-QS (Hellma Analytics) with an optical path of 10mm and with thick side to reduce the inner filter effect. PDZ1ext and P1P2 were prepared in Phosphate Buffer pH 6.6 (50mM) at 6µM and 2µM respectively, while PDZ2ext was prepared in Tris HCl pH 7.5 (50mM) at 5µM. Variations in the quantum yield of fluorescence of tryptophans from the sample were monitored upon progressive addition of buffered ADGRV1pbm peptide. After peptide addition, the sample is homogenized by in-spectrometer magnet agitation for 30 seconds. The sample is left to rest for an additional 60 seconds before measurement. Excitation is set at 295nm and fluorescence was measured at the maximum emission wavelength (346-349nm) of the unbound form with a 5nm bandwidth for both. Titration curves were fitted using the following formula: $y=F_p*p_0+(F_x-F_p)*0.5*(R*I_0+(1-R)*x+K-((R*I_0+(1-R)*x+K)^2-4*R*(I_0-x)*x)^{1/2})$, where p_0 is the initial protein concentration, I_0 the peptide stock concentration, R the p_0/I_0 ratio, x the volume of added peptide in µl, F_p the fluorescence of PDZ constructs, F_x the fluorescence of the complex, and K the Kd. The values of p_0, I_0 and R are kept constant.

Microscale Thermophoresis (MST)

MST titrations were performed on a Monolith (Nanotemper) spectrometer using Monolith NT.LabelFree Premium Capillaries. The PDZ2 sample was prepared at 140µM in MES pH6 (50mM) and mixed (1:1) with a serial dilution (½) of ADGRV1pbm peptide (13,7mM stock). MST traces were read out at 20 seconds where signal to noise ratio was maximum and titration curves were fitted using the following formula: $y=U+(B-U)*(c+c_{max}+K-((c+c_{max}+K)^2-4c*c_{max})^{1/2}/(2c_{max})$. U is the F_{norm} signal of the target, B the F_{norm} signal of the complex, c the ligand concentration, c_{max} the final concentration of the ligand in the assay and K the dissociation constant.

NanoDifferential Scanning Fluorimetry (nanoDSF)

NanoDSF experiments were performed on a Prometheus (Nanotemper) spectrometer using Prometheus NT.48 Series nanoDSF Grade High Sensitivity Capillaries. All samples were prepared in MES pH6 (50mM) at micromolar concentrations. For the complexes,

ADGRV1pbm peptide was added at a final concentration of 756µM, diluting the samples by 5%.

Pull-down experiments

HeLa cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS). PDZD7 constructs correspond to the full-length murine sequence in fusion with a N-terminal strep tag and a C-terminal GFP, wild-type or carrying G103R/G228R variants, cloned in a PMT3 vector. The ADGRV1 construct corresponds to the murine β subunit of the receptor (5884-6298) cloned in a pRK5 vector. Plasmids were transfected in separate HeLa cultures using 25µg (PDZD7) and 10µg (ADGRV1) of DNA per 10cm dish using TurboFect™ Transfection Reagent (ThermoFisher Scientific) at 3µl per µg of transfected DNA. Mixes were prepared in volumes of 1ml/plate (in DMEM), vortexed and incubated up to 30 minutes at room temperature, cell media were renewed before transfection. After drop by drop addition of the transfection mix, cells were maintained at 37°C for 14 hours for protein expression. Cells were dissociated using Gibco™ Trypsine-EDTA (0,05 %), subsequently inactivated by addition of DMEM containing 10% FCS. Cells were counted, pelleted and resuspended at similar concentrations in PBS Tween 0.01%, Dodécyl-β-D-maltoside (DDM) 0.1% containing anti-proteases (buffer). Benzonase (Sigma) was added before sonication of the cells (40mA amplitude, 10 pulses of 12s with 18s intervals, in ice). An identical volume of ADGRV1 β transfected cell lysate was mixed to either PDZD7 WT, G103R or G228R transfected cell lysates. PDZ proteins have different expression levels, meaning that different amounts of sample are needed to saturate the dynabeads with the PDZ proteins. Pull down experiments have been performed several times allowing us to estimate the yield of expression of each PDZ construct for a given amount of cells. Therefore, volumes of added PDZD7 lysate were adjusted to compensate for expression disparities between WT and mutants and maintain a similar bait concentration in each mix. Lysates from non-transfected cells were added to each mix to level total cell content to ensure the same lysate concentration in the input fraction. Dynabeads™ Protein G (Invitrogen) with coupled α-GFP rabbit antibody (Invitrogen A11122) was added to each mix (equivalent to 15µl of bead stock per condition), then incubated 1h on wheel (4°C). The supernatant was removed (unbound fraction) and beads were washed twice with buffer containing 10% BSA and twice with buffer without BSA. A final 40µl wash of buffer was performed to check for remaining non-specific binding (wash fraction). The bait and bound proteins were eluted in 40µl Bolt™ LDS Sample Buffer 4X by heating (100°C, 5 minutes) and vortexing (15 sec).

Samples were loaded twice onto duplicate 4% polyacrylamide gels along with PageRuler™Plus protein ladder (ThermoFisher Scientific) and transferred to PVDF membranes. Membranes were blocked using 5% milk TBST. Target proteins were detected using mouse α-GFP (LivingColors™ JL-8, Takara) and rabbit α-ADGRV1 (custom, raised against the cytoplasmic domain) antibodies (1/2000 in 5% milk TBST), as well as secondary α-mouse and α-rabbit IgG antibodies coupled to HRP (1/2000 in 5% milk TBST). One duplicated gel was used for each primary antibody. Signal was generated using Immobilon Forte Western HRP substrate.

II.2.e – Results

The N-terminal PDZ domains of PDZD7 paralogous proteins harmonin and whirlin possess C-terminal extensions with conserved proline and tryptophan residues (Figure 1). These folded extensions modulate the affinity of associated PDZ domains for their partners and promote the formation of domain assembly. In harmonin, this extension folds as a hairpin followed by a short α-helix mediating the formation of a heterotopic supramodule between the first PDZ and the upstream Harmonin Homology Domain (HHD) and forming direct contacts with residues upstream of the partner's PBM[87] (PDB entry 3k1r). This supramodule displays ~10 fold increased affinity for protocadherin-15 compared to the isolated PDZ[55]. In whirlin, the first and second PDZ domains interact via their respective C-terminal extensions, structured as hairpins, promoting the formation of a homotypic supramodule that improves the binding capacity of the PDZ tandem by ~3 fold[269]. In PDZD7, homologous extensions with conserved P and W residues are also identified downstream the two N-terminal PDZ domains (Figure 1).

We designed four constructs of the isolated N-terminal PDZ domains of mouse PDZD7 to decipher the role of the PDZ downstream sequences in the interaction with the C-terminal motif PBM of ADGRV1 (ADGRV1pbm). These constructs correspond to the first PDZ domain (PDZ1 construct; A81 to G167), the first PDZ with its C-terminal extension (PDZ1ext construct; A81 to E192), the second PDZ (PDZ2 construct; G206 to G292) and the second PDZ with its C-terminal extension (PDZ2ext construct; G206 to G313).

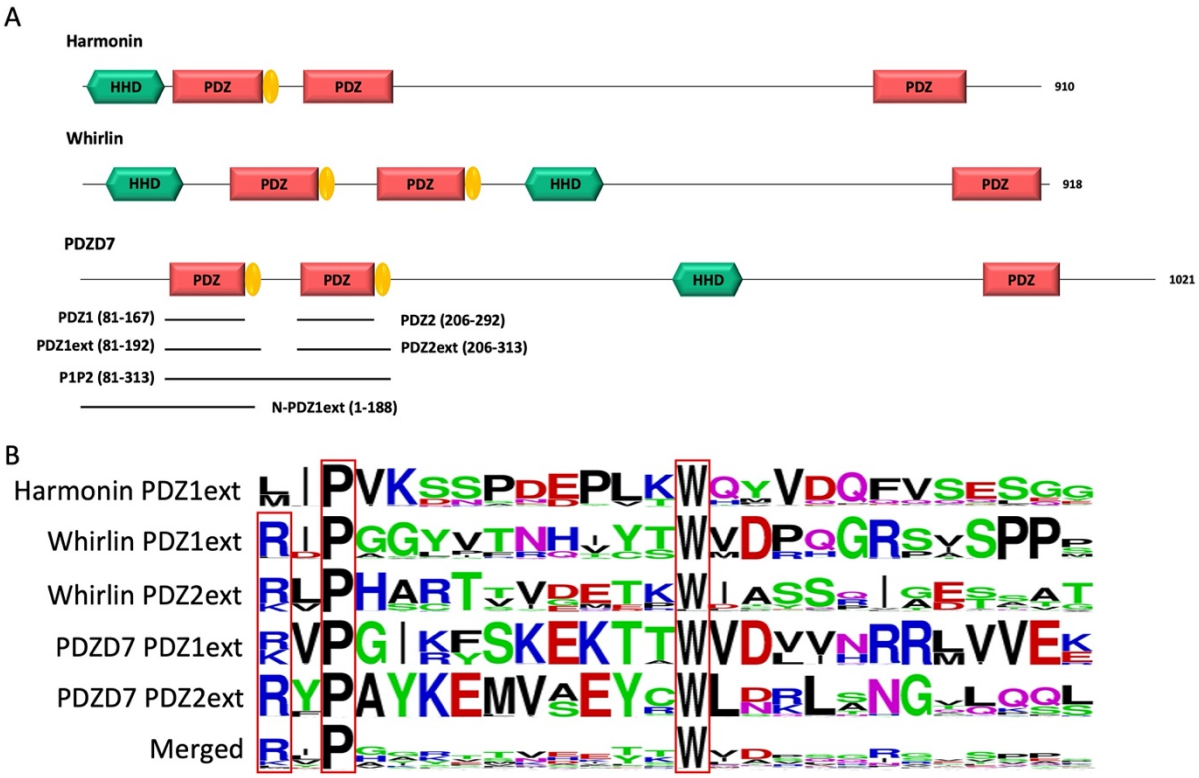


Figure 1 - (A) Schematic representation of the modular organization of the three paralogous proteins harmonin, whirlin and PDZD7. PDZ domain extensions are indicated in yellow. The 6 constructs of PDZD7 used in our in vitro study are indicated by black lines aligned to the corresponding fragment of PDZD7. Indicated delimitations correspond to the mouse protein sequence. (B) Comparison of PDZ domain extension sequences from harmonin, whirlin and PDZD7. For each extension, all UniprotKB sequences of the corresponding protein were aligned. The result is displayed using a logo representation, where the height of each residue one-letter code translates to its conservation at the corresponding position in the sequence alignment[379].

Folding of the PDZ1 extension

We produced a ^{15}N and ^{13}C labeled sample of PDZ1ext to characterize the folding state of the PDZD7 PDZ1 C-terminal extension. We successfully assigned 98%, 99% and 100% of the (H, N), CO and (CA,CB) backbone resonances of PDZ1ext, respectively. Secondary structure analysis derived from HN, N, CA, CB and CO chemical shifts predicts backbone torsion angles consistent with a triple β strand fold of the PDZ1 extension (TALOS-N prediction server; Shen and Bax, 2013)(Figure 2A). These three strands display lower propensities than those in the PDZ core (around 0.6). Consistently, the predicted order parameters are lower than in the rigid secondary structure elements of the core ($S^2_{pred} < 0.6$), but higher than expected for a flexible unfolded tail. Moreover, the peak intensities of the β strand extension in the ^{1}H-^{15}N HSQC spectrum at pH 6.5 are similar to the ones of the PDZ core (data not shown), but are significantly reduced at higher pH (7.4) due to increased HN/H_2O exchange. Altogether, these results are consistent with the formation of transient C-ter β-strands.

Furthermore, the mapping of Chemical Shift Perturbations (CSP) calculated for backbone resonances between PDZ1 and PDZ1ext constructs highlights an area encompassing β2 to β4 strands and the α1-β4 loop of the PDZ core (Figures 2B,C). This result is in agreement with the model of PDZ1ext predicted by AlphaFold[175] (Figure D), where the triple β strand extension is packed onto the inner leaflet of the PDZ, although a lower confidence level is found in the C-terminal extension (lDDT around 60%, Supplementary 1).

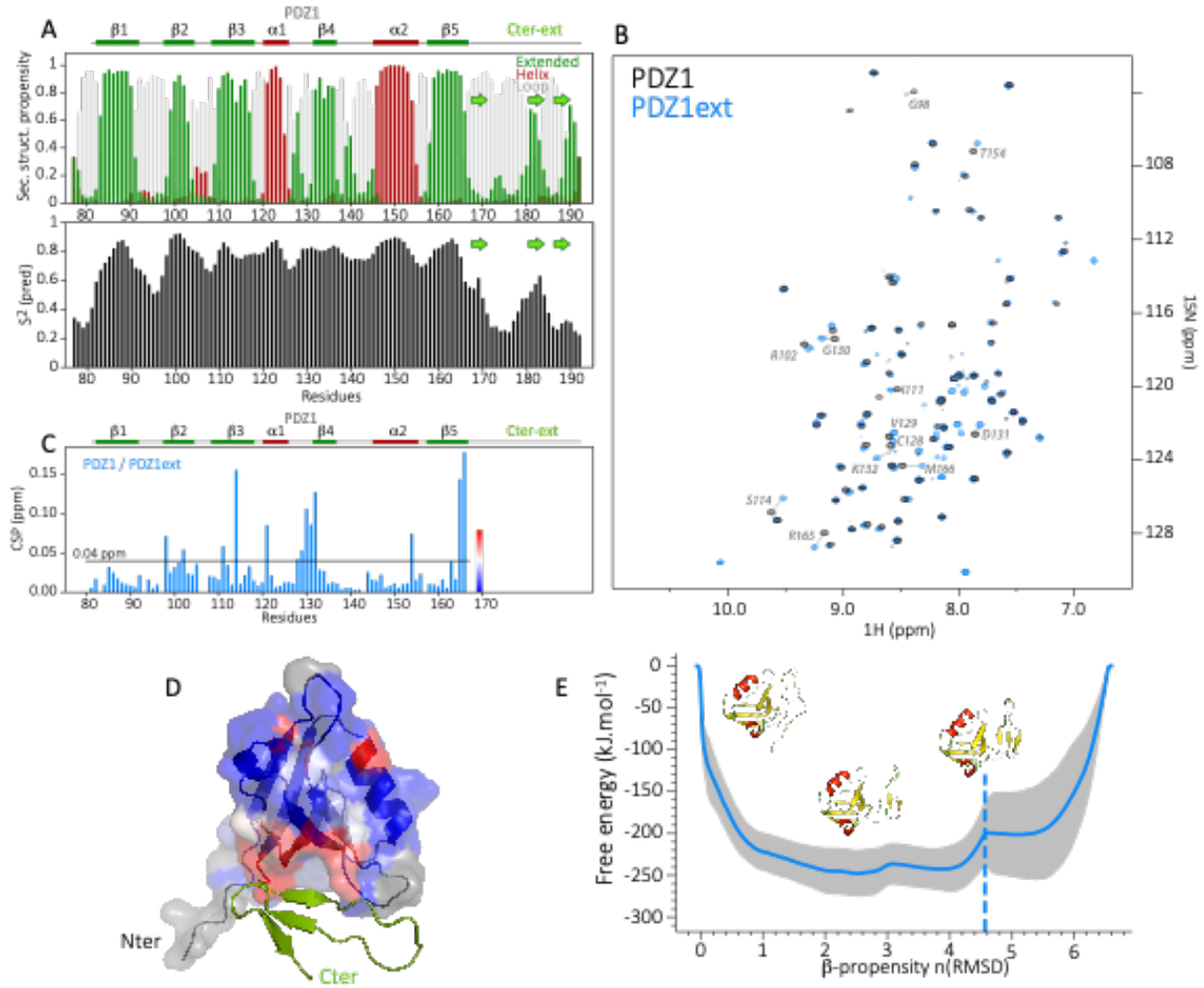


Figure 2 - (A) Secondary structure prediction from TALOS-N, based on the backbone assignment of PDZ1ext, showing the propensity of β-strands in green and α-helices in dark red (at the top) and the predicted order parameters S^2 (at the bottom). In the Cter extension, the consensus position of the three b-strands (green arrows) is inferred from the TALOS order parameters and the AlphaFold predicted model. (B) Superimposed ^{1}H-^{15}N HSQC spectra of PDZ1 (black) and PDZ1ext (blue). Assignment is indicated in gray for the most perturbed PDZ1 residues (CSP > 0.04 ppm). Blue peaks without assignment correspond to residues of the Cter extension. (C) NMR chemical shift perturbations (CSP) on PDZ1 induced by the Cter extension, i.e., between PDZ1 and PDZ1ext. Secondary structure elements are indicated at the top. (D) Structure of PDZ1ext AlphaFold model in ribbon, with the PDZ1 domain displayed as surface and the Cter extension in green. CSP induced by the Cter extension are mapped on the AlphaFold model of PDZ1ext. The color coding from blue (no CSP) to red (large CSP) as indicated in (C) highlights the packing of the Cter extension onto PDZ1, in agreement with the AlphaFold model. Residues depicted in grey have no associated chemical shift (missing peak in PDZ1ext). (E) Energy landscape obtained by metadynamics computing. The blue line indicates the average energy measured for each β propensity across 4 trajectories. Standard deviation is indicated by the grey outline. Displayed snapshots correspond to conformations extracted from the trajectory. The vertical dashed line indicates the starting point of the simulations.

Finally, we used metadynamics[360] enhanced sampling to depict the free-energy landscape of PDZD7 PDZ1 extension and assess the relative stability of its folded states. We used the Beta Collective Variable[361] on the β strand extension to enlarge the conformational space sampled for this region. The β strand scoring (bounded between 0 and 1) was computed for each residue of the considered region. The sum of this score thus reflects the overall β propensity of the extension throughout the simulation. The free energy associated with each sampled conformation was calculated[359]. The metadynamics showed that multiple metastable folded states of the β strand extension are explored with a β-propensity between 2 and 4 for the lowest free-energies, consistent with the partially stable β strands predicted by NMR. In conclusion, the C-terminal extension of PDZD7 PDZ1 adopts a dynamic triple β strand conformation, transiently in contact with the PDZ core.

Folding of the PDZ2 extension

The biochemical behavior of PDZD7 PDZ2 drastically differs from PDZ1, with or without its C-terminal extension. Rapid aggregation of both purified PDZ2 and PDZ2ext is observed at temperatures higher than 15°C despite extensive screening of buffer conditions (including buffer, pH, salt and additional ions). The ^{15}N-labelled PDZ2 constructs were stable for a few hours only, preventing the acquisition of 3D heteronuclear NMR experiments and the assignment of PDZ2 and PDZ2ext for an analysis at the residue level. Nonetheless, we recorded 2D HSQC spectra on PDZ2 and PDZ2ext samples (Figure 3A) and we observed that the presence of the PDZ2 extension induces perturbations of resonances on the PDZ core comparable in number and amplitude to those detected between PDZ1 and PDZ1ext (Figure 3B). The subsequent CSPs ranked by decreasing order indicate that 16.3% of PDZ2 residues have CSPs higher than 0.04 ppm and 15.3% in the case of PDZ1 in agreement with perturbations of comparable amplitude induced by the extensions on both PDZs. In order to further establish the accessible folding states of the PDZ2 extension, we used the same metadynamics approach as for PDZ1ext and probed its conformational energy landscape (Figure 3C). Again, we observe accessible metastable folded states of the PDZ2 extension, with β propensities between 1 and 4 and associated free energies comparable to the ones of PDZ1ext. Altogether, these results suggest an exchange between unfolded and β conformations of PDZD7 PDZ2 extension.

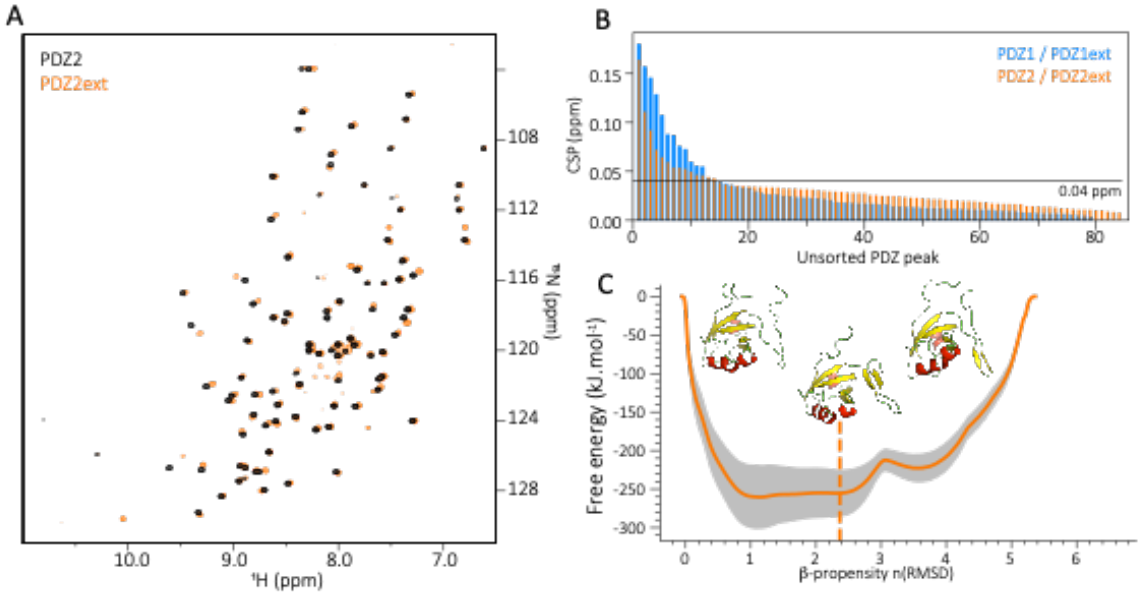


Figure 3 - (A) Superimposition of PDZ2 (black) and PDZ2ext (orange) ^{1}H-^{15}N HSQC spectra. (B) Unsorted Chemical Shift Perturbations (CSP) measured for each residue of the PDZ core when comparing PDZ1 to PDZ1ext (blue) and PDZ2 to PDZ2ext (orange). (C) Energy landscape obtained by metadynamics computing. The orange line indicates the average energy measured for each β propensity across 4 trajectories. Standard deviation is indicated by the grey outline. Displayed snapshots correspond to conformations extracted from the trajectory. The vertical dashed line indicates the starting point of the simulations.

Supramodular conformation of the PDZ tandem

As previously stated, folding of PDZ extensions downstream PDZ1 and PDZ2 domains of harmonin and whirlin is necessary for the supramodular organization of the proteins. In particular, the two N-terminal PDZ domains of whirlin adopt a compact conformation with a hairpin-mediated interface. In PDZD7, our results showed that PDZ1 and PDZ2 also possess C-terminal folded extensions. We collected Small Angle X-ray Scattering (SAXS) data and determined geometrical parameters of the PDZD7 PDZ tandem (P1P2 construct; A81 to G313), with a derived radius of gyration (Rg) of 27.8 Å and interatomic maximum distance (Dmax) of 109.5 Å. These parameters are comparable to what we previously measured on the PDZ tandem supramodule in whirlin (Rg=27.5Å; Dmax=115Å)[269]. The estimated molecular weight of 24 kDa derived from the PDZD7 SAXS data is also consistent with the theoretical value of 25.8 kDa. As illustrated by the dimensionless Kratky representation, PDZD7 PDZ tandem adopts a rather compact, yet elongated conformation in solution, similar to the PDZ supramodule of whirlin (Supplementary 2).

The superimposition of HSQC spectra of PDZ1ext, PDZ2ext and P1P2 in figure 4 show well overlapped peaks of the isolated PDZ domains and of the PDZ tandem (Figure 4A). This results in small CSPs between P1P2 and each PDZ domain, indicating only slight changes

of environment of each PDZ domain in the context of the tandem. Only the last 2 β strands of PDZ1 extension are significantly perturbed in the PDZ tandem (CSPs > 0.04 ppm) (Figure 4B, C), with reduced peak intensities likely reflecting a loss of flexibility and a potential stabilization of the C-terminal extension. Similarly, only few residues of the second PDZ exhibit significant CSPs (> 0.04 ppm) in the presence of the PDZ1 (Figure 4D). In conclusion, the two N-terminal PDZ domains of PDZD7 are not independent in solution, with an interplay likely mediated by folded C-terminal sequences and with little perturbations relayed to the core of each PDZ domain. Altogether, the PDZ tandem of PDZD7 appears to have a similar behavior as the one of whirlin, which is in equilibrium between partially opened and closed conformations.

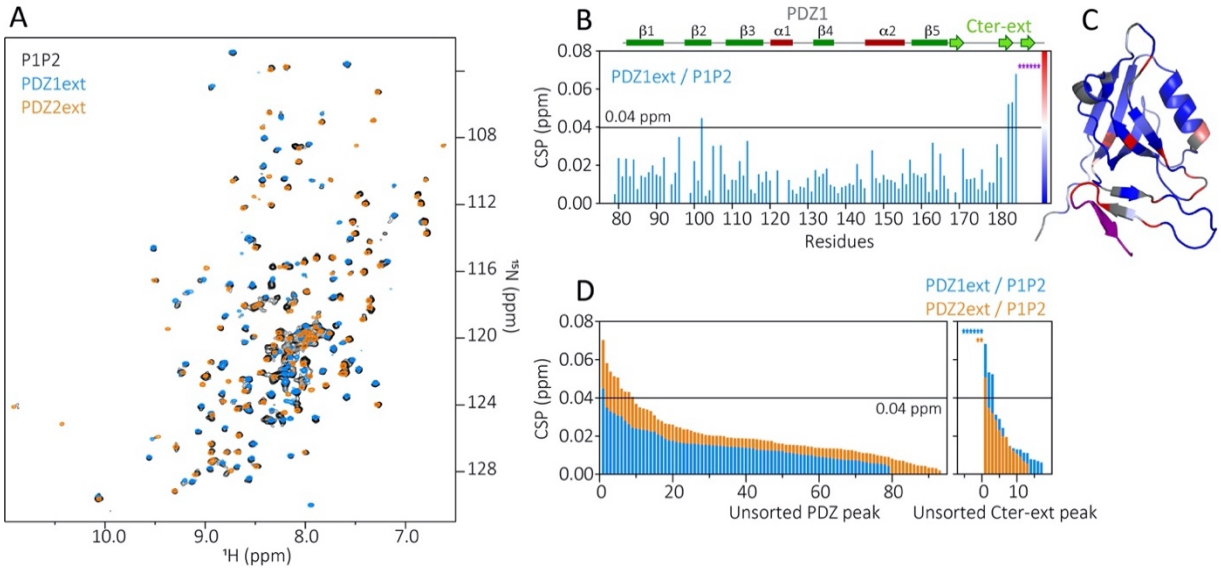


Figure 4 - (A) Superimposition of P1P2 (black), PDZ1ext (blue) and PDZ2ext (orange) ^{1}H-^{15}N HSQC spectra. (B) Chemical Shift Perturbations (CSP) measured for each residue of PDZ1ext when comparing PDZ1ext to P1P2. Purple stars correspond to residues of the extension which can not be assigned in P1P2 due to large CSP or missing peaks, indicating large changes of environment. Secondary structures are represented above the barplot. (C) Mapping of the PDZ1ext/P1P2 CSPs on PDZ1ext AlphaFold model with the color coding indicated in (B). (D) Unsorted CSPs measured for each residue when comparing PDZ1ext to P1P2 (blue) and PDZ2ext to P1P2 (orange). Colored stars correspond to residues which can not be assigned in P1P2 due to large CSP or missing peaks, indicating large changes of environment.

Affinity of the PDZ domains and tandem towards ADGRV1pbm

We monitored the interaction between PDZD7 constructs and a peptide (ADGRV1pbm) corresponding to the last 13 residues of ADGRV1, encompassing its PBM (sequence VELRRIPIADTHL$_{\text{-COOH}}$; PBM underlined). We measured apparent Kd values of our constructs towards ADGRV1pbm using several biophysical methods (NMR, fluorescence, MicroScale Thermophoresis (MST)) (Table 1; Supplementary 3-5). First, we showed that the isolated PDZ1 and PDZ2 of PDZD7 bind to ADGRV1pbm with weak affinities of 1313µM and 313µM, respectively. However, addition of their respective extensions induces a dramatic increase of the affinity towards ADGRV1pbm, with Kd values of 47µM for PDZ1ext and 3.1µM for PDZ2ext. In both cases, these extensions appear critical for the interaction, with PDZ2ext being the preferential binder. As monitored by NMR, addition of ADGRV1pbm to PDZ1ext induces large chemical shift changes for residues in the canonical binding site defined by the β2 strand and α1 helix and the loss of peaks corresponding to residues in the PDZ extension due to intermediate exchange processes. This underlines a change in dynamics of the pre-formed triple strands upon peptide binding. Further shifts of residues from the β3 and β4 strands also suggest a potential stabilization of the interface between the folded extension and the inner leaflet of the PDZ core. The contribution of this extension to the increase of affinity towards ADGRV1pbm could thus arise from direct contacts with the peptide as well as distal structural/dynamic perturbations of the PDZ core. Altogether, these results support a structural contribution of the hairpin sequence of PDZD7 PDZ1 to the interaction with ADGRV1pbm.

We also showed that PDZ2ext has a similar affinity (Kd=3.1µM) for ADGRV1pbm compared to the PDZ tandem (P1P2; Kd=1.6µM), but 15-fold higher than PDZ1ext (Kd=47 µM). Taking advantage of the good superimposition of the isolated PDZ spectra on the tandem spectrum, we followed the saturation of each PDZ in the tandem as a function of ADGRV1pbm concentrations (data not shown). We confirmed that both PDZ domains in the tandem bind to ADGRV1pbm and that PDZ2 is saturated at lower ligand concentration than PDZ1, in agreement with the affinities measured for the isolated domains. Finally, we assessed if the presence of ligand further stabilizes a compact conformation of the P1P2 supramodule using SAXS and Analytical UltraCentrifugation (AUC)(Supplementary 6). For the bound P1P2, we derived a Rg of 27.4 Å and a Dmax of 104 Å from SAXS curves, as well as a sedimentation coefficient of 1.7s from AUC. Compared to the unbound PDZ tandem (Rg=27.8 Å, Dmax=109.5 Å, S=1.6s), these results reflect a minor, if significant, compaction of the supramodule without major structural reorganization of the protein.

*Table 1: Kd values measured for PDZD7 PDZ constructs in interaction with the last 13 residues of ADGRV1. PDZ1 and PDZ1ext by NMR titration; PDZ2 by MST titration; PDZ2ext and P1P2 by fluorescence titration). Errors indicated with * derive from the quality of the fit. For PDZ1 and PDZ1ext, errors derive from the standard deviation of measured Kd values for at least ten significantly shifting peaks.*

Construct	Kd
PDZ1	1313 ± 81µM
PDZ2	313 ± 52µM*
PDZ1ext	47 ± 6µM
PDZ2ext	3.1 ± 0.2µM*
P1P2	1.6 ± 0.3µM*

Stability of the PDZ domains and tandem

We measured the thermostability of PDZD7 constructs using nanoDifferential Scanning Fluorimetry (nanoDSF). As previously observed during the production and biochemical/biophysical characterizations of the domains, PDZ1ext is drastically more stable than PDZ2ext, with melting temperatures of 61.6°C and 38.5°C, respectively. Interestingly, the binding of ADGRV1pbm strongly increases the thermal stability of the unstable PDZ2ext (Tm increased from 38.5°C to 47.2°C), while the binding has no significant effect on the more stable PDZ1ext (Tm increased from 61.6°C to 61.8°C)(Table 2). The melting temperature of P1P2 also seems to be driven by the PDZ2, with a low Tm for the unbound form (41°C) and a strong stabilization upon ADGRV1pbm binding (Tm of 57.2°C). Overall, these results showed that ADGRV1pbm, primarily interacting with the second PDZ domain, largely stabilizes the PDZ supramodule.

Table 2: Melting temperatures (TM) measured for PDZD7 PDZ constructs, without ligand (unbound) or in complex with the last 13 residues of ADGRV1 (bound).

Construct	Tm (unbound)	Tm (bound)
PDZ1ext WT	61.6°C	61.8°C
PDZ1ext G103R	56.8°C	56.3°C
PDZ2ext WT	38.5°C	47.2°C
PDZ2ext G228R	~30°C	~30°C
P1P2 WT	41°C	57.2°C
P1P2 G103R	34.9°C	53.7°C

Other determinants of the PDZD7/ADGRV1 interaction

The P1P2 tandem of PDZD7 is preceded by 80 residues with no predicted secondary or tertiary structure, but which could contribute to the interaction with ADGRV1. Similarly, the 152 residues of ADGRV1 cytoplasmic domain upstream its PBM exhibit several stretches of conserved positions and could participate in the interaction. We first aimed at identifying potential secondary structures in both PDZD7 N-terminal region and ADGRV1 cytoplasmic domain. To this end, we produced both an extended PDZ1ext construct, including the N-terminus of the protein (N-PDZ1ext construct; M1-E192) and the whole ADGRV1 cytoplasmic domain of 152 residues.

On the overlay of PDZ1ext and N-PDZ1ext HSQC spectra, there is no significant chemical shift changes on PDZ1ext due to the presence of the N-terminal region (CSPs < 0.04 ppm). The latter is characteristic of a disordered region, with about 70 peaks (80 expected) of poor dispersion of HN chemical shifts (8-8.5 ppm) (Supplementary 7). This supports that the N-terminal 80 residues are unfolded and independent of the downstream PDZ1ext domain. We confirmed using SAXS experiments that the upstream N-terminal region is extended, flexible and behaves independently from the folded PDZ1 domain (Supplementary 2). For ADGRV1 cytoplasmic domain, the HSQC spectrum also displays a weak dispersion of signals (8-8.5 ppm), corresponding to a predominantly disordered protein (Supplementary 8) and we confirmed the absence of stable secondary structures in solution using circular dichroism that showed a spectrum largely governed by random coil conformations (data not shown). We then aimed at deciphering the potential contribution of these flexible sequences to the interaction between ADGRV1 and PDZD7 (Table 3). We used the entire cytoplasmic domain of ADGRV1 as partner of PDZD7 in NMR and fluorescence titrations experiments. First, we quantified the interaction with PDZ1ext, obtaining an apparent Kd value of 41 ± 14µM comparable to the value 47 ± 6µM obtained with the ADGRV1pbm peptide. Similarly, we used the N-terminal PDZD7 sequence N-PDZ1ext and measured Kd values of 241 ± 115µM and 270 ± 21µM for ADGRV1 cytoplasmic domain and PBM peptide, respectively. In addition, no significant CSPs have been observed in the N-terminal sequence upstream PDZ1 upon ADGRV1 binding. The presence of the long and unfolded N-terminus region induces a 6-fold decrease in affinity for ADGRV1 (PBM peptide or cytoplasmic domain) that might be due to an entropic cost on the PBM binding. Finally, we measured Kd values in the few micromolar range between the tandem P1P2 and the whole cytoplasmic domain and the last 13 residues of ADGRV1 by fluorescence experiments. Altogether, these results showed that

the interaction between ADGRV1 cytoplasmic domain and PDZD7 do not involve other determinants besides the PBM-PDZ interaction.

*Table 3: Kd values measured for PDZD7 constructs against ADGRV1 whole cytoplasmic domain (cyto) or last 13 residues (peptide) (PDZ1ext and N-PDZ1ext by NMR titration; P1P2 by fluorescence titration). For fluorescence, the error derives from the quality of the fit (marked as *). For NMR, errors derive from the standard deviations of the measured Kd values for at least ten significantly shifting peaks.*

Construct	Kd (cyto)	Kd (peptide)
PDZ1ext	41 ± 14µM	47 ± 6µM
N-PDZ1ext	241 ± 115µM	270 ± 21µM
P1P2	8.1 ± 1.1µM*	1.6 ± 0.3µM*

Pathogenicity of the mutants

Two deafness-causing mutations of the same amino acid substitution are located at the same position in the binding groove of the first (G103R) and second (G228R) PDZ domains in PDZD7[299]. We first assessed by nanoDSF the effect of these mutations on the thermal stability of PDZD7 constructs (Table 2).

The mutation in PDZ1 (G103R) induces a significant destabilization of both isolated PDZ domains (PDZ1ext, Tm decreased by 4.8°C for the apo form) and PDZ tandem (P1P2, Tm decreased by 6.1°C for the apo form). The second mutation (G228R) markedly destabilizes the isolated PDZ2ext, with a slow transition of fluorescence ratio (330/350nm) initiated below 15°C, indicating progressive unfolding of the protein at room temperature, with a Tm that can be roughly estimated in the 30°C range (Supplementary 9). As a consequence, we could not express the PDZ tandem carrying the G228R mutation. Globally, both deafness mutations significantly decrease the thermostability of PDZ domains with a more pronounced effect on PDZ2, initially more unstable. Interestingly, addition of ADGRV1pbm peptide still stabilizes the P1P2 tandem when the PDZ1 is mutated (by 19°C in Tm), consistent with a functional binding site in the PDZ2 and an overall melting temperature of P1P2 driven by this second PDZ.

We characterized the potential effect of the two substitutions on the interaction with ADGRV1pbm using a panel of techniques. First, we confirmed by NMR that PDZ1ext G103R is well-folded, with significant perturbations (CSPs > 0.04 ppm) only detected for residues spatially close to the G103R mutation (Supplementary 10). In addition, we observed an

increase in peak intensity in the C-terminal β strands facing the mutated residue, suggesting a potential higher flexibility of the extension. This G103R substitution results in a 25-fold decrease in affinity (Kd from 47 ± 6µM to 1190 ± 140µM for WT and G103R PDZ1ext, respectively) showing that this mutation interferes with the binding of PDZD7 to its partners through PDZ1ext. A similar affinity in the mM range has been determined for PDZ1 lacking the β strand extension (Kd of 1310 ± 80µM), suggesting that the G103R mutation most likely affect the interface between the PBM pocket and the C-terminal extension, making it incompetent for binding. The substitution of a small, neutral side chain (glycine) by a large, cationic side chain (arginine) residue could also induce steric hindrance in the binding pocket. Similar local perturbations are observed in the P1P2 G103R tandem, but WT and G103R mutated constructs have similar affinities measured by fluorescence, with Kd values of 1.6 ± 0.3µM and 1.9 ± 0.4µM respectively. This first mutation thus alters direct PBM binding to the first PDZ, but has no long-distance impact on the binding of ADGRV1pbm to the second PDZ. This result confirms the main role of PDZ2 in the interaction with ADGRV1. We used SAXS to monitor global conformational changes of the PDZ supramodule carrying the G103R mutation. We calculated Rg=28 Å and Dmax=118.2 Å for the mutant tandem without ligand, and Rg=27.7 Å and Dmax=123.6 Å in presence of ADGRV1pbm peptide. These values are slightly larger than in the WT P1P2, which might indicate a shift of the closed/open conformer equilibrium toward a relaxed state of the tandem. This could result from the increased flexibility of the PDZ1 β strand extension required to form the interface in the supramodule.

We could not measure the interaction between PDZ2 G228R and ADGRV1pbm, due to a low yield and severe aggregation of the mutated domain during biophysical experiments, consistent with its low melting temperature. For the same reasons, we are unable to produce the PDZ tandem carrying the G228R substitution, preventing us from measuring its affinity for ADGRV1pbm.

To conclude, the substitution G103R prevents the first domain of the tandem from binding ADGRV1, but does not affect the binding capacity of the second domain. On the other hand, it significantly alters the overall thermostability of the supramodule even in complex with ADGRV1pbm. The other substitution G228R in PDZ2 has an even greater effect by targeting the more unstable PDZ domain and the best ADGRV1 binder in the tandem.

Full-length proteins in cell assays

Finally, we used pull-down assays to test the interaction between full-length PDZD7 wild-type and mutants (in fusion with a C-terminal GFP), and ADGRV1 β subunit, encompassing the seven transmembrane segments and cytoplasmic domain. Proteins are expressed separately in HeLa cells and lysates are mixed after sonication. Addition of G sepharose beads coupled to an anti-GFP antibody (see Material and methods) allowed to bind PDZD7. Three experiments were performed in parallel for each mutant; first a pull-down of ADGRV1 without bait to rule out unspecific binding (negative control), then with WT PDZD7 (positive control) and finally with the considered mutant. For each experiment, the same samples corresponding to the input, unbound, wash and elution fractions are loaded in equal amounts into two twin gels (top and bottom parts). Each gel is used to detect the presence of either PDZD7-GFP (Figure 5, top part) or ADGRV1 (Figure 5, bottom part) on the beads after washing by western blotting. Each mutant is therefore tested in independent experiments in comparison with the wild-type (Figure 5 left part for G103R and right part for G228R). In both cases, little to no ADGRV1 is detected bound to the beads without bait, while a clear band is observed with WT PDZD7 as bait. Thus, full-length PDZD7 interacts with ADGRV1 β subunit, as expected. When using either G103R or G228R mutants of PDZD7, a decrease of ADGRV1 quantity bound to the beads was observed in comparison with wild-type PDZD7, similar to the experiment performed without bait, although the amount of bait on the bead was comparable (Supplementary 11). Therefore, both G103R or G228R variants impair binding of full-length PDZD7 to ADGRV1. To note, the two PDZD7 variants are expressed in lower quantities compared to the wild-type when transfecting an equal amount of cells, and have a higher tendency for degradation, resulting in multiple bands of low molecular weight in the western blot (Supplementary 11). This observation reflects the lower stability of the variants paired with the loss of binding capacity to ADGRV1, in agreement with the results obtained *in vitro* using purified mutated PDZ constructs.

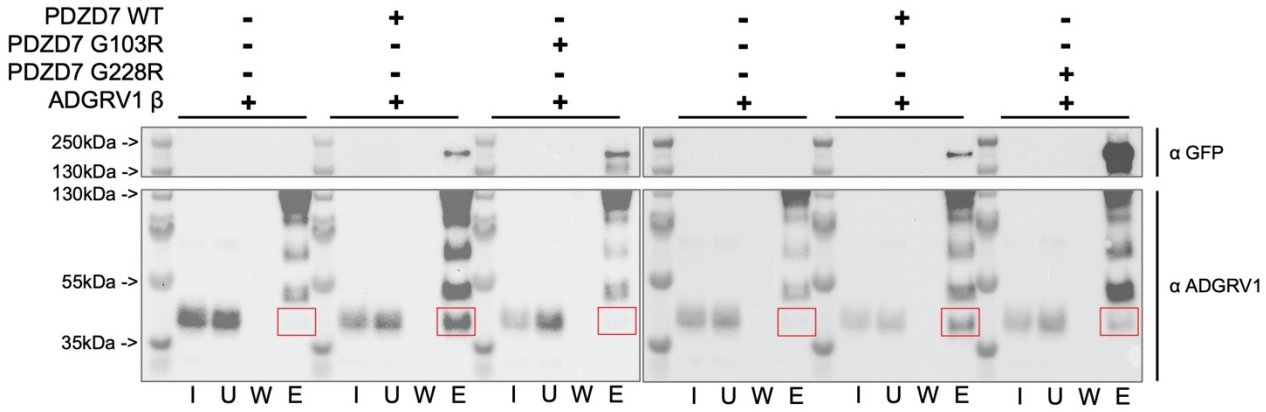


Figure 5 - Western blot detection of ADGRV1 β subunit (prey) pulled down by PDZD7-GFP (bait) constructs from HeLa cell lysates. PDZD7 constructs are detected with an antibody directed against the fused GFP. ADGRV1 is detected using an antibody directed against its cytoplasmic domain. For each condition, fractions corresponding to the input (I), unbound fraction (U), wash (W) and elution (E) are loaded. Red squares indicate the expected position for ADGRV1 β in the elution fraction.

II.2.f – Discussion

PDZD7 and ADGRV1 are the main candidate proteins for regulating the signaling associated with the Usher2 complex. PDZD7 is primarily a scaffolding protein, with several protein-protein interaction domains (three PDZ domains and one HHD) likely recruiting several partners at the ankle links anchoring sites[322]. Interestingly, PDZD7, but not its homologue whirlin, downregulates ADGRV1 GPCR activity in cellular models in a PDZ-PBM dependent manner[191].

Here we show that both N-terminal PDZ1 and PDZ2 domains of PDZD7 can interact with the C-terminal PDZ binding motif of ADGRV1. Our results highlight a critical contribution of atypical PDZ domain C-terminal extensions to the binding, tuning affinity towards partners. Indeed, the presence of these extensions results in a ~20 to 100 fold increase in affinity for each PDZ towards ADGRV1pbm. In addition, the ADGRV1 motif preferentially interacts with PDZD7 PDZ2, with a ~15 fold higher affinity (Kd=3.1µM) compared to PDZ1 (Kd=47µM). Therefore, ADGRV1pbm binds to PDZ2 in the higher range of affinity than what is commonly found for PDZ domain interactions (1-100µM). As monitored by NMR on PDZ1, the C-terminal extension of the domain adopts an original 3 β strands conformation in solution, that is at least partially in contact with the core of the PDZ (β2, β3 strands), consistent with AlphaFold's prediction. Interestingly, addition of the ADGRV1pbm peptide induces a change in dynamics of the extension, potentially favoring the folded state, and a compaction onto the PDZ core. The increase of affinity induced by the extension could arise from a modulation of the dynamics of the PDZ core in favor of the bound state, or from additional contacts between

the extension and the partner upstream of its PBM, as observed in the homologous protein harmonin in complex with the protein SANS[87] (PDB entry 3k1r). PDZD7 PDZ2 also possesses a functional C-terminal extension responsible for a drastic increase in affinity of the PDZ towards ADGRV1pbm. We showed that the C-terminal extension of PDZ2 has an energy minimum corresponding to a folded β hairpin conformation. Furthermore, the presence of this extension induces perturbations of the ^{1}H-^{15}N spectra of the PDZ2 similar to the ones observed in PDZ1, suggesting comparable contacts with the PDZ core. Lastly, complex formation with ADGRV1pbm also leads to a change in dynamics of the extension for both PDZ, resulting in a loss of signals by NMR. Altogether, these results support an accessible 2-3 β strands conformation of each PDZ extension in solution.

In the homologous proteins whirlin and harmonin, PDZ extensions mediate the formation of supramodules. In harmonin, only the first PDZ possesses a C-terminal hairpin extension, mediating an interface with the preceding HHD domain[87]. In whirlin, the N-terminal PDZ tandem is in equilibrium between open states and a close conformation. In this compact supramodule, each PDZ extension folds as a β hairpin packed onto the inner leaflet of its adjacent PDZ. The two hairpins then form the interface between the two PDZs[269] without direct contact between the cores of the two domains. PDZD7 PDZ tandem also adopts a rather compact and elongated conformation in solution, possibly in exchange with more open conformations, with SAXS parameters comparable to the ones of whirlin. Addition of ADGRV1pbm peptide seems to favor the closed conformation of the PDZ tandem through the stabilization of the β extensions. However, the tandem conformation has limited effect on affinity towards the partner when compared to PDZ2 alone (Kds of 1.6 ± 0.3µM and 3.1 ± 0.2µM, respectively). PDZD7 PDZ tandem still exhibits two functional binding sites, allowing it to potentially bind two ADGRV1pbm in the highly confined environments of the stereocilia, which would be compatible with the high protein densities found at the ankle links anchoring sites. PDZD7 PDZ1 could also have other natural binders instead of ADGRV1. PDZD7 PDZ2 displays a histidine in position 1 of the α2 helix lining the binding groove, canonically interacting with type 1 (-S/T-X-Φ_{-COOH},X being any residue and Φ a hydrophobic residue) PBM such as the one of ADGRV1 (-**T**HL$_{-COOH}$). By contrast, PDZD7 PDZ1 unusually exhibits a methionine at position 1 of its α2 helix (only 4 amongst the 266 human PDZ domains), more suitable for type 2 (-Φ-X-Φ_{-COOH}) PBM binding. Therefore, PDZD7 PDZ1 might recruit other partners in the vicinity of ADGRV1[322] or more generally promote the assembly of ternary proteins within the Usher2 complex. Furthermore, the difference in thermostability of the two PDZ domains could also underlie distinct functions in protein assemblies. PDZ1 is highly

stable with and without partner. Conversely, the salvageability of PDZ2 instability by the binding of ADGRV1 PBM possibly regulates temporal maintenance of the protein by promoting its recycling in the absence of ligand. The temporal expression of ADGRV1 and PDZD7 at the ankle links has been poorly described and events triggering expression, persistence and degradation of the two proteins remain unclear. PDZD7 destabilization in absence of PDZ2 ligand might help coordinate disassembly of the complex in maturing hair cells.

To probe for additional determinants of interaction between PDZD7 and ADGRV1, we enlarged our study to PDZD7 N-terminal region (80 first residues upstream the PDZ tandem) and ADGRV1 whole cytoplasmic domain (152 intracellular residues). We confirmed that PDZD7 residues upstream of the tandem do not exhibit tertiary structures, are largely extended and flexible in solution and do not induce perturbations onto the PDZ1 core. Similarly, we have shown that the isolated cytoplasmic domain of ADGRV1 is mostly unfolded and flexible. In terms of interaction, we measured similar affinities for ADGRV1 PBM or whole cytoplasmic domain with the PDZD7 constructs we tested. In conclusion, the PBM of ADGRV1 and the PDZ tandem of PDZD7 are the only determinants of interaction between the two proteins in our experiments. Altogether, our results highlight the critical role of flanking regions in PDZ interactions and supramodular organization, which has not been systematically documented thus far. Interestingly, five out of the six PDZ located in the N-terminal regions of the three paralogous proteins harmonin, whirlin and PDZD7 have their binding capacity modulated by domain extensions. These sequences have been conserved during evolution and could play a unique role as modulator of affinity and specificity. More generally, PDZ domains are known to form hetero and homotypic supramodules. One of the most extensively studied supramodule is the three-domains PDZ3-SH3-GK from the protein PSD-95[362]. In addition, about 40% of human PDZ domains have predicted secondary structures within 50 residues of their C-termini[271]. However, available experimental data do not reflect such propensity. Here our results highlight the need to include these predicted extensions in structural studies, as well as in high-throughput screening assays knowing their potential role on PDZ stabilization and binding properties. A thorough bioinformatics study of the PDZ domain family, aimed at identifying conserved and co-evoluted stretches of residues outside of their canonical fold, could already unveil several regulatory mechanisms for protein-protein interactions, scaffolding and signaling mediated by PDZ containing proteins.

The regulation of GPCR trafficking by PDZ-containing proteins has been well documented[363]. However, to our knowledge, a regulation of GPCR activity through a PDZ/PBM interaction is unique with the ADGRV1/PDZD7 molecular assembly, but the underlying molecular mechanisms remain elusive. ADGRV1 cytoplasmic domain might behave differently in the context of the entire GPCR β subunit, including intracellular loops and plasma membrane, opening other regulatory mechanisms. The regulation by PDZD7 could also arise from direct interaction with proteins mediating GPCR signaling (G protein, arrestin), or by recruiting other G protein signaling inhibitors such as members of the Regulators of G-protein Signaling (RGS) family[364]. Such regulation of ADGRV1 by PDZD7 could be affected by pathogenic mutations identified in patients. We have characterized the effect of two autosomal recessive non-syndromic hearing loss variants of PDZD7, with glycine to arginine substitutions in the lining groove of both the first (G103R) and second (G228R) PDZ domains. By probing the thermostability, folding, and binding capacity of the mutated PDZ domains, we showed that each mutation induced a decrease in thermostability of the protein, as well as a decrease of affinity towards ADGRV1. Consistently, our pull-down experiments showed a loss of binding capacity of the full-length PDZD7 to ADGRV1 β subunit with the two variants, likely abrogating regulatory potential. We propose that the destabilization of PDZD7 by the two variants could affect both the formation of the Usher 2 complex and the regulation of ADGRV1, inducing physiopathogenesis. The effect of these variants on Usher 2 proteins localization in hair cell stereocilia, and ADGRV1 mediated signaling remains to be characterized.

II.2.g – Additional comments

In this article, we showed that both N-terminal PDZ domains of PDZD7 were able to recognize the C-terminal PBM of ADGRV1. This interaction depends on the presence of C-terminal extensions of each PDZ, adopting a transient β conformation (two or three strands). The second PDZ is the preferential binder for ADGRV1pbm, even though the first PDZ could bind with medium affinity, with a 15 fold difference in Kd (in the tenths of micromolar). We detected no additional contacts between ADGRV1 cytoplasmic domain and the PDZ tandem or PDZD7 N-terminal disordered region. However, the mechanism by which PDZ extensions mediate the stabilization of the PDZ-PBM complex remains elusive. In particular, structural data is lacking.

The study of the PDZ domain extensions has proven more difficult than we expected. Notably, their contribution to the interactions results in a complex molecular mechanism

mixing conformational exchange and binding events. To overcome this issue, we used several methods to monitor complex formation (Table 4).

Table 4: Kd values obtained for the interaction between ADGRV1 and PDZD7 constructs using NMR, Fluorescence and MST experiments.

Apparent Kd (µM)	Method	ADGRV1pbm	ADGRV1cyto
PDZ1	NMR	1313 ± 81	ND
PDZ1ext	NMR	47 ± 6	41 ± 14
	Fluorescence	21 ±1	-
	MST	> 6410	-
PDZ1ext G103R	NMR	1187 ± 138	-
	Fluorescence	40 ± 23	-
	MST	> 5740	-
PDZ2	MST	313 ± 52	-
PDZ2ext	Fluorescence	3.1 ± 0.2	-
	MST	> 8	-
P1P2	Fluorescence	1.6 ± 0.3	8.1 ± 1.1
	MST	2.9 ± 1.1	-
P1P2 G103R	Fluorescence	1.9 ± 0.4	-
	MST	3.97 ± 0.03	-
N-PDZ1ext	NMR	241 ± 115	270 ± 21

Using NMR, we could not quantitatively study the interaction between ADGRV1 and constructs encompassing the second PDZ of PDZD7. For these constructs, broadening of the peaks prevented to accurately follow perturbations of residues environment depending on ligand concentration. This likely results from a change in the equilibrium between folded and flexible conformations of the extension, resulting in a shift towards intermediate exchange rates, though no protein dynamics parameters were measured. In fluorescence experiments, we monitored the emission of tryptophan located in the extensions that may only indirectly probe the complex formation. Therefore, the fluorescence signals are affected by both ligand binding and conformational changes of the hairpins. In the same way, the analysis of MST results was complicated by a potential change of the extension conformation upon binding may compensate the peptide binding. This may be the reason why no binding is observed for PDZ1ext constructs at low peptide concentrations, resulting in Kd values in the millimolar range.

Metastable extensions are also detrimental for structural characterization. No crystals could be obtained for PDZD7 constructs encompassing PDZ extensions, despite several attempts using various protein concentrations, buffers and ligand concentrations. By NMR, peak broadening observed for the second PDZ is incompatible with the recording of 3D spectra required for secondary structure propensity calculation and structural determination. Finally, the PDZ tandem alone is too small and likely too dynamic to be characterized by cryo-EM.

We showed that the two N-terminal PDZ of PDZD7 can recognize the PBM of ADGRV1 exposed at the C-terminus of its unfolded and flexible cytoplasmic domain. However, identifying the determinants involved in the two proteins interaction is not sufficient to understand the mechanisms underlying ADGRV1 regulation by PDZD7. Additional structural information is required, notably on the stachel-activated GPCR in complex with its primary effectors. Such protein assembly may involve the conserved stretches of residues located along ADGRV1 cytoplasmic domain, some of which have predicted helical structures. Deciphering the structural specificities associated to the active receptor may help understand the negative regulation promoted by PDZD7 binding to the otherwise seemingly isolated PBM. Therefore, we decided to characterize the structure of the entire GPCR β subunit, in complex with signaling partners such as its associated G protein. This should shed light on structural context in which the PDZ-mediated regulation occurs. The first steps of this ongoing project are described in the next section.

II.3 – EM study of ADGRV1

In order to characterize the structure of ADGRV1β subunit, we needed to express and purify the membrane receptor. The expression and purification procedure described below has initially been implemented in the laboratory by Mariano Genera and further optimized by Yasmine Achat. Yasmine Achat now performs protein expression and purification, as well as cloning of new constructs and mutants of ADGRV1. I am in charge of grid preparation, Cryo-EM data collection and image processing.

We have cloned and expressed several constructs of ADGRV1β, with N-terminal and C-terminal Flag and His tags, respectively, and encompassing or not the C-terminal cytoplasmic domain. These constructs have been systematically generated with either a wild-type ($_{NH2}$SVYAVYA-) or a mutation-inactivated stachel ($_{NH2}$SV<u>A</u>AV<u>A</u>A-) peptide at their N-terminus, substituting residues (Y to A) at positions critical for stachel activation in other adhesion GPCRs[201–204]. These constructs were cloned in a pRK5 vector for expression in mammalian cells through means of transient transfection. They were also cloned in pFastBac1 plasmids for insertion by recombination (in DH10Bac *E. coli* cells) in a bacmid (Bac-to-bac Baculovirus Expression System, *Thermofisher*). This bacmid is used to infect Sf9 cells and produce a baculovirus stock, which can in turn be used to infect Sf9 insect cells for protein expression. After preliminary tests, we showed that the protein is well expressed both in mammalian (HEK) and insect (Sf9) cells. Insect cells having a higher protein yield, we chose this expression system for the production of our construct. The expression and purification

protocols implemented by Mariano Genera and Yasmine Achat are presented in the next paragraph.

For protein expression, a culture of Sf9 cells is infected at MOI 2.5 for 42 hours at 28°C. After centrifugation, the cell pellet is resuspended in Tris HCl pH7.5 50mM, NaCl 300mM, TCEP 0.5mM, complemented with Benzonase (*Sigma*) and tablets of anti-proteases (*Sigma*) and lysis is performed using an Ultra-turrax homogenizer system (*IKA*). Membranes are retrieved by ultracentrifugation at 200,000g for two hours. Membrane pellet is solubilized using the same buffer as for the lysis, complemented with 1% LMNG (Lauryl Maltose Neopentyl Glycol), 0.1% CHS (Cholesteryl Hemisuccinate) and 5% glycerol. Solubilized proteins are subsequently retrieved by ultracentrifugation (200,000g for 30 minutes). 20mM of pH-adjusted imidazole is added to the supernatant before loading on a HisTrap (*Cytiva*) column. During washing, the imidazole concentration is increased to 50mM while the LMNG is progressively decreased to 0.01%. The protein is eluted in Tris HCl pH7.5 50mM, NaCl 300mM, TCEP 0.5mM, anti-proteases, 0.01% LMNG and 0.001% CHS, complemented with 200mM imidazole. The same buffer (without imidazole) is used for a gel filtration on a S200 10/300 increase (*Cytiva*). The elution profile of the gel filtration features four peaks Figure 52). Then, we mainly used negative staining to qualitatively assess the homogeneity of our samples, as well as Analytical UltraCentrifugation to quantitatively determine the receptor oligomeric state which underlines the peak multiplicity in the gel filtration chromatogram. We showed that the first two peaks correspond to partially aggregated forms of the protein. The third peak corresponds to larger objects, potentially GPCR dimers. The last and larger peak corresponds to the monomeric ADGRV1β. Further characterization using mass spectrometry, and nanoDSF confirmed the integrity of the protein and its good thermostability (melting temperature of 45°C). Therefore, we have successfully expressed and purified ADGRV1β

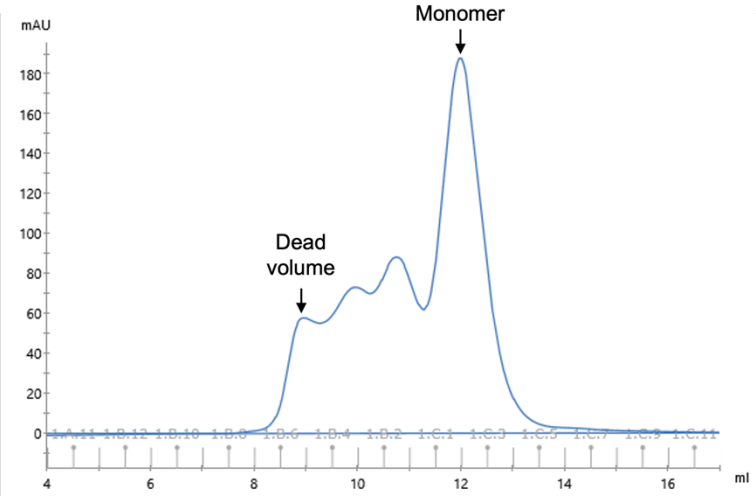


Figure 52 – Elution profile from the gel filtration column for ADGRV1β expressed in Sf9 cells.

from the baculovirus/insect cells SF9 system and purified them in LMNG/cholesterol micelles, as a monomer with a protein yield of 2mg per liter of transfected Sf9 cells.

As a first approach for cryo-electron microscopy using our sample, we directly used our gel filtration elution (in that case ~5µM equivalent to ~0.3mg/mL of protein) in order to screen freezing conditions and grid types. After several trials, we found that freezing is best performed when loading 3.5µL of sample onto the grid (glow discharged at 15mA for 25 seconds), blotting at 100% humidity and 12°C for 4 seconds with a blot force of 0. When using regular quantifoil or lacey grids, we observed adsorption of the protein onto the carbon surfaces, with few particles in the holes. Using UltrAufoil R1.2/1.3 grids solved this issue, maintaining ADGRV1β in the vitrified buffer filling the grid holes. UltrAufoil grids are made of a gold foil (R1.2/1.3) laying on a gold mesh grid, which dramatically reduced the adsorption of the protein on the grid structure. Data collection on a glacios 200keV microscope equipped with a falcon3 camera working in linear mode (NanoImaging Platform of the Institut Pasteur, Paris, France) showed that a homogeneous particle layer can be observed in grid squares of rather thick ice. In grid squares with thinner ice, however, a particle-depleted area is found in the middle of the holes (Figure 53).

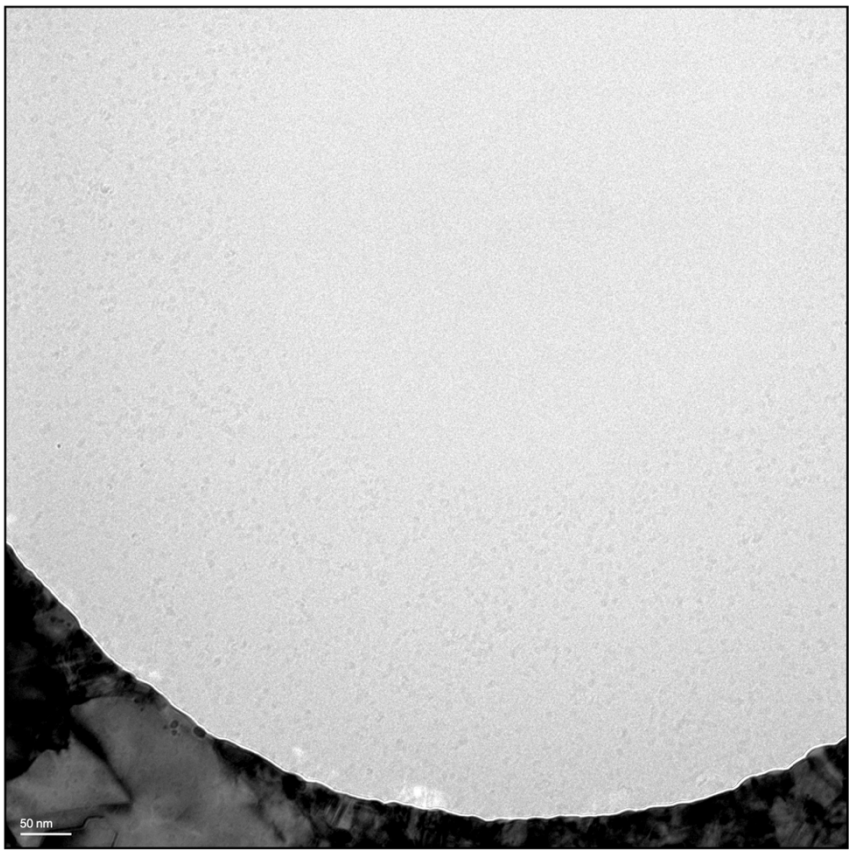


Figure 53 – cryo-EM image (73k magnification) of purified ADGRV1β at 0.01% LMNG and 0.001% CHS on an UltrAufoil grid. Image was collected on a glacios 200keV microscope equipped with a falcon3 camera working

This is detrimental for the quality of the generated datasets, as thick ice may yield more background noise, less contrasted particles and is more likely to be partially non-vitrified. For these experiments on ADGRV1β without effector, the sample concentration used is satisfying.

Another factor that may contribute to an increased local noise around particles is the detergent concentration used in our buffers. At 0.01%, LMNG is above its critical micelle concentration (CMC; 0.001% for LMNG in water). Proteins are therefore likely included in larger micelles, contributing to local noise, and empty micelles (without proteins) are likely to be found in the sample, which have to be removed during processing. Therefore, we decided to adjust our final gel filtration buffer to 0.00075% LMNG (0.000075% CHS), a detergent concentration used in the four articles recently published on adhesion GPCRs activation mechanism study by cryo-EM[201–204]. At this concentration, LMNG is below its CMC and only detergent monomers should be found bound to our protein. Elution profiles and yield are similar using high (0.01%) and low (0.00075%) detergent concentrations. The method for grids preparation is the same as previously described. We collected new images on a glacios 200keV microscope, recently equipped with a Falcon4i camera working in counting mode (Figure 54). Using this preparation method, we observed a similar particle density in our grids compared to the previous samples. Decreasing the LMNG concentration in the buffer also allowed for a homogeneous distribution of particles in holes even of very thin ice. Therefore, we decided to keep these conditions for the study of the complex between ADGRV1β subunit and its G protein partner.

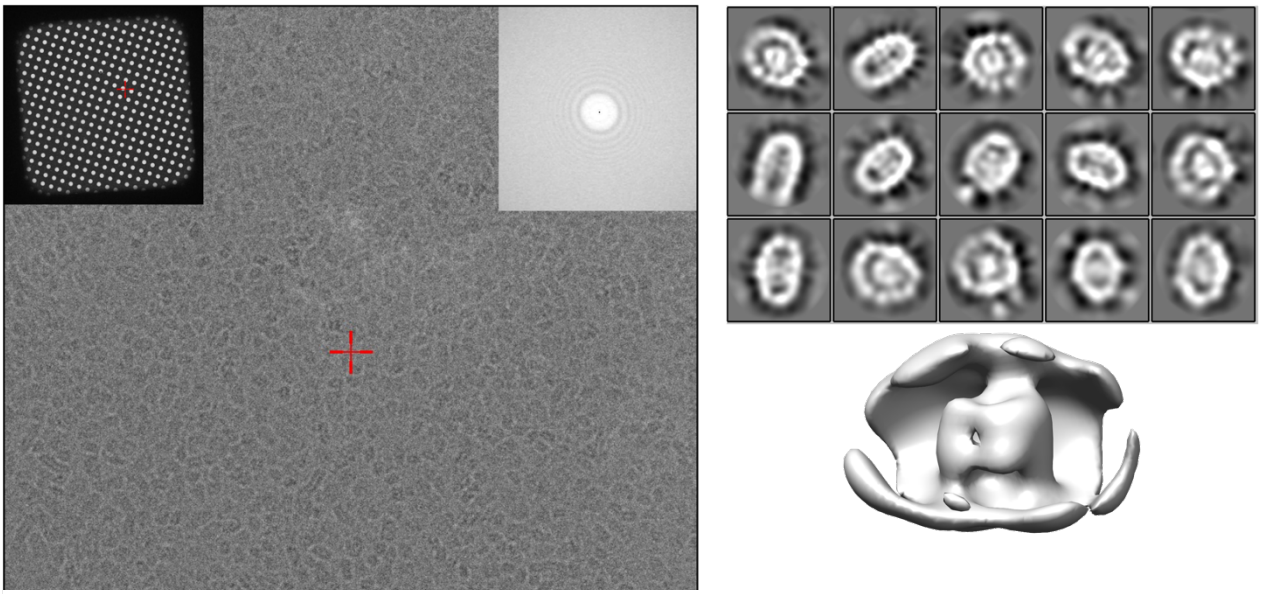


Figure 54 – cryo-EM image (150k magnification) of purified ADGRV1β at 0.0075% LMNG on a UltrAufoil grid. Image was collected on a glacios 200keV microscope equipped with a falcon4i camera working in counting mode. The defocus of the presented image is of -3 microns. After 2D classification (top right) an initial model can be obtained, with a density likely corresponding to the receptor 7TM embedded in LMNG (102967 particles used for model building).

According to the limited literature on the signaling of ADGRV1, the receptor would associate to multiple G proteins[189–191], with $G_{\alpha I}$ proteins as recurring partners[189,191]. Therefore, we decided to firstly use $G_{\alpha i}$ as partner for the study of ADGRV1 in complex. We received a first sample of $G_{\alpha i-1}$ purified by Alexandre Bouyssou at the Institut de Génomique Fonctionnelle (IGF) in Montpellier. We used mass photometry to assess complex formation at different receptor to G protein ratios. Mass photometry measures particles sizes in solution based on light scattering by single molecules. From this experiment, the GPCR appeared over-saturated at a 3 times excess of G protein.

As a first test for grid preparation, we decided to directly mix our purified ADGRV1β sample to the received G protein sample without further purification of the complex. For this complex, the final concentration of GPCR was kept the same as in our previous tests (5µM) and the G protein was added at a 15µM, for a final molar ratio of one to three. The complex was let one hour on ice before grid preparation. Again, we collected images on a glacios 200keV microscope equipped with a Falcon4i working in counting mode (Figure 55). This first test confirmed that the complex at least partially forms. This is illustrated in the 2D classification featuring the G protein alone, the GPCR alone, and the GPCR with an additional density not previously observed most likely corresponding to the GPCR/G protein complex. While not converging towards a high-resolution structure, the 3D models highlight two globular densities for the two proteins forming this complex.

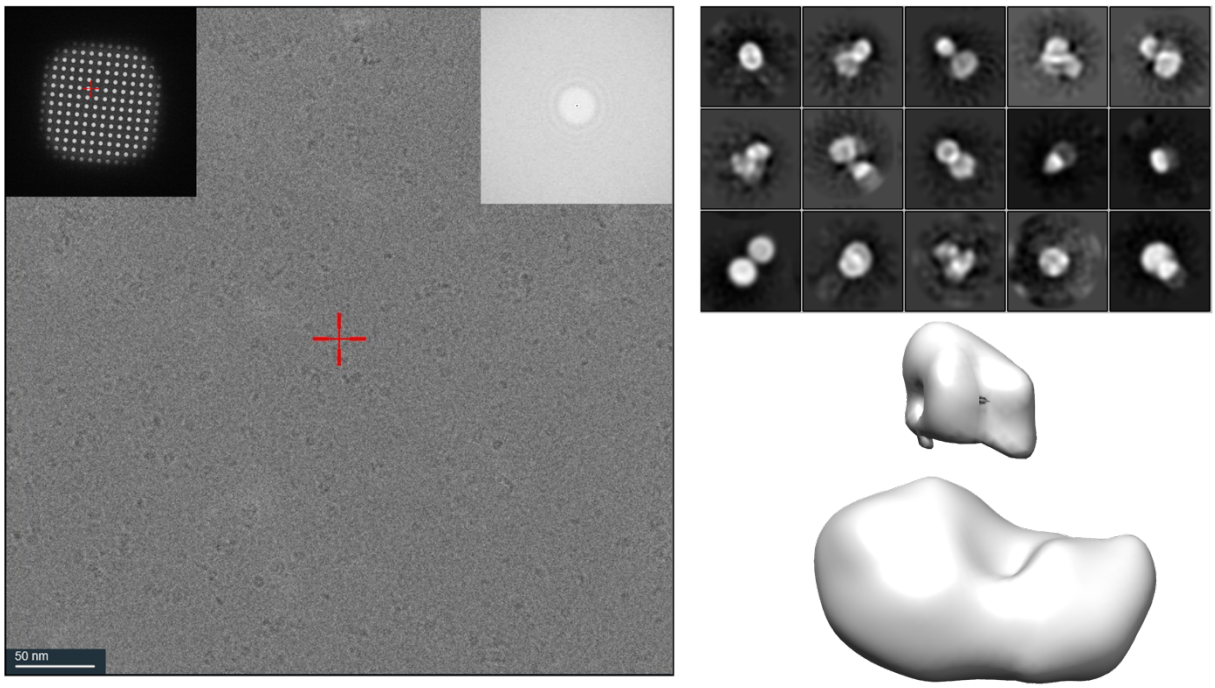


Figure 55 - cryo-EM image (150k magnification) of purified ADGRV1β at 0.00075% LMNG/0.000075% CHS in complex with a 1.5 excess of G protein on a UltrAufoil grid. Image was collected on a glacios 200keV microscope equipped with a falcon4i camera working in counting mode. The defocus of the presented image is of -3 microns. After 2D classification (top right) an initial model can be obtained, with two densities likely corresponding to the receptor 7TM and G protein (324818 particles used for model building).

However, the apparent particle density on the grids dramatically decreased compared to ADGRV1β alone, while the total protein concentration in the sample was higher. We did not observe protein aggregates on the grid either. This behavior is reproducible from one sample to another, and we have no obvious explanation to how it occurs. Therefore, we needed to produce a more concentrated sample of the complex in order to achieve satisfactory particle density on the grid. This conclusion is in line with sample preparation protocols available in the literature, with GPCR complex concentration often in the tenths of milligram per milliliter. The preparation of a concentrated sample is in progress, thanks to the team of Sébastien Granier (IGF, Montpellier), who sent us more purified G protein, as well as baculovirus stocks for us to express and purify the G protein subunits in insect cells. Yasmine Achat is currently purifying ADGRV1β/G protein in LMNG/CHS.

Further work on this project includes the characterization of the ADGRV1β subunit in complex with its G protein effector. This first step should provide insights on the activation mechanism of ADGRV1, which agonist stachel sequence diverges from other adhesions GPCRs. Such structure may also provide an insight on its cytoplasmic domain behavior, notably whether it remains free in solution (no corresponding density in the map), if it makes contact with the G protein or adopts a stable conformation. Such observation would provide potential explanations for the PDZD7-mediated regulation of the GPCR activity. Such approach may also be applied to complexes with proteins of the arrestin family, which could interact with the constitutively active ADGRV1, even though arrestin-mediated signalization or internalization is to be demonstrated. Finally, ternary complexes with other Usher 2 proteins constructs will also be considered in order to get closer to understanding the full protein network**.** In particular, a ternary assembly with PDZD7 PDZ tandem will allow to document the downregulation of ADGRV1 activity. However, if such approach may yield a structural description of the proteins interplay, it is unlikely to capture the network assembly in the unique environment of the hair cell stereocilia. The next section will treat of *in situ* imaging techniques aiming at describing the Usher 2 complex localization over the course of the inner ear development, as well as describing its molecular organization in the context of the cell.

III – *In situ* characterization of the Usher 2 complex

In mouse cochlear hair cells, the stereocilia ankle-links are detected between postnatal days 2 to 12 (P2 to P12)[164], while the inner ear is considered fully matured at P25. These extracellular links are thought to consist of the large extracellular regions of usherin and ADGRV1. At the anchoring site of these links are notably found the cytoplasmic proteins whirlin and PDZD7. The fine spatial localization of the complex components has poorly been characterized, as generally described localizing at the base of all stereocilia. Moreover, the localization of the different complex components over time has not been studied. In particular, the potential dissociation of ADGRV1 α (extracellular region) and β (GPCR 7 TM and cytoplasmic domain) subunits has not been confirmed *in vivo*. If it occurs, this may correspond to a timed activation of the adhesion GPCR, potentially in response to a specific stimulus.

Here I will first present preliminary fluorescence microscopy experiments aiming at characterizing the localization of Usher 2 complex components at different stages of the mouse cochlear development. Thanks to the specificity of the probe, this imaging technique allows for the identification of the structures of interest. Taking advantage of a relatively simple sample preparation, we can thus study the distribution of individual proteins in the whole hair bundle of both IHCs and OHCs at different mouse ages. The main objective is to identify fine localization of the four Usher 2 proteins, as well as monitor expression changes during the maturation of the cochlea. This includes determining which proteins may form the fibrous ankle-links, the distribution of those links between stereocilia and the distribution of associated cytoplasmic proteins, in particular PDZD7. Finally, we aim at monitoring the potential autoproteolysis of ADGRV1 in hair cells.

Then, I will describe the sample preparation steps performed for soft X-ray and cryo-EM tomographies on dissected tissue at P5. The objective is to characterize the Usher 2 network organization within hair cell stereocilia. These techniques perform a full 3D imaging of the sample, highlighting all features present. While powerful methods, additional information is often required to interpret the data and identify the structures of interest. Therefore, to better understand the obtained results in the context of a maturing cell, we will combine them with our *in vitro* structural data and *ex vivo* fluorescence imaging. Moreover, these techniques work with samples, tissues in our study, that are very challenging to prepare.

For all experiments, the sample primarily consists of mouse (OF1 strain) organ of Corti, which encompasses the hair cells. I will here briefly describe the dissection protocol. For mouse pups, euthanasia is directly performed by decapitation. For adult mouse, we

performed euthanasia by CO_2 inhalation or cervical dislocation prior to decapitation. The subsequent steps are identical for juvenile and adult individuals. After decapitation, the skull is opened and the brain removed to easily visualize inner ears on both sides of the head. The inner ears are packed onto the skull (Figure 56A). Pieces of skull where the inner ears are sitting are cut and placed in PBS (Figure 56B). The cochlear bony labyrinth consists of a spiralling central section (Modiolus) containing the membranous labyrinth, closed on the outer turn by a shell-like bone enclosing (Figure 56C) the whole structure. The first step is to remove this outer shell, exposing the membranous labyrinth (Figure 56D). The membranous labyrinth is a closed tube, that can be opened by removing its outer half, the *stria vascularis* (SV) and exposing the organ of Corti (OC). The organ of Corti may then be removed from the Modiolus (Figure 56E). The isolated mouse organ of Corti is a spiralling organ of about 2 millimeters in length, encompassing the hair cells, which can be fixed and labelled depending on the needs of individual experiments (Figure 56F). Before imaging, the last step consists in removing the tectorial membrane covering the hair cells. This can be performed at any point before the sample is mounted for fluorescence microscopy or frozen for further preparation steps required for tomography experiments. An example of organ of Corti labelled with an antibody against ADGRV1 extracellular region is presented in Figure 56G.

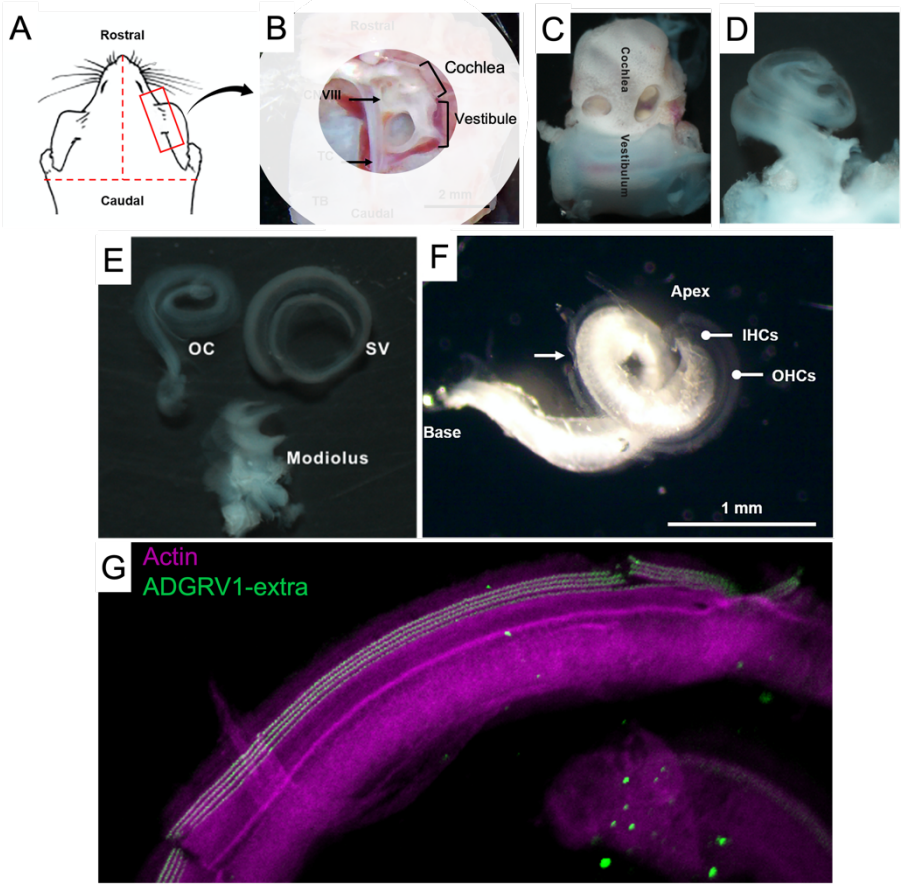


Figure 56 - The dissection steps of a mouse organ of Corti.

III.1 – Fluorescence microscopy

Our fluorescence microscopy approach has two main objectives. The first one is to finely describe the distribution of Usher 2 proteins in hair cell bundles. The second one is to monitor changes in protein expression and localization upon disappearance of the ankle links.

To achieve a nanoscale resolution, we decided to use state-of-the-art STimulated Emission Depletion (STED) super-resolution microscopy, in collaboration with Gaël Moneron (Neuroscience Department, Institut Pasteur). STED microscopy is a scanning imaging technique, which resolution are inherently limited by the size of the illuminated area. It makes use of confocal microscopes instrumentation in which the sample is illuminated in one location at a time and out of focus fluorescence emission is blocked by a pinhole. This approach allows to record the fluorescence emission of a probe with a planar resolution around 180nm and an axial resolution around 500nm[365]. Confocal microscopy, however, is limited by the diffraction limit of light. In STED microscopy, the size of the emitting sample area is further reduced using a ring-shaped laser overlapping the excitation light. Molecules in the excited state that are within the STED laser are immediately brought back to their resting state by emission of a photon at the laser frequency through a phenomenon called stimulated emission. It is therefore possible to only filter the light (spontaneous emission) emitted by molecules in the center of the ring, inherently increasing the final resolution. STED microscopy, however, induces multiple excitation and stimulated emission of fluorophores as it scans the sample, dramatically bleaching the probes. The smaller the hole in the STED laser (corresponding to a potentially higher resolution), the more bleaching is induced. As a consequence, the measured intensity of fluorescence is usually lower than in conventional confocal microscopy, which constitutes a limitation of the technique. The final resolution thus also depends on fluorophores bleaching rate. In our experimental conditions, a planar resolution of ~30nm is expected. The gain in resolution from confocal to STED microscopy is illustrated Figure 57.

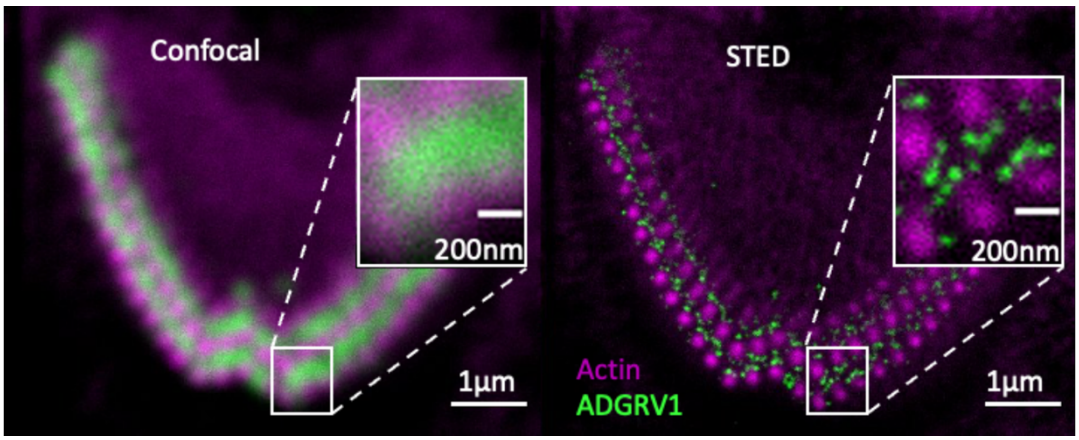


Figure 57 – Confocal and STED imaging of a same hair bundle labelled with phalloidin and an antibody against ADGRV1 extracellular region.

Preparation of the sample is as follow. After dissection, organs of Corti are fixed in an excessive volume of PBS + 4% ParaFormAldehyde (PFA) for at least 2 hours at 4°C, or until use. Before immunolabelling, the dissected tissue is washed of fixative using PBS, then incubated 30 minutes in PBS +20% Nature Goat Serum (NGS) and 0.1% Triton X-100 for blocking and permeation. After washing using PBS, the adequate primary antibody is added at a concentration of 0.5µg/µL in PBS + 4% Bovine Serum Albumin (BSA) and incubated overnight at 4°C. After washing with PBS, a secondary Anti-Rabbit IgG antibody coupled to a STAR RED dye (*Abberior*) diluted to 0.5µg/µL in PBS + 4% BSA, as well as 3 units (~0.5µM) of phalloidin coupled to an Alexa 594 dye (Invitrogen A12381) are added to the sample and incubated for 2 to 4 hours at 4°C. Phalloidin labels actin filaments, allowing to delineate the shape of the stereocilia. After extensive washing with PBS, the sample is mounted on regular microscopy slides using Mowiol supplemented with DABCO.

Before super-resolution data collection, we used conventional confocal microscopy to assess the quality of our samples. In particular, the labelling must yield enough signal to be detected by STED microscopy. We screened a series of commercial and custom antibodies against the four mouse Usher 2 proteins, usherin, whirlin, PDZD7 and ADGRV1.

For usherin, two commercial antibodies directed towards the extracellular region of the mouse protein were tested (sc-515555 from Santa Cruz Biotechnology; AP55368PU-N from Origene). In confocal microscopy, the signal was well located at the base of hair cell stereocilia, but the yielded signal was faint. Using STED microscopy, no signal could be detected. More antibodies directed against usherin orthologs will be tested.

For whirlin, we tested two commercial antibodies (sc-271508 from Santa Cruz Biotechnology; H00025861-M05 from Novus Biologicals), as well as six custom antibodies provided by Vincent Michel (Institut Pasteur). None of the eight antibodies gave satisfactory results, with faint or no labelling in hair bundles and high background noise in confocal and STED microscopy. We currently have no validated probe for detecting whirlin in mouse cochlear hair cells. As for usherin, more antibodies should be tested, despite being raised against non-mouse sequences.

Two commercial antibodies were tested for PDZD7 (bs-9042R from Bioss Antibodies; orb459604 from Biorbyt). Only Bioss Antibodies allowed for localized labelling at the ankle-links region. Intensity of the labelling is low, but still detectable by STED microscopy.

In order to monitor the potential hydrolysis of ADGRV1 in cochlear hair cells, we needed antibodies directed against the extracellular or intracellular regions of the protein. For the extracellular region, we tested two commercial antibodies (PA5-85761 from Invitrogen;

A32340 from Boster Bio) and one custom antibody provided by Nicolas Michalski (Institut de l'Audition, Paris). Only the antibody from Invitrogen yielded satisfactory labelling, with very localized and high intensity signal. For the cytoplasmic domain of ADGRV1, only one commercial antibody was available, which did not yield satisfactory labelling of the protein and is no longer for sale. Therefore, we ordered a custom antibody (*Agro-bio*) generated in rabbit against our purified construct of ADGRV1 cytoplasmic domain (full-length domain expressed in *E. coli*). After serum selection, we could retrieve an antibody yielding localized labelling in the hair bundles, despite some background noise remaining.

From this antibody screening, we focused on studying ADGRV1 and PDZD7. Using our antibodies, we aim at characterizing the localization of the ankle links (labelling of ADGRV1 extracellular region), monitor the autoproteolysis of the GPCR (comparing labelling of ADGRV1 extracellular and intracellular regions during development) and the potential regulation of either form of the receptor by PDZD7 (labelling of PDZD7 at different stages of development). Furthermore, we can readily compare the results obtained in inner (IHCs) and outer (OHCs) hair cells. This information might already shed light on the temporal regulation of Usher 2 mediated signaling.

III.1.a – IHC labelling

The labelling obtained on IHCs at P5 with our three probes is presented Figure 58. Antibody labelling is colored in green and actin labelling using phalloidin colored in magenta. In these top views, each magenta spot corresponds to a stereocilia, clearly delineating the only two rows of grown stereocilia of IHCs. The focus is adjusted to yield the highest signal for the antibody directed against the protein of interest. We observed labelling for ADGRV1 extracellular region in between stereocilia of the outer row, as well as between stereocilia of different rows (Figure 58A). No signal was detected between stereocilia of the inner row. This result indicates that ankle links are always connected on one end to a stereocilia of the outer row. Furthermore, using our antibody against ADGRV1 cytoplasmic domain, we obtained a similar labelling close to the actin core of outer row stereocilia (Figure 58B). These results indicate that, in IHCs, the GPCR 7TM is inserted only in the outer row stereocilia, with the extracellular region oriented towards stereocilia of either row. This may suggest a role for the GPCR in hair bundle row differentiation of IHCs. Finally, the antibody against PDZD7 yielded a similar crown-like labelling of the outer row only, supporting a role for the association of the PDZ-containing protein with the 7TM and cytoplasmic domain of the GPCR.

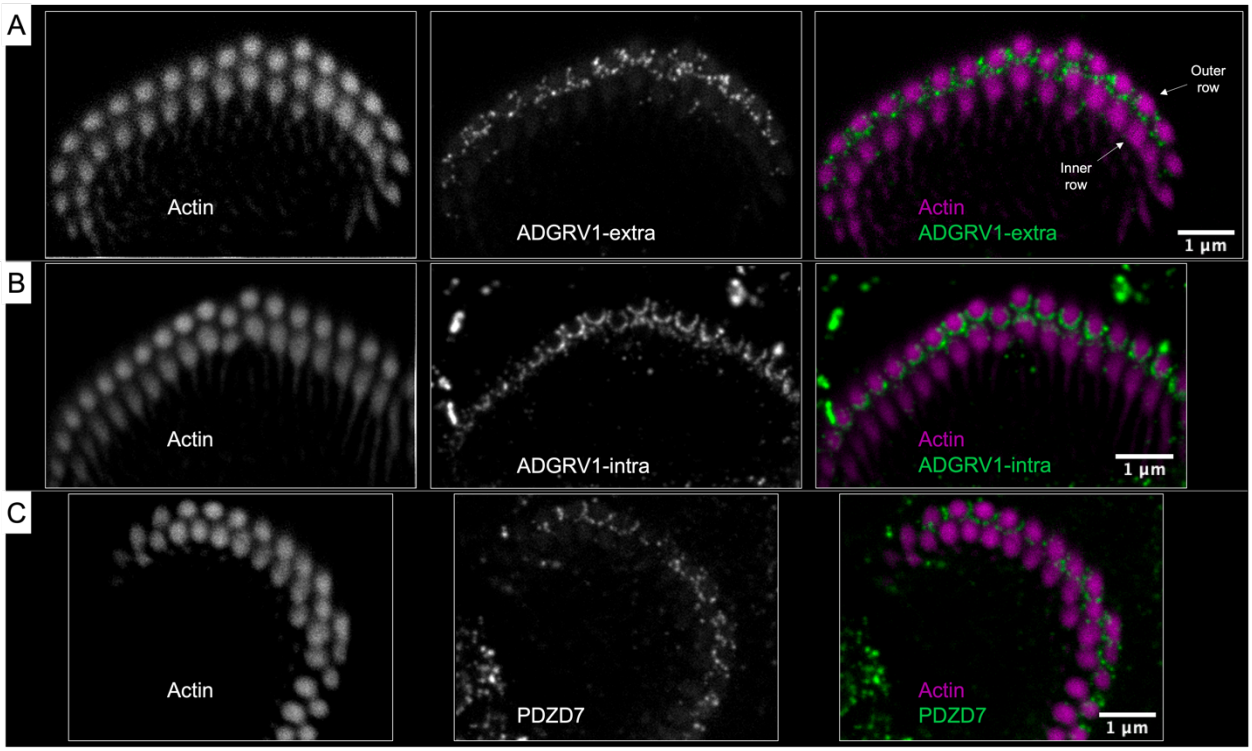


Figure 58 – Immunolabelling of ADGRV1 extracellular region (A), ADGRV1 intracellular region (B) and PDZD7 (C) in P5 mouse IHCs. Top views.

A side view of a IHC hair bundle is presented Figure 59. This image highlights the height difference between the two rows, with the outer row being approximately twice the size as the inner row. In green is displayed the labelling obtained with the antibody against ADGRV1 extracellular region. Here, the corresponding signal consists in only one band spanning across the hair bundle, consistent with ankle-links settling at a given distance from the surface of the cell. For the shorter row, labelling is located along the second quarter of the stereocilia length, from ~25% to ~50% of their height. For the outer row, this corresponds to the second eight of their length (from ~12.5% to 25%). In both cases, this corresponds to a position located just above the taper region.

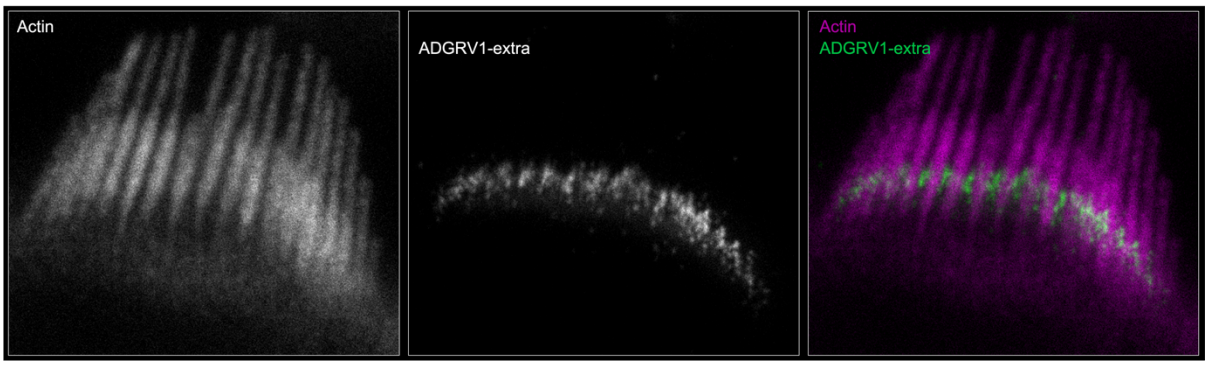


Figure 59 - Immunolabelling of ADGRV1 extracellular region in P5 mouse IHCs. Top view.

III.1.b – OHC labelling

We used the same approach and labelled OHCs with our three selected antibodies. OHCs hair bundles consist in three rows of stereocilia, increasing in height from the inner to the outer row. For top views (Figure 60), all three antibodies yield labelling between the inner and middle row, as well as between the middle and outer row of stereocilia. Even though our images are of lower quality for ADGRV1 extracellular region and PDZD7 labelling (Figure 60A, C), none of the three antibodies yield the ring-like structures observed around the taller row in IHCs. This indicates that the ankle links span mainly between stereocilia of different rows. Our antibody directed against ADGRV1 cytoplasmic domain illustrates this result (Figure 60B), where discrete spots can be resolved on each side of the inter-row spaces, but not between stereocilia of a same row. Such labelling indicates that the 7TM and cytoplasmic domain of the GPCR are found at each end of the links, with the protein oriented in both directions. This result also differs from IHCs, where the GPCR 7TM was only found in stereocilia from the outer row.

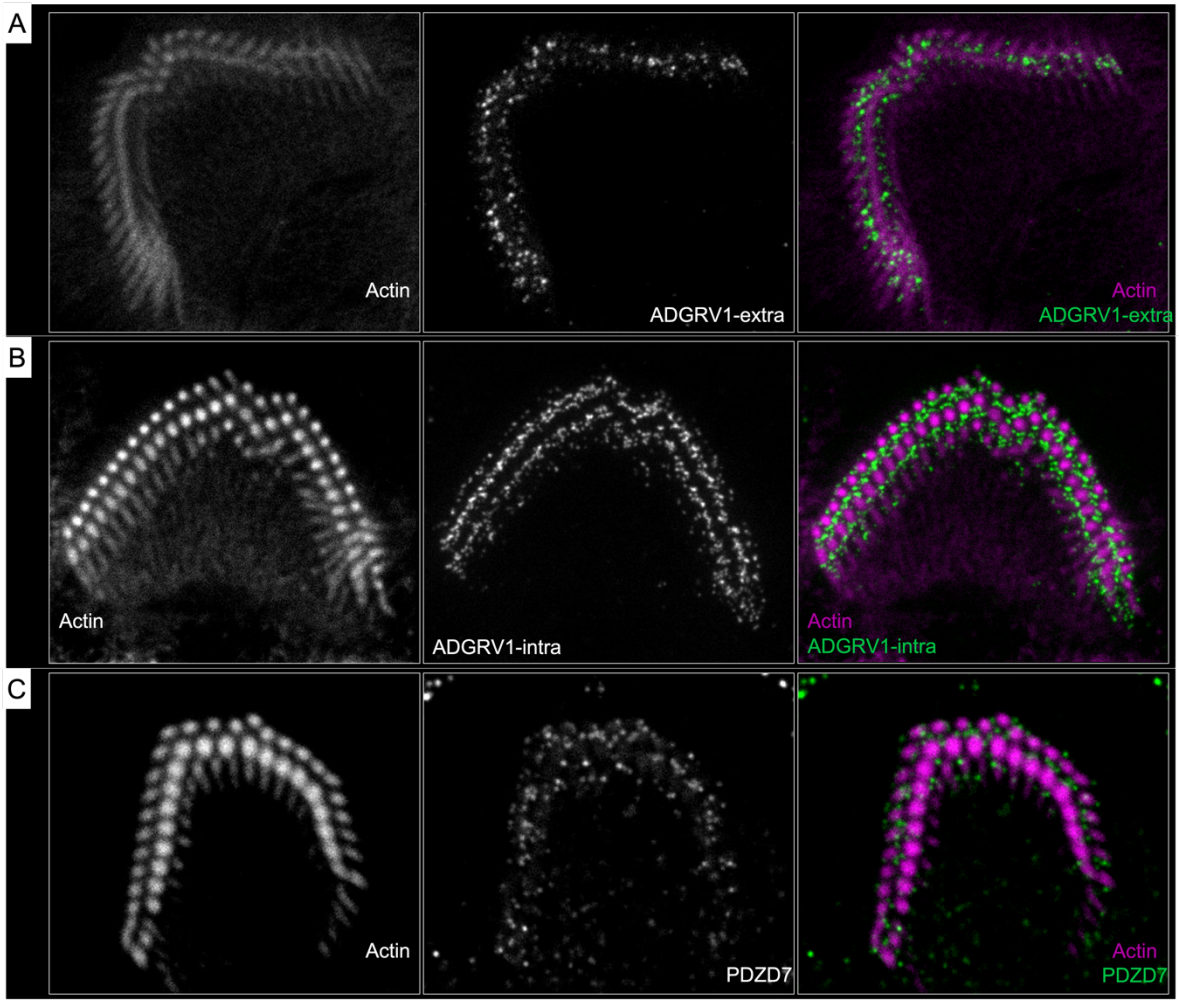


Figure 60 – Immunolabelling of ADGRV1 extracellular region (A), ADGRV1 intracellular region (B) and PDZD7 (C) in P5 mouse OHCs. Top views.

Finally, a partially tilted view of an OHC with ADGRV1 extracellular region labelling also highlights two bands of signal (Figure 61). However, due to the three rows and difficulty to obtain an ideal OHC side view, signal localization is more difficult to decipher using 2D images compared to IHCs. In this case, generating 3D stacks may help, though no gain in resolution for the z axis is obtained using 2D STED compared to confocal microscopy. However, obtained images can be discussed qualitatively. It appears that labelling of ADGRV1 extracellular region is still located just above the taper region. The localization of the ankle links relative to the cell surface is similar to what was observed in IHCs, though it appears higher relative to the stereocilia height, spanning from the middle of the smaller inner row.

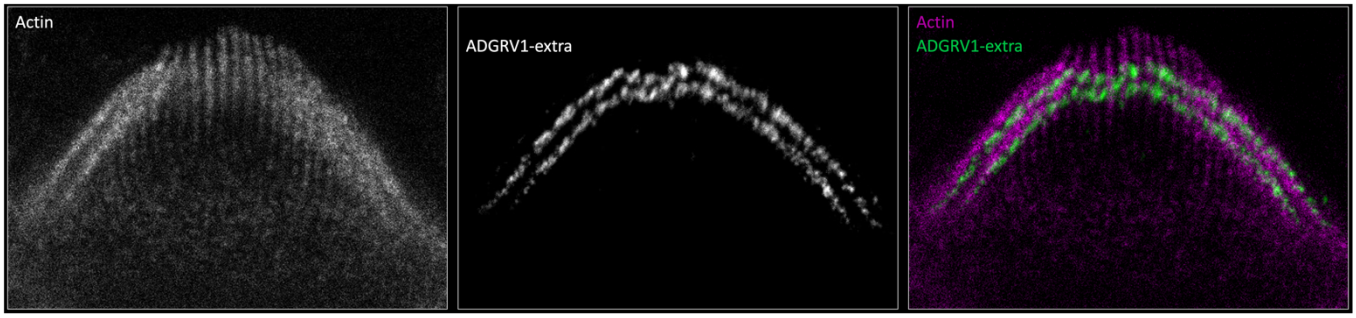


Figure 61 - Immunolabelling of ADGRV1 extracellular region in P5 mouse OHCs. Top view.

Across all our images, ADGRV1 and PDZD7 proteins labelling always localize above the taper region, rather than at a fixed position along the stereocilia. Presence of the ankle links at a fixed location from the cell surface might confirm an interplay of the Usher 2 complex with proteins expressed at the taper, consistent with some interactions demonstrated *in vitro*. To go further and confirm differences between protein expression patterns between different cells, we need to develop a method for image analysis allowing to quantitatively describe the distribution of our probes within the hair bundles. We are currently working on an automatized approach to average intensity profiles for a large number of cells. Having a robust approach would allow to quantitively describe changes of protein localization over a time course from P0 (birth) to P25 (accepted cochlear maturity in mouse). Notably, we are interested in comparing the labelling obtained with our antibodies directed against the extracellular and intracellular regions of ADGRV1 throughout this period. According to preliminary experiments, we showed that the β subunit of ADGRV1 is still detected in mature cochleas (Figure 62), while the extracellular region disappears. This observation confirms a hydrolysis of the receptor, which most likely corresponds to the specific cleavage of ADGRV1 in the GAIN domain. Cleaved receptor may display an increased activity in stereocilia, as described for several stachel-activated aGPCRs. This is the first evidence of the cleaved form of

ADGRV1 being detected in mature cochlear hair cells. In the literature, Usher 2 proteins are described as being exclusively expressed during the development of the cochlea. The most likely explanation is that ADGRV1 β subunit is maintained for some time after disappearance of the ankle links. As PDZD7, an ADGRV1 regulator, is also downregulated during the later stages, this may open avenues for GPCR signaling mediated by the receptor in the early mature stages.

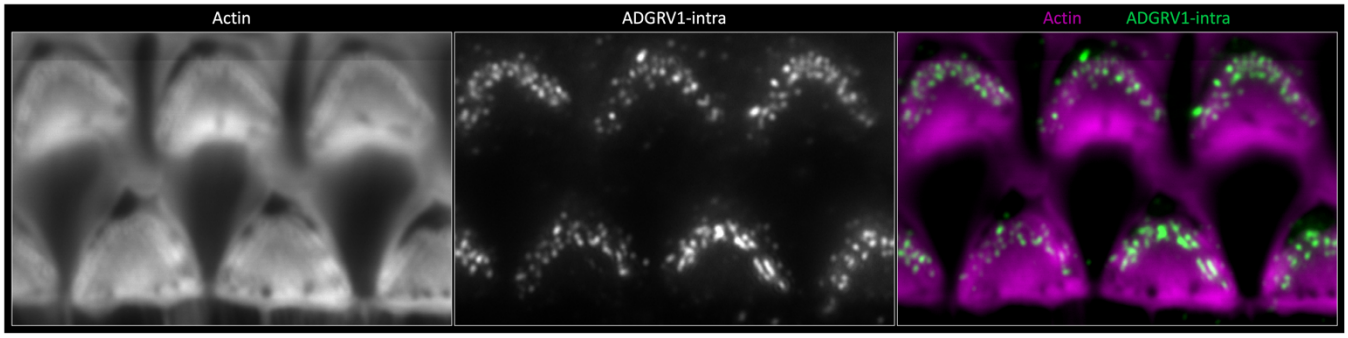


Figure 62 – Immunolabelling of ADGRV1 intracellular region in mature mouse OHCs by confocal microscopy. Top views.

III.2 – Soft X-ray and cryo-EM tomographies

With experiments presented thus far, we aimed at describing Usher 2 proteins interactions at the molecular level, as well as describing the cellular context for the complex at the cellular scale. However, these results are not sufficient to decipher the complex assembly from a structural standpoint. To bridge the gap between our cellular and *in vitro* experiments, we sought to use soft X-ray and cryo-EM tomographies approaches for studying macromolecular complexes *in situ*. Soft X-ray tomography allows for the imaging of large field of views, up to a thickness of 10-15 microns, with a resolution in the tenths of nanometers (25-50nm)[366]. Cryo-EM tomography is much more limited in terms of field of view, with sample thickness ideally below 200nm. The obtainable resolution, however, is in the nanometer range (2-10nm). Further processing of a given (and relatively abundant) feature by sub-tomogram averaging may yield resolutions in the range of 5-10Å, sometimes even higher. By imaging the Usher 2 complex at different resolutions, we ambition to have a near-complete structural description of its main components in the cellular environment. However, *in situ* imaging of protein complexes within cells is challenging and approaches are still being developed in multiple laboratories. Here, we ambition to image a protein complex within a dissected organ from mouse. This has, to our knowledge, never been previously published in the literature, with few available examples of tissues from *Caenorhabditis elegans* or

Drosophila melanogaster larvae[367,368]. This work is made in collaboration with Pierre Montaville, Frédéric Jammé and Pierre Legrand from the SOLEIL synchrotron (Gif-sur-Yvette, France), Eva Pereiro Lopez from ALBA synchrotron (Barcelona, Spain) and Stéphane Tachon from the NanoImaging Platform of the Institut Pasteur (Paris, France).

The sample preparation procedure for the two imaging techniques is essentially the same until the final thinning of the sample. The organ of Corti is dissected from juvenile mice and fixed in PB pH7 20mM containing PFA. At this step, immunolabelling of the sample may be performed as described for fluorescence microscopy experiments. In any case, a 10 minutes incubation with fluorescent phalloidin must be performed in order to label the hair bundles. The dissected tissue is then incubated with 20% glycerol (cryoprotectant, in PB pH7 20mM) and all following steps are performed in this same buffer. As mentioned, soft X-ray and cryo-EM tomographies both make use of vitrified samples. For vitrification of tissues, the sample must be frozen at high pressure. For high-pressure freezing (HPF), the sample is placed in a glow-discharged 3mm type B HPF carrier recess subsequently filled with buffer, and closed with the flat side of a type A HPF carrier coated with hexadecene (Figure 63). The jointed two carriers form a liquid-filled cavity containing the sample, with one side hydrophilic (glow-discharged) and the other hydrophobic (covered with hexadecene). Vitrification of the sample is performed using Leica's EM ICE instrument, allowing to reach liquid nitrogen temperatures (-196°C) in less than 20 milliseconds and thus preventing the formation of ice crystals.

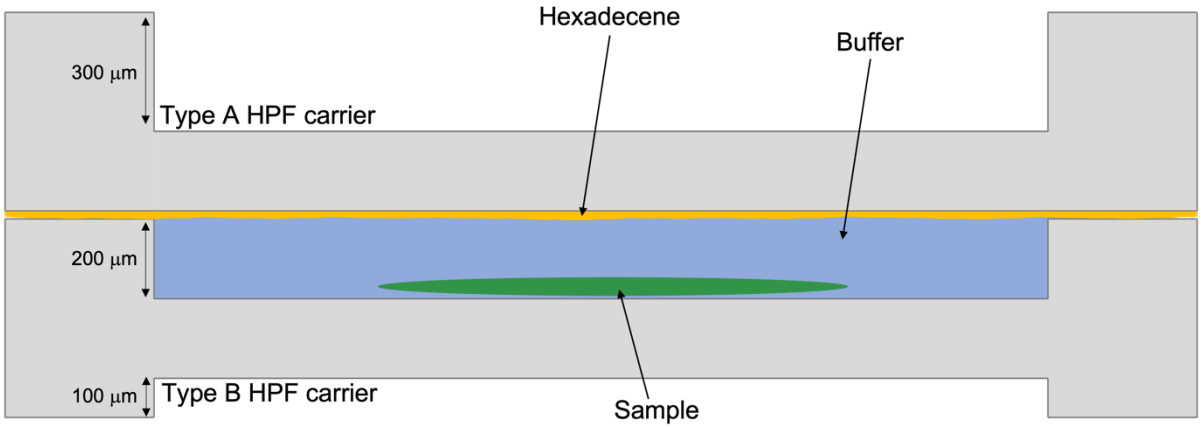


Figure 63 – Scheme of a High Pressure Freezing sample montage between two carriers.

Carriers are then retrieved and opened in liquid nitrogen. The type B carrier is stored in classical grid boxes until use. 3mm HPF carriers can directly be loaded in an Aquilos cryo-FIB microscope (*Thermofisher*). The microscope associates scanning electron microscopy (SEM) imaging, fluorescence imaging (integrated Fluorescence Light Microscope; iFLM) and Flux Ion Beam (FIB) milling. SEM imaging allows to monitor procedures performed on the

specimen within the microscope chamber. The iFLM is necessary for the identification of the structures of interest within the vitrified ice. FIB milling is used to remove parts of the sample with high accuracy using a focused beam of gallium ions. It is a required step to extract blocs of sample using the lift-out procedure (Figure 64). In this approach, the sample around the region of interest is milled away using the flux of gallium. The sample is then glued, using a sputter of platinum (Pt), to a micron-thin needle manipulated at high precision within the chamber of the microscope (easyLift system). The needle can then be lifted away from the sample, taking a piece of tissue containing the structures of interest (lift-out). The bloc is finally glued again on a half EM-grid to be thinned and polished before imaging. Again, the final thickness of the lamella depends on the imaging method and can be of about 15 microns for soft X-ray tomography, while of a maximum of 200nm for cryo-EM tomography.

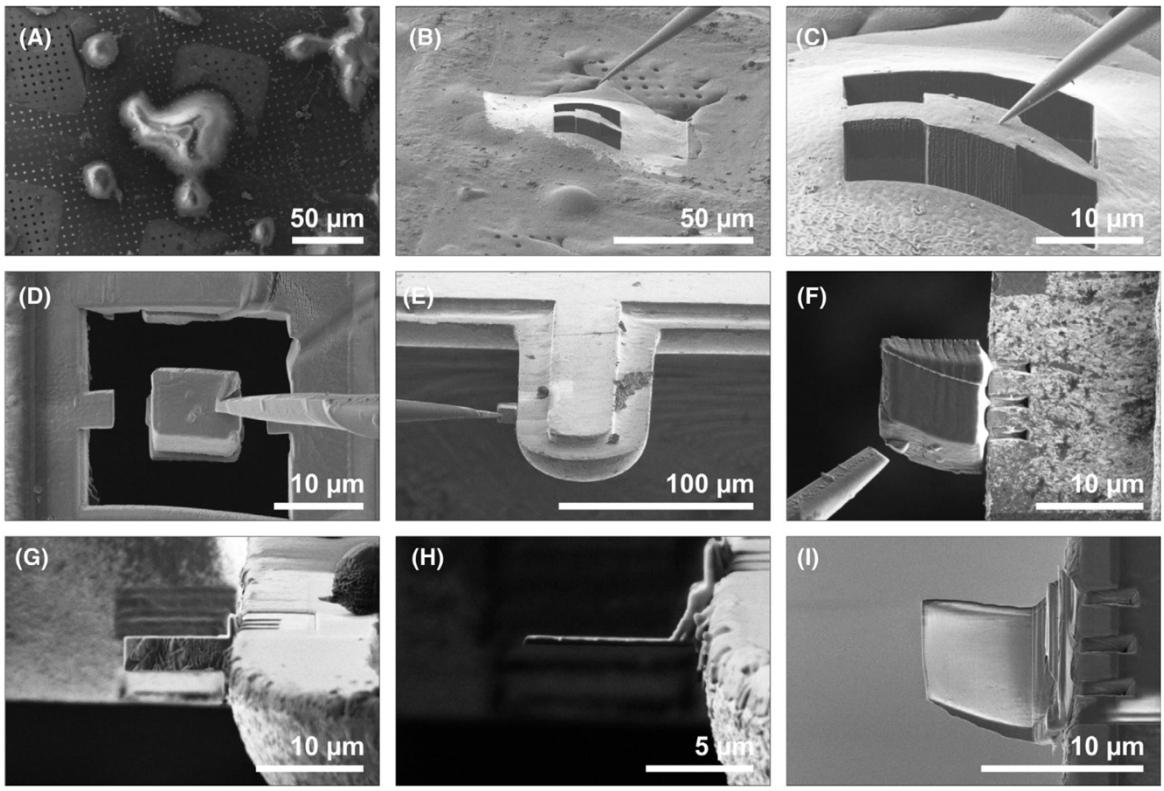


Figure 64 – Step-by-step cryo-lift-out procedure. (A) Vitrified cells. (B) Milling of a chunk for lift-out. (C) Attaching the EasyLift to the coarse lamella by cold Pt deposition. (D) Cutting the coarse lamella. (E) Attaching the coarse lamella to the post of a half-moon grid. (F) Cutting the needle free. (G) Thinning the coarse lamella. (H) Final thinned cryo-lift-out lamella. (I) Top view of cryo-lamella. Adapted from Kuba et al. (2021)[369].

For the first tests, P5 mouse organ of Corti were fixed overnight in 2% PFA and 0.5% GA (GlutarAldehyde). Before freezing, the sample was incubated one hour in PB containing 20% glycerol as cryo-protectant. Pieces of organ of Corti were embedded in 2% ultra-low melting temperature agar (A5030, *Sigma*) placed in a 3mm type B HPF carrier (Figure 63). Type B carriers have a 200 microns recess on one side, and 100 microns recess on the other side. The pieces were layered in the inside of the 200 microns recessed after brief glow discharge of the carrier. The recess was filled with PB and closed using the flat side of a 3mm type A HPF carrier dipped in hexadecene (type A carriers have a 300 microns recess on one side and are flat on the other). The sample theoretically remains in the 200 microns cavity of the type B carrier.

After HPF, Martin Sachse (Ultrastructural BioImaging facility, Institut Pasteur, Paris) and Moara Lemos (Structural Image Analysis Unit, Institut Pasteur, Paris) performed the freeze substitution of the sample for imaging on a Tecnai F20 EM-microscope (*FEI*) at room temperature. Using this approach, they confirmed that the freezing procedure maintains the structures of the stereocilia and that the sample was well vitrified, without damages induced by formation of crystalline ice (Figure 65).

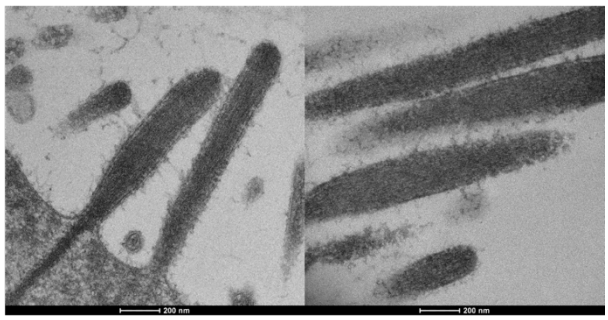


Figure 65 – Room temperature imaging of high-pressure frozen stereocilia after freeze substitution. Scale: 200nm.

After this initial test, I froze new samples and proceeded with the experiments in cryogenic conditions. For this second test, P5 cochleas were fixed for 2 hours in 4% PFA and then immunolabelled using the antibody directed against ADGRV1 extracellular region to further guide FIB milling. Labelling of extracellular epitopes allows to not permeabilize the sample and preserve membrane morphology, critical for high-resolution imaging. The labelling protocol is otherwise identical to our fluorescence imaging experiments. In brief, the sample was blocked using 20% NGS in PB for 30 minutes, labelled overnight using Invitrogen PA5-85761 primary antibody against ADGRV1 extracellular region (1/100 in PB +4% BSA), washed, and incubated 2 hours with Sigma anti-rabbit IgG secondary antibody coupled to Alexa 488 (1/100 in PB +4% BSA), as well as atto 565-conjugated phalloidin. Extensive wash using PB was performed before incubation in PB glycerol 20%. We observed during the first test was that a single whole organ of Corti would fit in the HPF carrier and would be big enough to remain still without agar embedding. For this second round of freezing, I thus froze the organ as a whole, allowing to better orient the sample within the carrier even without agar

and maintain its global morphology. The obtained samples were loaded in Institut Pasteur's Aquilos cryo-FIB microscope.

To our surprise, the HPF carriers appeared mostly empty, while we were expecting a full layer of ice containing the organ. Only one piece of tissue could be found and identified as a remaining tectorial membrane with imprints of hair cell stereocilia at the surface (Figure 66A). While not containing our structure of interest, we successfully implemented the whole lift-out procedure on several random blocs of tissue (Figure 66B). To our knowledge, this is the first time it is successfully applied to an animal sample besides *Caenorhabditis elegans*[367,370]. Yet, our sample preparation method did not appear satisfactory in order to preserve the hair cells and we proceeded to another assay of high pressure freezing.

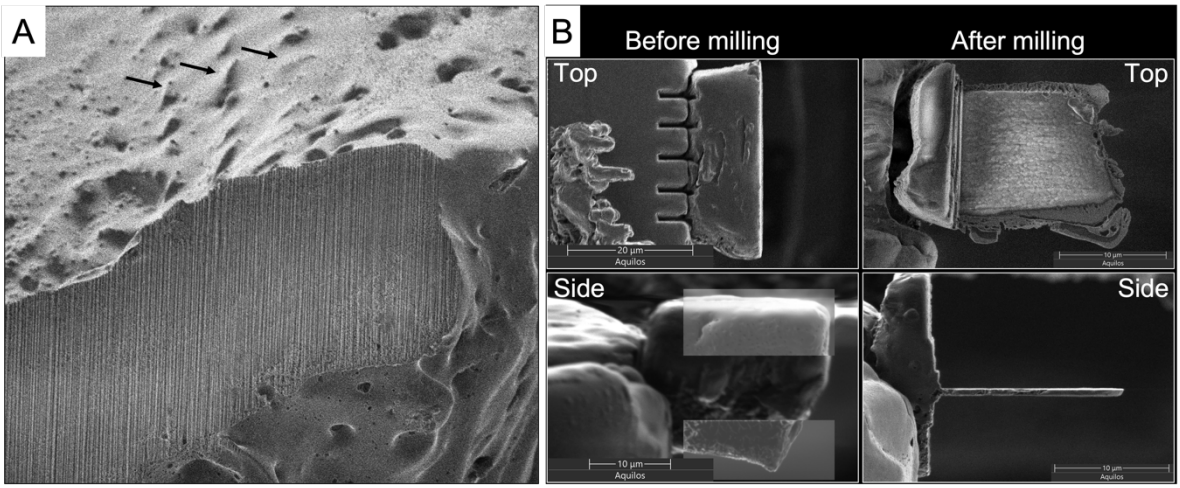


Figure 66 - A) Scanning Electron Microscopy image of a tectorial membrane. Imprints made by OHCs tips inserting in the membrane are indicated by arrows. B) SEM images of a block extracted from the tectorial membrane glued to the EM half grid, before and after milling. Presented lamella was milled to a thickness of 800 nanometers.

For this third test, we used P5 mouse cochlea fixed in 4% PFA for 2 hours and immunolabelled with our antibody directed against ADGRV1 extracellular domain, as well as phalloidin. The procedure used is largely the same as for the previous test, except that the entire cochlea is embedded in agar at the bottom of the HPF carrier recess. After carefully removing the tectorial membrane, the sample is positioned within the carrier. The buffer is gently removed by capillarity before immediate addition of a drop of 2% ultra-low melting temperature agar (in PB 20% glycerol). The agar shortly solidifies around the organ and a drop of ice-cold buffer is added to fill the recess completely. The protocol for subsequent freezing is unchanged.

Again, samples were loaded in the Aquilos microscope. These new samples visually met our expectations. By scanning electron microscopy, a rather flat and homogeneous surface can be observed, nearly filling the carrier recess. The organ itself can no longer, however, be visualized. To locate the organ within the bloc of ice, we used fluorescence imaging (iFLM), correlating 1) the image of the organ in fluorescence, 2) the image of the surface in fluorescence, 3) the image of the surface in SEM. Superimposing these three images allows to pinpoint the localization of the tissue with respect to the surface of the bloc (Figure 67). Imaging of the organ is possible using our antibody against ADGRV1 extracellular region or phalloidin to label actin filaments. Overall phalloidin gives a more intense signal, but ADGRV1 labelling confirms the preservation of the structures of interest.

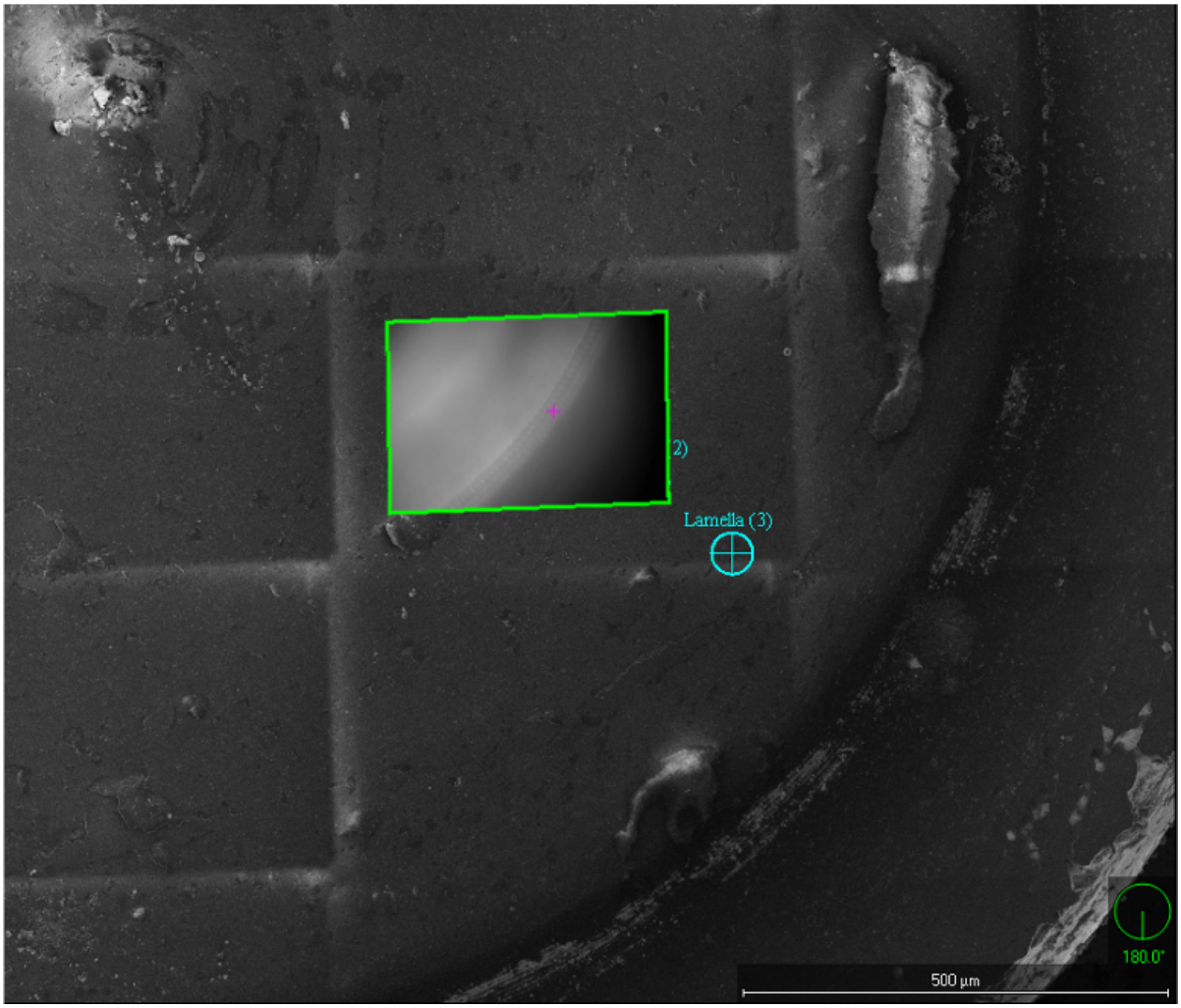


Figure 67 - Scanning electron microscopy image with correlated fluorescence image taken at the organ focus. Phalloidin labelling is detected. Individual hair bundles can be observed with surprising clarity.

Theoretically, the sample is ready to be used for the lift-out procedure. In practice, however, the accessibility of the tissue is compromised by the volume of vitrified buffer filling the carrier. Using the in-chamber iFLM imaging, we could estimate that the stereocilia are

located deeply within the sample, about 130 microns below the surface of the ice. Extracting a block from the organ with the easyLift needle in this condition is difficult and extremely time consuming, with the high risk of damaging the tissue. On this sample, we could lift a single block of tissue in three days of milling (about 10µm long, 6µm wide, and 5µm thick), which will be sent to our collaborators in SOLEIL and ALBA synchrotrons for imaging. For soft X-ray tomography, this may be sufficient as it is possible to image a large "lamella" where the probability of finding the region of interest is high (several cells in one block). For cryo-EM tomography, however, the probability of finding the region of interest within a 200nm thick lamella decreases substantially. A direct approach to obtain a thin lamella containing a structure of interest would be to increase the number of lamellas prepared. This will also be useful if we want to perform sub-tomogram averaging, which requires a large volume of data. However, our current preparation method does not allow to prepare multiple lamellas in a reasonable amount of time.

Now that the proof of concept is achieved, we have to optimize our sample preparation to facilitate the lift-out procedure and increase the sample preparation throughput. The most straightforward approach is to freeze the sample in the 100 microns recess of the HPF carrier, instead of the 200 microns one. If performed successfully, this should reduce the thickness of the vitrified ice above the sample from about 130 microns to 30 microns, which will dramatically reduce the time spent on milling buffer away and the risk to damage the tissue. Another solution would be to trim the surface using a cryo-ultramicrotome, but this adds sample handling hazards to the process and can be performed for the central region of the carrier only. Lastly, in combination with one of the two previous approaches, we may use soft X-ray tomography imaging as a preliminary step for cryo-EM tomography sample preparation. In that case, after extraction of a few microns thick block, the volume may be mapped by soft X-ray tomography prior to block thinning for cryo-EM tomography. This has the advantages to both provide cellular context to the EM tomograms, as well as accurately locate the region of interest for further FIB thinning. The main limitation of this approach may be damages caused by soft X-rays to the high-resolution structures, which could be detrimental if performing sub-tomogram averaging. Yet such data has never been produced.

Discussion

In this manuscript, I presented my results obtained on proteins associated to the Usher syndromes, which combine hearing and vision loss. Both cochlear hair cells and retinal photoreceptor cells are highly specialized. They have an apical structure mediating stimuli sensing and directly interface with the nervous system at their base by means of a ribbon synapse. Retinal cells are abundant, with millions of cells in an organ, and progress has been made in growing retinal-like cells *in vitro*. Cochlear hair cells, however, have not been successfully cultivated and no cellular models captures their highly organized cilia structure, which is critical for auditory protein functions. In the organ of Corti, only few thousands of hair cells are found (about 20 thousand in human), further divided in inner and outer hair cells of distinct functions. Lastly, each hair cell type is morphologically different depending on the localization along the axis of the cochlea, reflected by a differential expression of several proteins. The study of auditory protein interactions has therefore been conducted in cellular models, which fail to recapitulate the peculiar environment of hair cell stereocilia. Firstly, the hair bundle is unique in its architecture, which is reflected in the expression patterns of stereocilia proteins. Secondly, stereocilia are filled with a dense actin network, leaving little space for submembrane complex assembly and resulting in high local protein concentrations.

The proteins expressed in the stereocilia are also very specialized. They include very large transmembrane proteins that bridge stereocilia (cadherin-23, protocadherin-15, usherin, ADGRV1) and scaffold proteins (harmonin, SANS, whirlin, PDZD7) that assemble local cytoplasmic complexes, in turn interacting with multiple partners. Cadherin23 (370kDa) and protocadherin15 (216kDa) make the tip-link, a single link at the stereocilia apex responsible for triggering the opening of the mechanotransduction channel upon stimulation of the hair bundle. Usherin (576kDa) and ADGRV1 (693kDa) form the ankle-links at the base of the stereocilia. These links appear fibrous, and may serve as extracellular scaffolding for the stereocilia, as well as mediate signaling potentially in response to mechanical triggers. In knock-out animal models, lack of ankle-links is associated to hair bundle disorganization. The larger protein networks are found at the anchoring sites of these links, as well as at the tip of the stereocilia, as monitored by densities on electron micrographs. Interestingly, each of these complexes involve at least one of the three paralogous scaffolding proteins harmonin, whirlin and PDZD7. These proteins encompass two domain types, PDZ and HHDs, and have a similar polarized modular organization. Their N-terminal region contains a PDZ tandem, notably mediating the interaction with transmembrane proteins. The binding profiles of these PDZ tandems, however, differs. No ligand has been found for whirlin PDZ2 thus far, while both PDZ domains of PDZD7 can interact with ADGRV1, simultaneously binding to two

molecules. The N-terminal region also encompasses one or two HHDs, flanking the PDZ tandem (harmonin, whirlin) or in the central part of the protein (PDZD7). Their C-terminal region is mostly disordered and flexible, ending with a third PDZ domain that notably mediates interactions with actin-binding proteins, allowing for the tethering of the transmembrane proteins to the actin cytoskeleton of the stereocilia. PDZ domains are found in many proteins (152 in humans) and have been the subject of extensive studies. They are known to mediate homotypic and heterotypic protein-protein interactions, as well as lipid binding. On the other hand, HHDs are an orphan type of domain found in only 6 proteins, with relatively few functional and structural data. They derived from a shared ancestor with PAH domains, forming a subclass of helix bundles. For Usher proteins, HHDs also mediate homotypic and heterotypic protein-protein interactions, as well as lipid binding that reinforces the polarization of whirlin and PDZD7 proteins with their N-terminal regions oriented toward the membrane. HHDs and PDZ domains are prone to auto-association and allow for a similar array of interactions, each recognizing a different type of secondary structure from the partners (helix for HHDs, strand for PDZ domains). The two domain types also mediate interactions in a similar affinity range, with Kd values typically between few micromolar and few tenths of micromolar. Together, these domains can promote the oligomerization of the protein, interact with multiple partners and fine-tune the specificity of adjacent domains. At high local protein concentrations, as in the stereocilia, the multiplicity of interactions may allow for a dense, but dynamic, interaction network. This results in the formation of submembrane condensates by phase separation, with potentially stable interactions despite the medium-range affinities measured *in vitro*[276].

These interaction networks are notably thought to promote the stable anchoring of the inter-stereocilia links, ensuring a structural resilience to the mechanical constraints applied onto the hair bundle. The use of transient interactions, with middle range affinities, may allow for a mechanical adaptability of the protein complexes, as well as regulation and fine-tuning of the interactions, notably in a concentration dependant manner. In the Usher 2 complex, this may dynamically regulate signaling mediate by the aGPCR ADGRV1. Such signaling depends on the autoproteolysis state of the receptor[189,191], the G proteins and arrestins localized in the stereocilia compartment, and the expression level of other regulatory proteins, notably PDZD7. As previously shown *in vitro*, ADGRV1 associates to G_s/G_q proteins in its uncleaved conformation[189]. It is unclear if this is a basal property of the receptor, or if a yet unknown extracellular ligand may promote receptor activation in hair cells. After autoproteolysis, ADGRV1 could be activated in a stachel-mediated manner and associates to G_i proteins[191].

Structural characterization of the receptor in its different states, as well as functional assays, will shed light on its signaling activity. Lastly, autoproteolysis of the receptor may promote the interaction with arrestin proteins, with an independent set of possible signaling pathways. Therefore, monitoring the cleavage of the receptor and expression of the other complex components is critical to define the turning-points in the receptor signaling during the development of hair cells. In addition, to fully understand the function of the Usher 2 complex, there is a dire need of data concerning signaling pathways initiated by ADGRV1 and the elicited cellular response.

While ADGRV1 and PDZD7, only located at the ankle links, mediate both scaffold and signalling functions, the role of the other two Usher 2 proteins, usherin and whirlin, is unclear. Usherin is also only found at the ankle-links. It has no identified contribution to the signaling mediated by ADGRV1. The organization of the usherin extracellular region suggests a main role in structural scaffolding of the inter-stereocilia space. This extracellular scaffolding may be necessary for sensing inter-stereocilia distances during hair bundle morphogenesis, as well as provide an increased mechanical resistance against external constraints for the hair bundle, although no data is available. Usherin may also assist in structuring ADGRV1 extracellular domain, potentially making heterocomplexes crossing the inter-stereocilia space, in a similar fashion as to cadherin23 and protocadherin15 forming the tip-link. This potential role is sustained by our results on ADGRV1 distribution in IHCs, where the protein reaches only from the taller row of stereocilia, with an anchoring in the facing shorter row that is unclear. Usherin may contribute to the ankle-links bridging the inter-stereocilia space. Studying the protein localization at high resolution, as we performed for ADGRV1 and PDZD7 by STED microscopy, should help determining if that could be the case. Further structural characterization by cryo-EM tomography may also provide a more precise description of the ankle-links composition and structure. From the protein organization, whirlin appears partially redundant to PDZD7 combining HHD, PDZ and disordered domains. However, whirlin does not regulate ADGRV1 signaling[191]. In contrast to PDZD7, whirlin localizes in several locations of the stereocilia[275], which is reflected by its interaction network[34]. At the tip of the stereocilia, whirlin participates in a phase-separated condensate involving several actin-binding proteins and thought to regulate the elongation of the stereocilia[276]. At the lower anchoring site of the tip-link, it notably interacts with protocadherin15[55], likely tethering it to the actin cytoskeleton. At the base of the stereocilia, whirlin interacts with proteins of the Usher 2 complex[292], but also proteins of the taper region[34]. It has been suggested that whirlin would act as an intermediate to transport molecules

up in the stereocilia, by loading cargo onto myosin molecules. This indicates another potential function for proteins of the Usher 2 complex and taper region in regulating the import of molecules in the stereocilia. Such mechanism has been described for tubulin-based cilia[371,372], and could be necessary due to the confined environment and limiting diffusion from the cell soma, as well as the specialized nature of the cilium compartment itself. This is also consistent with the supposed function of usher proteins in retinal photoreceptor cells, where they are thought to bridge the tubulin-based transport machinery of the inner segment to the actin-based transport machinery of the outer segment. A transport machinery is required to translocate a variety of molecules in the stereocilia. Various isoforms of key proteins are expressed at different locations of the stereocilia and at different times during the development of the hair cells. These includes proteins of the mechanotransduction machinery, such as cadherin23, protocadherin15 and proteins forming the MET channel, with a switch from TMC2 as a core channel component at early stages to TMC1 in mature cells. After the hair cell development, the MET complex is thought to be poorly renewed, with damages causing definitive hearing loss in human.

Usher 2 is the most prevalent subtype of the syndrome. Variants of the four Usher 2 proteins have been identified. These variants often consist in protein truncations, abrogating part of the protein function or inducing expression defects. Interestingly, single-point mutations have also been identified in both PDZD7 and ADGRV1. For PDZD7, single-point mutations are located on the N-terminal region of the protein responsible for binding to ADGRV1. We showed that two of these deafness variants strongly weaken the interaction with the receptor, potentially upregulating ADGRV1 activity. On ADGRV1, human pathogenic variants include truncations of the cytoplasmic domain (one of them 12 residues upstream the C-terminal), as well as single point mutations in the 7TM core of the receptor, close to the ligand binding site. Truncation of the cytoplasmic domain likely affects the interaction with cytoplasmic proteins like PDZD7 or arrestins. Mutations in the 7TM, near the ligand (molecule, peptide or stachel sequence) binding site would affect the activation mechanism of the receptor. Finally, several mutations of unknown pathogenicity have been identified just upstream of the stachel sequence, including the histidine (H^{6037} in human sequence) required for autoproteolysis of ADRGV1. These mutations may thus prevent autoproteolysis, modifying mediated signaling and preventing internalization, but their associated physiopathogenesis is to be confirmed. Overall, the effect of the mutations on receptor function and/or expression and trafficking remain to be explored.

The Usher 2 complex as a whole appears to have multiple functions, including extracellular scaffolding, protein signaling and potentially protein transport. Therefore, in order to characterize the assembly of the Usher 2 complex, how it fulfils its varied functions, as well as further understand the effect of pathogenic mutations, we must use an integrative structural biology approach, complemented by functional assays:

- *in vitro* studies to characterize structural mechanisms underlying receptor activation and protein interactions at high resolution (NMR, X-ray diffraction, single-particle cryo-EM…),
- *in situ* structural approaches to decipher the complex assembly (soft X-ray and cryo-EM tomographies).
- Finally, *in situ* imaging at the cellular level to monitor alterations in protein distribution and hair bundle morphology (confocal and STED fluorescence microscopies).

I have attempted to implement such approach, using mouse as an animal model and multiple projects are still ongoing. The mouse organ of Corti is a great sample for such integrative projects. For developing tissue tomography approaches, it has a rather simple organization and a repeated pattern of cells of interest at the surface of the organ, which can be easily labelled. The simple geometry of the cell distribution should be beneficial for later increasing the efficiency for preparing cochlea lamellas. For fluorescence imaging, the high degree of hair bundle organization allows for unambiguously describing protein localizations and how they are affected by pathological mutations. To go further in understanding the function of Usher proteins, one should also consider initiating a similar work in the retina. Unfortunately, due to main morphological differences of the photoreceptor cells, mouse is a poor model for studying human vision. Zebrafish have proven to be a good model for mimicking the human vision phenotype associated to Usher syndrome and its genetics is relatively easy to manipulate compared to mammalian models. This may prove invaluable to decipher the physiopathogenesis of pathogenic variants.

Annex

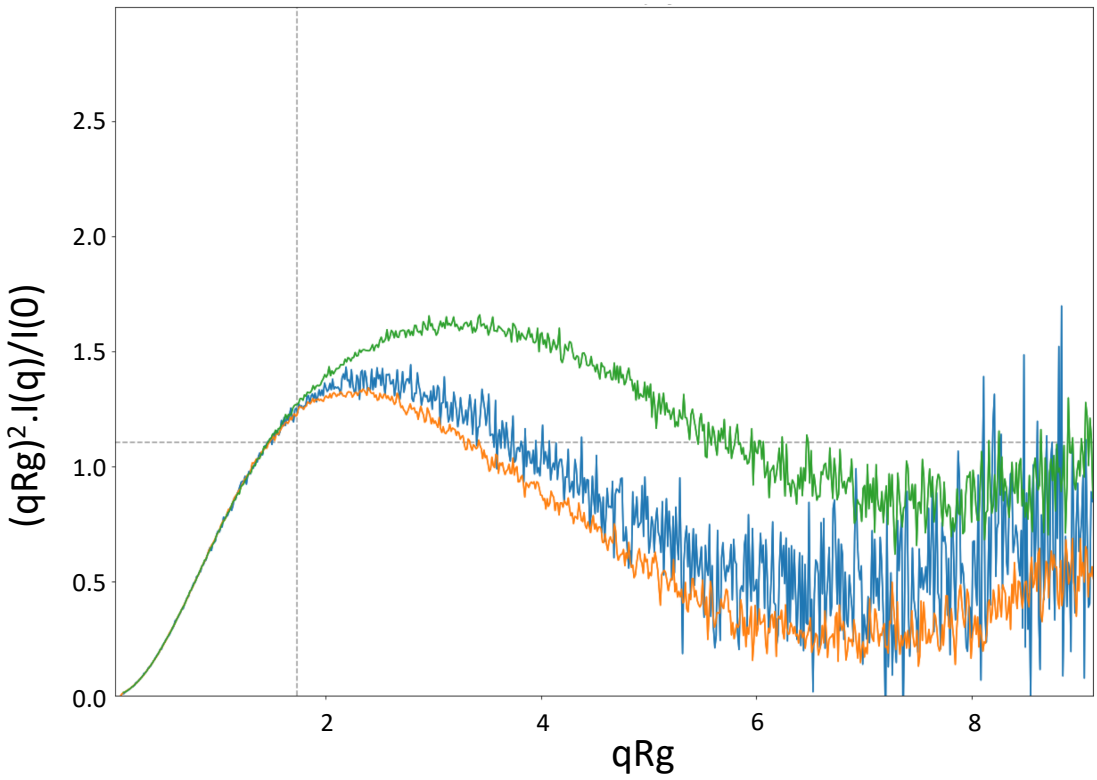
I – Annex 1: Phylogenetic analysis of Harmonin Homology Domains (Supplementary Information)

Harmonin HHD	Whirlin HHD1	Whirlin HHD2	PDZD7 HHD
V5G	A37D	R423P	L564P
V13A	A37V	Q429R	T570A
D14G	N38S	R431W	T578P
A21S	L42P	H439L	R579H
E22G	T47A	M442V	R583Q
D24E	Y65H	Y444C	Y584S
Y25H	R68C	Y445H	H586Q
Y25C	F72I	Y445C	L598M
L26F	F72L	R450C	P604Q
D28N	V75L	R450H	R614T
R31G	T77P	V456L	M628T
R31Q	T77I	A458T	V629M
M32T	V80A	A462V	
Q35H	R89H	T469P	
M37T	P92S	A471G	
V39M	M93T	P473S	
V41M	R95G	L475P	
L42F	R95L	S477P	
V43M	P99R	I483V	
D45A	R100C	R490L	
D45E	S101C	D492N	
L48V	D102Y	D492H	
V49L	L105P		
N51S	D107Q		
E52D	T110A		
S54R			
A55C			
L56M			
I62V			
R63Q			
I66S			
R69Q			
H70Y			
V72M			
D75E			
L77Q			
T78N			

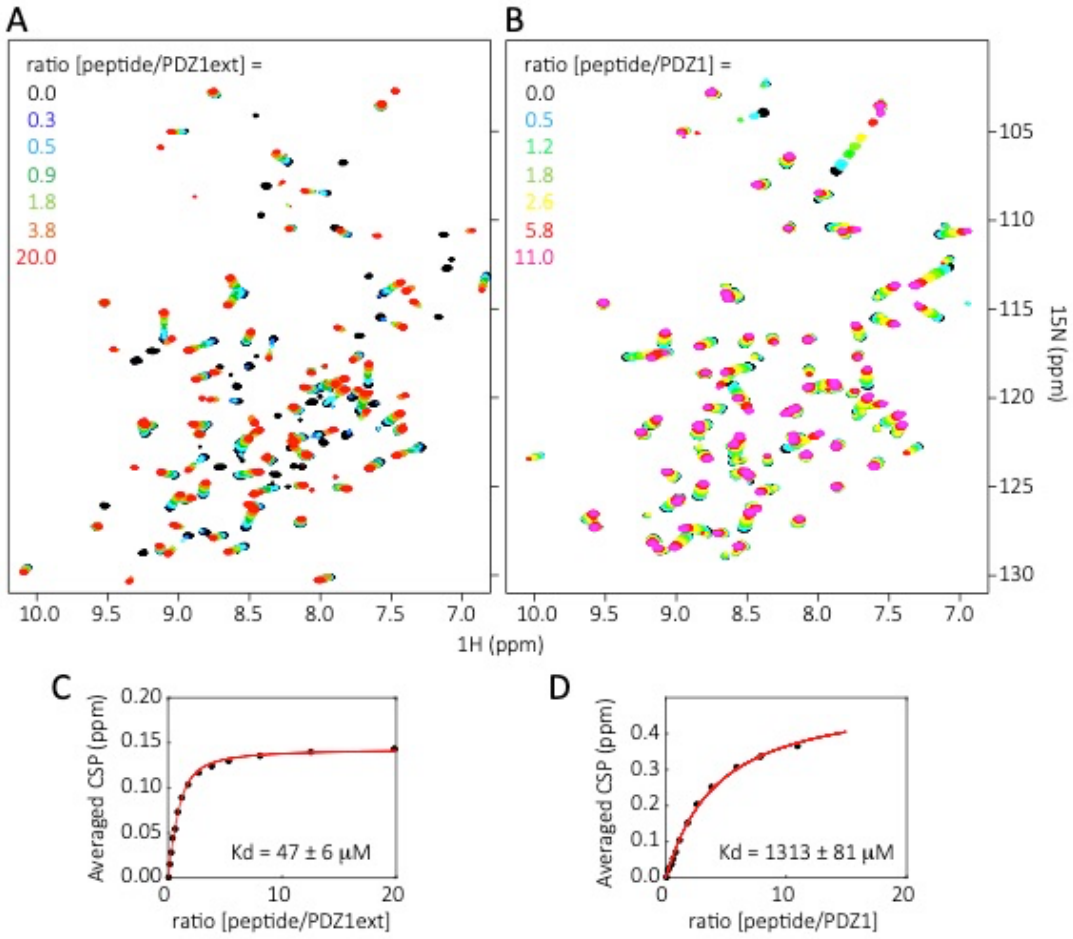
Table S1: All identified variants of unknown significance

II – Annex 2: Deciphering the molecular interaction between the adhesion G protein-coupled receptor ADGRV1 and its PDZ-containing regulator PDZD7 (Supplementary Information)

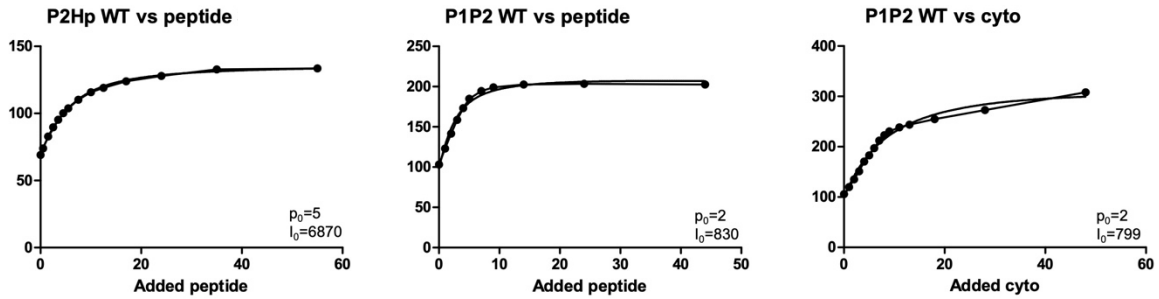
Supplementary 1: Alphafold model obtained for the PDZ1ext construct of PDZD7 (rank 1 displayed). Predicted lDDTs are mapped on the model, highlighting a lower confidence for the triple β strand extension than the PDZ core.



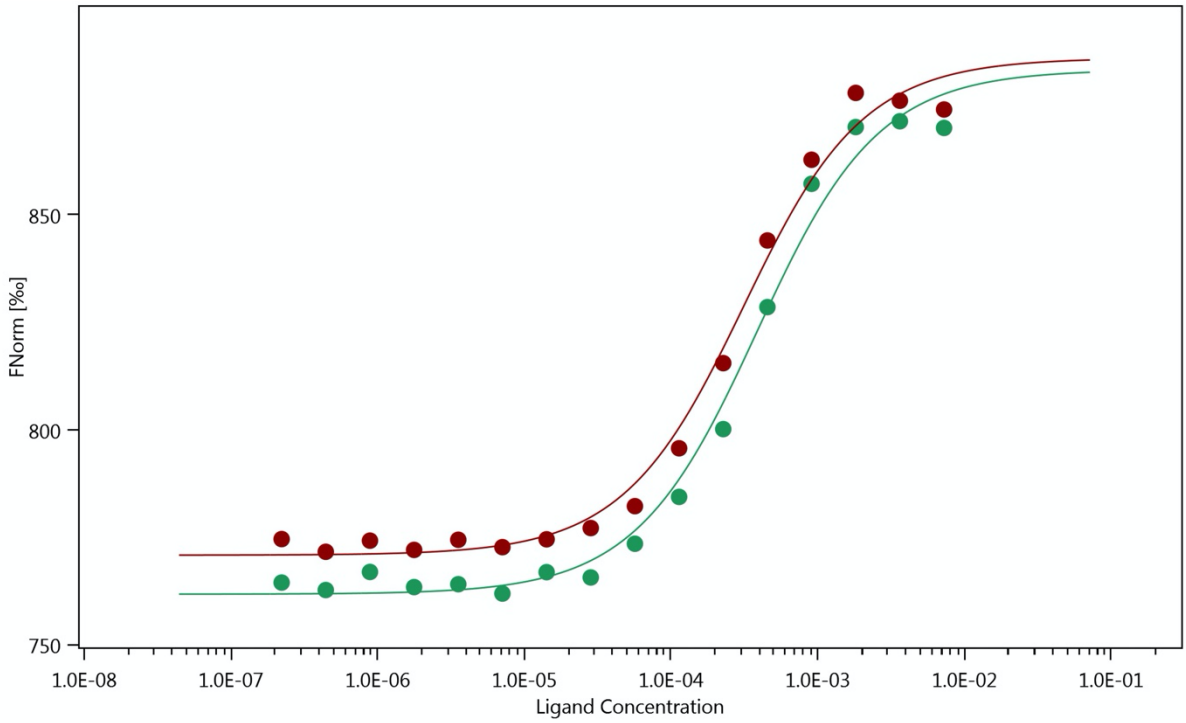
Supplementary 2: Dimensionless Kratky plots of the P1P2 tandem of whirlin (colored in orange), of the P1P2 tandem of PDZD7 (colored in blue) and of the N-PDZ1ext of PDZD7 (colored in green). The dashed lines indicate the peak position (~ 1.7) and peak height (~ 1.1) for a perfectly globular protein.



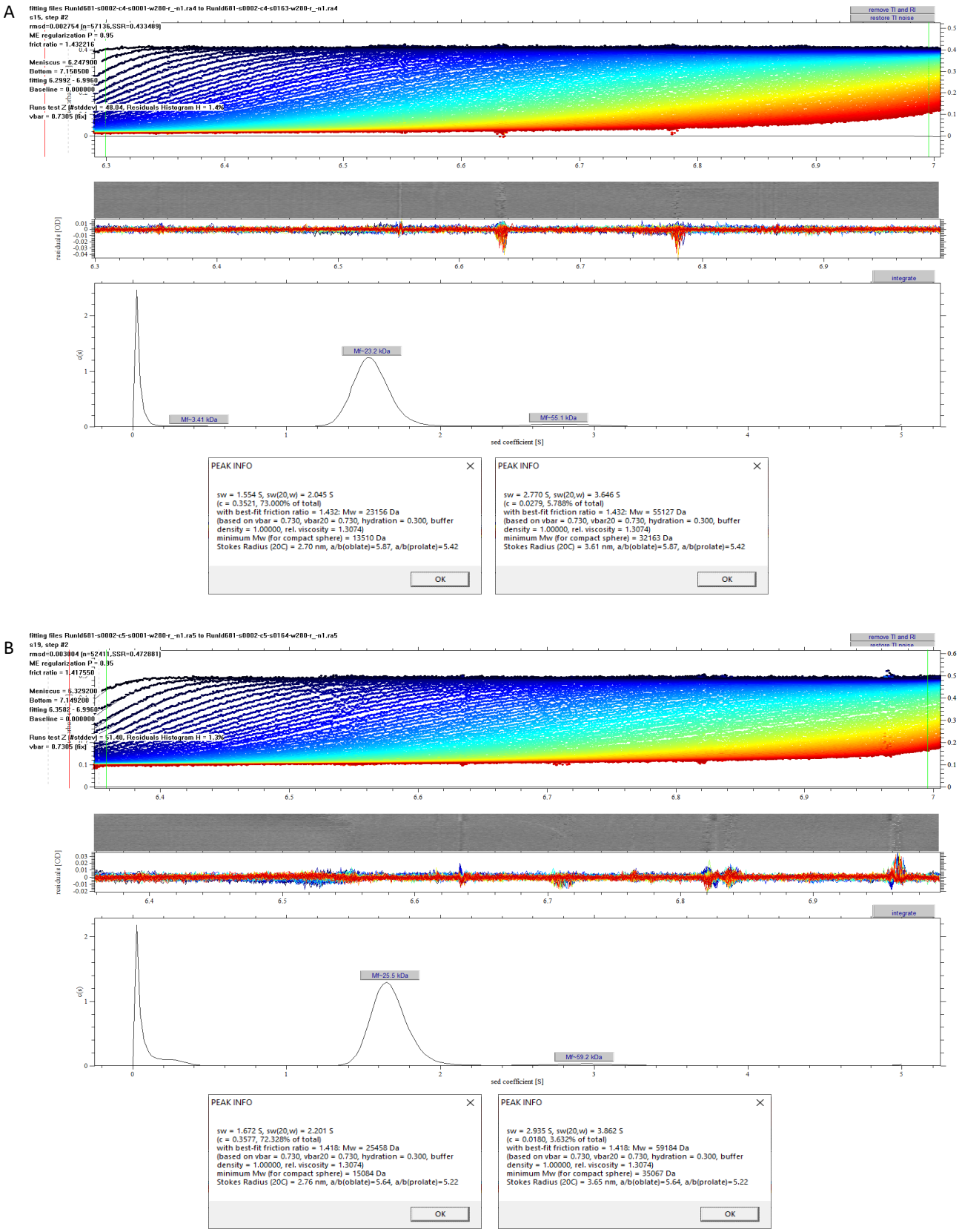
Supplementary 3: Determination of the dissociation constants (Kd) of ADGRV1pbm/PDZ1ext (C) and ADGRV1pbm/PDZ1 (D) interactions: the averaged CSP (over about 10 significantly shifting peaks) are plotted as a function of the molar ratio (peptide/PDZ) and fitted using nonlinear regression and assuming a simple complex formation model. The error on Kd values comes from the standard deviation of Kd calculated from individual fits on each of the significantly shifting peaks. Raw data are indicated by dots and fits by red curves.



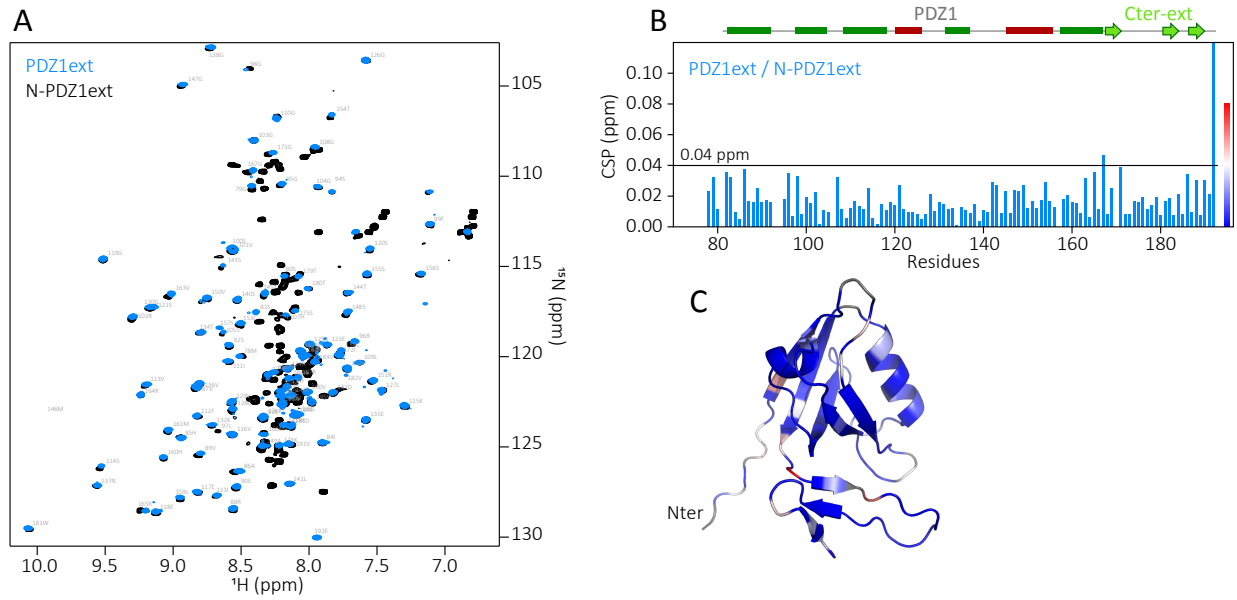
Supplementary 4: Fluorescence titration of P2HpWT, P1P2 WT with ADGRV1pbm peptide, and P1P2WT with ADGRV1 cytoplasmic domain. Raw data are indicated by connected points and fits by curves.



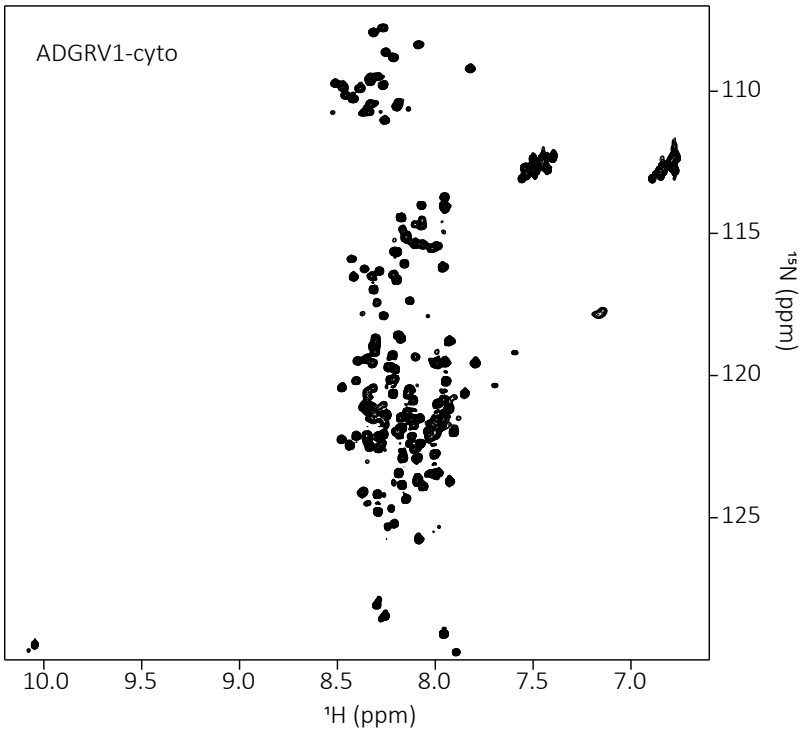
Supplementary 5: MicroScale Thermophoresis titrations (duplicate) of P2 WT with ADGRV1pbm peptide. Raw data are represented by dots and fits by curves.



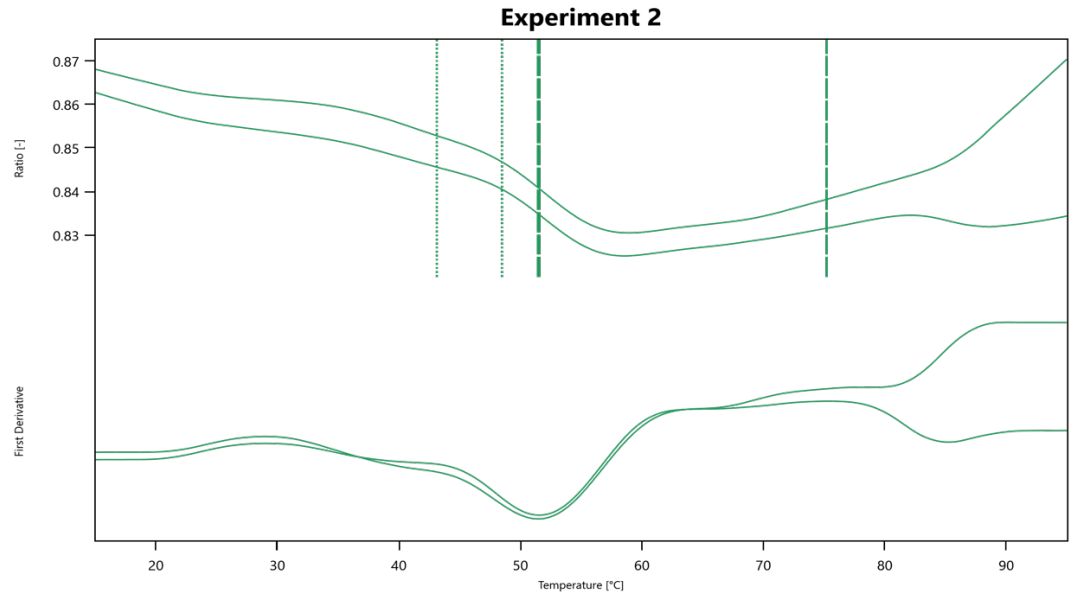
Supplementary 6: Analytical UltraCentrifugation data obtained for PDZD7 P1P2 without ligand **(A)** and in complex with the last 13 residues of ADGRV1 **(B)**.



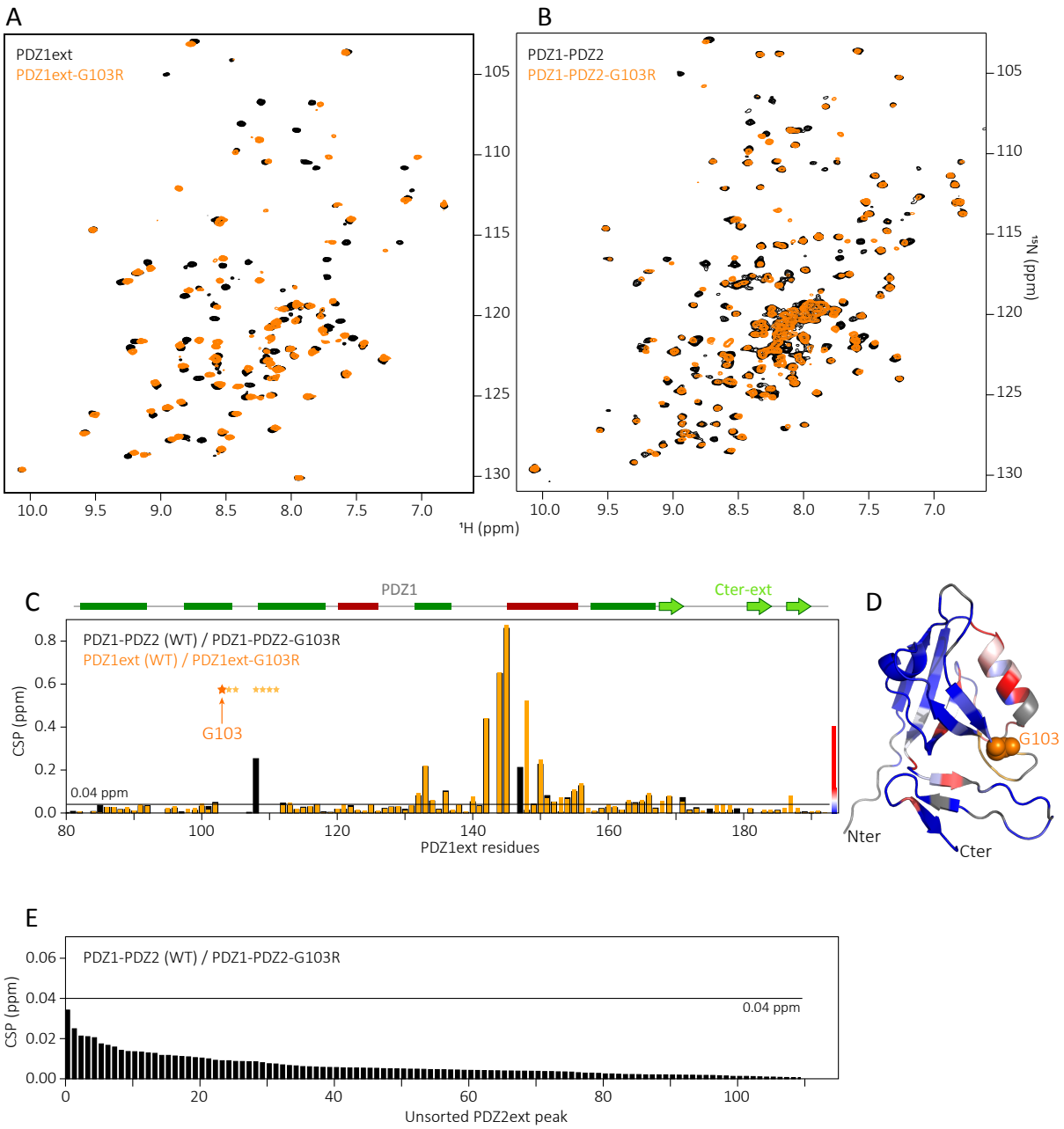
Supplementary 7: **(A)** Superimposition of PDZ1ext (black) and N-PDZ1ext (blue) ^{1}H-^{15}N HSQC spectra. **(B)** Chemical Shift Perturbations (CSP) measured for each residue of PDZ1ext when comparing PDZ1ext to N-PDZ1ext. Secondary structures are represented above the barplot, green for β strands of the PDZ core (from SSP), dark red for α helices of the PDZ core (from SSP) and light green arrows for β strands of the extension (from alphafold). **(C)** CSP are mapped on the alphafold model for PDZ1ext, from blue (no perturbation) to red (large perturbation). Residues depicted in grey have no associated chemical shift (missing peak in N-PDZ1ext).



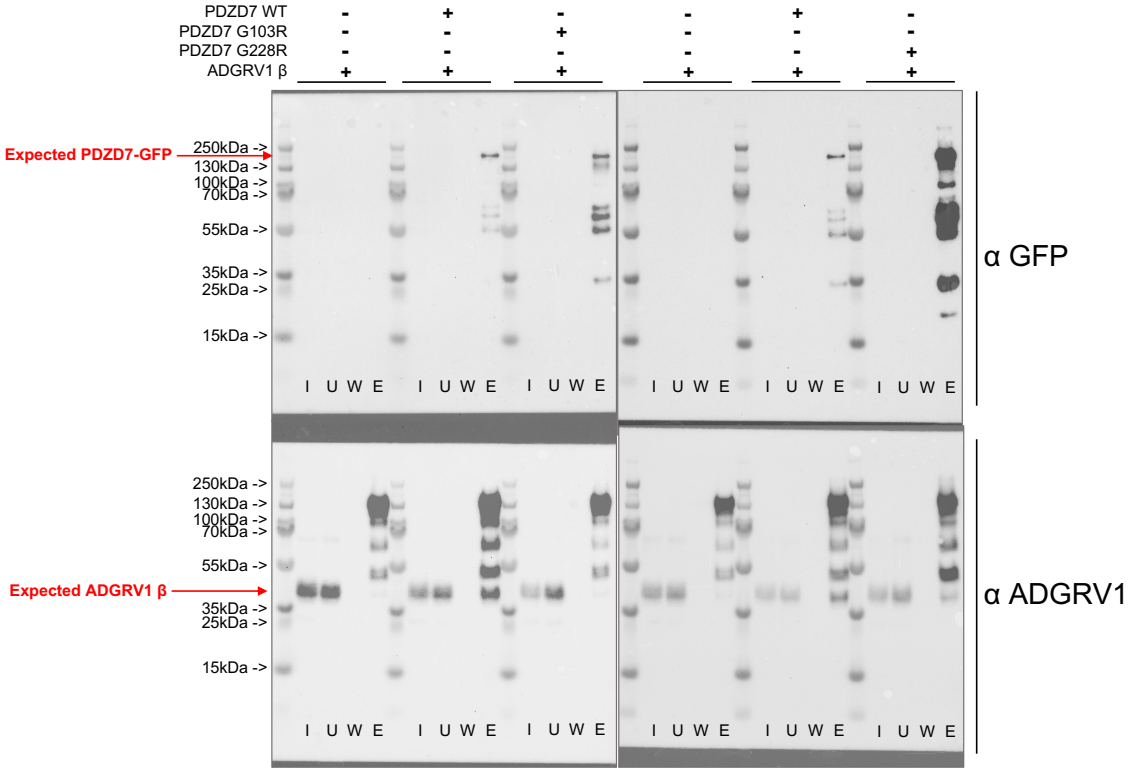
Supplementary 8: ^{1}H-^{15}N HSQC spectra of ADGRV1 whole cytoplasmic domain (152 residues).



Supplementary 9: nanoDSF denaturation curves of PDZD7 PDZ2ext G228R, unbound and after addition of a peptide corresponding to the last 13 residues of ADGRV1. The transition of fluorescence emission ratio (330/350nm) starts below the temperature range of the spectrometer (15°C).



Supplementary 10: **(A)** Superimposition of PDZ1ext WT (black) and PDZ1ext G103R (orange) ^{1}H-^{15}N HSQC spectra. **(B)** Superimposition of P1P2 WT (black) and P1P2 G103R (orange) ^{1}H-^{15}N HSQC spectra. **(C)** Chemical Shift Perturbations (CSP) measured for each residue of PDZ1ext when comparing P1P2 WT to P1P2 G103R (black) and PDZ1ext WT to PDZ1ext G103R (orange). Secondary structures are represented above the barplot, green for β strands of the PDZ core (from SSP), dark red for α helices of the PDZ core (from SSP) and light green arrows for β strands of the extension (from alphafold). **(D)** Unsorted Chemical Shift Perturbations (CSP) measured for each residue of PDZ2ext when comparing P1P2 WT to P1P2 G103R.



Supplementary 11: Full membranes corresponding to the figure 5 of the main text. Top membranes are read using an antibody directed against the GFP (fused to PDZD7). Bottom membranes are read using a custom antibody directed against the cytoplasmic domain of ADGRV1.

References

1. Fritzsch, B. & Beisel, K. W. Evolution and development of the vertebrate ear. *Brain Research bulletin* **55**, 711–721 (2001).
2. Carroll, S. B. Genetics and the making of Homo sapiens. *Nature* **422**, 849–857 (2003).
3. Shukla, A. *et al.* Hearing Loss, Loneliness, and Social Isolation: A Systematic Review. *Otolaryngology - Head and Neck Surgery (United States)* vol. 162 622–633 Preprint at https://doi.org/10.1177/0194599820910377 (2020).
4. Idstad, M., Tambs, K., Aarhus, L. & Engdahl, B. L. Childhood sensorineural hearing loss and adult mental health up to 43 years later: results from the HUNT study. *BMC Public Health* **19**, (2019).
5. Bigelow, R. T. *et al.* Association of Hearing Loss with Psychological Distress and Utilization of Mental Health Services among Adults in the United States. *JAMA Network Open* **3**, (2020).
6. Cosh, S., Helmer, C., Delcourt, C., Robins, T. G. & Tully, P. J. Depression in elderly patients with hearing loss: Current perspectives. *Clinical Interventions in Aging* vol. 14 1471–1480 Preprint at https://doi.org/10.2147/CIA.S195824 (2019).
7. Amieva, H., Ouvrard, C., Meillon, C., Rullier, L. & Dartigues, J. F. Death, Depression, Disability, and Dementia Associated with Self-reported Hearing Problems: A 25-Year Study. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* **73**, 1383–1389 (2018).
8. Livingston, G. *et al.* Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet* vol. 396 413–446 Preprint at https://doi.org/10.1016/S0140-6736(20)30367-6 (2020).
9. World Health Organization (WHO). *WORLD REPORT ON HEARING*. https://www.who.int/publications/i/item/world-report-on-hearing (2021).
10. Bialek, W. & Wit, H. P. Quantum limits to oscillator stability: theory and experiments on acoustic emissions from the human ear. *Physics Letters* **104A**, 173–178 (1984).
11. The Middle Ear. in *Neuroscience* (eds. Purves, D. et al.) (Sinauer Associates, 2001).
12. Koch, R. W., Ladak, H. M., Elfarnawany, M. & Agrawal, S. K. Measuring Cochlear Duct Length - A historical analysis of methods and results. *Journal of Otolaryngology - Head and Neck Surgery* **46**, (2017).
13. Leibovici, M., Safieddine, S. & Petit, C. Chapter 8 Mouse Models for Human Hereditary Deafness. in *Current Topics in Developmental Biology* vol. 84 385–429 (2008).
14. Helpard, L., Li, H., Rask-Andersen, H., Ladak, H. M. & Agrawal, S. K. Characterization of the human helicotrema: Implications for cochlear duct length and frequency mapping. *Journal of Otolaryngology - Head and Neck Surgery* **49**, (2020).
15. Raphael, Y. & Altschuler, R. A. Structure and innervation of the cochlea. *Brain Research Bulletin* **60**, 397–422 (2003).
16. Delpire, E. & Mount, D. B. Human and murine phenotypes associated with defects in cation-chloride cotransport. *Annual Review of Physiology* **64**, 803–843 (2002).
17. Wangemann, P. K + Cycling and Its Regulation in the Cochlea and the Vestibular Labyrinth. *Audiology Neuro Otology* **7**, 199–205 (2002).
18. Parsa, A., Webster, P. & Kalinec, F. Deiters cells tread a narrow path-The Deiters cells-basilar membrane junction. *Hearing Research* **290**, 13–20 (2012).
19. Merchan, M. A., Merchan, J. A. & Ludena, M. D. Morphology of Hensen's cells. *Journal of Anatomy* **131**, 519–523 (1980).

20. Anniko ', M. & Wrbblewski, R. Ionic environment of cochlear hair cells. *Hearing Research* **22**, 279–293 (1986).
21. Teudt, I. U. & Richter, C. P. Basilar membrane and tectorial membrane stiffness in the CBA/CaJ mouse. *Journal of the Association for Research in Otolaryngology* **15**, 675–694 (2014).
22. Emadi, G., Richter, C. P. & Dallos, P. Stiffness of the Gerbil Basilar Membrane: Radial and Longitudinal Variations. *Journal of Neurophysiology* **91**, 474–488 (2004).
23. Gavara, N., Manoussaki, D. & Chadwick, R. S. Auditory Mechanics of the Tectorial Membrane and the Cochlear Spiral. *Current Opinion in Otolaryngology & Head and Neck Surgery* **19**, 382–387 (2011).
24. Helmholtz, H. *On the sensations of tone as a physiological basis for the theory of music*. (1877).
25. Manoussaki, D. *et al. The influence of cochlear shape on low-frequency hearing*. *PNAS* vol. 105 www.pnas.orgcgidoi10.1073pnas.0710037105 (2008).
26. LeMasurier, M. & Gillespie, P. G. Hair-cell mechanotransduction and cochlear amplification. *Neuron* **48**, 403–415 (2005).
27. Avan, P. *et al.* Otogelin, otogelin-like, and stereocilin form links connecting outer hair cell stereocilia to each other and the tectorial membrane. *PNAS* **116**, 25948–25957 (2019).
28. Verpy, E. *et al.* Stereocilin connects outer hair cell stereocilia to one another and to the tectorial membrane. *Journal of Comparative Neurology* **519**, 194–210 (2011).
29. Richardson, G. P. & Petit, C. Hair-bundle links: Genetics as the gateway to function. *Cold Spring Harbor Perspectives in Medicine* **9**, (2019).
30. Hakizimana, P. & Fridberger, A. Inner hair cell stereocilia are embedded in the tectorial membrane. *Nature Communications* **12**, (2021).
31. Rueda, J., Cantos, R. & Lira, D. J. Tectorial membrane-organ of Corti relationship during cochlear development. *Anat Embryol* **194**, 501–514 (1996).
32. Pacentine, I., Chatterjee, P. & Barr-Gillespie, P. G. Stereocilia rootlets: Actin-based structures that are essential for structural stability of the hair bundle. *International Journal of Molecular Sciences* **21**, (2020).
33. Liu, C., Luo, N., Tung, C. Y., Perrin, B. J. & Zhao, B. GRXCR2 Regulates Taperin Localization Critical for Stereocilia Morphology and Hearing. *Cell Reports* **25**, 1268-1280.e4 (2018).
34. Zhu, Y. *et al.* Deciphering the Unexpected Binding Capacity of the Third PDZ Domain of Whirlin to Various Cochlear Hair Cell Partners. *Journal of Molecular Biology* **432**, 5920–5937 (2020).
35. Du, T. T. *et al.* LMO7 deficiency reveals the significance of the cuticular plate for hearing function. *Nature Communications* **10**, (2019).
36. Pollock, L. M., Gupta, N., Chen, X., Luna, E. J. & McDermott, B. M. Supervillin is a component of the hair cell's cuticular plate and the head plates of organ of Corti supporting cells. *PLoS ONE* **11**, (2016).
37. Ashmore, J. Cochlear outer hair cell motility. *Physiological Reviews* **88**, 173–210 (2008).
38. He, D. Z. Z., Zheng, J., Kalinec, F., Kakehata, S. & Santos-Sacchi, J. Tuning in to the amazing outer hair cell: Membrane wizardry with a twist and shout. *Journal of Membrane Biology* **209**, 119–134 (2006).
39. Appler, J. M. & Goodrich, L. v. Connecting the ear to the brain: Molecular mechanisms of auditory circuit assembly. *Progress in Neurobiology* **93**, 488–508 (2011).
40. Santos-Sacchi, J. & Dilger, J. P. Whole cell currents and mechanical responses of isolated outer hair cells. *Hearing Research* **35**, 143–150 (1988).

41. Kalinec, F., Holleyt, M. C., Iwasa, K. H., Lim, D. J. & Kachar, B. A membrane-based force generation mechanism in auditory sensory cells. *PNAS* **89**, 8671–8675 (1992).
42. Okoruwa, O. E. *et al.* Evolutionary insights into the unique electromotility motor of mammalian outer hair cells. *Evolution and Development* **10**, 300–315 (2008).
43. Zheng, J. *et al.* Analysis of the oligomeric structure of the motor protein prestin. *Journal of Biological Chemistry* **281**, 19916–19924 (2006).
44. Schaechinger, T. J. & Oliver, D. Nonmammalian orthologs of prestin (SLC26A5) are electrogenic divalent/chloride anion exchangers. *PNAS* **104**, 7693–7698 (2007).
45. Iwasa, K. H. A Two-State Piezoelectric Model for Outer Hair Cell Motility. *Biophysical Journal* **81**, 2495–2506 (2001).
46. Schaechinger, T. J. *et al.* A synthetic prestin reveals protein domains and molecular operation of outer hair cell piezoelectricity. *EMBO Journal* **30**, 2793–2804 (2011).
47. Jones, E. M. C., Gray-Keller, M. & Fettiplace, R. The role of Ca2+-activated K+ channel spliced variants in the tonotopic organization of the turtle cochlea. *Journal of Physiology* **518**, 653–665 (1999).
48. Ramanathan, K., Michael, T. H. & Fuchs, P. A. Beta Subunits Modulate Alternatively Spliced, Large Conductance, Calcium-Activated Potassium Channels of Avian Hair Cells. *The Journal of Neuroscience* **20**, 1675–1684 (2000).
49. Ramanathan, K. & Fuchs, P. A. Modeling hair cell tuning by expression gradients of potassium channel β subunits. *Biophysical Journal* **82**, 64–75 (2002).
50. Ramanathan, K., Michael, T. H., Jiang, G.-J., Hiel, H. & Fuchs, P. A. A Molecular Mechanism for Electrical Tuning of Cochlear Hair Cells. *Science (1979)* **283**, (1999).
51. Kachar, B., Parakkal, M., Kurc, M., Zhao, Y.-D. & Gillespie, P. G. High-resolution structure of hair-cell tip links. *PNAS* **97**, 13336–13341 (2000).
52. Choudhary, D. *et al.* Structural determinants of protocadherin-15 mechanics and function in hearing and balance perception. *PNAS* **117**, 24837–24848 (2020).
53. Lanie Tobin, M., Chaiyasitdhi, A., Michel, V., Michalski, N. & Martin, P. Stiffness and tension gradients of the hair cell's tip-link complex in the mammalian cochlea. *Elife* (2019) doi:10.7554/eLife.43473.001.
54. Ahmed, Z. M. *et al.* The tip-link antigen, a protein associated with the transduction complex of sensory hair cells, is protocadherin-15. *Journal of Neuroscience* **26**, 7022–7034 (2006).
55. Michel, V. *et al.* Interaction of protocadherin-15 with the scaffold protein whirlin supports its anchoring of hair-bundle lateral links in cochlear hair cells. *Scientific Reports* **10**, (2020).
56. Maeda, R. *et al.* Tip-link protein protocadherin 15 interacts with transmembrane channel-like proteins TMC1 and TMC2. *Proc Natl Acad Sci U S A* **111**, 12907–12912 (2014).
57. Pan, B. *et al.* TMC1 Forms the Pore of Mechanosensory Transduction Channels in Vertebrate Inner Ear Hair Cells. *Neuron* **99**, 736-753.e6 (2018).
58. Erickson, T. *et al.* Integration of Tmc1/2 into the mechanotransduction complex in zebrafish hair cells is regulated by Transmembrane O-methyltransferase (Tomt). *Elife* (2017) doi:10.7554/eLife.28474.001.
59. Pacentine, I. v. & Nicolson, T. Subunits of the mechano-electrical transduction channel, Tmc1/2b, require Tmie to localize in zebrafish sensory hair cells. *PLoS Genetics* **15**, (2019).
60. Cunningham, C. L. *et al.* TMIE Defines Pore and Gating Properties of the Mechanotransduction Channel of Mammalian Cochlear Hair Cells. *Neuron* **107**, 126-143.e8 (2020).
61. Giese, A. P. J. *et al.* CIB2 interacts with TMC1 and TMC2 and is essential for mechanotransduction in auditory hair cells. *Nature Communications* **8**, (2017).

62. Michel, V. *et al.* CIB2, defective in isolated deafness, is key for auditory hair cell mechanotransduction and survival. *EMBO Molecular Medicine* **9**, 1711–1731 (2017).
63. Liang, X. *et al.* CIB2 and CIB3 are auxiliary subunits of the mechanotransduction channel of hair cells. *Neuron* **109**, 2131-2149.e15 (2021).
64. Wang, Y. *et al.* Loss of CIB2 causes profound hearing loss and abolishes mechanoelectrical transduction in mice. *Frontiers in Molecular Neuroscience* **10**, (2017).
65. Wan Yoon, K. *et al.* CIB1 functions as a Ca 2-sensitive modulator of stress-induced signaling by targeting ASK1. *PNAS* **106**, 17389–17394 (2009).
66. Huang, H., Bogstie, J. N. & Vogel, H. J. Biophysical and structural studies of the human calcium- and integrin-binding protein family: understanding their functional similarities and differences. *Biochemistry and Cell Biology* **90**, 646–656 (2012).
67. Beurg, M., Nam, J.-H., Chen, Q. & Fettiplace, R. Calcium Balance and Mechanotransduction in Rat Cochlear Hair Cells. *Journal of Neurophysiology* **104**, 18–34 (2010).
68. Corns, L. F. *et al.* Mechanotransduction is required for establishing and maintaining mature inner hair cells and regulating efferent innervation. *Nature Communications* **9**, (2018).
69. White, P. M., Doetzlhofer, A., Lee, Y. S., Groves, A. K. & Segil, N. Mammalian cochlear supporting cells can divide and trans-differentiate into hair cells. *Nature* **441**, 984–987 (2006).
70. Zheng, J. L. & Gao, W.-Q. Overexpression of Math1 induces robust production of extra hair cells in postnatal rat inner ears. *Nature* **3**, 580–586 (2000).
71. Izumikawa, M. *et al.* Auditory hair cell replacement and hearing improvement by Atoh1 gene therapy in deaf mammals. *Nature Medicine* **11**, 271–276 (2005).
72. Woods, C., Montcouquiol, M. & Kelley, M. W. Math1 regulates development of the sensory epithelium in the mammalian cochlea. *Nature Neuroscience* **7**, 1310–1318 (2004).
73. Hirono, M., Denis, C. S., Richardson, G. P. & Gillespie, P. G. Hair Cells Require Phosphatidylinositol 4,5-Bisphosphate for Mechanical Transduction and Adaptation. *Neuron* **44**, 309–320 (2004).
74. Zhao, B. *et al.* TMIE is an essential component of the mechanotransduction machinery of cochlear hair cells. *Neuron* **84**, 954–967 (2014).
75. Yu, X. *et al.* Deafness mutation D572N of TMC1 destabilizes TMC1 expression by disrupting LHFPL5 binding. *PNAS* **117**, 29894–29903 (2020).
76. Beurg, M. *et al.* Variable number of TMC1-dependent mechanotransducer channels underlie tonotopic conductance gradients in the cochlea. *Nature Communications* **9**, 2185 (2018).
77. Fettiplace, R. & Kim, K. X. The Physiology of Mechanoelectrical Transduction Channels in Hearing. *Physiological Reviews* **94**, 951–986 (2014).
78. Li, S. *et al.* Myosin-VIIa is expressed in multiple isoforms and essential for tensioning the hair cell mechanotransduction complex. *Nature Communications* **11**, 1–15 (2020).
79. Siemens, J. *et al.* The Usher syndrome proteins cadherin 23 and harmonin form a complex by means of PDZ-domain interactions. *PNAS* **99**, 14946–14951 (2002).
80. Boëda, B. *et al.* Myosin VIIa, harmonin and cadherin 23, three Usher I gene products that cooperate to shape the sensory hair cell bundle. *The EMBO Journal* **21**, 6689–6699 (2002).
81. Pan, L., Yan, J., Wu, L. & Zhang, M. Assembling stable hair cell tip link complex via multidentate interactions between harmonin and cadherin 23. *PNAS* **106**, 5575–5580 (2009).

82. Wu, L., Pan, L., Zhang, C. & Zhang, M. Large protein assemblies formed by multivalent interactions between cadherin23 and harmonin suggest a stable anchorage structure at the tip link of stereocilia. *Journal of Biological Chemistry* **287**, 33460–33471 (2012).
83. Bahloul, A. *et al.* Cadherin-23, myosin viia and harmonin, encoded by usher syndrome type I genes, form a ternary complex and interact with membrane phospholipids. *Human Molecular Genetics* **19**, 3557–3565 (2010).
84. Verpy, E. *et al.* A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C. *Nature Genetics* **26**, 51–55 (2000).
85. Adato, A. *et al.* Interactions in the network of usher syndrome type 1 proteins. *Human Molecular Genetics* **14**, 347–356 (2005).
86. Yu, I. M. *et al.* Myosin 7 and its adaptors link cadherins to actin. *Nature Communications* **8**, (2017).
87. Yan, J., Pan, L., Chen, X., Wu, L. & Zhang, M. The structure of the harmonin/sans complex reveals an unexpected interaction mode of the two Usher syndrome proteins. *Proc Natl Acad Sci U S A* **107**, 4040–4045 (2010).
88. He, Y., Li, J. & Zhang, M. Myosin VII, USH1C, and ANKS4B or USH1G Together Form Condensed Molecular Assembly via Liquid-Liquid Phase Separation. *Cell Reports* **29**, 974-986.e4 (2019).
89. Holt, J. R. & Corey, D. P. Two mechanisms for transducer adaptation in vertebrate hair cells. *Proc Natl Acad Sci U S A* **97**, 11730–11735 (2000).
90. Eatock, R. A. Adaptation in Hair Cells. *Annual Review in Neuroscience* **23**, 285–314 (2000).
91. Schwander, M., Kachar, B. & Müller, U. The cell biology of hearing. *J Cell Biochem* **190**, 9–20 (2010).
92. Ricci, A. J. & Fettiplace, R. The effects of calcium buffering and cyclic AMP on mechano-electrical transduction in turtle auditory hair cells. *Journal of Physiology* **501**, 24 (1997).
93. Cheung, E. L. M. & Corey, D. P. Ca2+ Changes the Force Sensitivity of the Hair-Cell Transduction Channel. *Biophysical Journal* **90**, 124–139 (2006).
94. Holt, J. R. *et al.* A Chemical-Genetic Strategy Implicates Myosin-1c in Adaptation by Hair Cells. *Cell* **108**, 371–381 (2002).
95. Phillips, K. R., Tong, S., Goodyear, R., Richardson, G. P. & Cyr, J. L. Stereociliary myosin-1c receptors are sensitive to calcium chelation and absent from cadherin 23 mutant mice. *Journal of Neuroscience* **26**, 10777–10788 (2006).
96. Lu, Q., Li, J., Ye, F. & Zhang, M. Structure of myosin-1c tail bound to calmodulin provides insights into calcium-mediated conformational coupling. *Nature Structural and Molecular Biology* **22**, 81–87 (2015).
97. Giese, S. *et al.* Mechanochemical properties of human myosin-1C are modulated by isoform-specific differences in the N-terminal extension. *Journal of Biological Chemistry* **296**, (2021).
98. Bosher, S. K. & Warren, R. L. Observations on the electrochemistry of the cochlear endolymph of the rat: a quantitative study of its electrical potential and ionic composition as determined by means of flame spectrophotometry. *Proceedings of the Royal Society B* **171**, 227–247 (1968).
99. Ghanem, T. A., Breneman, K. D., Rabbitt, R. D. & Mack Brown, H. Ionic Composition of Endolymph and Perilymph in the Inner Ear of the Oyster Toadfish, Opsanus tau. *The Biological Bulletin* **214**, 83–90 (2008).

100. Ruan, Q. *et al.* Topological and developmental expression gradients of Kir2.1, an inward rectifier K+ channel, in spiral ganglion and cochlear hair cells of mouse inner ear. *Developmental Neuroscience* **30**, 374–388 (2009).
101. Kubisch, C. *et al.* KCNQ4, a Novel Potassium Channel Expressed in Sensory Outer Hair Cells, Is Mutated in Dominant Deafness. *Cell* **96**, 437–446 (1999).
102. Kollmar, R., Montgomery, L. G., Fak, J., Henry, L. J. & Hudspeth, A. J. Predominance of the 1D subunit in L-type voltage-gated Ca 2+ channels of hair cells in the chicken's cochlea. *Neurobiology* **94**, 14883–14888 (1997).
103. Issa, N. P. & Hudspeth, A. J. Clustering of Ca2+ channels and Ca2+-activated K+ channels at fluorescently labeled presynaptic active zones of hair cells. *PNAS* **91**, 7578–7582 (1994).
104. Lee, S., Briklin, O., Hiel, H. & Fuchs, P. Calcium-dependent inactivation of calcium channels in cochlear hair cells of the chicken. *Journal of Physiology* **583**, 909–922 (2007).
105. Dumont, R. A. *et al.* Plasma Membrane Ca2+ -ATPase Isoform 2a Is the PMCA of Hair Bundles. *The Journal of Neuroscience* **21**, 5066–5078 (2001).
106. Chen, Q. *et al.* The development, distribution and density of the plasma membrane calcium ATPase 2 calcium pump in rat cochlear hair cells. *European Journal of Neuroscience* **36**, 2302–2310 (2012).
107. Schnee, M. E., Santos-Sacchi, J., Castellano-Muñoz, M., Kong, J. H. & Ricci, A. J. Calcium-Dependent Synaptic Vesicle Trafficking Underlies Indefatigable Release at the Hair Cell Afferent Fiber Synapse. *Neuron* **70**, 326–338 (2011).
108. Magupalli, V. G. *et al.* Multiple RIBEYE-RIBEYE interactions create a dynamic scaffold for the formation of synaptic ribbons. *Journal of Neuroscience* **28**, 7954–7967 (2008).
109. Regus-Leidig, H. *et al.* Identification and Immunocytochemical Characterization of Piccolino, a Novel Piccolo Splice Variant Selectively Expressed at Sensory Ribbon Synapses of the Eye and Ear. *PLoS ONE* **8**, (2013).
110. Khimich, D. *et al.* Hair cell synaptic ribbons are essential for synchronous auditory signalling. *Nature* **434**, 889–894 (2005).
111. Seal, R. P. *et al.* Sensorineural Deafness and Seizures in Mice Lacking Vesicular Glutamate Transporter 3. *Neuron* **57**, 263–275 (2008).
112. Heidrych, P. *et al.* Otoferlin interacts with myosin VI: Implications for maintenance of the basolateral synaptic structure of the inner hair cell. *Human Molecular Genetics* **18**, 2779–2790 (2009).
113. Roux, I. *et al.* Myosin VI is required for the proper maturation and function of inner hair cell ribbon synapses. *Human Molecular Genetics* **18**, 4615–4628 (2009).
114. Dulon, D., Safieddine, S., Jones, S. M. & Petit, C. Otoferlin is critical for a highly sensitive and linear calcium-dependent exocytosis at vestibular hair cell ribbon synapses. *Journal of Neuroscience* **29**, 10474–10487 (2009).
115. Goutman, J. D. & Glowatzki, E. Time course and calcium dependence of transmitter release at a single ribbon synapse. *PNAS* **104**, 16341–16346 (2007).
116. Magistretti, J., Spaiardi, P., Johnson, S. L. & Masetto, S. Elementary properties of Ca2+ channels and their influence on multivesicular release and phase-locking at auditory hair cell ribbon synapses. *Frontiers in Cellular Neuroscience* **9**, (2015).
117. Tang, J. *et al.* A Complexin/Synaptotagmin 1 Switch Controls Fast Synaptic Vesicle Exocytosis. *Cell* **126**, 1175–1187 (2006).
118. Nouvian, R. *et al.* Exocytosis at the hair cell ribbon synapse apparently operates without neuronal SNARE proteins. *Nature Neuroscience* **14**, 411–413 (2011).
119. Johnson, S. L. *et al.* Synaptotagmin IV determines the linear Ca 2+ dependence of vesicle fusion at auditory ribbon synapses. *Nature Neuroscience* **13**, 45–52 (2010).

120. Michalski, N. *et al.* Otoferlin acts as a Ca2+ sensor for vesicle fusion and vesicle pool replenishment at auditory hair cell ribbon synapses. *Elife* **6**, (2017).
121. Glowatzki, E. & Fuchs, P. A. Transmitter release at the hair cell ribbon synapse. *Nature Neuroscience* **5**, 147–154 (2002).
122. Ruel, J., Chen, C., Pujol, R., Bobbin, R. P. & Puel J L. AMPA-preferring glutamate receptors in cochlear physiology of adult guinea-pig. *Journal of Physiology* **518**, 667–680 (1999).
123. Ohna, T. L. *et al.* Hair cells use active zones with different voltage dependence of Ca2+ influx to decompose sounds into complementary neural codes. *Proc Natl Acad Sci U S A* **113**, E4716–E4725 (2016).
124. Rutherford, M. A., von Gersdorff, H. & Goutman, J. D. Encoding sound in the cochlea: from receptor potential to afferent discharge. *Journal of Physiology* **599**, 2527–2557 (2021).
125. Yoshinaga-Itano, C., Sedey, A. L., Coulter, D. K. & Mehl, A. L. Language of Early- and Later-identified Children With Hearing Loss. *Pediatrics* **102**, 1161–1171 (1998).
126. Morton, C. C., Nance, W. E. & St, E. M. Newborn Hearing Screening-A Silent Revolution. *The New England Journal of Medicine* 2151–2164 (2006).
127. Korver, A. M. H. *et al.* Congenital hearing loss. *Nature Reviews Disease Primers* **3**, 16094 (2017).
128. Smith, R. J. H., Bale, J. F. & White, K. R. Sensorineural hearing loss in children. in *The Lancet* vol. 365 879–890 (Elsevier B.V., 2005).
129. van Rompaey, V. *et al.* Pathogenic Variants in the COCH Gene That Lead to Hearing Loss and Vestibular Deficit: Update on DFNA9 and Introducing DFNB110. *B-ENT* (2022) doi:10.5152/B-ENT.2022.21791.
130. Denoyelle, F. *et al.* Clinical features of the prevalent form of childhood deafness, DFNB1, due to a connexin-26 gene defect: implications for genetic counselling. *The Lancet* **353**, 1298–1303 (1999).
131. Jagger, D. J. & Forge, A. Connexins and gap junctions in the inner ear – it's not just about K+ recycling. *Cell and Tissue Research* **360**, 633–644 (2015).
132. Kimberlin, D. W. *et al.* Effect of ganciclovir therapy on hearing in symptomatic congenital cytomegalovirus disease involving the central nervous system: A randomized, controlled trial. *Journal of Pediatrics* **143**, 16–25 (2003).
133. Swanson, E. C. & Schleiss, M. R. Congenital Cytomegalovirus Infection. New Prospects for Prevention and Therapy. *Pediatric Clinics of North America* **60**, 335–349 (2013).
134. Ville, Y. & Leruez-Ville, M. Managing infections in pregnancy. *Current Opinion in Infectious Diseases* **27**, 251–257 (2014).
135. Xu, J. & Han, W. Improvement of adult BTE hearing aid wearers' front/back localization performance using digital pinna-cue preserving technologies: An evidence-based review. *Korean Journal of Audiology* **18**, 97–104 (2014).
136. Tjellstrom, A., Hikansson, B. & Granstrom, G. Bone-anchored hearing aids. *Otolaryngologic clinics of north america* **34**, 337–364 (2001).
137. Golub, J. S., Won, J. H., Drennan, W. R., Worman, T. D. & Rubinstein, J. T. Spectral and temporal measures in hybrid cochlear implant users: On the mechanism of electroacoustic hearing benefits. *Otology and Neurotology* **33**, 147–153 (2012).
138. Mianné, J. *et al.* Correction of the auditory phenotype in C57BL/6N mice via CRISPR/Cas9-mediated homology directed repair. *Genome Medicine* **8**, (2016).
139. Akil, O. *et al.* Restoration of Hearing in the VGLUT3 Knockout Mouse Using Virally Mediated Gene Therapy. *Neuron* **75**, 283–293 (2012).
140. Iizuka, T. *et al.* Perinatal GJB2 gene transfer rescues hearing in a mouse model of hereditary deafness. *Human Molecular Genetics* **24**, 3651–3661 (2015).

141. Askew, C. *et al.* Tmc gene therapy restores auditory function in deaf mice. *Science Translational Medicine* **7**, (2015).
142. Chien, W. W. *et al.* Gene therapy restores hair cell stereocilia morphology in inner ears of deaf whirler mice. *Molecular Therapy* **24**, 17–25 (2016).
143. Geng, R. *et al.* Modeling and preventing progressive hearing loss in usher syndrome III. *Scientific Reports* **7**, (2017).
144. Yu, Q. *et al.* Virally expressed connexin26 restores gap junction function in the cochlea of conditional Gjb2 knockout mice. *Gene Therapy* **21**, 71–80 (2014).
145. Crane, R., Conley, S. M., Al-Ubaidi, M. R. & Naash, M. I. Gene Therapy to the Retina and the Cochlea. *Frontiers in Neuroscience* **15**, (2021).
146. Lee, M. Y. & Park, Y. H. Potential of Gene and Cell Therapy for Inner Ear Hair Cells. *BioMed Research International* **2018**, (2018).
147. Ryals, B. M. & Rubel, E. W. Hair Cell Regeneration After Acoustic Trauma in Adult Coturnix Quail. *Science (1979)* **240**, 1774–1776 (1988).
148. Umemoto, M. *et al.* Hair cell regeneration in the chick inner ear following acoustic trauma: ultrastructural and immunohistochemical studies. *Cell Tissue Research* **281**, 435–443 (1995).
149. Menendez, L. *et al.* Generation of inner ear hair cells by direct lineage conversion of primary somatic cells. *Elife* **9**, 1–33 (2020).
150. Du, X. *et al.* Regeneration of mammalian cochlear and vestibular hair cells through Hes1/Hes5 modulation with siRNA. *Hearing Research* **304**, 91–110 (2013).
151. Atkinson, P. J., Wise, A. K., Flynn, B. O., Nayagam, B. A. & Richardson, R. T. Hair cell regeneration after ATOH1 gene therapy in the cochlea of profoundly deaf adult guinea pigs. *PLoS ONE* **9**, (2014).
152. Coate, T. M., Scott, M. K. & Gurjar, M. Current concepts in cochlear ribbon synapse formation. *Synapse* **73**, (2019).
153. Hartong, D. T., Berson, E. L. & Dryja, T. P. Retinitis pigmentosa. *The Lancet* **368**, 1795–1809 (2006).
154. Yang, J. Usher Syndrome: Genes, Proteins, Models, Molecular Mechanisms, and Therapies. in *Hearing Loss* (2012). doi:10.5772/32663.
155. Delmaghani, S. & El-Amraoui, A. The genetic and phenotypic landscapes of Usher syndrome: from disease mechanisms to a new classification. *Human Genetics* **141**, 709–735 (2022).
156. French, L. S., Mellough, C. B., Chen, F. K. & Carvalho, L. S. A Review of Gene, Drug and Cell-Based Therapies for Usher Syndrome. *Frontiers in Cellular Neuroscience* **14**, (2020).
157. Abbott, J. A. *et al.* The Usher Syndrome Type IIIB Histidyl-tRNA Synthetase Mutation Confers Temperature Sensitivity. *Biochemistry* **56**, 3619–3631 (2017).
158. Dulon, D. *et al.* Clarin-1 gene transfer rescues auditory synaptopathy in model of Usher syndrome. *Journal of Clinical Investigation* **128**, 3382–3401 (2018).
159. Williams, D. S. *et al.* Harmonin in the murine retina and the retinal phenotypes of Ush1c-mutant mice and human USH1C. *Investigative Ophthalmology and Visual Science* **50**, 3881–3889 (2009).
160. Maerker, T. *et al.* A novel Usher protein network at the periciliary reloading point between molecular transport machineries in vertebrate photoreceptor cells. *Human Molecular Genetics* **17**, 71–86 (2008).
161. Zou, J. *et al.* Deletion of PDZD7 disrupts the Usher syndrome type 2 protein complex in cochlear hair cells and causes hearing loss in mice. *Human Molecular Genetics* **23**, 2374–2390 (2014).

162. Grati, M. *et al.* Localization of PDZD7 to the stereocilia ankle-link associates this scaffolding protein with the usher syndrome protein network. *Journal of Neuroscience* **32**, 14288–14293 (2012).
163. Morgan, C. P. *et al.* PDZD7-MYO7A complex identified in enriched stereocilia membranes. *Elife* **5**, (2016).
164. Goodyear, R. J., Marcotti, W., Kros, C. J. & Richardson, G. P. Development and properties of stereociliary link types in hair cells of the mouse cochlea. *Journal of Comparative Neurology* **485**, 75–85 (2005).
165. Zou, J. *et al.* Individual USH2 proteins make distinct contributions to the ankle link complex during development of the mouse cochlear stereociliary bundle. *Human Molecular Genetics* ddv398 (2015) doi:10.1093/hmg/ddv398.
166. Whatley, M. *et al.* Usher Syndrome: Genetics and Molecular Links of Hearing Loss and Directions for Therapy. *Frontiers in Genetics* **11**, (2020).
167. Bhattacharya, G., Miller, C., Kimberling, W. J., Jablonski, M. M. & Cosgrove, D. Localization and expression of usherin: a novel basement membrane protein defective in people with Usher's syndrome type IIa. *Hearing Research* **163**, 1–11 (2002).
168. Darribère, T. & Schwarzbauer, J. E. Fibronectin matrix composition and organization can regulate cell migration during amphibian development. *Mechanisms of Development* **92**, 239–250 (2000).
169. Odenthal, U. *et al.* Molecular analysis of laminin N-terminal domains mediating self-interactions. *Journal of Biological Chemistry* **279**, 44504–44512 (2004).
170. Engel, J. EGF-like domains in extracellular matrix proteins: localized signals for growth and differentiation? *FEBS* **251**, 1–7 (1989).
171. Mayer, U. *et al.* Low nidogen affinity of laminin-5 can be attributed to two serine residues in EGF-like motif g2III4. *FEBS Letters* **365**, 129–132 (1995).
172. Takagi, J., Yang, Y., Liu, J., Wang, J. & Springer, T. A. Complex between nidogen andlaminin fragments reveals aparadigmaticb-propeller interface. *Nature* **424**, 969–974 (2003).
173. Wells, H. R. R. *et al.* GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. *American Journal of Human Genetics* **105**, 788–802 (2019).
174. Hohenester, E. Laminin G-like domains: dystroglycan-specific lectins. *Current Opinion in Structural Biology* **56**, 56–63 (2019).
175. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
176. Weston, M. D. *et al.* Genomic Structure and Identification of Novel Mutations in Usherin, the Gene Responsible for Usher Syndrome Type IIa. *Am. J. Hum. Genet* **66**, 1199–1210 (2000).
177. Bjarnadóttir, T. K., Fredriksson, R. & Schiöth, H. B. The Adhesion GPCRs: A unique family of G protein-coupled receptors with important roles in both central and peripheral tissues. *Cellular and Molecular Life Sciences* **64**, 2104–2119 (2007).
178. Harmar, A. J. Protein family review Family-B G-protein-coupled receptors. *Genome Biology* **2**, (2001).
179. Sun, J.-P. *et al.* The Very Large G Protein Coupled Receptor (Vlgr1) in Hair Cells. *Journal of Molecular Neuroscience* **50**, 204–214 (2013).
180. Suzuki, K. *et al.* A synthetic synaptic organizer protein restores glutamatergic neuronal circuits. *Science (1979)* **369**, (2020).
181. Gewurz, H., Zhang, X.-H. & Lint, T. F. Structure and function of the pentraxins. *Current Opinion in Immunology* **7**, 54–64 (1995).
182. Yamagata, A. *et al.* Structural basis of epilepsy-related ligand-receptor complex LGI1-ADAM22. *Nature Communications* **9**, (2018).

183. Staub, E. *et al.* The novel EPTP repeat defines a superfamily of proteins implicated in epileptic disorders. *Trends in Biochemical Sciences* **27**, 441–444 (2002).
184. Myers, K. A. *et al.* ADGRV1 is implicated in myoclonic epilepsy. *Epilepsia* **59**, 381–388 (2018).
185. Libé-Philippot, B. *et al.* Auditory cortex interneuron development requires cadherins operating hair-cell mechanoelectrical transduction. *Proc Natl Acad Sci U S A* **114**, 7765–7774 (2017).
186. Langenhan, T., Piao, X. & Monk, K. R. Adhesion G protein-coupled receptors in nervous system development and disease. *Nature Reviews Neuroscience* **17**, 550–561 (2016).
187. Lin, H. H. *et al.* Autocatalytic cleavage of the EMR2 receptor occurs at a conserved G protein-coupled receptor proteolytic site motif. *Journal of Biological Chemistry* **279**, 31823–31832 (2004).
188. Araç, D. *et al.* A novel evolutionarily conserved domain of cell-adhesion GPCRs mediates autoproteolysis. *EMBO Journal* **31**, 1364–1378 (2012).
189. Knapp, B. *et al.* Affinity Proteomics Identifies Interaction Partners and Defines Novel Insights into the Function of the Adhesion GPCR VLGR1/ADGRV1. *Molecules* **27**, 3108 (2022).
190. Shin, D., Lin, S. T., Fu, Y. H. & Ptácek, L. J. Very large G protein-coupled receptor 1 regulates myelin-associated glycoprotein via Gαs/Gαq-mediated protein kinases A/C. *Proc Natl Acad Sci U S A* **110**, 19101–19106 (2013).
191. Hu, Q.-X. *et al.* Constitutive Gαi Coupling Activity of Very Large G Protein-coupled Receptor 1 (VLGR1) and Its Regulation by PDZD7 Protein. *Journal of Biological Chemistry* **289**, 24215–24225 (2014).
192. Fredriksson, R., Lagerström, M. C., Lundin, L.-G. & Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Molecular Pharmacology* **63**, 1256–1272 (2003).
193. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nature Reviews Drug Discovery* **16**, 19–34 (2017).
194. Wacker, D., Stevens, R. C. & Roth, B. L. How Ligands Illuminate GPCR Molecular Pharmacology. *Cell* **170**, 414–427 (2017).
195. Munk, C. *et al.* GPCRdb: the G protein-coupled receptor database – an introduction. *British Journal of Pharmacology* **173**, 2195–2207 (2016).
196. Kendall, R. T. & Luttrell, L. M. Diversity in arrestin function. *Cellular and Molecular Life Sciences* **66**, 2953–2973 (2009).
197. Cahill, T. J. *et al.* Distinct conformations of GPCR-β-arrestin complexes mediate desensitization, signaling, and endocytosis. *Proc Natl Acad Sci U S A* **114**, 2562–2567 (2017).
198. Zhou, X. E. *et al.* Identification of Phosphorylation Codes for Arrestin Recruitment by G Protein-Coupled Receptors. *Cell* **170**, 457-469.e13 (2017).
199. Dwivedi-Agnihotri, H. *et al.* Distinct phosphorylation sites in a prototypical GPCR differently orchestrate beta-arrestin interaction, trafficking, and signaling. *Sci. Adv* **6**, 8368–8379 (2020).
200. Chen, H., Zhang, S., Zhang, X. & Liu, H. QR code model: a new possibility for GPCR phosphorylation recognition. *Cell Communication and Signaling* **20**, (2022).
201. Qu, X. *et al.* Structural basis of tethered agonism of the adhesion GPCRs ADGRD1 and ADGRF1. *Nature* **604**, 779–785 (2022).
202. Ping, Y. Q. *et al.* Structural basis for the tethered peptide activation of adhesion GPCRs. *Nature* **604**, 763–770 (2022).

203. Xiao, P. *et al.* Tethered peptide activation mechanism of the adhesion GPCRs ADGRG2 and ADGRG4. *Nature* **604**, 771–778 (2022).
204. Barros-Álvarez, X. *et al.* The tethered peptide activation mechanism of adhesion GPCRs. *Nature* **604**, 757–762 (2022).
205. Ping, Y. Q. *et al.* Structures of the glucocorticoid-bound adhesion receptor GPR97–Go complex. *Nature* **589**, 620–626 (2021).
206. Weis, W. I. & Kobilka, B. K. The Molecular Basis of G Protein-Coupled Receptor Activation. *Annual Review of Biochemistry* **87**, 897–919 (2018).
207. Stephenson, J. R. *et al.* Brain-specific angiogenesis inhibitor-1 signaling, regulation, and enrichment in the postsynaptic density. *Journal of Biological Chemistry* **288**, 22248–22256 (2013).
208. Okajima, D., Kudo, G. & Yokota, H. Brain-specific angiogenesis inhibitor 2 (BAI2) may be activated by proteolytic processing. *Journal of Receptors and Signal Transduction* **30**, 143–153 (2010).
209. Paavola, K. J., Stephenson, J. R., Ritter, S. L., Alter, S. P. & Hall, R. A. The N terminus of the adhesion G protein-coupled receptor GPR56 controls receptor signaling activity. *Journal of Biological Chemistry* **286**, 28914–28921 (2011).
210. Mhaouty-Kodja, S. *et al.* Constitutively Active Alpha-1b Adrenergic Receptor Mutants Display Different Phosphorylation and Internalization Features. *The American Society for Pharmacology and Experimental Therapeutics* **55**, 339–347 (1999).
211. Kennedy, M. B. Origin of PDZ (DHR, GLGF) domain. *Trends in Biochemical Science* **20**, 350 (1995).
212. Cho, K.-O., Hunt, C. A. & Kennedy, M. B. The Rat Brain Postsynaptic Density Fraction Contains a Homolog of the Drosophila Discs-large Tumor Suppressor Protein. *Neuron* **9**, 929–942 (1992).
213. Luck, K., Charbonnier, S. & Travé, G. The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Letters* **586**, 2648–2661 (2012).
214. Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *Journal of Cell Science* **114**, 3219–3231 (2001).
215. Fan, J. S. & Zhang, M. Signaling complex organization by PDZ domain proteins. *NeuroSignals* **11**, 315–321 (2002).
216. Sheng, M. & Sala, C. PDZ Domains and the Organization of Supramolecular Complexes. *Annual Review in Neuroscience* **24**, 1–29 (2001).
217. Garner, C. C., Nash, J. & Huganir, R. L. PDZ domains in synapse assembly and signalling. *Trends in Cellular Biology* **10**, 274–280 (2000).
218. Bikkavilli, R. K. *et al.* Dishevelled3 is a novel arginine methyl transferase substrate. *Scientific Reports* **2**, 805 (2012).
219. Jané, P. *et al.* Interactomic affinity profiling by holdup assay: Acetylation and distal residues impact the PDZome-binding specificity of PTEN phosphatase. *PLOS ONE* **15**, e0244613 (2020).
220. Adey, N. B. *et al.* Threonine phosphorylation of the MMAC1/PTEN PDZ binding domain both inhibits and stimulates PDZ binding. *Cancer Res* **60**, 35–7 (2000).
221. Chung, H. J. Regulation of the NMDA Receptor Complex and Trafficking by Activity-Dependent Phosphorylation of the NR2B Subunit PDZ Ligand. *Journal of Neuroscience* **24**, 10248–10259 (2004).
222. Birrane, G., Chung, J. & Ladias, J. A. A. Novel Mode of Ligand Recognition by the Erbin PDZ Domain. *Journal of Biological Chemistry* **278**, 1399–1402 (2003).

223. Raghuram, V., Hormuth, H. & Foskett, J. K. A kinase-regulated mechanism controls CFTR channel gating by disrupting bivalent PDZ domain interactions. *Proceedings of the National Academy of Sciences* **100**, 9620–9625 (2003).
224. Voltz, J. W. *et al.* Phosphorylation of PDZ1 Domain Attenuates NHERF-1 Binding to Cellular Targets. *Journal of Biological Chemistry* **282**, 33879–33887 (2007).
225. Truschel, S. T., Zhang, M., Bachert, C., Macbeth, M. R. & Linstedt, A. D. Allosteric Regulation of GRASP Protein-dependent Golgi Membrane Tethering by Mitotic Phosphorylation. *Journal of Biological Chemistry* **287**, 19870–19875 (2012).
226. Shao, X. *et al.* Threonine 82 at the PDZ domain of PICK1 is critical for AMPA receptor interaction and localization. *Neurochemistry International* **56**, 962–970 (2010).
227. Raman, A. S., White, K. I. & Ranganathan, R. Origins of Allostery and Evolvability in Proteins: A Case Study. *Cell* **166**, 468–480 (2016).
228. Ivarsson, Y. *et al.* Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes. *Proc Natl Acad Sci U S A* **111**, 2542–2547 (2014).
229. Vincentelli, R. *et al.* Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nature Methods* **12**, 787–793 (2015).
230. Hillier, B. J., Christopherson, K. S., Prehoda, K. E., Bredt, D. S. & Lim, W. A. Unexpected Modes of PDZ Domain Scaffolding Revealed by Structure of nNOS-Syntrophin Complex. *Science (1979)* **284**, 812–815 (1999).
231. Tochio, H. *et al.* Formation of nNOS/PSD-95 PDZ dimer requires a preformed β-finger structure from the nNOS PDZ domain. *Journal of Molecular Biology* **303**, 359–370 (2000).
232. Penkert, R. R., DiVittorio, H. M. & Prehoda, K. E. Internal recognition through PDZ domain plasticity in the Par-6-Pals1 complex. *Nature Structural and Molecular Biology* **11**, 1122–1127 (2004).
233. Lemaire, J. F. & McPherson, P. S. Binding of Vac14 to neuronal nitric oxide synthase: Characterisation of a new internal PDZ-recognition motif. *FEBS Letters* **580**, 6948–6954 (2006).
234. Banerjee, M. *et al.* Specificity and promiscuity in human glutaminase interacting protein recognition: Insight from the binding of the internal and c-terminal motif. *Biochemistry* **51**, 6950–6960 (2012).
235. Erlendsson, S. *et al.* Protein interacting with C-kinase 1 (PICK1) binding promiscuity relies on unconventional PSD-95/discs-large/ZO-1 homology (PDZ) binding modes for nonclass II PDZ ligands. *Journal of Biological Chemistry* **289**, 25327–25340 (2014).
236. Awadia, S. *et al.* SGEF forms a complex with Scribble and Dlg1 and regulates epithelial junctions and contractility. *Journal of Cell Biology* **218**, 2699–2725 (2019).
237. Ali, M. *et al.* Integrated analysis of Shank1 PDZ interactions with C-terminal and internal binding motifs. *Current Research in Structural Biology* **3**, 41–50 (2021).
238. Mu, Y., Cai, P., Hu, S., Ma, S. & Gao, Y. Characterization of diverse internal binding specificities of PDZ domains by yeast two-hybrid screening of a special peptide library. *PLoS ONE* **9**, (2014).
239. Lenfant, N. *et al.* A genome-wide study of PDZ-domain interactions in C. elegans reveals a high frequency of non-canonical binding. *BMC Genomics* **11**, (2010).
240. Meerschaert, K. *et al.* The PDZ2 domain of zonula occludens-1 and -2 is a phosphoinositide binding domain. *Cellular and Molecular Life Sciences* **66**, 3951–3966 (2009).

241. Pan, L. *et al.* Clustering and synaptic targeting of PICK1 requires direct interaction between the PDZ domain and lipid membranes. *EMBO Journal* **26**, 4576–4587 (2007).
242. Ivarsson, Y. *et al.* Cooperative phosphoinositide and peptide binding by PSD-95/discs large/ZO-1 (PDZ) domain of polychaetoid, Drosophila zonulin. *Journal of Biological Chemistry* **286**, 44669–44678 (2011).
243. Wu, H. *et al.* PDZ Domains of Par-3 as Potential Phosphoinositide Signaling Integrators. *Molecular Cell* **28**, 886–898 (2007).
244. Zimmermann, P. *et al.* PIP 2-PDZ Domain Binding Controls the Association of Syntenin with the Plasma Membrane. *Molecular Cell* **9**, 1215–1225 (2002).
245. Chen, Y. *et al.* Genome-wide Functional Annotation of Dual-Specificity Protein- and Lipid-Binding Modules that Regulate Protein Interactions. *Molecular Cell* **46**, 226–237 (2012).
246. Ivarsson, Y. *et al.* Prevalence, Specificity and Determinants of Lipid-Interacting PDZ Domains from an In-Cell Screen and In Vitro Binding Experiments. *PLoS ONE* **8**, (2013).
247. Chang, B. H. *et al.* A systematic family-wide investigation reveals that ∼30% of mammalian PDZ domains engage in PDZ-PDZ interactions. *Chemistry and Biology* **18**, 1143–1152 (2011).
248. Cierpicki, T., Bushweller, J. H. & Derewenda, Z. S. Probing the supramodular architecture of a multidomain protein: The structure of syntenin in solution. *Structure* **13**, 319–327 (2005).
249. Kang, B. S. *et al.* PDZ tandem of human syntenin: Crystal structure and functional properties. *Structure* **11**, 459–468 (2003).
250. Grembecka, J. *et al.* The binding of the PDZ tandem of syntenin to target proteins. *Biochemistry* **45**, 3674–3683 (2006).
251. Grootjans, J. J., Reekmans, G., Ceulemans, H. & David, G. Syntenin-syndecan binding requires syndecan-synteny and the co-operation of both PDZ domains of syntenin. *Journal of Biological Chemistry* **275**, 19933–19941 (2000).
252. Bach, A. *et al.* A high-affinity, dimeric inhibitor of PSD-95 bivalently interacts with PDZ1-2 and protects against ischemic brain damage. *PNAS* **109**, 3317–3322 (2012).
253. Long, J. *et al.* Supramodular Nature of GRIP1 Revealed by the Structure of Its PDZ12 Tandem in Complex with the Carboxyl Tail of Fras1. *Journal of Molecular Biology* **375**, 1457–1468 (2008).
254. Feng, W., Shi, Y., Li, M. & Zhang, M. Tandem PDZ repeats in glutamate receptor–interacting proteins have a novel mode of PDZ domain–mediated target binding. *Nature Structural & Molecular Biology* **10**, 972–978 (2003).
255. Zhang, Q., Fan, J.-S. & Zhang, M. Interdomain Chaperoning between PSD-95, Dlg, and Zo-1 (PDZ) Domains of Glutamate Receptor-interacting Proteins. *Journal of Biological Chemistry* **276**, 43216–43220 (2001).
256. Mishra, P. *et al.* Dynamic Scaffolding in a G Protein-Coupled Signaling System. *Cell* **131**, 80–92 (2007).
257. Maisonneuve, P. *et al.* Regulation of the catalytic activity of the human phosphatase PTPN4 by its PDZ domain. *FEBS Journal* **281**, 4852–4865 (2014).
258. Chen, K. E. *et al.* Reciprocal allosteric regulation of p38γ and PTPN3 involves a PDZ domain-modulated complex formation. *Science Signaling* **7**, ra98 (2014).
259. Maisonneuve, P. *et al.* Molecular basis of the interaction of the human protein tyrosine phosphatase non-receptor type 4 (PTPN4) with the mitogen-activated protein kinase p38γ. *Journal of Biological Chemistry* **291**, 16699–16708 (2016).

260. Wilken, C., Kitzing, K., Kurzbauer, R., Ehrmann, M. & Clausen, T. Crystal Structure of the DegS Stress Sensor: How a PDZ Domain Recognizes Misfolded Protein and Activates a Protease. *Cell* **117**, 483–494 (2004).
261. Wilken, C., Kitzing, K., Kurzbauer, R., Ehrmann, M. & Clausen, T. Crystal Structure of the DegS Stress Sensor: How a PDZ Domain Recognizes Misfolded Protein and Activates a Protease. *Cell* **117**, 483–494 (2004).
262. Sohn, J., Grant, R. A. & Sauer, R. T. Allosteric Activation of DegS, a Stress Sensor PDZ Protease. *Cell* **131**, 572–583 (2007).
263. Zhu, J., Shang, Y. & Zhang, M. Mechanistic basis of MAGUK-organized complexes in synaptic development and signalling. *Nature Reviews Neuroscience* **17**, 209–223 (2016).
264. Zeng, M. *et al.* A binding site outside the canonical PDZ domain determines the specific interaction between Shank and SAPAP and their function. *Proc Natl Acad Sci U S A* **113**, E3081–E3090 (2016).
265. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *PNAS* **106**, 18249–18254 (2009).
266. Charbonnier, S. *et al.* The structural and dynamic response of MAGI-1 PDZ1 with noncanonical domain boundaries to the binding of human papillomavirus E6. *Journal of Molecular Biology* **406**, 745–763 (2011).
267. Ramírez, J. *et al.* Disorder-to-order transition of MAGI-1 PDZ1 C-terminal extension upon peptide binding: Thermodynamic and dynamic insights. *Biochemistry* **54**, 1327–1337 (2015).
268. Bhattacharya, S. *et al.* Ligand-induced dynamic changes in extended PDZ domains from NHERF1. *Journal of Molecular Biology* **425**, 2509–2528 (2013).
269. Delhommel, F. *et al.* Structural Characterization of Whirlin Reveals an Unexpected and Dynamic Supramodule Conformation of Its PDZ Tandem. *Structure* **25**, 1645-1656.e5 (2017).
270. Chi, C. N. *et al.* Interactions outside the boundaries of the canonical binding groove of a pdz domain influence ligand binding. *Biochemistry* **51**, 8971–8979 (2012).
271. Wang, C. K., Pan, L., Chen, J. & Zhang, M. Extensions of PDZ domains as important structural and functional elements. *Protein and Cell* **1**, 737–751 (2010).
272. Faure, G., Revy, P., Schertzer, M., Londono-Vallejo, A. & Callebaut, I. The C-terminal extension of human RTEL1, mutated in Hoyeraal-Hreidarsson syndrome, contains Harmonin-N-like domains. *Proteins: Structure, Function and Bioinformatics* **82**, 897–903 (2014).
273. Fisher, O. S. *et al.* Structure and vascular function of MEKK3-cerebral cavernous malformations 2 complex. *Nature Communications* **6**, (2015).
274. Lin, L. *et al.* Structure and Membrane Targeting of the PDZD7 Harmonin Homology Domain (HHD) Associated With Hearing Loss. *Frontiers in Cell and Developmental Biology* **9**, (2021).
275. Mathur, P. D. & Yang, J. Usher syndrome and non-syndromic deafness: Functions of different whirlin isoforms in the cochlea, vestibular organs, and retina. *Hearing Research* **375**, 14–24 (2019).
276. Lin, L. *et al.* Phase separation-mediated condensation of Whirlin-Myo15-Eps8 stereocilia tip complex. *Cell Reports* **34**, (2021).
277. Belyantseva, I. A., Boger, E. T. & Friedman, T. B. Myosin XVa localizes to the tips of inner ear sensory cell stereocilia and is essential for staircase formation of the hair bundle. *PNAS* **100**, 13958–13963 (2003).
278. Belyantseva, I. A. *et al.* Myosin-XVa is required for tip localization of whirlin and differential elongation of hair-cell stereocilia. *Nature Cell Biology* **7**, 148–156 (2005).

279. Disanza, A. *et al.* Eps8 controls actin-based motility by capping the barbed ends of actin filaments. *Nature Cell Biology* **6**, 1180–1188 (2004).
280. Manor, U. *et al.* Regulation of stereocilia length by myosin XVa and whirlin depends on the actin-regulatory protein Eps8. *Current Biology* **21**, 167–172 (2011).
281. Yap, C. C. *et al.* CIP98, a novel PDZ domain protein, is expressed in the central nervous system and interacts with calmodulin-dependent serine kinase. *Journal of Neurochemistry* **85**, 123–134 (2003).
282. Mburu, P. *et al.* Whirlin complexes with p55 at the stereocilia tip during hair cell development. *PNAS* **103**, 10973–10978 (2006).
283. Okumura, K. *et al.* Protein 4.1 expression in the developing hair cells of the mouse inner ear. *Brain Research* **1307**, 53–62 (2010).
284. Mauriac, S. A. *et al.* Defective Gpsm2/Gαi3 signalling disrupts stereocilia development and growth cone actin dynamics in Chudley-McCullough syndrome. *Nature Communications* **8**, 1–17 (2017).
285. Tadenev, A. L. D. *et al.* GPSM2-GNAI Specifies the Tallest Stereocilia and Defines Hair Bundle Row Identity. *Current Biology* **29**, 921-934.e4 (2019).
286. Hsueh, Y.-P., Wang, T.-F., Yang, F.-C. & Sheng, M. Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2. *Nature* **404**, 298–205 (2000).
287. Tibbe, D., Pan, Y. E., Reißner, C., Harms, F. L. & Kreienkamp, H. J. Functional analysis of CASK transcript variants expressed in human brain. *PLoS ONE* **16**, (2021).
288. Chen, Q., Zou, J., Shen, Z., Zhang, W. & Yang, J. Whirlin and PDZ domain-containing 7 (PDZD7) proteins are both required to form the quaternary protein complex associated with Usher syndrome type 2. *Journal of Biological Chemistry* **289**, 36070–36088 (2014).
289. Riazuddin, S. *et al.* Alterations of the CIB2 calcium-and integrin-binding protein cause Usher syndrome type 1J and nonsyndromic deafness DFNB48. *Nature Genetics* **44**, 1265–1271 (2012).
290. Liu, X., Udovichenko, I. P., Brown, S. D. M., Steel, K. P. & Williams, D. S. Myosin VIIa Participates in Opsin Transport through The Photoreceptor Cilium. *The Journal of Neuroscience* **19**, 6267–6274 (1999).
291. Sorusch, N. *et al.* Characterization of the ternary Usher syndrome SANS/ush2a/whirlin protein complex. *Human Molecular Genetics* **26**, 1157–1172 (2017).
292. Yang, J. *et al.* Ablation of whirlin long isoform disrupts the USH2 protein complex and causes vision and hearing loss. *PLoS Genet* **6**, e1000955 (2010).
293. Overlack, N. *et al.* Direct interaction of the Usher syndrome 1G protein SANS and myomegalin in the retina. *Biochimica et Biophysica Acta - Molecular Cell Research* **1813**, 1883–1892 (2011).
294. Papal, S. *et al.* The giant spectrin βV couples the molecular motors to phototransduction and Usher syndrome type I proteins along their trafficking route. *Human Molecular Genetics* **22**, 3773–3788 (2013).
295. El-Amraoui, A. & Petit, C. The retinal phenotype of Usher syndrome: Pathophysiological insights from animal models. *Comptes Rendus - Biologies* **337**, 167–177 (2014).
296. Schneider, E. *et al.* Homozygous disruption of PDZD7 by reciprocal translocation in a consanguineous family: A new member of the Usher syndrome protein interactome causing congenital hearing impairment. *Human Molecular Genetics* **18**, 655–666 (2009).

297. Du, H. *et al.* Lack of PDZD7 long isoform disrupts ankle-link complex and causes hearing loss in mice. *The FASEB Journal* **34**, 1136–1149 (2020).
298. Vona, B. *et al.* Confirmation of PDZD7 as a Nonsyndromic Hearing Loss Gene. *Ear & Hearing* **37**, e238–e246 (2016).
299. Booth, K. T. *et al.* PDZD7 and hearing loss: More than just a modifier. *American Journal of Medical Genetics Part A* **167**, 2957–2965 (2015).
300. Guan, J. *et al.* Novel recessive PDZD7 biallelic mutations in two Chinese families with non-syndromic hearing loss. *American Journal of Medical Genetics, Part A* **176**, 99–106 (2018).
301. Wu, D. *et al.* Clinical and genetic study of 12 Chinese Han families with nonsyndromic deafness. *Molecular Genetics and Genomic Medicine* **8**, (2020).
302. Luo, H. *et al.* Novel recessive PDZD7 biallelic mutations associated with hereditary hearing loss in a Chinese pedigree. *Gene* **709**, 65–74 (2019).
303. Lee, S. Y. *et al.* Identification of a potential founder effect of a novel PDZD7 variant involved in moderate-to-severe sensorineural hearing loss in Koreans. *International Journal of Molecular Sciences* **20**, (2019).
304. Fahimi, H., Behroozi, S., Noavar, S. & Parvini, F. A novel recessive PDZD7 bi-allelic mutation in an Iranian family with non-syndromic hearing loss. *BMC Medical Genomics* **14**, (2021).
305. Ebermann, I. *et al.* PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. *Journal of Clinical Investigation* **120**, 1812–1823 (2010).
306. Li, J. *et al.* Structure of Myo7b/USH1C complex suggests a general PDZ domain binding mode by MyTH4-FERM myosins. *Proc Natl Acad Sci U S A* **114**, E3776–E3785 (2017).
307. Zou, J. *et al.* Whirlin replacement restores the formation of the USH2 protein complex in whirlin knockout photoreceptors. *Invest Ophthalmol Vis Sci* **52**, 2343–51 (2011).
308. Schietroma, C. *et al.* Usher syndrome type 1 – associated cadherins shape the photoreceptor outer segment. *The Journal of Cell Biology* **216**, 1849–1864 (2017).
309. Mathur, P. & Yang, J. Usher syndrome: Hearing loss, retinal degeneration and associated abnormalities. *Biochim Biophys Acta* **1852**, 406–420 (2015).
310. Richardson, G. P., Monvel, J. B. De & Petit, C. How the Genetics of Deafness Illuminates Auditory Physiology. *Annu Rev Physiol* **73**, 311–334 (2011).
311. Kremer, H., van Wijk, E., Märker, T., Wolfrum, U. & Roepman, R. Usher syndrome: molecular links of pathogenesis, proteins and pathways. *Hum Mol Genet* **15 Spec No**, R262-70 (2006).
312. Keats, B. J. & Corey, D. P. The usher syndromes. *Am J Med Genet* **89**, 158–66 (1999).
313. Zou, J. *et al.* The roles of USH1 proteins and PDZ domain-containing USH proteins in USH2 complex integrity in cochlear hair cells. *Human Molecular Genetics* **26**, 624–636 (2017).
314. Zou, J. *et al.* Deletion of PDZD7 disrupts the Usher syndrome type 2 protein complex in cochlear hair cells and causes hearing loss in mice. *Human Molecular Genetics* **23**, 2374–2390 (2014).
315. Reiners, J. *et al.* Scaffold protein harmonin (USH1C) provides molecular links between Usher syndrome type 1 and type 2. *Hum Mol Genet* **14**, 3933–3943 (2005).
316. Mathur, P. D. & Yang, J. Usher syndrome and non-syndromic deafness : Functions of different whirlin isoforms in the cochlea, vestibular organs, and retina. *Hearing Research* **375**, 14–24 (2019).
317. Sorusch, N. *et al.* Characterization of the ternary Usher syndrome SANS / ush2a / whirlin protein complex. *Hum Mol Genet* **26**, 1157–1172 (2017).

318. Wu, L., Pan, L., Zhang, C. & Zhang, M. Large protein assemblies formed by multivalent interactions between cadherin23 and harmonin suggest a stable anchorage structure at the tip link of stereocilia. *J Biol Chem* **287**, 33460–33471 (2012).
319. Bahloul, A. *et al.* Cadherin-23, myosin viia and harmonin, encoded by usher syndrome type I genes, form a ternary complex and interact with membrane phospholipids. *Human Molecular Genetics* **19**, 3557–3565 (2010).
320. Adato, A. *et al.* Interactions in the network of Usher syndrome type 1 proteins. *Hum Mol Genet* **14**, 347–356 (2005).
321. Du, H. *et al.* Lack of PDZD7 long isoform disrupts ankle-link complex and causes hearing loss in mice. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **34**, 1136–1149 (2020).
322. Du, H., Ren, R., Chen, P., Xu, Z. & Wang, Y. Identification of Binding Partners of Deafness-Related Protein PDZD7. *Neural Plasticity* **2018**, 1–10 (2018).
323. Grati, M. *et al.* Localization of PDZD7 to the Stereocilia Ankle-Link Associates this Scaffolding Protein with the Usher Syndrome Protein Network. *Journal of Neuroscience* **32**, 14288–14293 (2012).
324. Fisher, O. S. *et al.* Structure and vascular function of MEKK3-cerebral cavernous malformations 2 complex. *Nature Communications* (2015) doi:10.1038/ncomms8937.
325. Pan, L., Yan, J., Wu, L. & Zhang, M. Assembling stable hair cell tip link complex via multidentate interactions between harmonin and cadherin 23. *Proceedings of the National Academy of Sciences* **106**, 5575–5580 (2009).
326. Wang, X. *et al.* Structural Insights into the Molecular Recognition between Cerebral Cavernous Malformation 2 and Mitogen-Activated Protein Kinase Kinase Kinase 3. *Structure* **23**, 1087–1096 (2015).
327. Bahloul, A. *et al.* Conformational switch of harmonin, a submembrane scaffold protein of the hair cell mechanoelectrical transduction machinery. *FEBS Letters* **591**, 2299–2310 (2017).
328. Yan, J., Pan, L., Chen, X., Wu, L. & Zhang, M. TEST 2 The structure of the harmonin/sans complex reveals an unexpected interaction mode of the two Usher syndrome proteins. *Proceedings of the National Academy of Sciences* **107**, 4040–4045 (2010).
329. Delhommel, F. *et al.* Structural plasticity of the HHD2 domain of whirlin. *The FEBS Journal* **285**, 3738–3752 (2018).
330. The Uniprot Consortium. UniProt : a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**, 506–515 (2019).
331. The Uniprot Consortium. UniProt : the universal protein knowledgebase. *Nucleic Acids Research* **45**, 158–169 (2017).
332. Spronk, C. A. E. M. *et al.* The Mad1-Sin3B interaction involves a novel helical fold. *Nat Struct Biol* **7**, 1100–1104 (2000).
333. Fisher, O. S. *et al.* Structural studies of cerebral cavernous malformations 2 (CCM2) reveal a folded helical domain at its C-terminus. *FEBS Letters* **587**, 272–277 (2013).
334. Lamaze, A., Jepson, J. E. C., Akpoghiran, O. & Koh, K. Antagonistic Regulation of Circadian Output and Synaptic Development by JETLAG and the DYSCHRONIC-SLOWPOKE Complex. *Iscience* **23**, 100845 (2020).
335. Jepson, J. E. C. *et al.* Regulation of synaptic development and function by the Drosophila PDZ protein Dyschronic. *Development* **141**, 4548–4557 (2014).
336. Jepson, J. E. C. *et al.* dyschronic, a Drosophila Homolog of a Deaf-Blindness Gene, Regulates Circadian Output and Slowpoke Channels. *PLoS Genet* **8**, e1002671 (2012).
337. Li, W., Jaroszewski, L. & Godzik, A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17**, 282–283 (2001).

338. Walne, A. J., Vulliamy, T., Kirwan, M., Plagnol, V. & Dokal, I. Constitutional Mutations in RTEL1 Cause Severe Dyskeratosis Congenita. *The American Journal of Human Genetics* **92**, 448–453 (2013).
339. Bonnet, C. *et al.* An innovative strategy for the molecular diagnosis of Usher syndrome identi fi es causal biallelic mutations in 93 % of European patients. *European Journal of Human Genetics* **24**, 1730–1738 (2016).
340. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).
341. Eddy, S. R. Accelerated Profile HMM Searches. *Plos computational biology* **7**, e1002195 (2011).
342. Camacho, C. *et al.* BLAST + : architecture and applications. *BMC Bioinformatics* **10**, 421–429 (2009).
343. Miele, V., Penel, S. & Duret, L. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* **12**, 116–124 (2011).
344. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, complex systems* **1695**, 1–9 (2006).
345. Blondel, V. D., Guillaume, J., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).
346. Nguyen, L., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE : A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2014).
347. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder : fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587–589 (2017).
348. Lemoine, F. *et al.* Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**, 452–456 (2018).
349. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **54**, (2016).
350. Van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server : User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* **428**, 720–725 (2016).
351. Popot, J.-L. *et al.* Amphipols: polymeric surfactants for membrane biology research. *Cellular and Molecular Life Sciences (CMLS)* **60**, 1559–1574 (2003).
352. Delhommel, F. *et al.* Structural plasticity of the HHD2 domain of whirlin. *The FEBS Journal* **285**, 3738–3752 (2018).
353. Lin, L. *et al.* Structure and Membrane Targeting of the PDZD7 Harmonin Homology Domain (HHD) Associated With Hearing Loss. *Frontiers in Cell and Developmental Biology* **9**, (2021).
354. Demberg, L. M. *et al.* Activation of adhesion G protein-coupled receptors: Agonist specificity of Stachel sequence-derived peptides. *Journal of Biological Chemistry* **292**, 4383–4394 (2017).
355. Stoveken, H. M., Hajduczok, A. G., Xu, L. & Tall, G. G. Adhesion G protein-coupled receptors are activated by exposure of a cryptic tethered agonist. *Proceedings of the National Academy of Sciences* **112**, 6194–6199 (2015).
356. Zou, J. *et al.* The roles of USH1 proteins and PDZ domain-containing USH proteins in USH2 complex integrity in cochlear hair cells. *Human Molecular Genetics* ddw421 (2016) doi:10.1093/hmg/ddw421.

357. Fuster-García, C. *et al.* High-throughput sequencing for the molecular diagnosis of Usher syndrome reveals 42 novel mutations and consolidates CEP250 as Usher-like disease causative. *Scientific Reports* **8**, 17113 (2018).
358. Eswar, N. *et al.* Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics* **15**, (2006).
359. Bonomi, M. *et al.* PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **180**, 1961–1972 (2009).
360. Branduardi, D., Bussi, G. & Parrinello, M. Metadynamics with Adaptive Gaussians. *Journal of Chemical Theory and Computation* **8**, 2247–2254 (2012).
361. Pietrucci, F. & Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *Journal of Chemical Theory and Computation* **5**, 2197–2201 (2009).
362. Laursen, L., Karlsson, E., Gianni, S. & Jemth, P. Functional interplay between protein domains in a supramodular structure involving the postsynaptic density protein PSD-95. *Journal of Biological Chemistry* **295**, 1992–2000 (2020).
363. Dunn, H. A. & Ferguson, S. S. G. PDZ Protein Regulation of G Protein–Coupled Receptor Trafficking and Signaling Pathways. *Molecular Pharmacology* **88**, 624–639 (2015).
364. Masuho, I. *et al.* A Global Map of G Protein Signaling Regulation by RGS Proteins. *Cell* **183**, 503-521.e19 (2020).
365. Fouquet, C. *et al.* Improving Axial Resolution in Confocal Microscopy with New High Refractive Index Mounting Media. *PLOS ONE* **10**, e0121096 (2015).
366. Weinhardt, V. *et al.* Switchable resolution in soft x-ray tomography of single cells. *PLOS ONE* **15**, e0227601 (2020).
367. Schaffer, M. *et al.* A cryo-FIB lift-out technique enables molecular-resolution cryo-ET within native Caenorhabditis elegans tissue. *Nature Methods* **16**, 757–762 (2019).
368. Bäuerlein, F. J. B., Pastor-Pareja, J. C. & Fernández-Busnadiego, R. Cryo-electron tomography of native Drosophila tissues vitrified by plunge freezing. doi:10.1101/2021.04.14.437159.
369. KUBA, J. *et al.* Advanced cryo-tomography workflow developments – correlative microscopy, milling automation and cryo-lift-out. *Journal of Microscopy* **281**, 112–124 (2021).
370. Mahamid, J. *et al.* A focused ion beam milling and lift-out approach for site-specific preparation of frozen-hydrated lamellas from multicellular organisms. *Journal of Structural Biology* **192**, 262–269 (2015).
371. Bhogaraju, S., Engel, B. D. & Lorentzen, E. Intraflagellar transport complex structure and cargo interactions. *Cilia* **2**, 10 (2013).
372. Engel, B. D. *et al.* The role of retrograde intraflagellar transport in flagellar assembly, maintenance, and function. *Journal of Cell Biology* **199**, 151–167 (2012).
373. Maria, P. L. & SOghalai, J. *Anatomy and Physiology of the Auditory System. Sataloff's Comprehensive Textbook of Otolaryngology: Head & Neck Surgery: Otology/Neurotology/Skull Base Surgery*. vol. 1 (2015).
374. Fettiplace, R. & Nam, J. H. Tonotopy in calcium homeostasis and vulnerability of cochlear hair cells. *Hearing Research* **376**, 11–21 (2019).
375. Nayak, G. D., Ratnayaka, H. S. K., Goodyear, R. J. & Richardson, G. P. Development of the hair bundle and mechanotransduction. *Int J Dev Biol* **51**, 597–608 (2007).
376. Corey, D. P., Akyuz, N. & Holt, J. R. Function and dysfunction of TMC channels in inner ear hair cells. *Cold Spring Harbor Perspectives in Medicine* **9**, (2019).
377. Dionne, G. *et al.* Mechanotransduction by PCDH15 Relies on a Novel cis-Dimeric Architecture. *Neuron* **99**, 480-492.e5 (2018).

378. Bruce, I. C., Erfani, Y. & Zilany, M. S. A. A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. *Hearing Research* **360**, 40–54 (2018).
379. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator: Figure 1. *Genome Research* **14**, 1188–1190 (2004).