Etude structurale de la Whirline, protéine modulaire cruciale dans les mécanismes de la vision et de l'audition

Florent Delhommel

▶ **To cite this version:**

Florent Delhommel. Etude structurale de la Whirline, protéine modulaire cruciale dans les mécanismes de la vision et de l'audition. Biophysique. Université Pierre et Marie Curie - Paris VI, 2017. Français. NNT : 2017PA066151 . tel-02169200v2

HAL Id: tel-02169200

https://theses.hal.science/tel-02169200v2

Submitted on 1 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unité de RMN des Biomolécules
Institut Pasteur
28, rue du Docteur Roux
75724 PARIS Cedex 15

THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité

Biophysique moléculaire
Ecole doctorale ED515 Complexité du vivant

Présentée par

M. Florent Delhommel

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE – Paris VI

Sujet de la thèse :

Etude structurale de la Whirline, protéine modulaire cruciale dans les mécanismes de la vision et de l'audition

soutenue le 29 Juin 2017

devant le jury composé de :

Dr. Nicolas WOLFF	Directeur de thèse
Dr Martin Blackledge	Rapporteur
Dr Gilles Travé	Rapporteur
Pr Germain Trugnan	Examinateur
Dr Sophie Zinn-Justin	Examinateur
Dr Pascale Zimmermann	Examinateur

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc-nd/3.0/

Remerciement

Je voudrais profiter des premières pages de ce manuscrit afin de manifester ma gratitude envers les nombreuses personnes qui ont contribué directement ou indirectement à cette thèse. Bien que seul mon nom apparaisse sur la couverture de ce manuscrit, son contenu résulte de la contribution, de l'influence et de l'aide de personnes que je ne remercierai jamais assez.

Je tiens tout d'abord à remercier les membres du jury, Dr. Gilles Travé et Dr. Martin Blackledge en tant que rapporteurs ainsi que Dr. Pascale Zimmermann, Dr. Sophie Zinn-Justin et Pr. Germain Trugnan en tant qu'examinateurs, d'avoir accepté de dédier de leur temps à l'évaluation de ce travail.

Je remercie également Muriel Delepierre de m'avoir accueilli, à plusieurs reprises, au sein de son unité pendant ces quelques années. Tu as réussi la tâche difficile d'établir un laboratoire équilibré et basé sur l'échange bien que composé de personnes d'horizons largement différents, qui fournissent un environnement de travail idéal. Je tiens également à te remercier pour ton soutien constant au cours de cette thèse, ainsi que pour nos discussions sur la science d'avant ma génération, qui m'auront permis d'apprécier pleinement nos conditions de travail moderne (en plus d'apprendre à intégrer une courbe sans ordinateur).

Je tiens tout particulièrement à remercier Nicolas pour ces années à travailler à ses cotés. Je te dois premièrement ce superbe sujet d'étude que tu m'as proposé à la fin de mon master et sur lequel tu as eu le courage de m'accorder ta confiance. Ta disponibilité de tous les instants, ta gentillesse et le soin particulier que tu portes aux personnes qui t'entourent auront rendu ces années de thèse particulières et importantes à mes yeux. Ce que j'ai pu apprendre à tes cotés me servira bien au delà de ma carrière scientifique.

Je tiens aussi à profondément remercier Florence avec qui j'ai travaillé étroitement pendant toutes ces années. Je te remercie de m'avoir initié à la RMN et de m'y avoir fait prendre goût, en me faisant découvrir toute la diversité que celle-ci offre. Merci aussi pour ta patiente, lors de mes nombreuses apparitions inopinées dans ton bureau, pour discuter d'une analyse ou proposer de nouvelles expériences bizarres.

Un grand merci à Elouan pour m'avoir introduit à la biophysique lors de mon tout premier stage au laboratoire. Ton encadrement durant ce stage aura été extrêmement influent sur ma décision de continuer pour une thèse dans ce domaine. De même, je remercie Célia pour son soutien, sa détermination et son franc parler. Nos interactions m'auront permis d'envisager la recherche comme un métier en plus d'une passion. Aussi, merci à Pierre pour sa constante positivité et bonne humeur. Je voudrais aussi remercier Juan "El zapoteco" Ramos pour ses discours ardents du vendredi soir ; ainsi que Virginie pour ton attitude toujours décontractée et la bonne ambiance que tu as propagée au sein du laboratoire (aussi, et de façon très importante, pour m'avoir fait découvrir "Charlie the unicorne").

Je remercie aussi l'ensemble des membres du laboratoire que j'ai côtoyé pendant des années : Catherine, merci de m'avoir patiemment appris à utiliser les spectromètres, au tout début de ma thèse. Inaki, merci pour tes conseils avisés en biochimie et tes suggestions sur le projet. Nadia, merci pour ta bonne humeur et pour prendre en charge l'organisation de la vie du laboratoire. Merci Bruno pour les nombreux coups de main informatiques et pour tes théories toujours plus farfelues et stimulantes! Merci Aracelys pour ton calme et ton attitude décontractée malgré les nombreuses heures de travail. Enfin merci à Marissa, Mariano & Borja pour insuffler un vent nouveau, et hispanisant, dans le laboratoire. Je vous souhaite une bonne continuation et de profiter au maximum de vos années de thèse. Je remercie aussi Idir et Ariane, que je n'aurais pas côtoyés longtemps mais qui ont toujours été bienveillants et disponibles.

Je veux enfin particulièrement remercier Yanlei, qui prendra ma relève. J'espère que tu auras autant de plaisir que moi à travailler sur ce magnifique sujet. Bonne continuation pour la suite, à la vue de la quantité de résultats que tu as déjà accumulée, je ne me fais aucun souci!

Je voudrais remercier les quatre étudiants que j'ai eu le plaisir d'encadrer pendant cette thèse, et qui auront été mes cobayes bien malgré eux. Merci à Madeleine, tu as été ma première étudiante et tu m'aura rendu la tache extrêmement facile par ton efficacité et ton sérieux. J'espère que tu trouveras un laboratoire qui te correspond pour la suite ! Merci à Pierre-Louis, ton passage dans notre groupe aura été de courte durée, mais extrêmement formateur et riche en émotion, pour ma part. Merci aussi à Baptiste, ta détermination et ta curiosité ne pourront que t'amener à réussir, peu importe ce que tu entreprendras, je n'ai pas de doute qu'on se recroisera bientôt ! Enfin merci à Lucas, ton avidité de découvrir des sujets

extrêmement variés et ta capacité à les intégrer et à les comprendre sont une force qui te servira toujours. J'espère que ta thèse se continuera bien, ça a l'air d'être en bonne voie !

Le travail de cette thèse n'aura pas pu être possible sans la contribution de nos nombreux collaborateurs.

Je voudrais tout d'abord remercier les membres de l'unité de génétique et physiologie de l'audition. Tout particulièrement Elise, pour ton énergie et pour le temps que tu auras dédié à m'expliquer tous les détails du fonctionnement de l'audition, me permettant d'être efficace rapidement, au début de cette thèse. Merci aussi à Sylvie pour sa constante bonne humeur et pour sa capacité surhumaine de venir à bout de n'importe quel problème de biologie moléculaire. Merci aussi à Alice pour avoir toujours été disponible pour m'aider ou pour simplement discuter. Je voudrais enfin remercier Amel et le Christine Petit pour avoir amené ce sujet à notre laboratoire.

Je voudrais ensuite dire merci aux membres de l'unité de bioinformatique structurale. Notamment à Benjamin et à Guillaume pour leur patience, malgré mes demandes constantes pour de nouvelles analyses. Ce projet n'aurait jamais pu aboutir sans leur précieuse contribution. Je voudrais aussi remercier Bradley et Thérèse pour cette collaboration inattendue, sur des aspects plus théoriques du traitement de données.

Je voudrais enfin remercier l'ensemble des plateformes et collaborateurs plus ponctuels. Notamment Sébastien Brulé, Bruno, Sylviane, Patrick Weber et Cédric. Je voudrais aussi particulièrement remercier Sébastien Brier, Alexandre, Marco, Patrick England, Bertrand Raynal, Ahmed Haouz et Frederick qui m'auront beaucoup appris tant sur les plans pratiques que théoriques, sur de nombreuses techniques de biophysique. Travailler à vos cotés aura été un plaisir, votre curiosité et votre envie de toujours aider représentent, pour moi, la philosophie même de l'Institut Pasteur.

Je remercie aussi François-Xavier Canterelle et Aurélien Thureau pour nous avoir reçu à de nombreuses reprises sur les plateformes nationales de RMN et de SAXS et qui ont toujours été d'un grand soutien.

Je tiens aussi particulièrement à remercier Henri pour son soutien et pour nos nombreuses discussions constructives et enrichissantes sur sa vision globale de la biologie. Je tiens aussi à remercier Alain. Merci pour tes nombreux conseils et ton aide sur d'innombrables techniques. Ta rigueur tant dans l'expérimentation que dans l'interprétation des résultats resteront, pour moi, un modèle à suivre. Je remercie aussi François-Xavier pour

ton honnêteté, ta bienveillance et tes nombreux conseils. De même, merci à Pierre-Jean et Aziz qui auront suivi ce sujet depuis le début et qui m'auront permis de progresser par leurs conseils éclairés.

Je tiens aussi à remercier tous les habitants du 1er étage du bâtiment Lwoff. Merci à Denise pour ta bonne humeur, ton efficacité et ton aide (indispensable) pour toutes les demandes administratives. Je tiens à remercier Jojo, Geneviève et Chantal pour leur travail silencieux mais essentiel au fonctionnement du laboratoire. Merci aux membres d'UPV, Isa, Daniela, Moran, Fanny et Madhura. Je voudrais particulièrement remercier Mathieu et Philippe, membres d'honneurs du club omnisport de l'étage ainsi que de la buvette du vendredi soir. Merci à Philippe pour toujours être disponible afin de partager ses connaissances encyclopédiques des … équipements … de l'Institut Pasteur. Merci à Mathieu *a.k.a.* l'homme qui apparaît comme par magie à 17h dans le bureau. On a "grandi" ensemble depuis notre stage de M2, c'est maintenant le moment de fermer ce chapitre, bon courage pour ta rédaction !

Enfin, un grand merci à Eleonora avec qui je partage ma vie. En plus de m'avoir supporté durant ces derniers mois intenses, tu es toujours là pour me rassurer et me redonner confiance dans les moments de doute. Je suis impatient de découvrir le futur que nous construirons ensemble !

Pour finir je voudrais particulièrement remercier mes parents. En plus de m'avoir inculqué les valeurs et la discipline qui m'auront permis de réaliser cette thèse, vous m'avez toujours soutenu et fait le maximum pour que mes choix se réalisent, depuis le collège jusqu'à ce jour. Je ne peux que vous être infiniment reconnaissant pour tout ce que vous avez fait et tout ce que vous continuez de faire. Vous êtes et avez toujours été une source d'inspiration continuelle.

A la mémoire de Georgios Kyriacos Panayiotou

Etude structurale de la Whirline, protéine modulaire cruciale dans les mécanismes de la vision et de l'audition

Résumé

La vue et l'ouïe sont permises par des cellules capables de rapidement traduire une onde, lumineuse ou sonore, en un message électrochimique transmissible au cerveau. La fonction de ces cellules repose sur des morphologies uniques : les photorécepteurs sont hautement compartimentalisés et nécessitent le renouvellement constant de leurs protéines photoréceptrices, transmises le long d'un cil connecteur; les cellules de l'audition sont équipées de touffes de stéréocils mécanosensibles défléchies par l'onde sonore. Les mutations de onze gènes sont la cause des syndromes Usher, associant surdité congénitale et cécité. Les protéines Usher codées par ces gènes sont indispensables à l'architecture de ces deux types cellulaires ; elles forment de multiples complexes dont les interactions clés sont maintenues principalement par des domaines PDZ. L'une des protéines centrales de ce réseau est la Whirline une grande protéine multi-domaine contenant trois domaines PDZ.

Pour comprendre les bases moléculaires des syndromes Usher, nous nous sommes concentrés sur la caractérisation biochimique et biophysique de la Whirline. Nous avons d'abord identifié un nouveau domaine HHD2 de la Whirline dont nous avons obtenue la structure à haute résolution et déterminé le comportement en solution, isolé et avec les domaines adjacents. Nous avons ensuite caractérisé, dans la Whirline, un supramodule transitoire entre deux domaines PDZ, maintenus par des extensions structurées en épingle-ß de chacun des domaines. Nous avons déterminé la structure de la conformation compacte unique de ce complexe et étudié son équilibre avec un ensemble de conformations étendues. Nous avons enfin caractérisé *in vitro* le réseau d'interaction des domaines PDZ de la Whirline avec la majorité des protéines impliquées dans la vision et l'audition. Nous avons ainsi montré que l'interaction du premier PDZ du supramodule avec ses partenaires est favorisée par la présence de son extension en épingle-ß et que le second domaine PDZ est quant à lui inactif.

L'ensemble de nos résultats sur la structure modulaire et l'interactome de la Whirline permet de mieux comprendre le rôle de la Whirline dans les différents complexes Usher et expliquer les conséquences de ses mutations sur les mécanismes moléculaires de l'audition et de la vision.

Mots clés : Supramodule, PDZ, Whirline, cellules ciliées, Syndrome Usher, photorécepteurs

Adresse du laboratoire : Unité de RMN des Biomolécules
Institut Pasteur
25-28 Rue du Dr. Roux
75724 Paris Cedex 15

Summary

Vision and hearing rely on the capacity of cells to rapidly transduce electromagnetic waves or sound waves into chemical messages that are transmissible to the brain. The function of these cells requires unique morphologies: photoreceptor cells are highly compartmentalized and require a constant renewal of their photoreceptive proteins, along a connecting cilium; hearing cells are equipped with mechanosensitive bundle of stereocilia, deflected by the sound wave. The mutations of eleven genes are responsible for Usher syndromes, associating congenital blindness and deafness. The Usher proteins encoded by these genes are pivotal to the architecture of the photoreceptor and hearing cells. They form multiples complexes in which the critical interactions are mainly maintained by PDZ domains. One of these central proteins is Whirlin, a large multi-domain protein encompassing three PDZ domains.

To understand the molecular basis of the Usher syndromes, we focused our project on the biochemical and biophysical characterization of Whirlin. We first identified a new HHD2 domain on Whirlin, for which we solved the structure at high resolution and determined the behavior in solution, isolated or with adjacent domains. We then identified a transient supramodule between two of Whirlin PDZ domains, maintained by PDZ extensions structured in ß-hairpins. We determined the structure of the compact and unique conformation of this tandem and we characterized its equilibrium with an ensemble of more extended conformations. Finally, we characterized *in vitro* the network of interaction of the PDZ domains of Whirlin, with the majority of the proteins involved in hearing and vision. We demonstrated that the interaction of its first PDZ domain with its partners is enhanced by its ß-hairpin extension stabilized in the tandem and we showed that the second PDZ is inactive as binder.

Our results on the modular structure and the interactome of Whirlin get insight into the role of Whirlin in the numerous complexes formed by the Usher proteins and allow to better explain the consequences of its mutation on the molecular mechanisms of hearing and vision.

Table des matières

Remerciement ... *3*
Résumé ... *7*
Summary ... *7*
Table des matières ... *10*
Introduction ... **15**
Préambule ... *16*
1 De l'onde sonore au message neuronale, les mécanismes de l'audition. ... *18*
1.1 De l'onde sonore à l'oreille interne ... 18
1.1.1 Onde sonore et capacités de l'oreille ... 18
1.1.2 Oreille externe ... 18
1.1.3 Oreille Moyenne ... 19
1.2 Oreille Interne ... 20
1.2.1 Le vestibule ... 20
1.2.1.1 Les canaux semi-circulaires ... 21
1.2.1.2 Les organes otolithiques ... 22
1.2.2 La cochlée ... 22
1.2.2.1 Anatomie ... 22
1.2.2.2 Traitement des vibrations dans la cochlée ... 24
1.3 Les cellules ciliées ... 26
1.3.1 Cellules ciliées externes (CCE) ... 27
1.3.2 Cellules ciliées internes (CCI) ... 28
1.3.2.1 Morphologie et physiologie ... 28
1.3.2.2 Synapse en ruban ... 29
1.4 Les touffes ciliaires ... 29
1.4.1 Description structurelle ... 30
1.4.2 Description de la mécano-transduction ... 31
1.4.3 Développement de la touffe ciliaire ... 32
2 Maladies du système auditif et visuel ... *35*
2.1.1 Surdités ... 36
2.1.2 Surdités congénitales ... 37
2.1.3 Syndrome Usher ... 38
2.1.4 Morphologie des photorécepteurs ... 40
3 Domaines d'interaction des protéines Usher ... *43*
3.1 Domaines et modules ... 44
3.1.1 Les domaines ... 44
3.1.2 Les modules ... 45
3.1.3 Modules d'interaction et motifs linéaires ... 45
3.2 Complexité des assemblages moléculaires de la cellule ... 47
3.2.1 Régulation des voies cellulaires par les modules ... 47
3.2.2 Echafaudage et complexe multi-protéines ... 48
3.3 HHD (Harmonin Homology Domain) ... 49
3.3.1 Structure ... 49
3.3.2 Interaction ... 50
3.4 PDZ (PSD-95, Dlg1, zo1) ... 51
3.4.1 Structure ... 53
3.4.2 Interaction ... 54

3.4.2.1 Motifs canoniques des ligands peptidiques 54
3.4.2.2 Classes des couples PDZ/ligand 56
3.4.2.3 Motifs internes 58
3.4.2.4 Interactions avec les lipides 59
3.4.2.5 Dimérisation 60
3.4.3 Modulation de l'interaction des PDZ 62
3.4.3.1 Chemins énergétiques dans les domaines PDZ 62
3.4.3.2 Extensions 63
3.4.3.3 Tandems PDZ 65
3.4.3.4 Supramodules hétérotopiques 69
4 Protéines de la touffe ciliaire 71
4.1 Généralités 73
4.1.1 Etude des protéines Usher 73
4.1.2 Caractéristiques structurales 73
4.2 Protéines Usher1 74
4.2.1 Cadhérine-23 (USH1D – DFNB12) 74
4.2.2 Protocadhérine-15 (USH1F-DFNB23) 75
4.2.3 Lien apical 76
4.2.4 Myosine-7a (USH1B – DFNA11 – DFNB2) 77
4.2.5 SANS (USH1G) 78
4.2.6 Harmonine 79
4.2.7 Supramodule hétérotypique de l'Harmonine et de SANS 79
4.2.8 Point supérieur d'ancrage du lien apical 81
4.2.9 CIB2 (USH1J, DFNB48) 82
4.3 Protéines membranaires et canal mécanosensible 83
4.3.1 LHFPL5 (DFNB66/67) 83
4.3.2 TMIE (DFNB6) 83
4.3.3 TMC1 (DFNB7/11 ; DFNA36) & TMC2 83
4.4 Protéines Usher2 et protéines associées 85
4.4.1 Usherine (USH2A, RP39) 85
4.4.2 Vlgr1 (USH2C) 85
4.4.3 PDZD7 (USH modifier ; DFNB à attribuer) 86
4.4.4 Myosine 15a (DFNB3) 86
4.4.5 Whirline (USH2D, DFNB31) 87
4.4.6 Liens basaux 87
5 La Whirline, acteur clé du développement et de la structure des touffes ciliaires. 90
5.1 Topologie de la Whirline 90
5.2 Les isoformes de la Whirline et leurs expressions différentielles 93
5.2.1 L'isoforme N-terminale courte 93
5.2.2 L'isoforme C-terminale courte 94
5.2.3 L'isoforme longue 94
5.3 Les phénotypes résultant des mutations de la Whirline 96
5.3.1 DFNB31 96
5.3.2 USH2D 97
5.4 Implications connues de la Whirline dans des complexes multi-protéiques 99
5.4.1 Complexe Usher2 : Usherine / Vlgr1 / PDZD7 99
5.4.2 Complexe Eps8/Myosine15 101
5.4.3 Autres interactions connues 102
5.5 Implication de la Whirline dans d'autres types cellulaires 103
6 Problématique 106
Résultats 110

7 Résultat 1 : Délimitation et production des constructions de la Whirline ... 111
7.1 Identification de structures conservées ... 111
7.1.1 Domaines HHD ... 111
7.1.2 Extensions des PDZ ... 113
7.1.3 Délimitations et production des constructions ... 115
7.2 Production et purification du Nter-PDZ1 ... 117
8 Résultats 2 : Structure of the second HHD of Whirlin ... 121
8.1 Résumé ... 121
8.2 Article : *High Resolution structure of the HHD2 of Whirlin* ... 122
8.2.1 Introduction ... 124
8.2.2 Results ... 126
8.2.2.1 HHD containing proteins comprehensive list ... 126
8.2.2.2 The canonical fold of whirlin HHD2 ... 127
8.2.2.3 The HHD2 can form a swapped dimer ... 129
8.2.2.4 HHD2 is independent from the N-terminal supramodule of whirlin ... 132
8.2.3 Discussion ... 133
8.2.4 Materiel and methods ... 135
8.2.4.1 Protein & sample production ... 135
8.2.4.2 Small Angle X-ray Scattering ... 135
8.2.4.3 Crystallography ... 136
8.2.4.4 NMR ... 136
8.3 Commentaire et expérience supplémentaires ... 138
8.3.1 Changement de conformation ... 140
8.3.2 Tests d'interaction avec Eps8 ... 141
9 Résultat 3 : Structural characterization of whirlin reveals an unexpected and dynamic supramodule conformation of its PDZ tandem ... 143
9.1 Résumé : ... 143
9.2 Article : *^{1}H, ^{13}C and ^{15}N backbone resonance assignments and dynamic properties of the PDZ tandem of Whirlin*. Biomol. NMR Assign. (2016) 10 : 361-365 ... 144
9.2.1 Abstract ... 145
9.2.2 Biological context ... 146
9.2.3 Methods and experiments ... 147
9.2.4 Extent of assignments, secondary structure elements and dynamics ... 148
9.2.5 Acknowledgments ... 151
9.3 Article : *Structural characterization of whirlin reveals an unexpected and dynamic supramodule conformation of its PDZ tandem* ... 153
9.3.1 Abstract ... 153
9.3.2 Introduction ... 154
9.3.3 Results ... 156
9.3.3.1 The tandem PDZ1-2 of whirlin forms a supramodule ... 156
9.3.3.2 PDZ1 and PDZ2 possess hairpin extensions that are required for PDZ1-2 supramodule 157
9.3.3.3 The structural model of PDZ1-2 reveals an unexpected arrangement ... 159
9.3.3.4 The perturbations induced by PDZ1-2 mutations confirm the supramodular structural model 162
9.3.3.5 PDZ1-2 is in exchange equilibrium between closed and open conformations 165
9.3.3.6 PDZ1 hairpin folding improves its binding affinity to cellular partners ... 167
9.3.4 Discussion ... 169
9.3.5 Materiel & methods ... 172
9.3.5.1 Sample preparation ... 172
9.3.5.2 Small angle X-ray Scattering ... 172

9.3.5.3 Fluorescence ... 173
9.3.5.4 NMR measurement ... 173
9.3.5.5 Structure calculations ... 175
9.3.5.6 Structure ensemble from MD simulations ... 176
9.3.5.7 *Ab initio* prediction of residual dipolar couplings ... 176
9.3.6 Genetic algorithm ... 177
9.3.6.1 Hydrogen/Deuterium Exchange - Mass spectrometry ... 177
9.3.7 Acknowledgements ... 178
9.3.8 Supplementary figures ... 179
9.4 Commentaire & expériences complémentaires ... 188
9.4.1 Difficultés rencontrées avec l'étude des épingles-ß ... 189
9.4.2 Rôle fonctionnel du tandem ... 190
9.4.3 Rôle de la séquence inter-domaine ... 191
9.4.4 Contexte de la protéine entière ... 191
10 Résultats 4 : Interactions de la Whirline ... 193
10.1 Criblage des interactions des domaines de la Whirline ... 193
10.1.1 Résonance Plasmonique de Surface ... 193
10.1.1.1 Implémentation ... 193
10.1.1.2 Résultats ... 194
10.1.2 Hold-Up ... 197
10.1.2.1 Implémentation ... 197
10.1.2.2 Résultats ... 199
10.2 Interactions avec SANS ... 200
10.2.1 Gain d'affinité mesuré par RMN et fluorescence ... 201
10.2.2 Affinité avec Sans SAM-PBM ... 202
10.3 Interactions avec les lipides ... 204
Perspective & Discussion ... 207
11 Perspective ... 208
11.1 Les complexes de l'isoforme courte ... 209
11.2 Le complexe Usher2 ... 210
12 Discussion ... 212
Appendices ... 216
13 Appendices ... 217
13.1 Séquences de la Whirline murine isoforme 3 ... 217
13.2 Liste des résidus en échange dans le tandem PDZ1-2 ... 218
13.3 Interactants décris de la Whirline ... 219
13.4 Delhommel et al. (2015) Deciphering the unconventional peptide binding to the PDZ domain of MAST2. Biochem J. Vol :469, 159-168 ... 219
13.5 Caillet-Saguy et al. (2015) Strategies to interfere with PDZ-mediated inteactions in neurons : What we can learn from the rabies virus. *Prog. Biophys. Mol. Biol. Vol.119 (1) p53-59* ... 231
Liste des Abréviations ... 238
Liste des Figures ... 241
Liste des Tables ... 243
Bibliographie ... 244

Introduction

Préambule

Les mammifères interagissent avec leur environnement aux travers des sens. Le système auditif contenu dans l'oreille nous permet de percevoir les vibrations transmises par l'air, de localiser leurs provenances et de les traduire en un signal interprétable par le cerveau. En plus de nous permettre de profiter de la diversité sonore contenue dans notre environnement, l'audition est à la base de notre faculté à communiquer avec nos semblables. Les capacités exceptionnelles de l'oreille humaine offrent un outil puissant pour transmettre des informations précises mais aussi des idées, des concepts et des sensations, au travers l'utilisation de la parole et du langage. L'aptitude de l'homme à communiquer est à la base de notre société [1].

La transmission et la traduction des ondes sonores en un message chimio-électrique neuronal sont des processus éminemment mécaniques. Elles impliquent des morphologies précises et contrôlées à l'échelle de l'organe, de la cellule et des molécules. Le bon fonctionnement de la transduction auditive peut être aisément perturbé à tous ses degrés d'organisation. Les surdités ont des origines très variées et constituent un problème majeur de santé publique. En effet, l'Organisation Mondiale de la Santé évalue qu'environ 5% de la population mondiale est affecté par des déficiences auditives incapacitantes [2].

Le travail présenté dans cette thèse se concentre sur la Whirline, une protéine d'échafaudage cellulaire impliquée dans le maintien des structures nécessaires à la perception du son. Dans cette introduction, je décrirai le trajet de l'onde sonore depuis le milieu extérieur jusqu'aux structures cellulaires périphériques des neurones auditifs, premiers effecteurs de la stimulation nerveuse. Nous nous concentrerons ensuite sur les protéines et structures moléculaires effectrices de la mécano-transduction auditive. Nous porterons enfin une attention toute particulière aux fonctions et partenaires de notre protéine d'intérêt, la Whirline.

1 De l'onde sonore au message neuronale, les mécanismes de l'audition.

1.1 De l'onde sonore à l'oreille interne

1.1.1 Onde sonore et capacités de l'oreille

Le son est généré par tout changement de l'environnement susceptible de déplacer les molécules du milieu, la vibration d'un objet par exemple. Il consiste en une onde longitudinale propagée au travers d'un milieu élastique, typiquement l'air, dans lequel il se déplace à une vitesse de 340 m/s.

Notre oreille est capable de percevoir des variations de pression à partir de 20 µPa (0 dB), correspondant à un déplacement d'atomes du milieu de l'ordre de la dizaine de picomètre, à peine supérieur au mouvement Brownien [3]. La limite supérieure de pression est d'environ 20 Pa (120 dB). Nous percevons la pression selon une échelle logarithmique : chaque incrément en intensité perçue correspond à une augmentation d'environ 10 fois de la pression de l'onde sonore. On définit donc couramment le niveau de bruit en décibels (dB), une unité logarithmique exprimant le rapport entre deux puissances. En plus de sa large gamme d'amplitudes perceptibles, l'oreille humaine est capable de percevoir des vibrations sur une large gamme de fréquence allant de 20 Hz à 20 kHz. L'oreille est un détecteur non-linéaire, l'intensité du son perçu diffère selon les fréquences. Elle est aussi active, la perception de certaines fréquences peut être favorisée par un système de rétrocontrôle positif induit par le cerveau [4].

Dans les prochains chapitres, je décrirai comment l'oreille est capable d'avoir de telles capacités de perception.

1.1.2 Oreille externe

L'oreille externe a pour fonction principale de capturer les ondes sonores de l'environnement. Elle est formée par le pavillon et le conduit auditif (Figure 1). Le pavillon, la plus partie la plus exposée de l'oreille, est un pli cartilagineux recouvert de peau. Ses formes complexes ont pour rôle de collecter les sons, de les mitiger et de les concentrer dans le conduit auditif. Les formes du pavillon auditif ont aussi un rôle important dans la

localisation du son sur l'axe vertical [5] et favorisent les sons perçus devant la personne. Les ondes sonores concentrées traversent ensuite le conduit auditif jusqu'au tympan, une membrane vibratile marquant la limite entre l'oreille externe et l'oreille moyenne et délimitant le milieu extérieur. Le conduit auditif mesure environ 25 mm de long pour un diamètre de 7 mm, et est délimité en deux sections. La partie externe est largement recouverte de peau et contient de nombreux follicules pileux et glandes sébacées produisant le cérumen servant à la protection contre le milieu extérieur. La fin du canal est isolée par une partie osseuse conduisant peu les vibrations et permettant de concentrer celles-ci sur le tympan. La membrane tympanique est la première composante de la machinerie de transduction du son. Elle a une forme conique très similaire à la membrane d'un haut-parleur, permettant de focaliser l'énergie induite par la vibration en un point central. La membrane tympanique couvre une ouverture d'environ 10 mm de diamètre et mesure moins de 0.1 mm d'épaisseur bien qu'elle soit composée de trois tissus de différentes rigidités.

1.1.3 Oreille Moyenne

L'oreille moyenne est composée d'une large cavité emplie d'air, fonctionnant comme caisse de résonance pour la membrane tympanique, ainsi que de la trompe d'Eustache et des osselets, les trois plus petits os du squelette (Figure 1).

Les osselets : le marteau, l'enclume et l'étrier connectent la membrane tympanique à l'oreille interne et sont articulés entre eux par des ligaments. La partie centrale du cône de la membrane tympanique est connectée à la base du marteau. Sa vibration entraine un mouvement de celui-ci qui est transmis à l'enclume, puis à l'étrier. Le mouvement de ces trois os est complexe et dépend de la fréquence et de l'intensité du son perçu. En simplifiant, on peut décrire le marteau et l'enclume comme deux levier interconnectés et l'étrier comme un piston. La poussée de l'enclume, induite par le marteau, pousse alternativement l'étrier dans une cavité le connectant au milieu liquide de l'oreille interne et permet de créer une impulsion répétée. D'une part, le tympan et les osselets permettent de concentrer l'énergie de vibration, issue d'un mouvement faible d'une grande surface, en un mouvement puissant d'une surface réduite. D'autre part, ils convertissent la vibration de l'air en une vibration transmise en milieu solide, qui sera alors efficacement transmise au milieu liquide baignant l'intérieur de la cochlée. Sans cette première étape de traduction, l'énergie transmise directement depuis l'air à un milieu liquide serait principalement réfléchie. Par ailleurs, deux muscles, connectés aux osselets, se contractent dans le cas d'une vibration de trop forte amplitude, réduisant ainsi la

mobilité des osselets pour étouffer la vibration et ne pas endommager les structures de la cochlée [6].

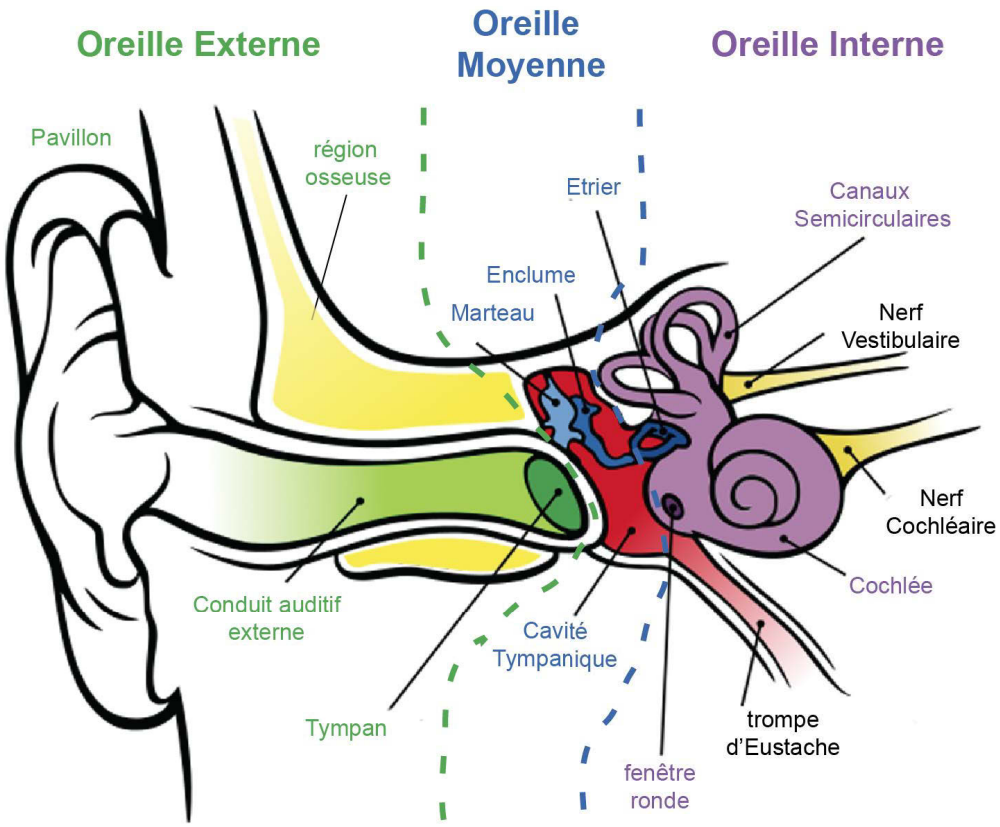


Figure 1 : Anatomie de l'oreille. Représentation schématique de l'organisation de l'oreille humaine

1.2 Oreille Interne

L'oreille interne est composée d'un labyrinthe osseux contenant l'organe de l'audition, la cochlée et l'organe de l'équilibre, le vestibule (Figure 1).

1.2.1 Le vestibule

Le système vestibulaire est un organe multi-sensoriel permettant l'analyse de la position et des mouvements de la tête dans l'espace, la stabilisation du regard et le contrôle de l'équilibre. Il est aussi impliqué dans notre capacité à naviguer dans l'espace et à s'orienter. Le vestibule possède deux structures majeures lui permettant de mesurer ces paramètres complémentaires.

1.2.1.1 Les canaux semi-circulaires

Le vestibule contient trois canaux semi-circulaires, chacun orienté perpendiculairement aux deux autres, permettant de couvrir les trois axes de l'espace (Figure 2A). A l'une des extrémités de chaque canal est localisé un renflement nommé l'ampoule, contenant les cellules sensorielles. Ces trois canaux sont capables de mesurer l'accélération angulaire de la tête le long des trois axes et sont donc impliqués dans l'équilibre dynamique. Lors d'un mouvement de tête, les canaux sont déplacés tandis que l'endolymphe liquide qu'ils contiennent reste stationnaire. L'inertie de mouvement de ce liquide crée une pression sur la cupule, masse gélatineuse contenue dans le canal au niveau de l'ampoule [7]. Dans cette cupule sont localisées des cellules sensitives équipées de touffes de stéréocils mécanosensibles (Figure 2B). La pression dévie les stéréocils dans une direction et une amplitude directement liées aux mouvements crâniens. Le mouvement des stéréocils déclenche la dépolarisation des membranes des cellules sensorielles, induisant le relargage de neurotransmetteurs du côté synaptique de la cellule et déclenchant ainsi l'influx nerveux, comme nous le verrons ensuite pour les cellules sensitives de la cochlée.

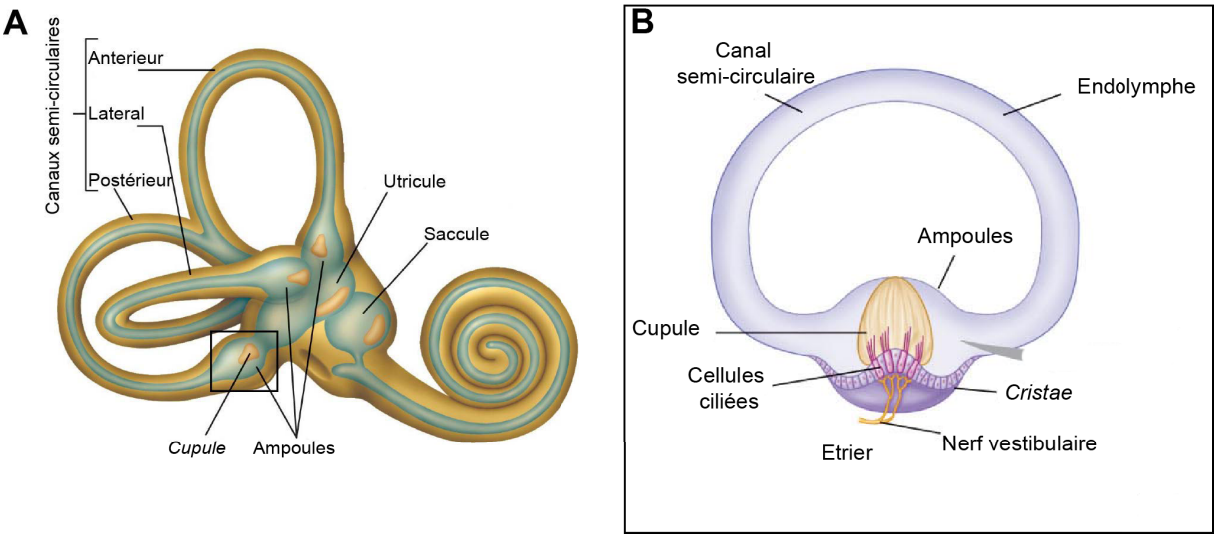


Figure 2 : Le système vestibulaire. (A) Morphologie du système vestibulaire localisé dans l'oreille interne humaine. **(B)** Structure d'une ampoule, à l'extrémité des canaux semi-circulaires, contenant les cellules sensorielles responsables de la mesure d'accélération angulaire. Figure adaptée de http://droualb.faculty.mjc.edu [8].

1.2.1.2 Les organes otolithiques

L'utricule et la saccule forment les organes otolithiques, sensibles aux accélérations latérales linéaires. Ces deux organes sont orientés de façon perpendiculaire sur l'axe vertical et sont eux aussi composés de cellules mécanosensibles. Les stéréocils de ces cellules sont baignés dans un gel extracellulaire incluant des cristaux de carbonate de calcium de haute densité comparée à celle de l'endolymphe les entourant. Lors d'un mouvement, la substance gélatineuse lourde est déplacée par la force de pesanteur, entrainant avec elle les stéréocils des cellules mécanosensibles, et déclenchant alors l'influx nerveux.

1.2.2 La cochlée

La cochlée est l'organe sensoriel de l'audition. Elle amplifie, décompose et traduit les vibrations induites par les ondes sonores.

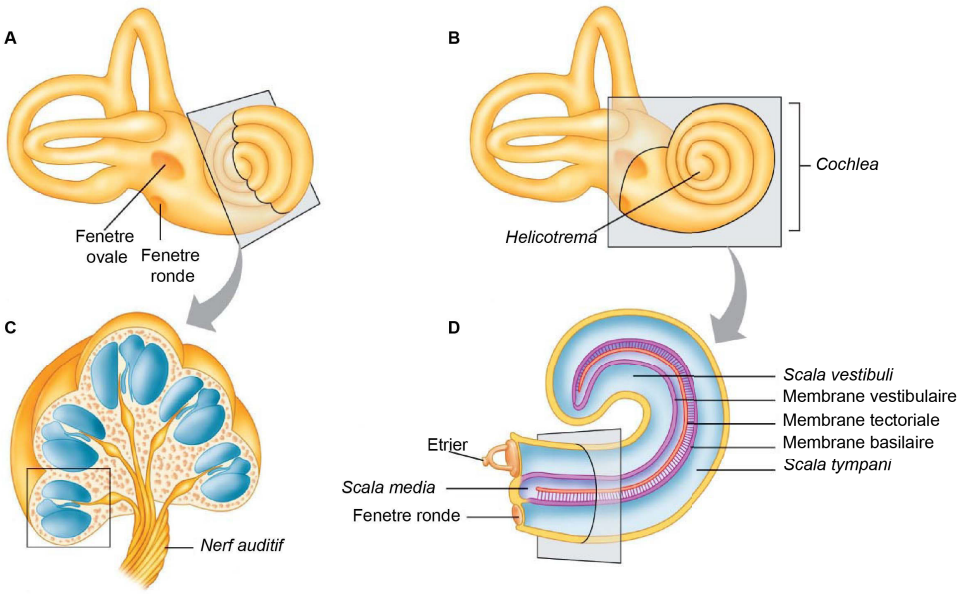


Figure 3 : Morphologie de la cochlée. (A) et (C) morphologie de la cochlée et coupe transverse de l'organe. **(B) et (D)** Coupe frontale de la cochlée

1.2.2.1 Anatomie

La cochlée est un organe osseux délimitant une cavité compartimentée, enroulée en spirale autour d'un axe central, dont la forme rappelle la coquille d'un escargot, d'où son nom (Figure 3 A, C). La cavité de la cochlée est scindée en trois canaux distincts, la rampe

vestibulaire (*scala vestibuli*), la rampe tympanique (*scala tympani*) et le canal cochléaire (*scala media*).

La rampe vestibulaire s'étend depuis la fenêtre ovale, au contact de l'étrier, jusqu'à l'extrémité apicale de la cochlée où elle se connecte à la rampe tympanique par une ouverture étroite, l'hélicotrème. La rampe tympanique suit la cochlée depuis son apex jusqu'à sa base où elle finit au niveau de la fenêtre ronde (figure 3 B, D). Ces deux rampes sont remplies de périlymphe, un liquide proche de la composition des milieux extracellulaires ($[K^+]$ = 4,2 mM, $[Na^+]$ = 145 mM, $[Ca^{2+}]$ = 1 mM).

Le canal cochléaire est enchâssé entre ces deux rampes **(**Figure 4 A, B). Il est rempli d'endolymphe riche en potassium ([K+] = 150 mM) mais pauvre en sodium ([Na+] = 1,3 mM) et en calcium ([Ca2+] = 0,03 mM). Il est séparé de la rampe vestibulaire par la membrane de Reissner (ou membrane vestibulaire), une membrane fine, permettant la diffusion de certains nutriments depuis la périlymphe. De l'autre côté, le canal cochléaire est séparé de la rampe tympanique par la membrane basilaire, sur laquelle repose la machinerie de détection des vibrations, l'organe de Corti.

L'organe de Corti est une structure répétée environ 6000 fois le long des 33 mm du canal cochléaire (*Scala media*). Chaque répétition est formée de quatre rangées de cellules ciliée mécanosensibles, complétées par des cellules de support et des cellules piliers. Les cellules ciliées sont séparées en deux groupes, les trois rangées de cellules alignées à l'extérieur de l'axe sont nommées cellules ciliées externes (CCE) et les cellules de la rangée la plus isolée, vers le centre de l'axe, sont nommées cellules ciliées internes (CCI) **(**Figure 4 C, D). Cette structure repose sur la membrane basilaire tandis qu'une membrane tectoriale rigide est au contact direct des protubérances ciliées des cellules sensorielles.

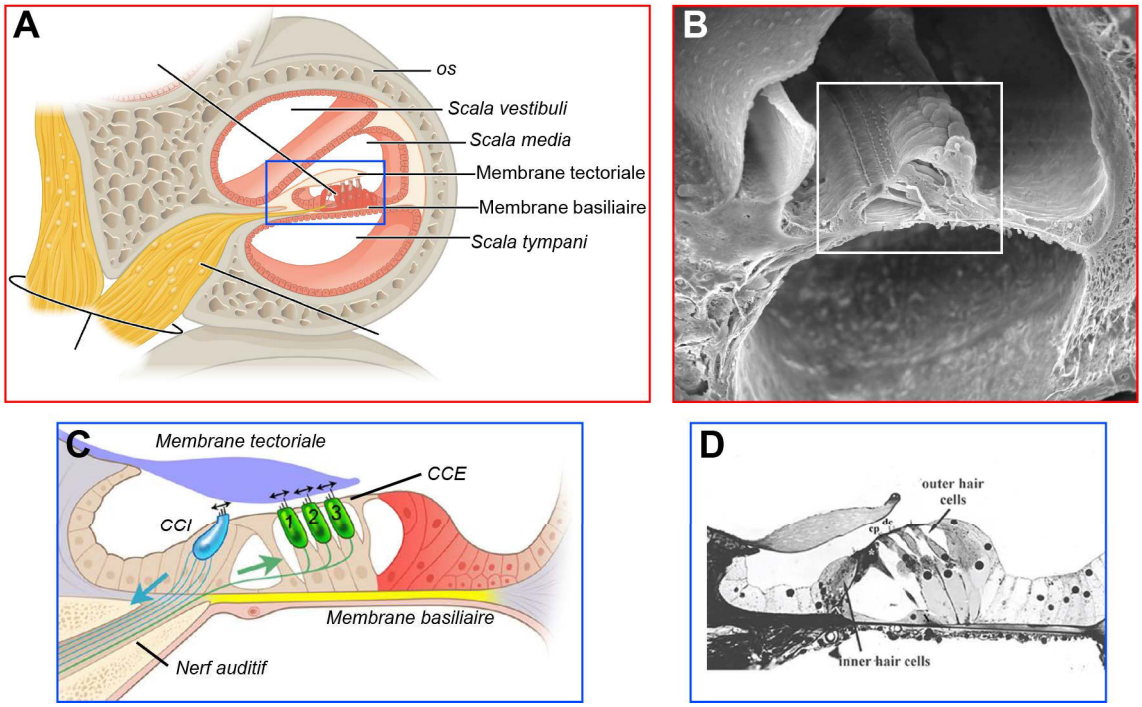


Figure 4 : Organe de Corti. (A) Schéma représentant une coupe transversale d'une section de cochlée. **(B)** Image par microscopie à balayage d'une coupe transversale d'une section de cochlée. Dans cette image, la plaque tectoriale est relevée afin de montrer la surface de l'organe de Corti (du laboratoire de B. Kachar) **(C)** Schéma représentant l'organe de Corti. **(D)** Image de microscopie optique d'une section d'organe de corti (de keele.ac.uk/emunit/gallery).

1.2.2.2 Traitement des vibrations dans la cochlée

Depuis l'oreille moyenne, la vibration est transmise par l'étrier à la rampe vestibulaire, au travers de la fenêtre ovale. Elle se propage tout au long de la rampe vestibulaire, dans le milieu liquide incompressible qu'est la périlymphe. Les parois osseuses des rampes externes limitent la diffusion de l'énergie de compression et maintiennent une force unidirectionnelle au travers la rampe vestibulaire. A l'apex de la cochlée, la compression passe au travers de l'hélicotrema, et revient vers la base de la cochlée à travers de la rampe tympanique jusqu'à la fenêtre ronde. Cette fenêtre est obturée par une membrane flexible permettant de diffuser l'énergie de l'onde, après son trajet au travers de la cochlée **(**Figure 5 A, B).

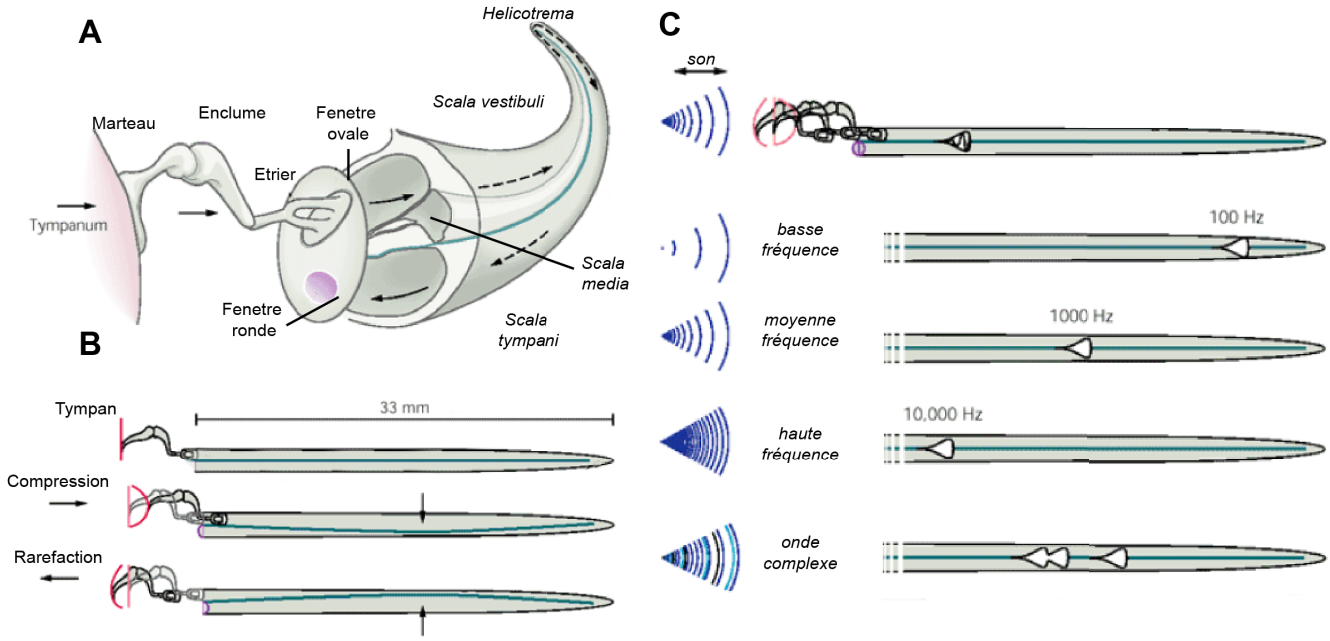


Figure 5 : traitement de l'onde sonore par la cochlée. (A) Schéma du trajet de la compression induite par l'onde sonore au travers de la cochlée. **(B)** Schéma simplifié d'une cochlée dépliée constituée de deux canaux connectés (les rampes) séparés par une membrane élastique (la membrane basilaire). Les deux schémas inférieurs montrent les effets de la propagation de l'onde propagée sur la membrane basilaire, de façon largement exagérée. **(C)** Illustration de la décomposition du son effectuée par la cochlée sur le modèle schématique simplifié. (figure adaptée de Kandel et al. [6])

L'analyse de la vibration complexe, induite par le son, commence par la décomposition des fréquences qui le composent, le long de l'axe de la cochlée. La clé de cette analyse du son provient des caractéristiques particulières de la membrane basilaire. Cette membrane sur laquelle repose l'organe de Corti est constituée de fibres tendues, parallèles et orientées perpendiculairement à l'axe de la cochlée. La composition de la membrane basiliaire varie et sa largeur diminue constamment depuis la base de la cochlée jusqu'à l'apex, diminuant sa rigidité progressivement le long de l'axe, jusqu'à atteindre une différence de largeur de cinq fois entre les deux extrémités. Chaque position de la cochlée possède ainsi une fréquence caractéristique de résonance unique, en partant des hautes fréquences (20000 Hz) à la base de l'organe, où la membrane est la plus rigide, jusqu'aux sons basses fréquences (20 Hz) à l'apex **(**Figure 5 C). Lors du trajet de l'onde le long de la membrane basilaire, les fréquences composant le son complexe résonnent à leur position spécifique, et ne progressent pas d'avantage dans la cochlée. L'amplitude de l'onde propagée le long de la membrane basilaire est proportionnelle à l'intensité de la stimulation sonore. On parle d'une carte tonotopique de la cochlée associant une position de l'axe à une fréquence de son. Sa découverte par les travaux de G. von Békésy publiés en 1960 [9] fut l'objet du Prix Nobel de médecine l'année suivante. Cette caractéristique exceptionnelle permet à l'oreille de

décomposer le son pour détecter ses fréquences séparément, le cerveau reconstituant ensuite la superposition de notes pures. La carte tonotopique de l'oreille n'est pas linéaire mais plutôt logarithmique en base dix ; ainsi la résolution en fréquence de l'oreille est variable et dépend de la plage de fréquence perçue.

La résonance d'une position spécifique de la cochlée va entrainer la vibration de l'organe de Corti à cette position. Ce mouvement va créer une force de cisaillement entre les protubérances des cellules ciliées mécanosensibles et la membrane tectoriale, permettant de détecter la stimulation à cette fréquence [10].

1.3 Les cellules ciliées

Une cochlée humaine compte environ 15000 cellules ciliées (Figure 6), 3000 internes et 12000 externes, reliées à plus de 30000 neurones. En plus de la tonotopie de la membrane basilaire, ces cellules ciliées sont toutes accordées à une fréquence spécifique. A titre de comparaison, chaque cellule ciliée a un accord en fréquence variant de 0.2% avec sa plus proche voisine, pour une différence d'environ 6% entre deux touches d'un piano [6,11]. Le long de la cochlée, les CCI détectent les vibrations et sont connectées à de nombreux nerfs afférents, alors que les CCE ont un rôle de filtre et d'amplificateur et sont innervées par des nerfs efférents.

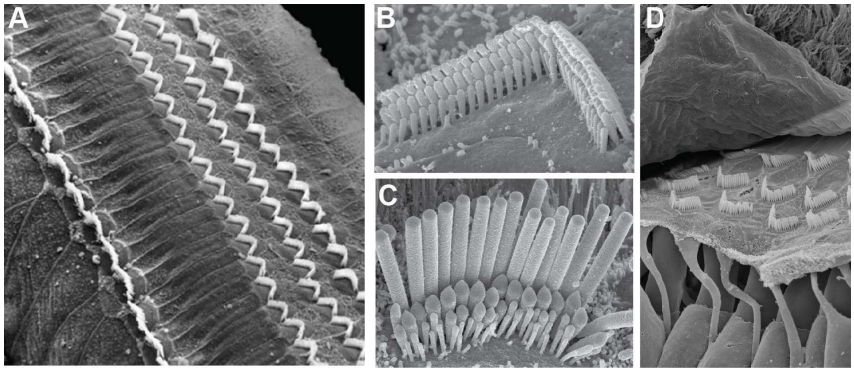


Figure 6 : Cellules ciliées. (A) image de microscopie électronique de la surface de l'organe de Corti après suppression de la plaque tectoriale. (Image de auditoryneuroscience.com) **(B)** Touffe ciliaire d'une CCE de rat observée par microscopie électronique **(C)** Touffe ciliaire d'une CCI de rat observée par microscopie électronique (images B et C adaptées de Fettiplace & Kim [12]). (**D**) Touffe ciliaire d'une CCE de souris avec la plaque tectoriale au dessus d'elles. On peut observer l'empreinte des CCE dans la plaque tectoriale. (Image de www.theoriesensorielle.com)

1.3.1 Cellules ciliées externes (CCE)

Le système auditif est capable de traiter les sons sur une large gamme d'intensité, de 20 µPa à 20 Pa. La cochlée utilise des détecteurs, les CCI, nécessitant un système d'amplification, les CCE.

Les CCE sont des cellules mécanosensibles contractiles avec un corps cellulaire cylindrique. A leur sommet, elles sont équipées de touffes ciliaires composées de trois rangées de stéréocils de taille décroissante, organisées en une forme caractéristique en "V" ou "W" **(**Figure 6 B) et dont la plus haute rangée est solidement ancrée dans la plaque tectoriale **(**Figure 6D) [13]. La taille des cellules entières diminue progressivement depuis la base de la cochlée où elles mesurent environ 75 µm, jusqu'à l'apex où elles ne mesurent plus que 10 µm [14]. Leurs touffes ciliaires ont aussi un degré variable d'ouverture de leur forme en "W" [15] le long de l'axe de la cochlée. La modification de la morphologie des CCE et de leurs touffes ciliaires semble permettre une seconde étape de sélection de fréquences, ajoutée à la tonotopie de la membrane basale, cependant il existe peu de preuves expérimentales pour confirmer cette corrélation structure/fonction [16].

Le rôle principal des CCE est d'amplifier la vibration induite par les sons de faible à moyenne intensité, à la position de résonance de la cochlée [17]. Le mouvement vers le haut de la membrane basilaire lors du passage de l'onde déplace l'organe de Corti. Ce mouvement, même faible, va pousser les touffes ciliaires des CCE contre la membrane tectoriale, créant ainsi une force de cisaillement faisant pivoter les stéréocils des touffes ciliaires vers leur plus grande rangée. Ce mouvement active le système mécanosensible de la cellule et induit la dépolarisation de sa membrane. Nous décrirons plus amplement ce mécanisme moléculaire dans la partie 1.4 de cette introduction. Les CCE sont électromotiles, la longueur de leur corps cellulaire est modifiée selon leur voltage intracellulaire. Dans les faits, les cellules se contractent d'environ 5% de leurs longueurs lors du changement de polarisation [18]. Cette contraction induit un mouvement de l'organe de Corti, et augmente la vibration dans la région stimulée [17]. Cette fonction unique des OHC est due à la Prestine, une protéine membranaire classée comme transporteur anionique voltage dépendant, surexprimée à la membrane des CCE. Bien que faisant partie d'une famille de transporteur, la Prestine n'est pas capable de transporter des ions au travers la membrane. Elle possède cependant une activité piézoélectrique : elle change sa conformation selon le voltage intracellulaire. La protéine oligomériserait à la membrane, et ferait partie d'un complexe plus large de protéines [18]. L'amplification produite par les CCE est complexe et mal comprise encore à ce jour. Il est

aussi probable que des mouvements coordonnés de sa touffe ciliaire soient impliqués dans la fonction d'amplification des CCE [19].

Les CCE ne sont quasi-exclusivement innervées que par des nerfs efférents, dont l'influx provient du cerveau. Ces neurones représentent environ 5% de l'innervation totale de la cochlée, un neurone pouvant contacter jusqu'à vingt CCE [20]. Le cerveau possède un système de rétrocontrôle capable de modifier l'électromotilité des CCE, exaltant ou inhibant l'amplification attribuée à une fréquence donnée. Cette fonction du cerveau est illustrée par notre capacité à se concentrer sur la voix de quelqu'un lors d'une discussion, même dans un environnent particulièrement bruyant. En synchronisant les sens, le cerveau est capable d'anticiper le temps d'arrivée du son et de sur-stimuler les CCE correspondant aux fréquences de la voix [10].

1.3.2 Cellules ciliées internes (CCI)

1.3.2.1 Morphologie et physiologie

Les CCI sont les véritables cellules sensorielles de la cochlée. Elles sont stimulées par la vibration induite par les CCE, ou par la vibration induite par le son, si son amplitude est suffisante. Les CCI sont globalement similaires aux CCE, leur corps cellulaire est cependant en forme de bouteille et leur orientation dans l'organe de Corti est plus horizontale.

Les touffes ciliaires sont morphologiquement différentes de celles des CCE, elles prennent une conformation en U amplement ouverte, au point d'être quasiment linéaires (Figure 6 C). La différence de taille entre la grande rangée et les deux plus petites est bien supérieure à celles dans les CCE et l'apex des stéréocils de la grande rangée n'est pas ancrée dans la membrane tectoriale [21]. Les touffes ciliaires des CCI ne sont donc pas déviées par la pression contre la membrane tectoriale, mais par le déplacement de fluide laminaire induit par la contraction des CCE. Bien que le mécanisme d'activation soit différent, la mécano-sensibilité des CCI est identique à celle des CCE. Le mouvement des stéréocils déclenche l'entrée d'ion dans la cellule entrainant la dépolarisation de la membrane plasmique. Cette dépolarisation se propage le long du corps cellulaire par des canaux cationiques voltage-dépendants, jusqu'à l'extrémité synaptique de la cellule, où elle induit l'ouverture de canaux calciques voltage-dépendants déclenchant l'exoctytose de vésicules synaptiques contenant les neurotransmetteurs. Chaque CCI est connectée à plusieurs dizaines de neurones afférents lui étant attribués. Ces neurones ont une bande passante, un seuil d'activation et une cadence

d'influx différents, permettant de retranscrire la large gamme d'amplitudes détectables par les CCI [20,22].

1.3.2.2 Synapse en ruban

Les CCI sont équipées de synapses en ruban, capables de répondre au signal variable caractéristique des cellules sensorielles. Dans ce type de synapses, une structure dense aux électrons, vue en microscopie, sert d'amarrage aux vésicules synaptiques, remplies de glutamate. Elle permet, d'une part, qu'une partie des vésicules soit au plus proche de la membrane, prêtes à une exocytose rapide. D'autre part, elle permet de délivrer les vésicules "les unes après les autres" et d'assurer un renouvellement continuel de ces vésicules. Ainsi, le ruban permet de démarrer l'influx nerveux le plus rapidement possible après stimulation et permet de maintenir l'influx dans le cas d'une stimulation longue et continue [22].

Dans les CCI de basse fréquence (jusqu'à quelques kHz), l'exocytose de glutamate est synchronisée sur plusieurs neurones avec la fréquence de la vibration stimulatrice, conservant ainsi l'information de phase dans l'influx nerveux. Le déphasage des ondes perçues par chaque oreille est essentiel pour la localisation du son dans le plan par le cerveau [22]. Chaque neurone afférent n'est associé qu'à une seule CCI et la carte tonotopique cochléaire est donc conservée dans les neurones afférents et dans le cortex auditif du système nerveux central (SNC) qui possède lui aussi une organisation tonotopique [23].

1.4 Les touffes ciliaires

Les touffes ciliaires sont les structures mécanosensibles des cellules ciliées internes et externes, capables de traduire une force mécanique en un message électrochimique transférable et interprétable par le système nerveux. Leurs structures et leurs fonctions sont similaires dans les CCE et dans les CCI, nous ne différencierons donc plus ces deux types cellulaires dans cette partie.

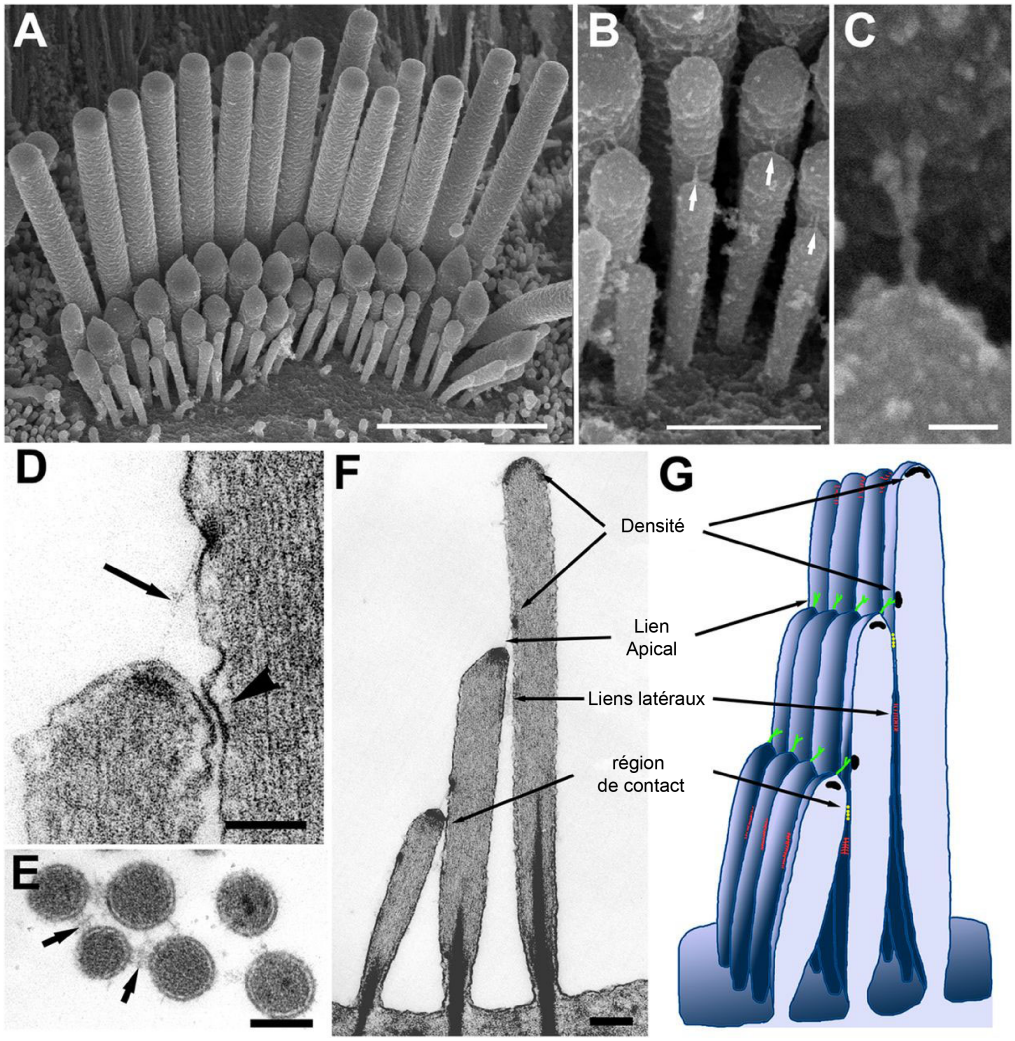


Figure 7 : Structures des touffes ciliaires. (A) Image par microscopie électronique à balayage d'une touffe ciliaire de CCI de rat. **(B)** Image par microscopie électronique à balayage montrant les liens inter-rangées de stéréocils. **(C)** Agrandissement du lien apical. **(D)** Image de microscopie électronique à transmission d'une section horizontale de touffe ciliaire de rat, montrant le lien apical et les densités cytoplasmiques. **(E)** Image de microscopie électronique à transmission d'une coupe horizontale de touffe ciliaire de rat montrant le cœur d'actine des stéréocils et les liens entre stéréocils. **(F)** Autre vue d'une touffe ciliaire de cochon d'inde montrant les trois rangées. **(G)** schéma récapitulatif des liens inter-stéréocils. Figure adaptée de Hackney & Furness [24].

1.4.1 Description structurelle

Les touffes ciliaires sont composées de stéréocils, des microvillosités d'actine rigides (Figure 7)[15]. Les stéréocils ont un diamètre entre 200 et 500 nm, variant selon la position le long de la cochlée. La tige, partie centrale des stéréocils, est constituée d'une région dense de plus de 3000 filaments d'actine parallèles, massivement interconnectés et structurés en motifs hexagonaux répétés (Figure 7 E) [25]. Cette région est exceptionnellement stable et rigide [26]. A leur base, les stéréocils sont ancrés dans la plaque réticulaire, une région rigide du corps cellulaire riche en actine. Peu avant le point d'insertion, le diamètre du stéréocil diminue

drastiquement jusqu'à ne plus contenir que quelques dizaines de fibres d'actine, qui plongent profondément dans le corps cellulaire et ancrent le stéréocil (Figure 7 F). Cette région de moindre rigidité sert de pivot aux stéréocils lors de l'inflexion des touffes ciliaires, le reste de la structure restant parfaitement linéaire et rigide.

La cohésion de la touffe de stéréocils est maintenue par une multitude de liens extra-cellulaires protéiques connectant entre eux les stéréocils [24] (Figure 7 F G). Ce réseau extracellulaire complexe donne un mouvement cohésif de la touffe ciliaire lors de l'inflexion et permet l'activation simultanée de la machinerie mécanosensible de tous les stéréocils [24].

Les filaments d'actine s'étendent jusqu'au sommet des stéréocils, à quelques nanomètres de la membrane apicale (Figure 7 D) [26]. Au sommet des stéréocils des rangées inférieures, une densité électronique est présente, ainsi qu'un lien protéique le reliant au côté du stéréocil le plus proche de la rangée supérieure, dans lequel une densité électronique est aussi visible. On parle de lien apical et de points d'ancrage inférieur et supérieur. Comme en témoigne la forme de la membrane au niveau du point d'insertion inférieur, ce lien apical extracellulaire de 185 nm est constamment sous tension, et constitue l'élément sensitif de la mécano-transduction (Figure 7 B C D) [27].

1.4.2 Description de la mécano-transduction

Lors de la déflection des stéréocils vers la rangée de plus grande taille, la tension au niveau du lien apical est augmentée. Cette tension est propagée à un complexe de protéine membranaire couplé à un canal cationique. L'augmentation de tension induit un changement de conformation de ce canal conduisant à son ouverture (Figure 8). Chaque stéréocil des rangées inférieures possède de un à deux canaux, proches du point d'ancrage inférieur du lien apical. La taille, le flux ionique et l'ensemble de leurs caractéristiques biophysiques sont identifiés. Cependant les protéines composant le canal de mécano-transduction sont toujours inconnues [28]. Ce système est extrêmement sensible au déplacement des stéréocils, les modélisations permettent d'estimer que l'ouverture du canal peut être déclenchée par des mouvements sub-nanométriques de l'apex des stéréocils [29].

L'ouverture du canal déclenche l'entrée indifférenciée d'ion cationique dans la cellule, principalement K^+ et Ca^{2+} de part la composition du milieu. Après l'ouverture initiale, le complexe s'adapte pour conserver sa sensibilité. Il est suspecté que les ions calcium se lient à la partie cytoplasmique du canal dans sa conformation ouverte et qu'ils induiraient son retour dans sa conformation fermée et inactivable après environ 20 µs. Après la fermeture du canal,

la concentration calcique intracellulaire diminue et induirait le retour du canal dans son état initial et activable après quelques millisecondes [28]. Un autre mécanisme d'adaptation lente a lieu après quelques dizaines de millisecondes, réduisant la tension sur le câble apical. Ce mécanisme serait contrôlé activement au niveau du point d'ancrage supérieur du stéréocil. La tension maintenue sur le lien apical déclencherait la descente des myosines le long du filament d'actine, entrainant l'ensemble de la machinerie vers le bas. Ce mouvement réduit de ce fait la tension du lien et permet la fermeture du canal [30], le rendant de nouveau apte à recevoir une stimulation de plus forte amplitude.

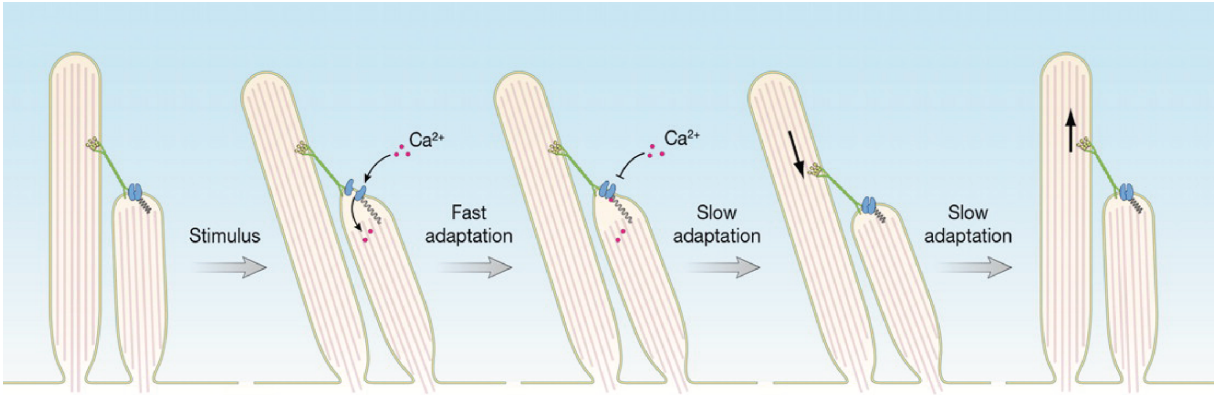


Figure 8 : Mécanisme de la mécano-transduction. Schéma du fonctionnement de la mécano-transduction. Figure de Schwander et al.[30]

1.4.3 Développement de la touffe ciliaire

Au cours du stade embryonnaire, de courtes villosités commencent à peupler la surface des cellules ciliées. Au centre de l'apex se dresse un unique cil constitué d'un cytosquelette microtubulaire, le kinocil. Pendant le développement, le kinocil migre à la périphérie de la cellule, où il servira alors de guide pour les microvillosités à sa proximité, qui s'allongent et formeront la grande rangée. Les microvillosités proches des stéréocils déjà formés vont elles-mêmes s'agrandir, guidées par cette rangée de taille supérieure. Ces étapes successives conduisent à l'organisation en marche d'escalier caractéristique des touffes ciliaires. Le kinocil sert donc de guide à l'élongation des stéréocils et donnera la polarisation planaire de la touffe ciliaire (Figure 9 A). Les villosités non comprises dans la touffe ciliaire en développement se résorberont. S'ensuit une étape d'élargissement des stéréocils par addition de brin d'actine, et d'enracinement dans le corps cellulaires *via* les fibres centrales du stéréocil. Au cours de cette étape, le kinocil se résorbera lui aussi, laissant place à la touffe ciliaire mature. La vitesse de maturation des rangées de cellules ciliées dépend de leur

position dans l'organe de Corti : plus celles-ci sont internes, plus elle arriveront à maturité rapidement [31,32].

Pour maintenir la cohésion de la touffe ciliaire en développement, une variété de liens interstéréocils se forme transitoirement (Figure 9 B) [24]. A partir du stade embryonnaire, des liens latéraux mal organisés se forment et connectent les stéréocils entre eux de façon peu discriminative. Les stéréocils de la plus grande rangée interagissent avec le kinocil par des liens plus organisés, concentrés à l'apex. Au stade post-natal, les liens inter-stéréocils se regroupent de façon discrète. Des liens latéraux migreront notamment jusqu'à l'apex où certains formeront le lien apical, et d'autres deviendront les connecteurs supérieurs, reliant stéréocils d'une même rangée entre eux et avec les stéréocils de la plus grande rangée. Quelques liens resteront le long du tronc des stéréocils, et des liens particuliers, dits basaux, apparaitront de façon transitoire à la base des stéréocils. Ces liens sont indispensables à la forme en "V" ou "W" des touffes ciliaires [33]. Au stade mature, seuls les liens apicaux et les liens supérieurs resteront en place. Les liens apicaux font partie intégrante de la machinerie de mécano-transduction, et les liens supérieurs assurent la cohésion de l'ensemble des stéréocils de la touffe ciliaire, permettant leur mouvement coordonné lors de la déflection [24].

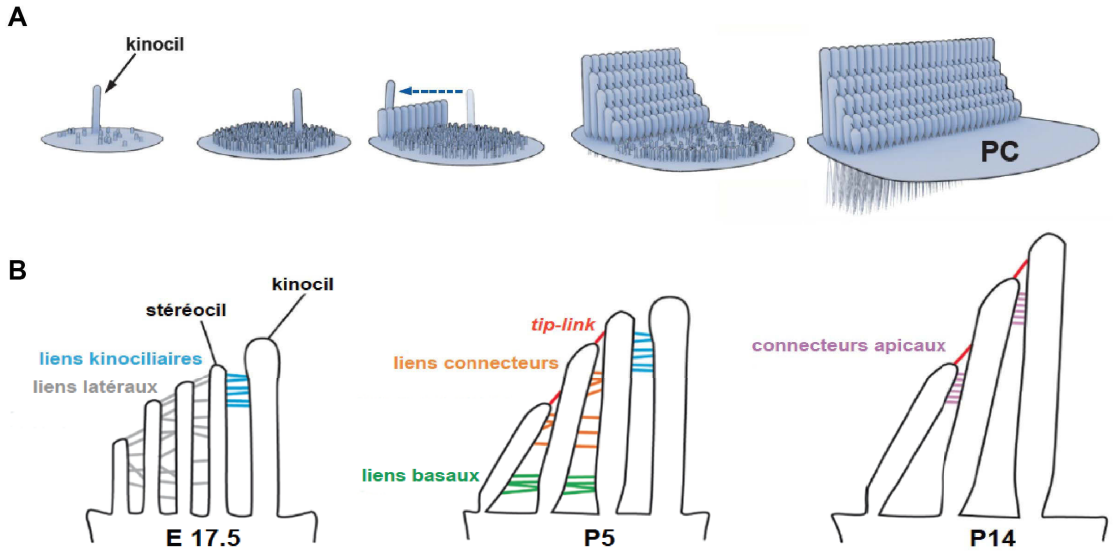


Figure 9 : Développement de la touffe ciliaire. A Représentation schématique des différentes étapes du développement de la touffe ciliaire, chez les oiseaux. Figure adaptée de Schwander et al. [30] **B** Schéma des différents liens inter-villosités de CCE de souris, à plusieurs stades de développement : E17.5 stade embryonnaire ; P5 stade post-natal ; P14 stade mature. Figure adaptée de Michalski & Petit [32].

La mise en place de la morphologie de la touffe ciliaire nécessite l'expression contrôlée, dans l'espace et dans le temps, des différentes isoformes de plus de 50 protéines [34]. Certaines, comme celles constituant les liens basaux, ne seront exprimées que pour une très courte période et ne persisteront pas au stade mature. D'autres seront présentes tout le long du développement mais changeront plusieurs fois de localisation. Elles formeront des complexes transitoires à différentes localisations des stéréocils, extrêmement encombrés par le réseau dense d'actine, l'ensemble étant orchestré par un groupe de myosines atypiques assurant la localisation correcte de ces complexes.

2 Maladies du système auditif et visuel

2.1.1 Surdités

La surdité est le défaut sensoriel le plus répandu chez l'homme. Elle affecte près de 360 millions de personnes dans le monde [2] . La majorité des personnes atteintes de surdité handicapante sont âgées de plus de 65 ans [35], cependant environ 10 % des surdités concernent des individus de moins de 15 ans. Une surdité est définie comme handicapante à partir d'une perte supérieure à 40 dB à la fréquence de moins bonne audition. Les surdités sont classées en quatre catégories selon le degré de perte d'audition à la fréquence de moins bonne audition [35]:

- Légère pour des pertes de 20 à 40 dB.
- Modérée pour des pertes de 41 à 70 dB.
- Sévère pour des pertes de 71 à 90 dB.
- Profonde pour des pertes de plus de 91 dB.

Les surdités peuvent être causées par différentes déficiences [36]:

- Surdités de transmission : Elles correspondent à des pathologies de l'oreille externe ou moyenne, affectant le transfert mécanique de l'onde sonore jusqu'à l'oreille interne. Généralement, le déficit est modéré et affecte préférentiellement la transmission des ondes de basses fréquences. Elles sont généralement causées par des infections du conduit auditif, du tympan ou de l'oreille moyenne, ou par des problèmes au niveau des osselets, malformations ou séquelles suite à un trauma crânien.
- Surdités neurosensorielles : Elles concernent la grande majorité des cas (>90%). Dans ce type de surdité, l'onde atteint la cochlée, mais l'information n'atteint pas le cortex cérébral. Elles sont le plus souvent dues à des dommages de l'organe de Corti ou du nerf auditif, mais elles peuvent aussi être causées par des problèmes au niveau du cortex auditif. Il existe trois causes acquises majeures de perte d'audition. Le traumatisme acoustique : l'exposition à un son de très forte intensité (>120dB) peut endommager les cellules ciliées et induire une exocytose importante de glutamate, cytotoxique à forte concentration pour la synapse. La prise de certains médicaments peut induire des pertes d'audition, notamment des anti-tumoraux et des antibiotiques. Enfin les cellules ciliées et l'ensemble de la cochlée accumulent des défauts avec les années, causant à terme des pertes variables d'audition.

2.1.2 Surdités congénitales

Les surdités congénitales sont définies par leur présence dès la naissance, ou par extension, par la présence de leurs causes à la naissance, pour des raisons génétiques ou environnementales. Dans les pays développés, leur prévalence à la naissance est de 1.33 ‰, pour une déficience supérieure à 30 dB et progresse jusqu'à 3.5 ‰ en considérant les enfants de moins de 15 ans. Leur prévalence est bien supérieure dans les pays moins développés et semble être corrélée exponentiellement au revenu national brut par habitant du pays. Elle peut ainsi atteindre plus de 20 ‰ dans les régions les plus touchées, notamment en Asie du sud [35,37]. Environ la moitié des surdités provient de l'environnent et ont été acquises pendant la grossesse (Figure 10). Les causes les plus communes sont les infections par un virus : cytomégalovirus, virus de la Rubéole, virus Zika, ou la prise de médicaments ototoxiques.

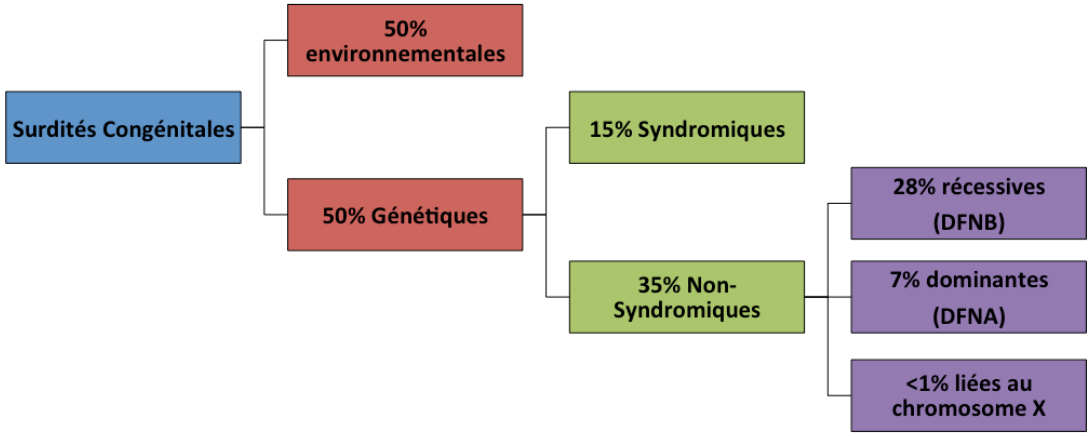


Figure 10 : Causes de surdités congénitales. Graphique hiérarchique des causes de surdités congénitales.

L'autre moitié des surdités congénitales est due à une cause génétique, le plus souvent par la mutation d'un seul gène. Avec le développement des grands projets génomiques humains, des avancées considérables ont été faites dans l'identification des causes génétiques de surdité et par extension dans la compréhension des mécanismes moléculaires de l'audition. A ce jour, des mutations à 120 *loci*, sur 92 gènes, ont été identifiées comme responsables de surdité. On distingue deux types de surdité génétique, celles non-syndromiques, c'est à dire dont le seul symptôme est la perte d'audition, et celles syndromiques, si d'autres symptômes sont associés à la surdité.

Les surdités non-syndromiques représentent la majorité des cas. Selon leur mode de transmission, elles sont autosomiques dominantes (DFNA) : 60 *loci* et 36 gènes identifiés ; autosomiques récessives (DFNB) : 84 *loci et* 66 gènes identifiés ; ou liées au chromosome sexuel X (DFNX): 6 *loci* et 5 gènes ; et mitochondriaux : 2 gènes [38]. Les *loci* et par extension

les gènes identifiés, sont généralement nommés selon leur mode de transmission, ils sont ensuite numérotés chronologiquement selon la date de leur découverte (*e.g.* gène de la Prestine : DFNB61).

Les surdités syndromiques sont très variées, les symptômes supplémentaires peuvent toucher des systèmes divers: la vision, l'équilibre, les systèmes musculo-squelettique, cardio-vasculaire, urogénital, nerveux, endocrinien ou encore digestif. On dénombre plus de 300 syndromes de surdité [39]. Il n'est pas rare qu'un même gène soit associé à différents syndromes, ou à une forme non-syndromique de surdité, selon la position et le type des mutations causatives.

2.1.3 Syndrome Usher

Le syndrome Usher est une forme syndromique de surdité, associée à une perte progressive de la vision et potentiellement à des troubles de l'équilibre. L'association de ces symptômes est décrite depuis le 19^{ième} siècle, mais il faudra attendre le siècle suivant pour que le syndrome soit clairement identifié par un ophtalmologiste Ecossais, Charles Usher qui donnera son nom au syndrome. Le syndrome Usher est la première cause de surdité-cécité, représentant environ 50% des cas et responsable de 3 à 6 % des surdités de l'enfant [40]. L'évaluation de la prévalence du syndrome Usher est difficile, notamment à cause du nombre important d'erreurs de diagnostique, les symptômes visuels apparaissant à l'adolescence, après les symptômes auditifs. Les premières études rapportent une prévalence d'environ 1/25000. Cependant de nouvelles estimations comptabilisant les patients avec des symptômes moins marqués ou non-diagnostiqués proposent une prévalence atteignant 4.25/25000 [41]. A l'enfance, le syndrome Usher est identique à une surdité non-syndromique, son diagnostique est donc difficile et nécessite le séquençage exonique des gènes en cause. Même après séquençage, le diagnostique peut être difficile, les gènes responsables du syndrome pouvant aussi être mutés dans des surdités non-syndromiques. Avant le développement des méthodes de diagnostique moléculaire, de nombreux individus ont été diagnostiqués de façon erronée et n'ont pas été préparés à la perte d'un sens supplémentaire. Un patient ayant appris le langage des signes perdra ainsi son moyen de communication principale lors de l'apparition des symptômes visuels.

Syndrome	Proportion	Surdité	Début de la Rétinite Pigmentaire	Troubles de l'équilibre
Usher 1	≃ 40%	Profonde, congénitale	Pré-pubertaire	Sévères
Usher 2	≃ 60%	Modérée à sévère, congénitale	Post-pubertaire	Absents
Usher 3	< 3%	Modérée et progressive	Variable	Variables

Table 1 : Caractéristiques cliniques des syndromes Usher. Adapté de El-Amraoui et Petit [41].

Le syndrome Usher est cliniquement hétérogène et comprend trois types selon la gravité des symptômes et les organes affectés (Table 1). La forme Usher type 1 est la plus sévère ; elle implique une surdité bilatérale profonde à la naissance, associée à un dysfonctionnement vestibulaire, responsable de troubles de l'équilibre important dès la naissance. La surdité profonde ne permet pas aux patients d'apprendre l'usage de la parole. Les patients développeront ensuite, aux alentours de 17 ans, une rétinite pigmentaire, une dégénérescence de la rétine commençant à sa périphérie et progresse vers son centre, réduisant progressivement l'angle de vision jusqu'à une cécité complète à un âge tardif [40] . Le syndrome Usher type 2 est la forme la plus fréquente de la maladie. Les patients naissent avec une surdité bilatérale moyenne à sévère. La rétinite pigmentaire est similaire à celle observée chez les patients atteints du syndrome de type 1 mais apparait plus tardivement. Les patients atteints du type 2 peuvent apprendre à parler et la gravité des symptômes est plus variable que dans le type 1 [42]. Le syndrome Usher type 3 est beaucoup moins courant. Il est associé aux symptômes les moins sévères et les plus variables. Les patients naissent avec une audition normale qui progresse vers une surdité modérée à sévère avant leur 30^{ième} année, conjointement avec l'apparition de symptômes visuels liés à la rétinite pigmentaire [41].

Les symptômes auditifs des syndromes Usher sont bien contrecarrés par le port d'un implant cochléaire. Ce dispositif permet de shunter les oreilles externe, moyenne et interne, en utilisant un microphone externe connecté à une électrode flexible, introduite, par la fenêtre ronde, dans la *scala tympani* et disposée le long de l'axe tonotopique de la cochlée. Le microphone transmettra l'information auditive par stimulation électrique aux positions tonotopiques des fréquences perçues, stimulant directement les neurones afférents sans utiliser l'organe de Corti. Ce système très efficace a cependant des limitations : il rend difficile la localisation des sons, et il ne permet pas d'amplification active, comme celle générée par les cellules ciliées externes, rendant l'audition difficile dans un milieu bruyant.

La classification des types de syndrome Usher est basée sur les caractéristiques cliniques des patients mais les gènes causatifs sont clairement associés à un unique type

(Table 2). Le syndrome Usher type 1 est causé par des mutations sur un des six gènes codant pour les protéines: Myosine7a, Harmonine, Cadhérine-23, Protocadhérine-15, SANS et CIB2. Le syndrome Usher type 2 implique des mutations des gènes associés à trois protéines, Usherine, Vlgr1 et Whirline. Enfin le syndrome Usher type 3 est causé par des mutations de la Clarine-1.

Type	*Locus*	Protéine	Surdité ou cécité non-syndromatique	Modèle murin
Usher type 1	USH1B	Myosine 7a	DFNB2 – DFNA11	*Shaker-1*
	USH1C	Harmonine	DFNB18A	*Deaf circler*
	USH1D	Cadhérine 23	DFNB12	*Waltzer*
	USH1F	Protocadhérine 15	DFNB23	*Ames*
	USH1G	SANS	-	*Jackson shaker*
	USH1J	CIB2	DFNB48	-
Usher type 2	USH2A	Usherine	RP39	$Ush2a^{-/-}$
	USH2C	Vlgr1	-	*Frings*
	USH2D	Whirline	DFNB31	*Whirler*
Usher type 3	USH3A	Clarine 1	RP61	$CLRN1^{N48K}$

Table 2 : protéines associées aux différents types de syndrome Usher.

Ces protéines, exprimées dans les photorécepteurs ou cellules ciliées, forment des complexes associés aux types de syndrome. L'étude de ces protéines est essentielle pour la compréhension des mécanismes moléculaires à la base de la vision et de l'audition. Les modèles murins ont été décisifs dans l'identification des protéines Usher mais aussi des autres protéines affectant l'audition. Les souris ont un système cochléaire extrêmement proche de celui de l'humain, avec des protéines effectrices très bien conservées. Le système rétinien des souris est malheureusement plus éloigné de l'homme, probablement due à la différence de rythme journalier des deux espèces. L'étude du rôle des protéines Usher dans la vision est donc plus difficile chez les souris.

Dans le prochain chapitre, nous décrivons en détail les fonctions des protéines Usher dans les stéréocils des touffes ciliées.

2.1.4 Morphologie des photorécepteurs

Bien que les cellules ciliées et les cellules photoréceptrices soient toutes deux des neurones sensitifs, leurs fonctionnements et leurs morphologies sont drastiquement différents. Dans cette partie, nous présentons brièvement la morphologie des photorécepteurs.

La rétine est une couche innervée du fond de l'œil (Figure 11 A). Elle contient les cellules photoréceptrices capables d'absorber l'énergie des photons et de la convertir en un message chimio-électrique interprétable par le cerveau [43]. La rétine possède deux types de cellules ; les bâtonnets, utilisés en condition de pénombre et composant moins de 6% des photorécepteurs. Le reste étant composé des cônes, utilisés en période diurne. Ces deux types cellulaires ont cependant des morphologies proches.

Au plus proche de la surface de la rétine est localisé le segment externe, composé de centaines de disques de membranes empilées. Ces membranes sont saturées en pigment, des oligomères de rhodopsine et sont entièrement renouvelées tous les 10 jours, afin d'éviter l'accumulation d'éléments oxydatifs, nécessaires à la signalisation de la vision mais toxiques pour les cellules [44]. Ce renouvellement constant est permis par le segment interne, contenant des mitochondries et l'appareil de Golgi qui produit constamment des nouvelles vésicules contenant la machinerie photosensible. Les deux segments ne sont connectés que par le cil connecteur, une région fine (300 nm) apparentée à un cil microtubulaire non-motile. Cette région est dense en trafic vésiculaire mais aussi en diffusion de protéine soluble. Chez l'homme, une couronne de microvillosités basées sur l'actine, les calyceal process, entoure la région du cil connecteur [45] (Figure 11 B C). Cette région pourrait être apparentée aux stéréocils, mais très peu d'information existe quant à sa fonction. Le reste de la cellule est composé du corps cellulaire contenant le noyau et de la région synaptique équipée d'une synapse à ruban au contact des neurones afférents.

Les cellules ciliées et les photorécepteurs nécessitent donc toutes deux une morphologie très particulière pour effectuer leur fonction. Dans les deux cas, cette morphologie repose sur un ensemble de cils et de stéréocils. L'étude de l'implication des protéines Usher dans la vision est difficile, car il n'y a pas de bon modèle animal des cellules photoréceptrices humaines. Cependant, il semble maintenant clair que les protéines Usher sont impliquées dans le maintien de la morphologie de la région du cil connecteur ; au niveau des calyceal process pour les protéines Usher1, et au niveau de la membrane adjacente au cil connecteur pour les protéines Usher2. Leurs mutations induisent une déstabilisation de la morphologie du cil connecteur, perturbant à terme le bon renouvellement des disques membranaires du segment externe. L'accumulation de produits oxydants cytotoxiques induit progressivement la mort des cellules photoréceptrices [41,46]. Ces dernières sont également équipées de synapses glutamiques à ruban similaires à celles des cellules ciliées cochléaires [47]. Au moins une isoforme de chaque protéine Usher est exprimée au niveau des synapses à

ruban des cellules ciliés et photoréceptrices. Il est donc probable que les mutations des protéines Usher aient aussi un impact sur le fonctionnement de ces synapses [41].

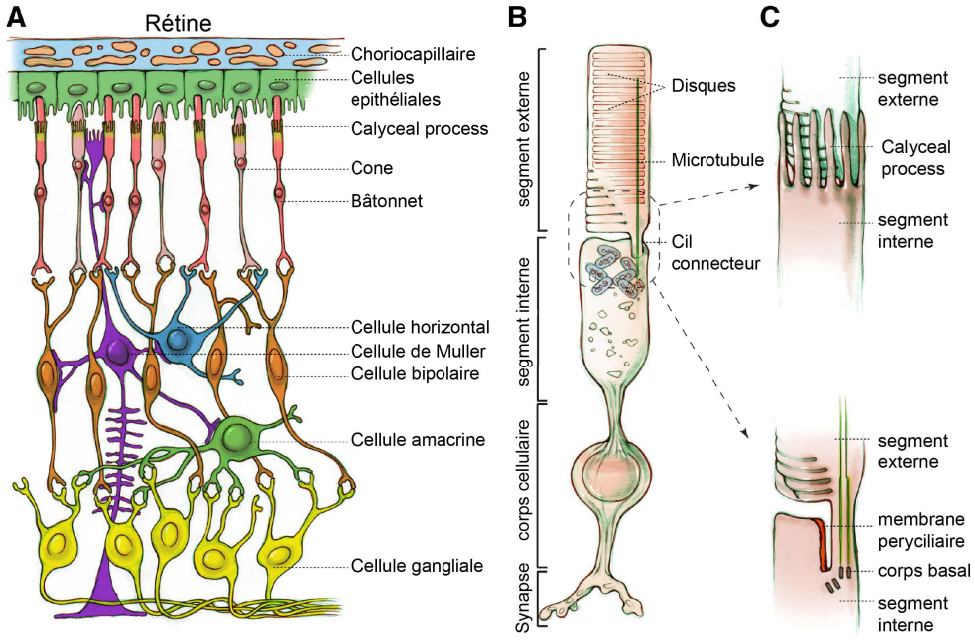


Figure 11 : Morphologie des photorécepteurs humains. A Schéma de l'organisation cellulaire de la rétine. **B** morphologie d'un bâtonnet humain. **C** Agrandissement de la région du cil connecteur, avec et sans les calyceal process. Figure adaptée de Mathur et Yang [46].

3 Domaines d'interaction des protéines Usher

3.1 Domaines et modules

3.1.1 Les domaines

La première apparition dans la littérature du terme domaine date de la fin des années 60 [48]. Elle émane de l'étude structurale à basse résolution des chaines lourdes d'anticorps, par l'équipe du Docteur Edelman. Les données collectées par microscopie électronique, par diffusion des rayons X aux petits angles et par sédimentation, indiquaient que de par leur volume, les anticorps se replient nécessairement en domaines globulaires reliés par des régions plus étendues et flexibles. Cette première utilisation du terme est donc basée sur les caractéristiques structurales d'une région incluse dans le contexte d'une séquence protéique plus longue.

Le terme domaine est maintenant largement utilisé mais il se réfère à une variété d'objets librement définis selon des critères variables:

- Une entité globulaire capable de se replier de façon indépendante du reste de la séquence polypeptidique [48].

- Une région possédant une séquence continue, identifiable car similaire aux domaines de la même famille [49].

- Une entité d'unité fonctionnelle définie et partagée avec les autres membres de la famille.

Ces critères sont à priori liés, mais peuvent parfois être en contradiction et ne sont donc pas considérés de façon stricte. Les domaines kinases sont par exemple formés de deux lobes de repliement indépendant, mais seul leur assemblage est fonctionnel. Au contraire, certains membres d'une famille de domaine ne possèdent pas la fonction canonique associée à leur famille [50]. Le critère de séquence peut aussi être remis en cause par l'existence, dans une famille, de domaines séparés en deux séquences non-continues mais formant une entité globulaire unique. C'est le cas de deux domaines partiels PH de l'α-syntrophine, séparés séquentiellement par une région contenant un domaine PDZ [51].

Plus de 90% des familles de repliements comprennent entre 20 et 200 acides aminés, avec une taille moyenne autour de 100 résidus. Cette limitation dans la taille des domaines est

en partie liée à la vitesse de repliement induite par la compaction des régions hydrophobes des domaines globulaires [52].

3.1.2 Les modules

Le terme de module provient du monde de l'ingénierie ; il définit une unité standardisée, de fonction indépendante et pouvant être intégrée dans une construction plus complexe. En biologie, les modules sont un sous-ensemble de domaines définis comme des éléments structuralement distincts, utilisés fréquemment comme blocs de construction dans des protéines de fonctions diverses [53]. La définition de module est intimement liée à l'identification des exons dans le génome et à la notion de mobilité génétique [54]. Les modules étaient alors définis comme des entités indépendantes d'évolution, souvent incluses dans un seul exon, et capables de s'intégrer rapidement dans un nouveau gène au cours de l'évolution. Cette définition a cependant évolué dans les années 90, comme nous allons le présenter dans le paragraphe suivant.

3.1.3 Modules d'interaction et motifs linéaires

Dans les années suivant leur définition, les modules ont été particulièrement étudiés dans le contexte de la signalisation cellulaire, notamment par les équipes des Drs. T. Pawson et W. Lim. Leurs travaux se concentrent sur la variété des arrangements de modules et leurs capacités à modifier la fonction des protéines les contenant par le recrutement de protéines partenaires. Ils définissent alors une nouvelle sous-catégorie, les modules d'interaction ; ceux-ci sont des domaines non-catalytiques interagissant avec des éléments caractéristiques et communs à une famille de module [55]. Les modules d'interaction reconnaissent des phospholipides, des acides nucléiques, ou plus communément, une séquence peptidique courte pouvant porter une modification post-traductionnelle (Figure 12). Cependant, chaque membre d'une famille de module peut avoir des interactions additionnelles avec des motifs non canoniques. De manière générale, les modules sont plus courts que les domaines catalytiques, ils sont définis comme possédant des extrémités N-terminale et C-terminale rapprochées et un site actif éloigné de celles-ci, facilitant ainsi leur insertion dans des chaines polypeptidiques de composition variée. Il existe cependant de nombreux contre-exemples à ces caractéristiques.

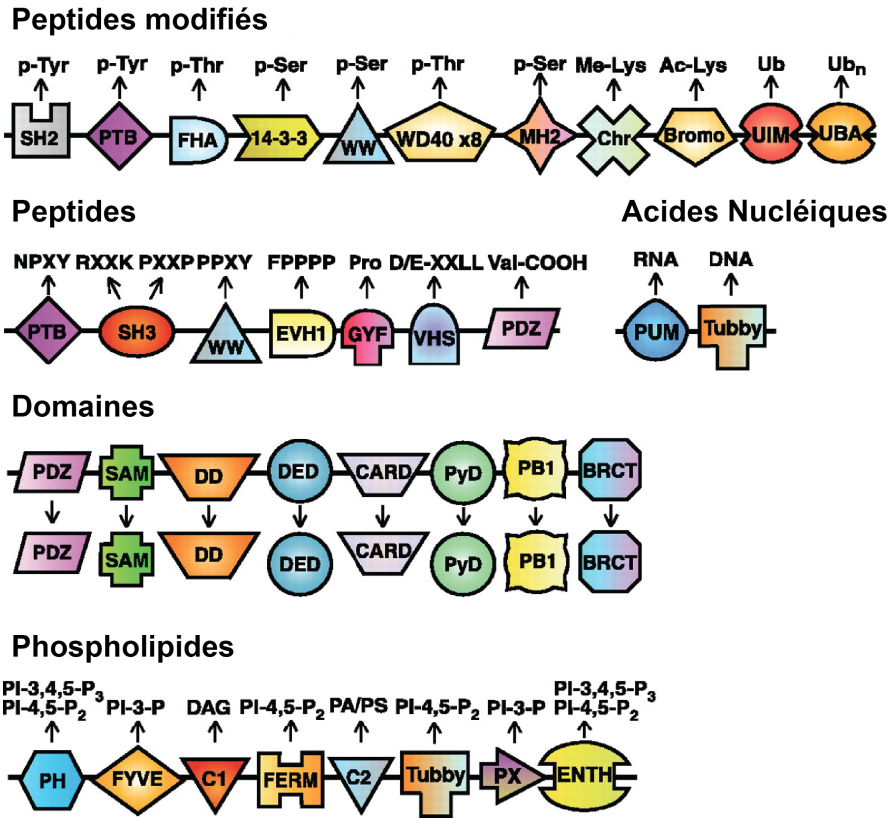


Figure 12 : Les modules d'interaction et leurs motifs associés. Adaptée de Pawson & Nash [55].

La majorité des modules interagissent avec des motifs linéaires d'interaction. Il s'agit de courtes séquences spécifiques, de 2 à 11 résidus, appelés SLiM (Short Linear Motifs) [56] et localisées dans des régions non-structurées. Les interactions modules/motifs impliquent des surfaces d'interaction peu étendues produisant des affinités faibles, dans la gamme du micromolaire. Par définition, les motifs possèdent certaines positions conservées, essentielles pour l'interaction avec leur famille de modules associés. Les positions entourant ces résidus conservés ont un rôle important pour la spécificité de reconnaissance par un membre d'une famille de module, comme nous le verrons plus tard dans le cas des domaines PDZ. La longueur et la relative simplicité des SLiM favorisent leur apparition aléatoire dans de nouvelles séquences, pouvant avoir un impact fonctionnel fort pour un coût évolutionnaire faible. L'impact de l'apparition des SLiM dans l'évolution reste cependant à évaluer plus en détail. L'ensemble des motifs identifiés est répertorié dans la banque de données ELM (Eukaryotic Linear Motif ; elm.eu.org), incluant motifs d'interaction, sites de PTM (modification post-traductionnelle) et sites de clivage.

3.2 Complexité des assemblages moléculaires de la cellule

3.2.1 Régulation des voies cellulaires par les modules

En moyenne, les protéines eucaryotes sont plus longues et contiennent plus de domaines que les protéines procaryotes [57]. Dans le génome humain, les protéines métaboliques n'ont bien souvent qu'un domaine et sont plus courtes que les protéines impliquées dans la signalisation cellulaire [58]. La multiplication des modules d'interaction corrèle avec l'apparition de la multicellularité et la nécessité d'une communication cellulaire plus complexe. Les protéines eucaryotes de signalisation sont organisées par combinaison de modules de fonctions simples, permettant une fonction précise et finement contrôlée. La complexification des organismes passe par la recombinaison des modules et de leurs connexions ainsi que par la création de nouveaux gènes [59].

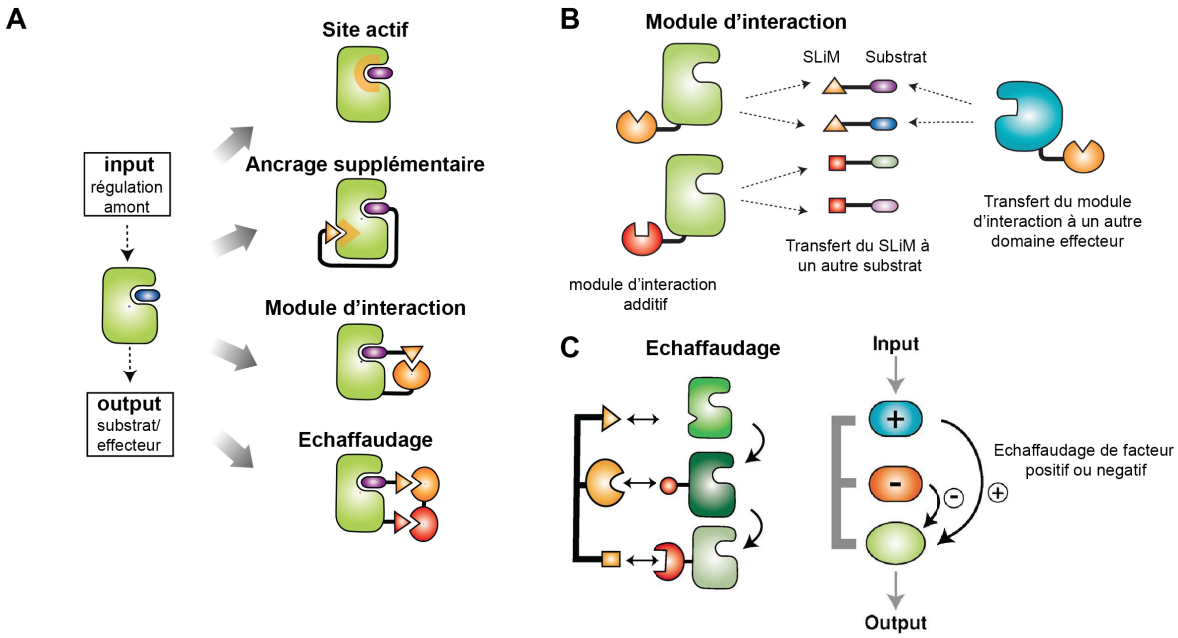


Figure 13 : Modes de régulation des voies de signalisation au niveau protéique. A Schéma résumant les différentes régulations disponibles pour un domaine catalytique dans son contexte protéique. **B L'**addition d'un module d'interaction proche d'un domaine catalytique peut modifier les partenaires de celui-ci. **C** Les protéines d'échafaudage rapprochent les acteurs d'une même chaine de régulation et permettent leur régulation croisée. Adapté de R. Bhattacharyya *et al.* [59]

Au niveau protéique, plusieurs mécanismes entrent en jeu pour assurer la régulation des domaines catalytiques des voies de signalisation (Figure 13). Dans le cas d'un domaine effecteur seul, l'allostérie permet de moduler l'activité de l'effecteur (output) en introduisant une régulation indépendante du substrat (input). La présence d'un ancrage supplémentaire du

substrat sur l'effecteur peut aussi jouer ce rôle, afin d'améliorer la spécificité de l'effecteur, ou de contrôler son activation.

L'addition d'un module d'interaction indépendant du domaine effecteur permet de séparer input et output sur deux entités distinctes, conduisant à une recombinaison bien plus importante des fonctions de la protéine, dans un système contraint par la variété limitée des repliements. Le domaine d'interaction peut avoir un rôle simple de recrutement du substrat ou de localisation des protéines dans leur compartiment, mais peut avoir un rôle plus complet "d'interrupteur" du domaine effecteur, lors de la liaison d'un input tiers [60]. Un domaine effecteur peut être régulé par plusieurs domaines d'interaction, permettant ainsi un contrôle fin. L'affinité faible des domaines d'interaction pour leur motif ainsi que leurs nombreux partenaires assurent la dynamique des complexes de signalisation, ne formant que des complexes transitoires, rapidement prêts à recevoir de nouveaux inputs, selon les changements d'environnement.

3.2.2 Echafaudage et complexe multi-protéines

Les protéines d'échafaudage sont composées de multiples modules d'interaction, sans domaine catalytique. Elles ont un rôle très important dans le maintien des machineries de contact et de communication intercellulaire. Elles permettent de maintenir les structures de contact cellulaire ancrées au cytosquelette, comme dans les jonctions serrées [61]. Elles ont aussi un rôle important dans la formation des complexes associés aux protéines membranaires, notamment au niveau de la synapse [62].

Les protéines d'échafaudage interviennent également au sein des systèmes de régulation, où elles peuvent recruter plusieurs effecteurs en une localisation cellulaire. En plus de ce rôle de regroupement, les protéines d'échafaudage permettent la localisation précise des effecteurs dans la cellule, notamment dans la région sous-membranaire. Leur présence peut regrouper les acteurs d'une même voie de signalisation au sein d'un même complexe, dans un volume confiné, favorisant ainsi la vitesse d'activation de catalyse ou d'interaction. Les affinités modérées entre modules et motifs sont encore une fois essentielles dans ce système, permettant de former et de déformer les complexes selon certaines stimulations.

3.3 HHD (Harmonin Homology Domain)

3.3.1 Structure

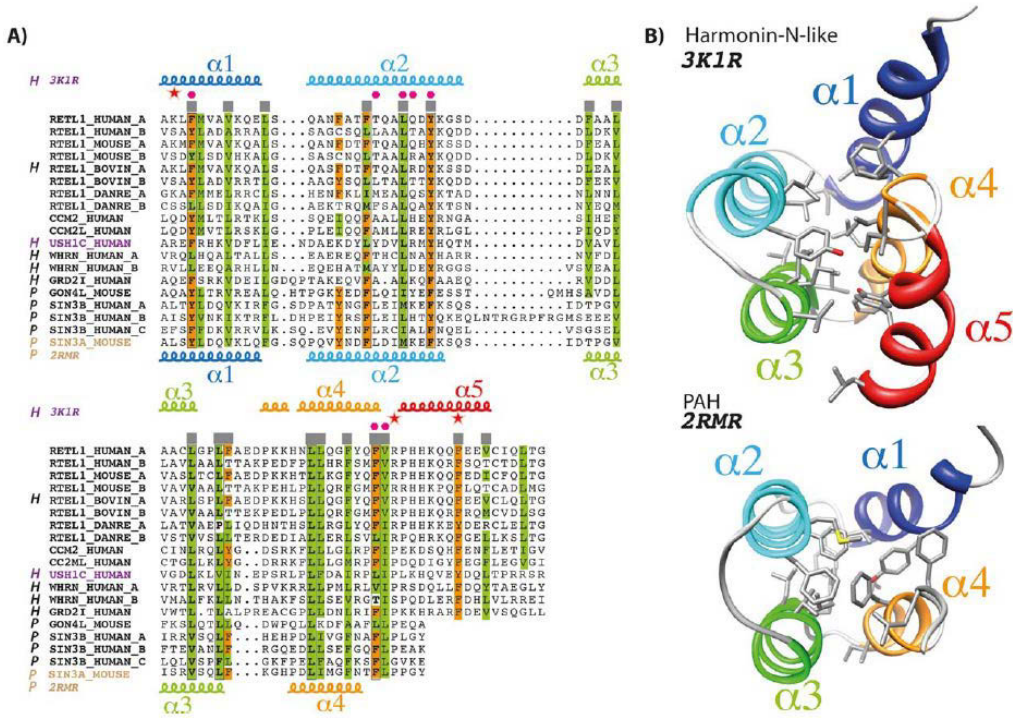


Figure 14 : Similarité de séquences entre les domaines HHD identifiés et des domaines PAH: (A) Alignement de séquences humaines et murines de l'ensemble des domaines HHD et PAH identifiés à ce jour. Les résidus conservés et hydrophobes sont indiqués en vert, les résidus conservés et aromatiques sont indiqués en orange. **(B)** Représentation en ruban du domaine HHD de l'Harmonine (pdb : 3K1R) ainsi que du premier domaine PAH de sim3A (pdb : 3RMR). Les résidus conservés du cœur hydrophobe sont indiqués en bâtonnet (Figure de Faure *et al*. 2014 [63]).

Les domaines HHD (Harmonin Homology Domain) sont de petits domaines d'interaction protéine-protéine intégrés dans de grandes protéines multidomaines d'échafaudage. Ils sont formés de 80 acides aminés repliés en un faisceau compact de 5 hélices α [64] (**Figure 14**). Bien que la structure soit bien conservée entre les domaines HHD d'une même espèce, leur identité de séquence est faible et ne concerne quasi uniquement que les résidus formant le cœur hydrophobe du repliement [63], rendant difficile leur identification par similarité de séquence (**Figure 14 A**). A ce jour, 9 domaines HHD ont été identifiés dans 6 protéines humaines : les quatre homologues Harmonine (une copie) [64], Whirline (deux copies), PDZD7 (une copie), Delphiline (deux copies), ainsi que dans la protéine CCM2 [65,66] (Cerebral Cavernous Malformation 2 ; une copie) et RTEL1 [63] (regulator of telomere elongation helicase 1, deux copies). Les quatre premières hélices des domaines HHD

partagent une forte similarité de structure et de séquence avec les domaines PAH (Paired Amphipatic Helix) (Figure 14 B) retrouvés en copie multiples dans les protéines humaines sin3a, sin3b et Gon4l.

3.3.2 Interaction

La résolution des structures du HHD de l'Harmonine [64] et de CCM2 [66,67] avec différents ligands a permis une meilleure compréhension des interactions de ces domaines. Ces domaines interagissent avec des hélices amphipathiques de quatre tours environ, dans un sillon formé entre l'hélice α1 et α2 et une surface hydrophobe dans la boucle reliant l'hélice α4 et α5. Dans les structures non liées à un partenaire, l'hélice α1 isole la surface hydrophobe d'interaction en se compactant sur l'hélice α2 et la boucle α4/α5. Dans les structures de HHD en complexe, l'hélice α1 est réorientée de quelques degrés pour accommoder l'hélice partenaire. Les hélices partenaires identifiées jusqu'à présent sont peu similaires entre elles et n'ont pour seuls points communs que leur caractère amphipathique et leur présence dans des régions non structurées des protéines. Il est actuellement difficile d'identifier un motif consensus d'interaction aux domaines HHD. Le sillon α1 et α2 du HHD de l'Harmonine est capable d'interagir avec deux hélices isolées et distinctes de la région cytosolique de la cadherine-23 [68] (Figure 15). L'un des motifs de la cadherine-23 est traduit par un exon différentiellement inclus dans ses isoformes, suggérant une modulation de l'interaction avec l'Harmonine selon l'isoforme exprimée.

Dans CCM2, le HHD se lie avec une faible affinité [65] à une hélice intramoléculaire isolée et présente dans la séquence en aval du domaine. Le partenaire cellulaire de CCM2, MEK3[66], possède également une hélice amphipathique isolée entrant en compétition avec le motif interne de CCM2 pour l'interaction avec son HHD. Les deux motifs hélicaux de partenaires de CCM2 interagissent avec une orientation similaire dans le sillon du domaine ; cependant l'un des motifs se positionne dans le sens opposé (Figure 15), suggérant que le caractère amphipathique des hélices partenaires des HHD est le principal moteur de l'interaction. En plus de leur repliement similaire, les domaines PAH interagissent aussi avec des hélices amphipatiques dans le sillon formé par leurs hélices α1 et α2 [69,70]. La structure du complexe ainsi formé est identique à celle observée chez les HHD, suggérant un lien évolutionnaire probable entre les deux familles de repliement.

Les HHD sont capables de former d'autres interactions que celles canoniques précédemment décrites (Figure 15). Le HHD de l'Harmonine forme un supramodule avec un

domaine PDZ adjacent positionné en C-terminal [71]. L'interaction se fait sur une surface faible impliquant la boucle α3/α4 et le début de l'hélice α4. En plus de l'interaction domaine/hélice entre CCM2 et MEK3, un domaine de dimérisation PB1 de MEK3 forme un deuxième ancrage sur le HHD de CCM2 [67]. Le complexe implique la surface exposée des hélices α2 et α3 du HHD avec un contact direct entre ces hélices de reconnaissance canonique et le domaine PB1 de MEK3. Une fois encore, les domaines PAH ont des capacités similaires d'interaction en dehors de leurs sillons canoniques. Le domaine PAH3 de Sin3a forme un second ancrage non-canonique entre ses hélices α2 et α3, similaire à l'interaction CCM2/MEKK3, avec un mini-domaine SAP formé de deux hélices en amont du motif d'interaction de son partenaire cellulaire SAP30 [70].

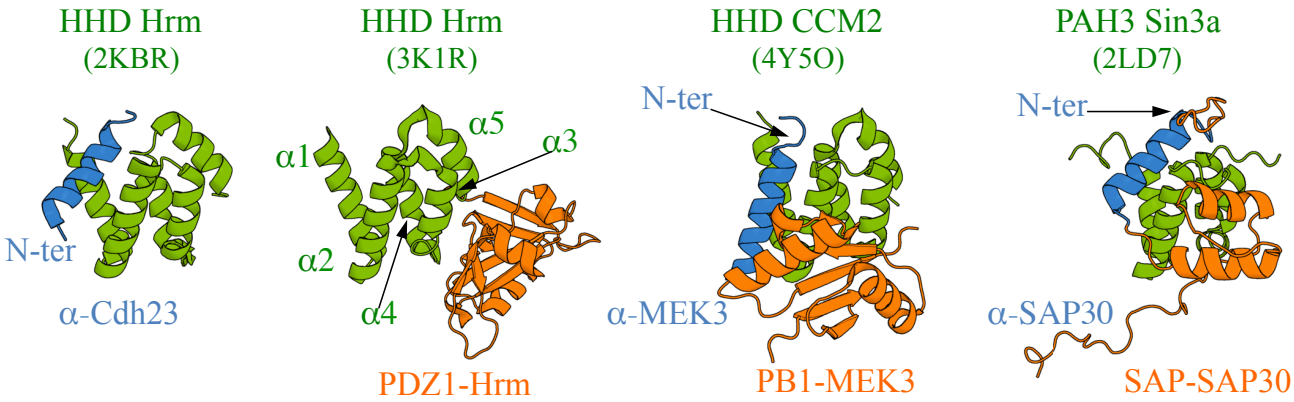


Figure 15 : Interactions canoniques et non-canoniques des domaines HHD. Représentation des complexes formés par les domaines HHD (vert) avec leurs motifs associés (bleu) ou sur des surfaces non-canoniques avec des domaines (orange).

3.4 PDZ (PSD-95, Dlg1, zo1)

Les domaines PDZ [72,73] composent une grande famille de modules d'interaction protéine-protéine ayant pour fonction principale la liaison à de courts motifs C-terminaux présents chez leurs protéines partenaires. Tout comme pour les modules SH2, l'apparition des domaines PDZ dans l'évolution semble intimement liée au développement de la multicellularité, et à l'expansion des protéines multi-domaines[74]. Les domaines sont caractéristiques des métazoaires et sont quasiment absents du génome des levures. La présence de quelques repliements PDZ dans les bactéries et les végétaux, en plus des levures, a amené à considérer l'hypothèse que ces domaines PDZ pourraient être la conséquence de transfert génétique provenant des métazoaires[75]. L'absence, à ce jour, de repliement similaire

aux domaines PDZ identifiés chez les archées, isolées des métazoaires, semble confirmer cette théorie[74,75].

Chez l'homme, on dénombre 266 domaines PDZ répartis sur 150 protéines contenant de 1 à 13 copies du module dans leurs séquences [76]. L'une des caractéristique frappante des domaines PDZ est leur propension à être présents en copie multiple dans les protéines. 28% des protéines humaines à PDZ contiennent des copies multiples, et 59% des domaines PDZ sont contenus dans des protéines à copie multiple du module. Cette occurrence est bien supérieure à celle de tous les autres modules humains d'adaptation (en excluant les domaines répétés) [74].

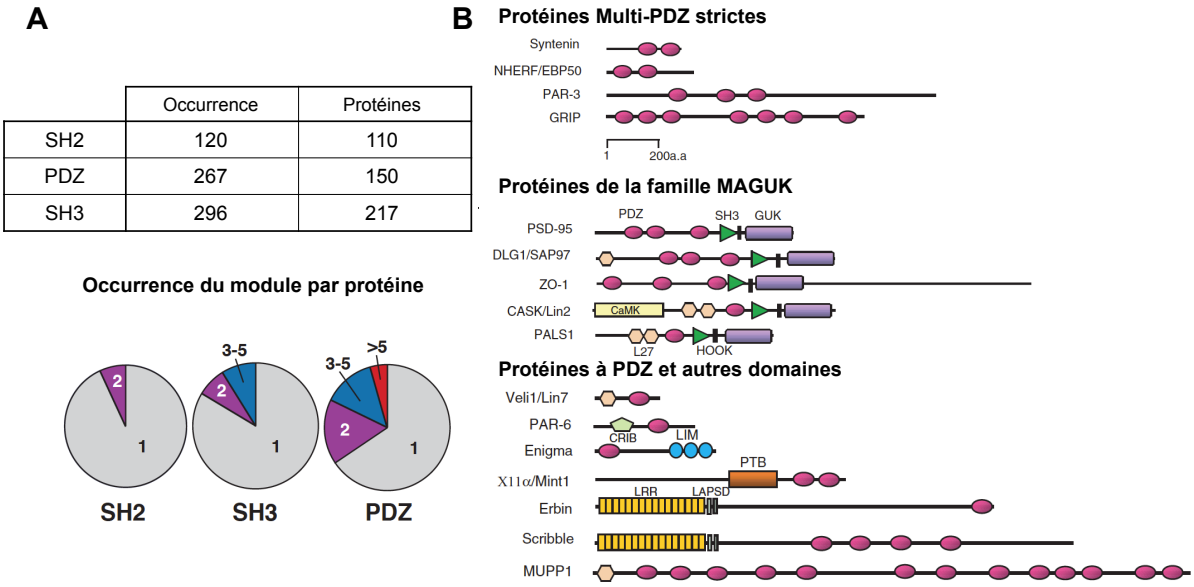


Figure 16 : Occurrence des domaines PDZ et diversité des protéines les contenant (A) Le tableau résume le nombre de domaine identifié dans le génome humain pour les 3 domaines d'adaptation les plus courants ainsi que le nombre de protéine contenant au moins l'un de ces domaines dans le génome. Les diagrammes en camembert montrent la part des protéines contenant une ou plusieurs copies de ces domaines d'adaptation dans le génome humain (adapté de Harris *et al.* 2001[74]). **(B)** Sélection de protéines à domaine PDZ classées selon leur organisation modulaire (adapté de Nourry *et al.* 2003[77]).

Le nom choisi pour cette famille de domaine correspond aux trois premières protéines identifiées contenant ces domaines, PSD-95 (Post Synaptic Density 95), Dlg1 (Disk Large Homolog 1) et ZO-1 (Zonula occludens 1) [72] et représente bien la grande diversité de fonctions couvertes par les domaines PDZ, de l'organisation post-synaptique[78] au maintien des jonctions cellulaires [79], en passant par la régulation des cascades de signalisation de la cellule [80]. Une majorité de protéines à PDZ ne possèdent pas de domaine catalytique et ont donc un rôle d'échafaudage strict [77,81], pouvant aller du simple maintien de complexes

fonctionnels au recrutement de protéines partenaires [82] notamment au niveau de récepteurs cellulaires [83]. Un nombre plus restreint de protéines à PDZ possède des domaines catalytiques et sont directement impliquées dans le maintien de l'homéostasie cellulaire [60,84].

3.4.1 Structure

Le repliement canonique des domaines PDZ consiste en 80 à 95 acides aminés formant un globule compact avec, de façon classique, leurs N- et C-terminaux proches dans l'espace. Ce repliement des PDZ contient de 5 à 6 brin β (de β1 à β6) formant un tonneau-ß antiparallèle incomplet, avec 2 hélices α (α1 et α2) couvrant les côtés ouverts du tonneau (Figure 17 B) [85,86]. En dehors des éléments de structures secondaires, trois régions des domaines PDZ sont très conservées. Les résidus "GD" en amont du brin ß4 ainsi que les résidus "NG" en aval de ce même brin sont extrêmement bien conservés dans les domaines PDZ (Figure 17 A). Leur présence est très probablement essentielle au maintien du repliement en tonneau-ß des domaines. La boucle "GLGF", située en amont du brin ß2, a une séquence conservée GϕGϕ, où ϕ indique un résidu hydrophobe. Le consensus montre une préférence marquée pour une Leucine et une Phénylalanine pour la première et la seconde positions hydrophobes respectivement, d'où le nom donné à la région. Un résidu R/K est conservé en amont de la boucle "GLGF" et stabilise une molécule d'eau structurale impliquée dans l'interaction avec les partenaires (Figure 17 C D). Cette région est conservée pour des raisons fonctionnelles ; en effet la boucle "GLGF" est essentielle à la liaison de l'acide carboxylique en C-terminal des motifs canoniques des domaines PDZ (Figure 17 C D).

A ce jour, 380 structures de domaines PDZ ont été déposées sur la PDB (de 145 domaines différents), couvrant les deux tiers des 152 protéines à domaines PDZ. Un tiers de ces structures a été résolu pour des domaines liés à un ligand et plus d'un tiers des structures de domaine PDZ a été déterminé par RMN, le plus souvent accompagné de données sur la dynamique de ces domaines. Cette somme de données structurales et dynamiques concernant les domaines PDZ permet une compréhension précise de leur fonctionnement et de leur régulation. (Informations recueillies en décembre 2016 sur le site de la banque de données des protéines de l'EMBL: https://www.ebi.ac.uk/pdbe).

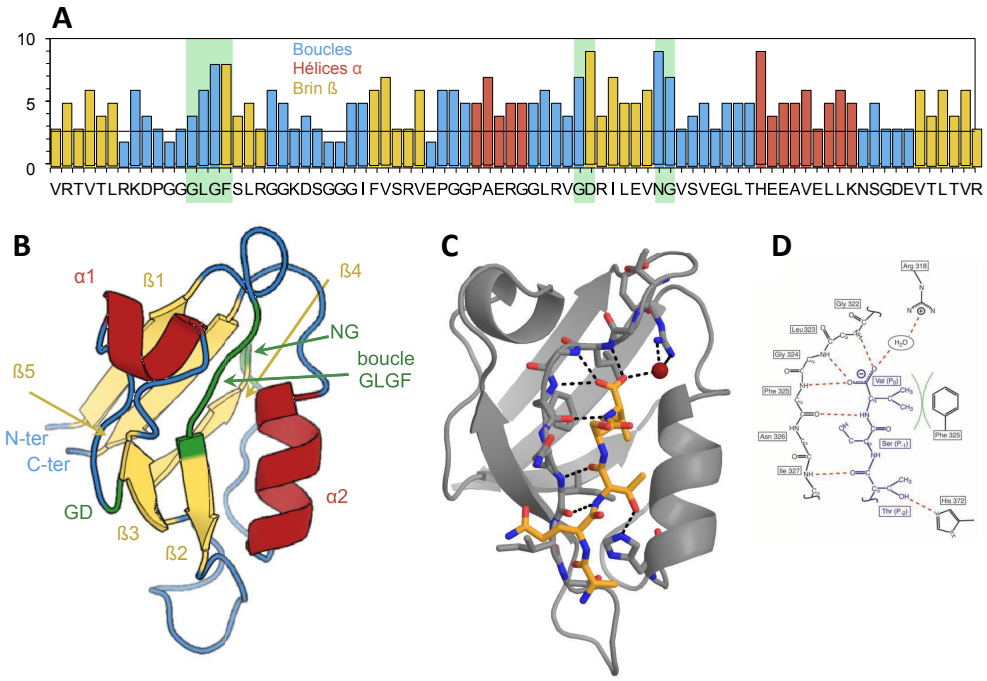


Figure 17 : Structure consensus des domaines PDZ et leur interaction canonique. (A) Le long de la séquence du PDZ est indiqué le résidu le plus probable pour chaque position basée sur la définition de la banque de données "conserved domain" (ncbi.nlm.nih.gov/Structure/cdd/) de la famille "PDZ-signaling" (PSSM : cd00992). Les structures secondaires consensus sont en jaune pour les brins ß, en rouge pour les hélices α et en bleu pour les boucles. Les régions strictement conservées sont surlignées en vert. **(B)** Structure du deuxième domaine PDZ de PTPN13 (pdb :1D5G) avec les brins ß en jaune, les hélices α en rouge et les régions conservées en vert. **(C)** Structure du troisième PDZ de PSD95 en complexe avec un peptide C-terminal de la protéine CRIPT (pdb : 1BE9). Le domaine PDZ est en gris et le peptide en orange. Les régions d'interaction sur le domaine et le peptide sont indiqués en bâtonnets avec une coloration par atome (oxygène en rouge et azote en bleu), les liaisons hydrogène protéine/peptide sont montrées en pointillés noirs. **(D)** Diagramme de la poche d'interaction avec le ligand, les liaisons hydrogène sont montrées en rouge et les interactions hydrophobes en vert (adaptée de Harris *et al.* 2001[74]).

3.4.2 Interaction

3.4.2.1 Motifs canoniques des ligands peptidiques

La fonction principale des domaines PDZ est d'interagir avec de courts motifs peptidiques linéaires, situés principalement en C-terminal de leurs protéines partenaires et nommés Motif d'Interaction aux domaines PDZ (PDZ Binding Motif ; PBM)[87,88]. La fixation du motif peptidique fait intervenir l'un des bords ouverts du tonneau ß constituant le cœur du domaine PDZ. Le ligand se positionne dans le sillon formé entre le brin ß2 et l'hélice α2 du domaine. Il forme un feuillet ß intermoléculaire en créant avec le brin ß2 du domaine un

réseau de liaisons hydrogène classique pour un feuillet antiparallèle (Figure 17 C D) [86]. L'extrémité C-terminal du PBM est stabilisée par un réseau de liaisons hydrogène se formant en haut du sillon avec les hétéroatomes du squelette de la boucle "GLGF" et une molécule d'eau structurale. Cette dernière est liée à la chaine latérale du résidu R/K en amont de la boucle [74]. La surface d'interaction du PBM avec le PDZ implique directement les résidus de la boucle "GLGF", l'ensemble du ß2, les chaines latérales de certains résidus de la boucle ß2-ß3 et du brin ß3 ainsi que les chaines latérales des résidus de la face interne de l'hélice α2. La structure globale des domaines PDZ n'est que très peu affectée par la liaison avec leurs ligands [89].

Sur la base de la reconnaissance C-terminale de leur ligand, les résidus des PBM sont numérotés en commençant par le dernier résidu (p0) en remontant vers le N-terminal (p(-1), p(-2) à p(-N)) :

- La position p0, en C-terminal, doit être un résidu hydrophobe [ILVCA] avec un fort biais vers [ILV]. La chaine latérale de ce résidu se positionne vers le cœur du domaine PDZ, dans une poche formée notamment par les deux résidus hydrophobes conservés de la boucle "GLGF" et un résidu du haut de l'hélice α2. L'acide carboxylique à l'extrémité du résidu forme un réseau de liaisons hydrogène avec le squelette peptidique de la boucle "GLGF" et une molécule d'eau structurale.
- La position p(-1) est sans préférence dans la définition du motif linéaire. Sa chaine latérale est dirigée vers le solvant et n'est généralement pas indispensable à la formation du complexe. Cependant il n'est pas rare que cette chaine latérale interagisse avec les chaines latérales des résidus des brins ß2 ou ß3. Bien que n'étant pas contrainte, cette position joue bien souvent un rôle important sur la sélectivité des motifs [88].
- La position p(-2) est strictement définie comme un résidu polaire de petite taille [ST], hydrophobe [ILVCF] ou chargé négativement [DE]. Sa chaine latérale interagit avec le premier résidu de l'hélice α2. Ces trois types différents de résidu définissent la classe du motif, que nous définirons plus loin dans le manuscrit.
- La position p(-3) ne fait pas strictement partie du motif linéaire de reconnaissance au domaine PDZ. Cependant, tout comme la position p(-1), la chaine latérale p(-3) peut interagir avec les résidus à la sortie du brin ß2 du PDZ. Cette position peut donc avoir un impact important sur l'affinité d'un motif pour des PDZ [88].

Les motifs de reconnaissance aux domaines PDZ sont strictement définis par les trois résidus C-terminaux de la protéine partenaire et notamment les positions p0 et p(-2). La

composition restrainte de ces deux positions est nécessaire au maintien de l'interaction avec les domaines PDZ mais ne permet pas une grande variété des motifs, essentielle à leur spécificité. La diversité est apportée par les résidus non-contraints du motif, ainsi que par les résidus en amont de la position p(-3). Du fait de ces disparités, la longueur donnée pour les PBM peut varier selon les publications.[76] Structuralement parlant, les quatre derniers résidus sont inclus dans le sillon du domaine PDZ, mais seuls les trois derniers résidus constituent le motif minimum, nécessaire à l'interaction. De nombreuses études ont aussi montré que les résidus en amont du PBM peuvent avoir un effet sur l'affinité et la spécificité, jusqu'à la position p(-12) pour certains domaines PDZ [84,90,91].

Les domaines PDZ ont classiquement une affinité comprise entre le micromolaire et la centaine de micromolaire pour les PBM [74,77]. Cette gamme d'affinités relativement faibles est comparable à celle observée pour les autres interactions domaine/motifs court. Elle permet une labilité des interactions et un renouvellement des partenaires, indispensable à la régulation des voies de signalisation cellulaire par exemple. Ces affinités peuvent être augmentées par des interactions non-canoniques additionnelles entre la protéine partenaire et le domaine PDZ [92,93], que nous présenterons plus tard dans ce manuscrit.

3.4.2.2 Classes des couples PDZ/ligand

Généralement, les PBM sont séparés en trois classes selon la nature du résidu en position p(-2) [74,88]. Par extension, les domaines PDZ ont été classés d'après leur capacité à interagir avec les trois classes de motif. Avec la résolution de nombreuses structures de PDZ, les déterminants de l'interaction favorisée envers une classe de PBM sont mieux compris et les domaines PDZ sont généralement classés selon la nature du premier résidu de leur hélice α2, indispensable à l'interaction avec la p(-2) du PBM (Figure 18).

- Les motifs de classe 1 ont une séquence X-[ST]-X-ϕ avec le résidu polaire en position p(-2) formant une liaison hydrogène entre son groupement hydroxyle et le noyau imidazole d'une Histidine en N-terminal de l'hélice α2 des domaines PDZ de classe 1. Cette classe contient la majorité des couples PDZ/PBM étudiés et environ la moitié des domaines PDZ.
- Les motifs de classe 2 ont une séquence X-[ILVF]-X-ϕ avec la position p(-2) formant un contact avec un résidu hydrophobe ou la partie aliphatique d'une Lysine, en N-terminal de l'hélice α2 du domaine. Bien que statistiquement plus probables ([H] vs [ILVAK]), les domaines PDZ de classe 2 sont bien moins nombreux que les domaines de classe 1. La

part de domaine possédant une chaine hydrophobe à cette position clé de l'hélice α2 ne représente qu'environ 20 % des domaines PDZ humains.

– Les motifs de classe 3 ont une séquence X-[DE]-X- ϕ avec la position p(-2) formant une liaison hydrogène avec le groupement hydroxyl d'une Tyrosine en N-terminal de l'hélice du sillon d'interaction et concerne un nombre encore plus limité de domaines.

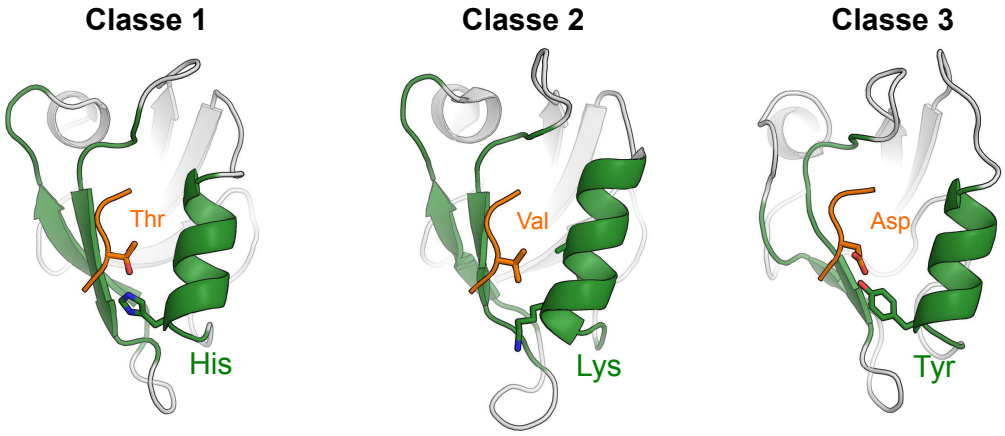


Figure 18 : Complexes PDZ/PBM caractéristiques des trois classes. Classe 1, PSD95-PDZ3 en complexe avec le peptide C-terminal de CRIPT (pdb : 1BE9). Classe 2, structure du domaine PDZ de PICK1 en complexe avec le peptide C-terminal GluR2 (pdb : 2PKU). Classe 3, structure du domaine PDZ de nNOS en complexe avec un peptide artificiel (pdb : 1B8Q).

En dépit de son utilisation très répandue dans les revues scientifiques, cette classification des domaines PDZ a été remise en cause par plusieurs groupes, en partie à cause de la grande plasticité des interactions PDZ/PBM. De nombreux domaines PDZ interagissent avec des ligands de différentes classes. L'une des classifications alternatives proposées se base sur les 5 derniers acides aminés du motif d'interaction et sépare les PBM en 16 catégories[94]. Une autre, se basant sur le résidu clé de l'hélice α2 et le résidu C-terminal du brin ß2 des domaines PDZ les ségrége en 25 catégories[95]. Les auteurs proposent de trier les domaines PDZ sur une matrice dépendante du type de résidus pour ces deux positions. Les auteurs démontrent la dépendance entre ces deux positions ainsi que la difficulté de ségréger les domaines PDZ ne faisant pas partie de la classe 1 en groupes distincts.

Dans un article récent, l'équipe du Dr. Ranganathan a montré qu'il était possible de passer d'un domaine PDZ de classe 1 (PSD95-PDZ3) à un domaine de classe 2 en ne mutant que ces deux positions[96,97]. Les auteurs ont ainsi déterminé le spectre d'interaction de quatre versions d'un domaine, (sauvage, H372A pour l'hélice α2, G330T du brin ß2 et le double mutant) contre 2000 ligands représentatifs de l'espace de séquence des PBM (Figure 19). Les auteurs montrent que la mutation G330T à la sortie du brin ß2 élargit le spectre d'interaction

du domaine de type 1 aux motifs X-[ILVF]-X-ϕ en plus des motifs classiques du domaine X-[ST]-X-ϕ. La mutation H372A de l'hélice α2 permet un échange quasi complet de la spécificité du domaine mais ne permet pas d'atteindre des affinités attendues pour un domaine spécifique à la classe 2. La double mutation G/H est nécessaire pour atteindre de telles affinités. Les auteurs se concentrent ensuite sur la capacité de la mutation G330T à faciliter le changement de classes des domaines PDZ sans contact direct avec le ligand. Ils concluent sur l'effet des mutations cryptiques et neutres comme facilitatrices de l'adaptation moléculaire rapide et comme élément important de l'évolution.

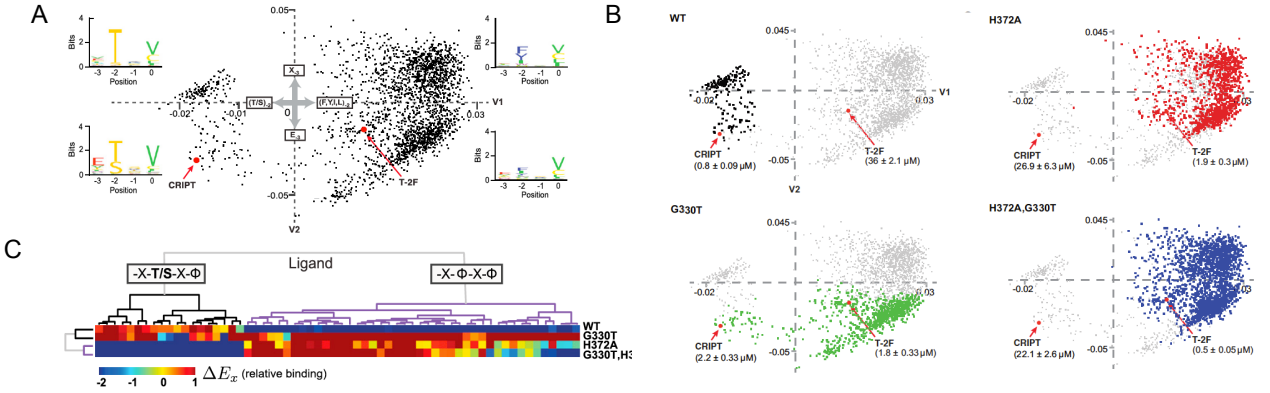


Figure 19 : Modification de classe de PSD95-PDZ3 par mutagénèse dirigée. (A) Espace des PBM testés dans cette étude et classés selon la ressemblance de profils d'interaction des peptides pour les quatre domaines. Les séquences logo montrent les consensus de chaque cadrant. (**B**) Espace de PBM exploré par chaque version du domaine. **(C)** Echantillons représentatifs de PBM avec leurs énergies relatives d'interaction pour chaque version du domaine. (Adaptée de Raman *et al.* 2016[97])

3.4.2.3 Motifs internes

Bien que l'interaction avec des motifs C-terminaux soit le signe distinctif des domaines PDZ, certains d'entre eux sont aussi capables d'interagir avec des régions internes de leurs protéines partenaires. Ces motifs internes sont très variés en terme de séquence et résistent à toute tentative de classification, rendant leur identification difficile. Le type de motif le plus décrit implique une séquence comparable aux motifs C-terminaux classiques complétée en aval par un Aspartate dont la chaine latérale s'oriente de façon à occuper la position normalement réservée à l'acide carboxylique C-terminal. Ce type de motif a été décrit pour au moins trois domaines PDZ, ceux de Par6[98], de NOS[99] et de Tip-1[100], liant des motifs internes de pseudo-types 1 et 2.

Une autre interaction décrite implique un pseudo-PBM inclus dans une épingle ß retrouvée en C-terminal du domaine PDZ de nNOS, interagissant avec les domaines PDZ de

la ß-syntrophine[101] ou de PSD95[102]. Dans cette structure, le coude de l'épingle ß se localise en haut du sillon, forçant la courbure de la séquence en aval et permettant d'éviter les incompatibilités stériques. D'une part, la Phénylalanine à l'origine de la courbure ß positionne sa chaine latérale dans la poche hydrophobe classique du domaine PDZ et, d'autre part, l'oxygène du squelette est contraint dans une position lui permettant de conserver une partie des liaisons hydrogène formées par un carboxylate classique. L'étude des variations d'énergies d'interaction de PBM classiques et internes mutés systématiquement a montré que les deux types de liaisons sont analogues [103]. Un exemple plus récent de motif interne nous vient du domaine PDZ de la protéine PICK1[90]. Dans ce cas, une population minoritaire de complexe se forme avec le motif C-terminal canonique se décalant de 2 résidus. Le résidu C-terminal du peptide se loge alors dans une poche hydrophobe formée au-dessus de la boucle "GLGF", décalant alors le peptide vers le haut.

Des études par double hybride et/ou bioinformatique ont proposé des listes de PBM internes potentiels[104,105]. Il est cependant difficile d'évaluer la pertinence de ces résultats à haut débit sans vérifications ultérieures *in vitro*. Il est cependant très probable que les quelques exemples de motifs internes connus à ce jour ne représentent pas la diversité offerte par les motifs internes et le nombre d'interactions domaine PDZ/PBM interne est probablement fortement sous-estimé [106].

3.4.2.4 Interactions avec les lipides

Une autre caractéristique non-canonique mais cependant partagée par de nombreux domaines PDZ est leur capacité à interagir avec les lipides et par extension avec la membrane plasmique [106–108]. Cette caractéristique est à rapprocher avec la présence fréquente de protéines à domaine PDZ dans la région sous membranaire des cellules et dans leur rôle récurrent dans l'organisation de complexes de signalisation des récepteurs membranaires.

Les interactions PDZ/membrane sont généralement contrôlées par attraction électrostatique entre une région fortement cationique à la surface du domaine PDZ et les têtes polaires des lipides anioniques, notamment des Phosphatidylinositol Phosphates de la face intérieure des membranes. Ce type d'interaction a été principalement décrit pour les Synténines 1 et 2 [109], PTP-BL, Par3 et PICK1 [110].

Deux études à haut débit [111;112], effectuées sur un panel de domaine PDZ, ont permis d'évaluer à 20-40% la population de domaines interagissant avec les membranes. Ces études ont de plus montré que tous les domaines capables de lier les membranes contiennent au moins une surface fortement cationique[111]. Les auteurs séparent les interactions en deux

classes principales, selon que le site cationique chevauche ou non le site de liaison au PBM du domaine[111]. Une étude sur culture de cellules *in vitro*[113] testant la quasi-totalité des domaines PDZ (246/266) a montré qu'un peu plus de 20% des domaines PDZ isolés avaient une localisation, discrète ou marquée, à l'une des membranes cellulaires, confirmant les études moléculaires précédentes.

Les domaines PDZ capables d'interagir avec les membranes sont généralement inclus dans les protéines multi-domaines contenant d'autres domaines d'interaction avec les lipides, comme les domaines BAR[107] ou PH[114]. La multiplication des points d'ancrage permet d'augmenter l'avidité de la protéine [106] pour la membrane, en partant de domaine d'affinité individuelle relativement faible. Dans le cas des domaines PDZ de Par3[112] et de PICK1[110], une insertion partielle des domaines dans la membrane plasmique a été décrite. Ces deux domaines PDZ possèdent des régions cationiques permettant l'adressage du domaine à la membrane, ainsi qu'un motif spécifique C-P-C favorisant l'insertion dans la membrane. L'interaction entre les domaines PDZ et les lipides a un rôle important dans la localisation de ces PDZ au niveau du compartiment membranaire.

3.4.2.5 Dimérisation

Une autre caractéristique commune à de nombreux domaines PDZ est leur tendance à former des oligomères homotypiques. Dans l'étude de référence sur le sujet [115], les auteurs ont testé l'ensemble des interactions inter-domaines possibles de 157 domaines PDZ de souris, évaluant ainsi plus de 12 000 complexes par polarisation de fluorescence sur biopuce. Les auteurs ont ainsi identifié la formation de 37 complexes impliquant 46 des 157 domaines PDZ testés. En considérant que le panel choisi de domaine PDZ est représentatif, les auteurs estiment que l'interaction homotypique concernerait jusqu'à 30% des domaines PDZ. Dans cette étude, les auteurs ne distinguent pas les homo-dimérisations (3 cas), interactions entre domaines adjacents (0 cas) et interactions entre domaines PDZ différents (43 cas).

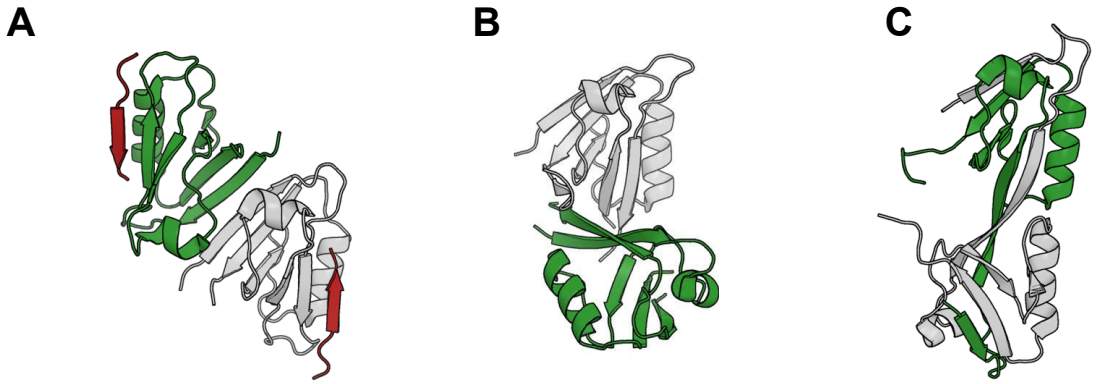


Figure 20 : Dimérisation des domaines PDZ. Un protomère est représenté en vert, le second en gris, le ligand est montré en rouge. **(A)** dimère du PDZ6 de la protéine GRIP1 par formation d'un feuillet inter-protomère ß1-ß5 (pdb : 1N7F) en complexe avec le PBM de la Liprine **(B)** dimère du domaine PDZ de GRASP par interaction des brins ß2-ß3 (pdb : 2EGK) **(C)** dimère par échange de brin du PDZ2 de ZO-1 (pdb : 2JWE).

La dimérisation des domaines PDZ a été décrite plus en détails dans quelques exemples. Parmi eux, plusieurs cas d'homo-dimérisation se font à l'opposé du site de liaison des PBM par formation d'un feuillet intermoléculaire des brins β1 et β5 (Figure 20 A). Cette structure a été observée par cristallographie pour le 6ème PDZ de GRIP1 [116] et l'unique PDZ de Shank1 [117]. Une autre forme de dimérisation connue implique les brins allongés β2/β3 de certains domaines PDZ comme celui de GRASP [118] et de MAST2 [91]. Dans ces deux cas, la fixation du PBM des partenaires favorise la monomérisation du domaine PDZ [91,118]. Les domaines PDZ des trois autres protéines de la famille des MAST ont été cristallisés sous une forme dimérique similaire à GRASP [119] (Figure 20 B). Cependant, leur cristallisation a été rendue possible par ajout d'un PBM en C-terminal des domaines, favorisant fortement l'association. Un dernier cas de dimérisation unique se fait par l'échange de structures secondaires, brins ß1 et ß2, de deux protomères du PDZ2 de Zo-1. Les domaines sont superposés, avec une rotation relative de 180°, permettant aux deux sillons d'interaction d'être alignés [120] (Figure 20 C).

3.4.3 Modulation de l'interaction des PDZ

Pour que les processus cellulaires soient correctement exécutés, il est indispensable que les complexes impliqués se forment au bon endroit et au bon moment. La boucle de régulation contrôlant l'expression des partenaires et l'adressage dans des sous-compartiments cytoplasmiques sont des moyens de régulation indispensables au niveau de la cellule. Peu de temps après l'identification des domaines PDZ, il est apparu que ceux-ci n'avaient pas qu'un simple rôle "d'architecte" des complexes sous-membranaires mais pouvaient aussi avoir un rôle plus dynamique dans la signalisation cellulaire [121,122]. Les domaines PDZ doivent interagir spécifiquement avec leurs partenaires en dépit du manque de diversité offerte par les motifs courts qu'ils reconnaissent. Avec la multiplication des domaines PDZ dans le génome lié à la complexification de l'architecture moléculaire, réduire la redondance des interactions et améliorer la spécificité des complexes PDZ/PBM est une nécessité qui a favorisée leur différenciation. Dans ce paragraphe, nous présenterons les stratégies évolutionnaires développées pour singulariser les différents domaines PDZ [123]. Ce sujet a déjà été abordé à plusieurs reprises dans d'excellentes revues, notamment Luck *et al.* 2012 [76] et Ye *et al.* 2013 [124].

3.4.3.1 Chemins énergétiques dans les domaines PDZ

Les domaines PDZ subissent de très faibles variations de structure lors de la liaison de leurs ligands. Néanmoins, il est apparu que le site de liaison de certains PDZ pouvait être énergétiquement lié à des régions distales, selon le modèle d'allostérie des domaines isolés de Smock et Gierasch [125] (Figure 21). Très tôt, des études de coévolutions des résidus de domaine PDZ ont montré des couplages entre certains résidus, formant des chemins au travers la structure des domaines [126]. Des années plus tard, ces travaux amèneront les auteurs à proposer la définition de "secteurs" comme des chemins conservés dans les familles de repliement, et capables de transmettre de l'information au travers des domaines [127]. En parallèle, de nombreux travaux ont montré que l'interaction d'un PBM pouvait propager une modification de la dynamique des chaines latérales de résidus depuis le site de liaison jusqu'à des régions distales du domaine, le chemin suivi correspondant au "secteur" décrit par l'étude de coévolution [128,129]. La réversibilité de ces chemins d'information a été observée à plusieurs reprises en évaluant l'impact de mutation dans les régions distales sur l'affinité des domaines PDZ [130,131]. Ces voies "allostériques" permettent une modulation de l'affinité des domaines

PDZ pour leurs ligands, par exemple par l'interaction non-canonique de structures additionnelles (extensions ou domaines) sur des régions éloignées du site actif. Ces cas seront présentés dans les parties suivantes.

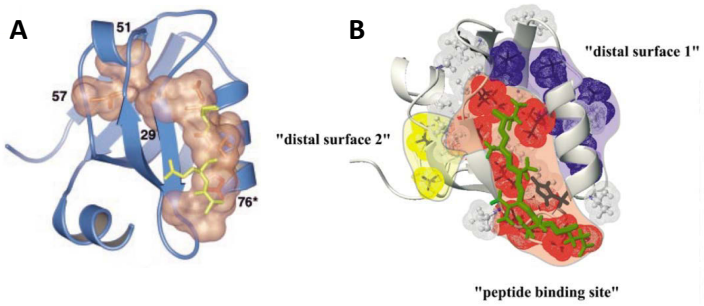


Figure 21 : Chemin énergétique conservé des domaines PDZ : (A) chemin déterminé par l'étude de coévolution (adaptée de Lockless *et al.* 1999[126]). **(B)** chemin de perturbation dynamique propagée depuis le site actif à des surfaces éloignées sur le PDZ2 de PTP1-BL (adaptée de Fuentes *et al.* 2004).

3.4.3.2 Extensions

Dans les premières années suivant leurs découvertes, les domaines PDZ ont été étudiés par l'approche traditionnelle de biologie structurale consistant à "diviser pour mieux régner", soit délimiter les domaines PDZ de la façon la plus stricte, notamment afin d'éviter les extrémités flexibles pouvant compromettre la cristallisation. Dans le cas spécifique des domaines PDZ, cette approche a été délétère, notamment avec la délimitation tronquée en N-terminal de l'ensemble des domaines PDZ dans certaines bases de données d'architecture modulaire tel que SMART.

Il apparaît maintenant clair qu'une grande partie des domaines PDZ possède des extensions structurées à leurs extrémités. En considérant 50 résidus supplémentaires aux deux extrémités du repliement canonique, environ 40% des domaines PDZ humains possèdent au moins une région prédite comme ordonnée et comme possédant des structures secondaires [132]. Ces extensions ont différents rôles, sur la stabilité ou les affinités du domaine considéré.

Malgré la présence fréquente d'extensions aux domaines PDZ, il existe peu de descriptions documentées de leur impact sur l'affinité. Le PDZ3 de PSD95 possède une hélice additionnelle en son C-terminal, trois résidus après le dernier brin du domaine[133]. Cette hélice additionnelle réduit la dynamique rapide (ns à µs) des chaines latérales des résidus de l'ensemble du domaine sans modifier sa structure (Figure 22 B). Cette diminution de la dynamique interne du domaine réduit le coût entropique de l'interaction avec le PBM de CRIPT, améliorant ainsi l'affinité d'un facteur 25. Une étude suivante a remis en cause la nature purement entropique de ce gain en démontrant que certains résidus en amont de motif canonique du partenaire était capable d'interagir directement avec l'extension hélicale[92]. Cette

interaction stabilisatrice entre le domaine PDZ et son extension peut être rompue par la phosphorylation d'une tyrosine de l'hélice localisée à l'interface. Cette phosphorylation permettrait donc de moduler l'affinité du domaine pour son ligand par découplage de son extension[134]. De façon relativement similaire, le deuxième PDZ de NHERF1 possède une extension C-terminale formée de deux hélices interagissant directement avec le domaine. Cette interaction favorise le repliement de celui-ci et réduit le coût entropique de la liaison au peptide partenaire [135] sans interaction directe avec le partenaire (Figure 22 B). Le premier PDZ de la protéine MAGI1 possède une extension C-terminale se rigidifiant lors de l'interaction avec le PBM de la protéine virale HPV E6 et pourrait ainsi avoir un rôle important dans le cycle cellulaire du papillomavirus [136,137]. Un exemple plus original concerne l'interaction du PDZ de la protéine Shank3 avec SAPAP3 (Figure 22 C). Dans ce cas, le domaine PDZ possède une épingle ß composée de six résidus en amont de son premier brin. Le peptide PBM SAPAP3 interagit de façon classique avec le premier protomère de Shank3, mais la région amont du ligand forme un feuillet ß intermoléculaire avec l'épingle ß d'un second protomère de Shank3, augmentant l'affinité d'un facteur 100 et formant ainsi un dimère entrecroisé dans le complexe [93].

Un sous-groupe de domaines PDZ possède des brins ß2-ß3 ou une boucle ß2-ß3 particulièrement allongés pouvant être considérés comme extensions. On retrouve notamment les PDZ des familles MAST[84,119], des Par[138] des Shank[117] ainsi que PICK1[139] et SNX27[140]. Malgré les différences structurales des extensions, dans de nombreux exemples, la région additionnelle comporte une surface supplémentaire d'interaction avec le ligand. Dans les cas de Par-3[138], MAST2[84] et PICK1[107], une région hydrophobe à la surface des brins ß2 et ß3 permet l'interaction avec un résidu aromatique localisé entre les positions p(-11) et p(-5) du partenaire (Figure 22 A). Ce point d'ancrage distal du site canonique améliore l'avidité des ligands comportant un tel résidu, pour le domaine PDZ[91]. Dans le cas de Shank1, une poche hydrophobe se forme par la grande boucle structurée ß2/ß3 qui interagit avec la chaine latérale du tryptophane p(-5) du partenaire ßPix[141] (Figure 22 A). Un autre cas d'interaction non-canonique de la région ß2-ß3 est illustré avec le domaine PDZ de l'Erbin. Dans ce cas, l'interaction entre ce domaine et le C-terminal de la protéine partenaire est contrôlée par une tyrosine phosphorylable en p(-7). La tyrosine se loge dans une poche supplémentaire formée par la longue boucle ß2/ß3. La phosphorylation de cette tyrosine abolit l'interaction dans cette poche et réduit l'affinité du ligand par 2.5 fois pour le domaine PDZ de l'Erbin [142].

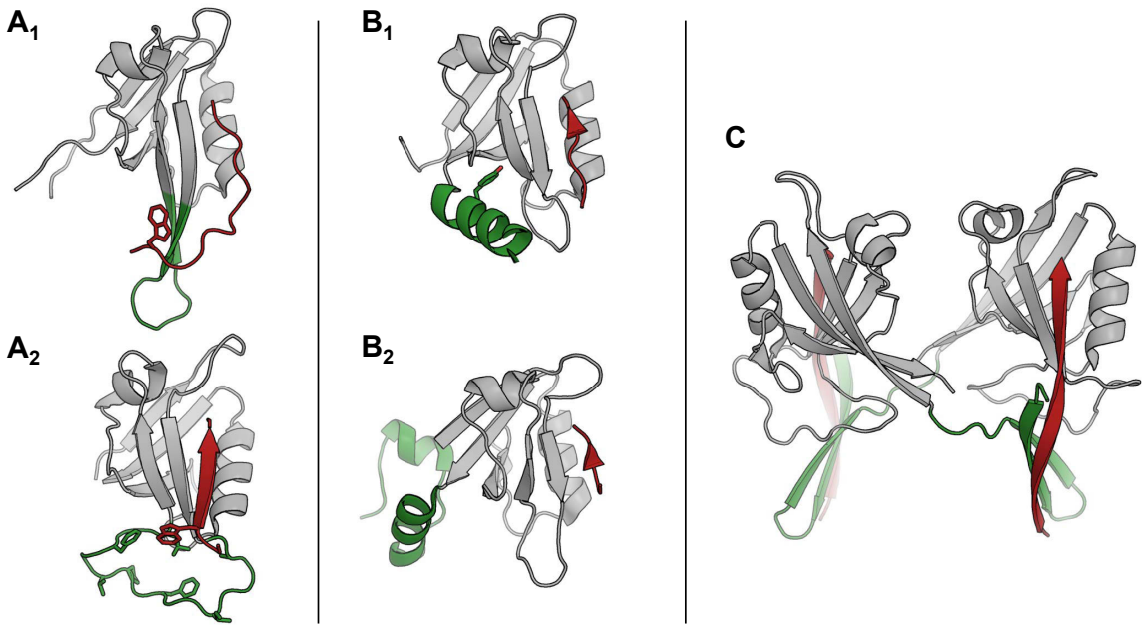


Figure 22 : Extensions des domaines PDZ. En gris est représenté le domaine PDZ canonique, en vert les extensions et en rouge le ligand. Certains résidus d'importance, discutés dans le texte, sont représentés en bâtonnets. **(A_1)** Extension des brins ß2 et ß3 amenant à un second point d'ancrage en p(-11) du peptide dans le complexe formé par le domaine PDZ de MAST2 et la protéine G du virus de la Rage (pdb : 2KQF). **(A_2)** Extension de la boucle ß2/ß3 du domaine PDZ de Shank1 en interaction avec le tryptophane p(-5) du C-terminal de ßPIX (pdb : 3L4F). **(B_1)** Extension C-terminale du domaine PDZ3 de PSD-95 formant une hélice interagissant avec le domaine PDZ par une Tyrosine phosphorylable et interagissant avec un peptide synthétique (pdb : 5D13). **(B_2)** Extension C-terminale du second PDZ de NHERF1 stabilisant le domaine et permettant l'interaction avec le ligand CFTR (pdb :2M0V). **(C)** Extension N-terminale formant une épingle ß du domaine PDZ de Shank3. L'extension crée une dimérisation croisée des complexes Shank3/SAPAP (pdb : 5IZU)

Plusieurs domaines PDZ avec extension ont été identifiés, leurs structures ont parfois été résolues avec ou sans ligand. Néanmoins, leur rôle n'est pas bien identifié et mérite d'être étudié en détail [76]. Des extensions sont également connues pour stabiliser les interactions avec d'autres domaines permettant de former des complexes supramodulaires, nous les présenterons aux cours des deux prochaines parties.

3.4.3.3 Tandems PDZ

Comme décrits précédemment, deux traits particuliers distinguant les domaines PDZ des autres modules d'interaction est leur propension à apparaître en copie multiple dans une protéine [74] ainsi que leur capacité à former des complexes homotypiques [115]. On dénombre environs 35 protéines possédant deux domaines PDZ successifs séparés par moins de 25 résidus, on parle alors de tandem de PDZ. Malgré cette abondance de tandem PDZ, il existe relativement peu de publication sur leurs organisations et leurs rôles dans la cellule. Une

grande partie des résultats de cette thèse concerne la structure du tandem de PDZ contenu dans la Whirline, nous allons donc détailler la littérature pour ce type d'organisation dans le prochain paragraphe.

La Synténine possède deux domaines PDZ, de faible identité de séquence, et séparés par un lien de seulement 5 résidus. Les deux domaines interagissent par des régions éloignées de leur sites actifs [143], sans conséquence sur l'accessibilité aux ligands. Une étude RMN du tandem en solution (RDC et CSP)[144] montre que les nombreuses structures cristallines du tandem sont semblables mais ne correspondent pas aux conformations plus diverses présentes en solution. Le tandem de la Synténine a une forte affinité pour son partenaire Syndecan. Les deux domaines PDZ de la Synténine sont tous deux capables d'interagir simultanément avec un dimère de Syndecan, augmentant les affinités des domaines isolés [145,146]. Dans ce modèle de structure, le ligand interagit de façon non-conventionnelle avec le PDZ1, en ne fixant que son carboxylate C-terminal sans autres contacts avec le site du PDZ [146].

La protéine GRIP1 est composée de 7 domaines PDZ organisés en deux régions et un domaine isolé en C-terminal: PDZ1-3, PDZ4-6 et le PDZ7 [147]. Dans chaque région, les deux domaines PDZ N-terminaux PDZ12 et PDZ45 s'organisent en tandem ; l'un des domaines PDZ ne possède qu'un rôle chaperon pour l'autre domaine endossant, lui, la fonction de liaison au PBM du tandem (Figure 23). Dans le cas de PDZ12, le second domaine PDZ a son site actif complétement occulté par la surface d'interaction avec le PDZ1, empêchant toute liaison à un partenaire [148]. Le PDZ2 ainsi que le lien peptidique entre les domaines sont indispensables au repliement natif et à la fonction du premier domaine. Les auteurs justifient le "sacrifice" de la fonction d'un domaine au profil d'un autre pour permettre une régulation plus fine de la protéine. Leur hypothèse est qu'une phosphorylation sur une tyrosine conservée du lien inter-domaine pourrait déstabiliser le supramodule et déplier le premier domaine. Cette hypothèse se base sur une la perte de la conformation en supramodule lors de la mutation de cette tyrosine[148]. Le phénomène n'a, en revanche, pas été montré dans la cellule. Pour le tandem PDZ45, les sites actifs des deux domaines sont accessibles au solvant, mais une déformation importante du site actif du PDZ4 le rend incapable d'interagir avec un PBM [149]. Comme pour le tandem PDZ12, le domaine PDZ5 ne peut se replier sans le domaine adjacent PDZ4.

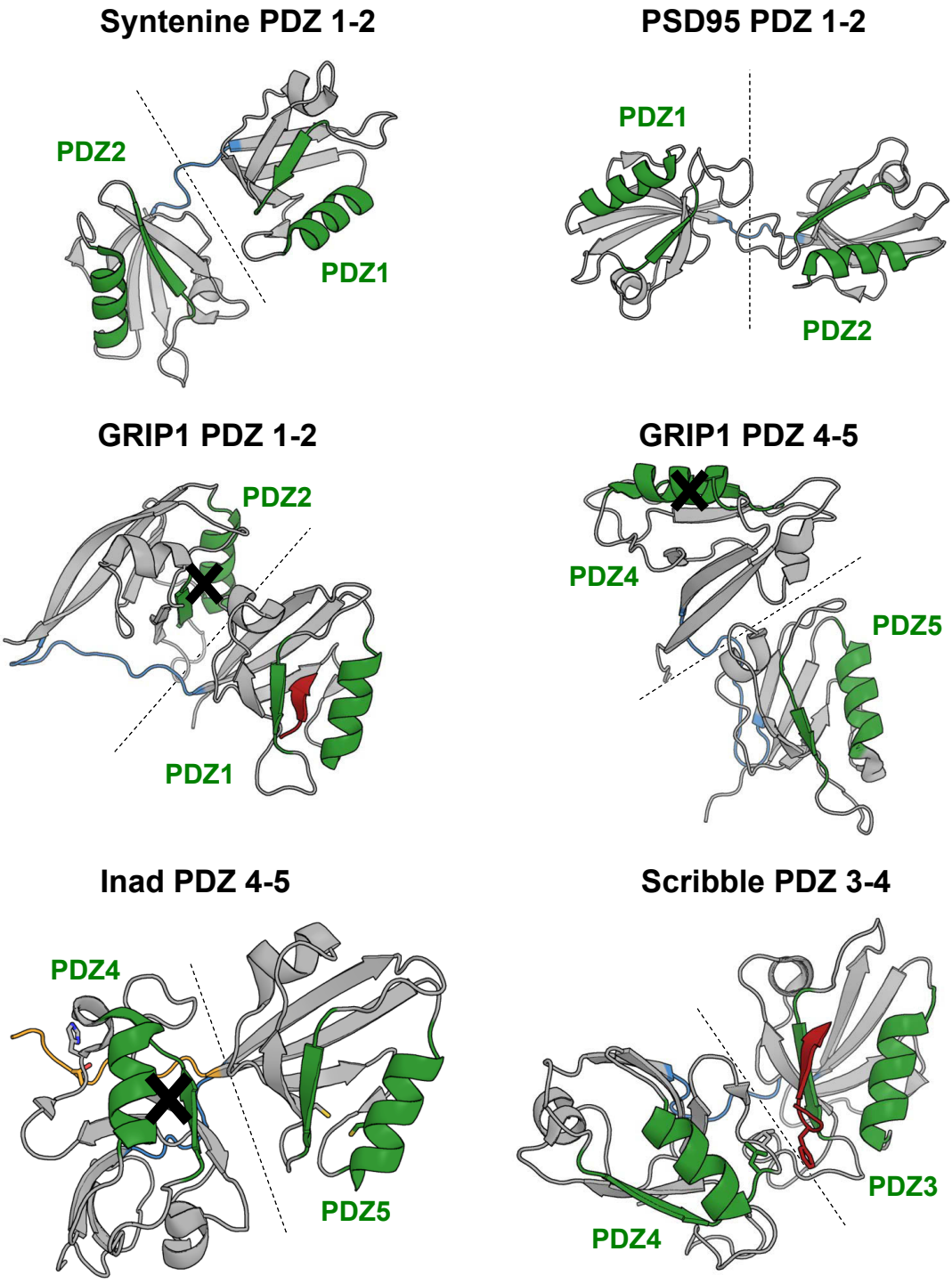


Figure 23 : Structures de tandem PDZ résolues à ce jour. Représentation des tandems de domaines PDZ dont la structure a été résolue à ce jour. Le sillon d'interaction des domaines est en vert (brin ß2 et α2) les ligands sont indiqués en rouge et les liens entre les domaines en bleu. L'extension C-terminale du PDZ5 de INAD est indiquée en orange. Synténine PDZ1-2 pdb :1N99 ; PSD95 PDZ1-2 pdb :3GSL ; GRIP PDZ1-2 pdb : 2QT5 ; GRIP PDZ4-5 pdb : 1P1D ; INAD PDZ4-5 pdb : 3R0H ; Scribble PDZ3-4 pdb : 4WYU.

La protéine PSD95 forme un tandem composé de ses deux domaines PDZ N-terminaux. Ce tandem est principalement maintenu par un lien inter-domaine court et rigide (4 résidus dont 2 Prolines) contraignant fortement la liberté d'orientation relative des deux domaines [150]. La contrainte permettrait de garder une orientation commune aux sites actifs des deux domaines, favorable à l'interaction avec des récepteurs membranaires multimériques au glutamate, partenaires de PSD-95. Cette orientation contrainte a permis de développer des petites molécules bidentates capables de cibler les deux sites du tandem [151]. Le cas de la protéine de drosophile INAD est un bel exemple de régulation des tandems PDZ. Dans cette protéine présente dans les cellules photoréceptrices, le site actif du PDZ5 est verrouillé par un pont disulfure. L'ouverture du sillon est régulée par la réduction réversible du pont disulfure induite par l'exposition du domaine à la lumière [152]. Ce PDZ5 forme un supramodule homotypique avec le PDZ4 (Figure 23). Cette conformation est notamment maintenue par une extension C-terminale du PDZ5 entourant le domaine PDZ4. L'interaction entre cette extension est maintenue par l'acceptation d'une liaison hydrogène par une Histidine protonable sur l'extension du PDZ5. La conformation modifie le potentiel redox de la liaison disulfure du PDZ5, gardant son sillon d'interaction ouvert. Un changement local de pH [153] pourrait changer l'état de protonation de l'Histidine clé de l'extension C-terminale du PDZ5, dissociant instantanément le supramodule homotypique PDZ45. La dissociation aurait alors un effet sur le potentiel redox du pont disulfure, refermant le site actif du domaine PDZ5 [50].

La protéine Scribble possède aussi un supramodule, formé par ses domaines PDZ3 et 4. La structure de ce tandem montre que les deux sites actifs des domaines sont accessibles dans le supramodule homotypique (Figure 23) [154]. L'interaction des deux domaines permet d'agrandir la poche d'interaction du PDZ3, en formant une interaction directe entre la Phenylalanine p(-4) du partenaire et la boucle ß4-α2 du PDZ4, améliorant d'un facteur 4 l'affinité. Il reste à démontrer si ce supramodule peut se dissocier pour moduler l'interaction avec le partenaire.

Ces exemples montrent bien la diversité des rôles des tandems de PDZ. De manière générale, les supramodules homotypiques ont deux rôles principaux; augmenter l'affinité pour un partenaire ou fournir un mode supplémentaire de régulation de l'interaction de l'un des domaines.

3.4.3.4 Supramodules hétérotopiques

De part leur nature, les domaines PDZ sont principalement trouvés dans des protéines multi-domaines, aux côtés de domaines d'interaction protéine-protéine ou des domaines catalytiques. Nous avons vu précédemment la capacité de tandem de domaine PDZ à former des supramodules homotypiques. Cependant, les domaines PDZ sont aussi capables de former des supramodules hétérotypiques avec des domaines adjacents différents [124].

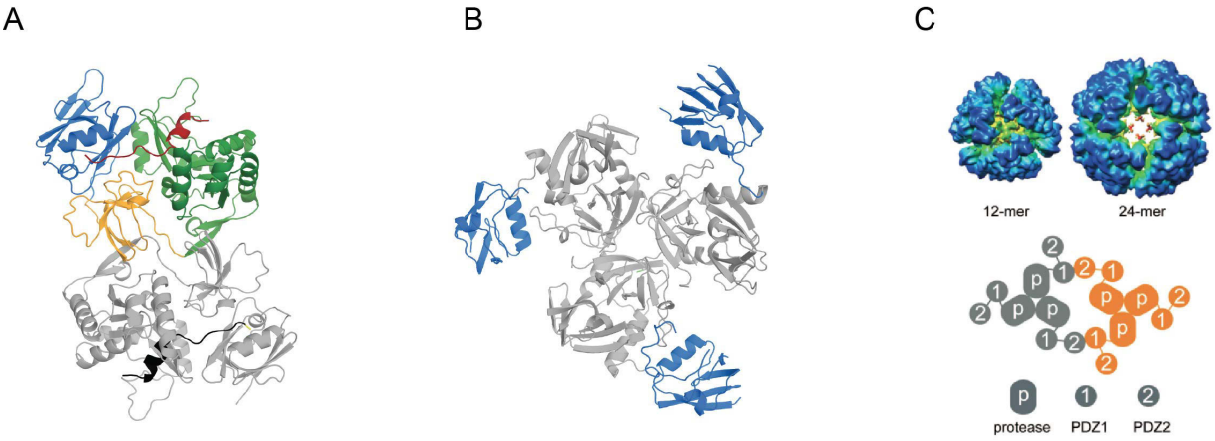


Figure 24 : Supramodules connus des domaines PDZ. A PDZ-SH3-GuK de Mpp5 en complexe avec le peptide C-terminal de Crumb (pdb : 4WSI). Le PDZ est en bleu, SH3 en orange, GuK en vert et ligand en rouge. Le second protomère du dimère swappé est en gris, avec le ligand en noir. **B** Structure trimérique de DegS. Le domaine protéolytique est en gris et le PDZ en bleu. (pdb : 4RR1) **C** Densité de microscopie électronique des formes 12-mère et 24-mère de la protéine DegS. Le schéma montre les contacts formés entre trimères, pour former les oligomères supérieurs. Figure adaptée de Jiang *et al.* [155].

La famille de protéines à PDZ MAGUK est la plus étudiée (Figure 24 A). Les protéines de cette famille, mis à part MAGI, possèdent une architecture conservée et formée par un PDZ, un domaine SH3 et un domaine Guanylate kinase sans activité catalytique (GuK) [156]. Ces trois familles de domaines ont des motifs définis et indépendants d'interaction: les SH3 interagissent avec des régions riches en proline et les domaines GuK interagissent avec des motifs phosphorylés. Ces trois domaines s'associent pour former une entité structurale unique, conservée entre MAGUK et nommée PSG, capable d'interagir en plusieurs points avec leurs protéines partenaires. Dans toutes les MAGUK, les domaines GuK et SH3 forment un ensemble structural, le PDZ pouvant parfois être relié par une région plus flexible [156].

Dans l'exemple de MPP5, le supramodule hétérotypique forme un dimère swappé, avec le domaine GuK d'un protomère en interaction avec le couple PDZ/SH3 du second [157]. Cette conformation permet l'interaction du C-terminal du partenaire Crumbs avec les trois domaines du PSG (Figure 24 A). Le PBM du domaine interagit avec le PDZ de façon

canonique. Les résidus directement en amont du PBM se positionnent dans une poche hydrophobe créée à l'interface du PDZ et du SH3. Cette surface d'interaction entre les domaines est présente dans d'autres MAGUK, mais les résidus la formant sont variables, permettant une bonne spécificité de cette poche pour un partenaire [156]. Une autre région en hélice de Crumbs, sept résidus en amont du PBM, se conforme en hélice et interagit avec le site de liaison du domaine GuK, sans phosphorylation, dans ce cas. La multiplication des points d'ancrage permet une spécificité forte de chaque MAGUK pour leurs partenaires. Bien que la structure du supramodule soit conservée, les modalités d'interaction sont différentes [158,159].

Les protéines homologues PTPN4 et PTPN3 possèdent un supramodule formé par un domaine PDZ et un domaine phosphatase actif [60,160]. L'interaction des deux domaines, bien que transitoire, a un effet inhibiteur sur l'activité catalytique du domaine phosphatase. La fixation d'un ligand par le domaine PDZ augmente la réorientation entre les deux domaines en solution. Cette modification de dynamique se traduit par une levée de l'inhibition du domaine phosphatase induite par l'interaction avec le domaine PDZ. Cette régulation pourrait permettre une activation spécifique du domaine catalytique, uniquement en présence d'une protéine partenaire contenant un PBM spécifique de son domaine PDZ [161].

La famille de protéase HtrA inclut des protéines bactériennes et mitochondriales à domaine PDZ [162]. Leur domaine catalytique et un ou plusieurs domaines PDZ associés forment un supramodule (Figure 24 B C). Ce supramodule lui-même trimérise par le domaine catalytique [163]. Dans certaines espèces, un oligomère fonctionnel de 12 ou 24 sous-unités peut se former, maintenant les sites protéolytiques dans une cavité au cœur de l'assemblage [155]. Dans le cas de DegS, le domaine PDZ inhibe la fonction du domaine protéolytique, tandis que la liaison du partenaire au domaine PDZ lève l'inhibition et permet une activité totale du domaine catalytique [164]. Avec les protéines PTPN3 et PTPN4, les HtrA sont un exemple unique de modulation de l'activité catalytique d'une protéine par l'interaction d'un domaine PDZ.

Malgré l'inclusion quasi systématique des domaines PDZ dans des protéines multi-domaines, un nombre très restreint de publications décrit les complexes intramoléculaires que ceux-ci peuvent former. Le faible nombre de complexes supramodulaires décrits reflète probablement plus un manque d'étude sur le sujet plutôt qu'un manque d'occurrence naturelle de tels ensembles.

4 Protéines de la touffe ciliaire

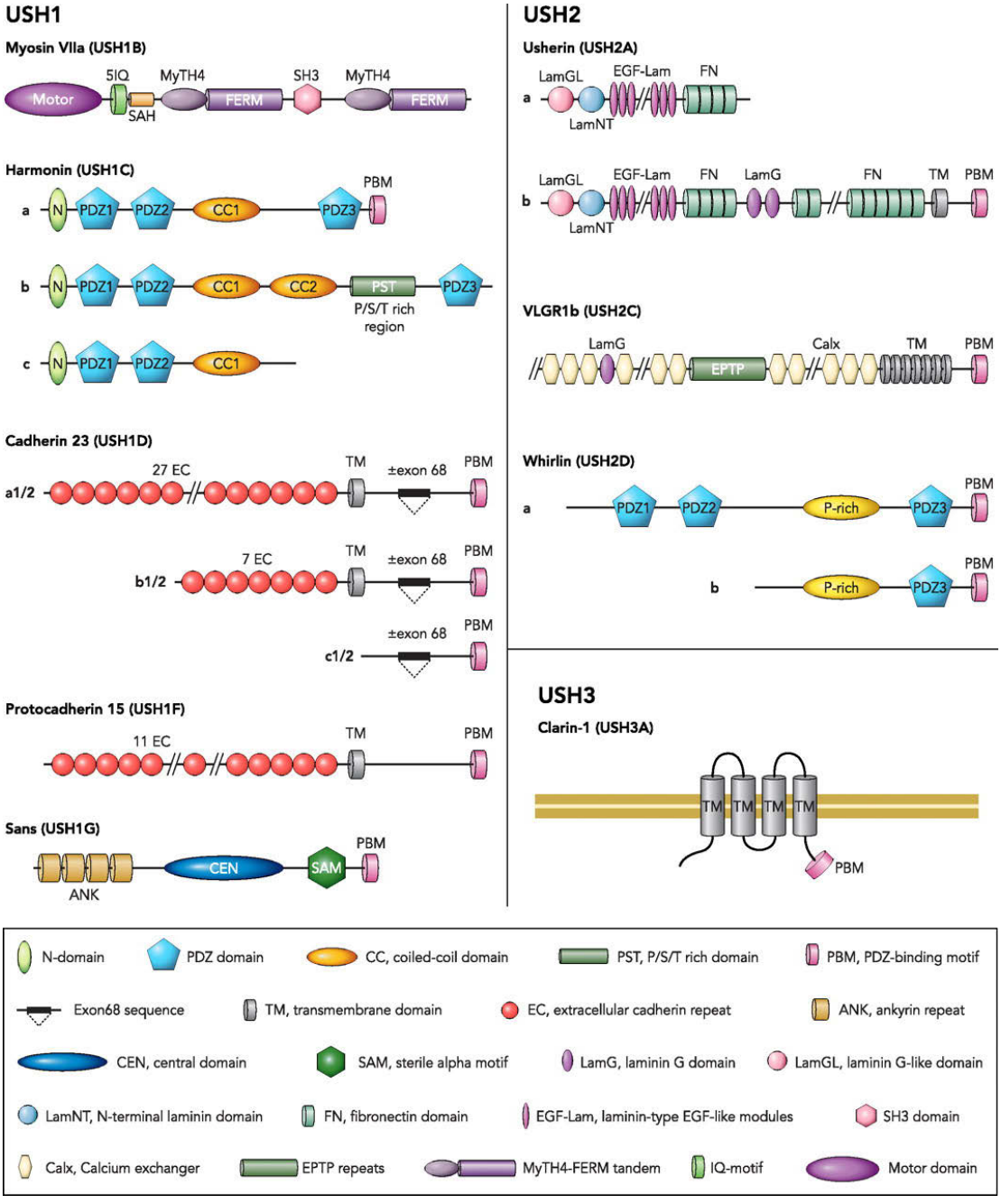


Figure 25 : Organisation en domaines de protéines Usher et de leurs isoformes identifiées. Les protéines Usher1 sont montrées à gauche, les protéines Usher2 et Usher3 sont montrées à droite. La légende des domaines est reportée sous les schémas.

4.1 Généralités

4.1.1 Etude des protéines Usher

Il n'existe actuellement pas de modèle des cellules ciliées et l'étude des protéines Usher *in situ* s'avère particulièrement difficile. La cochlée ne contient que quelques dizaines de milliers de cellules ciliées, dont la touffe ciliaire ne représente qu'une infime fraction du volume total de la cellule. L'ensemble des résultats biologiques sur la fonction des protéines provient donc essentiellement de la localisation, par microscopie confocale et par microscopie électronique, de ces protéines dans les cochlées prélevées chez des souris saines ou knock-out (KO). Les stéréocils sont de petites structures et les technologies d'imagerie n'ont permis que récemment une localisation précise des protéines Usher le long de leurs axes. La détermination de la fonction des protéines Usher et de leurs protéines associées est donc difficile. En plus des difficultés techniques, les protéines Usher sont exprimées en plusieurs isoformes, impliquées dans de multiples fonctions parfois difficiles à distinguer. Par ailleurs, la plupart de ces protéines participe au développement des touffes ciliaires ; leurs knock-out résultent bien souvent dans des touffes ciliaires morphologiquement aberrantes, rendant difficile la détermination de leurs fonctions dans les stéréocils matures.

4.1.2 Caractéristiques structurales

Les protéines Usher et leurs protéines associées sont des protéines multi-domaines sans fonction catalytique. Elles contiennent généralement plusieurs domaines d'interaction protéine-protéine leur permettant de former des complexes multi-protéines solidement connectés. De manière générale, on peut séparer les protéines Usher en trois catégories : les protéines membranaires inter-ciliaires, comportant des motifs extracellulaires répétés d'interaction et une région cytoplasmique courte et non-structurée ; les protéines cytoplasmiques adaptatrices, possédant de multiples modules d'interaction séparés par de grandes régions non-structurées ; enfin les myosines atypiques, interagissant avec l'actine et comportant plusieurs domaines d'interaction protéine-protéine (Figure 25).

Les protéines Usher et leurs protéines associées possèdent systématiquement un court motif d'interaction aux domaines PDZ à leur C-terminal, mis à part la Myosine 7a [165]. Cependant, seules trois protéines Usher et Usher associées, possèdent des domaines PDZ en copies multiples, ces protéines sont donc présentes dans plusieurs complexes impliquant divers partenaires (Figure 26).

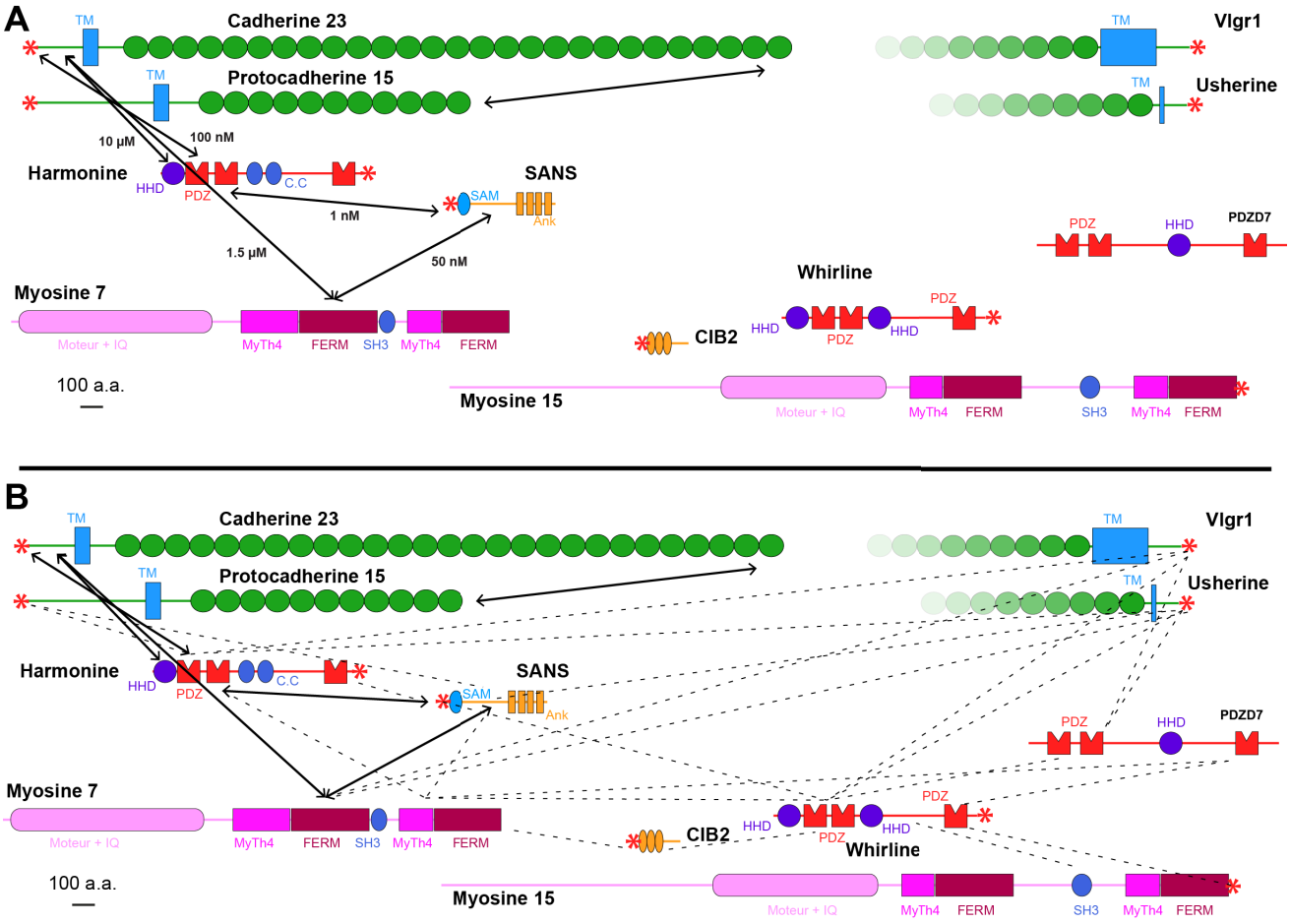


Figure 26 : Interactions entre protéines Usher et protéines associées. (A) Interactions confirmées par des structures de complexes, l'affinité (Kd) est indiquée si l'interaction a été quantifiée. L'astérisque rouge symbolise les PBM. Les parties extracellulaires de Vlgr1 et Usherine ne sont pas montrées entièrement à cause de leur longueur. **(B)** Ensemble des interactions reportées dans la littérature entre protéines Usher et protéines associées.

4.2 Protéines Usher1

4.2.1 Cadhérine-23 (USH1D – DFNB12)

La cadhérine-23 est une protéine d'adhésion cellulaire dépendante au calcium. Sa partie extracellulaire est particulièrement longue et contient 27 domaines cadhérines répétés [166]. Les domaines cadhérines répétés possèdent des sites de liaison aux ions calcium à l'interface entre deux domaines. Le calcium est indispensable pour rigidifier cette interface et confère la conformation étendue linéaire de la partie extracellulaire mesurant 130 nm [167]. Le segment extracellulaire de la cadherine-23 dimérise sur toute sa longueur, comme l'attestent de nombreux clichés de microcopie électronique [168]. Le reste de la protéine est composé d'un

segment transmembranaire et d'une partie cytoplasmique non structurée d'environ 150 résidus, terminée en son extrémité C-terminale par un PBM de type 1 (-ITEL).

La cadherine-23 est exprimée en deux isoformes dans l'audition dont la seule différence est la présence, ou non, de la séquence protéique correspondant à un exon. Pendant le développement de la touffe ciliaire, les deux isoformes sont exprimées indifféremment, tout le long des stéréocils, où elles sont impliquées dans la formation des liens latéraux et des liens kinociliaires. Au stade mature, la cadherine-23 est localisée à l'apex des stéréocils où elle forme la partie supérieure du lien apical, senseur de la mécano-transduction [167].

4.2.2 Protocadhérine-15 (USH1F-DFNB23)

La Protocadhérine-15 est aussi une protéine membranaire d'adhésion cellulaire. Sa partie extracellulaire est composée de 11 domaines cadhérine répétés, qui, comme la cadhérine-23, forme un dimère sur toute sa longueur et possède une longueur maximale d'environ 50 nm. Les sites inter-domaines de liaison au Ca^{2+} ne sont pas conservés entre toutes les répétitions du segment extracellulaire, lui conférant une certaine flexibilité [169]. La protocadhérine-15 est exprimée en 24 isoformes classées en trois groupes principaux, se distinguant par la séquence et la longueur de leur région cytoplasmique ; on les nomme CD1, CD2 et CD3. Ces isoformes ont toutes un segment cytoplasmique non-structuré, chacun terminé par un PBM différent de type 1 (-STSL, -NTAL, -MTKL). Les trois isoformes ont des localisations différentes ; CD1 est localisée tout le long des stéréocils, CD2 à l'apex des rangées de petites tailles et CD3 à l'apex de toutes les rangées.

Pendant le développement, les différentes isoformes de la protocadhérine-15 entrent dans la composition des liens latéraux, des liens kinociliaires et des liens apicaux. Au stade mature, l'isoforme CD2 forme la partie inférieure du lien apical et est directement reliée à la machinerie mécano-sensible membranaire [28].

4.2.3 Lien apical

Le lien apical mature est formé par un complexe de Cadhérine-23 et de Protocadhérine-15 CD2 [170]. *In vitro,* les parties extracellulaires des deux protéines forment des homodimères sur l'ensemble de leur longueur. La dimérisation est confirmée dans les touffes ciliaires, où il n'est pas rare d'observer un lien apical se séparant en deux avant le point d'insertion membranaire [168]. Les deux protéines homodimériques interagissent par leurs premiers domaines cadhérines répétés (Figure 27 A), formant un filament extracellulaire d'environ 180 nm. Même dans les touffes ciliaires aux repos, le lien apical est sous tension, comme en témoigne la déformation de la membrane au niveau du point d'ancrage inférieur (**Figure 27 C D**).

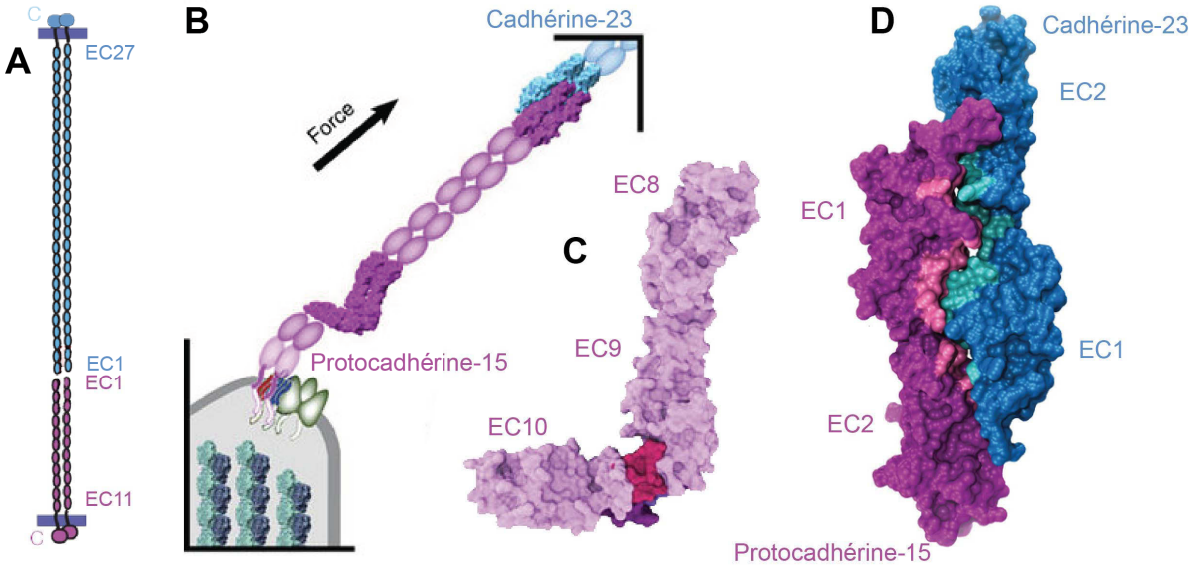


Figure 27 : Structure du lien apical. A Schéma montrant l'organisation en domaine de la Cadhérine-23 et de la Protocadhérine-15, formant le lien apical. **B** Structures connues des domaines cadhérines répétés de la Protocadhérine-15 reportées sur un schéma de lien apical. **C** structure des domaines répétés 8 à 10 de la Protocadhérine-15 (pdb :4XHZ). **D** Structure de complexe formé par les domaines répétés 1 et 2 de la Protocadhérine-15 et de la Cadhérine-23 (pdb :4AQ8)**.** Figures adaptées de Sotomayor *et al.*[171] et de Araya-Secchi *et al.* [169].

Les deux premiers domaines cadhérine répétés de chaque protéine forment un complexe dont la structure a été résolue par cristallographie (Figure 27 D) [171]. L'interaction est classique pour un couple de cadhérine, en "poignée de main", c'est à dire avec le premier domaine d'une protéine se logeant dans l'espace formé entre les deux domaines répétés de la seconde, et vice-versa. Ce complexe forme une grande surface d'interaction (>1000 $Å^2$) avec une affinité de l'ordre du micromolaire. Les structures montrent une interaction entre monomère de chaque protéine, inconsistantes avec des homodimères. Les auteurs proposent

que les derniers domaines EC de chaque protéine n'homodimèrisent pas, séparant les deux protomères au niveau du point d'interaction entre Cadhérine-23/Protocadhérine-15, où celles-ci formeraient deux complexes indépendants, multipliant ainsi l'affinité. Par dynamique moléculaire, les auteurs montrent que la résistance de ce complexe aux tensions est compatible avec la tension subie par le lien apical lors de la déflection des stéréocils [171].

Les études biophysiques de la machinerie de mécano-transduction ont montré que le système nécessite un ressort moléculaire entre le lien apical et le canal mécanosensible [29]. Le lien apical fait lui-même office de ressort dans le système [169,172]. Certains domaines répétés de la Protocadhérine-15 n'ont pas de sites de liaison inter-domaine au Ca^{2+}, augmentant la liberté conformationnelle de ces régions. Notamment, la répétition 9-10 de la Protocadhérine-15, proche du segment transmembranaire, ne fixe pas le calcium et adopte une conformation non étendue en solution (Figure 27 B C). Les modélisations d'application de force en dynamique moléculaire indiquent que la force nécessaire pour étendre la conformation de ce segment est supérieure à la tension de repos appliquée au lien apical. Le mouvement du stéréocil induirait donc une extension de cette région qui agirait comme un ressort avec une amplitude de 5 nm. La résistance de ce ressort est cependant supérieure à celle mesurée sur les touffes ciliaires, il est donc complété par un autre mécanisme encore inconnu. Enfin, la longueur du lien mesurée pour une touffe ciliaire complétement défléchie est proche de 200 nm, soit 20 nm plus grande que la longueur du lien apical au repos. Il est donc probable que certains domaines cadhérine se déplient complétement sous une forte tension.

4.2.4 Myosine-7a (USH1B – DFNA11 – DFNB2)

La Myosine-7a est une myosine non-conventionnelle de 2215 résidus, faisant partie du groupe des MyTH4-FERM. En son N-terminal, elle contient un domaine moteur responsable de la liaison à l'actine et du mouvement le long de celle-ci. Cette partie effectrice est suivie d'un "bras" formé d'une succession de cinq motifs de liaison à la calmoduline IQ répétés, formant chacun une hélice amphiphile de sept tours. Cette région sépare le domaine moteur de sa région d'interaction avec le cargo. Dans la Myosine-7a, la répétition de motifs IQ est directement suivie d'un motif SAH, un autre type d'hélice isolée stable. La succession d'hélices isolées crée un bras rigide d'environ 28 nm entre le moteur et les domaines suivants [173]. A la différence des myosines classiques, la partie C-terminale de la protéine n'est pas constituée d'une région coiled-coil de dimérisation mais par deux couples de domaines

MyTh4/FERM d'interaction avec les protéines partenaires, séparés par un domaine SH3 [174]. C'est la seule protéine Usher ne possédant pas de PBM C-terminal.

Isolée en solution, la Myosine-7a est monomérique, elle ne peut donc pas se déplacer le long de l'actine dans ces conditions. De plus, sa cinétique de libération de l'ADP est particulièrement lente, favorisant l'interaction longue de la protéine avec l'actine [175]. De part ces deux caractéristiques, la fonction proposée de la Myosine-7a est un rôle d'ancrage moléculaire, notamment au niveau du point supérieur d'insertion du lien apical, comme nous le verrons dans la partie **4.2.8**. La dimérisation de la Myosine-7a lui permet de se déplacer le long de l'actine [175]. Son déplacement, et celui d'autres Myosines MyTh4/FERM, sont contrôlés par l'interaction avec un partenaire cargo qui induit leur dimérisation [174,176]. Cette capacité particulière pourrait être essentielle pour la bonne localisation des différents complexes le long des stéréocils.

Dans les stéréocils, la Myosine-7a interagit avec SANS, l'Harmonine, PDZD7 et la Whirline. Elle a un rôle essentiel dans l'adaptation lente et dans la localisation des protéines cytoplasmiques des liens basaux [177–179].

4.2.5 SANS (USH1G)

SANS est une petite protéine d'échafaudage de 460 acide aminés. Elle contient en son N-terminal, trois domaines ankyrine répétés dont la fonction est inconnue. Le centre de la protéine est à priori non-structuré sur 250 acides aminés contenant plusieurs motifs linéaires d'interaction [180]. La protéine possède un domaine SAM, domaine globulaire de 70 résidus formé par cinq hélices α, suivi directement par un PBM de type 1 (-DTEL) [181]. Les domaines SAM sont connus pour oligomériser en solution, et pour interagir avec d'autres domaines protéiques. *In vitro,* le domaine SAM de SANS oligomérise, mais cette tendance n'a pas été confirmée *in vivo* [71].

SANS est une protéine pivot dans l'organisation des complexes sous-membranaires impliqués dans la structure et la fonction des touffes ciliaires et des photorécepteurs. Elle interagit avec la majorité des protéines Usher1 et Usher2. Dans les stéréocils matures, SANS se localise à l'apex des stéréocils de la petite et moyenne rangée où elle entre probablement en jeu dans les complexes de la mécano-transduction au point inférieur d'ancrage et au point supérieur d'ancrage du lien apical pour l'adaptation lente [182]. SANS est aussi impliquée dans la localisation du complexe des liens basaux pendant le développement [179]. Dans les photorécepteurs, SANS a été décrite à la fois dans les calyceal processes, aux côté des autres

protéines Usher1 [45] et dans le cil connecteur, au niveau de la localisation des protéines Usher2 [183].

4.2.6 Harmonine

L'Harmonine est une protéine d'échafaudage de 910 résidus comprenant plusieurs domaines d'interaction protéine-protéine [184]. Elle est très étudiée structuralement. En son N-terminal, l'Harmonine contient un domaine HHD [64] suivi par un premier domaine PDZ, possédant une extension C-terminale structurée en épingle-ß. Séparé par un lien peptidique de 20 résidus, suit le deuxième PDZ. Ce domaine PDZ possède lui aussi une extension C-terminale, mais formée, elle, par deux tours d'hélice α, en interaction avec les brins ß2-ß3 du domaine. Suivent deux régions coiled-coil de structure et de fonction inconnues et d'une grande région riche en Proline-Sérine-Thréonine de fixation à l'actine [166]. En C-terminal, se situe le troisième domaine PDZ directement suivi d'un PBM de type 1 (-LTFF).

L'Harmonine est exprimée en trois isoformes : l'isoforme-b correspond à la protéine entière décrite dans le paragraphe précédent ; l'isoforme-a ne contient pas la région PST et le second coiled-coil ; et enfin l'isoforme-c est tronquée après le PDZ2 (Figure 25). Dans les touffes ciliaires matures, l'isoforme b est spécifiquement localisée au niveau du point supérieur d'ancrage du lien apical [185]. Pendant le développement, l'Harmonine, sans précision d'isoforme, est localisée tout le long des stéréocils.

En plus de son rôle dans la rétine et la cochlée, l'Harmonine est un élément d'échafaudage important des microvillosités de l'épithélium intestinal. Dans ce système, l'Harmonine est au centre d'une interaction formée par des protéines similaires aux protéines Usher1 : deux protéines de jonction extracellulaire assimilées à des cadhérines (CdhR2, CdhR5), un paralogue de SANS (Anks4b) et un paralogue de la Myosine-7a (Myosine-7b). Le complexe ainsi formé est identique au complexe Usher1 [186].

4.2.7 Supramodule hétérotypique de l'Harmonine et de SANS

Depuis son identification [166,184], l'Harmonine a été largement étudiée tant au niveau de sa fonction que de sa structure. L'équipe du Pr. Mingjie Zhang, spécialisée dans les interactions non-canoniques des domaines PDZ, a beaucoup contribué à la compréhension des bases moléculaires de la fonction de cette protéine.

Le domaine HHD N-terminal, le PDZ1 et son extension forment un supramodule (Figure 28) [71]. L'extension C-terminale du domaine PDZ forme un mini-domaine de 24

résidus composé d'une épingle-ß et d'une hélice courte C-terminale. Le domaine HHD et le PDZ de l'Harmonine n'interagissent pas directement, l'ensemble des contacts implique l'épingle-ß de l'extension. Ce supramodule hétérotypique interagit avec le PBM de la protéine SANS. Cette interaction canonique PDZ/PBM est complétée par un large contact non-canonique entre une surface exposée du PDZ1 de l'Harmonine et le domaine hélical SAM de SANS, cinq résidus en amont de son PBM C-terminal. L'extension du domaine PDZ interagit aussi avec les chaines latérales des résidus en amont du PBM de la protéine partenaire, agrandissant d'autant la poche d'interaction (Figure 28). Le domaine SAM tronqué du PBM C-terminal lie le PDZ1 de l'Harmonine avec une faible affinité (Kd = 44 µM) alors que le PBM seul interagit, lui, avec le PDZ1 avec une affinité de l'ordre du micromolaire (Kd = 1.2 µM). Ces deux points indépendants de liaison entre SANS et l'Harmonine permettent de former un complexe stable de très haute affinité (Kd = 1 nM) .

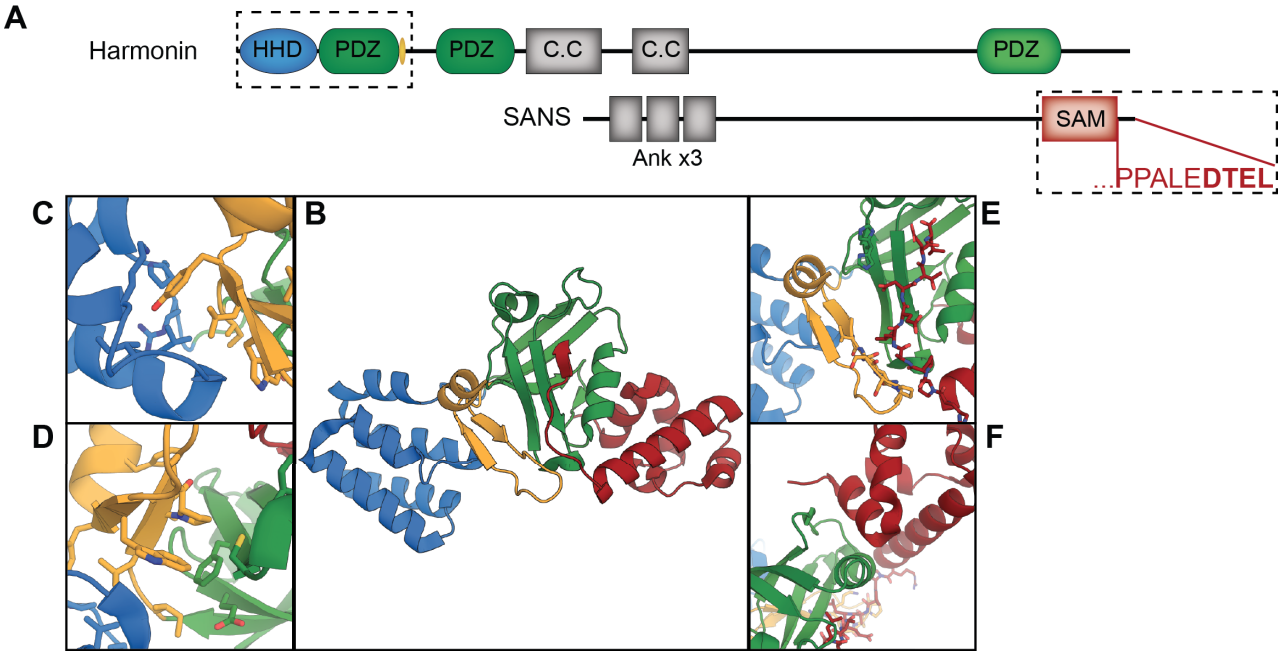


Figure 28: Supramodule hétérotypique de l'Harmonine. A Architecture de l'Harmonine et de SANS. Les régions impliquées dans le supramodule sont entourées en pointillé, la séquence du PBM C-terminal de SANS est indiquée. **B** Représentation en ruban du supramodule, contenant le HHD (bleu), le domaine PDZ (vert), l'extension du domaine PDZ (Orange), le domaine SAM et le PBM de SANS (rouge). **C** Agrandissement de l'interface entre l'extension du domaine PDZ et le HHD. **D** Agrandissement de l'interface entre le domaine PDZ et son extension C-terminale. **E** Agrandissement de l'interface entre l'extension du domaine PDZ et SANS. **F** Agrandissement de l'interface entre le domaine PDZ et SANS.

Il est important de préciser que pour obtenir ce complexe, les auteurs ont utilisé des mutants ponctuels de leurs domaines. Les séquences naturelles des deux protéines Harmonine et SANS forment des homo-oligomères. Afin de diminuer leur auto-association et obtenir des

cristaux, les auteurs ont inversé la charge d'une Lysine du domaine PDZ de l'Harmonine (K157E) et du domaine SAM de SANS (K437E). .

4.2.8 Point supérieur d'ancrage du lien apical

Les bases structurales des complexes formés par les protéines Usher1 ont été largement explorées, donnant une bonne vision de leur fonctionnement au niveau du point supérieur d'insertion des stéréocils, depuis le lien apical jusqu'au cytosquelette d'actine [165] (Figure 29).

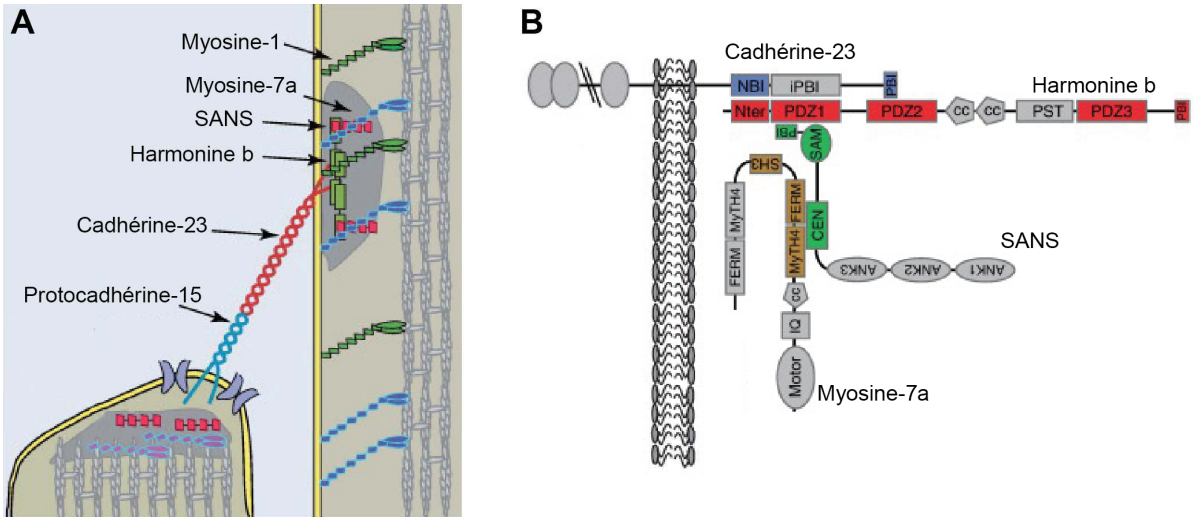


Figure 29: Interactions entre protéines Usher1. A Schéma de l'organisation des protéines Usher1, au niveau du point supérieur d'ancrage du lien apical **B** Schéma résumant, par domaines, les interactions identifiées par méthodes biophysiques entre protéines Usher1**.** Figure adaptée de Kazmierczak *et al.* [187]

L'équipe du Pr. Zhang a identifié la grande majorité des interactions entre protéines Usher1. Ils ont d'abord montré l'interaction multiple de la partie cytoplasmique de la Cadhérine23, la protéine de lien apical, avec l'Harmonine [64,68]. La Cadhérine-23 interagit par son motif PBM C-terminal avec le PDZ2 de l'Harmonine. Dans sa région cytoplasmique, la Cadhérine-23 possède également deux motifs distincts d'interaction avec le domaine HHD N-terminal de l'Harmonine. Ces deux motifs peuvent interagir de façon indifférenciée avec le HHD de l'Harmonine mais ne lient pas les domaines HHD d'autres protéines [64]. Les trois points d'ancrage distincts de la Cadhérine-23 sur une molécule d'Harmonine permettraient une interaction forte capable de résister aux tensions subies, lors de la déflection des stéréocils. Les auteurs proposent que la présence de deux motifs d'interaction au HHD provoque l'accumulation de l'Harmonine en dessous du lien apical par le branchement de plusieurs Harmonine sur une même région cytoplasmique de la cadherine23 [68]. Le rôle de cet

assemblage serait de diffuser les forces appliquées au lien apical, le long de la membrane. Cette théorie reste cependant toujours à vérifier expérimentalement.

Les mêmes auteurs ont complété ce travail sur les protéines Usher1 par la recherche d'éléments connectant les protéines Usher1 au cytosquelette. Deux segments conservés et non-structurés de la région centrale de SANS interagissent avec le premier domaine FERM de la myosine7 avec une affinité micromolaire (Kd = 2μM). La structure du complexe a été résolue [188].

De façon remarquable, l'ensemble des données structurales sur les protéines Usher1 montre un schéma clair et non-contradictoire de la formation de leur large complexe multiprotéine. Plusieurs motifs de chaque composante interagissent faiblement avec plusieurs domaines d'un partenaire, créant ainsi des contacts sur plusieurs points d'ancrage, conférant des interactions finales de fortes affinités, permettant de maintenir le lien apical fortement ancré au cytosquelette.

En plus de leur rôle évident dans l'ancrage du lien apical au cytosquelette, les protéines Usher1 joueraient aussi un rôle dans l'adaptation active. Dans ce modèle, la déflection de la touffe ciliaire et/ou l'ouverture du canal mécanosensible induit le déplacement du point supérieur d'ancrage vers le bas pour réduire la tension au lien apical et baisser la probabilité d'ouverture du canal [189]. Il n'existe actuellement aucune preuve moléculaire d'un tel mécanisme chez les mammifères et son existence est de plus en plus remise en cause [190].

4.2.9 CIB2 (USH1J, DFNB48)

La protéine CIB2 est à part dans les protéines Usher1, ne se localisant pas au niveau du point supérieur d'insertion. C'est une petite protéine de 180 acides aminés dont le tiers N-terminal ne contient pas de domaine identifié mais est toutefois prédit comme ordonné (IUpred). Le reste de la protéine est constitué de trois domaines main-EF de liaison au calcium. Seuls les domaines EF2 et EF3 sont capables de lier un ion Ca^{2+}. CIB2 possède un PBM de type 2 en son C-terminale (-HIRI), qui est inaccessible si le domaine EF3 est replié en présence de calcium.

CIB2 est localisée tout le long des stéréocils et préférentiellement à leur apex. Récemment identifiée, son rôle dans les cellules ciliées est mal compris ; elle pourrait être impliquée dans la signalétique cellulaire après ouverture du canal de mécano-transduction [191].

4.3 Protéines membranaires et canal mécanosensible

Les protéines formant la machinerie mécanosensible au niveau du point inférieur d'ancrage du lien apical sont mal connues, mais de nombreux progrès ont été faits ces trois dernières années. Un débat est toujours ouvert sur les protéines composant le canal mécano-transducteur [192,193], mais leur identification non-équivoque semble plus proche que jamais.

4.3.1 LHFPL5 (DFNB66/67)

LHFPL5 est une protéine membranaire de 200 résidus composée de quatre segments transmembranaires avec les segments N-terminal et C-terminal du côté cytoplasmique. Le segment cytoplasmique C-terminal se termine par un possible PBM de type 3 (-NEEV).

Cette protéine n'est pas nécessaire pour le développement des stéréocils. Elle est cependant indispensable à l'intégrité du lien apical et à la mécano-transduction [194]. Le dernier segment cytoplasmique de LHFPL5 interagit avec la Protocadhérine-15 par son segment transmembranaire ainsi que par une région cytoplasmique proche de la membrane, conservée dans les trois isoformes (Figure 30).

4.3.2 TMIE (DFNB6)

TMIE est une protéine membranaire de 156 résidus contenant deux segments transmembranaires. Le segment cytoplasmique C-terminal de 78 acides aminés est particulièrement chargé positivement, avec plus de 25% de sa séquence composée de lysines [195] . De façon étonnante, les trois mutations connues donnant lieu à une surdité non-syndromique, induisent toutes des substitutions d'arginine en résidu hydrophobe, dans ce segment cytoplasmique C-terminal.

TMIE interagit à la fois avec LHFPL5 et avec la région cytoplasmique de l'isoforme CD2 de la Protocadhérine-15 (Figure 30). L'injection de la région C-terminale de TMIE dans les cellules ciliées en culture abroge l'entrée de cation dans ces cellules, montrant que l'interaction entre les régions cytoplasmiques de TMIE et de Pcdh-15 CD2 est critique pour la mécano-transduction [195].

4.3.3 TMC1 (DFNB7/11 ; DFNA36) & TMC2

TMC1 et TMC2 (Trans-Membrane, Channel-like) sont deux protéines membranaires contenant chacune six segments transmembranaires. Une isoforme de TMC2 contient un

PBM de type 1 à son C-terminal (-VSAL). Plusieurs mutations de TMC1, mais aucune de TMC2, sont associées à des surdités.

Ces deux protéines sont suspectées depuis des années de former, au moins en partie, le canal de mécano-transduction [196]. La suppression d'une seule de ces protéines n'a pas d'effet sur la mécano-transduction, alors que celle simultanée de TMC1 et TMC2 empêche totalement l'entrée de cations au sommet des stéréocils, lors d'une stimulation. Cette redondance a rendu difficile l'identification de ces protéines comme canaux potentiels. TMC1 et TMC2 sont détectées au niveau du point inférieur d'insertion du lien apical, aux côtés de TMIE et de LHFPL5 [197]. TMC1 interagit avec LHFPL5 ainsi qu'avec la Protocadhérine-15, et TMC2 avec la Protocadhérine-15 [198] (Figure 30).

Le débat fut longtemps de savoir si les protéines TMC composent elles-mêmes le canal ou si celles-ci sont des sous-unités associées à la chaine de transmission de la force [192]. Ces doutes reposaient notamment sur la présence de canaux mécanosensibles additionnels à la surface de la cellule ciliée. Ces doutes ont été dissipés dans un article récent, montrant que ce canal additionnel est formé par la protéine Piezo-2 [199].

Les protéines connectant le complexe membranaire transducteur de la tension au cytosquelette restent inconnues. De par leurs localisations, CIB2 et la Myosine-15a ont été proposées, mais leur implication reste à prouver de façon directe [200].

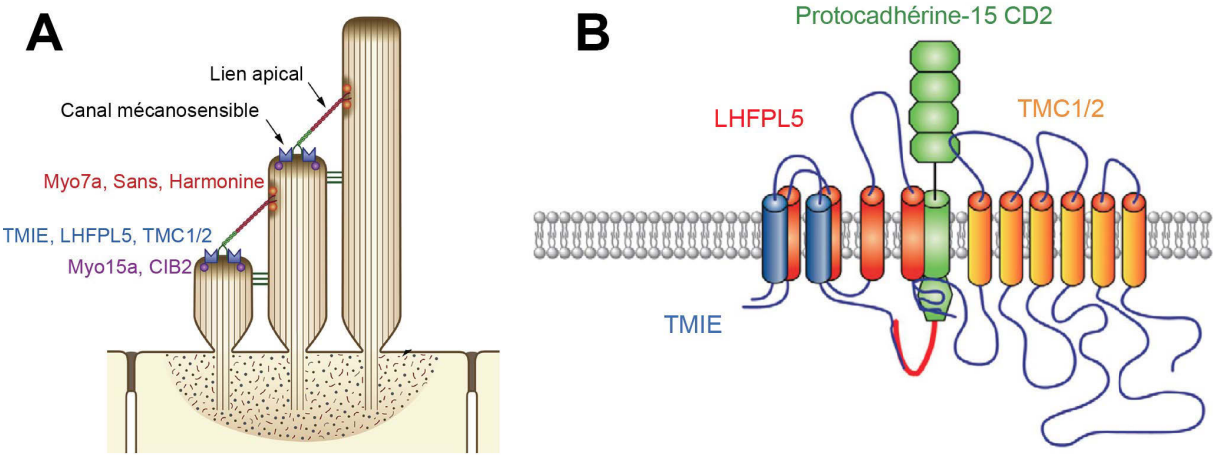


Figure 30: Organisation du complexe du point inférieur d'ancrage. A Représentation schématique d'une touffe ciliaire résumant les protéines impliquées dans la mécano-transduction. **B** Représentation schématique du complexe formé par les protéines membranaires du point inférieur d'ancrage.

4.4 Protéines Usher2 et protéines associées

4.4.1 Usherine (USH2A, RP39)

L'Usherine est une très grande protéine membranaire, contenant une région extracellulaire de 5000 résidus, un segment transmembranaire et une région cytoplasmique de 140 acides aminés. Sa partie extracellulaire est composée de 13 domaines associés aux laminines et de 36 domaines Fibronectines de type-3, des domaines associés aux interactions inter-cellule et cellule/matrice. La partie cytoplasmique est à priori non structurée et possède un PBM de type 1 à son C-terminus (-DTHL).

Les mutations du gène associé à l'Usherine sont responsables de la majorité des cas Usher [201], cependant sa fonction reste élusive. L'Usherine se localise à la base des stéréocils dans les premiers jours postnataux chez la souris, correspondant à la période de formation des liens basaux [202]. Elle est aussi exprimée dans le kinocil pendant le développement [203]. Les K.O variés des protéines formant le complexe basal confirment l'implication de l'Usherine dans ces liens [33].

4.4.2 Vlgr1 (USH2C)

Vlgr1 est une très grande protéine G, avec une partie extracellulaire composée de 5880 acides aminés, sept segments transmembranaires et une région C-terminale cytoplasmique de 150 résidus [204]. Sa région extracellulaire contient 35 domaines de liaison au calcium Calx-ß et 6 domaines répétés EAR, impliqués dans les interactions protéine-protéine et associés à l'épilepsie. La région cytoplasmique n'est à priori pas structurée, mis à part une hélice hydrophobe potentielle. Vlgr1 se finit par un PBM identique à celui de l'Usherine (-DTHL).

Vlgr1 est exprimée dans de nombreux tissus et tout particulièrement dans le cerveau. Son K.O augmente drastiquement l'apparition de crise d'épilepsie induite par le son chez la souris, soulignant un rôle probable dans le développement des neurones inhibiteurs [205]. Il induit aussi une malformation des touffes ciliaires des CCE qui perdent leur forme en "V" [203]. Tout comme l'Usherine, Vlgr1 a été identifiée à la base des stéréocils en développement et est impliquée dans les liens basaux [206]. Sa partie extracellulaire n'a pas de ligand connu et son implication dans un éventuel chemin de transduction associé aux protéines G n'est pas déterminée [207].

4.4.3 PDZD7 (USH modifier ; DFNB à attribuer)

PDZD7 est une protéine d'échafaudage de 1000 résidus, paralogue à l'Harmonine et à la Whirline, mais moins étudiée que celles-ci. La partie N-terminale contient deux premiers domaines PDZ proches dans la séquence. Le centre de la protéine est principalement constitué de deux grandes régions non-structurées entourant un domaine HHD isolé. Enfin, un troisième domaine PDZ est localisé proche du C-terminal de la protéine, qui ne contient pas de PBM. Deux isoformes N-terminales courtes de la protéine sont décrites : elles tronquent toutes deux la protéine dans la région non-structurée séparant le PDZ2 et le HHD et ne contiennent donc que les deux PDZ N-terminaux.

PDZD7 a été décrite à l'origine comme un "modificateur" du syndrome Usher2 [208]. Dans une famille atteinte du syndrome USH2 dû à une mutation de l'Usherine, les membres porteurs d'une mutation additionnelle sur le gène associé à PDZD7 ont des symptômes aggravés comparés aux membres non-porteurs. Récemment, une mutation créant une troncature de PDZD7 après le domaine HHD a été identifiée comme cause d'une perte non-syndromique récessive d'audition (DFNB), un numéro n'a cependant toujours pas été attribué au *locus* [209]. PDZD7 localise principalement, voir uniquement, au niveau des liens basaux, où la présence de son isoforme longue est nécessaire à la localisation des autres composantes du complexe [177]. Le K.O de PDZD7 induit une perte d'audition chez les souris, notamment en perturbant la morphologies des touffes ciliaires des CCE [210].

4.4.4 Myosine 15a (DFNB3)

La Myosine-15a est une myosine non-conventionnelle de 3500 résidus similaire à la Myosine-7a, faisant partie du groupe des MyTH4-FERM. Elle a la particularité d'avoir une région majoritairement non structurée de 1200 résidus en N-terminal de son domaine moteur. Ce dernier est suivi de seulement deux répétitions IQ, lui donnant un bras levier plus court que la Myosine-7a. La partie C-terminale de la protéine est composée de tandems MyTh4/FERM séparés par un domaine SH3, comme pour la Myosine-7a. Le dernier domaine FERM est à l'extrémité C-terminale de la protéine, et se termine par un PBM de type 1 (-ITLL).

Pendant le développement et au stade mature, la Myosine-15a est spécifiquement localisée au sommet de la grande rangée de stéréocils des touffes ciliaires. A cette localisation, la Myosine-15a est impliquée dans un complexe protéique comprenant entre

autres la Whirline et Eps8. Ce complexe est essentiel à l'élongation des stéréocils pendant le développement et à leur maintien en longueur au stade mature [211].

Comme la Myosine-7, la Myosine-15a ne contient pas de domaine de dimérisation, mais se déplace le long de l'actine dans sa forme dimérique, induite par une interaction spécifique avec son cargo [212]. La grande région en amont du domaine moteur n'a pas de fonction connue, cependant elle est indispensable à une fonction normale des cellules ciliées [213].

4.4.5 Whirline (USH2D, DFNB31)

La Whirline est une protéine cytoplasmique d'échafaudage paralogue à l'Harmonine et à PDZD7. La moitié N-terminale de la Whirline contient deux domaines HHD entourant les deux premiers domaines PDZ. S'ensuit une grande région riche en proline, allant jusqu'au troisième domaine PDZ qui est suivi par un PBM de type 2 en son C-terminal (-NVML).

La Whirline est exprimée en trois isoformes principales: la forme longue exprimée dans les liens basaux et les photorécepteurs. Une forme courte N-terminale contenant les deux domaines HHD et les deux premiers domaines PDZ est exprimée uniquement dans les photorécepteurs. Une forme courte C-terminale inclut le second HHD, la région riche en Proline, le domaine PDZ3 et le PBM C-terminal. La littérature portant sur la Whirline sera discutée en détails dans le chapitre suivant.

4.4.6 Liens basaux

Les liens basaux sont créés de façon transitoire pendant le développement des touffes ciliaires de mammifères. Dans toutes les autres espèces, les liens basaux sont présents au stade mature. Ces liens sont impliqués dans la morphologie des touffes ciliaires [33]. Les trois protéines Usher2 et PDZD7 forment un complexe au niveau des liens. Les deux protéines membranaires Vlgr1 et Usherine sont de bons candidats pour composer ces liens. La protéine Vlgr1 est largement plus abondante que l'Usherine dans la région basale et son K.O empêche la localisation des trois autres protéines du complexe dans la région basale. Vlgr1 est donc actuellement le meilleur candidat pour la composition des liens basaux [33,203].

La Whirline et PDZD7 sont toutes deux capables d'interagir avec les PBM identiques des deux protéines membranaires (Figure 21) [214]. Les interactions entre ces protéines formant le complexe des liens basaux seront détaillées au cours du chapitre suivant.

Il apparaît de plus en plus évident que les liens basaux sont localisés et maintenus à leur position par la Myosine-7a [177,179,203]. La myosine-7a interagit avec l'ensemble des protéines Usher2.

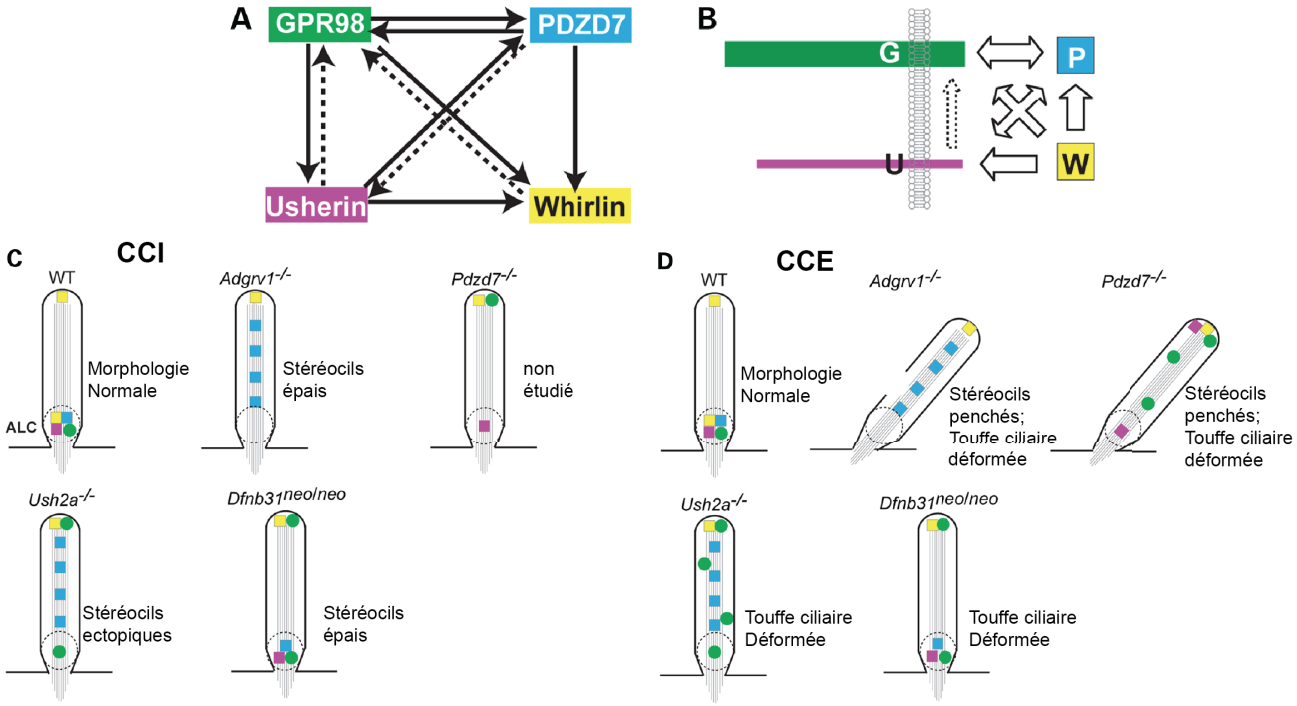


Figure 31: Organisation du complexe des liens basaux. A Interdépendance de localisation entre protéines du complexe des liens basaux. Les lignes pleines indiquent une dépendance complète, les lignes pointillées indiquent une dépendance partielle. **B** Interactions identifiées (flèches pleines) ou hypothétiques (flèches en pointillé) entre protéines du complexe des liens basaux. **C et D** Distribution des protéines du complexe des liens basaux après K.O de l'une d'entre elles, chez les CCI et les CCE, les conséquences sur la morphologie des touffes ciliaires sont indiquées. Figure adaptée de Zou *et al.*[33]

5 La Whirline, acteur clé du développement et de la structure des touffes ciliaires.

Le travail fourni dans cette thèse est concentré sur la Whirline, une protéine impliquée dans les processus moléculaires de l'audition et de la vision. Cette protéine est extrêmement bien conservée entre l'homme et la souris tant en séquence (84% d'identité sur l'ensemble de la séquence) qu'en fonction. Dans ce travail, la séquence référence utilisée est l'isoforme 3 de souris (Uniprot Id: Q80VW5-3), la plus proche de la séquence humaine.

Cette grande protéine de 907 résidus a de nombreux partenaires cellulaires avec lesquels elle interagit de façon transitoire au sein de différents complexes multiprotéines. Les isoformes de la Whirline sont impliquées dans différentes fonctions, à des stades de développement et à des localisations distinctes, dans les cellules ciliées et les cellules photoréceptrices.

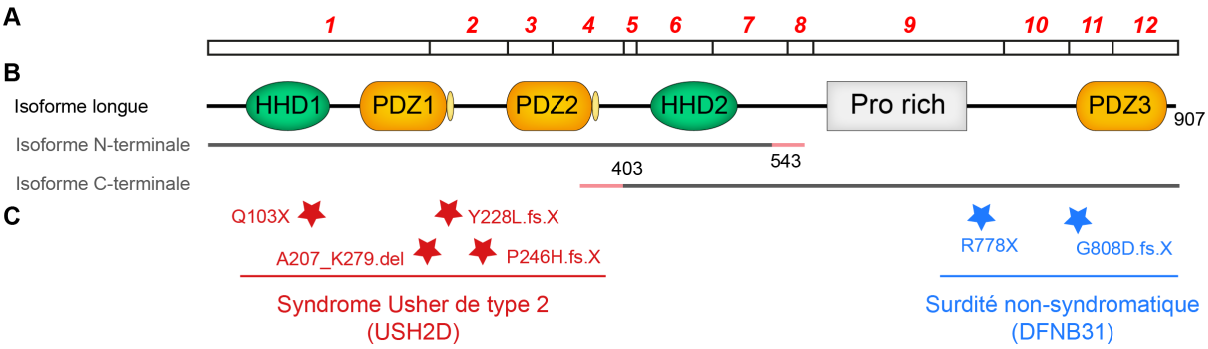


Figure 32: Isoformes et mutations symptomatiques de la Whirline. A Schéma de la séquence génomique de la Whirline. Les exons sont numérotés en rouge, les régions de la protéine correspondant à chaque exon sont indiquées. **B** Les trois isoformes principales de Whirline exprimées dans les cellules ciliées et les photorécepteurs. En rouge sont indiquées les régions additionnelles aux isoformes provenant d'exons alternatifs et donnant lieu à des séquences différentes de la forme longue. **C** Positions des mutations identifiées de la Whirline répertoriées sur OMIM (omim.org) ainsi que leurs maladies humaines associées.

5.1 Topologie de la Whirline

La Whirline est codée par le gène humain DFNB31 sur le 9ème chromosome humain au locus 9q32-34 et le gène murin *whrn* sur le 4ème chromosome [215,216]. Le gène codant pour la protéine contient au total 12 exons et est fortement épissé en transcrits multiples résultant en

plusieurs isoformes de la protéine. La forme complète de la protéine contient 5 domaines et une région riche en proline. (Figure 32)

Les 30 premiers acides aminés en N-terminal de la protéine constituent une région de faible complexité enrichie en Serine, Alanine et Glycine. Cette séquence est suivie de quatre domaines consécutifs, séparés les uns des autres par des liens flexibles de 30 à 40 résidus environ. Deux domaines HHD (Harmonin Homology domain) sont aux deux extrémités de cette région [217]. Ces HHD entourent un tandem de domaines PDZ. Le reste de la protéine est composé d'une grande région riche en prolines et d'un troisième domaine PDZ proche de l'extrémité C-terminale. Un PBM est présent à son extrémité C-terminale, quatre résidus après le dernier brin du PDZ3 (Figure 32).

Au commencement de cette thèse, peu d'information structurale était disponible sur la Whirline. Le domaine HHD1, les trois domaines PDZ et la région riche en Proline étaient identifiés, mais le domaine HHD2 n'avait pas encore été décrit.

La grande majorité des publications sur la Whirline ne considère toujours que la possibilité d'interaction par les domaines PDZ, en occultant complétement le rôle des domaines HHD. Ce manque d'intérêt reflète très probablement le manque de données relatives aux domaines HHD. En se basant sur le supramodule N-terminal de l'Harmonine, de nombreuses publications utilisent une construction contenant à la fois le HHD1 et le PDZ1, pour n'observer que l'interaction du domaine PDZ avec des ligands. Dans le cas du PDZ2, les constructions couramment utilisées pour les expériences pull-down contiennent un domaine HHD2 tronqué au milieu de sa séquence [214]. Il n'existe actuellement aucune donnée directe sur l'interaction des deux domaines HHD de la Whirline avec des partenaires potentiels. La très faible conservation des résidus de surface des HHD rend difficile la prédiction de motif d'interaction spécifique aux HHD de la Whirline en se basant simplement sur leurs séquences. Les structures des domaines HHD sont inconnues, seules celles de l'Harmonine et de CCM2 peuvent être utilisées pour créer des modèles d'homologie, en raison de la faible similarité de séquence entre HHD.

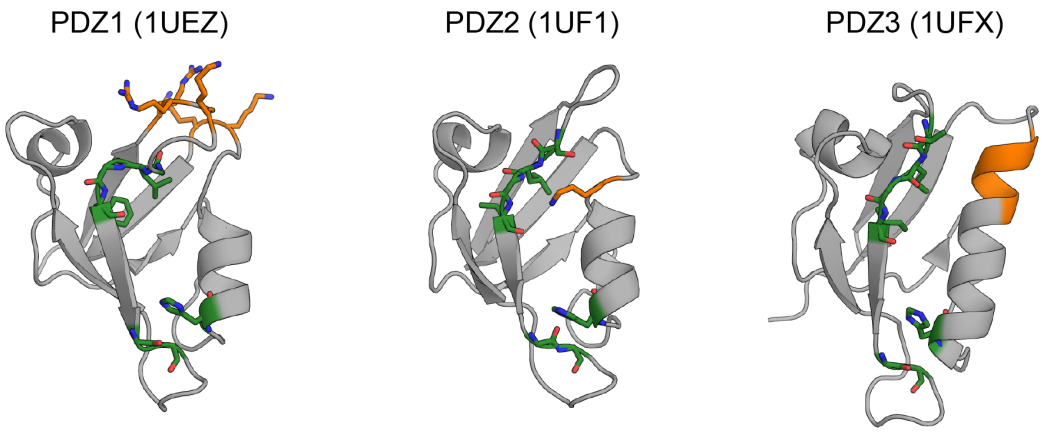


Figure 33 : Structures des domaines PDZ de la Whirline humaine. Structures RMN déterminées par le projet de génomique structurale RIKEN, sous le nom de protéine KIAA1526. En vert sont indiqués les résidus importants pour l'interaction avec les partenaires. En orange sont indiquées les particularités de chaque domaine. PDZ1 : patch cationique ; PDZ2 : Lysine dans la poche d'interaction ; PDZ3 : hélice α2 allongée.

Pour les domaines PDZ, plus d'informations sont disponibles (Figure 33). L'étude de leurs séquences les prédit tous trois comme type 1, avec une Histidine caractéristique au début de l'hélice α2 et une Glycine en sortie du brin ß2. Les structures des trois domaines PDZ de la Whirline humaine ont été résolues par RMN dans le cadre de l'initiative de génomique structurale de l'Institut RIKEN. Ces structures ont été publiées avant l'identification du gène et de la protéine, en conséquence, la protéine est enregistrée sous le nom de code KIAA1526. Les trois ensembles de structure adoptent le repliement classique des domaines PDZ avec cependant quelques particularités détaillées dans la Figure 33.

Le PDZ1 a un repliement proche du repliement PDZ. Sa plus forte particularité est son pI élevé de 9.9, conséquence d'un patch cationique formé de deux Arginines et trois Lysines au dessus de sa poche d'interaction, dans la boucle "GLGF" et dans la boucle α2/ß5[113]. Le pI élevé de ce domaine a attiré l'attention d'une équipe de chercheurs s'interessant aux interactions entre PDZ et lipides [113]. Les auteurs montrent que le PDZ1 de la Whirline est capable d'interagir spécifiquement avec certains lipides, notamment les phosphatidylinositole-phosphates. La mutation ponctuelle des résidus du patch cationique permet d'abroger l'interaction, confirmant l'implication de ce patch dans la liaison aux lipides. Cette étude a été réalisée avec une construction de PDZ1 tronquée des premiers résidus du brin ß1, ayant une tendance à oligomériser, selon les auteurs, compliquant de fait l'interprétation de ces résultats.

Le PDZ2 de la Whirline partage une forte identité de séquence avec le PDZ1 (57%) et un repliement similaire. De façon originale, la chaine latérale de la Lysine 347 est orientée vers l'intérieur du sillon d'interaction du domaine PDZ. Dans cette conformation, la chaine latérale recouvre la région hydrophobe du haut du sillon et ne permet pas l'accès à la "GLGF" pour un ligand. Malgré cette particularité structurale, plusieurs séquences C-terminales de protéines partenaires ont été identifiées par pull-down comme interagissant avec ce domaine PDZ2 [218–220].

Le PDZ3 a un repliement plus original. Son hélice α2 est formée de 15 résidus, allongée de plus d'un tour supplémentaire par rapport à celle d'un PDZ canonique. La longueur inhabituelle de cette hélice force la boucle "GLGF" dans une position plus éloignée du site de liaison et diminue sa flexibilité. Aucune information structurale n'existe sur une forme en complexe de ce domaine, il est donc difficile de spéculer sur l'impact de cette structuration sur la fonction du domaine. Ce domaine possède lui aussi un pI théorique élevé de 9.7, mais seule une interaction faible avec des phosphatidylinositol-phosphates a été décrite à ce jour [113].

5.2 Les isoformes de la Whirline et leurs expressions différentielles

Au cours des premières études de la Whirline, douze isoformes ont été prédites et déposées [216]. Récemment, une étude séminale a permis d'identifier les isoformes réellement exprimées parmi celles prédites. L'équipe de P. Mathur *et al.*[221] a identifié les trois formes principales de la Whirline effectivement exprimées dans les cellules ciliées et les photorécepteurs (**Figure 34**) et de les localiser dans les cellules ciliées.

5.2.1 L'isoforme N-terminale courte

La forme N-terminale courte inclut les 543 premiers résidus de la protéine. Elle contient les quatre premiers domaines de la Whirline, deux domaines HHD entourant un tandem de domaine PDZ. L'ARNm associé à cette isoforme courte N-terminale n'est pas détecté dans les cellules ciliées. Une étude précédente a montré l'expression de ce transcrit ainsi que l'expression de la protéine associée dans les cellules photoréceptrices de la rétine (Figure 34) [222,223]. Depuis, peu d'études se sont concentrées sur cette forme de la protéine qui semble être faiblement exprimée, comparativement à l'isoforme longue (protéine entière) avec qui elle colocalise systématiquement. Il est donc difficile de lui attribuer une fonction particulière et différente de l'isoforme longue.

5.2.2 L'isoforme C-terminale courte

La forme courte C-terminale de la protéine contient le HHD2, la région riche en proline, le PDZ3 et le PBM de la Whirline. L'expression de cette isoforme est contrôlée par un promoteur différent des deux autres, elle peut donc être exprimée de façon indépendante [221]. Cette isoforme C-terminale courte est uniquement exprimée dans les cellules ciliées. Lors du développement, elle se localise à l'extrémité des stéréocils de la plus grande rangée dans les CCI et les CCE [224,225]. A maturité, il semble que cette isoforme ne maintient sa localisation que dans les CCI [221]. Dans la littérature, une incohérence existe actuellement sur la présence, ou non, de la forme longue à l'extrémité des grandes rangées, aux côtés de la forme courte C-terminale [226]. L'étude de P. Mathur *et al.* ainsi que les travaux de nos collaborateurs, de l'unité de physiologie de génétique de l'audition de l'Institut Pasteur, corroborent la colocalisation de ces deux isoformes au sommet des stéréocils (Figure 34).

Indépendamment de la présence de la forme longue, l'isoforme C-terminale permet l'élongation des touffes ciliaires et leur maintien en longueur au stade mature [178,226,227]. Le mécanisme de cette régulation est inconnu. Il est cependant clair que l'interaction tripartite de la Whirline, de la Myosine15a et de Eps8 est indispensable. Les souris mutantes pour la Myosine-15a (*shaker*) et la Whirline (*whirler*) ont un phénotype identique, avec des touffes ciliaires extrêmement raccourcies et élargies [227]. D'autres protéines interagissent potentiellement avec la Whirline, à la localisation de cette isoforme courte. On peut notamment citer les travaux sur la Gelsoline et la protéine p55, toutes deux reliées à la condensation des fibres d'actine [225,228].

5.2.3 L'isoforme longue

L'isoforme longue de la Whirline contient l'ensemble de la séquence de la protéine. Morphologiquement, la protéine peut être décrite en deux parties, celle N-terminale contenant les deux HHD et le tandem de domaines PDZ, et la partie C-terminale contenant le PDZ3 et séparée par la région Proline-riche. L'isoforme longue est exprimée à la fois dans les photorécepteurs et en différentes localisations dans les cellules ciliées.

Dans les cellules photoréceptrices, l'isoforme longue est indispensable à la morphologie du cil connecteur entre le segment interne et externe. La partie N-terminale de l'isoforme longue régule le complexe formé par les protéines Usher2 (Usherine, Vlgr1, PDZD7) et SANS. Ce complexe crée des liens extracellulaires reliant le cytosquelette de microtubule du cil connecteur au cytosquelette d'actine du segment interne des

photorécepteurs[229]. Ainsi, la Whirline est un adaptateur clé dans la liaison des liens extracellulaires aux cytosquelettes des deux compartiments[183]. La perturbation de ce complexe entraine des variations d'écartement entre le cil connecteur et le segment interne [223], causant à terme des problèmes fonctionnels.

Dans les cellules ciliées, la forme longue de la Whirline est premièrement impliquée durant le développement. Cette isoforme longue est exprimée de façon transitoire à la base des stéréocils des CCI et des CCE, dans les régions des liens basaux formés par les protéines membranaires du complexe Usher2 [214,223]. Ce complexe temporaire est indispensable à la bonne morphologie des touffes ciliaires des CCE dans leur forme caractéristique en V [221,226]. La composition moléculaire de ce complexe semble très similaire à celle observée autour du cil connecteur des photorécepteurs, expliquant les syndromes Usher de type 2 touchant à la fois cellules photoréceptrices et cellules ciliées.

Il est maintenant établi que la forme longue de la Whirline est absente des CCE matures mais reste présente dans les CCI [221,226]. Cependant, sa localisation et sa fonction restent à définir. Comme précisé précédemment, les travaux de P.Mathur *et al.*[221] et de nos collaborateurs suggèrent que la forme longue localise à l'extrémité des stéréocils de la grande rangée. Cependant, une étude récente dans les CCI matures par microscopie super-résolue [226] montre une nouvelle localisation spécifique de l'isoforme longue au milieu de la moyenne rangée de stéréocils. Les auteurs associent à cette localisation un rôle de la forme longue dans la formation de liens inter-stéréocils importants pour la structure de la touffe ciliaire. Cette proposition de fonction est uniquement appuyée par l'étude de localisation et parait contradictoire avec la morphologie normale observée pour des touffes ciliaires de CCI de souris K.O spécifique de la forme longue. Bien que les auteurs excluent la localisation de la forme longue à l'extrémité des grands stéréocils, la qualité de leur marquage anticorps N-terminal et de leurs images ne semblent pas permettre de conclure définitivement. (Figure 34)

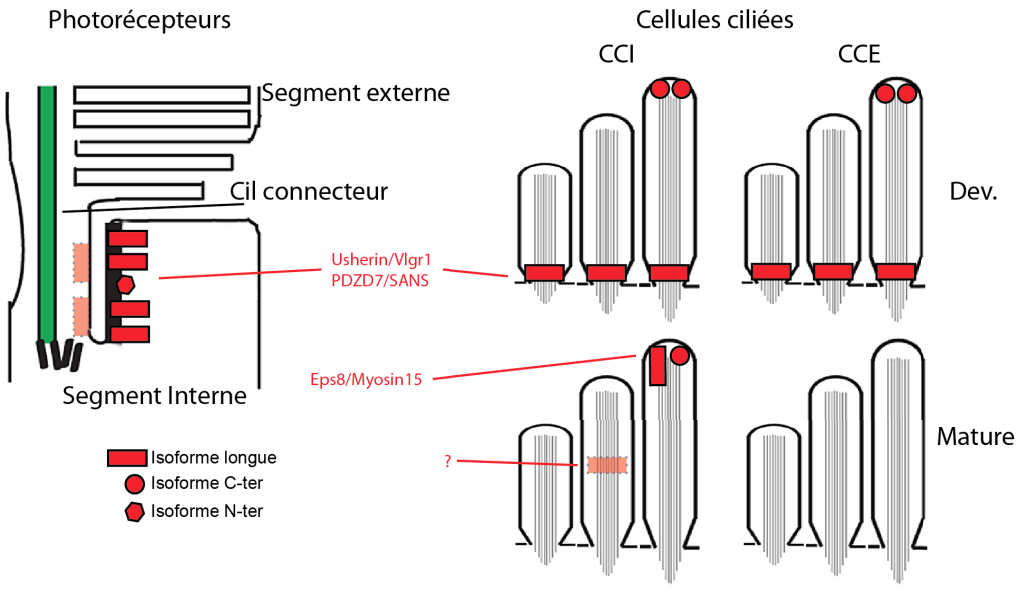


Figure 34: Localisation des isoformes de la Whirline dans les photorécepteurs et les cellules ciliées. Représentation schématique des localisations décrites des isoformes de la Whirline. En plus clair sont montrées les localisations reportées mais restant à confirmer. Figure adaptée de Mathur *et al.*[221]

5.3 Les phénotypes résultant des mutations de la Whirline

Les mutations de la Whirline peuvent induire des déficiences graves de la vision et de l'audition. Deux maladies majeures sont liées à la Whirline, la première affectant uniquement les fonctions de la cochlée et la seconde touchant à la fois les fonctions cochléaire et rétinienne. Les mutations du gène associé à la Whirline induisant l'une ou l'autre maladie sont ségrégées le long de la séquence. Les mutations causant uniquement une surdité congénitale sont localisées à l'extrémité C-terminale de la séquence alors que les mutations responsables de syndromes de surdité-cécité se situent dans la partie N-terminale de la protéine (Figure 32) [226].

5.3.1 DFNB31

La surdité DFNB31, causée par mutation de la Whirline est autosomale, récessive et non-syndromique. Elle est caractérisée par une perte profonde de l'audition sur l'ensemble des fréquences, dès la naissance[215]. Jusqu'à présent, toutes les mutations du gène associées à la Whirline liées à cette maladie induisent des formes tronquées dans la région C-terminale de la protéine (Figure 32).

Le modèle de souris *whirler* possède une délétion naturelle des exons 7-9 du gène associé à la Whirline et présente des symptômes similaires à la maladie humaine[230]. Cette mutation cause la perte de l'isoforme courte C-terminale et produit, à la place de la forme longue, une forme N-terminale de la Whirline tronquée au milieu du domaine HHD2 [216]. Cette lignée de souris a été abondamment étudiée pour comprendre les mécanismes à l'origine de la maladie humaine. Chez ce mutant naturel, les touffes ciliaires des CCI et des CCE de l'ensemble de la cochlée ne s'allongent pas, et ont même tendance à raccourcir après la naissance. Ce phénotype est maintenant associé à la perte de fonction de l'isoforme C-terminale de la protéine. Le second phénotype observé concerne les touffes ciliaires des CCE, qui adoptent une organisation en U plutôt que la forme traditionnelle en "V ou W". Ce phénotype est associé spécifiquement à la fonction de l'isoforme longue au niveau des liens basaux[221] (Figure 35).

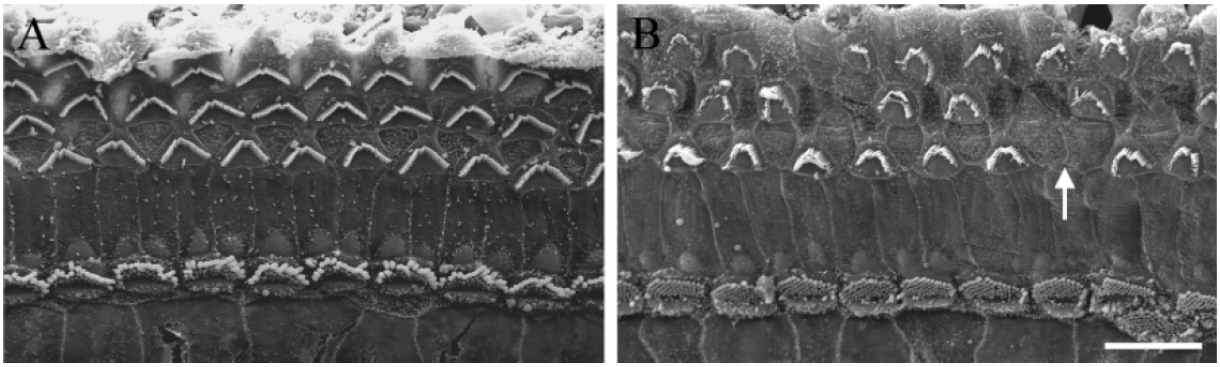


Figure 35 : Changement de morphologie des touffes ciliaires chez les souris *whirler* : A image de cochlée chez une souris adulte (P60) hétérozygote *+/wi.* **B** image de cochlée chez une souris adulte (P60) *wi/wi.*Barre d'échelle : 10µm ; figure 7 de Holme *et al.* 2002[230].

5.3.2 USH2D

Le syndrome Usher2 sous-type D est induit par des mutations de la partie C-terminale de la Whirline. Il est caractérisé par une perte légère à moyenne de l'audition et une dégénération progressive de la rétine, perceptible par le patient entre la vingtième et la trentième année [231,232]. La plupart des mutations Usher de la Whirline engendre une terminaison prématurée de la protéine, produisant des isoformes largement tronquées, peu informatives sur la fonction individuelle des domaines de la protéine. Une mutation est cependant plus informative : elle induit une délétion locale commençant dans le dernier brin du PDZ1 et finissant au début du PDZ2. Cette isoforme longue de la Whirline est complète, mis à part ses PDZ1 et PDZ2 fusionnés et probablement non fonctionnels [232]. Cette mutation est suffisante pour induire les symptômes d'audition réduite et de perte progressive de vision

du syndrome Usher, au moins l'un de ces deux PDZ est donc crucial pour la fonction de la Whirline dans ces deux systèmes.

Il existe deux lignées murines génétiquement modifiées sur le N-terminal (K.O de l'exon 1 ou de l'exon 4) de la Whirline reproduisant les symptômes Usher [223,226]. Ces deux modèles de souris mutées sur la partie N-terminale de la protéine montrent des déficits de vision et d'audition. Il est important de traiter avec précaution les données obtenues sur les cellules photoréceptrices de souris, leur morphologie étant différente des cellules photoréceptrices de primates. Au niveau des photorécepteurs, ces souris modèles ont une distance extrêmement variable entre le cil connecteur et la membrane du segment interne (Figure 36)[223]. La perte du complexe Usher2 induit par la mutation perturbe ainsi la régulation de cette distance et induit une fusion partielle du cil avec le segment interne. Au niveau des touffes ciliaires, des phénotypes variés ont été observés. Dans tous les exemples publiés, les CCE adoptent une forme en U, comme pour les souris *whirler*. Dans l'une des publications, une perte d'audition forte est présentée, avec comme symptôme supplémentaire un raccourcissement partiel des stéréocils des CCI [221]. Ces résultats n'ont pas été confirmés par une étude suivante, montrant une morphologie de stéréocils normale et une perte d'audition partielle (Figure 36) [226]. Il est possible que la perte de la forme des touffes ciliaires des CCE soit responsable de la perte d'audition observée, cette morphologie étant indispensable pour la stimulation correcte de ces cellules amplificatrices.

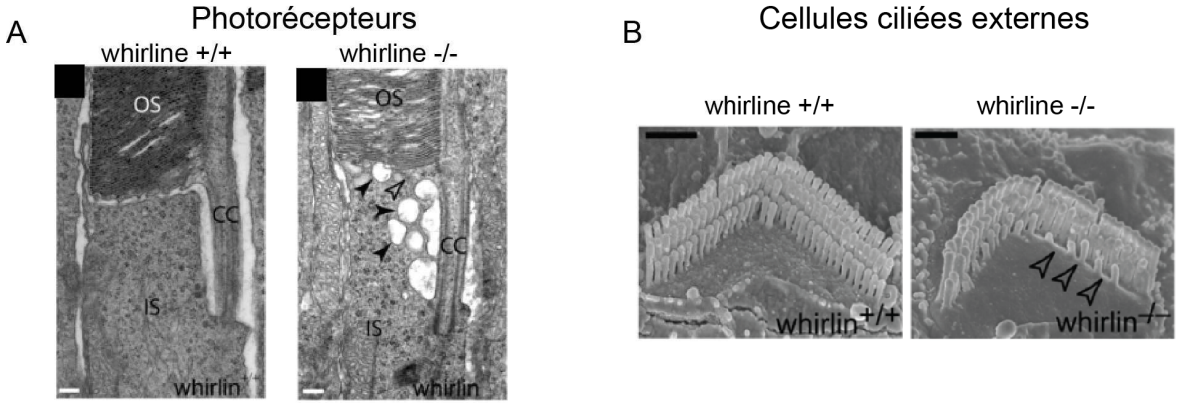


Figure 36 : Effet de mutations de l'isoforme longue de Whirline sur les photorécepteurs et les cellules ciliées externes. A Région du cil connecteur de souris sauvage (+/+) ou mutée (-/-). **B** Touffes ciliaires de cellules ciliées externes de souris sauvage (+/+) ou mutée (-/-). Adaptée de Yang *et al.* 2010.

5.4 Implications connues de la Whirline dans des complexes multi-protéiques

Depuis l'identification de la Whirline, de nombreux partenaires lui ont été attribués. Toutes ces interactions ont été montrées par des techniques indirectes, notamment la co-localisation par microscopie confocale ou par des techniques prompts aux faux-positifs comme le pull-down ou la co-immunoprécipitation. L'identification tardive des domaines HHD et la mauvaise définition des bornes des domaines PDZ dans les banques de données traditionnelles ont conduit à l'utilisation récurrente de constructions comportant des domaines tronqués et probablement sensibles à l'agrégation. L'ensemble de ces facteurs rend difficile l'identification fiable du réseau d'interaction de la Whirline (Figure 37). La multiplication des publications permet cependant de faire ressortir certains partenaires récurrents que nous présenterons dans cette partie.

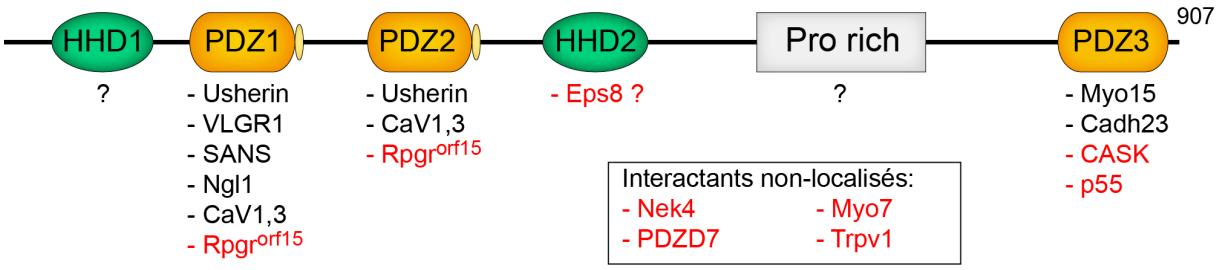


Figure 37 : Interactants de la Whirline, par domaine. Liste des interactants identifiés de la Whirline pour chaque domaine. Les partenaires listés en rouge ne possèdent pas de PBM à leur C-terminal.

5.4.1 Complexe Usher2 : Usherine / Vlgr1 / PDZD7

L'isoforme longue de la Whirline est impliquée dans la formation du complexe des protéines Usher2, au niveau des liens basaux des cellules ciliées et du cil connecteur des photorécepteurs. Le complexe est composé de deux protéines membranaires responsables des liens basaux, l'Usherine et Vlgr1, et de deux protéines cytoplasmiques paralogues, Whirline et PDZD7 [214,233,234]. Ces protéines membranaires possèdent toutes deux une région cytoplasmique courte avec un PBM identique (–DTHL). Ces PBM interagissent avec les deux PDZ N-terminaux de la Whirline et de PDZD7 [33,214], qui elles-mêmes interagissent entre elles pour ponter les protéines membranaires. Dans des études de compétition par pull-down, les

auteurs proposent que la Whirline ait une interaction préférentielle avec l'Usherine alors que PDZD7 ait une interaction favorisée avec Vlgr1, malgré les séquences identiques de leurs PBM [214]. Les auteurs montrent aussi une hétéro-dimérisation possible entre PDZD7 et la Whirline. Dans leurs expériences, tous les domaines PDZ d'une protéine sont capables d'interagir avec tous les domaines PDZ de la seconde. Les auteurs proposent aussi une homodimérisation de la Whirline par auto-association de son PDZ3 et de son tandem de PDZ N-terminal (Figure 38 A) [214]. Une telle interconnexion entre PDZD7 et la Whirline semble peu probable. Cependant, plusieurs études démontrent la nécessité d'un complexe entre ces quatre protéines au niveau des liens basaux [33,203,234,235]. Aucune autre interaction du PDZ3, de la région riche en proline ou du HHD2 n'a été décrite pour des protéines présentes au niveau des liens basaux. Cependant, les souris *whirler* présentent des touffes ciliaires de CCE en forme de U, caractéristique d'un défaut du complexe Usher2 [221]. Cette région C-terminale de la protéine est donc également impliquée dans la fonction de l'isoforme longue au niveau du complexe Usher2.

Dans les cellules photoréceptrices, PDZD7 n'est pas exprimée, mais la Whirline colocalise avec l'Usherine et Vlgr1 [236] au niveau du cil connecteur. La protéine Usher1 SANS colocalise aussi avec ce complexe Usher2 à la membrane du cil connecteur ainsi qu'à la membrane du segment interne. Montrée par pull-down, une interaction directe a lieu entre la partie centrale de SANS et l'Usherine ainsi qu'entre le PBM de SANS et les PDZ N-terminaux de la Whirline (Figure 38 B) [183,229]. Ces différentes interactions permettraient un ancrage multiple entre les protéines membranaires et le cytosquelette, favorable au maintien des liens extracellulaires.

La forme longue de la Whirline interagit également avec la protéine $Rgpr^{orf15}$. Cette protéine associée aux microtubules est nécessaire au bon fonctionnement du cil connecteur et ses mutations sont responsables de plus 70% des cas de *retinitis pigmentosa* [222]. Les auteurs ont montré, par des expériences de double hybride et de pull-down, l'interaction de la région C-terminale de $Rgpr^{orf15}$ avec le PDZ1 de la Whirline, et ce malgré l'absence de PBM canonique sur $Rgpr^{orf15}$ [222]. Cette interaction n'a cependant été reportée que dans ce seul article et mérite d'être confirmée avant d'intégrer $Rgpr^{orf15}$ dans le complexe Usher2 des photorécepteurs.

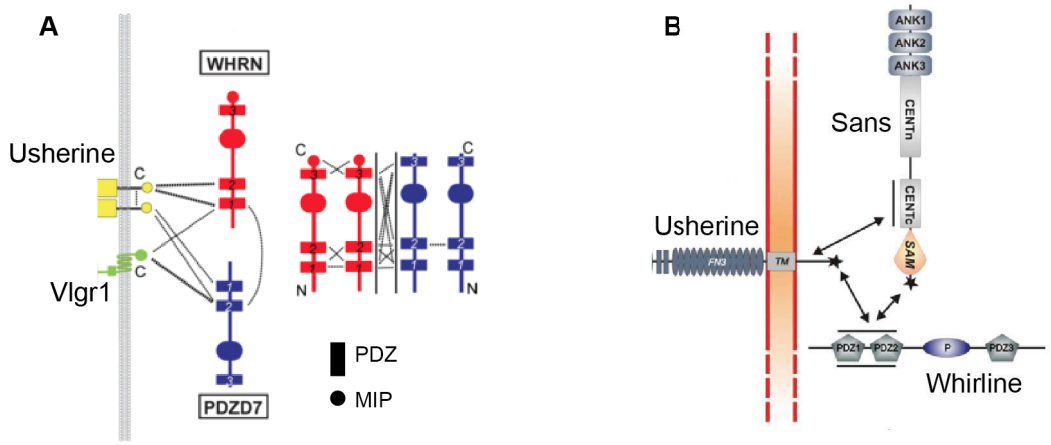


Figure 38 : Organisation moléculaire du complexe Usher2. A Interactions décrites dans le complexe Usher2 formé dans les liens basaux des cellules ciliées. Figure adaptée de Chen *et al.* 2014 [214] **B** Interactions décrites dans le complexe Usher2 formé dans le cils connecteur des photorécepteurs. Figure adaptée de Sorusch *et al.* 2017 [183].

5.4.2 Complexe Eps8/Myosine15

Dans les touffes ciliées matures et lors de leur développement, la forme courte de la Whirline se localise à l'apex des stéréocils de la grande rangée [178,225,226]. A cette localisation, la partie C-terminale de la Whirline fait partie d'un complexe impliquant, au minimum, la Myosine15a et la protéine Eps8. Ce complexe est indispensable à l'élongation et au maintien en longueur des stéréocils des touffes ciliaires [219,227,237].

Les premières indications de la formation de ce complexe viennent de la similarité phénotypique des souris modèles *shaker2* et *whirler,* respectivement mutées sur la Myosine15a et sur la Whirline [178] ainsi que de souris K.O de Eps8 [237], Elles possèdent toutes des touffes ciliaires formées de stéréocils courts et larges. La Whirline et Eps8 ne se localisent plus à l'apex des stéréocils dans des souris Myosine-15 K.O, suggérant une interaction des deux protéines avec la Myosine15 et proposant son rôle de cargo pour la bonne localisation du complexe. Le K.O de la Whirline ou de Eps8 n'affecte pas la localisation de la seconde. Les deux protéines interagissent chacune directement avec la Myosine15 [237]. La Myosine15 possède un PBM en son extrémité C-terminale et interagit avec l'isoforme C-terminale courte de la Whirline [226], la mutation *whirler* abolit cette interaction [178]. En cellule COS7 transfectées, la Whirline et la Myosine-15 colocalisent dans les filopodes. La suppression du PDZ3 de la forme longue de la Whirline ainsi que celle du PBM de la myosine15a sont toutes deux suffisantes pour empêcher la colocalisation des deux protéines [227]. Ce résultat souligne donc l'importance de l'interaction entre le PBM de la Myosine15a et le PDZ3 de la Whirline dans leur interaction. Par ailleurs, une autre interaction, entre la région riche en proline de la

Whirline et le second domaine SH3 de la Myosine-15a a aussi été reportée par pull-down [178]. L'interaction entre Eps8 et la Whirline implique quant à elle une grande région non-structurée au centre d'Eps8 et une région non-identifiée de la Whirline [237]. Eps8 ne possède pas de PBM et est capable d'interagir avec l'isoforme C-terminale courte de la Whirline de façon simultanée avec la Myosine-15a [226]. La région riche en proline, et le domaine HHD2, de partenaire inconnu, sont donc deux bons candidats pour moduler cette interaction. L'interaction entre Eps8 et la Myosin-15 est contrôlée par la région C-terminale d'Eps8 et par la région proche du domaine moteur de la Myosine-15a, sans recouvrement avec les régions d'interaction à la Whirline des deux protéines (Figure 39) [237].

La Gelsoline [228] et l'Espine [238] sont décrites comme interagissantes avec la Whirline et pourraient être associées à ce complexe, régulant la longueur des stéréocils. Ces deux protéines on clairement un rôle régulateur de la condensation de l'actine dans les stéréocils, cependant leur lien avec le complexe Myosine-15a/Eps8/Whirline reste encore à établir.

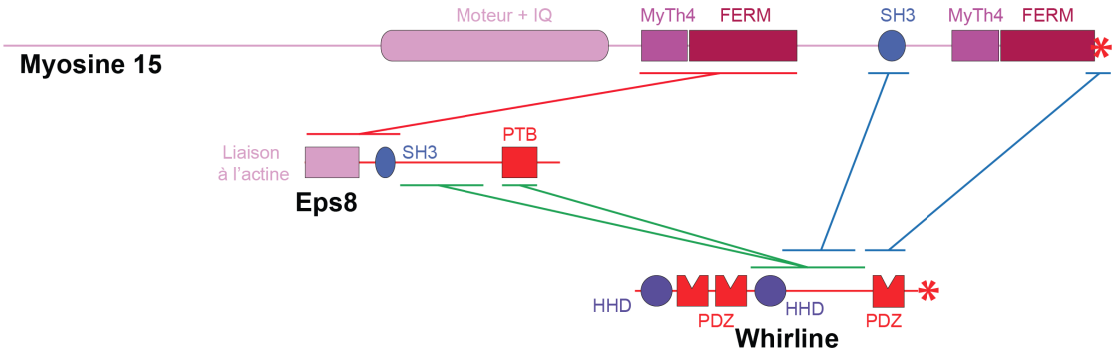


Figure 39 : Organisation moléculaire du complexe de l'apex des stéréocils. Interactions décrites entre les trois protéines principales du complexe de l'apex de la grande rangée de stéréocils. Les interactions Myosine15/Eps8 sont montrées en rouge ; Eps8/Whirline en vert ; Myosine15/Whirline en bleu.

5.4.3 Autres interactions connues

En plus des interactions attribuables à un complexe, une localisation et une fonction précise, plusieurs autres interactants ont été décrits pour la Whirline. C'est le cas de plusieurs protéines de la famille MAGUK, notamment CASK et MPP1 [239,240]. Dans ces deux cas, l'interaction est montrée par pull down entre leur séquence C-terminal et le PDZ3 de la Whirline, bien que les deux protéines ne possèdent pas de PBM. CASK et MPP1 possèdent elles-mêmes des domaines PDZ qui pourraient lier en retour le PBM C-terminal de la Whirline. Les mutations de ces protéines ne sont pas impliquées dans des déficiences de vision ou d'audition et n'ont pas directement été connectées aux réseaux Usher. CASK et

MPP1 sont cependant largement exprimées dans l'ensemble du cerveau. Il est possible que les interactions décrites soient impliquées dans les fonctions de la Whirline en dehors de la vision et de l'audition.

Des expériences de pull-down et de colocalisation dans les cellules photoréceptrices montrent que le tandem N-terminal de PDZ de la Whirline se lie au PBM C-terminal du canal calcique voltage dépendant CaV1,3 [241,242]. Ce dernier est exprimé tout le long des photorécepteurs ainsi que dans la synapse des cellules ciliées. Son interaction avec la Whirline a été remise en cause et est toujours débattue [242]. Cependant, la co-expression des deux protéines dans différents types cellulaires et l'interaction de CaV1,3 avec l'Harmonine [243] en font un bon candidat comme partenaire de la Whirline. Leur interaction pourrait notamment avoir un rôle au niveau de la synapse en ruban des cellules ciliées, où l'ensemble des protéines Usher est également localisé.

Un autre canal Trpv1, pourrait aussi être un partenaire de la Whirline. Leur interaction est modulée par la région N-terminale de domaines Ankyrines répétés de Trpv1, capable d'interagir avec l'ensemble des domaines PDZ de la Whirline[244]. Trpv1 est une protéine membranaire impliquée dans la sensation de douleur associée aux hautes températures. En cellule, la Whirline augmente le ciblage de cette protéine à la membrane, améliore son regroupement (clustering) au sein de la membrane et sa stabilité [244].

Enfin, une dernière interaction de la Whirline a été décrite avec deux isoformes de la protéine Nek4 [245]. Cette Serine/Thréonine kinase n'a pas de fonction connue actuellement, mais semble impliquée dans la stabilisation des microtubules et dans la régulation des cils primaires. L'interaction avec la Whirline a été montrée lors d'un criblage protéomique par spectrométrie de masse et par analyse informatique, sans validation ultérieure.

5.5 Implication de la Whirline dans d'autres types cellulaires

La présence et la fonction de la Whirline en dehors des cellules ciliées et cellules photoréceptrices ont été peu étudiées. La Whirline est cependant décrite dans d'autres systèmes sensoriels, ainsi que dans le maintien structurel des neurones.

Comme décrit dans le chapitre précédent, la Whirline est reliée à un système protéique impliqué dans la fonction des cellules nociréceptrices [244], et plus particulièrement dans la sensation de douleur liée aux hautes températures. Dans ces cellules, la Whirline n'est pas directement impliquée dans la transmission de la cascade de signalisation mais intervient dans l'organisation du complexe sous membranaire du canal thermosensible Trpv1. Les autres

composantes de ce complexe de signalisation restent inconnues, il est possible que les interactions de la Whirline avec les protéines MAGUK telles que CASK ou MPP1 entrent en jeu dans ce système.

L'autre système sensitif impliquant la Whirline est la proprioception, c'est à dire la perception de la position d'un membre ou d'une partie du corps, dans l'espace [246]. Cette sensation émane des muscles squelettiques et est modulée par leurs contractions. Les mécanismes moléculaires sont encore mal compris, ils impliquent cependant des protéines membranaires mécano-sensibles. Dans les neurones propriocepteurs, la Whirline est exprimée pendant le développement et durant le stade mature. Elle est nécessaire à la transduction du signal mécanique en un potentiel d'action neuronal [246]. Ses partenaires dans ce type cellulaire sont actuellement inconnus.

Enfin la Whirline a aussi été décrite dans l'organisation axonale de neurones non-sensoriels [247]. Les auteurs suspectent une fonction similaire à celle connue dans les neurones sensitifs, c'est à dire la connexion des protéines membranaires au cytosquelette, de microtubules dans ce cas.

Bien que la Whirline soit principalement étudiée dans les cellules régulant vision et audition, de plus en plus d'évidences montrent son implication dans d'autres types cellulaires, majoritairement reliés au système nerveux. Il semble notamment que des variations autour du complexe Usher2 soient utilisées dans les cellules du système nerveux pour contrôler le développement et la morphologie des neurones.

6 Problématique

L'objectif de cette thèse est d'élucider les mécanismes moléculaires associés aux fonctions de la Whirline dans la rétine et la cochlée. Pour se faire, nous nous sommes intéressés d'une part aux interactions de la Whirline avec les autres protéines Usher et les protéines de l'audition. D'autre part, nous avons caractérisé sa structure, en portant un intérêt particulier à son organisation modulaire.

Dans la rétine et la cochlée, les photorécepteurs et les cellules ciliées sont des neurones sensoriels dont la fonction unique est de traduire un stimulus physique, l'arrivée d'une onde électromagnétique ou sonore, en un message électrochimique transmissible au cerveau. Cette fonction unique nécessite de pouvoir être stimulée de façon constante, de capter l'information dans une large gamme d'amplitude et de fréquence et de la transmettre en une fraction de seconde. Ces contraintes nécessitent une morphologie cellulaire et des voies de signalisation exceptionnelles.

Les protéines Usher, dont la mutation des gènes associés provoque un syndrome de surdité/cécité, forment un ensemble constitué de protéines membranaires et de protéines d'échafaudage ne possédant aucun domaine catalytique. Associées à des protéines spécifiques du type cellulaire, elles assurent la mise en place et le maintien des morphologies et de la fonction des neurones sensitifs. Ces complexes se forment *via* les nombreux modules d'interaction et motifs contenus dans les protéines Usher. La grande majorité des protéines Usher et protéines associées possède des motifs d'interaction aux domaines PDZ, alors que seulement trois protéines du système possèdent de tels domaines, les orthologues Harmonine, PDZD7 et Whirline. La fonction de l'Harmonine et de ses partenaires moléculaires a été beaucoup étudiée au début des années 2010 [165]. PDZD7 n'a été caractérisée que récemment et son rôle semble limité au complexe des liens basaux. La Whirline est exprimée en plusieurs positions des stéréocils, pendant le développement et au stade mature, mais ses fonctions sont restées longtemps ignorées.

Au début de ce projet en 2013, la caractérisation de l'organisation modulaire de la Whirline était très parcellaire (Figure 40). Tout d'abord, les domaines HHD n'étaient pas identifiés, seul un article de bioinformatique de décembre 2014 repère les domaines HHD dans la séquence de la Whirline [214]. Dans les banques de données des protéines telles

qu'Uniprot et Interpro, la présence des domaines HHD dans la Whirline n'est toujours pas reconnue. Ce manque d'identification a eu pour conséquence l'utilisation de constructions supposées ne contenir qu'un domaine PDZ mais incluant partiellement ou entièrement les HHD, dans de nombreuses publications [178,214,238]. Ces constructions ont été utilisées en pull-down en systèmes hétérologues, où la construction est surexprimée sans contrôle de son agrégation. Ainsi dans le grand nombre d'interactants décrits au début du projet, il était particulièrement délicat de séparer partenaires et artefacts.

Enfin, au lancement du projet, la localisation proposée pour la Whirline le long des stéréocils était différente. Il était considéré que la Whirline localisait au complexe des liens basaux pendant le développement, au sommet des stéréocils de la grande rangée, mais aussi au sommet des plus petites rangées, au niveau du complexe de la mécano-transduction. Cette localisation au niveau du point inférieur d'ancrage du lien apical a été montrée dans un seul article [234] mais n'a jamais été reproduite ultérieurement. Elle a cependant était largement adoptée par la communauté et perdure encore aujourd'hui en dépit de l'absence de confirmation [200]. Cette localisation sous-entend un rôle direct de la Whirline dans la mécano-transduction, qui ne semble pas être avérée avec des résultats plus récents (Figure 40).

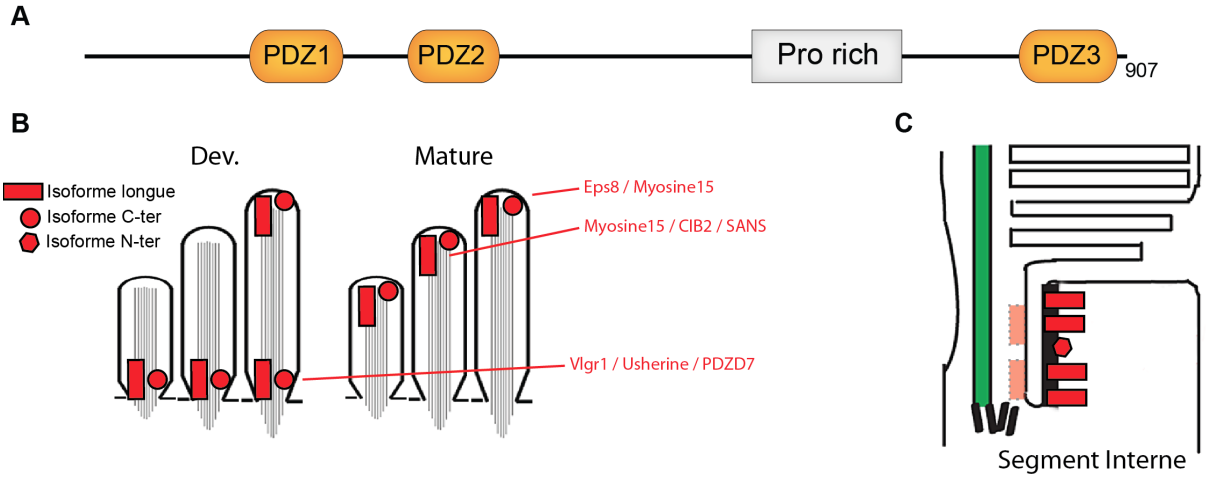


Figure 40 : Information sur la Whirline au début de la thèse. (A) Organisation en domaine de la Whirline connue en 2013. **(B et C)** Localisations des isoformes de la Whirline, identifiées en 2013, dans les stéréocils et les photorécepteurs.

L'objectif principal de cette thèse a été l'étude structurale des domaines de la Whirline, et plus particulièrement les structures supramodulaires que ceux-ci forment. Au début de cette thèse, aucune information sur des superstructures potentiellement formées par les domaines de la Whirline n'était disponible. En se basant sur les travaux effectués précédemment sur l'Harmonine, nous suspections cependant que celles-ci pouvaient exister et pouvaient moduler significativement la fonction des domaines isolés de la Whirline. J'ai donc

cloné, produit et purifié les constructions des domaines isolés, de toutes les combinaisons possibles de domaines deux à deux et plusieurs constructions contenant trois à quatre domaines. J'ai ensuite caractérisé ces constructions de la Whirline avec une approche systématique couplant cristallographie, SAXS et RMN ainsi que d'autres méthodes biophysiques, telle que l'émission de fluorescence. Nous avons ainsi pu identifier un supramodule transitoire formé entre les deux domaines PDZ N-terminaux de la Whirline pour lequel nous avons déterminé la structure, caractérisé son échange avec des formes non-compactes et déterminé l'impact de ces conformations sur l'affinité pour l'un des partenaires. Nous avons aussi résolu la structure atomique du second domaine HHD, nouvellement identifié et montré que celui-ci n'interagit pas avec le supramodule de domaine PDZ le précédant.

Le second objectif de cette thèse était de clarifier le réseau d'interaction de la Whirline au sein du système de l'audition en se concentrant particulièrement sur les interactions impliquant les domaines PDZ. Dans le système de l'audition, la Whirline, possédant à la fois PBM et domaines PDZ, est un candidat idéal pour être au cœur de complexes variés. Nous avons donc criblé par Résonance Plasmonique de Surface (SPR) les C-terminaux de toutes les protéines Usher et certaines autres protéines de l'audition avec les domaines PDZ de la Whirline, isolés ou avec leurs domaines adjacents. Nous avons ensuite confirmé les interactions nous paraissant les plus intéressantes par d'autres méthodes de titrage en solution, suivie par l'émission de fluorescence des tryptophanes ou par RMN. Nous avons ainsi obtenu une vue plus clair du réseau d'interaction formé par la Whirline.

Résultats

7 Résultat 1 : Délimitation et production des constructions de la Whirline

7.1 Identification de structures conservées

7.1.1 Domaines HHD

La première étape de ce projet a été l'analyse approfondie de la séquence de la Whirline. Cette étape initiale a été essentielle et déterminante pour l'ensemble des résultats qui seront présentés ultérieurement dans cette thèse, ainsi que pour les travaux futurs qui seront entrepris sur ce projet après mon départ.

Tous les travaux de cette thèse ont été effectués sur la séquence de l'isoforme-3 de la Whirline de souris (Uniprot : Q80VW5-3). Cette isoforme a été choisie car elle correspond à la forme la plus proche de l'isoforme-1 humaine, avec une très forte identité de séquence dans les domaines (supérieure à 90%). Le choix d'utiliser le gène de souris est lié au contexte scientifique, plus directement avec les expériences de biologie cellulaire de localisation et de K.O, effectuées quasi-exclusivement chez la souris.

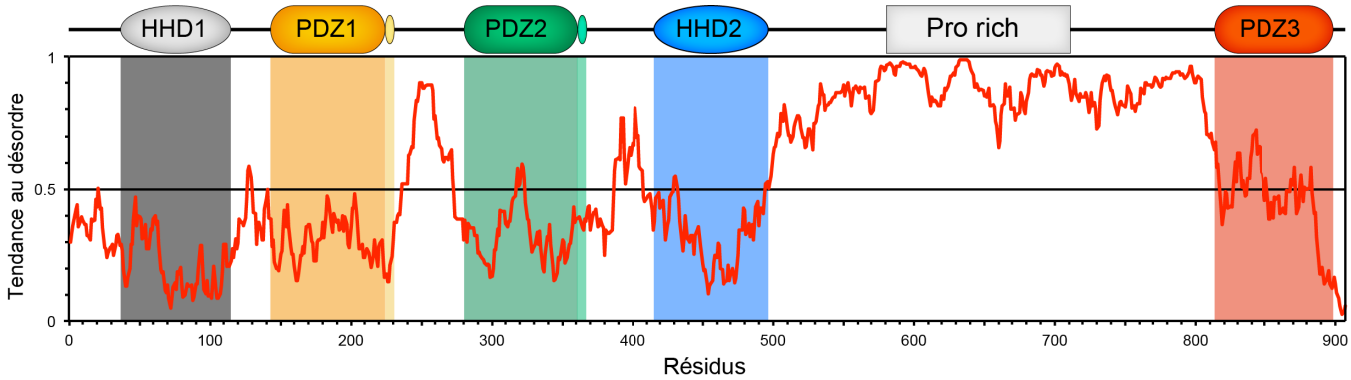


Figure 41 : Prédiction de tendance au désordre de la séquence de la Whirline. Comparaison de l'organisation en domaine de la Whirline que nous avons déterminée et la prédiction de tendance au désordre déterminée par IUpred (iupred.enzim.hu).

Comme nous l'avons précisé dans le chapitre précédent, seuls les domaines PDZ de la Whirline étaient identifiés au début du projet. Pourtant, la prédiction de désordre basée sur la séquence de la Whirline montre que d'autres régions tendent vers l'ordre et pourraient elles aussi être structurées (Figure 41). L'ensemble du N-terminal, des 2 à 125, est prédit comme ordonné (séquence en appendice, section 13.1). Les 32 premiers résidus de la séquence

forment une région de basse complexité, riche en S/A/G. Nous avons donc concentré notre travail sur la région du résidu 33 au résidu 125. L'Harmonine, paralogue à la Whirline, contient un domaine HHD en amont de son domaine PDZ1 ; nous avons étudié la présence d'un tel domaine dans la Whirline. L'alignement des deux régions montre une identité de 24%, concentrée exclusivement sur les résidus hydrophobes du cœur du domaine HHD de l'Harmonine. Cette identité, ainsi que les prédictions de structure secondaire, ont permis d'attribuer de façon non ambiguë la région 36-110 de la Whirline comme un domaine HHD.

La séquence du domaine HHD de l'Harmonine, à la base de la définition de la famille, est relativement atypique pour un domaine HHD (Table 3). L'identification d'autres domaines HHD, par alignement contre sa séquence est infructueuse. Au contraire, Le séquence du domaine HHD1 de la Whirline est, de loin, la plus canonique des domaines HHD, de son identification découle ensuite l'identification des autres domaines HHD impliqués dans l'audition (**Table 3**). La recherche de séquences similaires en utilisant un algorithme d'alignement itératif sur positions spécifiques (PSI-BLAST) a permis d'identifier, entre autres, le second HHD de la Whirline. Ce HHD2 (424-496) de la Whirline correspond à la région en C-terminal du PDZ2, prédite comme ordonnée mais qui n'était pas encore attribuée (**Figure 41**).

	Hrmn	Delph_1	Delph_2	Whrn_1	PDZD7	Whrn_2
Hrmn		28.9	18.7	22.7	18.7	17.3
Delph_1	28.9		22.7	26.7	21.3	24.0
Delph_2	18.7	22.7		26.7	18.7	17.3
Whrn_1	22.7	26.7	26.7		34.7	28.0
PDZD7	18.7	21.3	18.7	34.7		34.7
Whrn_2	16.0	24.0	17.3	28.0	34.7	

Table 3 : Taux d'identité entre HHD. Matrice montrant le pourcentage d'identité des domaines HHD, deux à deux. Les domaines sont ordonnés selon leurs identités respectives, Hrmn : Harmonine, Delph_n : Delphiline HHDn, Whrn_n : Whirline HHDn, PDZD7. Le code couleur est : <20% rouge, de 20 à 25% orange, >25% vert.

La matrice de score d'alignement par position (PSSM) des domaines HHD a ainsi été raffinée par l'ajout des nouvelles séquences de la Whirline. Elle a aussi permis de détecter l'unique HHD contenu dans PDZD7 et les deux HHD contenus dans la Delphiline. La recherche des domaines HHD montre que l'Harmonine, la Whirline, PDZD7 et la Delphiline partagent une organisation commune en domaine et sont, à priori, paralogues (**Figure 42**).

Bien que ces quatre protéines aient une région N-terminale commune comportant deux domaines PDZ et un ou deux domaines HHD, l'ordre et l'espacement entre domaines est cependant propre à chaque protéine. Les bases génétiques de cette plasticité de recombinaison seraient intéressantes à étudier plus en profondeur. L'Harmonine, la Whirline et PDZD7 sont toutes trois impliquées dans l'audition. La Delphinine, quant à elle, n'a pas été identifiée dans les cellules de l'audition, mais est impliquée dans la densité post synaptique (PSD), notamment associée aux récepteurs glutamate GluR1 et GluRδ2 [248–250]. En même temps que nous progressions sur cette étude, l'équipe du Dr. Callebaut a publié un article sur l'identification d'un nouvel domaine HHD dans RTEL1, en se basant sur un alignement de séquence contenant les HHD de la Whirline, de l'Harmonine et de l'un des HHD de la Delphiline, confirmant ainsi nos propres travaux. Bien que publiés, les domaines HHD de la Whirline ne sont que très rarement pris en considération dans les articles portant sur leur rôle dans l'audition.

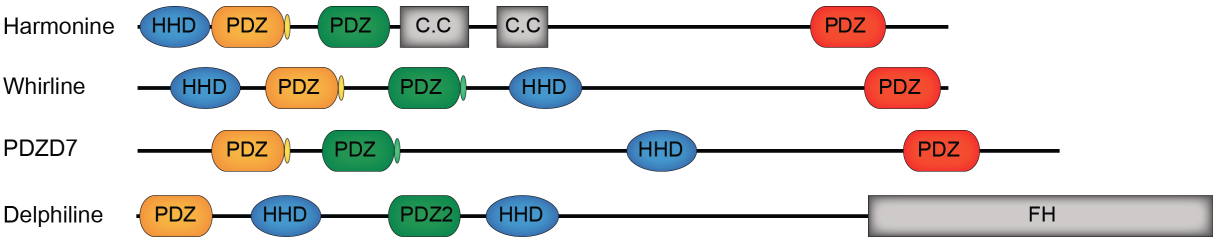


Figure 42 : Organisation en domaine des paralogues de la Whirline.

7.1.2 Extensions des PDZ

L'extension C-terminale du PDZ1 de l'Harmonine se structure en une épingle-ß. Cette structure additive est en interaction directe avec le PDZ1, avec le domaine HHD proximal ainsi qu'avec les ligands du domaine PDZ1. Le C-terminal du PDZ2 de l'Harmonine contient également une extension mais se structurant en une hélice courte de fonction inconnue, qui n'intervient pas directement dans la liaison aux partenaires du PDZ2. L'identification des HHD de la Whirline et de PDZD7 nous a encouragés à étudier la conservation d'autres structures entre ces protéines. Nous avons donc étudié les régions flanquantes des domaines PDZ de la Whirline et de PDZD7. L'alignement minutieux des séquences C-terminales des domaines PDZ1 et PDZ2 des trois protéines montre une conservation, faible. Trois résidus des brins de l'épingle-ß du PDZ1 de l'harmonine sont notamment conservés en C-terminal des PDZ1 et PDZ2 de la Whirline et de PDZD7; un résidu hydrophobe et une proline dans le premier brin, et un tryptophane dans le second brin (Figure 43). La conservation de ces

quelques résidus suggére leur structuration en épingle-ß dans les deux protéines. Dans l'Harmonine, les chaines latérales des trois résidus conservés sont en contact direct avec les chaines latérales des résidus des brins ß2/ß3 du domaine PDZ. L'orientation commune de ces chaines latérales renforce l'hypothèse du repliement en épingle-ß de ces régions et de leur localisation au contact du domaine PDZ, proche de la poche d'interaction (Figure 43). Les résidus constituant la surface d'interaction avec le domaine HHD ne sont eux pas conservés ; il est donc difficile de prédire l'interaction entre épingle-ß et HHD de la Whirline et de PDZD7.

La séquence de l'hélice en C-terminal de l'épingle-ß du PDZ1 de l'Harmonine n'est pas conservée chez la Whirline et PDZD7. Cette hélice, à l'extrémité C-terminale de la construction, n'est pas particulièrement stabilisée dans la structure du complexe Harmonine/SANS [71] et ses facteurs B sont largement supérieurs à ceux du reste de l'extension. Dans une récente structure de l'Harmonine en complexe avec un autre SAM/PBM, cette hélice est moins bien définie et est clairement dynamique [186]. Son rôle n'est pas formellement identifié et sa présence dans la construction de l'Harmonine pourrait être liée aux conditions de cristallisation et a l'empilement cristallin. Cette séquence en hélice n'est pas conservée en C-terminal des PDZ de la Whirline et de PDZD7 et semble unique dans cette famille de paralogues.

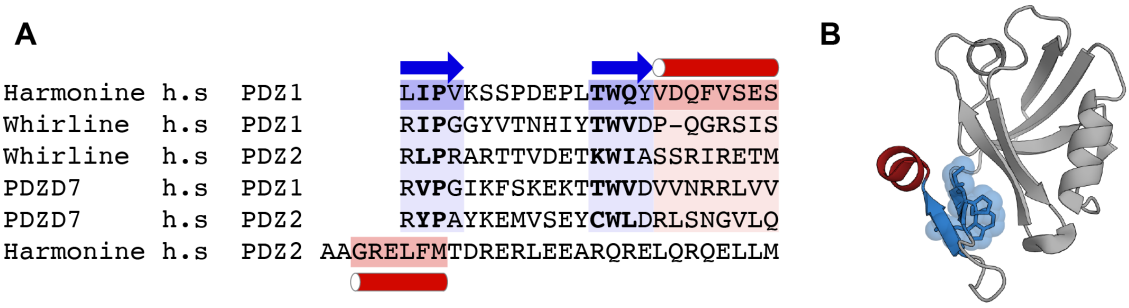


Figure 43 : Conservation des séquences épingles-ß de l'audition. (A) Alignement manuel des séquences flanquantes C-terminales des PDZ1 et 2 des trois protéines à PDZ de l'audition. Les structures secondaires indiquées correspondent à celles identifiées pour les PDZ de l'harmonine. **(B)** Structure du PDZ1 de l'Harmonine avec ses extensions en épingle-ß et hélice. Les résidus montrés en bâtonnets et sphères correspondent aux résidus conservés dans les deux PDZ de la Whirline et de PDZD7, (pdb : 3K1R).

J'ai ensuite étudié la conservation de la séquence des épingles-ß de la Whirline chez différentes espèces. En utilisant le serveur ConSurf (consurf.tau.ac.il/2016/), j'ai aligné et sélectionné manuellement 215 séquences d'orthologues de la Whirline. La visualisation de la conservation par position est ensuite faite en utilisant un générateur de logo de séquence

(weblogo.berkeley.edu/logo.cgi). Le résultat permet d'identifier sans ambiguïté les domaines, extrêmement bien conservés, et les régions les connectant bien moins conservées et de longueur variable (**Figure supplémentaire 12.2 fig3**). Cette visualisation de la conservation des séquences montre bien que les séquences en C-terminale des deux premiers domaines PDZ de la Whirline sont aussi bien conservées que celles des domaines, et bien mieux conservées que les séquences inter-domaines.

La similarité des résidus de l'extension C-terminale des domaines PDZ de la Whirline avec l'épingle-ß du PDZ1 de l'Harmonine, ainsi que leur conservation dans ses orthologues, nous a poussés à étudier le repliement de ces extensions et leur importance probable pour la fonction de la protéine [123]. Nous présenterons ces résultats dans le chapitre 9. La conservation des épingles-ß dans PDZD7 est aussi une question intéressante qui sera abordée dans les perspectives.

7.1.3 Délimitations et production des constructions

Au commencement du projet, aucune information n'était disponible sur l'interaction possible entre domaines de la Whirline. Nous avons donc pris le parti de définir des amorces délimitant chacun des domaines et nous permettant par la suite de cloner des constructions incluant toutes les combinaisons possibles de ces domaines. Nous avons ainsi cloné tous les domaines individuels ainsi que toutes leurs combinaisons deux à deux (**figure 44**). Pour les domaines PDZ1 et PDZ2, nous avons inclus leurs potentielles extensions C-terminales dans toutes les constructions. Nous avons ensuite complété cet ensemble initial par une construction incluant les quatre domaines N-terminaux de la protéine et correspondant à l'isoforme N-terminale courte de la Whirline, ainsi que par une construction incluant les deux premiers domaines PDZ et le HHD2. Au cours de la caractérisation des structures de la Whirline, nous avons généré de nombreux mutants ponctuels des constructions présentées, afin d'améliorer la stabilité des constructions, de tester le rôle fonctionnel de certains résidus ou d'introduire des sondes paramagnétiques pour la RMN. Au total, 43 constructions différentes auront été clonées, produites et purifiées en des quantités compatibles avec leurs caractérisations biophysiques (Figure 44).

Toutes les constructions sont clonées dans le vecteur pGST-parrallel 2, contenant une étiquette GST en amont de la séquence insérée et séparée de celle-ci par un site de clivage à la protéine TEV. Après clivage à la TEV, une séquence exogène GAMGS persiste en N-terminale de la séquence d'intérêt. Toutes les constructions sont exprimées chez *E. coli* et

purifiées en deux étapes de chromatographie : une première colonne d'affinité au glutathion suivie d'un clivage TEV sur colonne. La fraction éluée est ensuite passée au travers d'une colonne d'exclusion de taille, aux billes Sephacryl, S100 ou S200 selon la taille de la construction. Les domaines HHD1 et le PDZ1 ont un pI très élevé ; en conséquence, toutes les constructions incluant au moins l'un de ces domaines ont une forte tendance à lier les acides nucléiques de façon aspécifique. Pour éviter la contamination aux acides nucléiques, un lavage à 1M de NaCl a été effectué lorsque la protéine est fixée sur la colonne d'affinité. Aucune cystéine n'est présente dans la partie N-terminale de la Whirline ; en conséquence, les tampons de purification ne contiennent pas de réducteur.

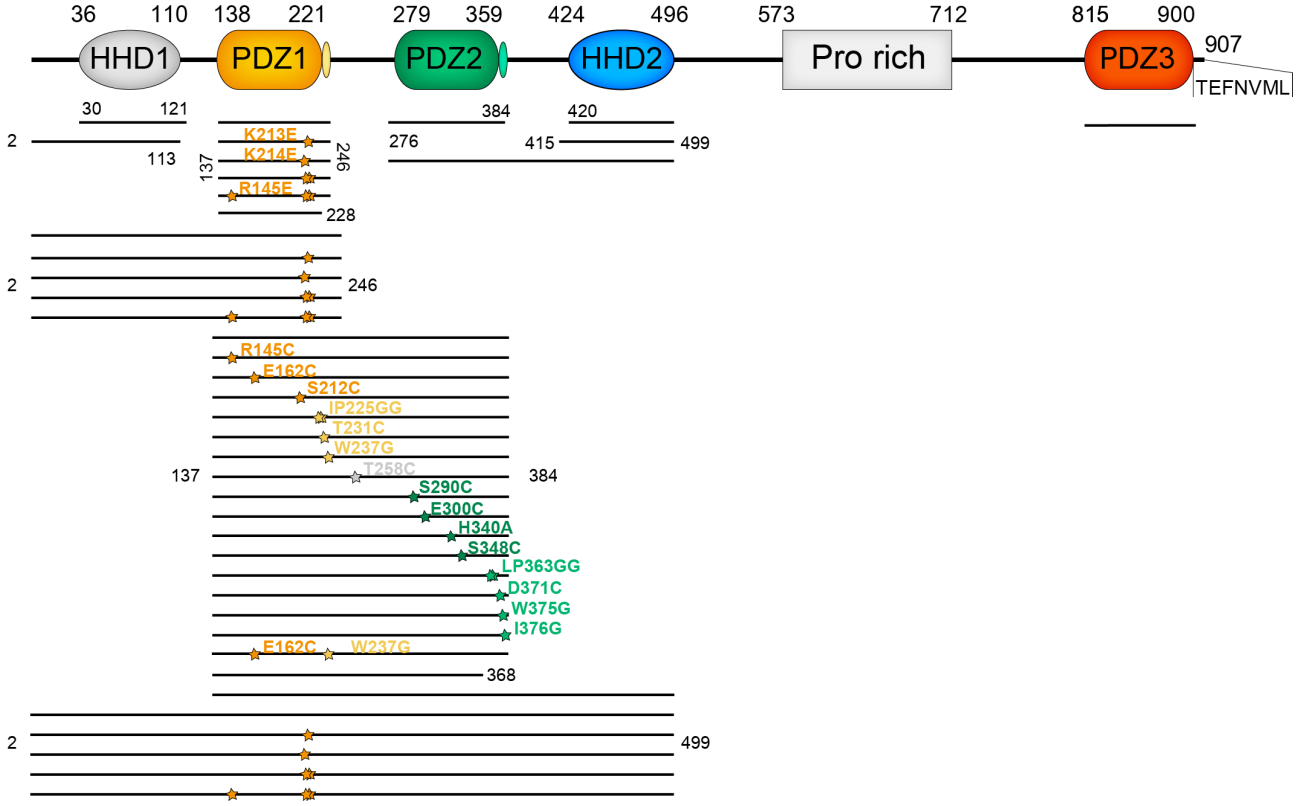


Figure 44 : Constructions produites au cours de la thèse. Séquence de l'isoforme 3 murine de la Whirline, la forme la plus proche de l'isoforme 1 humaine. Les étoiles représentent les substitutions générées par des mutations site-spécifiques afin d'attacher des sondes (via des Cystéines), améliorer la stabilité (inversions de charge) et étudier les surfaces d'interaction.

Cette purification simple permet d'avoir de très bons rendements pour toutes les constructions de la Whirline pour lesquelles 15 et 20 mg de protéine sont obtenus par litre de culture bactérienne. Les constructions de la Whirline sont particulièrement stables et peuvent être conservées pendant plusieurs mois à 4°C sans agrégation ou dégradation. La facilité de production et de conservation de la majorité des constructions de la Whirline aura permis de multiplier les expériences sans se soucier des quantités de protéines consommées, une situation particulièrement confortable tout au long de cette thèse.

7.2 Production et purification du Nter-PDZ1

Au début du projet, notre hypothèse de travail était que les deux couples HHD-PDZ de la Whirline forment des supramodules similaires à celui de l'Harmonine. Nous nous sommes donc d'abord concentrés sur l'étude des constructions bi-domaines PDZ2-HHD2 et HHD1-PDZ1. La production de cette dernière a cependant posé de nombreux problèmes.

La construction contenant strictement le HHD1 et le PDZ1-Hp1 (résidus 30-246) est entièrement retrouvée dans les corps d'inclusions chez *E. coli*. Nous avons testé différentes conditions d'expression, en faisant varier la température et le temps d'induction, ainsi que la quantité d'inducteur. Nous avons aussi modifié les conditions de lyse, en faisant varier le tampon (MES, Tris, Hepes), la concentration de sel (de 50 mM à 500 mM de NaCl ou KCl) et en ajoutant différents détergents (DDM et Triton X100); cependant, la construction n'a jamais pu être produite malgré nos efforts. Nous avons donc décidé de changer ses délimitations et de travailler avec une construction commençant au N-terminal de la protéine (Nter-PDZ1, 2-246), ajoutant la région de faible complexité en amont du HHD1. Cette construction s'est elle produite en large quantité dans nos conditions standards (Tris 50 mM ; NaCl 150 mM). Cependant, l'analyse par ultracentrifugation analytique (AUC) a très rapidement diminué nos espoirs (Figure 45).

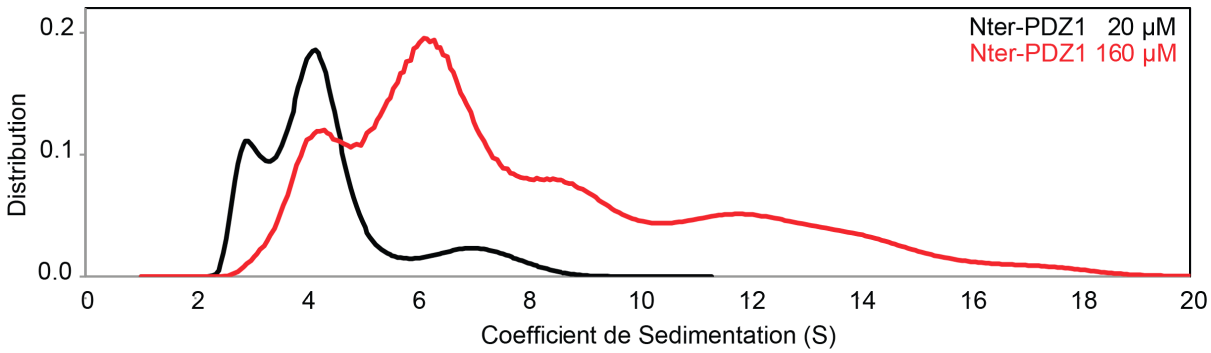


Figure 45 : Sédimentation de Nter-PDZ1. Résultats d'ultracentrifugation analytique de sédimentation de la construction Nter-PDZ1 de la Whirline à deux concentrations.

L'AUC montre que la construction à 20 µM est oligomérique en solution et que son état de polymérisation augmente avec la concentration. La position des pics de distribution des coefficients de sédimentation montre qu'à 20 µM, Nter-PDZ1 est majoritairement dimérique avec une plus faible proportion de monomére ainsi que des objets de plus grande taille. A 160 µM, de nombreuses espèces coexistent jusqu'à de grands degrés de polymérisation. Cette polydispersité ne conduit pas la construction vers l'agrégation mais rend cependant sa caractérisation biophysique difficile. Nous avons tenté d'améliorer la

dispersité de la solution de protéine en criblant les solvants de conservation. Comme pour la construction HHD1-PDZ1, le tampon, le pH, la concentration de sel et la présence de détergent ont été testés sans obtenir une condition où la construction est monodisperse. Le suivi de la taille des objets en solution par DLS nous aura cependant appris que la polymérisation est dépendante de la concentration en protéine et est réversible par dilution. Les fortes concentrations de sels (supérieures à 300 mM) et un pH faible (<5.5) permettent de réduire la taille des objets en solution, sans pour autant obtenir une forme monomérique ou au moins monodisperse. La polymérisation est ainsi probablement induite par des interactions polaires entre protomères.

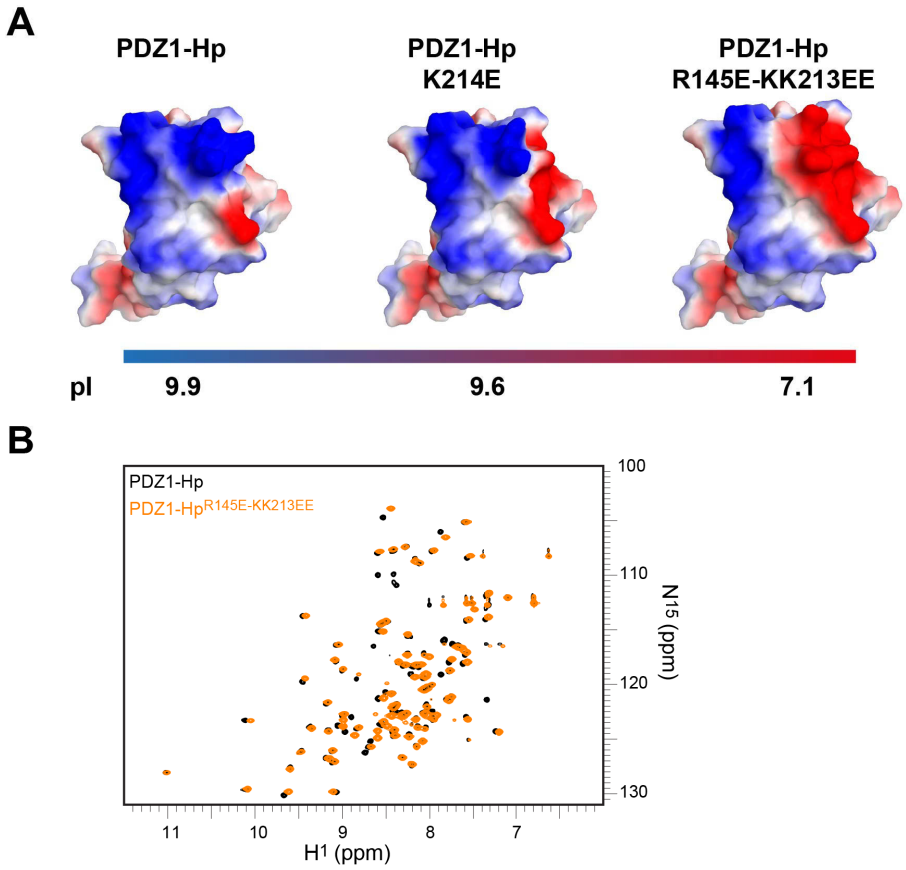


Figure 46 : Substitution sur le PDZ1 de la Whirline. (A) répartition des charges sur une représentation de la surface du PDZ1-Hp, d'une forme monosubstituée (PDZ1-Hp^{K214E}), ou d'une forme triplement mutée (PDZ1-Hp$^{R145E\text{-}KK213EE}$). Les valeurs théoriques de pI sont indiquées sous les représentations. **(B)** superposition des spectres ^{1}H-^{15}N HSQC de PDZ1-Hp (noir) et de PDZ1-Hp$^{R145E\text{-}KK213EE}$ (orange).

L'échec de l'optimisation de tampon nous a amené à considérer la mutagénèse comme stratégie pour stabiliser la construction. Dans leur travaux sur la structure de Nter-PDZ1 de l'Harmonine, l'équipe du Dr. Zhang a été confrontée à un comportement similaire de l'Harmonine, comme décrit dans les supplementary data de leur article [251] ("*The quality of the*

^{1}H-^{15}N HSQC spectrum of the fusion protein [NPDZ1] *was very poor, presumably due to nonspecific self-association of the protein. Screening of different sample conditions revealed that inclusion of high concentrations of salt in the sample buffer led to dramatic improvements of the sample behavior"*). Pour prévenir la polymérisation par interaction polaire, les auteurs ont systématiquement muté les résidus chargés de la construction, en inversant leurs charges. Cette approche leurs a permis d'identifier une mutation ponctuelle K157E, dans la boucle α2/β5 du PDZ1 de l'Harmonine, rendant la construction monomérique et monodisperse. Cette lysine est conservée dans le PDZ1 de la Whirline ; elle est cependant incluse dans un patch cationique qui n'est pas présent dans l'Harmonine. Ce patch cationique est composé d'une arginine (R145) dans la boucle "GLGF" et de deux lysines successives (K213, K214) dans la boucle α2/β5. Nous avons tenté quatre combinaisons d'inversion de charge de cette région : K213E, K214E, KK213EE, R145E-KK213EE sur les constructions Nter-PDZ1 et Nter-PDZ1-PDZ2-HHD2 (2-499) pour réduire leur degré de polymérisation (**Figure 46A**). Les mutations ont aussi été effectuées sur la construction PDZ1-Hp, pour vérifier leur impact sur la structure du PDZ1. La comparaison des spectres ^{1}H-^{15}N HSQC de PDZ1-Hp et de sa forme triplement mutée, PDZ1-Hp$^{R145E-KK213EE}$ permet de conclure que la structure globale du domaine n'est pas affectée par les trois mutations de ce patch cationique (**Figure 46B**).

Malheureusement, aucune de ces constructions n'a permis d'obtenir une forme monomérique de la construction Nter-PDZ1 ou de la construction contenant les quatre domaines N-terminaux. Les mutations diminuent le degré de polymérisation, mais elles ont aussi modifié le comportement de la construction vis à vis de la concentration en sel, ne permettant plus de réduire la polymérisation en augmentant la concentration saline. Après optimisation de tampon sur la construction triplement mutée, Nter-PDZ1$^{R145E-KK213EE}$, nous avons identifié une condition de tampon (MES 20mM, NaCl 100mM, pH6.0) où la construction est éluée peu après le volume mort de la colonne d'exclusion de taille, à la différence de toutes les autres conditions testées auparavant **(Figure 47 A,B)**. Nous avons caractérisé la construction dans ces conditions par SAXS. La solution protéique est monodisperse (pas de variation de Rg au cours de l'élution sur colonne, donc en fonction de la concentration), mais l'objet correspond probablement à du trimère (soit 3 *29 kDa) **(Figure 47 C,D)**.

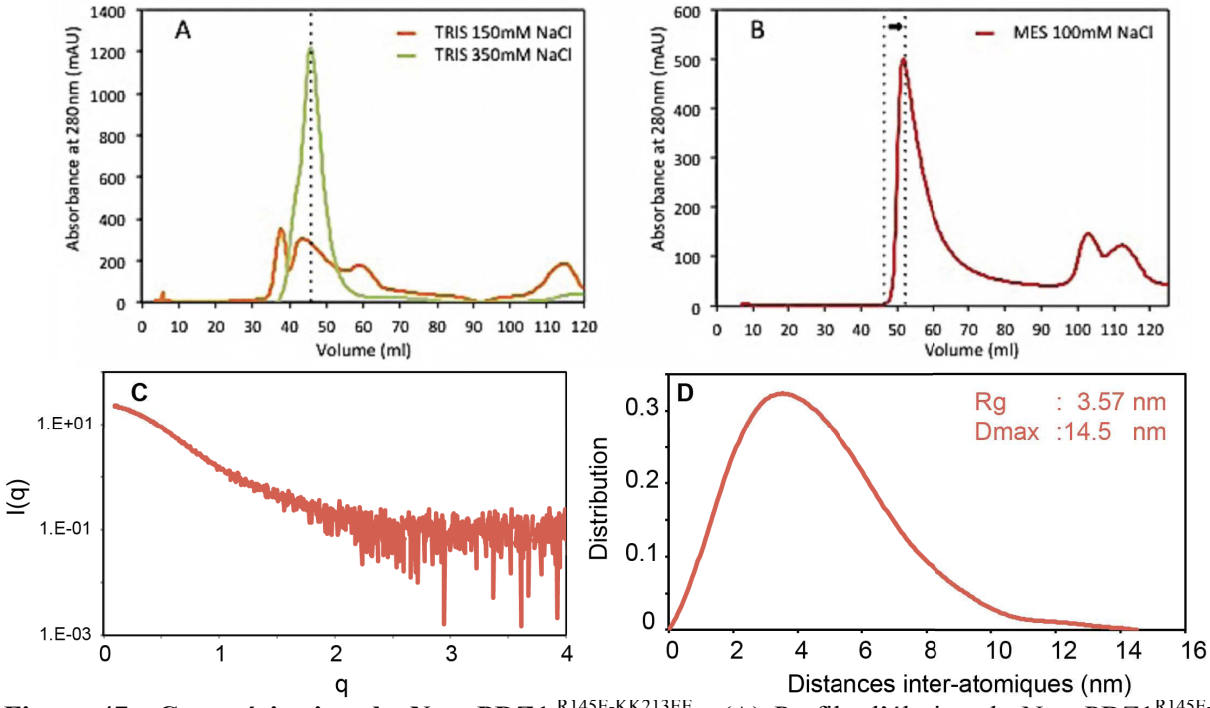


Figure 47 : Caractérisation du Nter-PDZ1 $^{R145E\text{-}KK213EE}$ **. (A)** Profils d'élution de Nter-PDZ1$^{R145E\text{-}KK213EE}$ sur colonne S200 en Tris-HCl pH7.5 à deux concentrations en sels, 150 mM NaCl (orange) et 350 mM NaCl (vert) **(B) profil d'élution** de Nter-PDZ1$^{R145E\text{-}KK213EE}$ sur colonne S200, après optimisation, MES 20 mM pH6.0 et NaCl 100 mM. (**C**) courbe de diffusion de lumière aux petits angles de Nter-PDZ1$^{R145E\text{-}KK213EE}$ en tampon optimisé. (**D**) Distribution de distance interatomique extraite de la courbe présentée en **C** et paramètres extraits Rg et Dmax.

La polymérisation, dans toutes les conditions testées *in vitro*, est une propriété intrinsèque de la construction contenant les quatre domaines C-terminaux de la protéines (2-499), cette séquence est similaire à l'isoforme N-terminale courte naturelle de la Whirline (2-543) exprimée dans les cellules photoréceptrices. La polymérisation de la Whirline pourrait-être plus qu'un simple artefact observée *in vitro* mais avoir un rôle physiologique. Cette propriété a d'ailleurs été suggérée dans le cas de l'Harmonine. Il a été proposé que l'Harmonine est condensée au niveau des points d'ancrage du lien apical, par son interaction multiple avec la Cadhérine-23 [68]. Le rôle de cette polymérisation serait de diffuser sur une large surface de membrane les forces appliquées sur le lien apical lors de la déflection des stéréocils. Un rôle similaire peut être proposé pour l'oligomérisation intrinsèque de la Whirline, même si aucune expérience cellulaire ne vient actuellement appuyer cette hypothèse.

8 Résultats 2 : *Structure of the second HHD of Whirlin*

8.1 Résumé

La Whirline est une protéine d'échafaudage impliquée dans le développement et le maintien de la structure de touffes mécanosensibles de stéréocils, dont la déflection par la vibration sonore déclenche l'influx nerveux. Dans sa moitié N-terminale, la Whirline contient une succession de quatre domaines rapprochés : deux domaines HHD entourant deux domaines PDZ. Cette première partie est suivie d'une grande région non structurée et riche en proline séparant le troisième domaine PDZ, isolé à l'extrémité C-terminale. Dans les stéréocils, elle est exprimée en plusieurs isoformes, avec des localisations et des partenaires différents. L'isoforme longue de la protéine, est impliquée dans la formation de liens inter-stéréocils transitoires nécessaires au développement des touffes ciliaires. La forme C-terminale courte commence avant le HHD2 et s'étend jusqu'au C-terminal ; son expression est contrôlée par un promoteur différent des autres isoformes. Elle est localisée au sommet de la plus grande rangée de stéréocils où elle est impliquée dans la régulation du cytosquelette d'actine en formant un complexe avec la Myosine-15a et Eps8.

Dans cet article, nous nous concentrons sur la structure du domaine HHD2, seul domaine contenu dans toutes les isoformes de la Whirline. Nous avons déterminé sa structure par cristallographie, dans une conformation à cinq hélices commune aux domaines HHD et dans une seconde conformation plane à trois hélices, et formant un dimère par échange de structures secondaires recréant ainsi le cœur hydrophobe du repliement globulaire monomérique. Par RMN, nous avons montré que le HHD2 de la Whirline a une conformation canonique à cinq hélices en solution. Nous nous sommes ensuite intéressés au rôle de ce domaine dans le contexte multi-domaine de la protéine. Ainsi HHD2 n'est pas en interaction avec le tandem de domaine PDZ présent en son N-terminal. Nous proposons ainsi que ce domaine est lié à la fonction de l'isoforme C-terminale courte. Le rôle du HHD2 dans l'isoforme C-terminale est aussi soutenu par le fait que celle-ci possède plusieurs partenaires interagissant simultanément avec elle et dont l'interaction ne peut pas être uniquement contrôlée par le domaine PDZ3. Eps8, l'un de ses partenaires, ne possède pas de PBM, mais interagit avec l'isoforme courte par une grande région non structurée possédant plusieurs hélices isolées amphipathiques, pouvant constituer des motifs de reconnaissance du domaine HHD2 de la Whirline.

8.2 Article : *High Resolution structure of the HHD2 of Whirlin*

Florent Delhommel[1,2,3], Florence Cordier[2,3], Frederic Saul[2,4], Lucas Chataigner[2,3], Christine Petit[5,6], Ahmed Haouz[2,4], Nicolas Wolff[2,3]

1. Unité de Résonance Magnétique Nucléaire des Biomolécules, Institut Pasteur, Paris, France

2. CNRS, UMR3528, Paris, France

3. Cellule Pasteur UPMC Univ. Pierre et Marie Curie, Paris, France

4. Plateforme de Cristallographie, Institut Pasteur, Paris, France

5. Unité de génétique et physiologie de l'audition, INSERM UMRS1120, Institut Pasteur, Paris, France

6. Collège de France, Paris, France

Abstract:

Whirlin is a protein essential to the function of sensory neurones, mutations of its associated gene are responsible for syndrome associating congenital deafness and progressive blindness. In the mechanosensing structure of sound-sensing cells, Whirlin is expressed in two isoforms of different localisations and functions. The full-length isoform anchor cell-junction proteins to the cytoskeleton while the short, C-terminal isoform takes part in a complex regulating cytoskeleton growth. Here, we describe the high-resolution structure of the HHD2 domain of Whirlin, the only domain contained in all its isoforms. We determined the structure of the domain in two conformations, a canonical five-helix monomer and a three-helix swapped-dimer and showed that in solution, only the five-helix monomer is observed. We also showed that the HHD2 does not interact with its preceding domains and that its function is most likely related to the short C-terminal isoform of the domains.

8.2.1 Introduction

The Usher syndrome is the leading form of hereditary hearing loss associated to blindness [46]. It is characterized by congenital deafness, progressive vision loss at varying onset and in some forms vestibular deficiency. This genetically heterogeneous syndrome is caused by mutations on the associated genes of 11 identified proteins, named Usher proteins [46]. They are involved in the molecular architectures of the specialized sensory cells of these two senses [252]. Hearing is based on hair cells utilizing mechanosensitive bundles of stereocilia that are deflected by sound-evoked vibrations [253], the Usher proteins are essential for the morphology of the mechanosensitive hair bundle and its development. Vision, on the other hand, requires light-sensitive cells that are highly compartmentalized. Their inner segment of the cell is dedicated to the constant production of rhodopsin containing vesicles transported through a small connecting cilium to regenerate the stacked photoreceptive membranes of the outer segment [254]. Usher proteins are heavily involved in structuring the intersegment connecting cilium controlling the transport of newly synthetized vesicles to the outer segment [45]. They are also found in the synaptic connection of those two sensory cell types. Molecularly, the Usher proteins are either extracellular link or multi-domains adaptor proteins that form large and intricate network of scaffolding.

Whirlin is a large Usher protein differentially expressed in three main isoforms in sensory cells [221]. Mutations of the gene associated to whirlin are responsible for Usher syndrome or non-syndromatic deafness in correlation with the position of the mutation along the sequence [216,232]. Mutations affecting the N-terminal half of the protein affect retinal and cochlear functions while C-terminal mutations are the cause of deafness without other associated symptoms [221]. The N-terminal short isoform contains two HHD (Harmonin Homology Domain) and two PDZ domains and is lowly expressed in photoreceptors, in which its function remain elusive and challenging to separate from the full-length isoform of whirlin with which it seems to share the localization [223]. The full-length isoform has an additional proline-rich region, a third PDZ domain and a motif of interaction with PDZ domains at its C-terminal extremity and is expressed in both sensory cell [46]. In photoreceptors, it is likely to anchor extracellular links formed by Vlgr1 and Usherin to the cytoskeleton by interacting with other adaptors proteins SANS and PDZD7 [183,214]. A similar complex between these proteins occurs transiently during the development of auditory hair cells and is critical for the correct morphology of the stereocilia bundle and its growth [223]. The localization of the

full-length isoform of whirlin in mature hair cells is more controversial, it has been reported either at the top of the tallest stereocilia row [221], or in the middle of the shaft of the intermediate size row [226]. The C-terminal short isoform starts before the second HHD and extend to the C-terminus of the protein. The expression of this isoform is controlled by a different promoter and is only present in hair cells [216,221], in which it localizes at the top of the tallest stereocilia row and is involved in a multi-protein complex with actin associated proteins Eps8 and Myosin15 [237].

The HHD2 is the only domain present in all isoforms of whirlin [221]. This small helical domain has been widely overlooked in previous studies due to the poor similarity between HHD, its identification was fortuitous and recent [63]. HHDs form a family of 9 identified domains over 6 human proteins (**Figure 1a**) [63]. They are small domains of 80 amino acids folded in a compact five alpha-helices bundle [64]. Structures of two HHD, from harmonin and CCM2 proteins, have been solved thus far [64,65,67]. HHD domains bind isolated amphipathic helix of their partners in a groove formed between their helices $\alpha 1$ and $\alpha 2$ and a conserved hydrophobic patch on the $\alpha 4/\alpha 5$ loop. In addition to this main mode of interaction, the two known structures of HHD are prompt at interacting with larger domains from other surfaces [67,71]. On one hand, the $\alpha 3/\ \alpha 4$ loop of HHD of harmonin binds to an hairpin extension of its adjacent PDZ domain, independently of the helix motif binding. On the other hand, the HHD domain of CCM2 binds its partner protein MEKK3 through its N-terminal helical motif and also interacts with the downstream PB1 domain on a large exposed surface of the helices $\alpha 2$ and $\alpha 3$. Thus, HHD seems to offer plasticity of interaction with their partners.

In this article, we present the high-resolution crystallography structures of the HHD2 domain of whirlin. This new member of the HHD family adopts two distinct conformations, a canonical 5-helix bundle and a 3-helix fold stabilized in a swapped-dimer. We confirmed the only conformation observed by solution NMR is the canonical 5-helix form of the domain. We also showed that in the larger molecular context of whirlin, the HHD2 was independent from the N-terminal supramodule and that its function is likely mainly associated to the C-terminal short isoform of whirlin.

8.2.2 Results

8.2.2.1 HHD containing proteins comprehensive list

Most HHDs have only been recently identified in their associated proteins. The weak sequence similarities of these 9 domains make them challenging to identify from sequence analysis only [217] (**Figure 1b**). Their residues exposed to the solvent are highly divergent and their sequence similarities are strictly limited to the residues forming the hydrophobic core (19/75 residues) [63]. The HHDs are mainly found in four paralog proteins that are related to hearing and synaptic processes, namely harmonin, whirlin, PDZD7 in hearing and delphilin in synapses. Strikingly, in these four proteins, the HHD modules are in close vicinity of PDZ domains with conserved hairpin extensions. However, their number of copies and their relative positions in the sequence are unique to each protein. In the case of harmonin, the sole HHD is a few residues apart from an extended PDZ domain and the two domains form an intramolecular complex via an hairpin extension of the PDZ domain [251]. PDZD7 also contains an HHD domain surrounded by low complexity regions and is furthest from another domain than any other HHD. Delphilin shows an alternation of HHDs and PDZ domains without extension. Whirlin has two HHD surrounding a PDZ tandem that forms a supramodule *via* the PDZ hairpins extensions. The co-occurrence of HHD and extended PDZ domain may indicate a complementarity of function and make it plausible that supramodular interaction similar to the one identified in harmonin might occur in this group of paralog proteins.

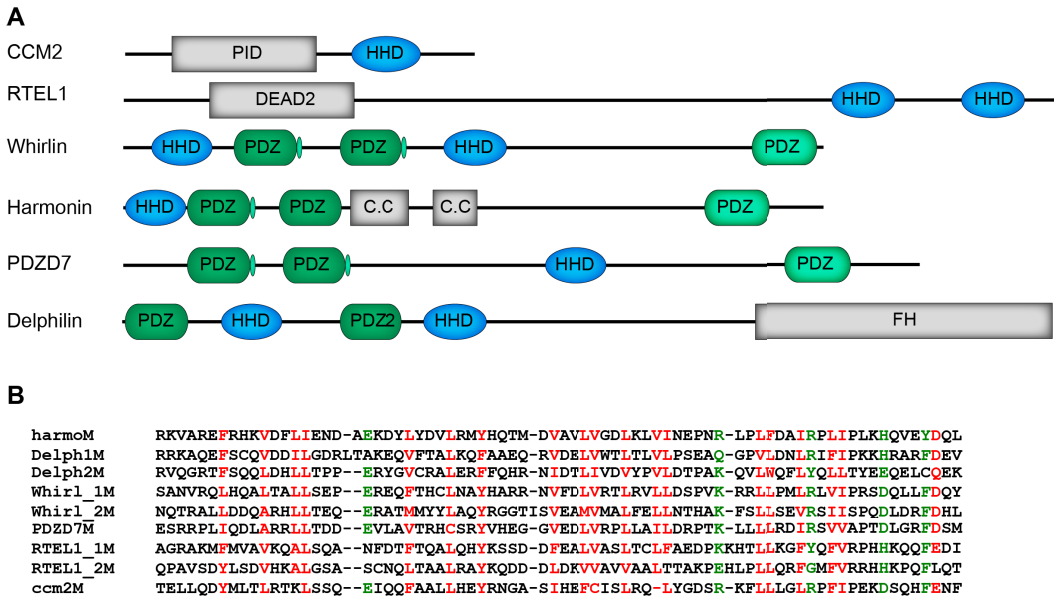


Figure 1 : HHD in proteins. A Domain organisation of the six known HHD-containing proteins from the mouse proteome. **B** Sequence alignment of the 9 known HHD from the mouse proteome. Family-defining position are shown in red and other well conserved positions are shown in green.

8.2.2.2 The canonical fold of whirlin HHD2

As for others HHD, the HHD2 of whirlin has not been investigated or reported in most studies. We determined its high-resolution structure by X-ray crystallography using SAD method and refined against a 1.7 Å native dataset (**table 1**). The asymmetric unit of HHD2 crystal structure contains six molecules organised in a dimer of trimer. The high number of copy per asymmetric unit and the slight differences of conformation compared to all other available structures of HHD made it unsolvable by molecular replacement and direct phase determination was performed using the anomalous scattering of seleno-methionine derivative.

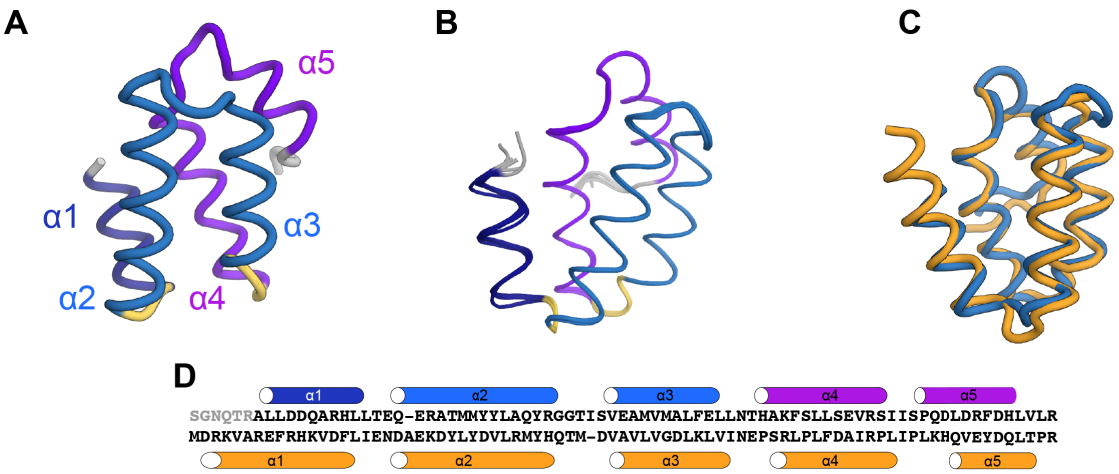


Figure 2: Structure of the HHD2 in the five-helix conformation. A Structure of the HHD2 of whirlin, from the chain D from the crystallography structure. **B** Superposition of the six HHD2 molecules from the asymmetrical unit. Root mean square deviations between molecules vary from 0.16 to 0.21. **C** Superposition of HHD2 of Whirlin with the apo form of the HHD of harmonin (pdb 2KBQ), showing a RMSD of 1.3. **D** Sequence alignment between the HHD2 of Whirlin (top) and the HHD of Harmonin (bottom). Secondary structures are reported.

The structures of the six copies in the asymmetrical unit show the canonical five-helix fold expected for this domain (**Figure 2a**). The four last helices and their connecting loops of the six subunits superimpose very well with a root mean square deviation on Ca ranging from 0.16 to 0.21, indicating that the crystal packing has no influence in the overall structure of the HHD2 domain (**Figure 2b**). The only slight divergences are observed for the first helix of the domain, which has an orientation varying of a few degrees between HHD2 in the asymmetric unit. In addition, the five N-terminal residues of the first helix are absent from the density map, underlying the dynamic nature of the first helix. The first helix only contains three turns, and is shorter than expected from sequence based secondary structure predictions (7 turns) and from structures of CCM2 (4 turns) and harmonin (5 turns) HHD domains (pdb: 4FQN and 2KBQ, respectively).

The helix orientation of the HHD2 of whirlin is similar to the other known structure of an unbound form of HHD, in harmonin (**Figure 2c**). Their five helices superimpose well with a RMSD on Cα of 1.3 Å. The unbound HHD of harmonin is more compact than its bound form (PDB code 2LSR). The helices α 1 and α 2 adopt a different orientation to bind the helical motif of the partner protein.

8.2.2.3 The HHD2 can form a swapped dimer

In addition to the five-helix conformation of the HHD2 of whirlin, we solved the structure of a non-canonical conformation of the domain at 2.1A resolution in other crystallographic conditions, containing 20% propan-2-ol and XXX. In this structure, the individual domain adopts a three-helix fold and dimerizes with another molecule (**Figure 3a**). Folding of the loops $\alpha 1/\alpha 2$ and $\alpha 3/\alpha 4$ in α-helical turns drives the global change of conformation. This loop folding merges the helices $\alpha 1/\alpha 2$ and $\alpha 3/\alpha 4$ that creates a planar three-helix conformation of each protomer

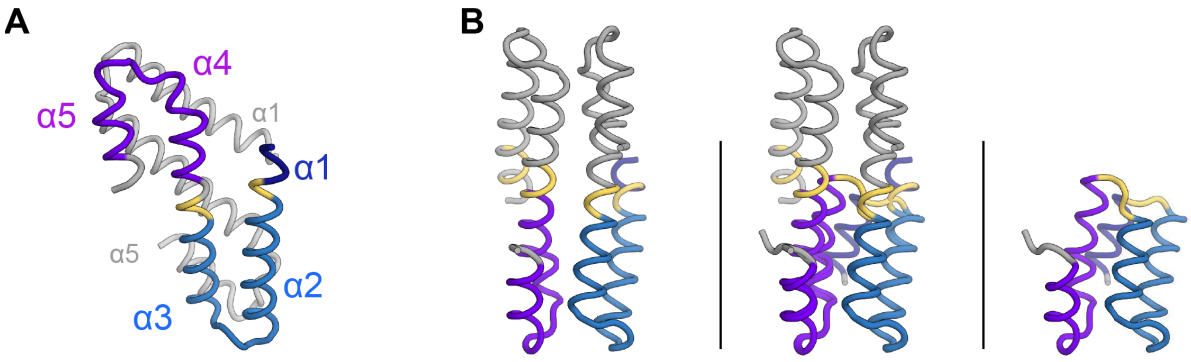


Figure 3 : HHD2 swapped-dimer conformation. A Structure determined by crystallography of the swapped-dimer form of the HHD2 of Whirlin. **B** Superposition of the five-helix conformation of whirlin and the three-helix swapped-dimer.

By precluding the globular fold of the domain, the three-helical conformation leaves the hydrophobic core of the protein accessible to the solvent. In the crystal, HHD2 forms a symmetric dimer with the $\alpha 2/\alpha 3$ of one protomer interacting with the $\alpha 4/\alpha 5$ of another (**Figure 3b**). This dimeric conformation protects the hydrophobic side chains of its core from the solvent and almost restores the hydrophobic contacts observed in the canonical five-helix fold. However, due to the different orientation of helix a1 in the swapped-dimer conformation and subsequently the absence of contact between helix a1 and helices a2 and a4, helix a4 is distorted and twisted by 1/3 of a helix turn before L477 to optimize the protection of L476 and L477. In this way, the hydrophobic core is rearranged at a minimal entropic cost.

Interestingly, most of the essential electrostatic interactions (salt bridges and H-bonds) maintaining the 5-helix bundle are still present in the 3-helix dimer. Those are located on the side of the bundle opposite to the hinge region. In the dimer, they are either strictly conserved, like R481…D493, or conserved but swapped between the two subunits, such as Y450…D488 and R491…E458. On the other side, due to the conformational rearrangement of the hinge

region, the strongly connected network of bifurcated H-bonds and salt bridges T436…E439…K473…N469 is lost in the dimer and replaced by weaker α-helical NH…O hydrogen bonds (T436…R432, E439…L435, K473…N469).

In summary, the monomer to dimer conformational change is accompanied by a rather compensated rearrangement of the hydrophobic and electrostatic forces, though the dimer might be slightly less favorable than the monomer and the energetic barrier might be substantial. Overall, the four helices (α2 to α5) of the swapped dimer superimposes well with helices α2-α5 of the monomer (RMSD = 0.39). However, the 12 N-terminal residues are undefined in the electron density of the dimer, showing an increased disorder or conformational heterogeneity of helix α 1 compared to the five-helix structure.

The 5-helix bundle is predominant is solutionTo determine the conformation of HHD2 in solution, we used NMR spectrometry and assigned the backbone resonances of the domain. 94% of the backbone resonances associated to the 79 residues of HHD2 were assigned (**Figure 4A**). The resonances of the four N-terminal residues (G420-T423) were too broad to be observed in the spectra, due to exchange in the intermediate regime. Analysis of the secondary structure elements was achieved using backbone chemical shifts with the TALOS-N prediction server. This dihedral angle prediction unambiguously confirms the fold of HHD2 in five distinct helices matching the canonical fold and not the three-helix fold of the swapped dimer (**Figure 4B**). To further confirm the canonical five-helix conformation observed by crystallography, we compared the residual dipolar coupling (RDC) measured on HHD2 with the theoretical ones calculated from the amide bond orientation in HHD2 crystal structure. Indeed, ^{1}H-^{15}N RDC are sensitive to the orientation of the N-H bond vector with respect to the alignment tensor of the molecule and is a powerful source of information on molecular conformation. Experimental RDC are compared to the theoretical RDC values predicted from the orientation of amide bonds from the five-helix or three-helix dimer conformations. The experimental data correlate with a coefficient of 0.93 for the five-helix monomer and of 0.67 for the three-helix dimer form, unambiguously confirming the five-helix conformation in solution (**Figure 4C, D)**.

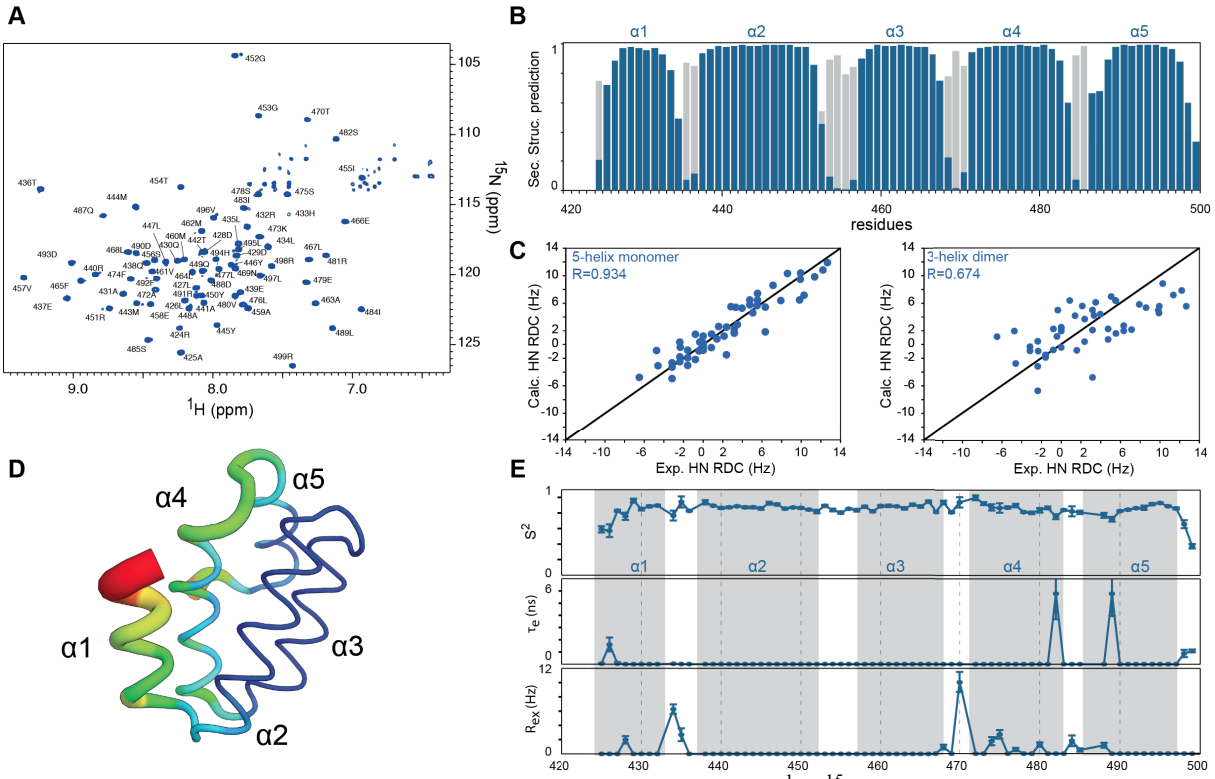


Figure 4 : HHD2 behaviour in solution and dynamic. A ^{1}H-^{15}N HSQC of HHD2 with the assignment. **B** Secondary-structure prediction per residues, based on the backbone assignment of the HHD2 of Whirlin, showing five independent helixes. **C** Correlation between measured RDCs in solution and the back-calculated values from two crystallographic structures: 5-helix monomer in the left panel and 3-helix swapped dimer in the right panel. **D** B-factor of the HHD2 represented on the chain D of the crystallography structure. B-factor are indicated by the size of the cartoon chain and the by the color, from low in dark blue to high in red. **E** ^{15}N relaxation data at 600 MHz of HHD2 at 25°C, as fitted by the model-free formalism of Lipari and Szabo The order factor S^2 along the sequence is shown on the top panel. The effective internal correlation time, τe is shown in ns in the middle panel. The relaxation rate due to chemical exchange Rex is shown in Hz in the bottom panel.

We concluded that the whirlin HHD2 adopts the canonical fold both in crystals and in solution. It is likely that the swapped conformation is induced by the strong crystallographic conditions in which the HHD2 crystal grew. The presence of 20% propan-2-ol in the solution might have promoted the helical folding of the loop regions. The induction of a-helical conformation by alcohols has been extensively described in the literature. The propan-2-ol might affect HHD2 conformation by weakening hydrophobic contacts and/or increasing local intramolecular polar bonds [255].

The dynamic behaviour of HHD2 was evaluated by NMR relaxation experiments, recording T1, T2 and heteronuclear NOE experiments. T_1, T_2 and NOE were analyzed in terms of internal motions, using the simple or extended model-free Lipari–Szabo approach.

Isotropic rotational correlation time (τc), estimated from an isotropic model in Tensor, is 5.8 ns. This value is in good agreement with the theoretical value of 4.7 ns calculated for a globular domain with a molecular weight of 9 kDa. Whirlin HHD2 is monomeric in solution. The domain is highly ordered from L427 to the R497 (**Figure 4E**). The first four residues of the structure are missing in the NMR spectra and the following two residues of helix α1 have lower order parameters than the rest of the structure. These results confirm the dynamic behavior of the first helix, which correlates with the disorder observed by X-ray diffraction with high values of B-factors. In contrast to other HHD domains, the first turn of the helix α1 for whirlin HHD2 is in exchange between conformational states as monitored by the low electron density from X-ray diffraction and the line-shape broadening of the resonances in NMR. Interestingly, the two other regions affected by significant conformational exchange in solution are the two "hinge" loops $\alpha 1/\alpha 2$ (433-437) and $\alpha 3/\alpha 4$ (470-472) that adopt an helical conformation in the swapped-dimer, despite being rigid on the fast time scale regarding B-factors in the crystal structure (**Figure 4D**).

8.2.2.4 HHD2 is independent from the N-terminal supramodule of whirlin

We then investigated, using small angle X-ray scattering and NMR, the potential intramolecular interaction of HHD2 with adjacent PDZ domains in the larger molecular context of whirlin. The two N-terminal PDZ domains of whirlin form a supramodule (PDZ1-2) that adopts a predominant close conformation in solution. The N-terminus of HHD2 is tethered to this supramodule by a 41-residues linker. The scattering curve of PDZ2-HHD2 yields large Rg (radius of gyration, 31.2 Å) and Dmax values (Maximum distance, 145 Å) (**fig.XX**). The distance distribution function P(r) shows a major population at small distances (31 Å) and slowly decreases up to long distances (145 Å), characteristic of independent domains linked by a flexible linker. By comparison, the larger supramodule construct PDZ1-2 has significant smaller Rg (27.5 Å) and Dmax (110 Å) values and a more compact distribution of interactomic distances (**Figure 5B**). The construct encompassing the three domains, PDZ1-PDZ2-HHD2, has similar size parameters than PDZ2-HHD2 (Rg: 33.2 Å and Dmax: 150 Å), the addition of the PDZ1 increasing only slightly the maximum interactomic distances. These results confirm that the three-domain construct does not form a three-domain supramodule. PDZ1 and PDZ2 still interact in the context of the three-domain construct and HHD2 is fully exposed at the end of a flexible region.

To confirm the global conformation information obtained by SAXS, the ^{1}H-^{15}N HSQC spectra of HHD2 and PDZ1-PDZ2-HHD2 constructs were compared. The good superposition of spectra and the marginal CSP (chemical shift perturbations, mostly < 0.02 ppm) observed for the resonances of the residues of HHD2 isolated or in the three-domain construct confirm that the domain is not interacting with the upstream PDZ tandem (**Figure 5 A, C**). This also confirms that HHD2 adopt the same conformation isolated or in the context of a larger construct. The only significant observed perturbations occur at the N-terminus of the protein, most likely explained by the impact of the PDZ2-HHD2 linker region. Thus NMR and SAXS data are fully consistent, HHD2 is linked to the PDZ2 domain by a flexible linker and does not interact with the N-terminal PDZ supramodule of whirlin.

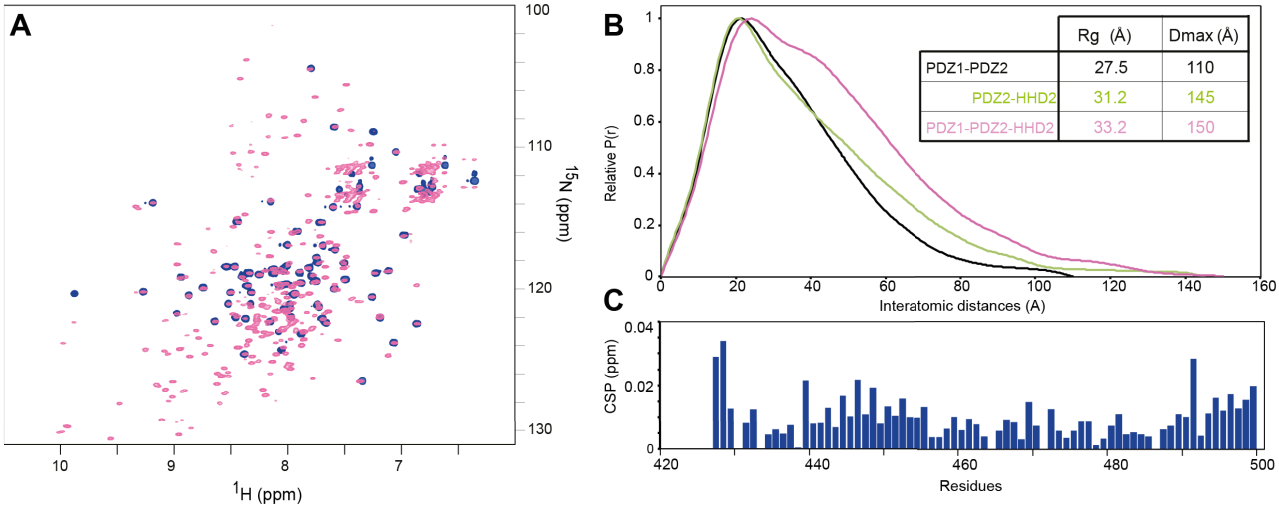


Figure 5 : HHD2 in the molecular context of whirlin. A Superimposed ^{1}H-^{15}N HSQC of HHD2 (blue) and PDZ1-PDZ2-HHD2 (pink). **B** Interatomic distance distribution P(r) of PDZ1-PDZ2 (black), PDZ2-HHD2 (green), PDZ1-PDZ2-HHD2 (pink) determined from small angle X-ray scattering data. The embedded table shows the size parameters Rg and Dmax of each of the three samples. **C** Chemical shift perturbations per residues of the ^{1}H-^{15}N resonances of HHD2 between its isolated construct in the tri-domains construct.

8.2.3 Discussion

Here, we report the structure of the second HHD of whirlin. Crystallography study revealed two potential conformations of the HHD2 domain of whirlin, a monomeric five-helix fold similar to the ones of other members of the HHD family and an original three-helix fold, forming a swapped dimer. Most of the electrostatic and hydrophobic interactions present in the monomer are preserved in the dimer. The only significant structural differences with the monomer are the conformation of the first helix and the orientation of the helix $\alpha 4$. The NMR data unambiguously established that HHD2 adopts a five-helix fold in solution, very

similar to the one determined by X-ray crystallography. Using NMR and X-ray scattering, we also showed that HHD2 is linked to the preceding PDZ supramodule by a flexible linker, that it does not interact with PDZ1 or PDZ2 and that it conserves the five-helix fold in multi-domain constructs of whirlin. The three-helix swapped dimer observed in crystallography is an interesting alternative conformation of HHD2. It is suspected that the PDZ-containing proteins involved in hearing form dimer or oligomer in solution to ensure their function [68,214].

The structure of the five-helix fold of whirlin revealed a shortened first helix compared to the other known structure of HHD. In the crystal, the first four residues of the construct are flexible. The observed B-factors correlate well with the NMR relaxation data that confirms that the region is also highly flexible and affected by conformational exchange in solution. In harmonin and CCM2, the first residues of the $\alpha 1$ are directly involved in ligand binding. In both cases, the HHD conformation changes upon peptide binding. The helix $\alpha 1$ of the HHD is closely packed to the helix $\alpha 2$. In peptide-bound structures, the orientation of the $\alpha 1$ is opened by a few degrees, exposing the hydrophobic residues of the helix $\alpha 2$ and of the loop the $\alpha 4/\alpha 5$. It is likely that upon peptide binding, the helix $\alpha 1$ of HHD2 adopts a helical conformation with its hydrophobic residues directly interacting with the helix of the partner.

HHD2 is the only domain present in the three main isoforms of whirlin; however its function and its partners are currently unknown. The long isoform of Whirlin is involved in maintaining the morphology of photoreceptors [256] and is required at the ankle-link complex during hair cells development [33]. At these two localizations, Whirlin interacts by its N-terminal PDZ tandem, with motifs contained in the cytoplasmic part of Usher membrane proteins forming the extracellular links of stereocilia [214]. We showed recently that in solution, these two N-terminal PDZ domains form a transient supramodule, increasing significantly the affinity of the individual PDZ for its partners. To obtain an insight on the molecular function of HHD2 in the different isoforms of whirlin, we tried to identify if the domain is interacting with this preceding PDZ tandem. We showed that in the three-domain construct, HHD2 is independent from the PDZ tandem and is separated by a 45-residue flexible linker. It is possible that the molecular interactions of HHD2 are more correlated with the function of the C-terminal isoform of whirlin. The short C-terminal isoform, which was almost exclusively studied through its PDZ3, is involved in a multi-protein complex by interacting with Myosin15 and Eps8 [227,237]. The C-terminal isoform interacts directly with more than one partners *in vitro* and in hair cells. It is unlikely that all interactions are modulated by PDZ3,

thus the HHD2 is a good candidate for being involved in the actin-related complex of whirlin, despite being separated from the PDZ3 by a large proline rich region.

8.2.4 Materiel and methods

8.2.4.1 Protein & sample production

All construct of whirlin are based on murine isoform 4 (Q80VW5-4): HHD2 (residues 420-499), PDZ2-HHD2 (276-499) PDZ1-2 (137-385), PDZ1-PDZ2-HHD2 (137-499). All are cloned into a pGST//2 vector (derived from pGEX-4T-1; Amersham). A TEV cleavage site is introduced between the N-terminal GST tag and the protein sequence. All mutants are generated using QuickChange-2XL (Agilent) kit and protocol. Plasmids verified for DNA sequence (Eurofins, Paris) are transfected to *E.coli* BL21 (DE3) star (Invitrogen). Bacteria are expressed in LB medium or M9 minimal medium supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose. Selenomethionine labelled proteins were first grown in in M9 complemented with L-methionine until $D.O_{600nm}$ 1.0, and then overexpressed in M9 complemented in L-seleno-methionine after a starvation phase. At DO_{600nm} of 0.8, protein expression is induced by the addition of 1 mM IPTG for 3 h at 30 °C. Bacteria are harvested by centrifugation, suspended in 50 mM Tris-HCl, 250 mM NaCl, 5 mM 2-mercaptoethanol, pH 7.5, buffer containing protease inhibitors (Complete, Roche Diagnostics) and disrupted by sonication. The extract is centrifuged at 15000 RCF for 1 h. Tagged proteins are purified by affinity chromatography using GSTrap (GE Healthcare) followed by TEV cleavage overnight at 4 °C and a last step of size-exclusion chromatography (SEC) using a Sephacryl S-100 HP 16/60 (GE Healthcare). All proteins are purified in 50mM Tris-HCl pH 7.5, NaCl 150mM.

8.2.4.2 Small Angle X-ray Scattering

X-ray scattering data are collected at the SWING beamline at Soleil (Saclay, France). 50 μl of sample is injected on a Superdex-200 5/150GL (GE) size exclusion column in-line with the SAXS measuring cell. Frames with a duration of 1 s are recorded during the whole elution time. Approximately 100 frames recorded in the dead volume of the column are averaged to generate the buffer subtraction curve and the 10 to 30 best frames corresponding to the top of the elution peak are average to generate the result curve. Primary data reduction was performed using FOXTROT. The one-dimensional scattering intensities were expressed as a function of the modulus of the scattering vector $Q = (4\pi / \lambda) \sin\theta$ with 2θ being the

scattering angle and λ the X-ray wavelength. Buffer intensities were subsequently subtracted from the respective sample intensities using software package PRIMUS [257]. The radii of gyration were evaluated using Guinier approximation [258]. The Dmax was determined from distance distribution function P(r) obtained with the program GNOM [259].

8.2.4.3 Crystallography

Initial screening of crystallization conditions was carried out for HHD2 by the vapour diffusion method with a Mosquito™ nanolitre-dispensing system (TTP Labtech). Sitting drops were set up using 400 nl of a 1:1 mixture of HHD2 and crystallization solutions (672 different commercially available conditions) equilibrated against 150 μ l reservoir in multiwell plates (Greiner Bio-One). The crystallization plates were stored at 18°C in a RockImager1000™ (Formulatrix) automated imaging system to monitor crystal growth. Two families of crystals were obtained. The 5-helix monomer conformation of the selenometionine labelled HHD2 was crystalized at 4°C in 100 mM Cacodylate and 1.26M ammonium sulfate. The swapped-dimer crystals conformation of HHD2 was obtained in 20% propan-2-ol and 10% PEG 4k. Single crystals were flash-cooled in liquid nitrogen using a mixture of 50% Paratone-N and 50% paraffin oil as cryoprotectant. X-ray diffraction data were collected on beamline PROXIMA-1 at Synchrotron SOLEIL (Saint-Aubin, France). The diffraction images were integrated with the program XDS and crystallographic calculations were carried out with programs from the CCP4 program suite. Five Se sites were located with SHELXCD and refined with autoSHARP. The initial model was built using Arp/wARP (19).

The structures were refined by alternate cycles of restrained maximum-likelihood refinement with the program Refmac5 and manual adjustments were made to the models with Coot.

8.2.4.4 NMR

All experiments were performed on a 900 MHz Bruker Avance III equipped with a cryogenically cooled triple resonance 1H{13C/15N} TCI probe, on a Varian NMR System 600 MHz equipped with a cryogenically cooled triple resonance 1H{13C/15N} PFG probe, and on a Varian NMR System 500 MHz equipped with a triple resonance probe 1H{13C/15N}. Data process and analysis were performed using NMRpipe [260], ccpNmr Analysis 2.4 [261]. Analysis of the secondary structure elements was achieved using proton and carbon chemical shifts with the TALOSN prediction server [262]. The sequence specific 1HN,

1Ha, 15N, 13Ca, 13Cb and 13CO resonance assignments was performed using the following experiments: 2D 1H–15N HSQC and 3D HNCO, HN(CA)CO, CBCA(CO)NH, HNCACB and HNHA. Chemical shift perturbations (CSP) were calculated as the weighted average (^{1}H, ^{15}N) chemical shift differences $\Delta\delta_{avg}$ between the wild-type and mutants PDZ1-2 as follow: $\Delta\delta_{avg} = ((\Delta\delta H)^2+(\Delta\delta N\times 0.159)^2)^{1/2}$. ^{1}H-^{15}N residual dipolar couplings were obtained by recording IPAP-HSQC experiments [263] on ^{15}N-labeled samples of HHD2 at concentrations ranging from 150 to 600 μM in the absence and in the presence of two different oriented media containing either 10 g/l Pf1-phages (ASLA biotech). The T1 and T2 relaxation and heteronuclear NOE experiments were recorded in an interleaved manner with a recycling time of 3 s and with 7 relaxation delays for T1 (20, 100, 200, 400, 600, 800, 900, ms) and ten for T2 (8.5, 17, 34, 51, 68, 85, 102, 119, 153, 187, ms). The relaxation data were analyzed using the software TENSOR [264], using the model-free formalism of Lipari and Szabo [265].

8.3 Commentaire et expérience supplémentaires

Dans cet article, nous avons présenté la structure du HHD2 de la Whirline dans deux conformations : une forme globulaire à cinq hélices ainsi qu'une forme dimérique induite par échange de structure secondaire (swapped-dimère), où chaque protomère prend une conformation plane, comportant trois hélices. La différence de conformation est induite par la structuration des deux boucles α1/α2 et α3/α4 en hélices, fusionnant ainsi les hélices qu'elles connectent. Dans le swapped-dimère, les contacts entre hélices α2/α3 et α4/α5 sont conservés dans un protomère tandis que les contacts entre ces deux couples d'hélices sont eux satisfaits par la conformation en dimère. L'hélice α1 est en partie déstructurée.

Nous avons ensuite montré que la conformation observée en solution est la forme monomérique à 5-hélices, en attribuant les résonances des atomes du squelette de la protéine. Les expériences de RDC et des expériences complémentaires de SAXS (Figure 52) sont deux techniques moyennant la contribution de toutes les conformations présentes en solution. Elles indiquent sans ambiguïté que la forme à 5-hélices est la seule présente en solution et que l'hypothèse d'une faible population de swapped-dimère est écartée jusqu'à présent.

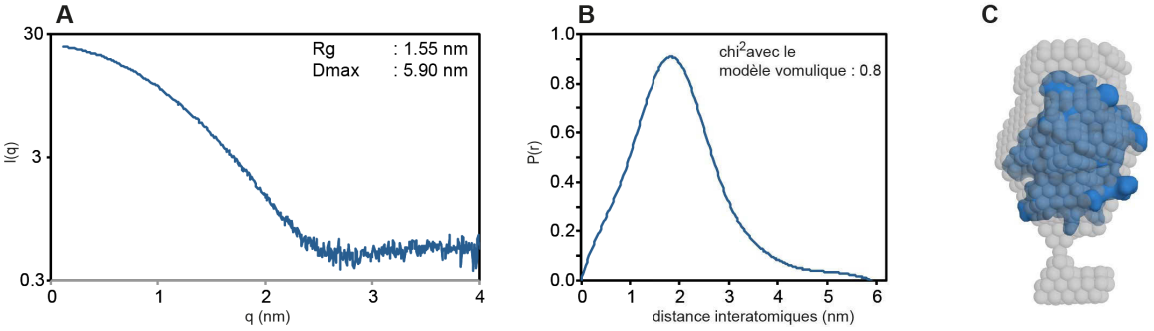


Figure 52 Diffusion aux petits angles de HHD2 en solution. (A) Courbe de diffusion des rayons X aux petits angles de HHD2 en solution. Le Rg et Dmax sont indiqués dans l'encadré. **(B)** Distribution de distances interatomiques déduites de la courbe SAXS de HHD2. Le chi^2 du volume présenté en C, avec ces données, est indiqué sur le graphique. **(C)** Volume calculé *ab initio* depuis la P(r) de HHD2 en solution, par le serveur Dammin (embl-hamburg.de/biosaxs/atsas-online/dammin).

Enfin, nous nous avons étudié la conformation et les interactions potentielles du domaine dans le contexte plus large de la protéine, pour obtenir des indications sur le rôle du domaine dans la Whirline. Nous avons montré par SAXS et RMN que la conformation du HHD2 n'est pas affectée par le tandem de PDZ le précédant. Dans une construction contenant les trois domaines, le HHD2 reste indépendant et séparé du tandem par une région flexible. Le N-terminal de la protéine forme un supramodule fonctionnel (**section 9**), l'absence d'interaction du HHD2 avec cette région pourrait indiquer que sa fonction est plutôt associée

à la partie C-terminale de la protéine, formant un complexe avec la Myosine-15a et Eps8. Nous nous intéresserons l'interaction avec Eps8 dans la section 8.3.2.

8.3.1 Changement de conformation

Après avoir obtenu les deux conformations de HHD2, nous avons tenté d'induire, en solution, la transition de conformation du monomère vers le swapped-dimère. Il est généralement admis que les protéines Usher dimérisent ou oligomérisent pour effectuer leur fonction dans les stéréocils [68,183]. Les complexes protéiques impliqués dans les stéréocils effectuent leurs fonctions sans modification post-traductionnelle ; il existe donc des "interrupteurs" conformationnels permettant la transmission de l'information au sein de la cellule. Le domaine HHD2 de la Whirline nous semblait un bon candidat pour endosser cette fonction "interrupteur", notamment dans l'isoforme courte de la Whirline. Nous avons fait varier le solvant contenant HHD2 pour obtenir des conditions favorables à la forme dimérique, sans succès jusqu'à présent. Voici une liste exhaustive des conditions testées :

- Encombrement : favorable aux conformations exposant une surface minimum, il pourrait favoriser la formation de swapped-dimère [266]. La Whirline étant exprimée dans un environnement confiné et très encombré, il était important de tester ces conditions, en simulant l'encombrement cellulaire. Nous l'avons testé en ajoutant 200 g/l de ficoll-70 dans la solution.
- Concentration en sel : lors de la déflection des stéréocils, le messager principalement de la mécano-transduction est l'entrée de cations dans la cellules. Nous avons donc testé des concentrations en NaCl variant de 150 mM à 500 mM.
- pH : dans la vision des drosophiles, un supramodule de domaine PDZ peut être désassemblé, libérant la fonction de l'un des domaines, lors d'un changement de pH lié à la transduction du signal [50,152]. Nous avons enregistré les spectres du domaine entre pH 8 et pH 3.5.
- Température : les boucles charnières $\alpha 1/\alpha 2$ et $\alpha 3/\alpha 4$ sont plus dynamiques que le reste de la protéine d'après notre étude RMN. Nous avons donc tenté d'intensifier cette dynamique, en augmentant la température, créant un état favorable au dimère. Nous avons enregistré les spectres du domaine de 10 à 50°C.
- Alcool : la conformation en swapped-dimère en cristal a été capturée en présence de propan-2-ol (20%) ou d'éthanol (5%). Il est connu que alcool favorise la conformation en

hélice des protéines [255]. Nous avons donc titré l'effet du propanol sur le domaine en solution.

- Oxydation : la stimulation trop longue des cellules ciliées induit l'accumulation d'espèces oxydantes en solution. Le HHD2 de la Whirline contient quatre méthionines oxydables dans son cœur hydrophobe. Nous avons oxydé la construction en présence de H_2O_2 afin de déstabiliser son cœur hydrophobe et induire la forme swapped-dimère.
- Acétylation : le HHD2 de la Whirline ne contient qu'une seule lysine (K473) dans $\alpha 4$. Dans la conformation à 5-hélice, celle-ci forme deux liaisons hydrogène avec E439 de l'hélice $\alpha 4$ et N469 de la boucle $\alpha 3/\alpha 4$. Ces liaisons ne sont pas conservées dans la forme dimérique (Figure 53), et sont localisées autour des régions charnières. Nous avons ajouté de l'anhydride acétique afin d'acétyler cette lysine, supprimer les liaisons hydrogène et promouvoir la forme dimérique.

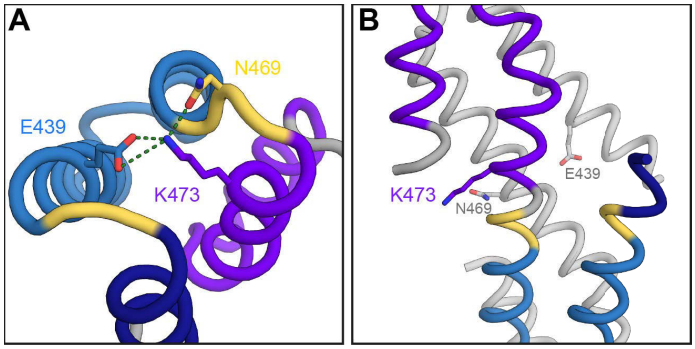


Figure 53 : K473 du HHD2. (A) Représentation de la K473 dans la conformation à 5-hélices, avec son réseau de liaisons hydrogène **(B)** Représentation de la K473 dans la conformation à 3-hélices.

Une dernière condition qu'il serait intéressant de tester est l'effet de la tension mécanique sur ce domaine. Comme nous l'avons présenté dans l'introduction, la mécano-transduction nécessite la présence de ressort moléculaire, encore non identifié. La flexibilité des régions charnières du HHD2 et la conformation alternative du domaine pourraient indiquer un changement conformationnel possible sous tension. Il serait intéressant de tester, expérimentalement ou par modélisation, l'impact sur sa conformation de la tension appliquée aux extrémités du domaine HHD2.

8.3.2 Tests d'interaction avec Eps8

Nous avons aussi étudié l'interaction possible de HHD2 de la Whirline avec Eps8. Les deux protéines interagissent et colocalisent en cellule. Ce complexe se forme entre le domaine

HHD2 ou la région riche en prolines de la Whirline et une grande région sans domaine identifié de Eps8 [211,226]. En règle générale, les domaines HHD interagissent avec des hélices amphiphiles isolées de leurs partenaires. Nous avons recherché dans cette région d'Eps8 des motifs potentiels d'interaction aux domaines. Nous avons identifié trois hélices d'une vingtaine de résidus chacune, que nous avons commandées par synthèse organique. Les premiers tests réalisés par RMN avec ces peptides n'ont pas montré d'interaction avec le domaine HHD. Nous sommes maintenant en train de délimiter des constructions plus grandes d'Eps8 afin d'identifier sa région d'interaction avec la Whirline, via le HHD2 et/ou avec la région riche en proline.

9 Résultat 3 : *Structural characterization of whirlin reveals an unexpected and dynamic supramodule conformation of its PDZ tandem*

9.1 Résumé :

Les protéines multidomaines composent la grande majorité du protéome humain. Un des avantages de ces organisations modulaires est leur plasticité de recombinaison fonctionnelle au cours de l'évolution. La succession de modules simples le long de la chaine polypeptidique permet l'émergence de nouvelles fonctions et de nouveaux modes de régulation par la formation de supramodule.

La Whirline est une protéine multidomaine impliquée dans les échafaudages moléculaires des structures mécanosensibles des neurones de l'audition. Dans les articles présentés dans cette partie, nous avons montré que le tandem de domaines PDZ, inclus dans l'isoforme longue de la Whirline, forme un supramodule impliquant des extensions structurées des deux domaines PDZ. La conformation en solution de ce tandem est en équilibre entre une forme compacte, dont nous avons déterminé la structure, et un ensemble de conformations plus ouvertes.

Pour ce faire, nous avons mis en place une approche de détermination de structure intégrant plusieurs techniques biophysiques de résolutions variées, moyennant les conformations sur différentes gammes de temps. Ce travail repose essentiellement sur des études par RMN ; l'étape initiale a été l'attribution des résonances des atomes du squelette peptidique de la construction de 27kDa contenant les deux domaines PDZ et leurs extensions, complétée par la mesure des temps de relaxations transversale et longitudinale (**partie 9.2**).

L'estimation du temps de corrélation de la protéine en solution ainsi que des études complémentaires de SAXS et d'émission de fluorescence ont confirmé la nature supramodulaire de ce tandem de domaines PDZ ainsi que l'implication des extensions dans la conformation de l'ensemble. Nous avons résolu la structure de la forme compacte de ce supramodule en utilisant des contraintes de distances obtenues en utilisant des sondes paramagnétiques et des contraintes d'orientation relative des domaines obtenues par l'introduction de milieux orientant dans le tube RMN. Nous avons confirmé cette structure par RMN et spectrométrie de masse en utilisant des mutants ponctuels à l'interface entre

domaines. Une analyse d'ensemble basée sur les données de SAXS et certaines données RMN nous a permis de déterminer les conformations explorées par le tandem en solution ; celui-ci est en équilibre entre la forme compacte déterminée précédemment et un ensemble de formes plus ouvertes **(partie 9.3)**. Nous avons enfin caractérisé l'impact de ce supramodule sur l'affinité de la Whirline pour ses partenaires.

9.2 Article : *^{1}H, ^{13}C and ^{15}N backbone resonance assignments and dynamic properties of the PDZ tandem of Whirlin*. Biomol. NMR Assign. (2016) 10 : 361-365

Florent DELHOMMEL[a,b,c], Nicolas WOLFF[a,b], Florence CORDIER[a,b] *

[a] Institut Pasteur, Unité de Résonance Magnétique Nucléaire des Biomolécules, Département de Biologie Structurale et Chimie, F-75015 Paris, France

[b] CNRS, UMR3528, F-75015 Paris, France

[c] Univ. Pierre et Marie Curie, Cellule Pasteur UPMC, rue du Dr Roux, 75015 Paris, France

9.2.1 Abstract

Mammals perceive sounds thanks to mechanosensory hair cells located in the inner ear. The stereocilia of these cells are tightly bound together in bundles by a network of cadherins and scaffolding proteins. Stereocilia deflection induces stretching of this network and is responsible for hair cell depolarization that triggers the neuronal message, transducing the mechanical signal into an electric signal transmissible to the brain. Nearly all proteins involved in this mechano-electrical transduction network contain short C-terminal motifs of interaction with PDZ domains (PSD-95, Discs Large, ZO-1). Interestingly only two of these proteins encompass PDZ domains: Harmonin and Whirlin. As our first step towards a comprehensive structural study of Whirlin, we have assigned the ^{1}H, ^{13}C and ^{15}N backbone resonances of a tandem formed by the first two PDZ domains of Whirlin, reported the secondary structure elements of this tandem as predicted by the TALOS+ server and evaluated its dynamics from ^{15}N relaxation measurements.

Keywords: Whirlin, backbone resonance assignments, PDZ, PDZ tandem, mechano-electrical transduction

9.2.2 **Biological context**

Mammals detect sound thanks to a sensory organ located in the inner ear, the cochlea. This organ made of three coiled ducts is capable of deconvoluating a complex sound to its pure components. Hair cells are located all along the cochlea; they possess mechanosensitive bundles of stereocilia that are deflected toward the tallest by the vibrations in the excited areas. The stereocilia are organised in rows of increasing length in which every stereocilium is connected at its apex to a stereocilium of a taller row by a protein link. This tip-link is anchored on both sides to large protein densities below the membrane [267]. When the hair bundle moves, the tension on this protein network increases, leading to the mechanical opening of a cation channel, the resulting influx of ion induces membrane depolarization and ultimately produces the neuronal message [268].

Remarkably, among all the identified proteins involved in this mechano-electrical transduction network, all except one contain short C-terminal motifs of interaction with PDZ domains (PBM), highlighting the significance of such a module in hearing mechanisms. Only two proteins located at the apex of the stereocilia encompass PDZ domains in several copies, Harmonin and Whirlin [166,216]. We are interested in the later, and more specifically in the tandem formed by its two N-terminal PDZ domains, PDZ1-PDZ2.

PDZ form a very abundant class of protein-protein interaction domains. In the human proteome, they are often found in multiple copies in large multi-domain proteins, among other scaffolding domains. They consist of 80 to 90 residues folded into 5 to 6 β-strands and 2 α-helices. PDZ mainly bind to specific motifs encompassing the four C-terminal residues of their partners. These short linear motifs insert into a groove of the PDZ located between its second α-helix and second β-strand to form an intermolecular anti-parallel β-sheet. A given PDZ usually has numerous potential partners with constants of dissociation in the micromolar range at best, explaining the broad plasticity of networks maintained by PDZ containing proteins. Even though the PDZ fold is highly conserved, the prevalence of structured extensions at the extremities of PDZ domain is high [124]. These additional structures are accountable for various effects. Some directly affect the binding, by increasing the size of the pocket of their proximal PDZ domain [92]. Others have been reported to tune the local dynamics, entropically improving the binding [133]. Some PDZ are able to form transient multi-domain assemblies [60,71]. It has also been widely reported that PDZ domain are able to form homo and hetero dimers [115]. Finally, in large multi-domain proteins, adjacent PDZ domains have been described to interact and form PDZ tandem [148,154].

We are presenting here the backbone resonance assignments of the PDZ tandem formed by the first and second PDZ domain of Whirlin PDZ1-PDZ2, as well as the determination of secondary structure elements and backbone dynamics inferred from ^{15}N relaxation measurements. This ^{1}H, ^{13}C and ^{15}N backbone assignment is the necessary and sufficient step toward a comprehensive structural study of Whirlin, such as the determination of a structural model of the tandem based on NMR data (i.e. ^{1}H, ^{15}N chemical shift mapping, residual dipolar coupling and paramagnetic relaxation enhancement). It will be further necessary for interaction studies (based on chemical shift mapping) with its adjacent HHD1 and HHD2 domains.

9.2.3 Methods and experiments

Sample preparation. The tandem of the first two PDZ domains of murine Whirlin isoform 4 (Identifier: Q80VW5-4) is cloned as a N-terminal GST-tagged protein. The construct covers the sequence from residue G137 to S384 and includes five N-terminal exogenous residues (G, A, M, G, S) from the expression vector for a total of 253 residues. The coding sequence of the construct is inserted into a pGST//2 vector (derived from pGEX-4T-1; Amersham) and then transfected to *Escherichia coli* BL21 (DE3) star (*Invitrogen*). For ^{15}N relaxation measurements, a sample of uniformly ^{15}N labelled Whirlin PDZ1-PDZ2 was produced in bacteria grown in isotopically labelled M9 minimal medium containing 1.0 g/l $^{15}NH_4Cl$ as sole nitrogen source. For the assignment experiments, a sample of U-^{15}N/^{13}C/^{2}H labelled PDZ1-PDZ2 was produced in a similar medium with the addition of 4.0 g/l of ^{13}C-glucose as sole carbon source, and the substitution of H_2O by D_2O. Protein expression was induced when the bacteria solution reached OD_{600nm} 0.8 by addition of 1 mM IPTG and incubation at 30°C for 3 hours to a final OD_{600nm} of 1.3. Harvested cells are suspended with buffer A (50 mM Tris-HCl, 250 mM NaCl, pH 7.5, 5 mM β-mercaptoethanol and 1 tablet of protease inhibitor cocktail cOmplete™ EDTA-free (Roche) for 80 ml of buffer) supplemented with 0.1 g/l of chicken egg lysozyme (Sigma-Aldrich). The solution is sonicated to disrupt the bacteria and benzonase® nuclease (Sigma-Aldrich) is added to degrade nucleic acids. Finally the lysate is clarified by centrifugation at 17000g for 1h. Clarified cell lysate is loaded on a GST column (GSTrap HP 5ml, *GE*) equilibrated with buffer A. The GST-tag is cleaved by TEV protease (1% M/M) directly injected on the column and incubated overnight at 4°C. The eluted Whirlin PDZ1-PDZ2 is loaded onto a size exclusion column (Sephacryl S-100 HP

16/60, *GE*) equilibrated with buffer B (50 mM Tris-HCl, 150 mM NaCl, pH7.5, 1 cOmplete™ (Roche) tablet for 150 ml of buffer). Size, purity and sequence of Whirlin PDZ1-PDZ2 samples were checked by SDS-Page and mass spectrometry. ^{15}N- and $^{15}N/^{13}C/^{2}H$-labelled samples of Whirlin PDZ1-PDZ2 were concentrated to 175 µM and 600 µM respectively, as estimated from its absorbance at 280 nm assuming a calculated ε_{280} of 19940 $M^{-1}.cm^{-1}$. The concentrated protein solution is complemented to 12% of D_2O for a final volume of 240 µl. For the $^{15}N/^{13}C/^{2}H$ sample used for the 3D NMR experiments, the pH was set to 6.0.

NMR spectroscopy. NMR experiments were recorded at 25°C on a 900MHz Bruker Avance III equipped with a cryogenically cooled triple resonance $^{1}H\{^{13}C/^{15}N\}$ TCI probe and on a Varian NMR System 600MHz equipped with a cryogenically cooled triple resonance $^{1}H\{^{13}C/^{15}N\}$ PFG probe. The sequence specific $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and ^{13}CO resonance assignments was performed using TROSY-based versions of the following experiments: 2D ^{1}H- ^{15}N HSQC and 3D HNCO, HN(CA)CO, CBCA(CO)NH and HNCACB. The ^{15}N relaxation times (T_1 and T_2) were measured at 25°C, at 600 MHz ^{1}H frequency by standard methods [269]. The T_1 and T_2 experiments were recorded using a TROSY version [270] in an interleaved manner with a recycling time of 3 s and with nine relaxation delays for T_1 (10, 50, 100, 300, 500, 700, 900, 1200, 1500 ms) and eleven for T_2 (5, 10, 15, 25, 30, 35, 45, 55, 65, 75, 90 ms). The spectra were processed with NMRPipe [260] and subsequently analysed in CcpNmr Analysis software [261]. Analysis of the secondary structure elements was achieved using proton and carbon chemical shifts with the TALOS+ prediction server [271].

9.2.4 Extent of assignments, secondary structure elements and dynamics

Over the whole endogenous sequence backbone atoms (239 non-Proline residues), we assigned 88% of the HN resonances, 90% of the Cα resonances, 88% of the Cβ resonances and 91% of the CO resonances. Our construct contains large unfolded regions and transiently folded domain extensions that are challenging to assign, namely the large linker between the two PDZ domains, from A222 to G276 and the C-terminal segment after the PDZ2 from V360 to S384. In PDZ1 and PDZ2 folded domains, we extended the assignments to 95% (157/165) for the HN resonances, 98% and 96% for the CA and CB resonances, respectively, and 97% for the CO resonances. In PDZ1, the resonances of the residue D194 and K213 are missing, while in PDZ2, the resonances of T294, G297, Y301, Y306, T308 and I327 are missing. The amide proton resonances of the 57-residue linker sequence that tethered the two

domains were assigned at 86% with the resonances of the residues H253, H255, S257 and R267 unassigned. The HN resonances of the C-terminal segment, from residues 363 to 384 remains largely unassigned with 13 missing out of the 20 expected, only the HN resonances of the residues A366, R367, T368 T369, V370, D371, T373, K374 and G382 could be assigned. The TALOS+ prediction delineates 5 β-strands and 2 α-helices for each domain, which is consistent with the canonical PDZ fold.

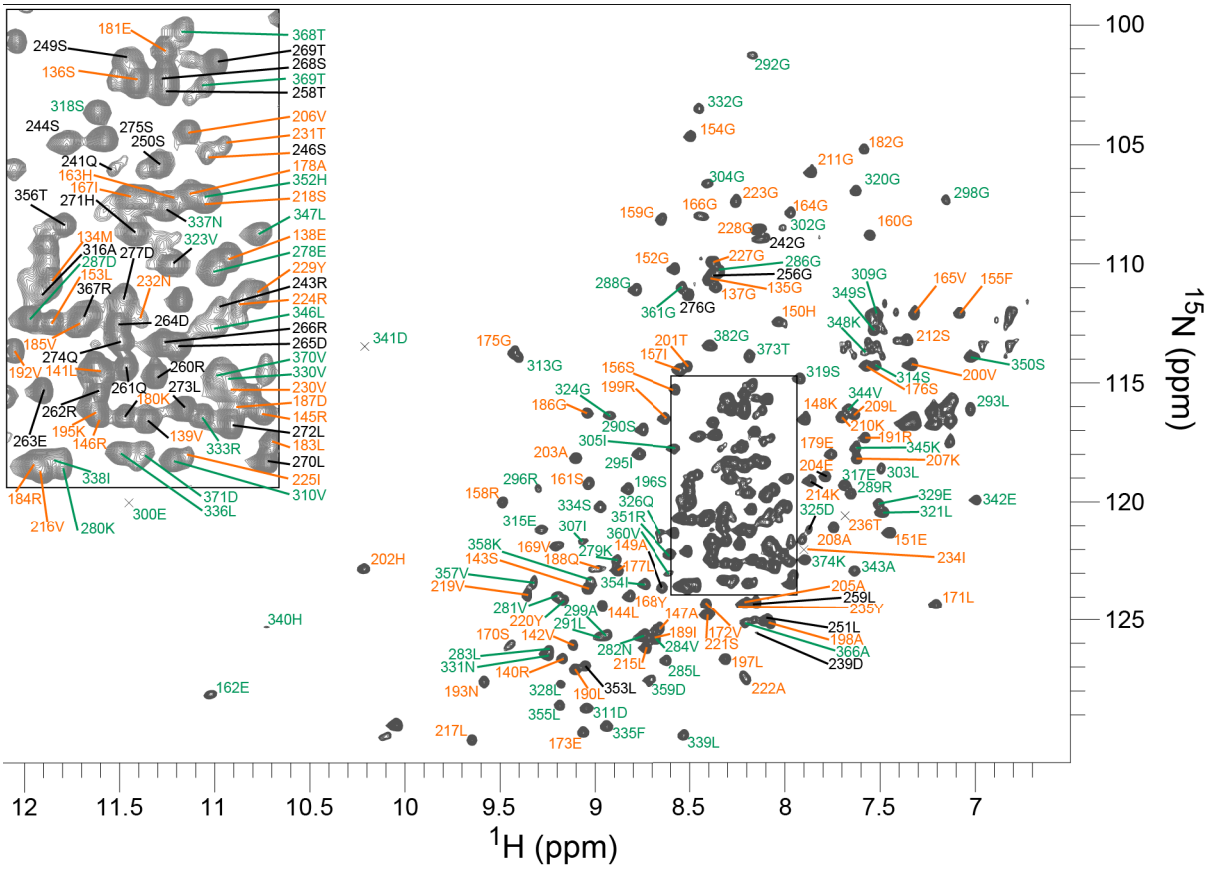


Figure 1 : ^{1}H-^{15}N HSQC spectrum of 0.6 mM ^{15}N/^{13}C/^{2}H-labelled Whirlin PDZ1-PDZ2 (pH 6.0, 25°C, 600 MHz)

The ^{15}N longitudinal relaxation time (T1) and transverse relaxation time (T2) were evaluated. Isotropic rotational correlation time (τc), estimated from the average T1/T2 ratio of non-flexible residues (16.4 ± 2.2), is 12.6 ± 0.1 ns. This value is in good agreement with the theoretical value of ~ 11.1 ns expected for a globular domain with a molecular weight of 27 kDa (23.1 Å stokes radius), suggesting that the two PDZ domains are not independently tumbling. The small difference between the experimental and the expected τc could be the result of an anisotropic shape of the PDZ1-PDZ2 tandem. Overall, the ^{15}N relaxation measurements denote the formation of two rather rigid PDZ domains connected by a flexible linker (residues A222-G276, with shorter T1, longer T2). Within each PDZ domain, increased flexibility on a fast timescale (picosecond-nanosecond) is observed in the loops connecting

β1-β2 (GLGF loop), β4-α2 and α2-β5. In addition, conformational exchange on the microsecond to millisecond time-scale can be detected by shorter T2 values in the loops connecting β2-β3, α1-β4 and β4-α2.

The chemical shifts have been deposited in the BioMagResBank under accession number 26788.

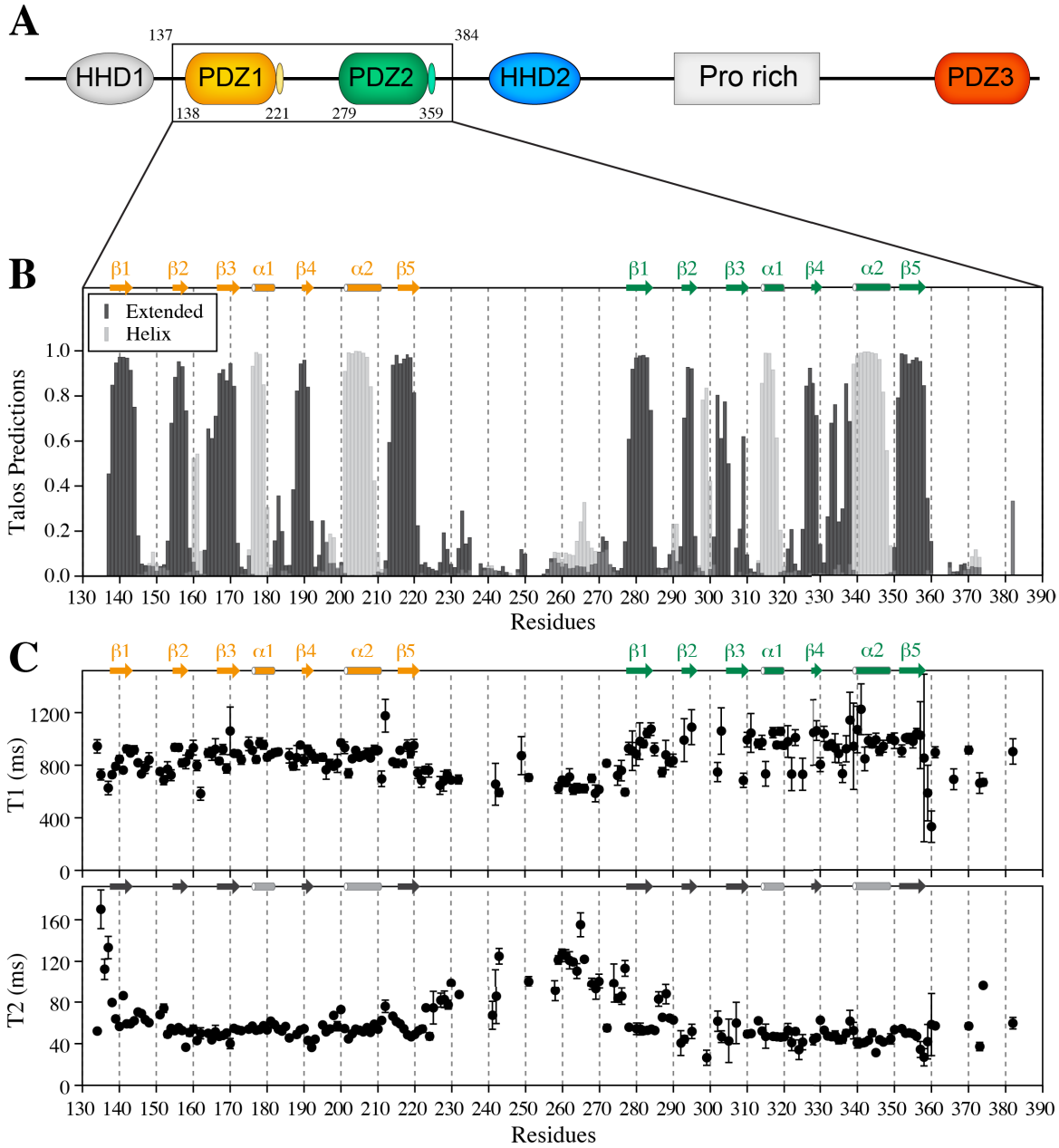


Figure 2 (A) Topology of murine Whirlin isoform 4 (Q80VW5-4). The black box represents the limitations of the construct PDZ1-PDZ2. **(B)** Values of TALOS+ predictions for Whirlin PDZ1-PDZ2 based on the backbone resonance assignments. Secondary structure elements inferred from these predictions are indicated at the top. **(C)** Backbone dynamics of Whirlin PDZ1-PDZ2 from ^{15}N relaxation data. ^{15}N longitudinal (T1) and transversal (T2) relaxation times are plotted in the top and bottom panel, respectively. Secondary structure elements are indicated at the top

9.2.5 Acknowledgments

We thank F.X. Cantrelle (UCCS-UGSF, Lille) and C. Simenel (NMR unit, Institut Pasteur) for their NMR technical expertise and helpful discussions. This work was supported by the Programme Transversal de Recherche from the Institut Pasteur (PTR grant no. 483) and the Ministère de l'Enseignement Supérieur et de la Recherche (grant no. 883/2013 to F.D.). Financial support from the TGIR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged.

9.3 Article : *Structural characterization of whirlin reveals an unexpected and dynamic supramodule conformation of its PDZ tandem*

Florent Delhommel[1,2,3], Florence Cordier[2,3], Benjamin Bardiaux[2,4], Guillaume Bouvier[2,4], Baptiste Colcombet-Cazenave[2,3], Sébastien Brier[2,5], Bertrand Raynal[2,6], Sylvie Nouaille[1,7,8], Amel Bahloul[1,7,8], Julia Chamot-Rooke[2,5], Michael Nilges[2,4], Christine Petit[1,7,8,9], Nicolas Wolff[2,3]

1. Sorbonne Universités, UPMC Univ Paris 06, Complexité du Vivant, F-75005 Paris, France
2. CNRS, UMR 3528, F-75015 Paris, France
3. Unité de Résonance Magnétique Nucléaire des Biomolécules, Institut Pasteur, F-75015 Paris, France
4. Unité de Bio-Informatique Structurale, Institut Pasteur, F-75015 Paris, France
5. Unité de Spectrométrie de Masse Structurale et Protéomique, Institut Pasteur, F-75015 Paris, France
6. Plateforme de Biophysique Moléculaire, Institut Pasteur, F-75015 Paris, France
7. Unité de Génétique et physiologie de l'audition, Institut Pasteur, F-75015 Paris, France
8. Unité Mixte de Recherche, UMRS 1120, Institut National de la Santé et de la Recherche Médicale (INSERM), F-75015 Paris, France
9. Collège de France, F-75005 Paris, France

9.3.1 Abstract

Hearing relies on the transduction of sound-evoked vibrations into electric signals, occuring in the stereocilia bundle of hair cells. The bundle is organized in a staircase pattern formed by rows of packed stereocilia. This architecture is pivotal to the transduction and involves a network of scaffolding proteins with hitherto uncharacterized features. Key interactions in this network are mediated by PDZ domains. Here we describe the architecture of the first two PDZ domains of whirlin, a protein involved in these assemblies and associated to congenital deaf-blindness. C-terminal hairpin extensions of the PDZ domains mediate the transient supramodular assembly which improves the binding capacity of the first domain. We determined a detailed structural model of the closed conformation of the PDZ tandem, and characterized its equilibrium with an ensemble of open conformations. The structural and dynamic behavior of this PDZ tandem provides key insights into regulatory mechanisms involved in the hearing machinery.

9.3.2 Introduction

Hearing relies on the capacity of highly specialized sensory hair cells to transduce sound-induced vibrations into electric potentials that are transmissible to the brain [253]. Hair cells possess an ensemble of actin-filled stereocilia tightly structured into staircase-shaped bundles deflected by sound-waves. Each stereocilium is connected to its taller neighbor by a single link of protein transferring sound-evoked tension to mechanically gated ion channels. Multiple extracellular links interconnect stereocilia allowing their concerted deflections [34] and thus simultaneously opening all mechanotransduction channels of the hair bundle. The cytosoplasmic regions of these connecting proteins are involved in large complexes anchoring them to the cytoskeleton [64,166]. Complexes of similar composition are also found in the photoreceptors cells [45]. The genes encoding these proteins are affected by mutations responsible for hereditary sensory diseases, notably the Usher syndrome, that associates deafness and progressive blindness. Usher proteins contain numerous scaffolding domains and protein-protein interaction motifs necessary to the intricacy of the network [34,187]. With a single exception, all Usher proteins possess short C-terminal motifs that interact with PDZ domains (known as PDZ binding motifs or PBM). Remarkably, only three Usher proteins contain multiples PDZ domains: harmonin [184], whirlin [216] and PDZD7 [272]. Through their PDZ domains, as well as other scaffolding domains, these three paralogous proteins are qualified to assemble large protein complexes.

Whirlin is a large Usher protein expressed in three isoforms found in different cells and subcellular compartments [221]. The C-terminal short isoform contains the HHD2 (Harmonin Homology Domain) [63], a large proline-rich region, the PDZ3 and a C-terminal PBM (**Fig. 1A**). This isoform is located at the very tip of the tallest row of stereocilia where it associates to the actin cytoskeleton and maintains stereocilia length [211,221]. The N-terminal short isoform contains the N-terminal HHD1, a tandem of PDZ domains and the HHD2 (**Fig. 1A**). The N-terminal short isoform is solely expressed in photoreceptor cells [221], in which whirlin maintains the connecting cilium [45,256]. Finally, the long isoform, corresponding to the full-length protein, has an expression more complex and less characterized. During hair-cell development, this isoform is involved in the Usher2 protein complex at the basis of the stereocilia, allowing for the correct growth of the hair bundle [214,224,225]. In mature hair-cells, it has been reported at the tip of the tallest stereocilia row, alongside the short isoform [221] and also at half-height of the middle stereocilia rows, where it is suspected to mediate inter-stereocilia links involved in hair-bundle cohesion [273]. Mutations of whirlin are responsible for

two diseases, non-syndromic deafness (DFNB31) and Usher syndrome type 2 (USH2D). Mutations of the C-terminal isoform may result in deafness exclusively, while N-terminal mutations are accountable for deaf-blindness syndromes [232]. A particular in-frame deletion responsible for Usher syndrome leads to a whirlin long isoform in which the PDZ tandem is truncated, underlying the importance of this region in the function of whirlin in photoreceptor cells and in cochlear hair cells [274].

PDZ domains are a large and diverse family of protein interaction modules. They frequently occur in multiple copies in large proteins of diverse cellular functions, including cell polarity, synaptic transmission or regulation of cell junctions [74,275,276]. PDZ domains interact with short linear motifs at the C-terminus of their partners called PBM [76]. For numerous PDZ domains, the specificity toward their partners is ensured by interaction outside the classical binding groove [91,138,277] or by folded extensions beyond the canonical PDZ fold [124]. These extensions enhance the normal function of their associated domains in various ways. Extensions can widen the PDZ binding pocket by forming direct interactions with residues upstream to the canonical PBM, and thereby increase the selectivity of binding to their partners [92,132]. Extensions that entropically favor the PDZ/partner binding, without interacting directly with the ligand have also been reported [133]. In addition to modulation through small extensions, PDZ domains are also prompt to form homo- or hetero-oligomers with PDZ domains [115] or with other adjacent domains, increasing the diversity of functions of the whole assembly by altering binding capacities [71,159] or enzymatic activities of the associated domains [60]. The paralog of whirlin, harmonin, illustrates the increase of diversity induced by the addition of adjacent structures. The PDZ1 of harmonin and the upstream HHD are connected by a C-terminal hairpin extension, establishing a supramodule and causing a drastic increase of affinity of PDZ1 for its ligand [71].

In this study, we aim at describing the molecular organization of whirlin by defining the structural model of the homotypic supramodule formed by its first two PDZ domains (subsequently called PDZ1-2). We show that this highly dynamic assembly is in equilibrium between a well-defined closed conformation in which the PDZ domains are interacting *via* their respective C-terminal hairpin extensions and an ensemble of more open conformations. To study this dynamic complex, an integrated approach based on site-directed mutagenesis combined with NMR spectroscopy, Mass Spectrometry and Small Angle X-ray Scattering was used to characterize and validate the structure and the conformational exchange of the PDZ1-2 tandem. We complete our description of the structure by evaluating its significance to the binding capacities of whirlin. The structures of the PDZ1-2 tandem of whirlin highlight

the dynamic nature of scaffolding complexes traditionally described as rigid assemblies. The structure of PDZ1-2 will also help understanding the functions of whirlin in hair cells and photoreceptors cells, that are altered in Usher syndrome.

9.3.3 Results

9.3.3.1 The tandem PDZ1-2 of whirlin forms a supramodule

We previously determined the dynamic behaviour of the two N-terminal PDZ domains of whirlin by ^{15}N NMR relaxation measurements [278]. The inferred correlation time of PDZ1-2 (12.6 ns) is in good agreement with the theoretical value of ~11.1 ns expected for a 27 kDa globular protein and is significantly higher than the correlation time of 7.3 ns measured for isolated PDZ1-Hp1 and PDZ2-Hp2 (**Fig. 1C**). The relaxation data indicate that the two PDZ domains are tumbling together rather than independently [279] despite the long linker of 57 residues tethering the domains. The X-ray scattering curve of PDZ1-2 yields a radius of gyration (Rg) derived from the Guinier approximation at low q-value of 27.5 Å and a maximum distance (Dmax) estimated from the pair-distance distribution function of 110 Å (**Fig.2D**), consistent with the two PDZ interacting in a mainly compact and globular conformation.

In addition to the main structure, PDZ1-2 explores other less populated conformations as highlighted by two distinct sets of resonances in the μs - ms exchange regime and a slight difference of correlation times between PDZ domains of the tandem with 12.4 ns and 13.7 ns for PDZ1-Hp1 and PDZ2-Hp2, respectively (**Supplementary Fig. 1)**. The flexible nature of PDZ1-2 prevented its crystallization and is directly observed on the dimensionless Kratky representation of its scattering intensity that shows a wide bell-shape curved with a maximum at 2.4 instead of the √3 expected for a rigid sphere [280,281] (**Supplementary Fig. 2**).

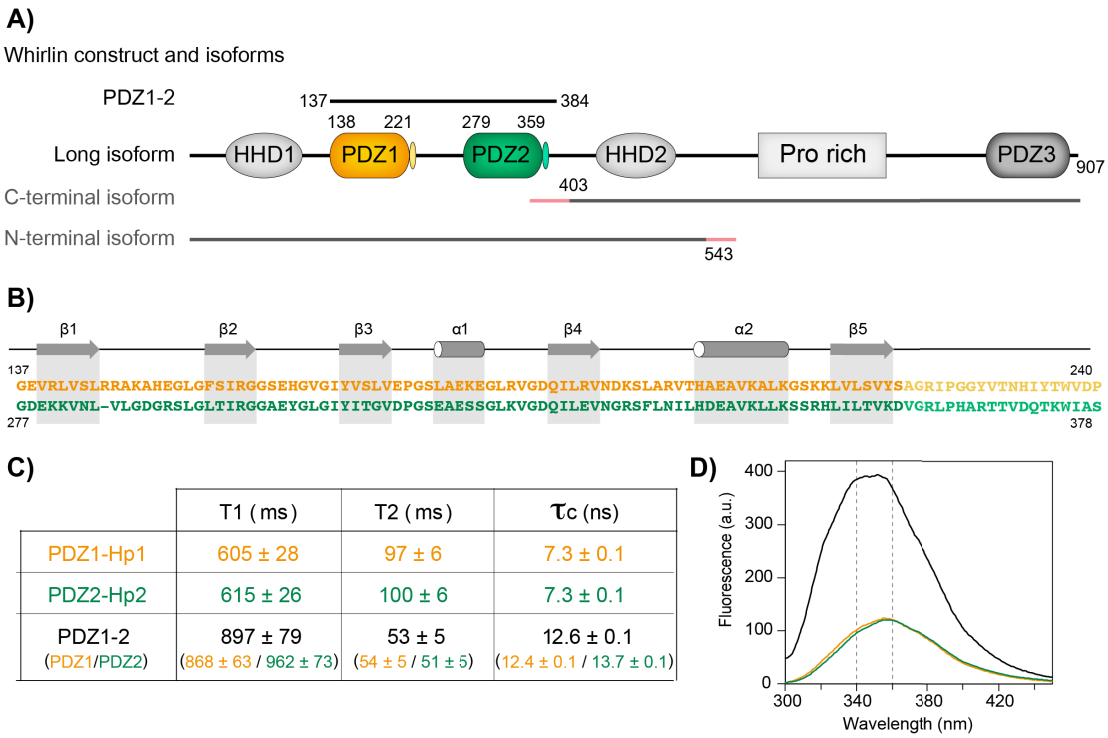


Figure 1 : Whirlin PDZ1-2 tandem dynamic in solution. (A) Domain organization of whirlin, from *Mus musculus* with the PDZ1-2^{WT} construct and the three main isoforms of whirlin. The red extremities of the short isoforms indicate sequence differences from alternative splicing **(B)** Alignment of whirlin PDZ1-Hp1 (orange) and PDZ2-Hp2 (green) with their secondary structures. **(C)** Averaged ^{15}N longitudinal and transverse relaxation times (25°C and 600MHz) and correlation times of PDZ1-Hp1, PDZ2-Hp2 and the tandem PDZ1-2. **(D)** Tryptophan emission fluorescence spectra of PDZ1-Hp1, PDZ2-Hp2 and PDZ1-2.

9.3.3.2 PDZ1 and PDZ2 possess hairpin extensions that are required for PDZ1-2 supramodule

The secondary structures identified from the chemical shift-based prediction of backbone torsion angles [262] match the expected five ß strands and two α helices of the two PDZ domains. In addition to their canonical fold, the 18 residues following the C-termini of both domains show predicted torsion angles favored in strands (**Supplementary Fig. 3**). The sequences of the two extensions are highly conserved among orthologs of whirlin (**Supplementary Fig. 3**) and are similar to the sequence of the known hairpin extension of harmonin PDZ1 (**Fig.2A**). In the predicted strands, the similarities with harmonin include a hydrophobic residue and a proline in the first strand and a tryptophan in the second strand (**Fig.1B**), these residues face the β3 strand and the α1-β4 loop of the PDZ1 in harmonin.

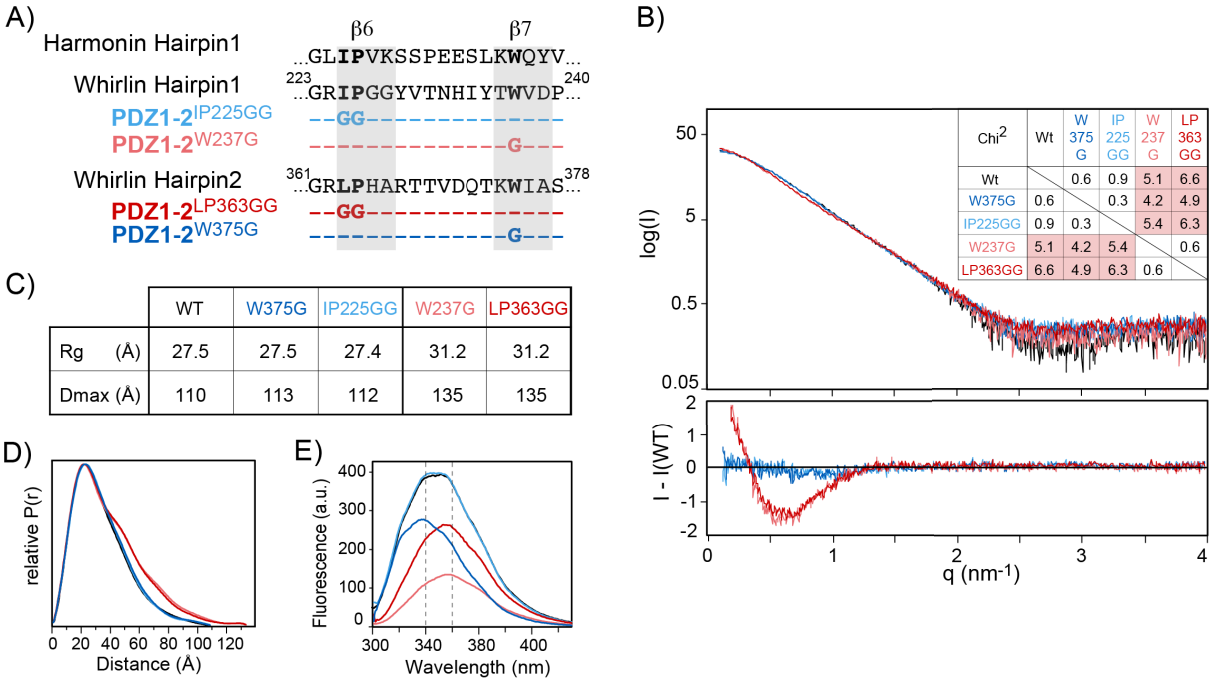


Figure 2 : Effect of mutation in the C-terminal hairpin extensions on the overall conformation of the PDZ1-2. (A) Substitutions performed on each PDZ hairpin (black for PDZ1-2^{WT}, light blue for PDZ1-2^{IP225GG}, light red for PDZ1-2^{W237G}, dark red for PDZ1-2^{LP363GG} and dark blue for PDZ1-2^{W375G}; the mutant color-coding is conserved throughout figures). The grey boxes delineate the expected strands from TalosN analysis and the homologous harmonin model (pdb: 3K1R). **(B)** Scattering curves of PDZ1-2 wild-type and PDZ1-2 mutants (top panel) and residual between the scattering curves of PDZ1-2 wild-type and PDZ1-2 mutants (bottom panel). The insert in the top panel is the similarity matrix showing the pairwise chi^2 between the SAXS profiles of each PDZ1-2 constructs (for q ranging from 0.119 to 4.007 nm^{-1}). **(C)** Radius of gyration (Rg) and Maximum distance (Dmax) values in Angstrom, extracted from the scattering curve of each PDZ1-2 constructs. It shows two families of constructs, one more compact (PDZ1-2^{WT}, PDZ1-2^{W375G} and PDZ1-2^{IP225GG}, highlighted in blue) and the other one more extended (PDZ1-2^{W237G} and PDZ1-2^{LP3635GG}, highlighted in red) **(D)** Distance distribution function extracted from SAXS data of the various constructs **(E)** Fluorescence emission spectra measured for PDZ1-2 wild-type and mutants, monitoring the local environment of the tryptophans located in the hairpins extensions of each PDZ domain.

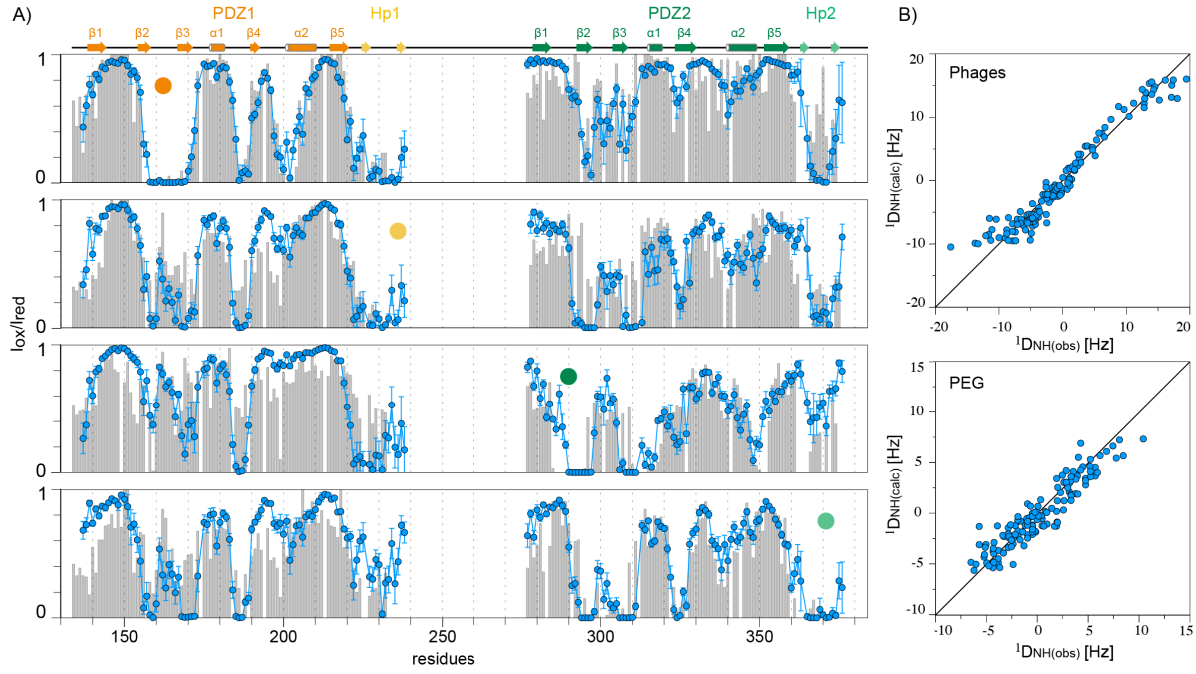
Whirlin PDZ1-2 possesses two tryptophans, one in each hairpin; their local environment can easily be probed by recording their emission of fluorescence. The emission spectra of isolated PDZ1-Hp1 and PDZ2-Hp2 shows maxima at 358nm that are consistent with fully exposed tryptophans (**Fig.1D**). However, the emission spectrum of PDZ1-2 shows two maxima at 358 and 340nm (**Fig.1D**), which indicates the protection from solvent of a subset of PDZ1-2 tryptophans and suggests a role of the hairpin extensions in PDZ1-2 structural arrangement. Four substitutions targeting the conserved residues of the first hairpin extension (PDZ1-2^{IP225GG} and PDZ1-2^{W237G}) or of the second hairpin extension (PDZ1-2^{LP363GG} and PDZ1-2^{W375G}) were designed to further explore their contribution in the conformation of the PDZ supramodule (**Fig.2A**). Out of the four mutants, PDZ1-2^{IP225GG} and

PDZ1-2^{W375G} have a compact conformation identical to the wild-type (i.e. Rg: 27.5 Å; Dmax: 110 Å) (**Fig.2B, C, D**). The fluorescence emission spectrum of the mutant PDZ1-2^{W375G} shows that the tryptophan of the first hairpin is buried in the tandem (**Fig.2E),** denoting that the tryptophan of the second hairpin remains exposed in the compact conformation of PDZ1-2. The two other substitutions of PDZ1-2^{W237G} and PDZ1-2^{LP363GG} induce a change of conformation resulting in significantly larger Rg (31.2 Å) and Dmax (135 Å) values than PDZ1-2^{WT} (**Fig.2B, C, D**), notably from increased contribution of interatomic distances beyond 50 Å (**Fig.2D**). In addition, their dimensionless Kratky plot shows a larger bell-shape shifted to the right (**Supplementary Fig. 2**) underlying a more extended and more flexible conformation of the mutants. The substitution PDZ1-2^{LP363GG} induces a change of the first hairpin environment in the extended conformation observed by a redshift of W237 contribution on the mutant fluorescence emission spectrum (**Fig.2E**) confirming its localization close to the PDZ interface (**Fig.2E**).

Altogether, these results confirm the essential role of the hairpin extensions in the PDZ1-2 conformation. The buried tryptophan W237 of the PDZ1 hairpin and the L363, P364 residues of the PDZ2 hairpin are essential in stabilizing the interaction between PDZ domains in the supramodule and are likely to be localized near the interface.

9.3.3.3 The structural model of PDZ1-2 reveals an unexpected arrangement

We used a combination of residual dipolar couplings (RDC) and paramagnetic relaxation enhancement (PRE) data [282] to determine an ensemble of PDZ1-2 structural models. PRE data were obtained by conjugation of a MTSL probe in four individually mutated positions of PDZ1-2 (E162C, T231C, S290C and D371C) to collect long-range distance information originating from each domain or extension (**Fig.3A**). From the resulting distance-dependent line-broadening, 313 intra-domain and 342 inter-domain long-range distances were extracted. RDC were obtained from two different alignment media, Pf1-phages and ether/alcohol liquid crystalline phases, (**Fig.3B, D**) leading to 148 and 133 orientation restraints for PDZ1 and PDZ2, respectively. In addition to the PRE and RDC data, φ and ψ dihedral angle values obtained from NMR chemical shifts were also used in the model calculation.



C)

	Global	E162C	T231C	S290C	D371C
Q-factor (intra / inter)	**0.35** (0.31 / 0.37)	**0.25** (0.19 / 0.29)	**0.39** (0.35 / 0.42)	**0.38** (0.37 / 0.38)	**0.40** (0.32 / 0.45)
Restraints (intra / inter)	**655** (313 / 342)	**164** (79 / 85)	**166** (84 / 82)	**162** (75 / 87)	**163** (75 / 88)

D)

	Phage	PEG
Q-factor (PDZ1 / PDZ2)	**0.24** (0.22 / 0.28)	**0.37** (0.45 / 0.33)
Restraints (PDZ1 / PDZ2)	**141** (81 / 60)	**166** (67 / 73)

Figure 3 : Correlation between the structural ensemble and the experimental PRE and RDC used as restraints in the docking calculation. (A) Experimental ratio of cross-peaks intensities in spectra corresponding to oxidized and reduced forms (grey bars) and back-calculated ratios derived from the structural ensemble of the 25 models of lowest energy (averages and standard deviations as blue dots and lines) for spin-label attached to 4 different positions indicated as spheres (E162C in orange, T237C in yellow, S290C in dark green and D371C in light green) **(B)** Correlation plots between experimental ^{1}N-^{15}N RDC and back-calculated RDC derived from the structural ensemble of the 25 best models (top panel: phages; bottom panel: ether/alcohol). **(C)** Quality score between experimental and back-calculated cross-peak intensity ratio from the structural ensemble for all intra- or inter-domains PRE restraints and for each individual set from different probe positions. **(D)** Quality scores between experimental and back-calculated RDC from the structural ensemble for the two alignment media.

The structures of the canonical PDZ1 and PDZ2 domains of human whirlin had already been determined by NMR (PDB id 1UEZ and 1UF1, respectively). Since the murine and human sequences of each PDZ domain are almost identical (100% and 98% identity, respectively), human PDZ1 and PDZ2 structures were used as initial models for determining the conformation of the tandem PDZ1-2. First, the structure of each canonical PDZ domain was completed with an hairpin model of its C-terminal extension and was refined against RDC, dihedral angle and intra-domain PRE restraints. Then, structural models of PDZ1-2 were calculated with HADDOCK [283] using all PRE, RDC and dihedral angle restraints. After

clustering of the docked solutions, the generated models converged toward a major cluster (72% of all solutions) exhibiting a single conformation with a backbone RMSD (Root-mean-square deviation) of 1.85Å **(Fig.3D; Supplementary Fig. 4).** The averaged back-calculated PRE and RDC data of the 25 lowest-energy models fit well with the experimental profiles, with a global PRE Q-factor of 0.35 and RDC Q-factors of 0.24 for phage medium and 0.37 for ether/alcohol mixture (**Fig.3C, D**). The PREs derived from probes located on the hairpins and PDZ2 have lower Q-factors, a likely consequence of higher dynamics in the substitution regions (higher T2 and lower T1 values than the PDZ1)**.** The hairpins have a larger structural variability due to sparser resonance assignment limiting the number of available restraints.

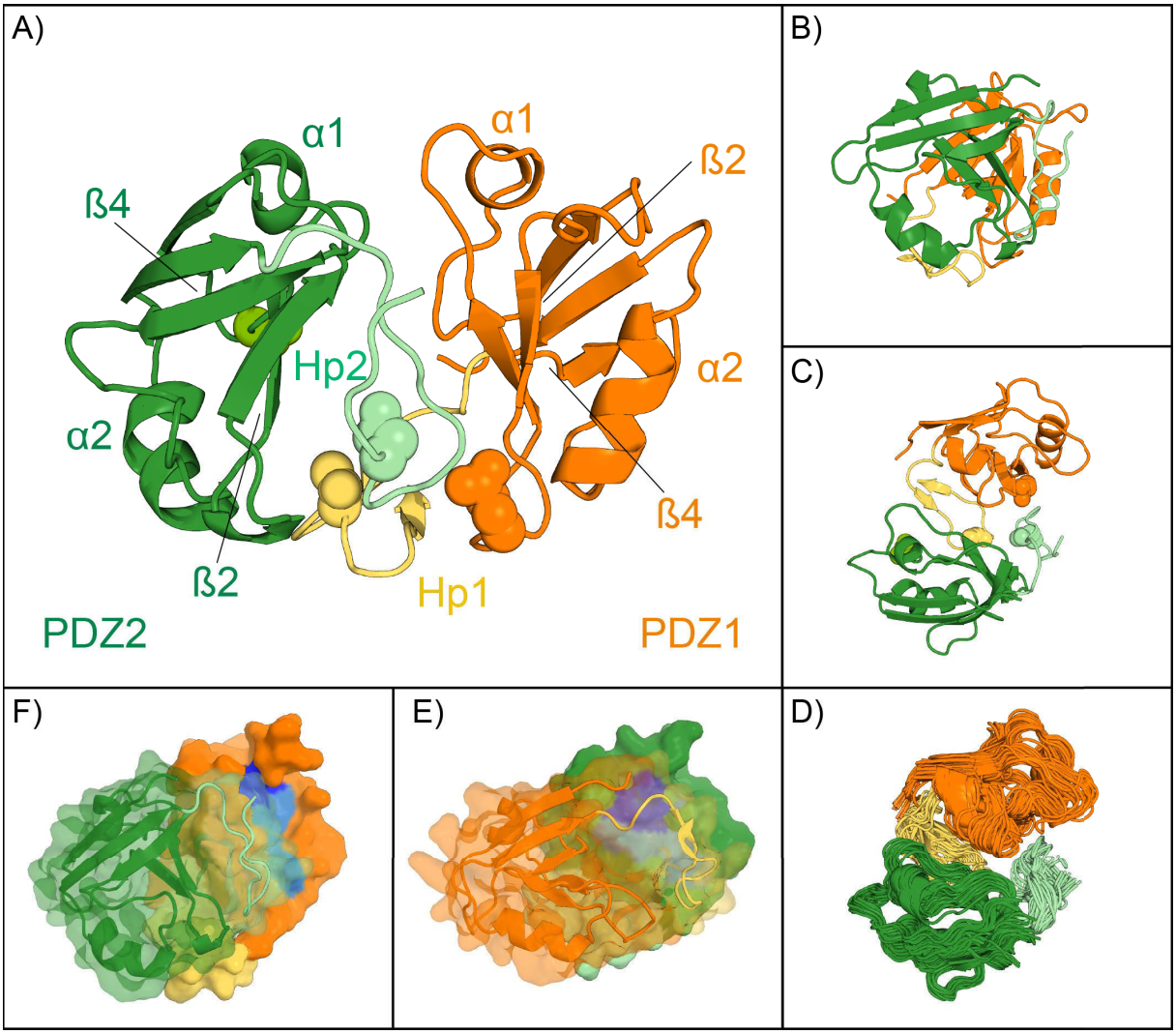


Figure 4 : Structural model of whirlin PDZ1-2 supramodule. (A, B, C) Cartoon representation of whirlin PDZ1-2 structural model in different orientations. The closest structure to the ensemble average of the 25 models is represented. PDZ1 is colored in orange with its associated hairpin in yellow, and the PDZ2 is represented in green, with its hairpin in light green. The residues in sphere are the 4 residues substituted to Cysteine for paramagnetic labeling (color coding of Fig. 3A). **(D)** Superimposition of the 25 models of lowest energy on the rigid secondary structures of both domains. **(E, F)** Volume representation of whirlin PDZ1-2 structural model showing the accessibility of the binding site (light blue) and the "GLGF" loop (dark blue) of the two PDZ domains

The PDZ domains adopt an original "face-to-face" conformation (**Fig.4**) with a symmetrical interaction of their binding site side with a relative tilt of 65°. The only direct contact made between domains involves the two flexible loops surrounding their respective α1 helix (ß3/α1 and α1/ß4 loops). All other inter-domain contacts involve the hairpins. In this conformation, the binding site of the PDZ2 is buried at the interface and would be hardly accessible to a potential ligand (**Fig.4E**), while the binding pocket of PDZ1 is directed in such a way that it remains exposed and unhindered in the tandem (**Fig.4F**).

9.3.3.4 The perturbations induced by PDZ1-2 mutations confirm the supramodular structural model

The NMR-derived structural model was validated by comparing the predicted PDZ interface to the perturbations induced by PDZ1-2 substitutions. Changes in local solvent accessibility were monitored by Hydrogen/Deuterium eXchange Mass Spectrometry (HDX-MS) **(Supplementary Fig. 5)** and perturbations of electronic environment at the residue level were evaluated by comparing the ^{1}H-^{15}N HSQC spectrum of the wild-type with spectra of mutated PDZ1-2 (**Supplementary Fig. 6)**.

The substitution of PDZ1-2^{W375G} does not modify hydrogen/deuterium exchange, chemical shifts or peak intensities of the tandem compared to PDZ1-2^{WT}, apart for residues on the second hairpin, where the mutation is located (**Fig. 5A, B, C**). This absence of perturbation confirms that the side chain of W375 is not localized at the interface between PDZ domains. The substitution of the mutant PDZ1-2^{IP225GG} solely affects the resonances of residues located on the ß2-ß3 strands (157-170) and the α1- ß4 loop (184-189) of its associated PDZ domain (**Fig.5B**). These perturbations delineate a surface corresponding to the PDZ1 residues facing the substituted side chains in the structural model, validating the position and conformation of the hairpin1. The absence of inter-domain perturbations or changes of deuterium uptake underlines the WT-like conformation of the PDZ1-2^{IP225GG} supramodule. The hairpin2 substitutions of PDZ1-2^{LP363GG} induce an increase of peak intensity on similar regions of its associated PDZ domain: the ß2-ß3 strands (291-300, 302-307), α1- ß4 loop (323-325) and the N-terminus of α2 helix (338-341) (**Fig.5B**). These differences of peak intensity confirm the defined position of the hairpin2 facing PDZ2, similar to the one observed for PDZ1/hairpin1. The peak intensity perturbation is associated to an increase in deuterium uptake in an overlapping region, indicating an increased solvent accessibility (**Fig.5A)**. Few residues of PDZ1 are distally affected in the PDZ1-2^{LP363GG} mutant they are localized on the ß2-ß3 strands (158-159; 166-170) of PDZ1 and on the first

hairpin (224-225; 232) **(Fig.5B)**. These small interdomain perturbations likely reflect the difference in conformation between the PDZ1-2^{LP363GG} and PDZ1-2^{WT} supramodules and are consistent with the PDZ1 interface determined in the structural model.

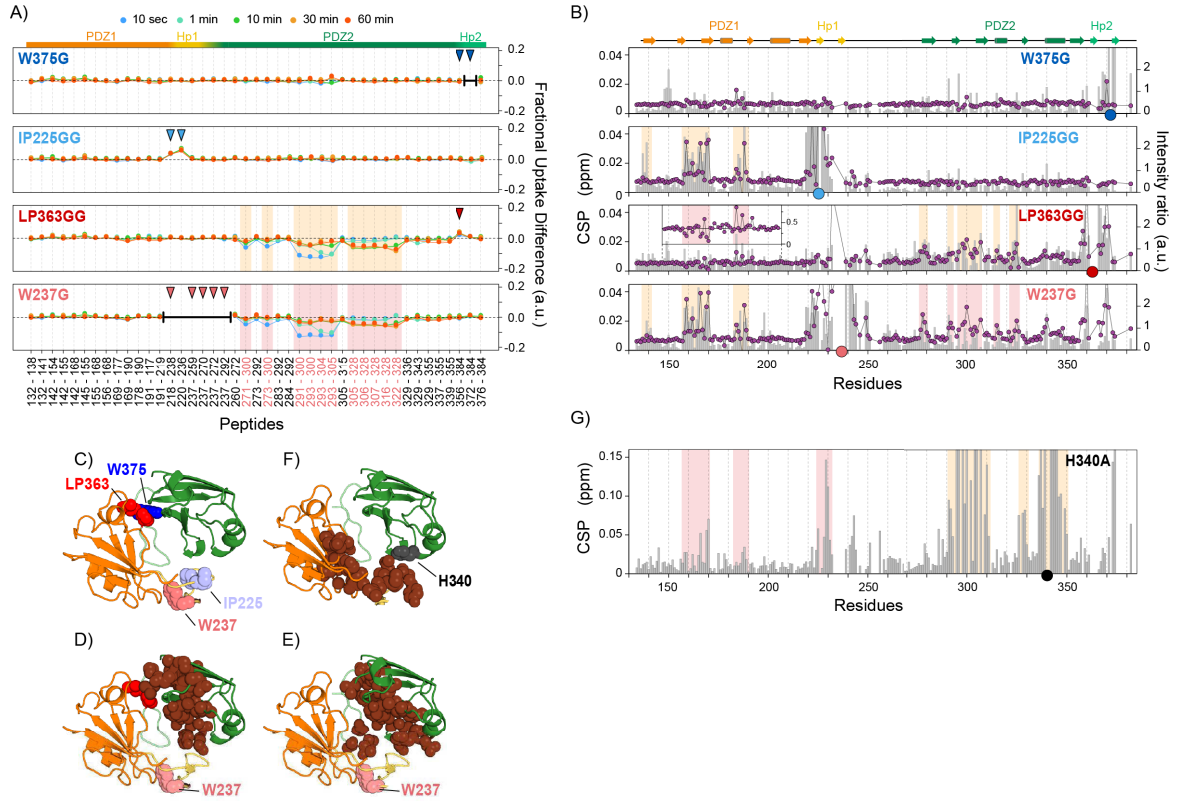


Figure 5 : Comparison of PDZ1-2 mutation effects with the PDZ1-2 supramodule model. (A) Triplicate average of differential fractional uptake of deuterium between the hairpin mutants and wild-type PDZ1-2 monitored by mass spectrometry after increasing duration of incubation (10, 60, 600, 1800 and 3600 seconds). Peptides containing the mutations are indicated by overhead triangles. Missing peptides induced by changes of enzymatic proteolysis in mutants are indicated by black segments. Peptides with significantly modified uptake are highlighted in red on the graph and the x-axis. **(B)** NMR chemical shift perturbations (grey bars) and peak intensity ratio (violet dots) between ^{1}H-^{15}N-HSQC spectra of hairpin mutants and wild-type PDZ1-2. Spheres using the mutant color-coding of figure 2A show mutation positions. The analysis of perturbations induced in compact mutants (blue) and extended mutants (red) delineates the PDZ interface (red transparent boxes). **(C)** whirlin PDZ1-2 structural model with the hairpin substitutions PDZ1-2^{IP225GG}, PDZ1-2^{W237G}, PDZ1-2^{LP363GG}, PDZ1-2^{W375G} shown in sphere Spheres using the color-coding of figure 2A **(D)** Mapping (dark red spheres) of the increase in Deuterium uptake observed by H/D exchange in (B) in the "open" tandem PDZ1-2^{W237G} and PDZ1-2^{LP363GG} onto the PDZ1-2 structural model. The PDZ1-2^{W237G} and PDZ1-2^{LP363GG} mutations are shown as light (W237G) and bright red (L363G-P364G) spheres, respectively. **(E)** Mapping onto whirlin PDZ1-2 structural model of the residues affected by inter-domain CSP and intensity changes (dark red spheres) observed with the PDZ1-2^{W237G} substitution (in light red spheres). **(F)** Mapping onto whirlin PDZ1-2 structural model of the residues affected by inter-domain CSP and intensity changes (dark red spheres) observed with the PDZ1-2^{H340A} substitution (in black). **(G)** NMR chemical shift perturbations (grey bars) induced by the H340A mutation in PDZ2 (black sphere). Inter-domain perturbations on PDZ1 and hairpin1 (highlighted in red) are due to the opening of mutant compared to the wild-type and are in perfect agreement with the structural model.

The other mutant in an extended conformation, PDZ1-2^{W237G}, combines the perturbations observed on the two previous mutants. PDZ1 residues display chemical shift perturbations identical to the ones observed on PDZ1-2^{IP225GG} mutants (**Supplementary Fig. 7**), resulting from direct modification of local environment due to the substitution. On PDZ2, An increase in peak intensities was also observed for similar residues than the ones observed in the mutant PDZ1-2^{LP363GG} (**Fig.5B**). This interdomain increase of peak intensity on PDZ2 and hairpin2 is associated with a significant increase of deuterium uptake in the regions identical to PDZ1-2 ^{LP363GG} (**Fig.5A, B, D, E**). This decrease in local exchange and increase in solvent accessibility on PDZ2 can only be assigned to the difference of global conformation between PDZ1-2^{W237G}and PDZ1-2^{WT}. The surface formed by the perturbed residues is fully consistent with the interface of the supramodule structural model (**Fig.5D, E**).

The residues of PDZ1 are scarcely affected in the mutants in extended conformation. To confirm the PDZ1 interface in the supramodule, a new substitution on PDZ2 was designed based on the structural model. In the NMR-derived model, H340 from the PDZ2 α2 helix is in close proximity to V230 of the PDZ1 hairpin (**Fig.5F**). We aimed at preventing this contact by substituting H340 to an alanine. The SAXS curve of PDZ1-2^{H340A} mutant is characteristic of an extended conformation and its spectrum of fluorescence emission shows a fully exposed W237 (**Supplementary Fig. 8**). As expected, the substitution affects PDZ2 (ß2 and a2) and hairpin2 but also affects PDZ1 resonances (**Fig.5G)**. Compared to the wild-type, significant chemical shifts were observed on the ß2-ß3 strands of PDZ1 (158-159; 165-170) and on the α1-ß4 loop (187-188). The affected residues match with the region of interdomain intensity changes observed on PDZ1-2^{LP363GG} (**Fig.5B, G**). All assigned residues of the hairpin1 are also affected by H340A substitution (221-232), with the largest CSP observed on residues Y229 and V230, the two residues predicted to be nearest to H340 in our structural model (**Fig.5F, G**).

The inter-domain CSP, variations of peak intensities and changes of solvent accessibility observed between wild-type PDZ1-2 and mutants PDZ1-2 confirm the NMR-derived structural model and underline the essential role of both hairpins in PDZ1-2 supramodule (**Fig.5D, E, F**). These combined results suggest that the hairpin1 is folded and interacts with PDZ1 regardless of the global conformation of the PDZ1-2 tandem and that the residues at PDZ2 interface are affected by fast to intermediate exchange, absent from extended mutants.

9.3.3.5 PDZ1-2 is in exchange equilibrium between closed and open conformations

The ^{1}H-^{15}N HSQC spectrum of the wild-type PDZ1-2 displays resonances affected by line-broadening characteristic of exchange in the 10^{-5} s to 10^{-4} s timescale. These broad resonances become narrow in the spectra of mutants having extended conformations, supporting the relation between exchange and global conformation of PDZ1-2. Inter-domain dynamics is also highlighted by the slight difference of τc between PDZ1 and PDZ2 in PDZ1-2 (12.4 ns and 13.7 ns respectively, **Fig.1C**). The PRE-derived ensemble represents the main, compact conformation of PDZ1-2 that is unlikely to characterize the whole conformational landscape explored by PDZ1-2 in solution [284]. The theoretical scattering curve calculated from the NMR-derived compact model does not reproduce the experimental scattering profile of PDZ1-2, underlying the presence of larger PDZ1-2 conformations in solution. To extract the information about the conformational landscape of PDZ1-2, we aimed at selecting ensembles of conformers in good agreement with our SAXS and RDC dataset [285–287].

To thoroughly cover the conformational space of PDZ1-2, a large pool of structures **(**160 000 conformers**)** was generated by Molecular Dynamic (MD) starting either from the NMR-derived structural model or from randomly generated conformations of the PDZ tandem. A genetic algorithm was subsequently used to find ensembles of conformations for which the average theoretical data fit best with the experimental o nes. Four sizes of ensembles (1, 5, 30 and 100) were used to determine the minimum number of conformers required for a good fitting quality. The ensembles allow repetitions of conformers; therefore the number of unique structures could be much lower than the preset ensemble size (**Supplementary Fig. 9**). The dataset were first fitted separately, for the SAXS experiment, the chi^2 decreased drastically from 3.73, by fitting a single conformation, to 0.99 for ensembles of 5 conformers and reached a plateau from that size (**Fig. 6A**). For the RDC data, the Q-factors decreased from 0.51 and 0.35 to 0.34 and 0.24 for RDC ether/alcohol mixture and phage medium, respectively, and also became constants after ensembles of 5 conformers (**Fig. 6B**). A single conformation is not sufficient to fit the data, however minimum ensembles of 5 conformations are appropriate and sufficient to fit all experimental data separately.

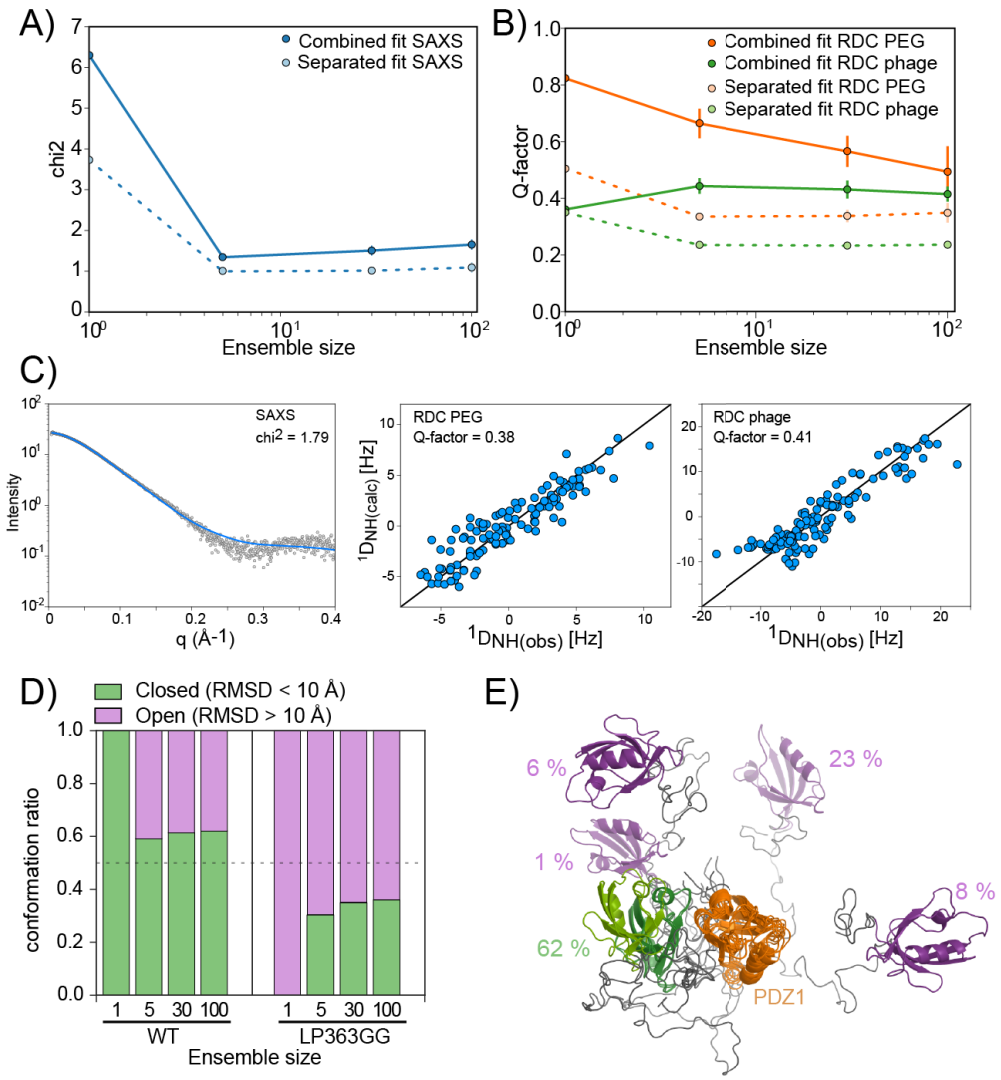


Figure 6: Selection of conformations describing PDZ1-2 equilibrium. (A) Fit improvement in chi^2 according to the ensemble size in the genetic algorithm selection for the SAXS data, when the selection is performed individually (dotted blue line) or combined with the RDC dataset (continuous blue line). **(B)** Fit improvement in Q-factor according to the ensemble size in the genetic algorithm selection for RDC ether/alcohol mixture and phage medium, when the selection is performed individually (dotted orange and green lines, respectively) or combined with the RDC and SAXS dataset (plain orange and green lines, respectively). The error bars are computed over 10 individuals run. **(C)** Left panel: back-calculated SAXS profile derived from the best ensemble (100 conformations preset, with 18 unique conformations selected, in blue) and the experimental profile (in black dot). Middle panel and right panel: correlation between the back-calculated RDC from this ensemble and the experimental RDC from ether/alcohol mixture and phage medium (middle and right, respectively). The chi^2 and Q-factors are indicated. **(D)** The conformations are clustered in two categories according to their RMSD with the NMR derived structural model, conformations with RMSD lower than 10 Å are considered "closed" (green bars) while the others are considered "open" (violet bars). The proportion of each cluster is represented on stacked bar chart for each ensemble size selected on the wild-type SAXS and RDC dataset or on PDZ1-2^{LP363GG} SAXS dataset. **(E)** Representative ensemble of PDZ1-2 superimposed on PDZ1 (orange) showing the PDZ2 in compact conformation (green) or in extended conformation (violet) with the proportion of each conformer.

The three dataset were then combined in the genetic algorithm by a global fitness score with individual weights determined from the separated fits, to find ensembles minimizing both chi^2 and Q-factors of SAXS and RDC data, respectively. In the combined dataset, the fitting score improves drastically between single conformations and ensembles of 5 conformers and continues to slightly improve with larger ensembles. The best global fitting score was obtained with ensembles of 100 conformers that encompassed only 18 unique conformations. The ensemble of best score had a final chi^2 of 1.79 for the SAXS dataset and Q-factors of 0.38 and 0.41 for the ether/alcohol mixture and phage medium RDC dataset, respectively (**Fig.6C**).

The weighted conformations obtained in all ensembles were clustered in two discrete groups according to their RMSD from the NMR-derived structural ensemble. Structures with an RMSD lower than 10 Å were grouped as "closed conformations", while the more diverging structures constituting a cluster of "open conformations". In all ensembles of all sizes, the proportions are similar, the closed conformation cluster is the main conformation with around 60% of weight, while the more open conformers cumulatively account for the remaining 40% (**Fig.6D**). The best fitting ensemble is composed at 62 % of conformers in a closed conformation, representative conformations of this ensemble are depicted on figure 6E. The same selection strategy was used to determine ensembles minimizing the chi^2 of SAXS data from the extended conformation mutant PDZ1-2^{LP363GG} (**Supplementary Fig. 10**) and an identical conformers clustering was performed. In this case, the contribution of closed conformation (to around 35%, depending on ensemble size) drastically reduced to the profit of the predominant open conformers (to around 65%).

The selection of a minimum ensemble of conformers fitting SAXS and RDC dataset from a large structural ensemble allowed uncovering low populated conformations of the PDZ1-2 tandem. PDZ1-2 is in equilibrium between a supramodular structure determined by the NMR-derived model and a large ensemble of extended conformations in which the two PDZ domains of whirlin do not interact. Hairpin mutations shift the conformational equilibrium by promoting open conformations of PDZ1-2 tandem.

9.3.3.6 PDZ1 hairpin folding improves its binding affinity to cellular partners

We could not identify a partner of PDZ2 when screening the PBM of proteins involved in the hearing network (**data not shown**). Thus, most PDZ-mediated interactions of whirlin N-terminal half are likely to involve PDZ1. The Usher protein Sans colocalizes with

whirlin in the connecting cilium of photoreceptor cells [183,256] and possesses a C-terminal PBM of type 1. The affinity of the C-terminal peptide of Sans was measured for several constructs to determine the impact of the hairpins and of the PDZ1-2 supramodule conformation on binding properties of whirlin.

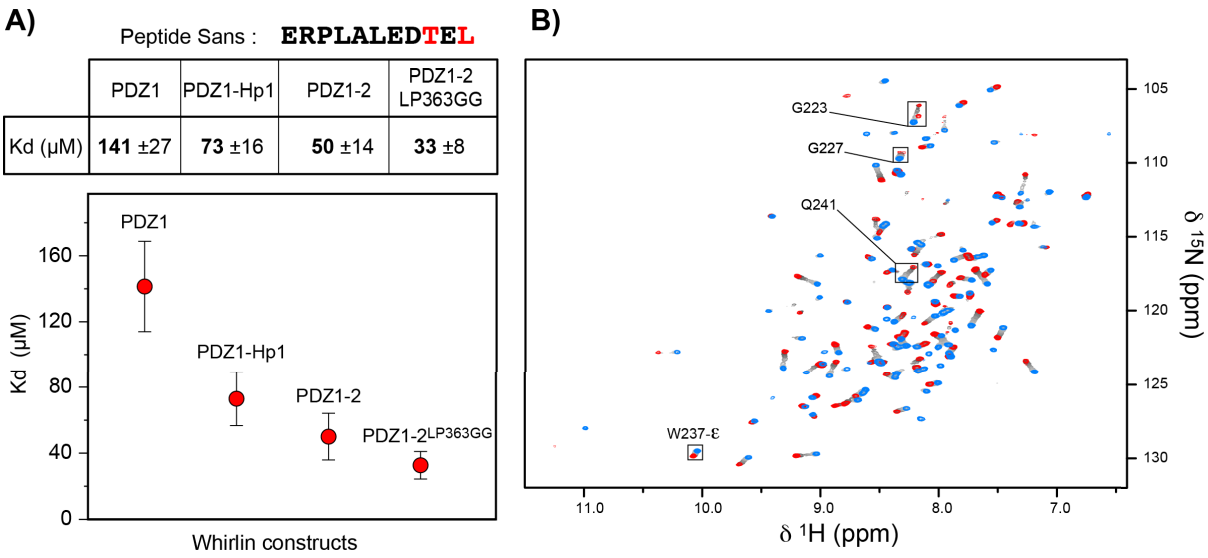


Figure 7: Interaction between whirlin and the peptide Sans (A) Table of affinity between Sans peptide and different constructs of whirlin and plot showing the decrease of dissociation constants according to the whirlin constructs. **(B)** ^{1}H-^{15}N HSQC spectra of the titration of PDZ1-Hp1 with Sans peptide. The PDZ1-Hp1 without peptide is shown in blue and the last point of the titration in red. The eight intermediate points are shown in shades of grey. Examples of residues affected by peptide binding, in the hairpin1, are shown in boxes.

PDZ1, without extension, binds to the Sans peptide with a constant of dissociation of 141 µM (**Fig.7A**), in the lower range of affinity commonly found for PDZ domains interactions. The addition of the C-terminal hairpin leads to a significant increase of affinity by two-fold (Kd of 73 µM, **Fig.7A**). The NMR titration shows shifts on the resonances of hairpin1 residues, confirming the direct involvement of this C-terminal extension in peptide binding (**Fig.7B**). An additional gain of affinity of 1.5 is obtained with the tandem PDZ1-2 (Kd of 50 µM, **Fig.7A**) although PDZ2 is not able to bind the ligand. On the ^{1}H-^{15}N HSQC spectrum, peptide binding only affects resonances of PDZ1 and its associated hairpin1 (**Supplementary Fig. 11**), indicating that the affinity gain is indirectly promoted by the linker and the downstream PDZ2 domain. Surprisingly, the PDZ1-2^{LP363GG} mutant, in predominant open conformations, has an affinity improved by another 1.5 fold (Kd of 33 µM, **Fig.7A**) for Sans peptide.

The unexpected improvement of affinity in an extended mutant shows that the subset of open conformations is more competent for Sans peptide binding than the closed

conformation. HDX and NMR experiments showed that the conformation of PDZ1 and the hairpin1 is only slightly affected by PDZ1-2^{LP363GG} substitution. The hairpin1 is predominantly folded and in contact of PDZ1 in the open-conformation, suitable to bind the C-terminus motif of a partner protein. The open-conformation might facilitate the access to PDZ1 binding groove, improving the "on rate" of binding. The improvement of binding affinity compared to PDZ1-Hp1 is likely due to the addition of the C-terminal sequence that lowers the conformational entropy of the hairpin and thus increasing the target-binding affinity of the PDZ1 domain.

The PDZ2 of whirlin might have evolved to favor PDZ1 stabilization at the expense of its own binding capacity, as described for the PDZ4 of Scribble [154], and for the PDZ2 of GRIP1 [148]. In addition to improving PDZ1 affinity by 3- to 4-fold, PDZ1-2 supramodule enables a fine-tuning of whirlin affinities for its partners. The difference of affinity between "closed" and "open" states allows modulating indirectly peptide binding by affecting tandem conformation.

9.3.4 Discussion

Supramodules play an important role in regulating the biomolecular processes of proteins. While the study of stable and large assemblies can be investigated by a variety of techniques [288], the structural determination of flexible and transient complexes remains challenging for most [289]. To characterize the structure and the energy landscape of such dynamic ensemble, the combination of NMR and SAXS data have proven to be particularly fruitful [290]. Here, we used multiple dataset from PRE and RDC to determine a structural model of the predominant conformation of two weakly interacting PDZ domains present at the N-terminal side of whirlin. The model reveals an original arrangement with two hairpin extensions maintaining the contact between PDZ domains. NMR also showed the difference in domain tumbling times and in the conformational exchange-type motions at the PDZ interface, revealing the existence of interconverting conformational states of the tandem PDZ1-2. In this case of fast to intermediate exchange, PRE data predominantly report on the shortest distances and thus on the most compact conformations [291]. To characterize the exchange equilibrium of PDZ1-2 in solution, we used dataset averaging all conformation on large timescale, from SAXS and RDC experiments. We determined ensembles fitting our experimental data using a genetic algorithm that revealed an equilibrium between a major compact state similar to the PRE-derived structure and an ensemble of less populated states of

PDZ1-2 tandem composed of extended conformations. The closed/open conformer equilibrium could be shifted toward the relaxed state using appropriate mutations of the C-terminal extensions of each PDZ domain.

The proximity of the two PDZ domains in the closed state is mainly mediated by the two structured extensions at the C-terminus of each domain. The hairpin of PDZ1 is preformed in the isolated domain and is stabilized in the supramodule. An additional function of this extension is to increase the affinity of the isolated PDZ1 for its partners, encoded by causal genes of Usher syndrome. The addition of the linker-PDZ2 further improve he affinity for a total gain of 4-fold compared to the isolated domain without directly interacting with the ligand. Similar modulations of target-binding affinity by lowering conformational entropy have previously been described in others PDZ and PDZ-containing proteins [133,135] and entropy modulation seems to be a recurring evolutionary approach to increase the affinity of a given domain for its partners [59,76,292].

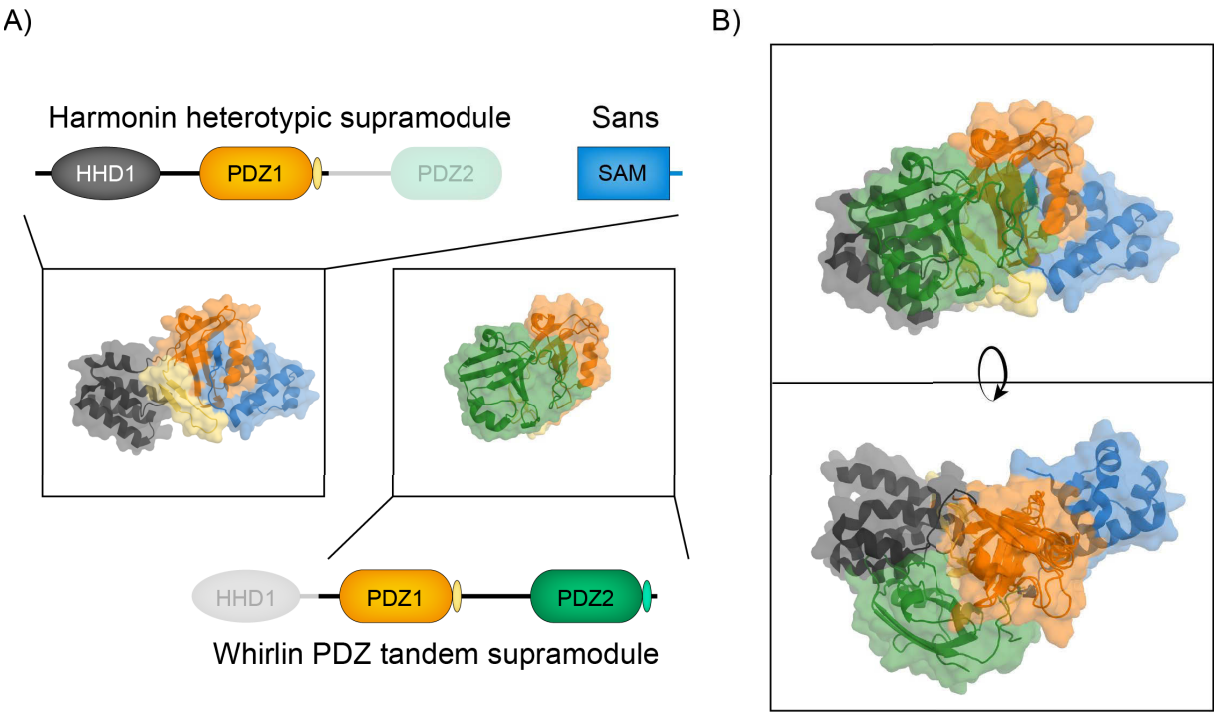


Figure 8: Superimposition of whirlin PDZ tandem with harmonin N-terminal supramodule. (A) The harmonin supramodule consisting of HHD1 (black) and PDZ1 (orange) in complex with the C-terminus of Sans (blue) and the whirlin supramodule PDZ1-2 consisting of PDZ1 (orange) and PDZ2 (green). **(B)** Superimposition of whirlin and harmonin PDZ1 domain within their respective supramodule.

Little is known about the hair bundle molecular complexes; our structural model brings a new insight into this highly intricate network that favors higher-order organization [68,180,211]. *In vitro* and *in vivo* evidences of interaction between the two first PDZ domains of

whirlin have been presented directly from pull-down of transfected mammalian cells [214] and indirectly from the inability to natively immunolabel the PDZ1-2 of whirlin in cochlear hair cells, likely from the presence of intra- or intermolecular complex (**Ebrahim et al. 2016**). Whirlin shares a common N-terminal topology with harmonin until the C-terminal end of PDZ2, after which whirlin contains an additional hairpin that is essential for the PDZ1-2 supramodule. In harmonin, the N-terminal HHD domain forms a heterotypic supramodule with the PDZ1 and its hairpin extension. When comparing harmonin heterotypic supramodule with whirlin homotypic supramodule (**Fig.8A**), we see that the interactions of PDZ1 with either the HHD of harmonin or the PDZ2 of whirlin are not mutually exclusive (**Fig.8B**). The three N-terminal domains of whirlin are included in the long isoform, in photoreceptors and during hair cells development but absent from the short C-terminal isoform expressed solely in hair cell bundles. Mutations of the N-terminal region of whirlin are responsible for deaf-blindness syndrome, while mutations of other regions only affect hearing [232]. An appealing proposition would be that the N-terminal half of whirlin organizes in a tri-partite heterotypic supramodule, with the specific function of coordinating the Usher2 complex.

Whirlin is one of the essential proteins of hair cells. These proteins are involved in stereocilia elongation and polar orientation during development; they are part of the mechanotransduction signaling, the stereocilia bundle cohesion and the neuronal connection at the ribbon synapse. Furthermore, they take part in the renewal of rhodopsin in photoreceptors [293] and whirlin was recently reported to play a role in the mechanosensory signaling of proprioceptors [246]. The common originality of these sensory cells is to possess exquisitely refined architectures transducing information almost instantly through network of scaffolding proteins without catalytic cascade. The thrift of protein regulating these organizations requires versatility and modularity at every level. Usher proteins are multi-domain proteins translated from numerous splice variants that are differentially expressed in space and time. Molecularly, the functions of their constituting domains are tuned by their arrangement in dynamic supramodules and oligomers that brings complementary partners together, promoting functional synergy [294]. The study of sensory cells has shown that scaffolding proteins can be much more than passive docking site, that they can be active components and dynamic modulators of signal transmission without the use of catalytic functions.

9.3.5 Materiel & methods

9.3.5.1 Sample preparation

All construct of whirlin are based on murine isoform 4 (Q80VW5-4): PDZ1 (137-228) PDZ1-Hp1 (137-246) PDZ2-Hp2 (275-358) PDZ1-2 (137-358). All are cloned into a pGST//2 vector (derived from pGEX-4T-1; Amersham). A TEV cleavage site is introduced between the N-terminal GST tag and the protein sequence. All mutants are generated using QuickChange-2XL (Agilent) kit and protocol. Plasmids verified for DNA sequence (Eurofins, Paris) are transfected to *E.coli* BL21 (DE3) star (Invitrogen). Bacteria are expressed in LB medium or M9 minimal medium supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose. At DO_{600nm} of 0.8, protein expression is induced by the addition of 1 mM IPTG for 3 h at 30 °C. Bacteria are harvested by centrifugation, suspended in 50 mM Tris-HCl, 250 mM NaCl, 5 mM 2-mercaptoethanol, pH 7.5, buffer containing protease inhibitors (Complete, Roche Diagnostics) and disrupted by sonication. The extract is centrifuged at 15000 RCF for 1 h. Tagged proteins are purified by affinity chromatography using GSTrap (GE Healthcare) followed by TEV cleavage overnight at 4 °C and a last step of size-exclusion chromatography (SEC) using a Sephacryl S-100 HP 16/60 (GE Healthcare). All proteins are purified in 50mM Tris-HCl pH 7.5, NaCl 150mM and TCEP 0.5 mM if a Cysteine is introduced by mutation.

9.3.5.2 Small angle X-ray Scattering

X-ray scattering data are collected at the BioSAXS beamline BM29 at the European Synchrotron Radiation Facility (Grenoble, France) and at the SWING beamline at Soleil (Saclay, France). 50 µl of sample is injected on a Superdex-200 5/150GL (GE) size exclusion column in-line with the SAXS measuring cell. All samples are dialyzed in the same buffer (Tris 50mM pH 7.5, NaCl 150mM) for 24h at 4°C and then concentrated to 8.8 to 11.2 g/l prior to the experiments. Frames with a duration of 1 s are recorded during the whole elution time. Approximately 100 frames recorded in the dead volume of the column are averaged to generate the buffer subtraction curve and the 10 to 30 best frames corresponding to the top of the elution peak are average to generate the result curve. Primary data reduction was performed using the dedicated beamline software BsxCuBE and FOXTROT. The one-dimensional scattering intensities were expressed as a function of the modulus of the scattering vector $Q = (4\pi/\lambda) \sin\theta$ with 2θ being the scattering angle and λ the X-ray wavelength. Buffer intensities were subsequently subtracted from the respective sample

intensities using software package PRIMUS [257]. The radii of gyration were evaluated using Guinier approximation [258]. The Dmax was determined from distance distribution function P(r) obtained with the program GNOM [259]. The similitude between curves of the different constructs was determined after scaling with SIBYLS SAXS structural comparison map (SCM) webapp [295] using chi^2 comparison on a q range of 0.119 – 4.007.

9.3.5.3 Fluorescence

Measurements of fluorescence emission were performed at 25°C in a LS50 spectrofluorometer (PerkinElmer) using a 109F cuvette (Hellma Analytics) with thick side to reduce the inner filter effect. Excitation is set at 295 nm and emission spectrum is recorded from 300 nm to 450 nm, with a 5 nm bandwidth for both. Proteins are diluted in the same buffer (Tris-HCl 50 mM pH 7.5, NaCl 150 mM) at concentrations ranging from 2 to 4 μM, depending on the number of tryptophans contained in the construct. Buffer fluorescence is subtracted from protein emission spectra and the spectra are corrected according to the measured concentration.

9.3.5.4 NMR measurement

All experiments were performed on a Varian Inova 600 MHz spectrometer equipped with a cryogenically cooled triple resonance $^{1}H\{^{13}C/^{15}N\}$ PFG probe. Data process and analysis were performed using NMRpipe [260], ccpNmr Analysis 2.4 [261].

Chemical shift perturbations: ^{15}N-labeled NMR samples of wild-type and mutants (I225G-P226G, W237G, L363G-P364G, W375G) PDZ1-2 were prepared at concentrations of 180 *m*M in 250 *m*l of a buffer containing 50 mM Tris-HCl, 150 mM NaCl and 12 % D_2O and at a carefully monitored pH of 6.0. ^{15}N-TROSY-HSQC spectra of wild-type and mutants PDZ1-2 were recorded at 25°C, with 120 (61 ms) and 1024 (121 ms) complex points (acquisition times) in ^{15}N and ^{1}H, respectively. Chemical shift perturbations (CSP) were calculated as the weighted average (^{1}H, ^{15}N) chemical shift differences $\Delta\delta_{avg}$ between the wild-type and mutants PDZ1-2 as follow: $\Delta\delta_{avg} = ((\Delta\delta H)^2+(\Delta\delta N\times 0.159)^2)^{1/2}$.

Residual dipolar couplings: ^{1}H-^{15}N residual dipolar couplings were obtained by recording IPAP-HSQC experiments [263] on ^{15}N-labeled samples of wild-type PDZ1-2 at concentrations ranging from 150 to 600 *m*M in the absence and in the presence of two different oriented media containing either 10 g/l Pf1-phages (ASLA biotech) and 300 mM

NaCl or a liquid crystalline mixture of 3 % dodecyl-tetra(ethylene glycol) (C12E4) and 2-hexanol at a C12E4/2-hexanol molar ratio of 1.36 [296].

Paramagnetic relaxation enhancements. To prepare each ^{15}N-labeled NMR sample of cysteine mutant PDZ1-2 (E162C, T231C, S290C and D371C) for paramagnetic relaxation experiment, TCEP was first removed from buffer solution using SEC on a 10DG column (Biorad). The thiol-reactive nitroxide label (1-oxyl-2,2,5,5-tetramethyl-D3-pyrroline-3-methyl (MTSL; Santa Cruz biotechnology)) was added to the solution at a molar ratio of 10:1 MTSL:protein and incubated for 24 h at room temperature (RT). Unbound MTSL was removed using another 10DG column. The four modified proteins were concentrated, pH is equilibrated to 6.0 and 12 % D_2O is added for a final volume of 240 µl leading to protein concentrations ranging from 150 µM to 350 µM for the oxidized (paramagnetic) sample. For the reduced (diamagnetic) sample, ascorbic acid (from a stock solution at 100 mM) was added to an excess of ten times with the protein concentration and incubated for 24 h at RT to ensure complete reduction of the spin label. Identical ^{15}N-TROSY-HSQC spectra were recorded with a recycle delay of 5s on each oxidized and reduced sample (8 to 12 h experimental times). PRE analysis was performed according to the method introduced by Battiste & Wagner [282] Intensity (peak heights) ratios of cross-peaks in oxidized versus reduced spectra were converted into distance restraints from the spin label, using estimations of the intrinsic transverse relaxation rates of each amide R2(^{1}H) from the peak line-widths in the ^{15}N-TROSY-HSQC spectra and a global correlation time τc of 12.6 ns [278].

Peptide titration : whirlin constructs were titrated using the biotinylated peptide corresponding to the last 11 residues of *Mus musculus* SANS. The peptide was synthesized on solid phase support using the Fmoc (9- fluorenyl methoxycarbonyl) strategy (ProteoGenix) The biotine was linked to the peptide sequence with a neutral TTDS linker (13-amino-4,7,10-trioxa-tridecayl-succinamic acid). NMR samples initially contained between 72 µM and 100 µM ^{15}N–labeled whirlin constructs. Unlabeled peptide was added stepwise to a final PDZ/peptide ratio of 1:7 to 1:10. ^{1}H-^{15}N HSQC spectra were recorded at 9 to 11 different titration points. Average (^{1}H, ^{15}N) chemical shift changes were calculated as $\Delta\delta_{avg} = ((\Delta\delta H)^2+(\Delta\delta N\times 0.159)^2)^{1/2}$. The dissociation constants of the respective complexes (Kd) were obtained by fitting the titration data with a model assuming a 1:1 complex formation and with nonlinear regression.

9.3.5.5 Structure calculations

Intensity ratios of cross-peaks in paramagnetic (oxidized) and diamagnetic (reduced) spectra (I_{ox}/I_{red}) were converted into distances as described in Battiste & Wagner, 2000. To account for measurement errors, distance bounds were derived assuming an error of 0.2. For ratio $I_{ox}/I_{red} \leq 0.2$, distances were caped to an upper bound corresponding to $I_{ox}/I_{red} = 0.4$ (i.e. 0.2 + 0.2) and $I_{ox}/I_{red} \geq 0.85$ were converted to a lower bound corresponding to a I_{ox}/I_{red} ratio of 0.65 (i.e. 0.85 - 0.2). Chimeric structures of each individual PDZ domain were constructed by attaching C-terminal hairpin extensions to the structures of human whirlin PDZ domains (PDB id 1UEZ and 1UF1) using residues S136-V219 and G276-T356 for PDZ1 and PDZ2, respectively. Since the 1UEZ and 1UF1 PDB entries are NMR structures, all deposited models were used to generated 20 chimeric structures for each PDZ domain. Atomic model of the hairpin extensions (T220-V238 for PDZ1 and V357-I376 for PDZ2) were obtained by homology modeling using the PDB structure of the PDZ1 domain of harmonin (PDB id 3K1R). Residues E162, T231, S290 and D371 were mutated to cysteine and MSTL spin-labels attached to these residues were added to the molecular definition. PRE distance restraints, corresponding to distance bounds derived from I_{ox}/I_{red} ratio, were generated between the amide hydrogen and the nitrogen atoms of the MTSL ring. In addition, backbone dihedral angles distance restraints predicted from chemical shifts with TALOS-N [262] were used. Residual dipolar couplings were converted into restraints using the SANI term in CNS [297]. For each medium, parameters of the anisotropy tensor (magnitude and rhombicity) were determined by best-fitting RDC data on the individual PDZ chimera using PALES [298] and confirmed by a Maximum Likelihood method [299]. First, individual structures of PDZ1 and PDZ2 chimera were refined using NMR data collected on the PDZ1-2 (intra-domain PRE restraints, RDC restraints and backbone dihedral angle restraints). The solvent refinement protocol implemented in ARIA/CNS [300] was used starting from 20 chimeric structures as initial conformations. Refinements were carried out with four replicas of each MTSL spin-label side-chain and PRE distances estimated with $\langle r^{-6} \rangle$ averaging [301]. In a second stage, structural models of PDZ1-2 were calculated using HADDOCK/CNS [283]. The 10 lowest-energy structures of each refined PDZ chimera were used as initial structures for the docking, resulting in 100 different starting combinations of PDZ1-2 structures. All available PRE restraints (intra- and inter-domain) together with RDC restraints (phage medium and ether/alcohol mixture) were used simultaneously. MTSL labeled residues were considered as

semi-flexible segments while the junctions between PDZs and hairpin extensions were fully flexible segments. Default values for the number of generated structures were increased to 2,000 for the rigid-body docking stage (default 1,000) and to 500 for the flexible and solvent refinement stage (default 200). Similarly, the number of molecular dynamics steps was doubled for the rigid-body and flexible stages of HADDOCK. During refinement of individual PDZ domains and the solvent refinement stage of HADDOCK, a network of Cα-Cα (< 12 Å) and N-O (< 4 Å) distances was also restrained using a log-harmonic potential to avoid large structural distortions of the canonical PDZ fold [302,303], where restrained distances were measured on each individual conformations before refinement. Additionally, hydrogen-bond and backbone dihedral angle restraints for the C-terminal hairpin extensions were included to maintain their local conformation. In all calculations, non-bonded interactions between MTSL side-chain replicas were turned off. Structures generated by HADDOCK after solvent refinement were clustered using a 4.5 Å cut-off.

9.3.5.6 Structure ensemble from MD simulations

The most stable structure in the cluster with the lowest energy determined by HADDOCK was used to generate an ensemble of 160 PDZ1-2 conformers by removing the inter domains restraints using Modeller. The set of conformations obtained were then used as starting points to run 160 independent molecular dynamics simulations. The starting conformations were parametrized with the Amber99SB force-field using the OpenMM molecular simulations library [304]. The simulations were performed in explicit solvent using the TIP3P water model. A Langevin dynamics integrator with a 2 fs time step and a 1 ps^{-1} collision rate was used, with a bath temperature of 300 K and bond lengths to hydrogen constrained. Long-range electrostatic interactions were calculated with the Particle Mesh Ewald protocol. Molecular dynamics trajectories were recorded over 1 ns every ps leading to 1,000 conformations recorded per trajectory. Hence, a total of 160,000 snapshots of conformations were produced. The obtained conformations were clustered in 1,000 structures.

9.3.5.7 *Ab initio* prediction of residual dipolar couplings

Theoretical residual dipolar couplings were predicted *ab initio* for full-length MD snapshots using the software PALES [298]. For each conformation, an order matrix (5 independent values of the alignment tensor) was predicted using the steric (simulating ether/alcohol medium) and electrostatic (simulating Pf1-phage medium) models in PALES.

Theoretical RDC were then calculated using the predicted order matrix and the N-H vectors taken form the RDC-refined models of PDZ domains after superimposition on the full-length conformation.

9.3.6 Genetic algorithm

A genetic algorithm was implemented to find the minimal ensemble that minimizes the error between the theoretical data computed from the structures resulting from the MD simulations and the heterogeneous experimental data. The python code is available on GitHub (https://github.com/bougui505/genetic_algorithm). 1,000 steps of genetic algorithm were performed to ensure convergence of the score. At each iteration the number of ensemble generated is set to 1,000 with a cross over frequency of 0.8 and a mutation frequency of 1. The separate fittings are used to determine the relative weighting of each dataset by defining their weight as the inverse of their fitness score (chi^2 or Q-factor). This weighting strategy normalizes the final fitness score to 1.0 upon convergence [305]. The genetic algorithm is repeated 10 times for every fit. The weights for the combined score are set to the inverse of the average fitness scores obtained for the individual fits.

9.3.6.1 Hydrogen/Deuterium Exchange - Mass spectrometry

Sample preparation. The concentration of each protein was adjusted to 8.5 µM in 50 mM Tris buffer, 150 mM NaCl prior to labeling. After 1 h equilibration at room temperature, the exchange reaction was initiated by adding 42 µl of D_20 buffer (50 mM Tris, 150 mM NaCl, pD 7.5) to 8 µl of protein. Aliquots of 9.5 pmol of labeled protein were removed at defined exchange time points (10 s up to 1 hour) and quenched upon mixing with an ice-cold 2% formic acid/4 M urea solution to decrease the pH to 2.6. Quenched samples were immediately snap frozen in liquid nitrogen and stored at -80 °C until analysis. Samples were prepared in triplicate for each time point and for each protein.

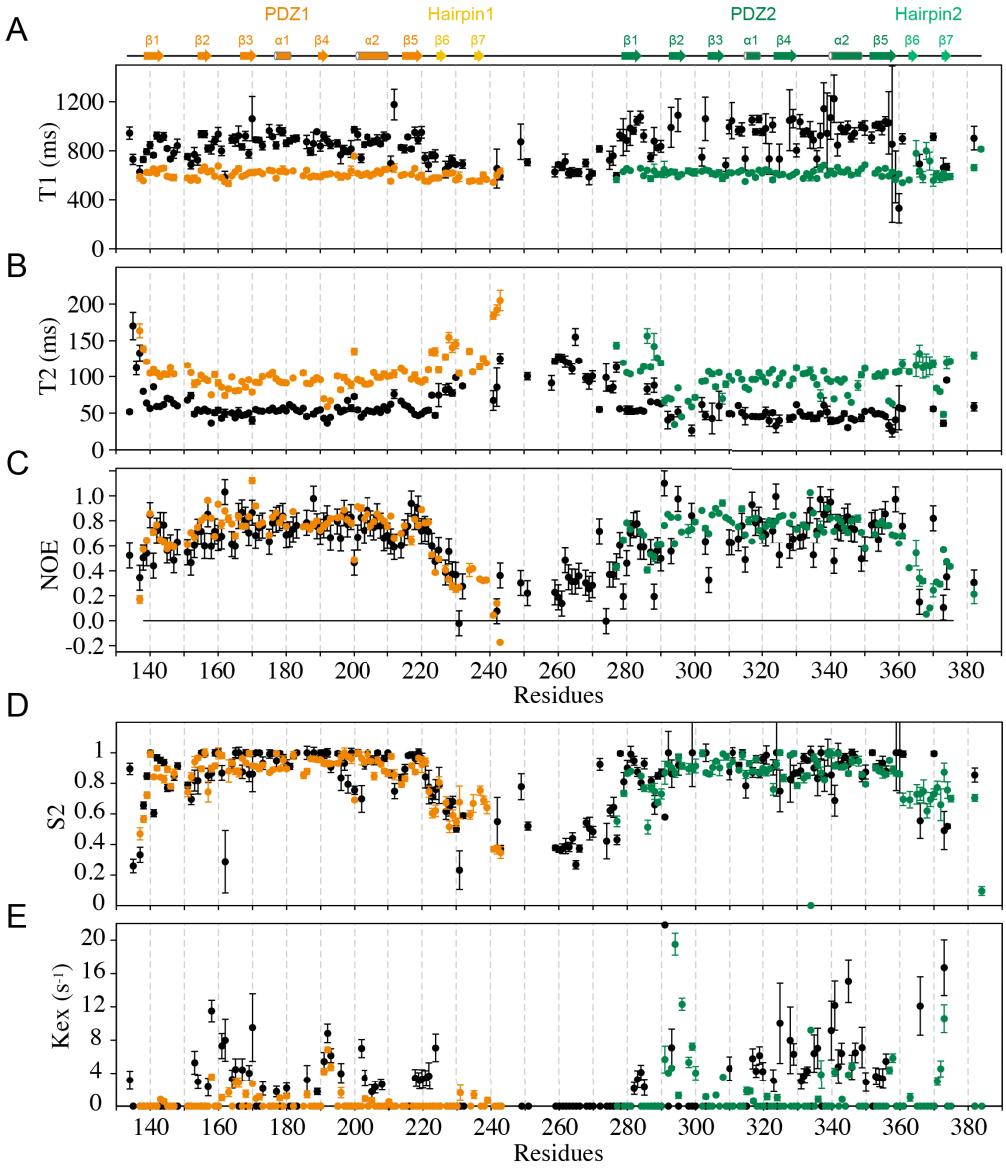
Data acquisition and processing. Quenched samples were rapidly defrosted on ice and immediately injected into a refrigerated nanoACQUITY UPLC system maintained at 0 °C. Samples (7.9 pmols) were on-line digested using an in-house prepared column of immobilized pepsin beads (2.0 x 20 mm, 66 µl bead volume) for 2 min at 100 µl/min mobile phase A (0.15 % formic acid) and 20 °C. Peptides were trapped, concentrated and desalted on a Vanguard C18 pre-column (BEH 1.7 µm, 2.1 x 5 mm, Waters) and eluted from the trap column to an analytical C18 column (nanoACQUITY UPLC CSH C18, 130Å, 1.7 µm, 1 x

100 mm, Waters) with a 7 min linear gradient of 5 % to 45 % mobile phase B (acetonitrile, 0.15 % formic acid) at 40 µl/min. After each run, the pepsin column was cleaned with a solution of 0.8 % formic acid, 5 % acetonitrile, 1.5 M guanidinium chloride, pH 1.7. Blank injections were performed between each run to confirm the absence of carry-over. Mass spectra were acquired in positive mode on a Synapt^{TM} G2-Si HDMS^{TM} mass spectrometer (Waters) equipped with a standard electrospray ionization (ESI) source. Peptides were fragmented in undeuterated samples using a data-independent acquisition scheme (MS^E) and identified by data-base searching in ProteinLynx Global Server 3.0 (Waters). Each fragmentation spectrum was manually inspected for assignment validation. DynamX 3.0 HDX software (Waters) was used to extract the centroid masses of all peptides selected for local HDX-analyses; only one charge state was considered per peptide. HDX-MS results are reported as relative deuterium uptake values expressed in mass unit or fractional uptake. A statistical analysis was carried-out with MEMHDX using a False Discovery Rate of 5% [306].

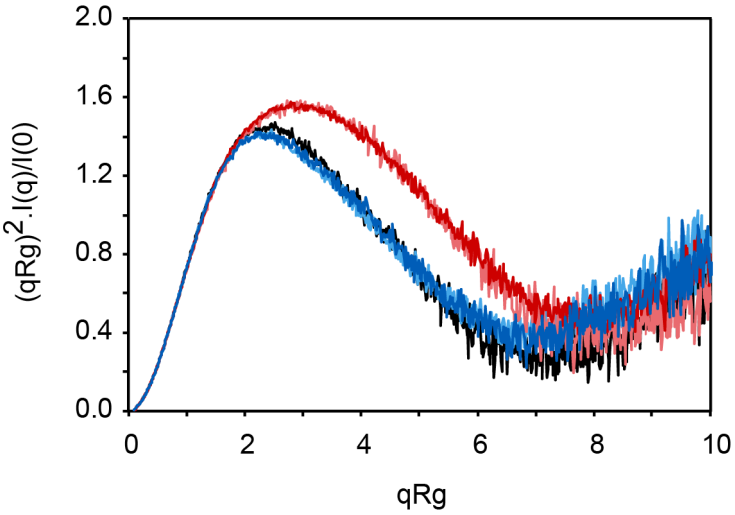
9.3.7 Acknowledgements

The authors are grateful to Bradley Worley for careful proofreading of the manuscript and to Muriel Delepierre for useful discussions. We thank Olivier Lequin, Patrick England and Alain Chaffotte for their technical expertise. Financial support from TGIR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged. We acknowledge SOLEIL and ESRF for provision of synchrotron radiation facilities and we would like to thank the staff of the SWING and BM29 BioSAXS beamlines for assistance during the SAXS measurements. This work was supported by the Programme Transversal de Recherche from the Institut Pasteur (PTR Grant No. 483 to N.W.), the European Union (FP7-IDEAS-ERC 294809 to M.N.), the Ministère de l'Enseignement Supérieur et de la Recherche (Grant No. 883/2013 to F.D) and of the Comité Berthe Fouassier – Maladies de l'oeil de la Fondation de France (No. 00071779 to F.D).

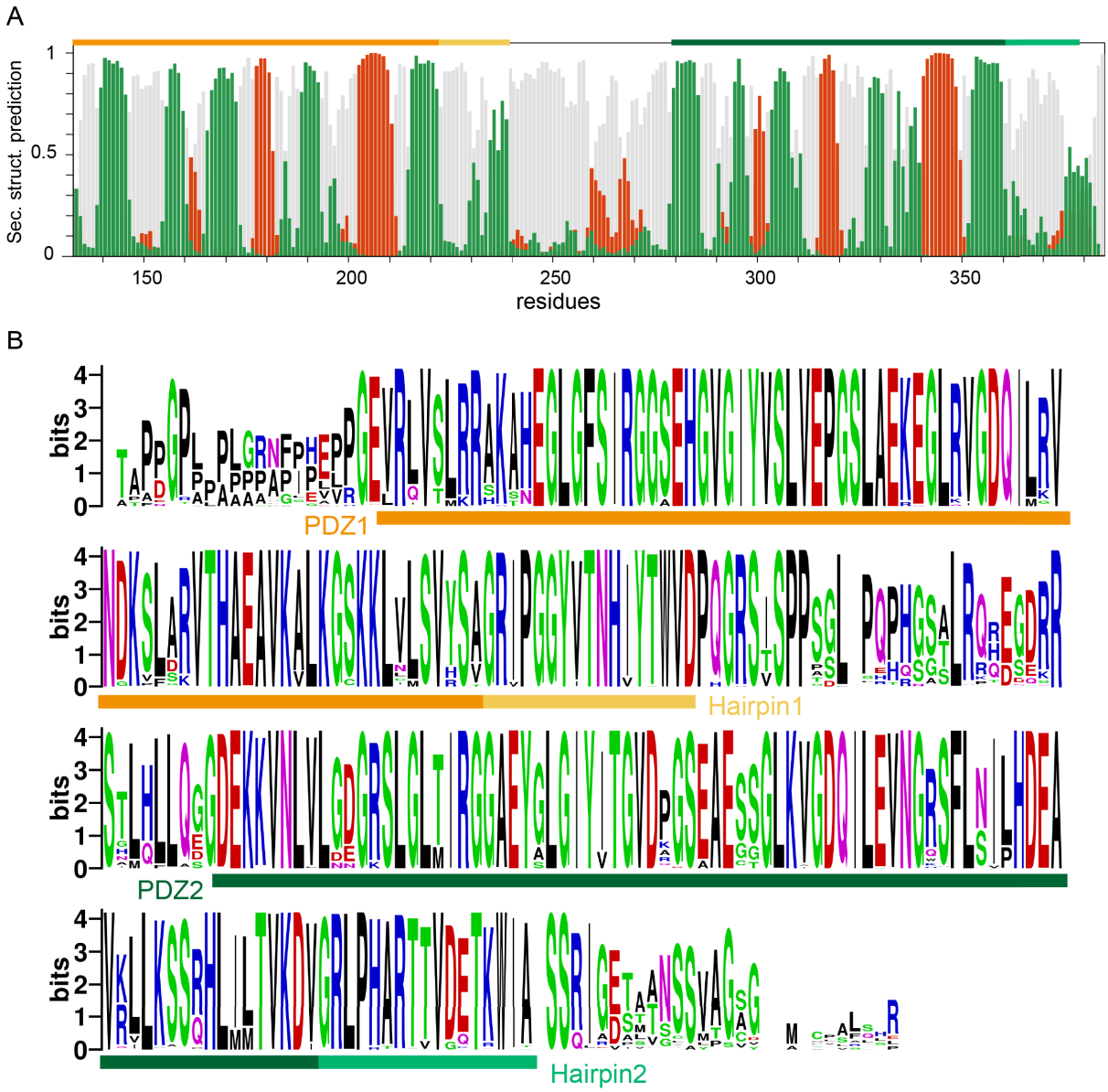
9.3.8 Supplementary figures



Supplementary Figure 1 : Backbone Dynamic of PDZ1-2. ^{15}N relaxation data, T1 **(A),** T2 **(B)** and ^{1}H-^{15}N NOE **(C)** and dynamic parameters using the Lipari-Szabo model-free approach, parameter S^2 (**D**) and chemical exchange R_{ex} (**E**), for PDZ1-2 (black), isolated PDZ1-Hp1 (orange) and isolated PDZ2-Hp2 (green) measured at 25°C at 600MHz.



Supplementary Figure 2 : Dimensionless Kratky plot of PDZ1-2 wild-type and mutants.) **PDZ1-2^{Wt}** in black and mutants: PDZ1-2^{IP225GG} in light blue, PDZ1-2^{W237G} in light red, PDZ1-2^{LP363GG} in dark red and PDZ1-2^{W375G} in dark blue.

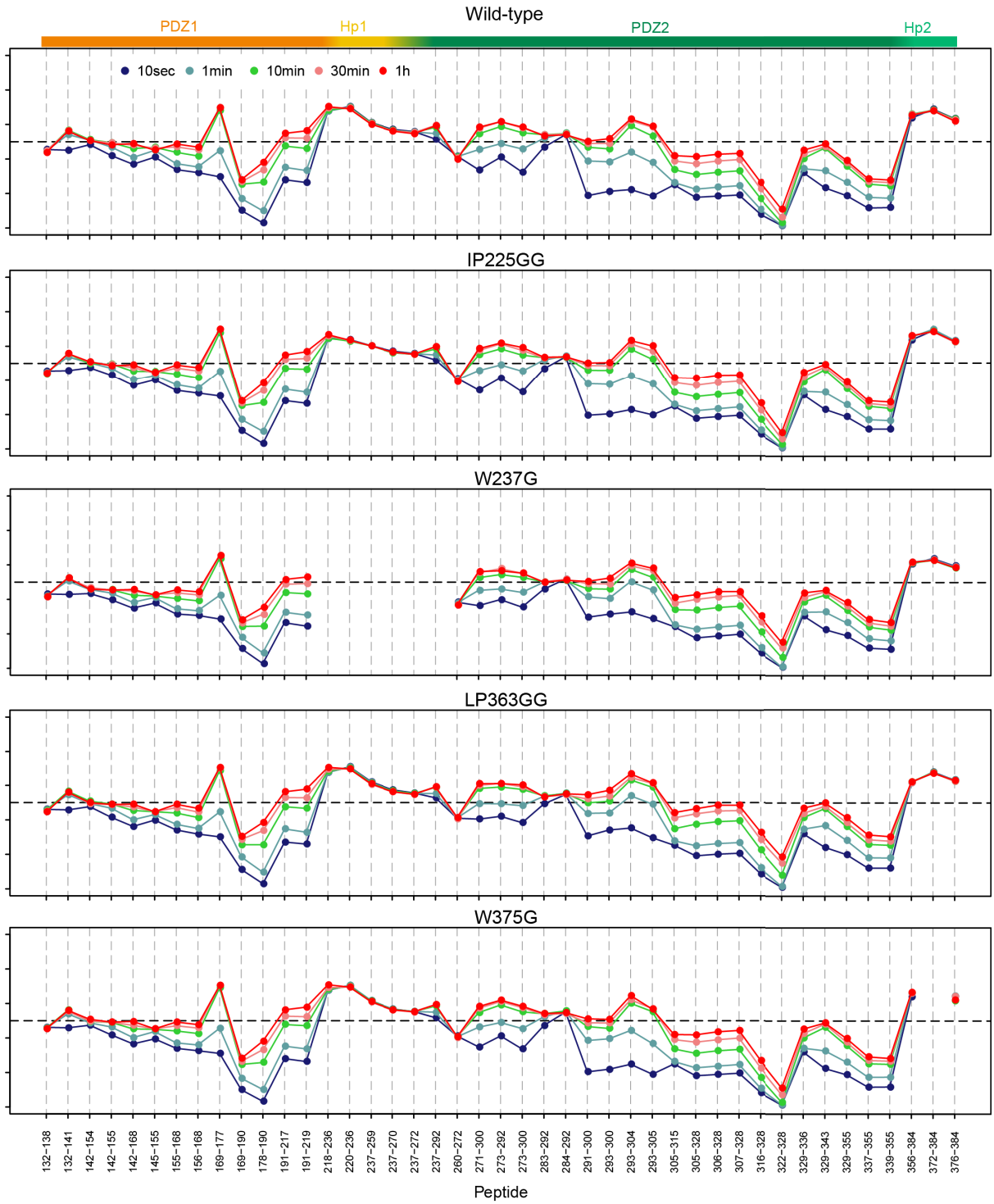


Supplementary Figure 3 : prediction and conservation of hairpin extensions of PDZ1-2. (A) Secondary structure predictions of Whirlin PDZ1-2 predicted from TALOS-N based on the backbone resonance assignments. **(B)** Sequence Logo representation (generated from http://weblogo.berkeley.edu) of 215 aligned sequences of PDZ1-2 Whirlin.

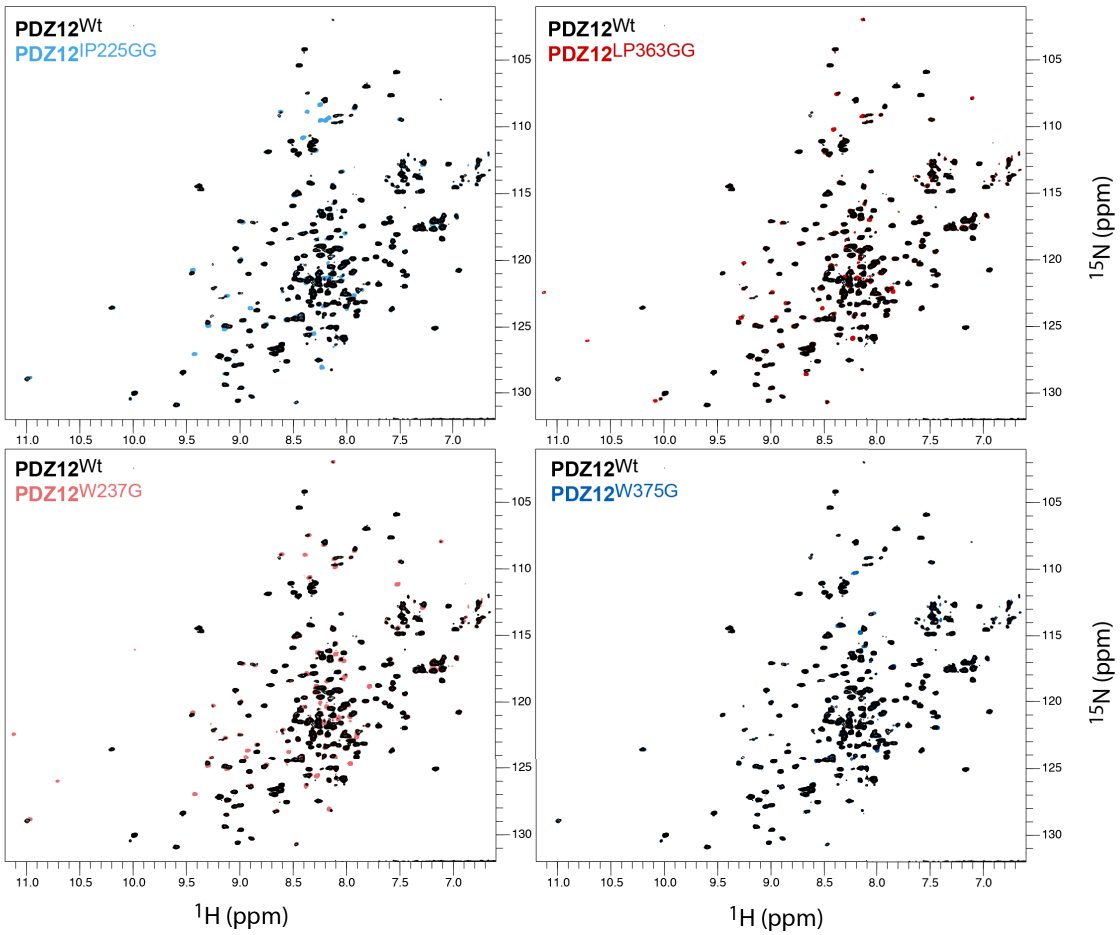
Restraints used	Population[a]	RMSD[b]	RMSD vs. PDZ1-2 model[c]	$Q^{RDC(phages)}$	$Q^{RDC(et./alc.)}$	Q^{PRE}
RDC_{phages} + $RDC_{et./alc.}$ + $PRE_{162,231,290,371}$	72 %	1.85 Å	-	0.24	0.37	0.35
$RDC_{et./alc.}$ + $PRE_{162,231,290,371}$	59 %	1.61 Å	0.71 Å	0.71	0.32	0.34
RDC_{phages} + $PRE_{162,231,290,371}$	69 %	2.10 Å	1.13 Å	0.23	0.73	0.34
$PRE_{162,231,290,371}$	55 %	2.10 Å	2.74 Å	0.62	0.78	0.34
RDC_{phages} + $RDC_{et./alc.}$ + $PRE_{231,290,371}$	67 %	1.78 Å	0.78 Å	0.24	0.37	0.36
RDC_{phages} + $RDC_{et./alc.}$ + $PRE_{162,290,371}$	64 %	1.74 Å	0.86 Å	0.23	0.37	0.37
RDC_{phages} + $RDC_{et./alc.}$ + $PRE_{162,231,371}$	63 %	1.69 Å	0.72 Å	0.24	0.37	0.35
RDC_{phages} + $RDC_{et./alc.}$ + $PRE_{162,231,290}$	74 %	2.25 Å	3.91 Å	0.23	0.33	0.38
RDC_{phages} + $RDC_{et./alc.}$ + $PRE_{162,231,290,371}$[d]	59 %	1.57 Å	1.71 Å	0.23	0.37	0.35

a Percentage of conformations belonging to the most populated cluster (among 500) .
b Average Root Mean Square Deviation (RMSD) of the 25 best-energy conformers in the cluster for backbone atoms.
c RMSD between the average conformation of the 25 best-energy conformers in the cluster and the PDZ1-2 model when superimposing on Cα atoms of PDZ1 and PDZ2.
d Experimental error for measured I_{ox}/I_{red} ratios assumed to 0.3 (instead of 0.2).

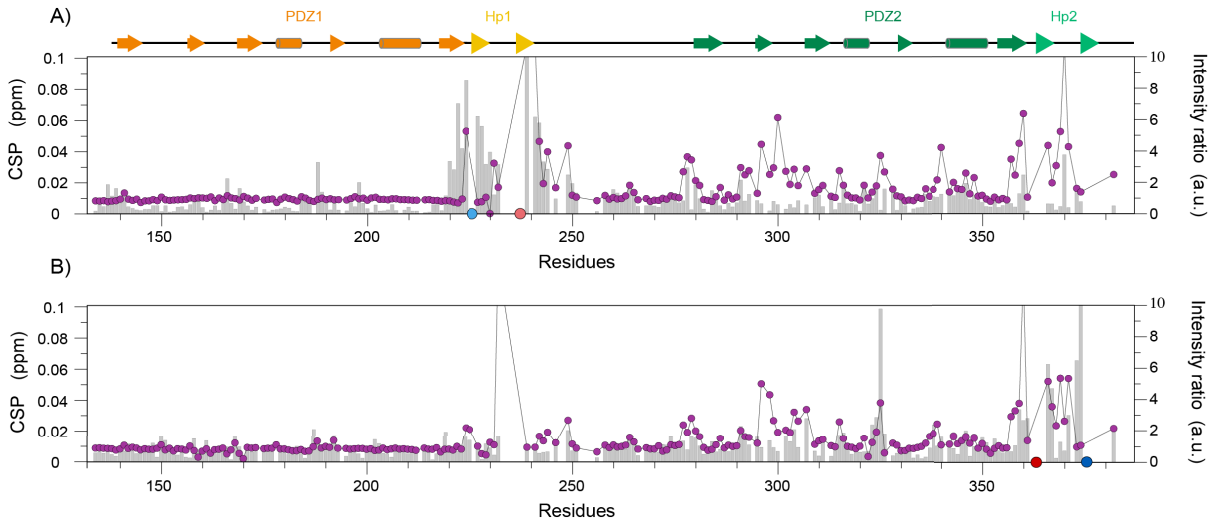
Supplementary Figure 4 : Statistics of the most populated clusters obtained with different restraint sets for PDZ1-2 structure calculations with HADDOCK



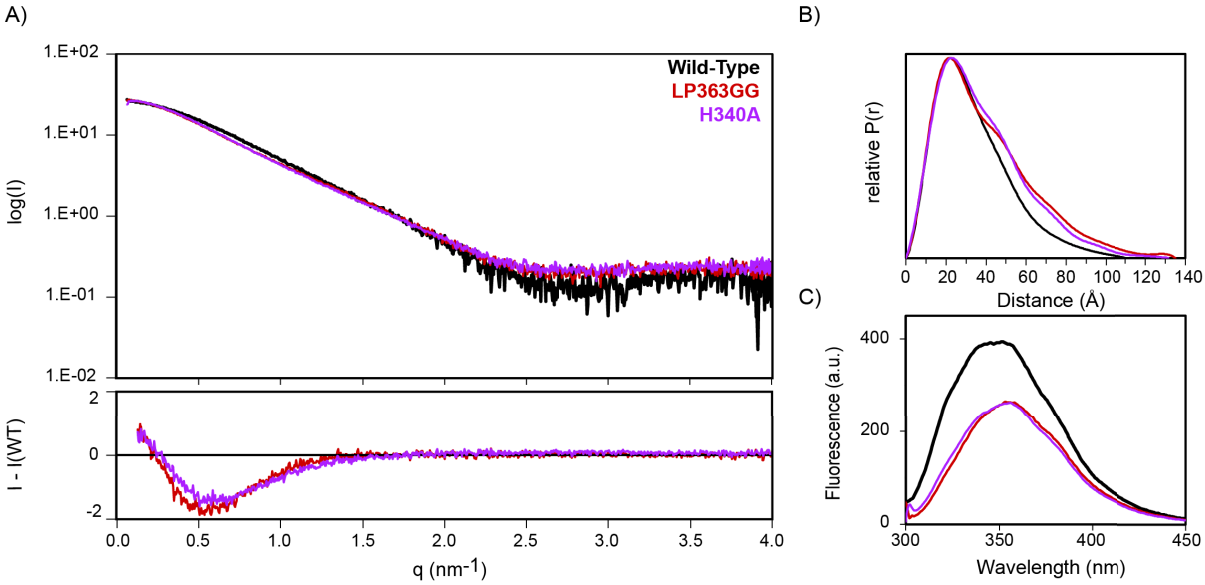
Supplementary Figure 5 : Deuterium Uptake of PDZ1-2 wild-type and mutants. Triplicate average of fractional uptake of deuterium between the hairpin mutants and wild-type PDZ1-2 monitored by mass spectrometry after increasing duration of incubation (10, 60, 600, 1800 and 3600 seconds).



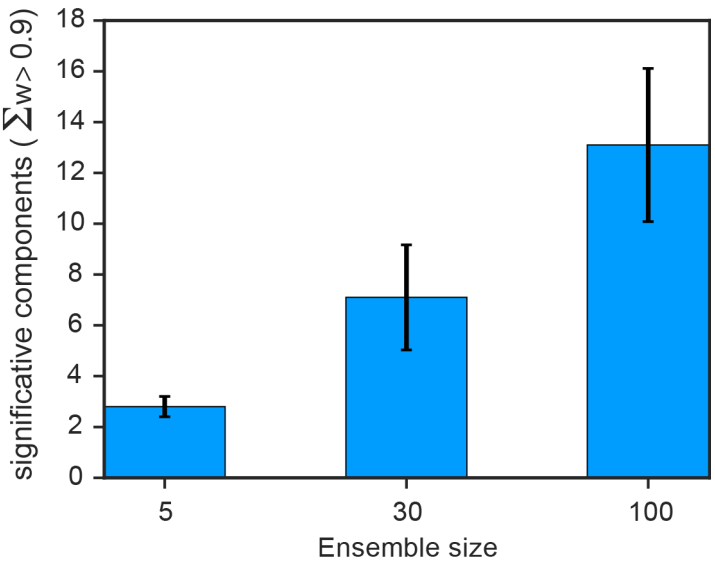
Supplementary Figure 6 : ^{1}H-^{15}N HSQC spectra superposition of PDZ1-2 wild-type with the mutants. Spectra recorded at 25°C at a ^{1}H frequency of 600MHz.



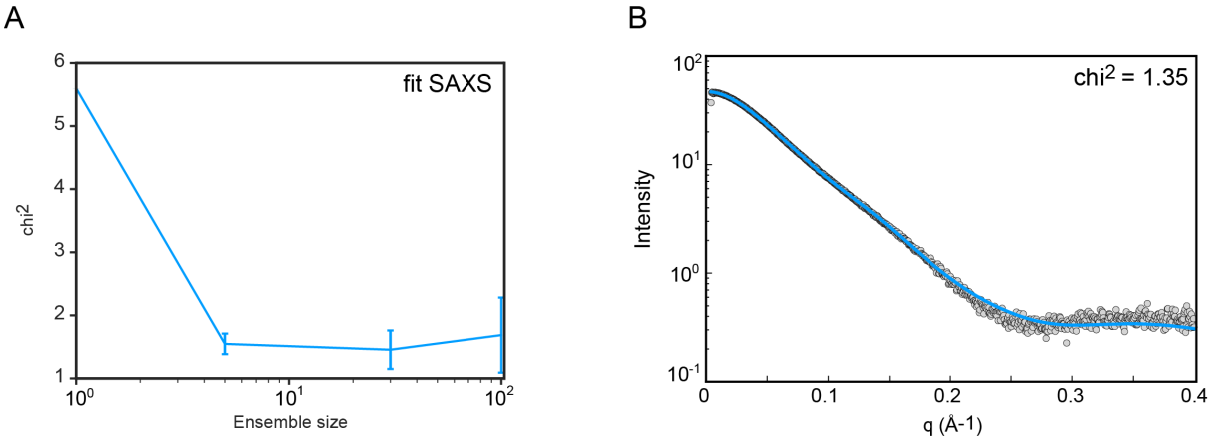
Supplementary Figure 7 : CSP and intensity ratio between mutants of same hairpin. (A) CSP and intensity ratio between mutants of the first hairpin PDZ1-2^{IP225GG} and PDZ1-2^{W237G}. **(B)** CSP and intensity ratio between mutant of the second hairpin PDZ1-2^{W375G} and PDZ1-2^{LP363GG}.



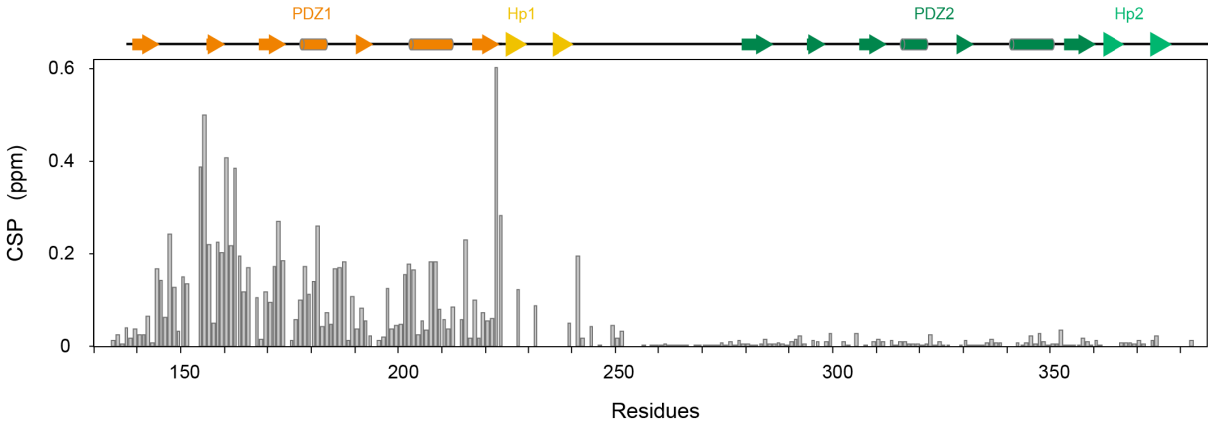
Supplementary Figure 8 : Effect of H340A mutation on the overall conformation of PDZ1-2. **(A)** Scattering curves of PDZ1-2^{H340A} (violet), PDZ1-2^{Wt} (black) and PDZ1-2^{LP363GG} (dark red) for comparison (Top panel) and residual between the scattering curves of PDZ1-2 Wild-type and PDZ1-2 mutants (bottom panel). **(B)** Distance distribution function extracted from SAXS data of PDZ1-2^{Wt}, PDZ1-2^{H340A} and PDZ1-2^{LP363GG} for comparison **(E)** Fluorescence emission spectra measured for PDZ1-2^{Wt}, PDZ1-2^{H340A} and PDZ1-2^{LP363GG}, monitoring the local environment of the two tryptophans located in the hairpin extensions of each PDZ domain.



Supplementary Figure 9 : Count of unique conformations for each ensemble size in the combined fit. Average number of unique conformations accounting for 90% of the total weight of the ensemble. Error bars are the standard deviation around the average for the 1000 ensembles generated.



Supplementary Figure 10 : ensemble selection on PDZ1-2^{LP363GG}. (A) Fit improvement in chi^2 according to the ensemble size in the genetic algorithm selection for the SAXS data. The error bars are computed over 10 individuals run. (**B**) Back-calculated SAXS profile derived from the best fitting ensemble (blue) and the experimental profile of PDZ1-2^{LP363GG} (black)



Supplementary Figure 11: CSP induced by Sans peptide binding on PDZ1-2^{Wt}. CSP between spectra of PDZ1-2^{WT} before and after addition of 800uM of Sans peptide.

9.4 Commentaire & expériences complémentaires

Avec ces deux articles, nous avons déterminé la structure d'un supramodule formé de façon transitoire entre deux domaines PDZ. A ce jour, seules six structures de supramodule homotypique de domaines PDZ ont été résolues, une faible proportion comparée à leur occurrence dans le protéome (42 tandems en considérant un lien inter-domaine inférieur à 25 acides aminés). Toutes ces structures ont été obtenues par cristallographie, favorable à la détermination des objets les plus rigides. L'étude de complexe transitoire est difficile d'un point de vue expérimental et demande l'association de données de plusieurs méthodes biophysiques, intégrées par modélisation. Il est probable qu'une partie des tandems de domaine PDZ forme des supramodules similaires à ce que nous avons présenté et dont le caractère transitoire nécessite une approche intégrative pour la détermination de leur conformation.

9.4.1 Difficultés rencontrées avec l'étude des épingles-ß.

L'étude du tandem PDZ1-2 s'est avérée plus compliquée qu'initialement prévue par la présence des extensions en C-terminale de chaque domaine. Ces extensions se localisent à l'interface entre domaines PDZ et ne se structurent que dans le tandem et non dans les domaines isolés. La dynamique d'échange conformationnelle des épingles-ß aura rendu leur attribution difficile dans la construction du tandem, notre couverture d'attributions pour ces régions est bien moindre que dans le reste des domaines.

La structuration de l'épingle-ß du PDZ1 est fluctuante. La conformation de celle-ci est différente dans le tandem et dans le PDZ1 isolé (émission de fluorescence) ; elle n'est cependant pas affectée par la conformation ouverte ou fermée du tandem (HSQC). Son repliement induit de l'échange lent sur les résidus du PDZ1 qui sont à son contact (**HSQC ; appendice 12.2**).

Le comportement de l'épingle-ß du PDZ2 est opposé à celui de l'épingle-ß du PDZ1. Sa présence induit de l'échange intermédiaire à la surface du PDZ2 dans le tandem. Par ailleurs, les mutations affectant la conformation du tandem ont un effet direct sur la dynamique conformationnelle de cette épingle-ß (**HSQC ; appendice 12.2**). Son tryptophane conservé (W375) est complétement exposé, et la mutation W375G n'affecte pas l'équilibre ouvert/fermé du tandem alors que la suppression de son second brin (PDZ1-2$^{\Delta 15}$) ou des résidus conservés du premier brin (PDZ1-2^{LP363GG}) supprime l'interaction entre domaines.

L'absence d'interaction forte et directe entre domaine a rendu extrêmement difficile l'identification des perturbations dues au changement de conformation des épingles-ß ou au changement global de conformation du tandem. Nous avons ainsi dû multiplier les mutants et les expériences pour comprendre les modes de repliement de ces épingles-ß avant de pouvoir conclure sur les résultats présentés dans le second article sur l'interface entre PDZ.

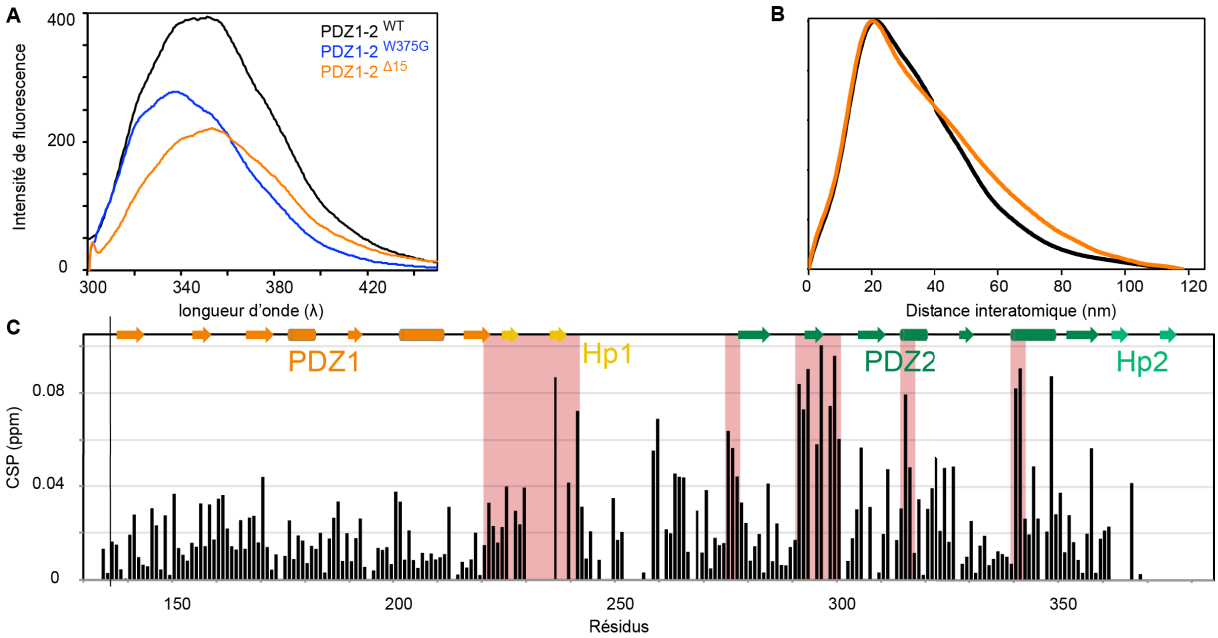


Figure 65: Effet de la mutation PDZ1-2$^{\Delta 15}$ sur la conformation du tandem PDZ1-2. (A) Emission de fluorescence de trois constructions de PDZ1-2 sauvage et mutantes. Le maximum d'émission de fluorescence du W237 dans la construction PDZ1-2 $^{\Delta 15}$ est à 360 nm, son environnement est donc différent de celui dans la conformation native. **(B)** Distribution de distances interatomiques normalisées. La contribution de longues distances est plus importante dans le mutant court PDZ1-2 $^{\Delta 15}$ que dans le sauvage. **(C)** CSP entre PDZ1-2 $^{\Delta 15}$ et PDZ1-2 ^{WT}. Les plus fortes perturbations sont observées sur le PDZ2, l'épingle-ß du PDZ1 est aussi particulièrement affectée.

9.4.2 Rôle fonctionnel du tandem

Une question à laquelle nous tentons de répondre dans cet article est le rôle du tandem PDZ1-2 sur la fonction des domaines PDZ le composant. La présence de l'épingle-ß du PDZ1 améliore l'affinité du PDZ1 pour son ligand SANS d'un facteur 3-4. Cet effet significatif observé est un premier pas vers la compréhension du rôle du tandem sur la fonction de la protéine, mais nous ne sommes pas en mesure de conclure si cette augmentation modeste est déterminante dans la fonction physiologique du domaine.

Nous n'avons pas pu identifier de partenaires du PDZ2 sur l'ensemble des PBM de l'audition, nous laissant supposer que ce domaine est inactif, jusqu'à preuve du contraire. D'un point de vue structural, la chaine latérale d'une lysine de l'hélice α2 traverse le sillon

d'interaction du domaine PDZ2, empêchant physiquement l'interaction avec un partenaire. Dans la moitié des tandems de PDZ dont la structure a été déterminée, l'un des deux domaines n'a pas de fonction d'interaction mais permet une régulation de la fonction de l'autre domaine [50,148,149]. Il est probable que PDZ2 ait lui-même un effet modulateur inductible sur la fonction du PDZ1, que nous n'avons pas pu identifier à ce jour. Pour étudier le rôle potentiel de l'équilibre ouvert/fermé sur la fonction de la Whirline, nous avons fait varier les conditions de solvant, dans le tube RMN. Les stéréocils des cellules ciliées sont un environnement particulièrement encombré, nous avons donc testé l'effet de l'encombrement stérique par ajout de ficoll (200g/l), sur les spectres HSQC du tandem, sans effet. Nous avons aussi testé l'effet de charge en variant la concentration de sel et le pH, cependant aucune de ces modifications du milieu n'a eu d'effet significatif sur l'équilibre ouvert/fermé du tandem.

9.4.3 Rôle de la séquence inter-domaine

La séquence connectant les deux domaines devrait être étudiée plus en détail. En effet, la structuration de l'épingle-ß du PDZ1, même dans le cas d'un mutant tandem ouvert, pourrait être une indication d'un rôle stabilisateur du lien. La prédiction de structure secondaire basée sur les résonances RMN, TALOS-N donne une faible prédiction d'hélice dans la partie C-terminale du lien, entre les résidus 260 et 270 ; cette région est aussi conservée dans les orthologues de Whirline. Des mutations, ponctuelles ou troncatures de cette région conservée, pourraient être envisagées pour évaluer son importance sur la conformation du supramodule. On peut notamment penser à substituer les résidus L259/R260 ou R267/S268 très bien conservés chez les orthologues afin de déstabiliser cette structure.

9.4.4 Contexte de la protéine entière

Une des questions récurrentes que nous nous posons concerne la relevance des observations faites sur ce tandem isolé dans le contexte de constructions plus grandes incluant notamment les domaines HHD adjacents.

Le travail présenté dans le chapitre précédent, sur le domaine HHD2, permet de répondre partiellement à cette question. Le domaine HHD2 est indépendant du tandem PDZ1-2. Une construction contenant les trois domaines ne présente pas de perturbations, ni sur les résonances des atomes du tandem PDZ1-2, ni sur les résonances des atomes du domaine HHD2, la conformation du tandem est donc très comparable dans le tri-domaine. D'un point

de vue fonctionnel, nous devrions mesurer l'affinité de la construction contenant les trois domaines pour l'un de ses partenaires et la comparer avec une valeur connue pour le tandem.

L'effet du HHD1 sur la structure du tandem est plus difficile à évaluer d'un point de vue expérimental. Comme décrit précédemment, celui-ci induit l'oligomérisation de toutes les constructions multi-domaines le contenant. Nous n'avons pas trouvé de conditions expérimentales satisfaisantes afin d'évaluer son effet sur la conformation du tandem PDZ1-2. Comme nous le présentons dans la figure de conclusion du second article, la superposition des structures du supramodule homotypique PDZ1-2 de la Whirline et du supramodule hétérotypique HHD-PDZ1 de l'Harmonine ne crée pas d'incompatibilité stérique entre le PDZ2 et le HHD. Il est donc possible qu'un supramodule de trois domaines se forme entre le HHD1 et le PDZ1-2 de la Whirline. Dans ce cas, l'équilibre conformationnel du tandem PDZ serait modifié.

Deux approches pourraient être entreprises pour étudier ce dernier point. Le suivi de la fluorescence des tryptophanes du tandem dans une construction plus large pourrait donner une première indication, mais la présence d'un tryptophane très certainement exposé dans le lien entre HHD1 et le PDZ1 pourrait compliquer l'interprétation des résultats. Nous envisageons également la production d'une construction HHD1-PDZ1-2. Si la conformation supramodulaire à trois domaines se confirme, il est probable que celle-ci rigidifie l'assemblage ainsi formé, qui pourrait avoir une plus grande chance de cristalliser. Sachant toutes les difficultés systématiquement rencontrées avec les constructions incluant HHD1, ce projet demandera encore beaucoup de mises au point biochimiques.

10 Résultats 4 : Interactions de la Whirline

10.1 Criblage des interactions des domaines de la Whirline

La plus grande partie du travail de cette thèse se concentre sur la détermination structurale de la Whirline, et notamment autour des propriétés émanant de son organisation modulaire, comme nous l'avons présenté dans les chapitres précédents. Nous nous sommes aussi intéressés au réseau d'interaction de la Whirline avec les protéines Usher et Usher associées. De nombreux interactants ont déjà été décrits pour la Whirline, mais la très grande majorité des partenaires est identifiée par expériences de pull-down et de co-immunoprécipitation en système hétérologue, avec une surexpression des protéines cibles. Ces approches ont un fort taux de faux-positifs, d'autant que les constructions utilisées possèdent bien souvent des domaines mal délimités.

Pour clarifier le réseau d'interaction *in vitro* de la Whirline, nous avons utilisé plusieurs techniques de criblage à moyen et haut débit des affinités des différentes constructions de la Whirline pour les peptides C-terminaux des protéines Usher. Nous avons par ailleurs cherché à comparer les capacités de fixation de nos différentes constructions de la Whirline. Les peptides PBM utilisés correspondent aux 11 acides aminés C-terminaux des protéines Usher. Les peptides sont synthétisés avec une étiquette biotine en N-terminal, séparée de la séquence exogène par un espaceur synthétique généralement constitué d'un triplet d'un motif polyéthylène glycol (PEG).

10.1.1 Résonance Plasmonique de Surface

10.1.1.1 Implémentation

La première approche que nous avons utilisée est la Résonance Plasmonique de Surface (SPR) sur un appareil ProteOn (Bio-Rad). La SPR repose sur la modification de l'indice de réfraction d'une surface, lors du changement de sa composition. Ainsi, l'un des composés de l'interaction est immobilisé à la surface d'une phase solide et son partenaire est contenu dans une phase liquide au contact de la surface. Si une interaction a lieu, le composé contenu dans la phase liquide s'adsorbe à la surface de la phase solide et modifie l'indice de réfraction de celle-ci, permettant ainsi le titrage en faisant varier les concentrations en ligand. Les phases d'association et de dissociation peuvent aussi être étudiées.

Dans notre cas, le peptide est fixé à la surface par son étiquette biotine N-terminale et les constructions de la Whirline sont contenues dans la phase liquide. Ce protocole a plusieurs avantages ; d'une part les constructions ont des masses moléculaires nettement supérieures à celle du peptide, leur fixation donnera donc un signal bien supérieur que si le peptide était contenu dans la phase liquide. D'autre part, en nettoyant correctement la surface après interaction, il est possible de mesurer, dans une même expérience, l'affinité d'une dizaine de constructions de protéine pour un peptide donné en modifiant simplement la phase liquide. Cette approche permet de titrer simultanément l'interaction de 5 peptides avec une dizaine de construction de la Whirline, un format bien adapté à nos objectifs de comparer les capacités de fixation des différentes constructions de domaines. Nous avons ainsi déterminé l'affinité par titrage à l'équilibre, avec 10 points par courbe, obtenus par dilution en série des constructions. Les expériences ont été effectuées une seule fois, les erreurs indiquées correspondent donc à l'erreur d'ajustage et non à plusieurs expériences indépendantes. Ces travaux ont été initiés par nos collaborateurs de l'Unité de physiologie et génétique de l'Audition, Elise Peppermans et Amel Bahloul, qui ont montré certaines de ces interactions. Cependant, tous les résultats présentés ici ont été obtenus par mes expériences.

Le ProteOn utilisé dans cette étude a une précision moindre que celle d'autres techniques de titrage en solution comme l'ITC, l'émission de fluorescence ou le suivi de perturbation de résonances par RMN. Une incertitude autour de la constante de dissociation est à garder à l'esprit, particulièrement dans notre cas où les expériences ne sont pas effectuées en triplicate. Les affinités mesurées sont donc ici considérées comme des ordres de grandeur plus que des mesures exactes. Avec nos constructions, nous avons systématiquement observé que les constantes d'affinités déterminées sur cette machine sont supérieures aux valeurs mesurées par techniques de titrage en solution, notamment RMN et émission de fluorescence. D'un part l'immobilisation d'un des partenaires sur une surface peut augmenter l'affinité par modification de l'entropie conformationnelle, d'autre part, l'environnement encombré à l'interface peut favoriser des interactions.

10.1.1.2 Résultats

Les affinités obtenues par SPR sont présentées dans le tableau 5. Tout d'abord, le PDZ2 de la Whirline n'interagit avec aucun peptide C-terminal des protéines Usher testées, même lorsque le domaine HHD2 est ajouté à la construction (Table 4). Bien que nous n'ayons testé que 13 peptides par cette technique, il semble probable que le PDZ2 de la Whirline ne soit pas capable de lier un partenaire de façon canonique. Dans la structure du PDZ2 de la

Whirline humaine, son sillon d'interaction est obturé au niveau de la boucle "GLGF" par la chaine latérale d'une lysine de l'hélice α2, empêchant l'interaction avec un partenaire canonique dans cette conformation. L'absence d'interaction du PDZ2 *in vitro* est en contradiction avec les résultats d'un article décrivant l'interaction du PBM de l'Usherine avec le PDZ2 de la Whirline, déterminée par double hybride, sans que les résultats ne soient cependant montrés [236]. Plusieurs autres articles décrivent l'interaction du PDZ2 avec le C-terminal de partenaires tels que Cav1,3 [241] , LRRC4C [219] et RPGR^{orf15} [222] (PDZ2 tronqué de son dernier brin et partenaire sans PBM) ; dans tous les articles décrivant une interaction avec le PDZ2, ces partenaires interagissent aussi avec le PDZ1. Pour réellement conclure sur la non-fonctionnalité du PDZ2 de la Whirline, il conviendrait de tester son affinité avec les peptides Cav1,3 et de LRRC4C.

		Kd (µM)					
Peptide	Séquence	PDZ1	PDZ1H	PDZ1-2	PDZ2H	P2-H2	PDZ3
Pcdh15-CD1	DVQPHSQ**STSL**	**130**	**35**	ND	>1000	>1000	>1000
Pcdh15-CD2	GANSEGY**NTAL**	**160**	**45**	ND	>1000	>1000	>1000
Pcdh15-CD3	WNSLHIP**MTKL**	**45**	**10**	ND	>1000	>1000	**250**
Cdh23	IMESPLE**ITEL**	**190**	**60**	ND	>1000	>1000	**5**
SANS	ERPLALE**DTEL**	**80**	**20**	**20**	>1000	>1000	>1000
Hrmn	PKEYDDE**LTFF**	ND	>1000	ND	>1000	>1000	**50**
CIB2	PDFLSTF**HIRI**	ND	>1000	>1000	>1000	>1000	>1000
Vlgr1	KEHTAFT**DTHL**	ND	**15**	**5**	>1000	>1000	>1000
Ush	LRRIPIA**DTHL**	ND	**40**	**20**	>1000	>1000	>1000
Whrn	DFLVTEF**NVML**	ND	>1000	ND	>1000	>1000	>1000
Myo15a	LTLPPSE**ITLL**	ND	>1000	ND	>1000	>1000	**5**
Clarine1	ETTNVAA**DLMY**	ND	>1000	>1000	>1000	>1000	>1000
TMHS	DDYKADG**TEEV**	ND	>1000	>1000	>1000	>1000	>1000

Table 4 : Affinités déterminées par SPR des domaines de la Whirline pour les peptides Usher. La séquence du peptide est indiquée avec les résidus canoniques du motif colorés en bleu. Les affinités supérieures à 100 µM sont indiquées en noir, en orange pour un Kd compris entre 20 et 100 µM et en vert pour un Kd inférieur ou égal à 20 µM, >1000 indique qu'aucun changement de réfraction n'a été observé ; ND = no data.

Le PDZ3 de la Whirline interagit avec un nombre de PBM restreint, mais avec une bonne affinité. Ce PDZ a une forte préférence pour les PBM de type 1 possédant un hydrophobe en position (-3), présent chez ses quatre partenaires identifiés. L'importance de cette position est soutenue par la différence drastique d'affinité du PDZ3 pour le PBM de la Cadhérine-23 (…LEITEL) et le PBM de SANS (…LEDTEL). Le PDZ3 lie la cadhérine-23 avec une constante de dissociation de l'ordre du micromolaire mais est incapable d'interagir avec SANS, dans la gamme de concentration étudiée. Il est difficile d'évaluer la préférence du domaine pour les positions (0) et (-2) à cause de la surreprésentation de motif T.X.L dans l'audition. La Myosine-15a est l'interactant le plus décrit du PDZ3 et a été montrée comme interagissant par son PBM avec la forme courte de la Whirline [227]. La colocalisation des deux protéines au sommet de la grande rangée de stéréocils a été plusieurs fois décrite. Nos résultats confirment leur interaction PDZ/PBM avec une affinité micromolaire. Les interactions observées avec la Cadhérine-23 et la Protocadhérine-15-CD3 sont plus surprenantes : celles-ci n'ont jamais été décrites et leur colocalisation avec la Whirline n'a pas été montrée. L'interaction avec la Cadherine-23 pourrait être due à la ressemblance de son PBM (…EITEL) avec celui de la myosine-15a (…EITLL), partenaire identifié du PDZ3 de la Whirline. Enfin, l'interaction avec l'Harmonine est intéressante, au moins d'un point de vue structural : le PBM est original, avec une Phénylalanine en position (0). Du point de vue biologique, cette interaction pourrait être un indice de la dimérisation hétérotypique de ces deux paralogues, dont la dimérisation homotypique est suspectée être importante pour leurs fonctions respectives.

Le PDZ1 de la Whirline interagit avec les PBM de toutes les protéines membranaires Ush1 et Ush2 ainsi que de la protéine d'échafaudage SANS. Il est tolérant aux variations des positions (-1) et (-3), cependant ses interactions avec la Cadhérine-23 et SANS, mais pas avec la Myosine-15a, illustrent une préférence pour les résidus polaires en position (-1). La présence d'un résidu hydrophobe en position (-5) semble aussi favoriser l'interaction sans pour autant être strictement définie. Tout comme pour le PDZ3, l'interaction du PDZ1 de la Whirline avec les protéines membranaires Ush1 est surprenante ; il n'y a pas de colocalisation décrite de ces protéines avec la Whirline, au stade mature et pendant le développement des touffes ciliaires. L'interaction avec les trois PBM très différents des isoformes principales de la Protocadhérine-15 fait suspecter un rôle physiologique de ces complexes, notamment pendant le développement, où toutes les isoformes de la Protocadhérine-15 sont exprimées. L'interaction de la Whirline avec les protéines membranaires Ush2 est bien décrite, les

protéines colocalisent ensemble pendant le développement des stéréocils, dans les liens basaux. Il est décrit que la Whirline, mais aussi PDZD7, interagissent toutes deux avec les parties C-terminales cytoplasmiques de l'Usherine et de Vlgr1. Il a cependant été proposé, par des études de compétitions en pull-down, que dans le contexte des liens basaux, la Whirline interagit préférentiellement avec l'Usherine et PDZ7 préférentiellement avec Vlgr1 [214]. Nos résultats indiquent cependant que la Whirline isolée est 2 à 4 fois plus affine pour Vlgr1 que pour l'Usherine. Il faudra comparer les affinités de la Whirline avec celles mesurées pour les domaines PDZ de PDZD7 pour pouvoir vérifier la composition du complexe Ush2. L'ensemble de ces résultats est présenté dans la figure 66.

Les affinités du PDZ1 sont systématiquement améliorées par l'ajout de structure au PDZ1 de la Whirline. L'ajout des 25 résidus de l'épingle-ß augmente globalement d'un facteur 4 l'affinité du domaine pour ses partenaires en participant directement à l'interaction. L'ajout du PDZ2 dans la construction améliore plus modérément les affinités mesurées.

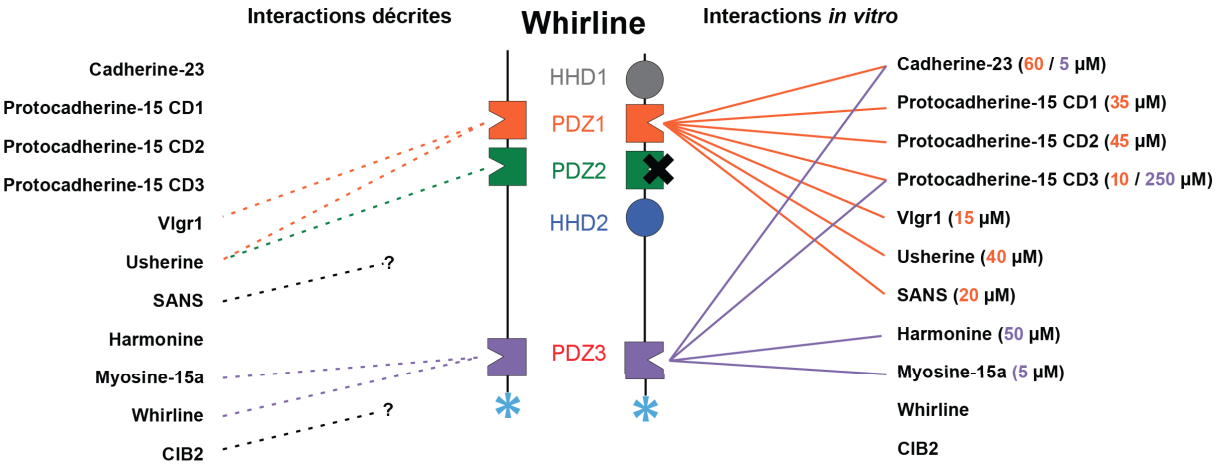


Figure 66 : Comparaison des interactions décrites dans la littérature avec celles observées *in vitro* en SPR. Entres parenthèses sont indiquées les affinités données pour le PDZ concerné. Les valeurs données pour le PDZ1 correspondent à celles mesurées pour PDZ1-Hp.

10.1.2 Hold-Up

10.1.2.1 Implémentation

La méthode hold-up a été développée par nos collaborateurs Gilles Travé et Renaud Vincentelli. Elle permet de mesurer l'interaction d'un PBM pour tous les domaines PDZ du protéome humain, de façon quantitative, permettant ainsi d'établir un profil complet une seule expérience [307]. Cette technique est basée sur une approche chromatographique traitée par robotique. Le peptide biotinylé du C-terminal de la protéine d'intérêt est fixé sur une résine qui est incubée (40 µM), sur microplaque 384 puits, avec les lysats bactériens contenant les

266 domaines PDZ dilués à 4 µM. Après incubation, la solution est filtrée et analysée par électrophorèse quantitative en micro-fluidique. Si le domaine PDZ en solution interagit avec le peptide, celui-ci restera fixé aux billes et ne sera pas contenu dans le filtrat. La proportion de domaine retenu dépend directement de la constante de dissociation entre le domaine PDZ et le peptide. L'interaction est mesurée par comparaison de la quantité de protéine non-retenue sur les billes couplées au peptide d'intérêt avec la quantité non-retenue sur une référence où les billes sont recouvertes de biotine sans peptide. L'interaction est donc faite dans les conditions similaires à la SPR, avec un peptide immobilisé sur une surface, ici des billes, et le domaine PDZ en solution.

Après analyse, le résultat est indiqué en fraction de l'intensité de liaison (Binding Intensity, BI) par rapport au puits de référence pour chaque PDZ. Une valeur de BI de 1 indique que la totalité du PDZ est lié au PBM. L'ordre de grandeur de la constante de dissociation peut être estimé directement à partir de la valeur du BI. Nous avons complété la banque des PDZ isolés (259 domaines exprimés et solubles sur les 266, dans la dernière banque produite) avec un ajout complémentaire de quelques constructions multidomaines ou possédant des extensions du système de l'audition, incluant les formes avec épingles-ß des PDZ de la Whirline, ainsi que le tandem PDZ1-2.

La mise en place de cette technique à l'Institut Pasteur a largement été effectuée par Virginie Girault, ingénieure au laboratoire. J'ai participé à la phase initiale du transfert de technologie, où je suis allé chez nos deux collaborateurs pour apprendre le protocole expérimental et l'utilisation du robot, ainsi que la partie analytique, incluant le traitement et l'analyse des données. J'ai ensuite été impliqué dans la conception des expériences, mais les résultats présentés dans cette partie ont été obtenus par les expériences de Virginie.

10.1.2.2 Résultats

		Intensité d'interaction					
Peptide	Séquence	PDZ1	PDZ1H	PDZ1-2	PDZ2	PDZ2H	PDZ3
Pcdh15-CD1	DVQPHSQSTSL	0.04	**0.30**	0.16	-0.08	0.02	-0.03
Pcdh15-CD2	GANSEGYNTAL	0.14	ND	ND	0.01	ND	0.01
Pcdh15-CD3	WNSLHIPMTKL	**0.39**	**0.85**	**0.82**	0.07	0.09	**0.26**
SANS	ERPLALEDTEL	**0.20**	**0.66**	**0.79**	-0.04	-0.03	0.19
Hrmn	PKEYDDELTFF	-0.03	-0.02	-0.01	-0.10	-0.04	**0.45**
CIB2	PDFLSTFHIRI	0.05	ND	ND	0.02	ND	0.00
Vlgr1	KEHTAFTDTHL	0.17	**0.79**	**0.91**	0.04	0.00	-0.03
Ush	LRRIPIADTHL	**0.24**	**0.95**	**0.70**	-0.02	-0.02	-0.11
Myo15	LTLPPSEITLL	-0.04	ND	ND	0.03	ND	**0.67**

Table 5 : Interactions déterminées par Hold-Up des domaines de la Whirline pour les peptides Usher. La séquence du peptide est indiquée avec les résidus canoniques du motif repérés en bleu. Les intensités d'interaction (BI) inférieures à 0.5 sont indiquées en noir, en orange pour un BI compris entre 0.5 et 0.75 et en vert pour un BI supérieur ou égal à 0.75. Les BI inférieurs à 0.2 sont considérés comme non significatif, leur signal/bruit étant trop faible pour distinguer les faux-positifs ; ND = no data.

Les résultats de Hold-Up confirment que le PDZ2 de la Whirline n'interagit avec aucun peptide testé, sur les protéines de l'audition (Table 5) mais aussi avec onze autres peptides contenant un PBM de type 1 et 2 préalablement testés par cette technique (résultats non montrés). Ce résultat tend à confirmer l'hypothèse que le PDZ2, avec ou sans épingle-ß, n'a pas de fonction d'interaction canonique propre.

Les résultats corrèlent bien avec ceux obtenus par SPR, en considérant les imprécisions respectives de chaque technique. Le peptide pcdh15-CD2 n'interagit pas, mais n'a pas pu être testé sur les constructions de PDZ1 contenant l'épingle-ß. Les résultats du peptide CD1 sont surprenants, car ce PBM est capable d'interagir avec le PDZ1-Hp mais pas avec le tandem. Le reste des interactions sont identiques à la SPR, avec des BI compatibles aux affinités mesurées.

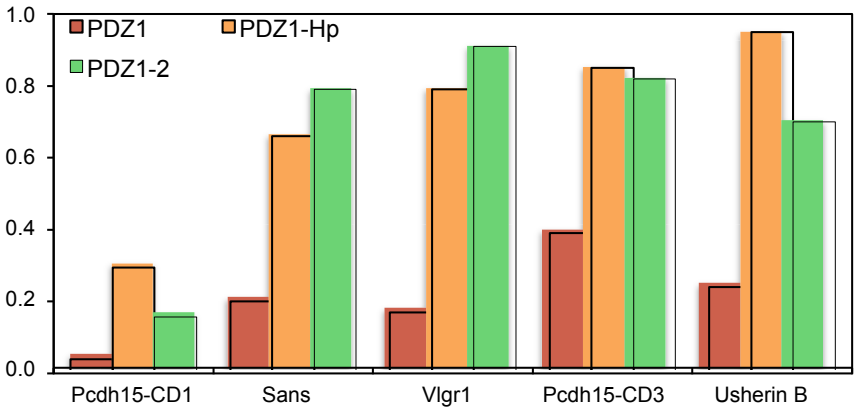


Figure 67: BI des constructions de Whirline pour différents peptides.

Avec la technique Hold-Up, l'amélioration systématique de l'interaction par ajout de structure additionnelle, notamment de l'épingle-ß, au domaine PDZ est confirmée (Figure 67). Dans trois des six ligands du PDZ1 de la Whirline, le signal Hold-up est en dessous de la limite de détéction pour le domaine isolé. L'ajout de l'extension de 25 acides aminés C-terminaux augmente drastiquement l'intensité d'interaction, en la faisant passer au-dessus de la limite de détection.

Dans notre banque de domaine actuelle, un tiers des PDZ ne lie aucun peptide avec une intensité supérieure au seuil de détection, sur 28 peptides testés ; nous ne sommes actuellement pas en mesure de déterminer si ces domaines ont une spécificité très marquée, s'ils nécessitent une structure additive pour leur fonction, ou si leur conformation n'est pas compatible avec une interaction en solution. Comme on le voit avec l'exemple extrême de Vlgr1, l'ajout de structure peut avoir un effet drastique, le domaine PDZ1 n'étant pas détecté lorsqu'il est isolé mais devient le PDZ de plus forte affinité lorsqu'il est inclus dans le tandem PDZ1-2. En conclusion, l'ajout de structure complémentaire, extensions ou domaine adjacent, peut avoir un effet drastique sur les affinités individuelles et donc sur le spectre de spécificité d'un peptide donné.

10.2 Interactions avec SANS

Nous nous sommes particulièrement intéressés à l'interaction entre le PDZ1 de la Whirline et la protéine SANS. Ces deux protéines sont toutes deux impliquées dans la formation du complexe Usher2 [183], dans la vision, mais aussi au niveau des liens basaux des stéréocils [179]. En plus de sa fonction biologique, l'interaction entre le peptide SANS et la

Whirline est l'une des plus fortes dans les peptides testés ; elle permet donc de tester efficacement le gain d'affinité du PDZ1 induit par la présence de structures additionnelles.

10.2.1 Gain d'affinité mesuré par RMN et fluorescence

	PDZ1	PDZ1-Hp	PDZ1-2	PDZ1-2^{LP363GG}
Kd (µM)	144 ± 27	73 ± 16	50 ± 14	33 ± 8

Table 6 : Affinités déterminées par titrage RMN entre SANS et la Whirline. L'erreur correspond à l'écart type entre les constantes de dissociation obtenues par le suivi de plusieurs pics.

Ces résultats sont déjà présentés dans le second article du chapitre précédent (Table 6). Ils montrent un gain d'affinité de trois à quatre fois entre le PDZ1 isolé (Kd = 144µM) et le PDZ1 dans le contexte du tandem (Kd = 50/33 µM). Une grande partie de l'amélioration est apportée par la présence de l'extension dans la construction PDZ1-Hp1 (Kd = 73 µM). Les résonances des résidus de l'extension du PDZ1 sont fortement perturbées par l'interaction du domaine avec le ligand, pointant, avec le gain d'affinité, à un rôle direct de celle-ci. Les perturbations de déplacement chimique reportées le long de la séquence montrent de façon non ambiguë que seul le PDZ1 est impliqué dans l'interaction avec SANS, les résonances du PDZ2 n'étant pas affectées par la présence du peptide en solution (section 9.3.8 ; figure supplémentaire 11). La forme ouverte du tandem a une affinité pour SANS similaire au tandem dans sa conformation fermée. L'épingle-ß du PDZ1 est dans la même conformation dans le tandem fermé ou ouvert, il semble donc que même dans le tandem, le gain d'affinité est exclusivement dû à la présence de l'épingle-ß. Le gain observé entre PDZ1-Hp1 et PDZ1-2 est probablement une conséquence de la stabilisation de l'épingle-ß dans un état conformationnel ou dynamique plus favorable à l'interaction avec le ligand.

L'émission de fluorescence des tryptophanes contenus dans l'épingle-ß du PDZ1 et du PDZ2 de la Whirline confirme le mécanisme d'interaction de celle-ci avec SANS (figure 68). Dans tous les cas, l'ajout du peptide SANS induit une exaltation du signal de fluorescence émis, que le tryptophane du premier épingle-ß soit exposé (PDZ1-Hp) ou non (PDZ1-2 ; PDZ1-2^{W375G}). Dans le cas du PDZ1-Hp1, l'interaction avec SANS induit une exaltation générale et une légère augmentation relative de l'intensité à 340 nm comparée à celle à 360 nm. Dans le tandem dans sa conformation ouverte (PDZ^{LP363GG}), les tryptophanes du PDZ1 et du PDZ2 sont exposés et donnent un signal avec un maximum à 360 nm. L'interaction avec le peptide SANS change complétement le profil de fluorescence, faisant apparaître un second maximum à 340 nm, donnant un profil très proche de celui de PDZ1-2^{WT}. L'interaction avec

le peptide protège le tryptophane du PDZ1 du solvant et montre que l'extension du PDZ1 est bien dans une conformation différente de celle dans le domaine isolé PDZ1-Hp. Bien que les deux domaines PDZ ne soient pas en interaction, la présence du lien et du PDZ2 permet bien de stabiliser l'épingle-ß du PDZ1 dans un état favorable pour l'interaction avec un peptide partenaire.

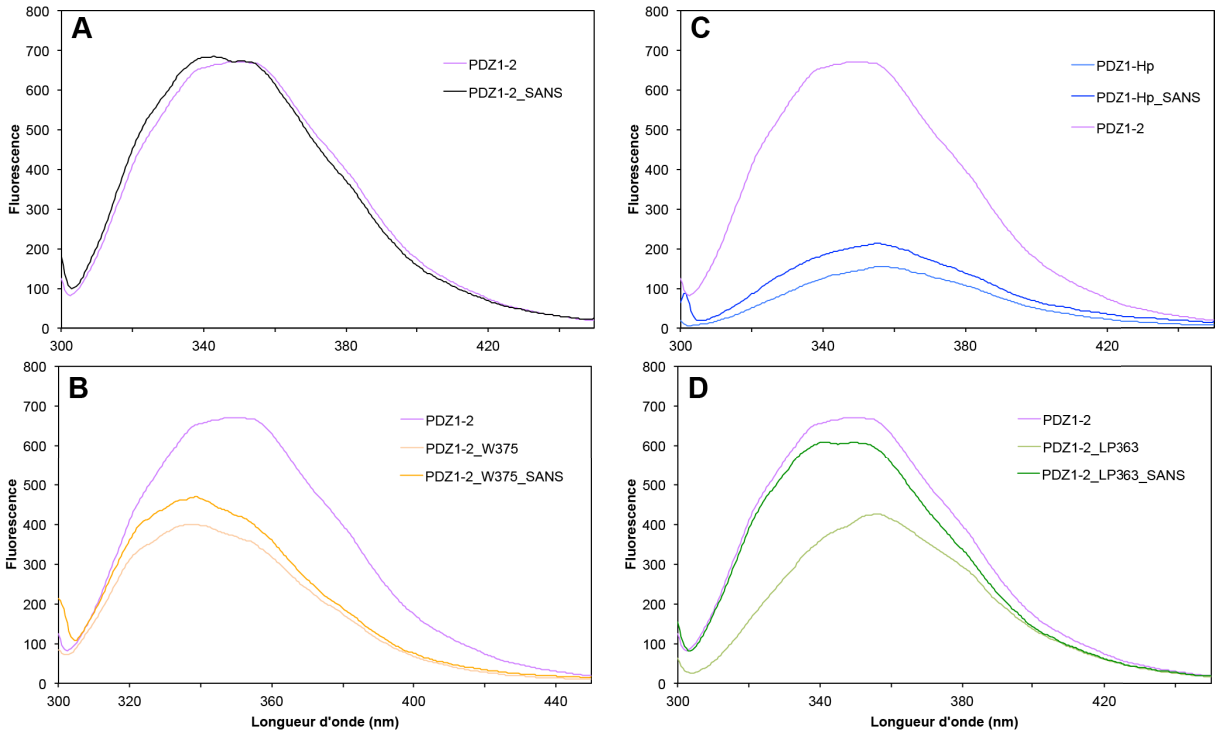


Figure 68: Emission de fluorescence des constructions de la Whirline en présence de SANS.

Les constantes de dissociation mesurées en solution par titrage RMN sont moins fortes que celles mesurées avec un peptide immobilisé. Comme nous l'avons discuté précédemment, ces différences sont probablement dues à l'environnement particulier à l'interface entre peptide et solution de protéine. Cependant, le gain relatif d'affinité entre PDZ1 et les constructions contenant l'épingle-ß est conservé. Le gain majeur, observé par toutes les techniques, provient de l'ajout de l'épingle-ß, le gain entre PDZ1-Hp et le tandem n'étant pas systématiquement observé selon les techniques utilisées.

10.2.2 Affinité avec Sans SAM-PBM

Dans la structure du complexe entre Harmonine et SANS, l'interaction entre le PDZ et le PBM est complétée par une interaction sur une large surface entre le domaine PDZ de l'Harmonine et le domaine SAM, précédant le PBM de SANS. Cette interaction supplémentaire augmente d'un facteur 1000 l'affinité entre l'Harmonine et le PBM isolé de

SANS. La superposition de la structure du PDZ1-2 de la Whirline et du complexe Harmonine/SANS montre que le PDZ1 de la Whirline peut simultanément accommoder le PDZ2 et le SAM de SANS. Nous voulions donc étudier l'impact de l'addition du domaine SAM de SANS sur son affinité avec la Whirline.

Le domaine SAM de SANS est difficile à purifier et a tendance à polymériser en solution. Nous utilisons le mutant K437E stabilisateur de SANS SAM-PBM, proposé par l'équipe du Dr. Zhang [71]. Cependant, en dépit de cette mutation, la construction est toujours oligomérique dans nos conditions. Nous avons donc utilisé une technique apparentée à la SPR (OctetRed384), nous permettant d'immobiliser sur une surface le domaine SAM-PBM avec une étiquette GST N-terminale. Cette approche nécessite une faible quantité de SAM-PBM et permet de réduire l'influence de son oligomérisation.

	PDZ1			PDZ1-Hp			PDZ1-2			PDZ1-2^{LP363GG}		
SAM PBM	**19** ±1.9	**21** ±3.7		**4.3** ±0.6			**4.6** ±0.5	**4.4** ±0.6	**4.6** ±0.7	**3.7** ±0.5	**3.2** ±0.5	
PBM	**28** ±1.6	**17** ±2.0	**16** ±1.7	**6.2** ±0.4	**6.6** ±0.6	**6.6** ±0.5	**2.2** ±0.2	**2.4** ±0.2	**2.6** ±0.2	**1.8** ±0.2	**1.9** ±0.2	**2.1** ±0.2

Table 7 : Affinités déterminées par titrage Octet entre SANS et la Whirline. L'erreur correspond à l'erreur l'ajustement de courbe pour chaque titration. Deux constructions de SANS sont utilisées, SAM-PBM avec une étiquette GST N-terminale et portant la mutation K437E, et PBM correspondant au peptide biotinylé composé des 10 résidus C-terminaux de SANS.

Nous avons comparé l'interaction du peptide incluant uniquement le PBM de SANS avec la construction contenant le SAM-PBM contre quatre constructions de la Whirline (Table 7) : le PDZ1 isolé, ou avec épingle-ß (PDZ1-Hp), le tandem compact (PDZ1-2) ou ouvert (PDZ1-2^{LP363GG}). Avant tout, les affinités observées entre SANS et le PDZ1 de la Whirline sont bien plus importantes que celles que nous avons mesurées avec les autres techniques. Ce phénomène semblant se confirmer lors de toutes nos utilisations de cet appareil (résultats non montrés).

Pour toutes les constructions de la Whirline, les affinités sont très similaires avec ou sans le domaine SAM. Ce domaine SAM n'interagit pas avec le PDZ1 ou le PDZ2 de la Whirline, à la différence de ce qui est observé pour l'Harmonine. Nos collaborateurs du laboratoire de Christine Petit ont testé cette interaction avec une construction Nter-PDZ1 de la Whirline et n'ont pas observé de changement significatif de l'affinité par l'ajout du domaine SAM (résultats non-publiés d'Elise Peppermans). L'interaction SAM/PDZ1 est donc unique à l'Harmonine.

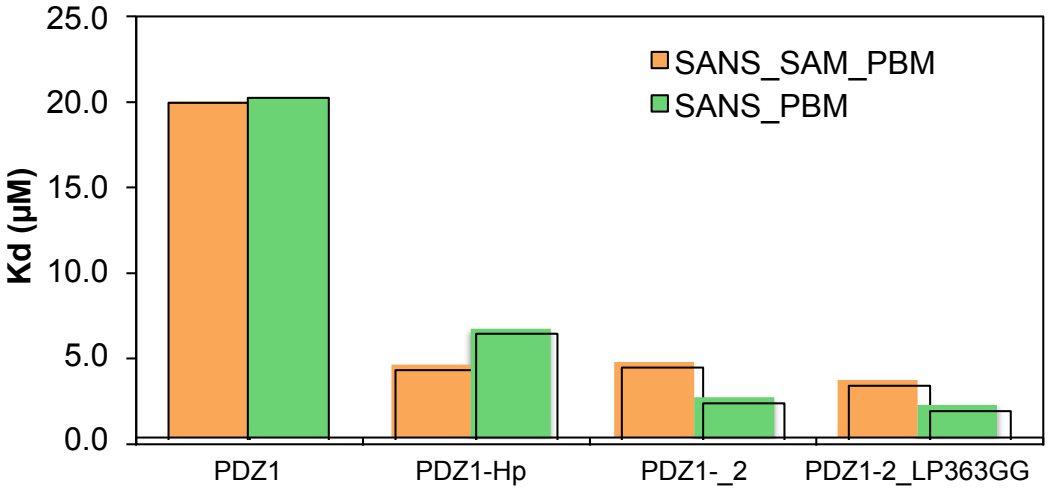


Figure 69 : Affinités déterminées par titrage Octet entre SANS et la Whirline. Deux constructions de SANS sont utilisées, SAM-PBM avec une étiquette GST N-terminale et portant la mutation K437E, , et le PBM biotinylé correspondant aux10 résidus C-terminaux de SANS.

10.3 Interactions avec les lipides

Le dernier type d'interaction que nous avons étudié est l'interaction du PDZ1 avec les phospholipides. Le PDZ1 de la Whirline a un pI très élevé ; la présence de nombreuses charges positives au pH physiologique pourrait être favorable à leur interaction avec les phospholipides anioniques. L'interaction du PDZ1 de la Whirline avec certains lipides, notamment les phosphatidyl-inositols, a été montrée précédemment dans une étude de l'équipe du Dr P. Zimmermann. L'article se concentre sur la prévalence des interactions lipides/PDZ et sur leurs techniques de détection [113]. Dans cette étude, la construction du PDZ1 est tronquée et a tendance à oligomériser, comme le reportent les auteurs.

Pour confirmer l'interaction entre le PDZ1 de la Whirline et certains lipides, nous avons utilisé des membranes de nitrocellulose incluant 15 lipides différents fixés à leur surface (Echelon, P-6001). En incubant la membrane avec une solution de protéine, et en révélant par anticorps après rinçage, il est possible de déterminer de façon qualitative l'interaction d'un lipide avec la construction. Dans cette expérience, toutes les constructions de la Whirline sont utilisées avec une étiquette GST à leur N-terminal pour permettre la révélation. Les membranes sont incubées dans une solution de protéine à 10^{-3} g/l pendant une heure et la révélation est faite avec un anticorps anti-GST direct (Abcam Ab3416), permettant un très faible bruit de fond dans ces expériences.

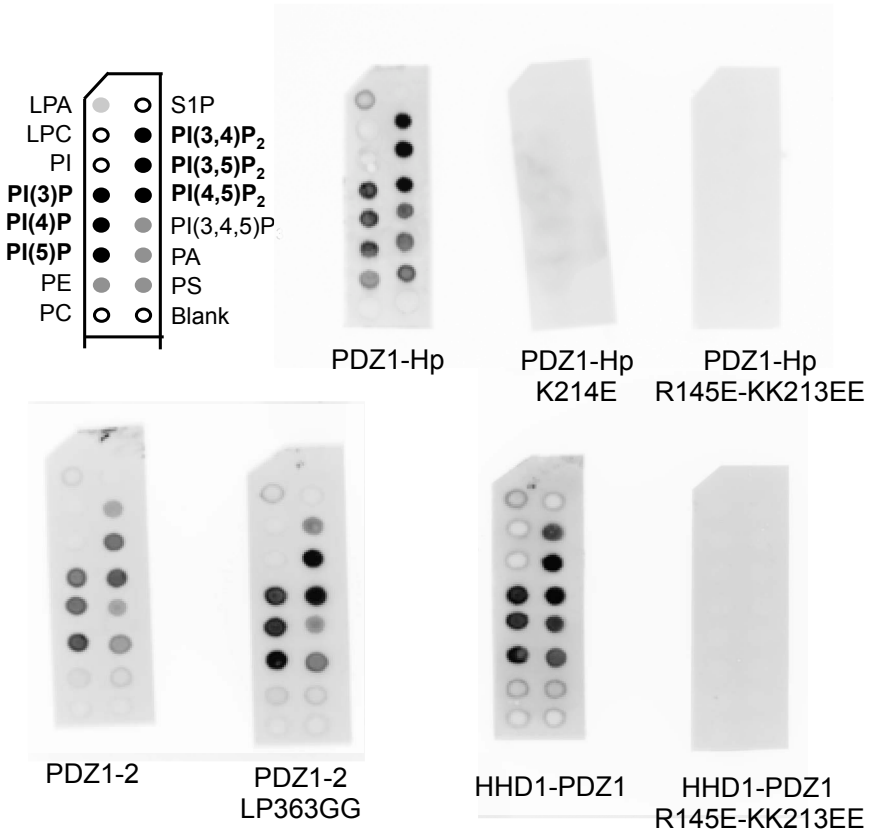


Figure 70: Interaction entre les lipides et le PDZ1 de la Whirline. LPA : Acide Lysophosphatidique ; LPC Coline Lysophosphatidique ; PI : Phosphatidyl-inositol ; PI(3)P :Phosphatidyl-inositol-3-Phosphate ; PI(4)P : Phosphatidyl-inositol-4-Phosphate ; Phosphatidyl-inositol-5-Phosphate ; PE : Phosphatidyl-étanolamine ; PC : Phosphatidyl-choline ; S1P = Sphingosine-1-Phosphate ; PI(3,4)P :Phosphatidyl-inositol-3,4-Biphosphate ; PI(3,5)P :Phosphatidyl-inositol-3,5-Biphosphate ; PI(4,5)P :Phosphatidyl-inositol-4,5-Biphosphate, PI(3,4,5)P :Phosphatidyl-inositol-3,4,5-Triphosphate ; PA : Acide Phosphatidique ; PS : Phosphatidyl-Serine.

Notre construction PDZ1-Hp1 interagit avec les lipides, comme observé dans l'article du Dr. Zimmermann [113]. Le domaine est particulièrement fixé sur les spots correspondants aux PIP_1 et PIP_2, sans préférence évidente pour la position de phosphorylation sur l'inositol. Le PI non phosphorylé est incapable d'interagir avec le PDZ1-Hp et le spot du PIP_3 est moins intense que les autres PIP_x suggérant une moins bonne affinité. De façon moins marquée, le PDZ1-Hp est aussi retrouvé interagissant avec les PA, LPA et PS, trois phospholipides anioniques, et avec le PE, cationique. Nous avons alors testé ces interactions dans des constructions contenant les domaines adjacents du PDZ1 (PDZ1-2 ; PDZ1-$2^{LP363GG}$; HHD1-PDZ1) ; dans toutes ces conditions, les interactions sont les mêmes que pour le domaine isolé, mis à part pour l'absence de liaison avec LPA, PE et PS, suggérant que ces interactions

pourraient être artéfactuelles. Dans ces constructions, le PDZ1 montre une légère préférence pour deux PIP_2 phosphorylés en position 5.

Pour déterminer la spécificité de l'interaction entre PDZ1 et les phosphatidyl-inositols, nous avons utilisé des mutations ponctuelles, à l'origine générées pour réduire le niveau d'oligomérisation de HHD1-PDZ1. Ces mutations inversent les charges d'une région particulièrement chargée positivement à la surface du PDZ1 (**section 7.2, figure 46A**). Ce patch cationique ciblé est localisé au début de la boucle "GLGF" ß1-ß2 et à la fin de la boucle α2/ß5 mais nous avons vérifié que les mutations ne perturbent pas la structure du PDZ1 (**section 7.2, figure 46B**). Le patch cationique ciblé est localisé au début de la boucle "GLGF" ß1-ß2 et à la fin de la boucle α2/ß5. L'inversion de la charge d'un seul des résidus de ce patch est suffisante pour supprimer l'interaction avec l'ensemble des phospholipides fixés à la membrane. L'inversion de charge sur PDZ1-Hp et sur HHD1-PDZ1 a le même effet. Ces résultats confirment que l'interaction du PDZ1 avec les PIP_x est liée à ce patch. Ces résultats confirment les effets de mutation sur le PDZ1 de la Whirline observés en SPR par Ivarsson *et al.* [113] .

Le N-terminal de la Whirline interagit principalement avec des protéines membranaires, organisant à priori un réseau sous membranaire de protéines d'adhésion au cytosquelette d'actine. L'interaction de cette partie de la Whirline avec les lipides aurait donc un sens, au moins du point de vue de la localisation et de la fonction de la protéine. Nous avons montré ici que la Whirline interagit spécifiquement avec les PIP mono et bi-phosphate par un patch cationique localisé sur son PDZ1. Ce patch est spatialement séparé du site d'interaction au PBM du domaine ainsi que de l'interface avec le PDZ2 et la position hypothétique du HHD1. La membrane des stéréocils est particulièrement enrichie en PI(4,5)P [308] , son interaction avec le PDZ1 pourrait donc avoir un rôle important pour l'orientation et la localisation de la Whirline à la membrane des stéréocils, proche des protéines membranaires partenaires.

Les membranes enrichies avec des lipides que nous avons utilisées ne sont pas le meilleur système pour étudier les interactions lipide/protéine. Elles sont en effet réputées pour produire de nombreux faux-positifs. L'absence d'interaction avec les constructions contenant le patch avec les charges inversées nous conforte dans l'interprétation de nos données, mais ces résultats nécessitent une confirmation par un système modèle mieux adapté. Il conviendra donc de suivre la localisation de la Whirline en cellules, ou du moins faire des études *in vitro* d'interaction avec des liposomes.

Perspective & Discussion

11 Perspective

Durant ces cinq dernières années, la compréhension de la fonction de la Whirline s'est grandement améliorée, notamment par l'identification non ambiguë de ses isoformes exprimées dans les cellules ciliées cochléaires, les photorécepteurs et les cellules ciliées vestibulaires[221,226,309], ainsi que par la détermination de leur localisation. L'apparente indépendance spatiale et fonctionnelle des isoformes de la Whirline est tout à fait particulière. Dans les stéréocils, chaque isoforme a des localisations discrètes qui ne se superposent pas. Il apparaît maintenant clair que l'isoforme longue de la Whirline est impliquée dans des complexes ne possédant aucune composante commune avec les complexes de l'isoforme courte. Les maladies induites par les mutations du gène associé à la Whirline sont tout aussi ségrégées, le syndrome Usher est exclusivement induit par des mutations contenues dans les 300 premiers résidus de la protéine et les surdités non-syndromique dans les 500 derniers résidus [274]. Il est important de noter que, bien que la localisation de la Whirline soit bien identifiée, les rôles des complexes dont elle fait partie restent largement incompris.

Pendant cette thèse, nous nous sommes employés à déterminer les caractéristiques biochimiques et biophysiques de la protéine Whirline *in vitro*, et à les confronter aux informations provenant des stéréocils et les photorécepteurs. Nos travaux ont permis l'identification et la détermination de la structure d'un complexe transitoire formé par le tandem de domaines PDZ de la Whirline. Nous avons aussi établi la structure du domaine HHD2 et montré que celui-ci se comporte indépendamment du supramodule de domaines PDZ le précédent (Figure 71). Nous avons enfin identifié *in vitro* les partenaires, protéiques ou lipidiques, des domaines PDZ de la Whirline dans le réseau des protéines Usher. Jusqu'à présent, les résultats structuraux que nous avons obtenus sur la Whirline sont à relier avec précaution avec la fonction cellulaire de la protéine. Cependant, la clarification récente de la localisation des isoformes de la Whirline ouvre de nouvelles possibilités pour l'étude structurale de la Whirline au sein de ses différents complexes.

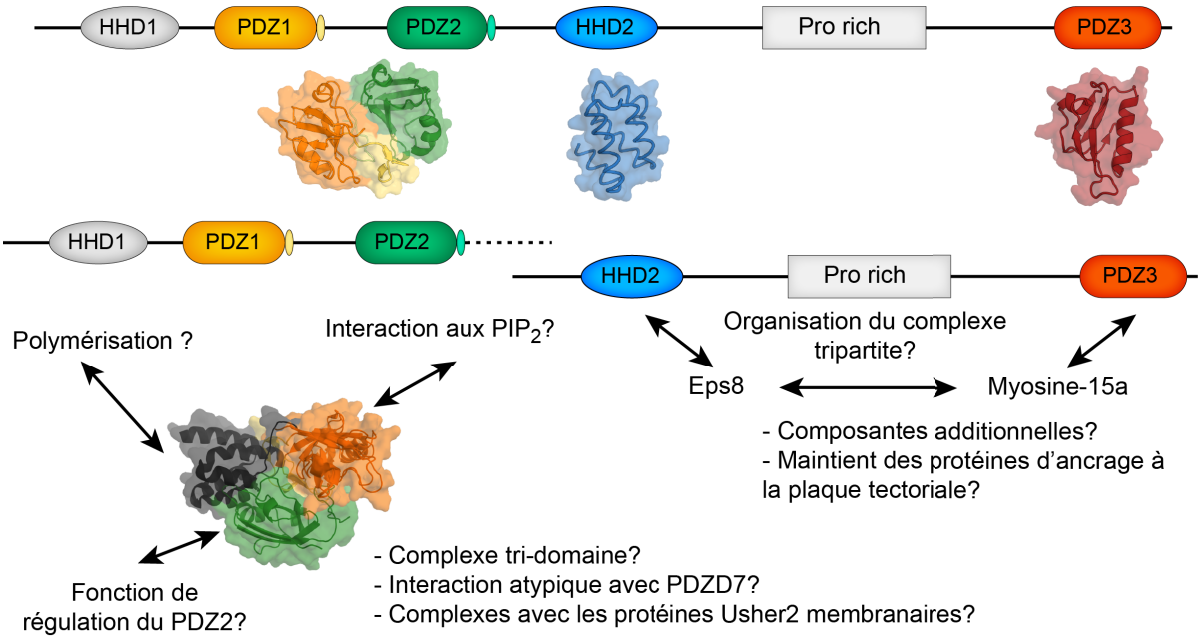


Figure 71: Questions ouvertes sur les complexes de la Whirline

11.1 Les complexes de l'isoforme courte

Le travail de cette thèse s'est principalement concentré sur l'étude des domaines N-terminaux de la protéine, au détriment de l'étude des composantes de l'isoforme courte C-terminale. Cette isoforme courte a une organisation en domaines tout à fait particulière, avec deux domaines aux extrémités séparés par une grande séquence non-structurée contenant une région riche en Proline. L'hypothèse simple que nous avons mise en avant jusqu'à présent est que cette région joue un rôle "d'espaceur " permettant d'isoler le PDZ3 et le PBM, connectés au cytosquelette [310]. Il serait cependant intéressant d'étudier la conformation de cette isoforme en solution pour déterminer si cette région restreint bien l'orientation relative des domaines à ses extrémités.

L'isoforme courte de la Whirline forme un complexe tripartite avec Eps8 et la Myosine-15a, ces trois protéines interagissant directement les unes avec les autres [211]. Eps8 est en partie non-structurée et possède plusieurs segments potentiellement en hélices amphiphiles pouvant être des motifs d'interaction au domaine HHD2 de la Whirline. Le domaine PTB (phospho-tyrosine binding) d'Eps8 pourrait lui aussi être impliqué dans cette interaction. Du côté de la Myosine-15a, nous avons confirmé que son PBM lie le PDZ3 de la Whirline [227]. Cependant, d'autres interactions entre le couple MyTH$_4$/FERM situé directement en amont du PBM de la Myosine-15a et la Whirline ont été reportées [219]. Très récemment, un article a montré l'interaction non canonique entre le PDZ3-PBM de l'Harmonine avec le

$MyTH_4$/FERM C-terminal de la Myosine-7a [311]. Le PDZ3 de la Whirline possède lui aussi un PBM proche de son PDZ3, une telle interaction avec la Myosine-15a est donc tout à fait possible. Enfin, la Myosine-15a et Eps8 possèdent toutes deux des domaines SH3 pouvant potentiellement interagir avec la région riche en proline de la Whirline. Eps8 interagit avec la région IQ et le premier couple $MyTH_4$/FERM de la Myosine-15a.

Les protéines Usher tendent à former des complexes où les protéines interagissent entre elles par plusieurs points d'ancrage indépendants [165]. Cette multiplication d'interactions le long des protéines permet de former des complexes de forte affinité et constitue autant de points de régulation de ces complexes. Il est probable que la Myosine-15a, Eps8 et la Whirline interagissent entre-elles par plusieurs de leurs domaines. Il sera intéressant d'étudier ces interactions, en commençant par les plus probables, entre le HHD2 et le N-terminal d'Eps8 et entre le PDZ3 et la séquence $MyTh_4$-FERM-PBM de la Myosine-15a. Le doctorant Yanlei Zhu, prenant la suite du projet, est actuellement en train de travailler sur ces interactions.

11.2 Le complexe Usher2

La Whirline est impliquée dans la formation du complexe Usher2 dans les stéréocils et les photorécepteurs. Son rôle suspecté est de servir de protéine adaptatrice, reliant les extrémités cytoplasmiques des protéines de jonction cellulaire entre elles, et de les ancrer au cytosquelette d'actine. Nous avons confirmé que le PDZ1 de la Whirline interagit avec Vlgr1 et l'Usherine ainsi qu'avec les PIP_2. Nous avons aussi montré qu'HHD1 induit l'oligomérisation de la protéine *in vitro*. Ces caractéristiques biochimiques sont favorables au rôle proposé de la Whirline ; elle s'oligomérise, interagissant avec la membrane et les protéines des liens basaux, permettant de concentrer les protéines de lien dans une région confinée des stéréocils, tout en les connectant au cytosquelette.

La Whirline est assistée dans cette fonction par la protéine PDZD7 [33,234], qui n'est pas présente dans le complexe Usher2 des photorécepteurs, où la Whirline semble suffisante pour joindre les protéines membranaires. Le rôle de PDZD7 et sa complémentarité avec la Whirline, ne sont pas bien compris. Actuellement, ces deux protéines sont décrites comme interagissant avec les mêmes partenaires [179,214,219], et sont capables de former un hétérodimère. Il serait intéressant d'étudier les complexes entre ces deux protéines, en présence des peptides C-terminaux des protéines membranaires Vlgr1 et Usherine. Une première piste pour l'interaction entre ces protéines est que le PDZ3 de PDZD7 est de type-2,

pouvant potentiellement interagir avec le PBM de la Whirline. PDZD7 a elle aussi un tandem de domaines PDZ, possédant chacun une extension C-terminale en épingle-ß conservée avec celles de la Whirline et de l'Harmonine ; est ce que ces domaines de PDZD7 peuvent former un supramodule en *cis*, voire en *trans,* avec le tandem de PDZ de la Whirline ? Caractériser le complexe potentiel entre PDZD7 et la Whirline pourrait permettre de déterminer la fonction apportée par la présence de PDZD7 dans ce complexe, qui n'est pas nécessaire dans la vision, mais dont la mutation du gène associé induit des problèmes morphologiques dus à la mauvaise formation des liens basaux [33,209].

Les régions cytoplasmiques des protéines membranaires Usher2 comprennent environ 150 résidus et sont à priori non-structurées. Il serait intéressant de rechercher des motifs d'interaction aux HHD dans ces régions, qui pourraient s'ancrer sur le HHD1, en plus d'interagir avec le PDZ1, comme dans le cas de la cadhérine-23 et de l'Harmonine [64]. Dans le cas de l'Usherine, sa région cytoplasmique interagit avec une région non structurée de SANS, formant ainsi un complexe tripartite avec la Whirline [183]. La formation de ce complexe soulève la question de l'interaction simultanée du PDZ1 de la Whirline avec les PBM de SANS et de l'Usherine, suggérant l'oligomérisation de la protéine.

Les mécanismes de l'apparition de la rétinite pigmentaire sont très mal compris. L'étude des complexes entre protéines Usher2 devrait aider à la compréhension de l'apparition de cette pathologie.

12 Discussion

Les domaines forment les unités structurales et fonctionnelles des protéines. Ils sont formés de séquences conservées de 30 à 300 résidus repliés en une conformation définie, donnant au domaine sa fonction, qu'elle soit ou non catalytique. Dans les voies métaboliques, ancestrales et partagées par tous les organismes, les protéines sont majoritairement composées d'un seul domaine catalytique. Les autres fonctions cellulaires reposant sur la transmission d'information impliquent des protéines contenant de multiples domaines. L'enrichissement en protéines multi-domaines est supérieur dans le génome des métazoaires par rapport aux génomes bactériens ; c'est une conséquence probable de la nécessité accrue de voies variées de communication et de régulation qu'implique la multicellularité [312–314]. La limite de taille des domaines conduit à un nombre limité de repliements et de fonctions fournies par un domaine isolé. Ainsi, les familles de repliement servent d'éléments de base, dont la combinaison permet l'enrichissement du répertoire de protéine [315]. Les protéines multi-domaines composent près de 80 % du protéome humain. Elles sont constituées par la combinaison de blocs simples, dont la fonction est modulée par leur contexte séquentiel [123] ou celui de leur partenaire, mais aussi par la composition spécifique du sous-compartiment cellulaire dans laquelle la protéine ou ses isoformes sont exprimées.

Nos travaux portent sur la compréhension des rôles moléculaires de la Whirline, une grande protéine d'échafaudage impliquée dans la mise en place et le maintien de la morphologie des cellules ultra-spécialisées responsables de la vision et de l'audition. Pour effectuer sa fonction, la Whirline interagit avec de nombreux partenaires, par ses multiples domaines d'interaction. Nous nous sommes intéressés aux interactions de ces modules, observés individuellement, ainsi qu'à la modulation de leur fonction dans le contexte de constructions multi-modulaires. Nous avons résolu la structure d'un supramodule dynamique formé par le tandem de domaines PDZ ainsi que la structure du second domaine HHD de la Whirline. Le tandem et le HHD2 sont reliés par une région flexible et n'interagissent pas entre eux, en solution.

Les deux domaines PDZ N-terminaux de la Whirline possèdent des extensions structurées à leur C-terminale qui sont conservées dans les autres protéines à domaine PDZ impliquées dans l'audition. Ces extensions ont la double fonction de maintenir des structures supramodulaires et d'améliorer drastiquement l'affinité des domaines PDZ adjacents pour leurs partenaires [186,251]. Dans la cellules, les 266 domaines PDZ du protéome humain

reconnaissent spécifiquement des motifs courts, dont l'espace de variabilité en séquence est extrêmement limité [316]. Les domaines PDZ sont souvent couplés à des extensions [132], fournissant aux domaines concernés un élément différenciateur, le démarquant des autres membres de la famille [76,124,123]. Ces extensions peuvent apporter au domaine un gain d'affinité en interagissant directement avec des résidus en amont du motif canonique des partenaires [92,251] ou en stabilisant le domaine dans une conformation favorable à l'interaction [135]. Elle peuvent aussi avoir un rôle activateur [134], ou un rôle dans la formation de supramodule, comme nous l'avons présenté. Les extensions sont un exemple simple de l'influence du contexte protéique sur l'activité d'un domaine.

Des successions de domaine sont retrouvées en plusieurs occurrences dans le génome : c'est notamment le cas des SH2-SH3, des PDZ-SH3-GuK ainsi que des tandems de domaine PDZ [317], on parle alors de supradomaines. Au vue de nos résultats, six des neuf domaines HHD identifiés à ce jour sont au voisinage immédiat d'un domaine PDZ, notamment dans les protéines de l'audition. On pourrait considérer la succession HHD-PDZ comme une classe de supradomaine, dans le sens de bloc évolutionnaire conservé. La combinaison de domaines successifs dans une même séquence peptidique promeut l'apparition de nouvelles fonctions [312]. Les domaines successifs d'un supramodule peuvent s'associer en un large complexe permettant d'augmenter la fonctionnalité de ses parties constituantes. Nous avons montré la présence d'une telle organisation entre les deux premiers domaines PDZ de la Whirline, accroissant l'affinité du domaine PDZ1 pour l'un de ses partenaires, alors que le PDZ2 n'a pas la capacité d'interagir avec un partenaire Usher. Cet effet *in vitro* ne représente probablement pas l'étendue de la fonction ajoutée par l'interaction de ces deux domaines. Sur la trentaine de tandems de domaines PDZ potentiels dans le génome, seul six ont une structure définie montrant une unité structurale [124,154]. Pour la moitié de ces tandems, l'un des domaines n'a pas de partenaire, il permet soit d'agrandir la région d'interaction du domaine actif [154], soit il agit comme interrupteur, permettant d'activer le domaine actif, en réponse à un stimulus [50,148,149]. Dans ce dernier cas, l'organisation en supramodule permet de transmettre une information en produisant un effet en réponse à une stimulation, en utilisant simplement des domaines non-catalytiques [318]. L'étude de la structure des domaines dans leur contexte séquentiel est donc essentielle pour comprendre l'ensemble de mécanisme de régulation les affectant.

La cellule est composée d'une multitude de compartiments enrichis en certaines protéines [319]. Cette compartimentalisation est notamment assurée par les protéines d'échafaudage, maintenant les protéines d'une même voie dans un espace confiné [292]. La

probabilité de rencontre entre protéines est donc augmentée, grâce à la concentration locale accrue des composantes. Ce phénomène est particulièrement important dans le cas des domaines PDZ ayant généralement plusieurs partenaires potentiels avec la même gamme d'affinité [97]. La localisation des protéines contenant des domaines PDZ est donc critique dans la composition de leurs complexes. Les isoformes de la Whirline sont différentiellement exprimées dans les stéréocils des cellules ciliées, un compartiment qui ne représente qu'une faible fraction du volume de la cellule [25]. Au sein des stéréocils, ces isoformes ont chacune une localisation discrète, réduisant le nombre de partenaires de la partie C-terminale de la protéine. Les PBM des protéines de l'audition sont similaires et nous avons vu que certaines interactions observées *in vitro* pourraient simplement être une conséquence de ce manque de diversité sans avoir une pertinence biologique. La localisation et le contrôle temporel de l'expression de ces protéines à PDZ sont donc des facteurs essentiels pour la composition des complexes que celles-ci forment.

Les cellules sensorielles sont ultra-spécialisées vers une seule fonction ; transmettre l'information de la déflection de son organe mécanosenseur le plus rapidement possible au premier neurone afférent. Au niveau moléculaire, la signalétique associée est entièrement mécanique, sans implication de domaine catalytique dont l'activité est trop lente. L'ensemble de leurs fonctions sensorielles est donc contrôlé par une succession de changements de conformation induits par des stimuli provenant du milieu extracellulaire. Nous avons montré que les PDZ de la Whirline interagissent de façon transitoire, en oscillation entre une forme compacte définie et un ensemble de conformations plus ouvertes. A ce jour, nous n'avons pas pu identifier de rôle fonctionnel associé à cette dynamique. De manière plus générale, la dynamique des protéines et leurs changements de conformation sont à la base de leurs fonctions. Le terme dynamique inclut les mouvements rapides locaux transmettant de l'information au sein d'un domaine [320,321], les grand mouvements conformationnels [322] ainsi que l'interaction transitoire entre partenaire, responsable de la dynamique des réseaux [292]. Les avancées récentes en biologie structurale permettent maintenant de se concentrer sur l'étude de la dynamique des protéines, essentielle pour la compréhension des mécanismes moléculaires de la cellule.

Les études structurales de domaines isolés, encore majoritaires il y a quelques années, font peu à peu place à l'étude plus ambitieuse de séquences multi-modulaires. Les grands programmes de génomique structurale ont permis de compléter les familles de structures uniques qui sont à présent en grande partie découvertes. Signe de ce succès, depuis 2010, seuls trois nouveaux repliements de domaine ont été déposés dans la PDB (rcsb.org). Un

catalogue quasi complet de structures permet maintenant de générer des structures de domaine en se basant sur celles de protéines avec une identité de séquence suffisante. Par ailleurs, l'amélioration et la multiplication des données génomiques permettent maintenant, en se basant sur la covariation de séquences entre génomes, de prédire les contacts inter-résidus. Ceux-ci peuvent être utilisés comme contraintes ambiguës en complément d'un nombre restreint de contraintes expérimentales pour générer des structures [323]. La biologie structurale se concentre maintenant sur l'étude d'objets plus complexes et plus pertinents. La cristallographie et la microscopie électronique se tournent vers l'étude à haute résolution de larges assemblages rigides [288]. L'étude structurale des objets flexibles et des complexes transitoires en solution est quant à elle plus difficile à réaliser par l'utilisation d'une seule technique, et par une représentation en conformation unique. L'utilisation des différents outils offerts par la RMN, intégrés aux données provenant de la diffusion de rayons X aux petits angles ou du FRET sur molécule isolée, s'avère particulièrement efficace dans la détermination structurale de ces objets flexibles [289,324–326]. La force de l'utilisation combinée de ces techniques réside dans l'étude des modifications de conformation sur une large gamme de temps et d'amplitude de mouvement, donnant ainsi accès à l'étude atomique des changements, à la base de la biologie. L'utilisation de techniques à basse résolution nécessite aussi l'utilisation d'outils bioinformatiques, les développements des méthodes de modélisation et l'augmentation de la puissance de calcul ont un rôle important dans l'essor de l'utilisation de ces techniques. L'évolution de la RMN a été quant à elle accélérée par les développements du marquage spécifique d'acides aminés et l'utilisation de sondes fluors. Ils permettent d'introduire un nombre limité de sonde sur une protéine, s'adressant à des objets physiologiquement plus pertinents au détriment de données structurales exhaustives, se rapprochant ainsi du FRET. La RMN est aussi l'une des très rares techniques permettant d'accéder directement aux états conformationnels faiblement peuplés des protéines [291,325,327].

Le travail fourni en biologie structurale pendant les trente dernières années a permis un accès à la conformation des domaines isolés et permet maintenant de se concentrer sur l'étude de la dynamique, des mouvements et des complexes transitoires à la base des changements et de la transmission d'informations en biologie.

Appendices

13 Appendices

13.1 Séquence de la Whirline murine isoforme 3

```
        10         20         30         40         50
MNAQLDGLSV SSSSTGSLGS AAAAAGGGGG AGLRLLSANV RQLHQALTAL
        60         70         80         90        100
LSEPEREQFT HCLNAYHARR NVFDLVRTLR VLLDSPVKRR LLPMLRLVIP
       110        120        130        140        150
RSDQLLFDQY TAEGLYLPAT TPYRQPAWAA PDGAGPGEVR LVSLRRAKAH
       160        170        180        190        200
EGLGFSIRGG SEHGVGIYVS LVEPGSLAEK EGLRVGDQIL RVNDKSLARV
       210        220        230        240        250
THAEAVKALK GSKKLVLSVY SAGRIPGGYV TNHIYTWVDP QGRSTSPPSS
       260        270        280        290        300
LPQPHGSTLR QREDDRRSTL HLLQSGDEKK VNLVLGDGRS LGLTIRGGAE
       310        320        330        340        350
YGLGIYITGV DPGSEAESSG LKVGDQILEV NGRSFLNILH DEAVKLLKSS
       360        370        380        390        400
RHLILTVKDV GRLPHARTTV DQTKWIASSR IGESVANSAG FPGDHTEEGT
       410        420        430        440        450
SKPGFYKGPA GSQVTLSSLG NQTRALLDDQ ARHLLTEQER ATMMYYLAQY
       460        470        480        490        500
RGGTISVEAM VMALFELLNT HAKFSLLSEV RSIISPQDLD RFDHLVLRRE
       510        520        530        540        550
IESMKARQPP GPGVGDTYSM VSYSDTGSST GSHGTSTTVS SARNTLDLEG
       560        570        580        590        600
TGETTQGSTN ALPDVSVDDV KSPSEDLPGI KPPPPPPPLA QGHDRLLGQP
       610        620        630        640        650
RKPGREDPAP LSSAAHSGIV FSAPRNRSPP PGTAPTPGPS SAQDSPSSPI
       660        670        680        690        700
YASISHANPS SRKPLDTHLA LVNQHPIGPF PRVQSPPHLK SPPAETPGAG
       710        720        730        740        750
ACLPPPSPSE HPDAVGANQH FVLVEVHRPD SEPDVNEVRA LPQTRTASTL
       760        770        780        790        800
SQLSDSGQTL SEDSGVDAGE TEASTSGRGR QTASAKNKNG KEQPRTERTA
       810        820        830        840        850
EGANKPPGLL EPTSTLVRVR KSAATLGIAI EGGANTRQPL PRIVTIQRGG
       860        870        880        890        900
SAHNCGQLKV GHVILEVNGQ TLRGKEHKEA ARIIAEAFKT KERDYIDFLV
    907
TEFNVML
```

Figure Supplémentaire 1 : Séquence de la Whirline avec son organisation en domaines. La séquence inclut : régions de complexité basse (violet) ; HHD1 (gris), PDZ1 (orange), épingle-ß1 (jaune), PDZ2 (vert foncé), épingle-ß2 (vert), HHD2 (bleu), PDZ3 (rouge).

13.2 Liste des résidus en échange dans le tandem PDZ1-2

Construction	WT	W375	IP225	W237	LP363	H340
Conf.	Fermé	Fermé	Fermé	Ouvert	Ouvert	Ouvert
159	✓	✓	-	-	✓	✓
161	✓	✓	-	-	✓	✓
162	✓	✓	-	-	✓	✓
166	✓	✓	-	-	✓	✓
169	✓	✓	-	-	✓	✓
170	✓	✓	-	-	✓	✓
188	✓	✓	-	-	✓	✓
222	✓	✓	-	-	✓	✓
228	✓	✓	-	-	✓	✓
241	✓	✓	✓	✓	✓	✓
242	✓	✓	✓	x	✓	✓
243	✓	✓	✓	x	✓	✓
279	✓	✓	✓	x	x	x
292	✓	✓	✓	✓	x	-
296	✓	✓	✓	x	x	x
300	✓	✓	✓	x	x	?
307	✓	✓	✓	✓	x	?
311	✓	✓	✓	x	x	x
315	✓	✓	✓	x	x	x
324	✓	✓	✓	x	x	x
340	✓	✓	✓	✓	✓	?
341	✓	✓	✓	✓	-	?
354	✓	✓	✓	✓	✓	✓
359	✓	✓	✓	x	x	x
360	✓	✓	✓	x	x	✓
371	✓	✓	✓	✓	x	x

Table 4 : résidus du PDZ1-2 en équilibre conformationel. Liste des résidus en équilibre conformationel dans le spectre HSQC de PDZ1-2 WT et leur état conformationel dans les mutants. Dans la nomenclature : ✓ = en équilibre; x = pic simple; - = modification de l'équilibre.

13.3 Interactants décrits de la Whirline

Protein	Domain	delim	PBM	sp.	iso	Whirl Domaim	delim	sp.	iso	Meth.	ref
CASK	Cter	702-909	PVSWVY	r.n		Pr-PDZ3	571-920	r.n		Y2H	Yap 2003
Cav1.3	Cter	1938-2181	MICITTL	h.s		PDZ1	138-233	h.s		1 pull down	Kersten 2010
Cav1.3	Cter	1938-2181	MICITTL	h.s		PDZ2	279-360	h.s		1 pull down	Kersten 2010
CIB2	Full length		STFHIRI	h.s		Full length	1-907	m.m		3 Co-IP	Riazuddin 2012
eps8	1-535			m.m		Full length				pull down	Manor 2011
Espin 3	Cter	x-871	KKGEIPK	m.m		ND1-tND2	3-472	m.m		3 Y2H	Wang 2012
LRRC4C	Cyto	549-640	VQETQI	m.m		ND1-PDZ1	1-263	m.m		4 pull down	Delprat 2005
LRRC4C	Cyto	549-640	VQETQI	m.m		PDZ2l	257-422	m.m		4 pull down	Delprat 2005
Myosin 15a	Cter Delt-ITLL	2536-3506		m.m		ND1-PDZ1	1-263	m.m		2 pull down	Delprat 2005
Myosin 15a	Cter Delt-ITLL	2536-3506		m.m		PDZ2l	257-422	m.m		2 pull down	Delprat 2005
Myosin 7a	Cter	847-2215	GSRSGK	h.s		Full length		m.m		4 pull down	Delprat 2005
p55	Cter	457-466	PVSWVY	h.s		Full length	1-907	h.s		1 Y2H	Gosens 2007
p55	Cter	323-466	PVSWVY	r.n		Pr-PDZ3	571-920	r.n		pull down	Mburu 2006
p55	Cter	335-466	PVSWVY	r.n		Pr-PDZ3	571-920	r.n		Y2H	Yap 2003
PDZD7	Full length	1-1021	LKDPHDPS	m.m		Full length	1-907	m.m		3 Co-IP	Zou 2013
RPGR ORF15	Full length	1-781 (Y2H 679-781)	LYLELK	m.m	5	PDZ1	141-216	m.m		3 pull down	Wright 2012
RPGR ORF15	Full length	1-781 (Y2H 679-781)	LYLELK	m.m	5	PDZ2	270-350	m.m		3 pull down	Wright 2012
SANS		PBM	DTEL	h.s		Full length		m.m		pull down	Sorusch 2017
SANS						PDZ1-PDZ2		h.s		1 Y2H	van Wijk 2006
TRPV1	Nter-Ank	1-433				PDZ1-PDZ2	?				Ciardo 2016
TRPV1	Nter-Ank	1-433				PDZ3	?				Ciardo 2016
USH2A	Cter	5213-5218	FTDTHL	h.s	b	Full length		m.m		4 Co-IP / pull d	Adato 2005
Ush2a	Cter		FTDTHL	m.m		Nter-PDZ1	1_247	m.m		Co-IP	Chen 2014
Ush2a	Cter		FTDTHL	m.m		PDZ2-ND2t	240-469	m.m		Co-IP	Chen 2014
Ush2a	Cter		FTDTHL	m.m		Nter-PDZ1-PDZ2	1-469	m.m		Co-IP	Chen 2014
USH2A	Cter	5064-5202	FTDTHL	h.s		PDZ1	138-233	h.s		1 pull down	van Wijk 2006
USH2A	Cter	5064-5202	FTDTHL	h.s		PDZ2	279-360	h.s		1 pull down	van Wijk 2006
USH2A	Cter	5053-5193	FTDTHL	m.m	1	ND1-tND2	3-472	m.m		3 pull down	Yang 2010
VLGR1	Cter		IADTHL	m.m		Nter-PDZ1	1_247	m.m		Co-IP	Chen 2014
VLGR1	Cter		IADTHL	m.m		Nter-PDZ1-PDZ2	1-469	m.m		Co-IP	Chen 2014
VLGR1	Cter	6157-6307	IADTHL	h.s		PDZ1	138-233	h.s		1 pull down	van Wijk 2006
Whirlin	Pr-PDZ3	443-906		m.m	4	Pr-PDZ3	443-906	m.m		4 pull down	Delprat 2005

13.4 Delhommel et al. (2015) Deciphering the unconventional peptide binding to the PDZ domain of MAST2. Biochem J. Vol :469, 159-168

View this PDF in Utopia Documents for more interactive functionality
Download FREE software now

Biochem. J. (2015) **469**, 159–168 doi:10.1042/BJ20141198

Deciphering the unconventional peptide binding to the PDZ domain of MAST2

Florent Delhommel*†‡, Alain Chaffotte*†, Elouan Terrien*†, Bertrand Raynal§†, Henri Buc‖, Muriel Delepierre*†, Florence Cordier*† and Nicolas Wolff*†[1]

*Département de Biologie Structurale et Chimie, Unité de Résonance Magnétique Nucléaire des Biomolécules, Institut Pasteur, F-75724 Paris, France
†UMR 3528, CNRS, F-75724 Paris, France
‡Université Pierre et Marie Curie, Cellule Pasteur UPMC, rue du Docteur Roux, 75015 Paris, France
§Plate-Forme de Biophysique des Macromolécules, Institut Pasteur, F-75724 Paris, France
‖Invited Scientist, Institut Pasteur F-75724 Paris, France

Phosphatase and tensin homologue (PTEN) and microtubule-associated serine threonine kinase 2 (MAST2) are key negative regulators of survival pathways in neuronal cells. The two proteins interact via the PDZ (PSD-95, Dlg1, Zo-1) domain of MAST2 (MAST2–PDZ). During infection by rabies virus, the viral glycoprotein competes with PTEN for interaction with MAST2–PDZ and promotes neuronal survival. The C-terminal PDZ-binding motifs (PBMs) of the two proteins bind similarly to MAST2–PDZ through an unconventional network of connectivity involving two anchor points. Combining stopped-flow fluorescence, analytical ultracentrifugation (AUC), microcalorimetry and NMR, we document the kinetics of interaction between endogenous and viral ligands to MAST2–PDZ as well as the dynamic and structural effects of these interactions. Viral and PTEN peptide interactions to MAST2–PDZ occur via a unique kinetic step which involves both canonical C-terminal PBM binding and N-terminal anchoring. Indirect effects induced by the PBM binding include modifications to the structure and dynamics of the PDZ dimerization surface which prevent MAST2–PDZ auto-association. Such an energetic communication between binding sites and distal surfaces in PDZ domains provides interesting clues for protein regulation overall.

Key words: intradomain energetic connectivity, PDZ domains, pre-steady-state kinetics, protein–protein interactions.

INTRODUCTION

Microtubule-associated serine threonine kinase 2 (MAST2) is a repressor of survival in neuronal cells [1]. We previously proposed that the function of MAST2 is mediated through its interaction with the phosphatase and tensin homologue (PTEN) [2], a major tumour suppressor acting on the Akt (protein kinase B) signalling pathway [3,4]. We showed that the complex formed by PTEN and MAST2 is disrupted upon rabies infection [5]. The viral glycoprotein competes with PTEN for interaction with MAST2, triggering neurosurvival, a key characteristic of rabies infection [6]. Both PTEN and viral proteins interact with the single PDZ domain of MAST2 (MAST2–PDZ) by their C-terminal sequences.

PDZ (PSD-95, post synaptic density 95; Dlg1, disk large homologue 1; Zo-1, zona occludens 1) domains are some of the most common protein–protein interaction domains. They are involved in a variety of cellular processes by assembling and regulating protein networks [7–9]. PDZ domains typically recognize C-terminal motifs of four residues called PBM (PDZ-binding motif). PDZ domains interact with PBM via a binding groove located between the PDZ strand $\beta 2$ and the helix $\alpha 2$. Two key residues of the PBM at positions 0 and -2 are responsible for the formation of an intermolecular sheet [10,11]. Even though most studies point out the crucial role of the four last residues of the PBM, the contribution of residues upstream from the canonical motif is increasingly documented in the literature [12–16].

The C-termini of PTEN and the viral glycoprotein dock on to MAST2–PDZ through a similar network of connectivity, defining an unusually large interaction surface [700 Å^2 (1 Å $= 0.1$ nm)] for a PDZ–PBM complex. This interface is composed of two distinct and non-overlapping regions (Figure 1). In addition to their canonical C-terminal motif, the peptides have an additional anchoring point involving the interaction of an aromatic residue at position -11 with a hydrophobic cluster exposed on the surface of the MAST2–PDZ $\beta 2/\beta 3$ sheet [2]. This unconventional mode of binding reinforces ligand selectivity and affinity for MAST2–PDZ. We previously showed that unliganded MAST2–PDZ domain is in self-association in solution [2]. PTEN and viral peptides exclusively interact with the monomeric form of MAST2–PDZ and prevent dimerization of the domain in solution. The molecular mechanism by which PBM binding interferes with MAST2 auto-association remains unknown.

Given the prominence of PDZ domains in biology, it is clear that understanding the molecular details of their specificity is of great significance. To decipher the role of each anchor, we dissect the interaction between MAST2–PDZ and peptides from endogenous and viral origins. Combining analytical ultracentrifugation (AUC), microcalorimetry, fluorescence and NMR, we document the dynamic and structural consequences of ligand binding to MAST2–PDZ. We show that peptide binding is driven by the canonical C-terminal PBM allowing the anchoring of a specific N-terminal determinant. As a functional consequence, our results reveal that the PBM, essential for the peptide binding,

Abbreviations: Akt, protein kinase B; AUC, analytical ultracentrifugation; $\Delta\delta$, chemical shift perturbation; Dlg1, disk large homologue 1; GRASP, GRP-1-associated scaffold protein; GRP-1, general receptor of phosphoinositides 1; IDP, intrinsically disordered protein; ITC, isothermal titration calorimetry; MAST, microtubule-associated serine threonine kinase; PBM, PDZ-binding motif; PDZ, PSD-95, Dlg1, Zo-1; PSD-95, post synaptic density 95; PTEN, phosphatase and tensin homologue; Zo-1, zona occludens 1.

[1] To whom correspondence should be addressed (email nicolas.wolff@pasteur.fr).

© 2015 Authors; published by Portland Press Limited

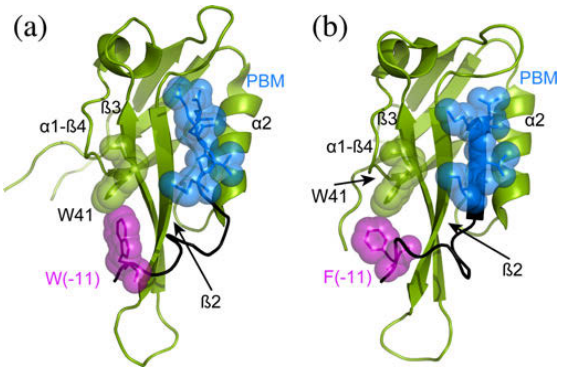


Figure 1 Structures of MAST2–PDZ in complex with $_{SW}$Cyto13$_{ETRL}$ and $_{PF}$PTEN$_{ITKV}$ peptides

Ribbon diagram of the structure of MAST2–PDZ in complex with a peptide containing the last 13 residues of the virulent form of the glycoprotein of rabies virus in (**a**) and of PTEN, one of MAST2's cellular partners in (**b**). MAST2–PDZ interacts similarly with both peptides through an unconventional network of connectivity involving two anchor points. The residues of the canonical PBM of $_{SW}$Cyto13$_{ETRL}$ and PTEN peptides are represented as blue spheres. The side chains of aromatic residues (tryptophan or phenylalanine) at position −11 are shown as purple spheres. MAST2–PDZ is represented in green (with Trp^{41} side chain as spheres).

prevents MAST2 auto-association by modifying the structure and the dynamics of the PDZ dimerization interface.

EXPERIMENTAL

Sample preparation

MAST2–PDZ, unlabelled, ^{15}N-labelled or ^{15}N/^{13}C-labelled (residues 1099–1193 renumbered as 2–96), is expressed and purified as previously described [17]. All peptides are synthesized on solid phase supports using the Fmoc (9-fluorenyl methoxycarbonyl) strategy (ProteoGenix; see Table 1 for sequences). All biophysics experiments were performed in Tris/HCl 50 mM, NaCl 150 mM, pH 7.5.

Equilibrium titration by fluorescence measurements

Static measurements of fluorescence were performed at 20 °C in a LS50 spectrofluorometer (Perkin Elmer). Excitation and emission wavelengths were set at 295 nm and 354 nm respectively, with a 10 nm bandwidth for both. One millilitre of a solution of MAST2–PDZ at 2 μM was titrated by successive additions of small aliquots from a highly concentrated peptide solution (140–5200 μM). After 2.5 min equilibration time, emission intensity was integrated for 30 s. All fittings of data were achieved with OriginPro 7.5 software (Origin Lab Corporation) using Lenvenberg–Marquardt minimization. Static fluorescence recorded data were fitted to the following equation:

$$F_i(p) = F_m R(p_0 - p) + F_p p + (F_x - F_m - F_p) \times \frac{K + Rp_0 + (1-R)p - \sqrt{[K + Rp_0 + (1-R)p]^2 - 4Rp(p_0 - p)}}{2}$$

Where p and $F_i(p)$ are the concentration of peptide and the fluorescence recorded after the i-th addition of peptide. F_m, F_p and F_x refer to specific fluorescence of MAST2–PDZ, peptide and MAST2–PDZ–peptide complex respectively, p_0 to the peptide concentration in the stock solution. R is the ratio of initial MAST2–PDZ concentration to peptide concentration in the stock solution and K is the equilibrium dissociation constant of the MAST2–PDZ–peptide interaction. F_m, R and p_0 are fixed.

Stopped-flow experiments

Kinetics of association between MAST2–PDZ and various peptides were acquired using SFM300 stopped-flow equipment with fluorescence detection (BioLogic). Typically, 150 μl of MAST2–PDZ solution were rapidly mixed with an equal volume of peptide solution. For each peptide, 3–5 concentrations were tested with a constant ratio MAST2–PDZ–peptide of 1:10 to respect pseudo first order conditions. The maximal concentration of MAST2–PDZ was 1 μM. A mixing time of 15 ms is used (total flow rate of 20 ml·s^{-1}), resulting in a 2.5 ms dead time. Data points recorded between 0 and 3 ms were excluded from the fitting process. Temperature was maintained at 20 °C by circulating water from a temperature-controlled water bath. The FC15 flow cell (0.15 × 0.15 cm) was illuminated by a monochromatic UV light at 295 nm (bandwidth 8 nm) and emitted fluorescence was collected perpendicularly to the incident beam through a 325-nm high pass filter. Kinetics was acquired on 0–0.5 s time scale with a 0.1-ms sampling period and a time constant of 0.1 ms (5001 data points). Series of 10 successive shots were recorded and averaged. Depending on the signal to noise ratio, 4–32 averaged traces were stored and processed independently. Controls of initial fluorescence intensity were performed under strictly identical conditions by recording fluorescence contributions of MAST2–PDZ and peptides separately upon 1:1 mixing with buffer. Averaged kinetic traces from stopped-flow experiments were fitted independently to the following mono-exponential equation: $F(t) = A \cdot e^{-kobs \cdot t}$ + offset. For each peptide concentration, the mean values of amplitude, the observed rate constant and the related S.D.s were calculated from the fitted values deduced from each series of average kinetics. Amplitudes deduced from the exponential fits were normalized to the concentration of MAST2–PDZ and variation of the resulting reduced amplitude with peptide concentration was fitted to the following equation:

$$A(p) = F \frac{\left[K_{app}\frac{R}{p} + R + 1 - \sqrt{\left(K_{app}\frac{R}{p} + R + 1\right)^2 - 4R}\right]}{2}$$

Where p is the peptide concentration; F, the maximum reduced amplitude; K_{app}, the apparent dissociation constant; and R, the ratio of MAST2–PDZ to peptide concentrations. R was maintained constant for a given peptide assay. For each peptide, intrinsic association and dissociation rate constants, k_{on} and k_{off}, were determined from the linear regression of the variation of k_{obs} as a function of peptide concentration.

Analytical ultracentrifugation

Sedimentation coefficients were determined using a Beckman Coulter XL-I centrifuge equipped with an AN60-Ti rotor. Samples were prepared by mixing MAST2–PDZ (100 μM) and peptides (200 μM). Samples were centrifuged for 17 h at 262 000 ***g***. Sedimentation scans were analysed by the continuous size distribution procedure c(S) with the program Sedfit 12.0 (available at analyticalultracentrifugation.com). All c(S) distributions were calculated with a fitted frictional ratio f/f0 and a maximum entropy regularization procedure with a confidence level of 0.95.

© 2015 Authors; published by Portland Press Limited

Table 1 Equilibrium (K_d) and intrinsic rate constants (k_{on}, k_{off}) of peptides with varying anchors on MAST2–PDZ

The apparent equilibrium dissociation constants K_{dapp} are deduced from the kinetic amplitudes. K_d values are obtained by titration monitored by static fluorescence. The K_d value marked with (*) ($_{SW}$PTEN$_{ITKV}$) has been deduced from ITC experiment.

Peptide	Sequence	k_{off} (s^{-1})	k_{on} ($s^{-1}\cdot\mu M^{-1}$)	k_{off}/k_{on} (μM)	$K_{d\,app}$ (μM)	Equilibrium titration K_d (μM)
$_{Ac}$Cyto4$_{ETRL}$	− − −ETRL	29.7 ± 0.2	12.7 ± 0.1	2.3 ± 0.1	1.5	1.9 ± 0.1
$_{Ac}$Cyto4 $_{QTRL}$	− − −QTRL	34.7 ± 0.4	9.5 ± 0.1	3.6 ± 0.1	4.2	3.2 ± 0.1
$_{SA}$Cyto13$_{ETRL}$	SAESHKSGGETRL	18.4 ± 0.7	7.7 ± 0.1	2.4 ± 0.1	1.2	1.5 ± 0.1
$_{SA}$Cyto13$_{QTRL}$	SAESHKSGGQTRL	23.0 ± 0.5	7.3 ± 0.2	3.2 ± 0.1	3.5	4.8 ± 0.7
$_{SW}$Cyto13$_{ETRL}$	SWESHKSGGETRL	9.4 ± 0.9	14.9 ± 0.3	0.6 ± 0.1	0.4	0.4 ± 0.1
$_{SW}$Cyto13$_{QTRL}$	SWESHKSGGQTRL	16.6 ± 0.7	8.5 ± 0.2	1.9 ± 0.1	0.7	1.5 ± 0.2
$_{PF}$PTEN$_{QTRL}$	PFDEDQHTQQTRL	6.4 ± 1.1	8.5 ± 0.2	0.8 ± 0.1	0.5	0.7 ± 0.1
$_{PF}$PTEN$_{ITKV}$	PFDEDQHTQITKV	9.7 ± 0.5	5.6 ± 0.1	1.7 ± 0.1	1.4	2.1 ± 0.1
$_{SW}$PTEN$_{ITKV}$	SWDEDQHTQITKV	7.5 ± 0.9	8.3 ± 0.2	0.9 ± 0.1	1.0	1.6*
$_{SW}$Cyto13Δ4	SWESHKSGG− −	Not measurable	Not measurable	-	Not measurable	Not measurable

NMR spectrometry

NMR spectra were recorded at 25 °C on a Varian Inova 600 MHz spectrometer (Agilent Technologies) equipped with a cryoprobe. ^{1}H-^{15}N HSQC spectra were recorded using 120 μM ^{15}N-labelled MAST2–PDZ and 700 μM peptide with 12 % $^{2}H_2O$. Spectra of apo MAST2–PDZ were obtained at a concentration of 400 μM. The ^{15}N relaxation times T_1 and T_2 and $\{^{1}H\}$-^{15}N heteronuclear NOE were measured by standard methods [18]. The spectra were recorded in an interleaved manner with six relaxation delays, from 10 to 1200 ms and from 10 to 110 ms for T_1 and T_2 experiments respectively.

RESULTS AND DISCUSSION

Both N-terminal and C-terminal extremities of peptide contribute to MAST2–PDZ binding

We previously studied the interaction of MAST2–PDZ with the C-terminus of its cellular partner PTEN and of viral rabies glycoproteins using 13-residue-long C-terminally-derived peptides. The peptides showed similar affinities in the micromolar range and shared a similar binding mode [2]. We dissect here the molecular determinants that allow ligands to bind to MAST2–PDZ. Several modified versions of the viral and endogenous sequences were used to investigate the contribution of each peptidic region to the affinity and specificity (Table 1).

The C-terminal sequences of viral glycoproteins, $_{SW}$Cyto13$_{ETRL}$ (from attenuated strain) and $_{SW}$Cyto13$_{QTRL}$ (from virulent strain) and of PTEN ($_{PF}$PTEN$_{ITKV}$) bind MAST2–PDZ with K_d of 0.4, 1.5 and 2.1 μM respectively, according to equilibrium titrations (Table 1). No binding of $_{sw}$Cyto13Δ4 (viral sequence deprived of its PBM) to MAST2–PDZ could be detected by ITC (isothermal titration calorimetry) [2] or by static fluorescence (Table 1), which shows that the canonical PBM is essential for the peptide-binding affinity. We then characterized the binding of 'complementary' sequences by testing the N-acetylated forms of the short PBM peptides $_{Ac}$Cyto4$_{ETRL}$ and $_{Ac}$Cyto4$_{QTRL}$. N-acetylation prevents the effects of the N-terminal positive charge [12]. Note that MAST2–PDZ binds $_{Ac}$Cyto4$_{ETRL}$ 5-fold more tightly than the non-acetylated form, $_{NH3}$Cyto4$_{ETRL}$ (result not shown), indicating that the N-terminal charge significantly disturbs the interaction between $_{NH3}$Cyto4 $_{ETRL}$ and MAST2–PDZ. The affinities displayed by MAST2–PDZ for $_{Ac}$Cyto4$_{ETRL}$ (1.9 μM) and $_{Ac}$Cyto4$_{QTRL}$ (3.2 μM) are close to the ones observed for two viral peptides containing a W(−11)A substitution, $_{SA}$Cyto13 $_{ETRL}$ (1.5 μM) and $_{SA}$Cyto13 $_{QTRL}$ (4.8 μM). Then, the W(−11)A substitution included in the natural viral sequences significantly decrease the overall affinities of these two peptides: $_{SW}$Cyto13s affinities are 2–5-fold higher than those of $_{Ac}$Cyto4s or $_{SA}$Cyto13s. This result corroborates the intermolecular NOE associated with the N-terminal anchor, which almost exclusively involve the tryptophan residue W(−11) (80 % of the N-terminal anchor intermolecular NOE; [2]). It seems therefore safe to conclude that in the presence of an efficient C-terminal anchor (QTRL or better ETRL), the N-terminal moiety of the peptides is only safely anchored if the tryptophan residue is present at position −11. We previously observed that viral peptides (44-residues long) containing Cyto13 with an N-terminal extension have an affinity ∼3-fold higher than that of shorter Cyto13s peptides for MAST2–PDZ [5]. In this case, the N-terminal extension most probably stabilizes the interaction of the aromatic residue W(−11) with MAST2–PDZ.

The canonical C-terminal anchor appears crucial for the interaction, being the main contributor to the peptide-binding affinity. An efficient N-terminal anchor most probably plays a crucial role in narrowing the specificity of interaction between MAST2 and its partners, since this type of interaction with the long $\beta2/\beta3$ loop of MAST2–PDZ has been reported for only one other PDZ domain [16].

The C-terminal binding drives the N-terminal anchoring in an overall one-step process

To further explore the relative contribution of the N- and C-terminal sequences of these peptides, we analyse binding kinetics at different concentrations of MAST2–PDZ and peptide using stopped-flow tryptophan fluorescence. To simplify the data processing, we apply pseudo first order conditions with a constant MAST2–PDZ–peptide ratio of 1:10. Under these conditions, interactions of all peptides, containing one or two anchor points, with MAST2–PDZ can satisfactorily be fitted to a single exponential decay on the 0–0.5 s interval. In every case, the observed rate constant and the amplitude both vary with peptide concentration. Figure 2 illustrates two examples of peptide–MAST2–PDZ association. In the case of $_{SA}$Cyto13$_{ETRL}$ (Figure 2a), fluorescence decreases with time. This decrease was systematically observed with long and short peptides containing the viral ETRL or QTRL C-terminal motifs. On the other hand, a time-dependent increase in fluorescence is observed with peptides terminated by the ITKV C-terminal motif (Figure 2b). For the 10 peptides studied, the variation of the observed rate constant with the peptide concentration fits to the linear function: $k_{obs} = k_{on}.[peptide] + k_{off}$. Therefore, the rate-limiting step

© 2015 Authors; published by Portland Press Limited

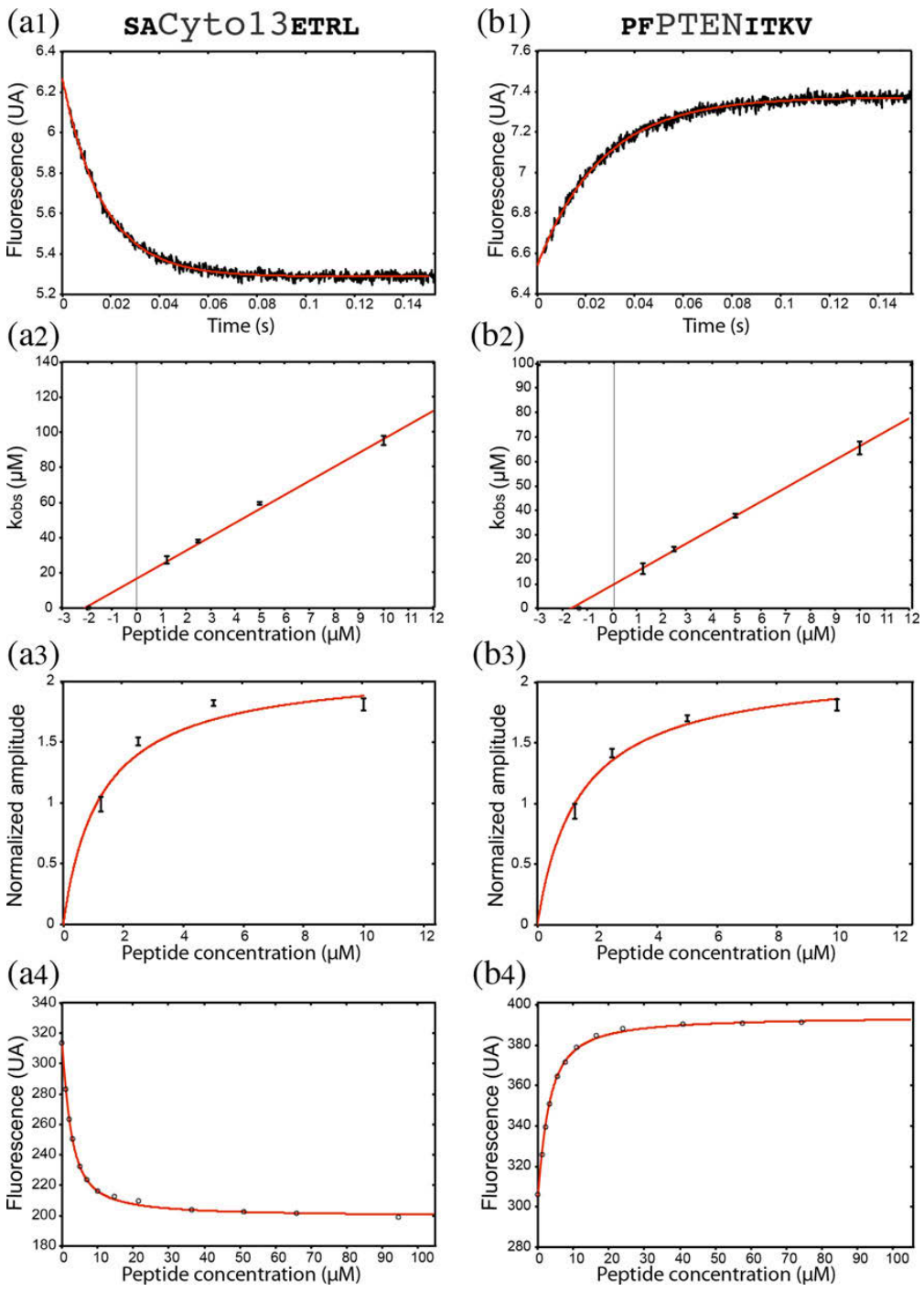


Figure 2 Kinetic and equilibrium binding of $_{SA}$Cyto13$_{ETRL}$ (a) and $_{PF}$PTEN$_{ITKV}$ (b) to MAST2–PDZ

(**a1** and **b1**) The averaged kinetic recordings of the interaction between MAST2–PDZ at 0.5 μM and $_{SA}$Cyto13$_{ETRL}$ at 5 μM or $_{PF}$PTEN$_{ITKV}$ at 5 μM (see 'Experimental' section for detailed experimental conditions). (**a2** and **b2**) Variation of k_{obs} with peptide concentration. (**a3** and **b3**) peptide concentration dependence of the kinetic amplitude normalized to the concentration of MAST2–PDZ. (**a4** and **b4**) Variation of fluorescence of MAST2–PDZ upon titration by $_{SA}$Cyto13$_{ETRL}$ and $_{PF}$PTEN$_{ITKV}$.

© 2015 Authors; published by Portland Press Limited

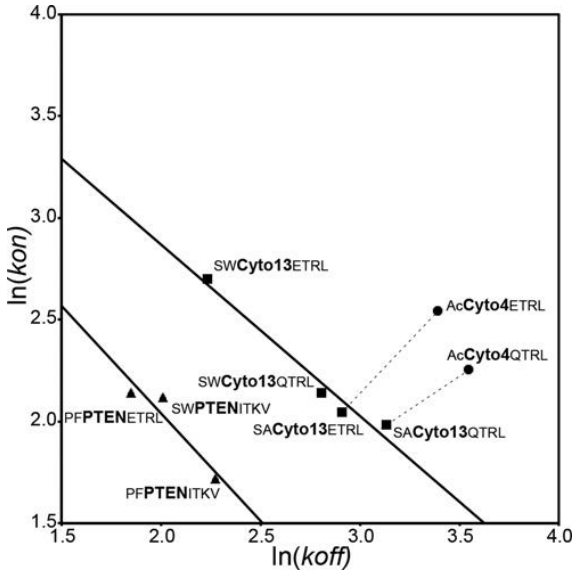


Figure 3 Log–log plot giving the variation of intrinsic association and dissociation rate constants with the sequence of peptides

In this representation, isoenergetic variations correspond to lines displaying a slope of $+1$. If the difference in free energy of binding between various partners affects, by the same factor, both the 'on' constant and the residence time ($1/k_{off}$), then the corresponding array of points will be aligned with a slope of -1. Such is apparently the case for the two sets of points characterizing the insertion of different C- or N-terminal sequences at each end of a linker corresponding either to the viral sequence (Cyto13s, in squares) or to the PTEN sequence (PTEN, in triangles). Points reporting the kinetic parameters of the two tetrapeptides Cyto4s (in circles) are located above the two continuous lines. Dotted lines indicate the changes observed when Cyto4s peptides are extended to tridecapeptides $_{SA}$Cyto13s.

detected by the fluorescence time course reflects a second order process, the binding of the peptide to MAST2–PDZ. The sign of the exponential amplitude strictly depends on the E/QTRL or ITKV motif, indicating that the binding event concerns as a priority the canonical C-terminal PBM sequence.

There are two implications of this simple behaviour: first, the apparent dissociation constant deduced from the variation of the kinetic amplitude normalized by MAST2–PDZ concentration must coincide with the back extrapolation of the linear plot of k_{obs} compared with peptide concentration. Second, the ratio k_{off}/k_{on} should match the equilibrium dissociation constant deduced independently by equilibrium titration of MAST2–PDZ with the same peptide. As shown in Figure 2, the back extrapolation to the abscissa axis of the linear plot of k_{obs} as a function of peptide concentration agrees well with both the dissociation constant deduced from the normalized amplitude and the equilibrium dissociation constant based on equilibrium titration. This is true for any of the MAST2–PDZ–peptide pairs tested (Table 1).

This set of binding data validates a simple two-state model for the anchoring of the C- and the N-terminal moieties of the peptides tested. The C-terminal anchoring event corresponds to the majority of the fluorescence change observed between isolated and assembled partners. The participation of the N-terminal sequence manifested itself through an increase of 20 %–25 % in the normalized amplitude when an A($-$11)W substitution is made in the case of $_{SW}$Cyto13$_{ETRL}$ and $_{SW}$Cyto13$_{QTRL}$. Therefore, in the concentration range explored for the peptide, the fact that the N-terminal anchoring does not give rise to a distinct subsequent phase means that it proceeds much more rapidly than the bimolecular event implying the C-terminal motif E/QTRL or ITKV.

Figure 3 shows values of k_{on} and k_{off}, plotted against each other on a double logarithmic scale. This representation allows a simple description of the effects of all the sequence modifications for the various peptides we have studied up to now. We first look at the effects of the lengthening of the strict C-terminal PBM by addition of an N-terminal sequence devoid of an anchor by comparing the peptides $_{Ac}$Cyto4$_{ETRL}$ and $_{Ac}$Cyto4$_{QTRL}$ to $_{SA}$Cyto13$_{ETRL}$ and $_{SA}$Cyto13$_{QTRL}$. Binding clearly proceeds without a significant change in the free energy of association. The 'on' and 'off' kinetic constants are affected by the same factor for each pair of peptide when the N-terminal sequence is added. The insertion of an aromatic residue A($-$11)W at the N-terminal of the Cyto13 sequences or the Q($-$3)E change at the C-terminal, have similar effects on the two kinetics constants. All the points represented for Cyto13s are on a line having a slope close to -1, implying that a modification of the peptide in either the N-terminal or the C-terminal anchor proportionally affects the residence time ($1/k_{off}$) and the 'on' constant roughly in the same proportion. When we analyse the kinetics behaviour of the sequences derived from PTEN, the effects of modifications on both the N-terminal and the C-terminal anchors display the same trend as observed before for the sequences derived from Cyto13. They affect the residence time ($1/k_{off}$) and 'on' constant by the same factor. However, the line representative of these PTENs points is shifted from the Cyto13s one by a significant amount: -0.3 kcal·mol^{-1} (1 kcal $\equiv$ 4184 J), as displayed on the graph. We suspect that this effect is due to the differences in the linker displayed by the two families of sequences. The Cyto13s sequences contain two glycines next to the canonical PBM that are absent from PTEN sequences. These two residues most probably increase the flexibility of the linker region between the two anchors favouring the exploration of the N-terminal, hence explaining the better affinities of Cyto13s peptides compared with PTENs. Our results are consistent with the extensive work of Huang and Liu [19] on the comparison of the binding kinetics of intrinsically disordered proteins (IDPs) and folded proteins. In their analysis, IDPs show higher k_{on} and k_{off} than folded protein, for a similar K_d, illustrating the effect of flexibility on kinetic constants.

Long-range structural and dynamic perturbations are induced by N-terminal anchoring

To investigate the structural and dynamic effects of peptide interactions with MAST2–PDZ, we recorded NMR ^{1}H-^{15}N HSQC spectra and measured backbone ^{15}N relaxation data. We first compared the ^{1}H and ^{15}N chemical shifts of MAST2–PDZ in complex with $_{SW}$Cyto13$_{ETRL}$ or with $_{Ac}$Cyto4$_{ETRL}$. Based on chemical shift perturbations (as $\Delta\delta$), we identified MAST2–PDZ structural modifications caused by the N-terminal anchor upon peptide binding (Figure 4a). Out of the 96 amino acids in the MAST2–PDZ construct, we were able to assign 85 in the MAST2–PDZ–$_{SW}$Cyto13$_{ETRL}$ spectrum and 80 on the MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ spectrum by direct comparison of the ^{1}H-^{15}N HSQC spectra. Spectra of both complexes are highly similar, indicating that the global fold and secondary structures of MAST2–PDZ are conserved upon binding of the two peptides. The ^{1}H-^{15}N $\Delta\delta$ values in the GLGF conserved loop of MAST2–PDZ (residues K^{16}–F^{19}) are weak, indicating a similar C-terminal-binding mechanism for both complexes. As expected, the structural variations of MAST2–PDZ induced by the binding of $_{SW}$Cyto13$_{ETRL}$ compared with that of $_{Ac}$Cyto4$_{ETRL}$ are mainly observed near the N-terminal

© 2015 Authors; published by Portland Press Limited

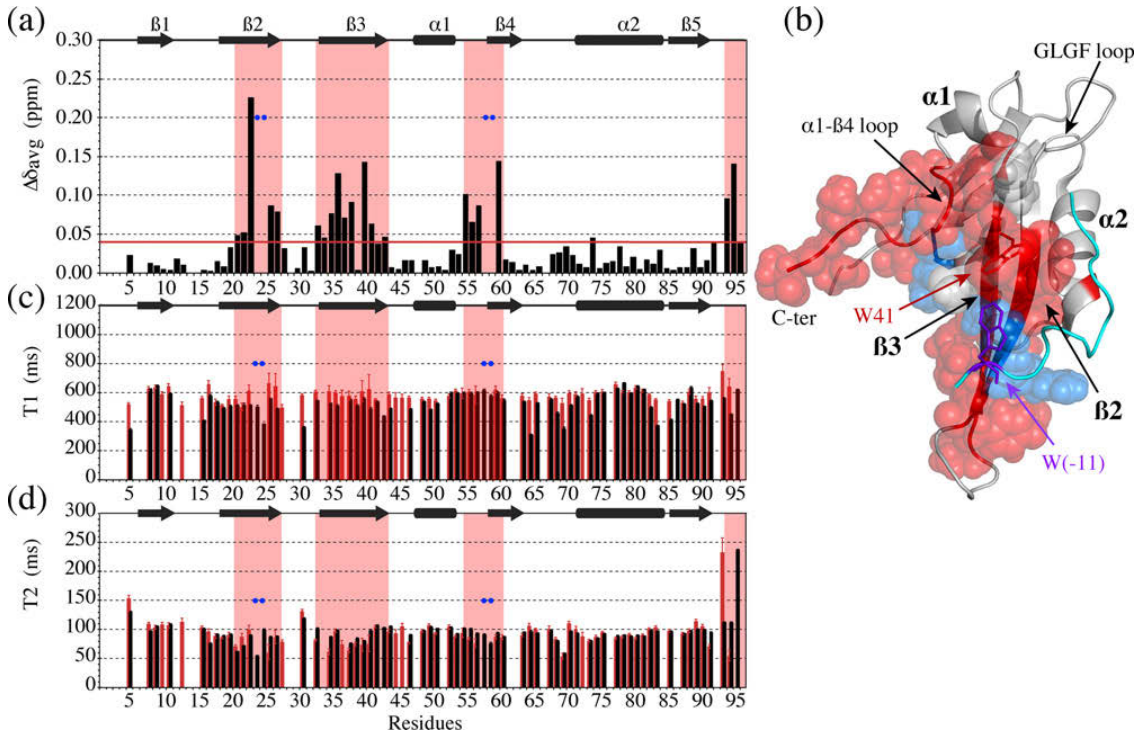


Figure 4 $\Delta\delta$ values and relaxation times differences between MAST2–PDZ–$_{SW}$Cyto13$_{ETRL}$ and MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$

(**a**) $\Delta\delta$ values (^{1}H,^{15}N) computed as $\Delta\delta = [(\Delta\delta H)^2 + (\Delta\delta N*0.159)^2]^{1/2}$. The red line represents the threshold used to determine perturbed region with $\Delta\delta > 0.04$ ppm; blue circles mark residues with no resonances in the MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ spectrum. (**b**) MAST2–PDZ–$_{SW}$Cyto13$_{ETRL}$ structure (PDB: 2KQF) with different colours depending on $\Delta\delta$ values. Grey: $\Delta\delta < 0.04$ ppm, red $\Delta\delta > 0.04$ ppm, blue: resonance absent from the MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ spectrum. Cyto13 is represented in cyan and the N-terminal anchor W(−11) is represented in purple. Side chains represented as spheres show a potential energy pathway between MAST2–PDZ $\beta 2$ strand and the C-terminal end. (**c**) Longitudinal relaxation times (T_1) of MAST2–PDZ–$_{SW}$Cyto13$_{ETRL}$ (black) and MAST2–PDZ/$_{Ac}$Cyto4$_{ETRL}$ (red). (**d**) Transverse relaxation times (T_2) of MAST2–PDZ–$_{SW}$Cyto13$_{ETRL}$ (black) and MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ (red).

anchor. Thus, the largest $\Delta\delta$ values ($\Delta\delta > 0.04$ ppm) are observed for MAST2–PDZ residues directly contacting the N-terminal region of the peptide, on the $\beta 2$ strand (residues L^{21}–M^{27}) and on the $\beta 3$ strand (residues V^{33}–W^{41}) (Figure 4b). Notably, the N-terminal anchor binds directly to the $\beta 3$ strand by an edge-to-face interaction of its tryptophan residue W(−11) with MAST2–PDZ W^{41} (PDB: 2KQL). Peaks of residues Ile^{24} and Arg^{25} are missing in the MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ spectrum. These residues of MAST2–PDZ are in the closest proximity to W(−11) of $_{SW}$Cyto13$_{ETRL}$. A microsecond–millisecond conformational exchange regime probably broadens the peaks of the $\beta 2$ strand beyond detection due to the missing N-terminal anchor.

Interestingly, in addition to these local perturbations, residues distal from the PBM-binding site are also affected. The resonances of the residues forming the $\alpha 1/\beta 4$ loop (residues R^{55}–I^{60}) and the C-terminal residues Leu^{94} and Glu^{95} of MAST2–PDZ are affected by the absence of the peptide N-terminal anchor in terms of either their shifts (Arg^{55}, Gln^{56}, Gly^{57}, Ile^{60}) or their intensities (Asp^{58}, Leu^{59}). These long-range perturbations are probably due to a re-orientation of the Trp^{41} side chain in the absence of the N-terminal anchor (Figure 4b). Trp^{41} is in direct contact with the side chain of Gln^{56}. This perturbation is most probably transmitted to MAST2–PDZ C-terminal residues Leu^{94} and Glu^{95} via the side chain of Arg^{55} in a 'domino-like' manner. The resonances of residues Asp^{58} and Leu^{59} are missing in the MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ spectrum indicating that the $\alpha 1/\beta 4$ loop is also affected by the PBM binding at the distal binding site.

To probe and compare the local dynamics of MAST2–PDZ complexed with the $_{SW}$Cyto13$_{ETRL}$ and $_{Ac}$Cyto4$_{ETRL}$ peptides, the ^{1}H-^{15}N longitudinal (T_1) and transverse (T_2) relaxation times of the domain were measured at 25 °C. The backbone ^{1}H-^{15}N heteronuclear NOE were also recorded but no significant differences could be observed between the two complexes (result not shown). The average T_1 values are 530 ± 79 ms for MAST2–PDZ–$_{SW}$Cyto13$_{ETRL}$ and 587 ± 44 ms for MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ (Figure 4c). The average T_2 values (Figure 4d) are 96 ± 22 ms for MAST2–PDZ–$_{SW}$Cyto13$_{ETRL}$ and 94 ± 24 ms for MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$. Backbone dynamics analysed from ^{1}H-^{15}N relaxation measurements revealed small but significant local changes in the dynamic behaviour of the two $_{SW}$Cyto13$_{ETRL}$ and $_{Ac}$Cyto4$_{ETRL}$ complexes. The most affected regions are the $\beta 2 - \beta 3$ loop and the C-terminal end and to a lesser extent the $\alpha 2 - \beta 5$ loop and the $\alpha 1 - \beta 4$ loop. As observed with the largest $\Delta\delta$ induced by the N-terminal anchor on MAST2–PDZ, the dynamical perturbations correspond to residues in the closest proximity to the N-terminal anchor but also to distal residues. In addition, these regions of MAST2–PDZ complexed to $_{SW}$Cyto13$_{ETRL}$ have decreased ^{1}H-^{15}N T_1 values and increased ^{1}H-^{15}N T_2 values in comparison with the values for MAST2–PDZ in complex with $_{Ac}$Cyto4$_{ETRL}$. This observation unambiguously shows that the N-terminal anchor of the PDZ ligand triggers

© 2015 Authors; published by Portland Press Limited

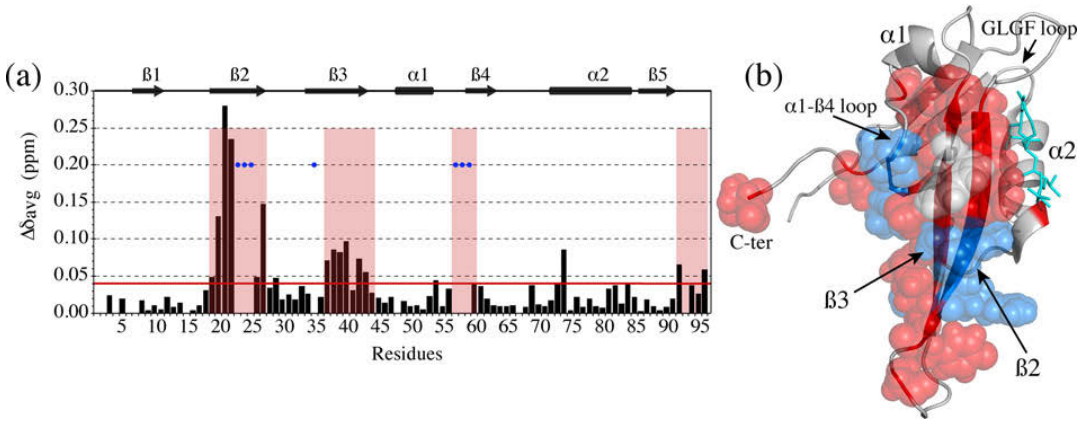


Figure 5 $\Delta\delta$ between MAST2 – PDZ–$_{Ac}$Cyto4$_{ETRL}$ and MAST2 – PDZ–$_{NH3}$Cyto4$_{ETRL}$

(**a**) $\Delta\delta$ values (^{1}H,^{15}N) computed as $\Delta\delta = [(\Delta\delta H)^2 + (\Delta\delta N^{*}0.159)^2]^{1/2}$. The red line represents the threshold used to determine the perturbed region with $\Delta\delta > 0.04$ ppm; blue circles mark residues with no resonance in MAST2 – PDZ–$_{NH3}$Cyto4$_{ETRL}$. (**b**) MAST2 – PDZ–$_{SW}$Cyto13$_{ETRL}$ structure (PDB: 2KQF) with different colours depending on $\Delta\delta$ values. Grey: $\Delta\delta < 0.04$ ppm, red $\Delta\delta > 0.04$ ppm, blue: resonances absent from the MAST2–PDZ–$_{NH3}$Cyto4$_{ETRL}$ spectrum. $_{Ac}$Cyto4$_{ETRL}$ equivalent is represented in cyan. Side chains represented in spheres show a potential energy pathway between MAST2–PDZ β2 strand and the C-terminal end.

significant changes in the conformation and dynamics at several local and distal regions of MAST2–PDZ in the fast picosecond-to-nanosecond timescale.

Long-range perturbations are also induced by modification of the canonical PBM

To study whether perturbations in MAST2–PDZ could be induced by other regions of the peptide, we compared the chemical shifts of MAST2–PDZ in complex with two forms of the C-terminal anchor. We looked at both the acetylated form of the short C-terminal peptide ($_{Ac}$Cyto4$_{ETRL}$) that mimics the PBM in a longer peptide chain and the non-acetylated form ($_{NH3}$Cyto4$_{ETRL}$). In the latter, the positive charge of the unprotected N-terminal amino group decreases the affinity of the peptide for MAST2–PDZ by a factor of five.

Similar to the chemical shift differences between $_{SW}$Cyto13$_{ETRL}$ and $_{Ac}$Cyto4$_{ETRL}$ complexes, the MAST2–PDZ β2 and β3 strands contain the most shifted resonances. All residues within the β2 strand (residues F^{19}–Y^{27}) either have a $\Delta\delta > 0.04$ ppm or are absent from the MAST2–PDZ–$_{NH3}$Cyto4$_{ETRL}$ spectrum (Figure 5). The microsecond–millisecond intermediate time-scale dynamics of β2 strand observed for MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ is exacerbated in the MAST2–PDZ–$_{NH3}$Cyto4$_{ETRL}$ complex with an additional missing resonance, namely that of residue Ala^{23}. Larger $\Delta\delta$ values on the β2 strand are observed compared with those induced by the binding of the N-terminal anchor. The residues His^{37} to Val^{43} of the β3 strand are also affected by the amino-charge of $_{Ac}$Cyto4$_{ETRL}$ ($\Delta\delta > 0.04$ ppm) but to a lesser extent compared with the effect of the N-terminal binding. In addition, perturbations in the $\alpha1/\beta4$ loop and MAST2–PDZ C-terminus are clearly observed. Gly^{57} from the $\alpha1/\beta4$ loop is even absent from the MAST2–PDZ–$_{NH3}$Cyto4$_{ETRL}$ spectrum, a consequence of microsecond-millisecond conformational exchange.

Overall, we observe that modification of the charge at peptide position (−3) induces long-range perturbations in the MAST2–PDZ similar to the effects of N-terminal anchoring. This result is consistent with the ^{1}H-$\Delta\delta$ we previously observed between the complexes of MAST2–PDZ with either $_{SW}$Cyto13$_{ETRL}$ or $_{SW}$Cyto13$_{QTRL}$ [2]. The conservative mutation E/Q is sufficient to induce perturbations that spread from the binding groove toward the $\alpha1/\beta4$ loop of MAST2–PDZ.

Both anchors contribute to the prevention of MAST2–PDZ dimerization

We previously observed that at high concentration in solution, both monomeric and multimeric forms of MAST2–PDZ coexist [2]. To characterize this self-association, we performed ITC dilution experiments of MAST2–PDZ. We obtained a K_d of 15 ± 7 μM (result not shown), consistent with the equilibrium dissociation constant of 29 ± 3 μM determined by ultracentrifugation [2]. The Gibbs free energy change associated with MAST2–PDZ dissociation is mainly driven by a large enthalpy contribution (13 kcal·mol^{-1}). We also demonstrated that the peptides $_{PF}$PTEN$_{ITKV}$ and $_{SW}$Cyto13$_{ETRL}$ promote full dissociation of MAST2–PDZ by binding exclusively to its monomeric form. We initially proposed that the N-terminal anchor of the PDZ ligand precludes the self-association of MAST2–PDZ by occluding the dimerization area of MAST2–PDZ. This hypothesis was supported by the fact that the N-terminal anchor of the peptides interacts with a hydrophobic cluster that is conserved in two closely related proteins, MAST1–PDZ and MAST3–PDZ. This patch located at the surface of the β2 and β3 strands is involved in the dimerization surface of MAST1–PDZ and MAST3–PDZ X-ray structures (PDB: 3PS4 and 3KHF respectively).

To assess whether this hypothesis is valid for all regions of the peptides, we monitored the oligomeric state of MAST2–PDZ in the presence of various peptides, with or without the N-terminal anchor. Our ultracentrifugation results (Table 2) show that, regardless of the N-terminal anchor, each peptide is equally capable of maintaining MAST2–PDZ in its monomeric form with sedimentation coefficients (S_{20w}) between 1.59 and 1.66 S $\pm$ 0.1 (2.0 S $\pm$ 0.2 for the dimeric form of MAST2–PDZ in the absence of PDZ ligand). Thus, the canonical PBM is sufficient to

© 2015 Authors; published by Portland Press Limited

Table 2 Sedimentation coefficients of MAST2–PDZ domain alone or in complex with various peptides

Sedimentation coefficients of MAST2–PDZ (100 μM) in the presence of peptides (200 μM) containing either one or two anchors are similar whereas the sedimentation coefficient of MAST2–PDZ alone is higher and consistent with a dimeric state.

Complex	Peptide sequence	Sedimentation coefficient (S)
MAST2–PDZ/–	–	2.07
MAST2–PDZ/$_{SW}$Cyto13$_{ETRL}$	SWESHKSGGETRL	1.59
MAST2–PDZ/$_{SA}$Cyto13$_{ETRL}$	SAESHKSGGETRL	1.66
MAST2–PDZ/$_{Ac}$Cyto4$_{ETRL}$	– – –ETRL	1.66

prevent MAST2–PDZ from dimerizing. Compared with the dimer structures of others PDZ domains of the MAST family, MAST1, MAST3 and MAST4, the short $_{Ac}$Cyto4$_{ETRL}$ peptide does not mask the corresponding dimerization surfaces on MAST2–PDZ. The dimerization of MAST2–PDZ is thus probably hampered by distal structural and dynamical changes induced by $_{Ac}$Cyto4$_{ETRL}$.

^{1}H-^{15}N HSQC spectrum of uncomplexed MAST2–PDZ at high concentration is consistent with MAST1–PDZ and MAST3–PDZ dimeric structures

To determine the PDZ/PDZ interface of MAST2, we recorded ^{1}H-^{15}N HSQC spectra of the uncomplexed MAST2–PDZ at high concentration (400 μM). On the basis of the monomer–dimer dissociation constant set at 15 μM, the dimeric form represents 96 % of the total population at this concentration. The resonances were broad and weak; a total of 57 out of the 91 expected resonances (for non-proline residues) were observed. We were able to assign 36 of them by direct comparison with the ^{1}H-^{15}N HSQC spectrum of the monomeric MAST2–PDZ in complex with $_{Ac}$Cyto4$_{ETRL}$. To determine the regions affected by dimerization, we compared the chemical shifts collected on the two spectra (Figure 6a). Sixteen have no chemical shift changes ($\Delta\delta < 0.08$ ppm) and 20 resonances are shifted ($\Delta\delta > 0.08$ ppm) in the spectrum. These $\Delta\delta$ could be due to dimerization and/or ligand binding. In addition, we identified 17 of the resonances that are absent from the dimer spectrum but are present in the MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ spectrum. The 16 assigned residues with $\Delta\delta < 0.08$ ppm are essentially located on the surface opposite to the PBM-binding groove (Figure 6a) and correspond to residues unaffected by both peptide binding and dimerization. This surface has been reported as the dimer interface for other PDZ [20,21], involving the formation of an intermolecular sheet via the β1 strand. Our data rule out this type of auto-association for MAST2–PDZ. We propose that the 17 missing resonances are due to the monomer–dimer exchange and could help us to delimit the interface of MAST2/PDZ auto-association. Out of 17 missing resonances, nine are located in the β2 strand and the β2/β3 loop between residues Arg^{22} and Ser^{31}. Out of those residues, only the Arg^{22} is in direct contact with the peptide [17]. Three residues of the α1/β4 loop and the β4 strand are also missing, Gly^{57}, Asp^{58}, Leu^{59}. Five residues (Gly^{14}, Tyr^{17}, Gly^{18}, Asp^{45} and Ser^{50}) in close contact are affected by exchange on the top part of the peptide-binding groove.

The dimeric structures of MAST1–PDZ (PDB: 3PS4) and MAST3–PDZ (PDB: 3KHF) were stabilized through the *cis* addition of a PBM at the C-terminal end of the PDZ domain, with each protomer binding to the PBM of the other. The surface of auto-association deduced from our NMR data overlaps the ones of liganded MAST1–PDZ, MAST3–PDZ and NHERF4

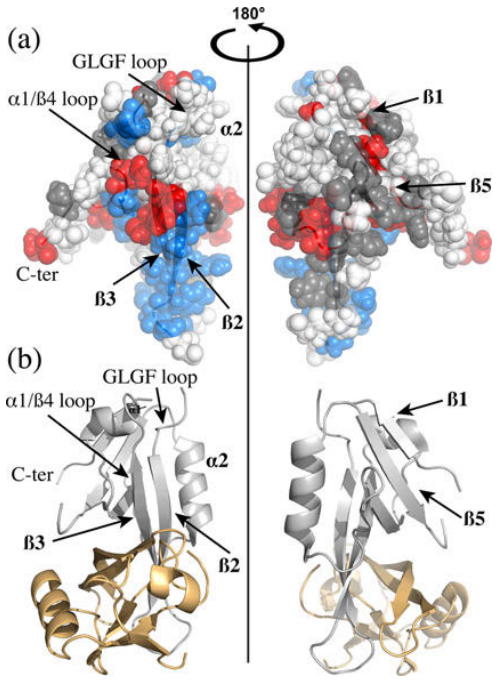


Figure 6 $\Delta\delta$ values between MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ and MAST2–PDZ in dimeric form, compared with MAST3–PDZ dimer structure

(**a**) MAST2–PDZ–$_{SW}$Cyto13$_{ETRL}$ structure coloured according to $\Delta\delta$ values between MAST2–PDZ–$_{Ac}$Cyto4$_{ETRL}$ and MAST2–PDZ dimer. White sphere: residues without assignment; grey sphere: the 16 residues with $\Delta\delta < 0.08$ ppm; red sphere: the 20 residues with $\Delta\delta > 0.08$ ppm; blue sphere: the 17 residues identified as absent from MAST2–PDZ dimer spectrum. (**b**) MAST3–PDZ dimer structure (PDB: 3KHF). Protomers interact mainly by their β2/β3 sheet, their α1/β4 loop and their C-terminus.

(Na^{+}/H^{+} exchange regulatory factor 4)–PDZ3 (PDB: 2V90) as well as of unliganded GRASP (GRP-1-associated scaffold protein)–PDZ (PDB: 2EGK). All these proteins display large β2-β3 strands sharing an exposed hydrophobic patch. We conclude that these PDZ domains could auto-associate following a similar mechanism.

CONCLUSIONS

In the present study, we report that the unconventional binding of peptide ligands to MAST2–PDZ controls the auto-association of this domain. Both viral and endogenous PDZ ligands bind to MAST2–PDZ in a similar manner, through two distinct anchoring points. By thoroughly analysing a set of peptides with different anchors, we showed that specific peptide binding necessarily implies their canonical C-terminal sequence, the major contributor to peptide affinities. The N-terminal sequence participates in the same binding event when it carries a specific aromatic side chain interacting with the β2-β3 hydrophobic cluster of the domain. However, our results differ from those found by Chi et al. [22] on the binding of a nine residue viral peptide on to SAP97 (synapse-associated protein 97)–PDZ2 [22]. They described a two-step process initiated by an unspecific interaction between the peptide and the PDZ domain surface leading to an isomerization connecting the PBM to the binding groove. Our combined results demonstrate the variety of mechanisms of interaction that domains such as PDZ can have. In the case of MAST2–PDZ it seems

© 2015 Authors; published by Portland Press Limited

likely that the N-terminal anchor is mainly beneficial to peptide specificity to MAST2–PDZ, whereas kinetics and affinity are mostly driven by the C-terminal PBM.

Furthermore, we have addressed the question of how the unconventional mode of binding of peptides to MAST2–PDZ affects the dynamics and conformation of the domain. We previously showed that PDZ ligand binding precludes the self-association of MAST2–PDZ [2]. At the time, we proposed that the N-anchoring of the peptide prevents dimerization by steric hindrance. Using NMR, we show in the present study that the dimerization surface involves residues from the $\beta 2$-$\beta 3$ strands, the $\alpha 1/\beta 4$ loop and the C-terminal of MAST2–PDZ. The interface of the dimer is consistent with the X-ray structure of MAST isoforms and with the dimeric structure of GRASP. In the case of GRASP, PBM binding and oligomerization modulate the cellular activity of the protein [23]. Contrary to our initial hypothesis, we show in the present study that the strict canonical PBM is by itself able to prevent dimerization of MAST2–PDZ although being solely confined within the binding groove. Thus, PBM binding must transmit effects through an interconnected network of residues on MAST2–PDZ. This network of side chains involves the $\beta 2$ and $\beta 3$ strands, the $\alpha 1/\beta 4$ loop and the C-terminus that are perturbed by the PBM binding. We propose that MAST2–PDZ auto-association is prevented by PBM binding through this network of energetically linked changes in the conformation of key residues.

The experimental evidence that we have collected over the years indicates that MAST2–PDZ sustains fine-tuned structural and dynamical changes upon ligand binding. A connected pathway signals the binding event to regions that are distal to the binding cleft, on the protein surface. These findings are compatible with a wealth of prior data on PDZ domains. Statistical coupling analysis and double mutant cycles suggested that energetic pathways within PDZ domains may support allostery [24–30]. A model of protein regulation controlled by the equilibrium between PBM binding and dimerization is particularly attractive since we now know that over 30 % of PDZ domains form dimers in solution [31]. Several types of heterologous complexes involving MASTs [32,33] have been documented; however, the auto-association of MAST2 in cells has not been studied yet. Allosteric perturbation of the dimeric interface of MAST2–PDZ upon peptide binding may provide interesting clues for MAST2 regulation overall. Valiente et al. [34] showed that binding of PTEN to MAST2–PDZ increases the rate of PTEN phosphorylation by MAST2 kinase. Thus, the regulation by PBM interaction might be involved at several levels in this system: recruiting kinase substrate and affecting protein auto-association. More specifically, dissociation of MAST2–PDZ might modify the catalytic activity of its kinase module. Such a mechanism acting on the substrate PTEN might lead to a modulation of neuronal proliferation and survival.

AUTHOR CONTRIBUTION

Florent Delhommel, Alain Chaffotte, Florence Cordier, Elouan Terrien and Nicolas Wolff planned the experiments. Florent Delhommel, Alain Chaffotte, Florence Cordier, Bertrand Reynal and Elouan Terrien performed the experiments. Florent Delhommel, Alain Chaffotte, Florence Cordier, Bertrand Reynal, Henri Buc and Nicolas Wolff analysed the data. Florent Delhommel, Alain Chaffotte, Florence Cordier, Muriel Delepierre, Henri Buc and Nicolas Wolff wrote the paper.

ACKOWLEDMENTS

We thank Jane Hardy and Terence Strick for careful proofreading of the manuscript and Patrick England for valuable discussions. The authors declare that they have no conflict of interest.

FUNDING

This work was supported by the Institut Pasteur and Institut Carnot Pasteur Maladies Infectieuses [grant number 13/10/120]; and the Ministère de l'Enseignement Supérieur [grant number 883/2013 (to F.D.)].

REFERENCES

1 Loh, S.H.Y., Francescut, L., Lingor, P., Bähr, M. and Nicotera, P. (2008) Identification of new kinase clusters required for neurite outgrowth and retraction by a loss-of-function RNA interference screen. Cell Death Differ. **15**, 283–298 CrossRef PubMed
2 Terrien, E., Chaffotte, A., Lafage, M., Khan, Z., Préhaud, C., Cordier, F., Simenel, C., Delepierre, M., Buc, H., Lafon, M. and Wolff, N. (2012) Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. Sci. Signal. **5**, 1–12 CrossRef PubMed
3 Manning, B.D. and Cantley, L.C. (2007) AKT/PKB signaling: navigating downstream. Cell **129**, 1261–1274 CrossRef PubMed
4 Sansal, I. and Sellers, W.R. (2004) The biology and clinical relevance of the PTEN tumor suppressor pathway. J. Clin. Oncol. **22**, 2954–2963 CrossRef PubMed
5 Préhaud, C., Wolff, N., Terrien, E., Lafage, M., Mégret, F., Babault, N., Cordier, F., Tan, G., Maitrepierre, E., Ménager, P. et al. (2010) Attenuation of rabies virulence: takeover by the cytoplasmic domain of its envelope protein. Sci. Signal. **3**, 1–10 PubMed
6 Préhaud, C., Dietzschold, B. and Lafon, M. (2003) Glycoprotein of nonpathogenic rabies viruses is a key determinant of human cell apoptosis. J. Virol. **77**, 10537–10547 CrossRef PubMed
7 Feng, W. and Zhang, M. (2009) Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. Nat. Rev. Neurosci. **10**, 87–99 CrossRef PubMed
8 Harris, B.Z. and Lim, W.A. (2001) Mechanism and role of PDZ domains in signaling complex assembly. J. Cell Sci. **114**, 3219–2931 PubMed
9 Hung, A.Y. and Sheng, M. (2002) PDZ domains: structural modules for protein complex assembly. J. Biol. Chem. **277**, 5699–5702 CrossRef PubMed
10 Saras, J. and Heldin, C.-H. (1996) PDZ domains bind carboxy.terminal sequences of target proteins. Trends Biochem. Sci. **4**, 455–458 CrossRef
11 Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., Chishti, A.H., Crompton, A., Chan, A.C., Anderson, J.M. and Canthley, L.C. (1997) Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science **275**, 73–77 CrossRef PubMed
12 Liu, X., Shepherd, T.R., Murray, A.M., Xu, Z. and Fuentes, E.J. (2013) The structure of the Tiam1 PDZ domain/ phospho-syndecan1 complex reveals a ligand conformation that modulates protein dynamics. Structure **21**, 342–354 CrossRef PubMed
13 Luck, K., Charbonnier, S. and Travé, G. (2012) The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. FEBS Lett. **586**, 2648–2661 CrossRef PubMed
14 Shepherd, T.R., Klaus, S.M., Liu, X., Ramaswamy, S., DeMali, K.A. and Fuentes, E.J. (2010) The Tiam1 PDZ domain couples to Syndecan1 and promotes cell-matrix adhesion. J. Mol. Biol. **398**, 730–746 CrossRef PubMed
15 Ye, F. and Zhang, M. (2013) Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. Biochem. J. **455**, 1–14 CrossRef PubMed
16 Erlendsson, S., Rathje, M., Heidarsson, P.O., Poulsen, F.M., Madsen, K.L., Teilum, K. and Gether, U. (2014) PICK1 (protein interacting with C-kinase 1) binding promiscuity relies on unconventional PDZ (PSD-95/Discs-large/ZO-1 homology) binding modes for non-class II PDZ ligands. J. Biol. Chem. **289**, 25327–25340 CrossRef PubMed
17 Terrien, E., Simenel, C., Prehaud, C., Buc, H., Delepierre, M., Lafon, M. and Wolff, N. (2009) 1H, 13C and 15N resonance assignments of the PDZ of microtubule-associated serine/threonine kinase 205 (MAST205) in complex with the C-terminal motif from the rabies virus glycoprotein. Biomol. NMR Assign. **3**, 45–48 CrossRef PubMed
18 Barbato, G., Mitsuhiko, I., Kay, L.E., Pastor, R.W. and Bax, A. (1992) Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible? Biochemistry **31**, 5269–5278 CrossRef PubMed
19 Huang, Y. and Liu, Z. (2009) Kinetic advantage of intrinsically disordered proteins in coupled folding–binding process: a critical assessment of the "fly-casting" mechanism. J. Mol. Biol. **393**, 1143–1159 CrossRef PubMed
20 Im, Y.J., Park, S.H., Rho, S.H., Lee, J.H., Kang, G.B., Sheng, M., Kim, E. and Eom, S.H. (2002) Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class II PDZ target recognition and PDZ domain-mediated multimerization. J. Biol. Chem. **278**, 8501–8507 CrossRef PubMed
21 Im, Y.J., Lee, J.H., Park, S.H., Park, S.J., Rho, S.-H., Kang, G.B., Kim, E. and Eom, S.H. (2003) Crystal structure of the Shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization. J. Biol. Chem. **278**, 48099–48104 CrossRef PubMed

© 2015 Authors; published by Portland Press Limited

22 Chi, C.N., Bach, A., Engström, A., Wang, H., Strømgaard, K., Gianni, S. and Jemth, P. (2009) A sequential binding mechanism in a PDZ domain. Biochemistry **48**, 7089–7097 CrossRef PubMed
23 Sugi, T., Oyama, T., Muto, T., Nakanishi, S., Morikawa, K. and Jingami, H. (2007) Crystal structures of autoinhibitory PDZ domain of Tamalin: implications for metabotropic glutamate receptor trafficking regulation. EMBO J. **26**, 2192–2205 CrossRef PubMed
24 Lockless, S.W. and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science **286**, 295–299 CrossRef PubMed
25 Fuentes, E.J., Der, C.J. and Lee, A.L. (2004) Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. J. Mol. Biol. **335**, 1105–1115 CrossRef PubMed
26 Hultqvist, G., Haq, S.R., Punekar, A.S., Chi, C.N., Engström, A., Bach, A., Strømgaard, K., Selmer, M., Gianni, S. and Jemth, P. (2013) Energetic pathway sampling in a protein interaction domain. Structure **21**, 1193–1202 CrossRef PubMed
27 McLaughlin, R.N., Poelwijk, F.J., Raman, A., Gosal, W.S. and Ranganathan, R. (2012) The spatial architecture of protein function and adaptation. Nature **491**, 138–142 CrossRef PubMed
28 Petit, C.M., Zhang, J., Sapienza, P.J., Fuentes, E.J. and Lee, A.L. (2009) Hidden dynamic allostery in a PDZ domain. PNAS **106**, 3–8 CrossRef
29 Van Den Berk, L. C. J., Landi, E., Walma, T., Vuister, G.W., Dente, L. and Hendriks, W. J. A. J. (2007) An allosteric intramolecular PDZ-PDZ interaction modulates PTP-BL PDZ2 binding specificity. Biochemistry **46**, 13629–13637 CrossRef PubMed
30 Zhang, J., Sapienza, P.J., Ke, H., Chang, A., Hengel, S.R., Wang, H., Phillips, G.N. and Lee, A.L. (2010) Crystallographic and nuclear magnetic resonance evaluation of the impact of peptide binding to the second PDZ domain of protein tyrosine phosphatase 1E. Biochemistry **49**, 9280–9291 CrossRef PubMed
31 Chang, B.H., Gujral, T.S., Karp, E.S., BuKhalid, R., Grantcharova, V.P. and MacBeath, G. (2011) A systematic family-wide investigation reveals that ~30 % of mammalian PDZ domains engage in PDZ-PDZ interactions. Chem. Biol. **18**, 1143–1152 CrossRef PubMed
32 Lumeng, C., Phelps, S., Crawford, G.E., Walden, P.D., Barald, K. and Chamberlain, J.S. (1999) Interactions between beta 2-syntrophin and a family of microtubule-associated serine/threonine kinases. Nat. Neurosci. **2**, 611–617 CrossRef PubMed
33 Yano, R., Yap, C., Yamazaki, Y., Muto, Y., Kishida, H., Okada, D. and Hashikawa, T. (2003) Sast124, a novel splice variant of syntrophin-associated serine/threonine kinase (SAST), is specifically localized in the restricted brain regions. Neuroscience **117**, 373–381 CrossRef PubMed
34 Valiente, M., Andrés-Pons, A., Gomar, B., Torres, J., Gil, A., Tapparel, C., Antonarakis, S.E. and Pulido, R. (2005) Binding of PTEN to specific PDZ domains contributes to PTEN protein stability and phosphorylation by microtubule-associated serine/threonine kinases. J. Biol. Chem. **280**, 28936–28943 CrossRef PubMed

Received 24 September 2014/17 April 2015; accepted 5 May 2015
Published as BJ Immediate Publication 5 May 2015, doi:10.1042/BJ20141198

© 2015 Authors; published by Portland Press Limited

13.5 Caillet-Saguy et al. (2015) Strategies to interfere with PDZ-mediated inteactions in neurons : What we can learn from the rabies virus. *Prog. Biophys. Mol. Biol. Vol.119 (1) p53-59*

Progress in Biophysics and Molecular Biology 119 (2015) 53–59

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

Review

Strategies to interfere with PDZ-mediated interactions in neurons: What we can learn from the rabies virus

Célia Caillet-Saguy [a], Pierre Maisonneuve [a], Florent Delhommel [a, b], Elouan Terrien [a], Nicolas Babault [a], Monique Lafon [c], Florence Cordier [a], Nicolas Wolff [a, *]

[a] Unité de RMN des Biomolécules (CNRS UMR 3528), Département de Biologie Structurale et Chimie, Institut Pasteur, Paris, France
[b] Université Pierre et Marie Curie, Cellule Pasteur UPMC, rue du Docteur Roux, 75015 Paris, France
[c] Unité de Neuroimmunologie Virale, Département de Virologie, Institut Pasteur, Paris, France

ARTICLE INFO

Article history:
Available online 5 March 2015

Keywords:
PDZ domain
Rabies virus
Neurons
Neuronal signaling
Phosphorylation
Enzymatic regulation

ABSTRACT

PDZ (PSD-95/Dlg/ZO-1) domains play a major role in neuronal homeostasis in which they act as scaffold domains regulating cellular trafficking, self-association and catalytic activity of essential proteins such as kinases and phosphatases. Because of their central role in cell signaling, cellular PDZ-containing proteins are preferential targets of viruses to hijack cellular function to their advantage. Here, we describe how the viral G protein of the rabies virus specifically targets the PDZ domain of neuronal enzymes during viral infection. By disrupting the complexes formed by cellular enzymes and their ligands, the virus triggers drastic effect on cell signaling and commitment of the cell to either survival (virulent strains) or death (vaccinal strains). We provide structural and biological evidences that the viral proteins act as competitors endowed with specificity and affinity in an essential cellular process by mimicking PDZ binding motif of cellular partners. Disruption of critical endogenous protein–protein interactions by viral protein drastically alters intracellular protein trafficking and catalytic activity of cellular proteins that control cell homeostasis. This work opens up many perspectives to mimic viral sequences and developing innovative therapies to manipulate cellular homeostasis.

© 2015 Elsevier Ltd. All rights reserved.

Contents

1. Introduction 54
   1.1. PDZ domains 54
   1.2. PDZ and neurons 54
   1.3. PDZ and viruses 54
   1.4. Rabies virus, a source of pathogenic mimicry of host PBM 54
2. Neuronal survival and the kinase MAST2 55
   2.1. RABV G protein competes with PTEN for binding to MAST2 55
   2.2. The phosphorylation of PTEN 56
3. Neuronal apoptosis and the phosphatase PTPN4 56
   3.1. RABV G protein acts as an apoptosis inducer by binding to PTPN4 56
   3.2. Allosteric regulation of the catalytic activity of PTPN4 by its PDZ domain 56

Abbreviations: CytoG, C-terminal cytoplasmic tail of G protein; MAST2, microtubule-associated serine and threonine kinase 2; PBM, PDZ binding motif; PDZ, PSD-95/Dlg/ZO-1; PSD, post-synaptic density; PTEN, phosphatase and tensin homolog deleted on chromosome 10; PTP, Protein tyrosine phosphatase; RABV, Rabies virus; SliM, small linear interaction motifs.

* Corresponding author. Unité de RMN des biomolécules, Institut Pasteur-URA 3528 CNRS, 28, rue du Dr Roux, 75724 Paris cedex 15, France. Tel.: +33 (0)1 45 68 88 72; fax: +33 (0)1 45 68 89 29.
E-mail address: nicolas.wolff@pasteur.fr (N. Wolff).

http://dx.doi.org/10.1016/j.pbiomolbio.2015.02.007
0079-6107/© 2015 Elsevier Ltd. All rights reserved.

4. Applications 58
5. Conclusions 58
Acknowledgments 58
References 58

1. Introduction

1.1. PDZ domains

PDZ domains represent a large family of protein–protein interaction domains found in prokaryotes and eukaryotes. About 270 PDZ domains, distributed in more than 150 proteins, have been identified in the human proteome (Weatheritt et al., 2012). They are involved in diverse cellular functions such as cell polarity and migration, intercellular junctions, cellular homeostasis, neuronal signal transmission or visual and auditory machinery (Nourry et al., 2003). PDZ domains adopt a globular fold of 80–90 amino-acids comprising 5 to 6 β-strands and 1 to 2 α–helices (Fig. 1) (Doyle et al., 1996). They recognize PDZ-binding motifs (PBM) typically located at the carboxyl terminus of proteins. In their bound state, these short linear motifs form an anti-parallel strand that inserts in a groove delimited by the β2 strand and the α2 helix of the PDZ domain (Fig. 1). Recognition of PBM by PDZ domains is based on two key PBM positions, the p0 (C-terminal residue) and p-2 (antepenultimate residue) (Fig. 1). However, an increasing number of studies indicates that residues up to p-11 as well PDZ extensions can participate to the complex formation (Fig. 1) (Terrien et al., 2012; Wang et al., 2010).

1.2. PDZ and neurons

The PDZ domain is the most common protein–protein interaction domain among the scaffold proteins present in the excitatory synapses, particularly in the post-synaptic density (PSD) (Feng and Zhang, 2009; Kim and Sheng, 2004). PDZ domains have been first discovered in the protein PSD-95, a main actor of the PSD, interacting with numerous PBM-containing neurotransmitter receptors such as glutamate receptors (Bard and Groc, 2011; Cho et al., 1992). Into the PSD, the abundance of PDZ domains ensures the dynamic response of synapses by targeting PBM located in the disordered carboxyl tail of their targets with weak binding affinities. PDZ and PBM-containing proteins are also key regulators of neuronal survival and death, and are involved into various signaling pathways. Unsurprisingly, mutations in PDZ/PBM-containing proteins are often linked to human neuronal pathologies. For instance, mutations in Harmonin cause Usher syndrome characterized by congenital deafness and blindness, mutations in Periaxin are associated with a severe form of neuropathy and mutations in phosphatase and tensin homolog deleted on chromosome 10 (PTEN) are related to Alzheimer's disease, Lhermitte-Duclos diseases and to a variety of cancers (Boerkoel et al., 2001; Chang et al., 2007; Griffin et al., 2005; Verpy et al., 2000).

1.3. PDZ and viruses

Because of their central role in cell signaling, PDZ-containing proteins are preferential targets of viruses to hijack cellular function to their advantages. Numerous small linear interaction motifs (SliM) such as PBM have been found in viral proteins where they mediate interactions between viral and host proteins (Davey et al., 2011). Indeed, viruses necessarily hijack various cellular functions for the accomplishment of their life cycle, such as entry and budding, cellular trafficking, transcriptional regulation, protein degradation, signal transduction and cell cycle regulation. The quick adaptation of viruses to their environment could then involve the extensive use of SLiM mimicry to subvert host functionality. The inhibition or perturbation of the function of cellular PDZ-containing proteins appears to be a strategy evolved by some viruses to enhance their replication, to disseminate in the host, and to transmit to new hosts (Javier and Rice, 2011).

1.4. Rabies virus, a source of pathogenic mimicry of host PBM

Rabies virus (RABV) is a pathogen that exclusively infects neurons. RABV causes rabies — an acute encephalitis — that is fatal in most mammal species including humans. Various RABV strains differentially affect the homeostasis of infected neurons through variations in the sequence of their glycoprotein (G protein). Pathogenic (virulent) RABV strains favor neuron survival preserving the cell integrity and therefore promoting viral dissemination (Lafon, 2011). Non-pathogenic (attenuated) strains lead to neuron apoptosis (Prehaud et al., 2003). The death/survival balance of infected neuroblastoma is governed by a single mutation located in the C-terminal cytoplasmic tail of G protein (CytoG) (Prehaud et al., 2010; Thoulouze et al., 1998). This critical mutation Q(-3)E is located in a PBM at the very C-terminus sequence of CytoG, and is sufficient to drastically affect the nature of the neuronal partners interacting with G protein (Fig. 2A). Neuron survival requires selective association of the viral PBM (-QTRL$_{COOH}$) with the PDZ

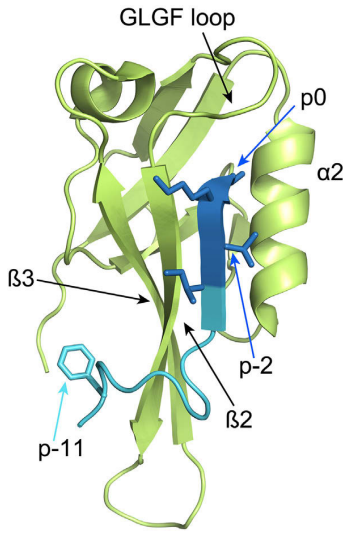


Fig. 1. Structure of the PDZ domain. Cartoon diagram of MAST2-PDZ (green) complexed with the C-terminus of PTEN (blue and cyan) (PDB id. 2KYL). The structure highlights two anchoring points: a canonical anchor with the PBM of PTEN (blue) and an unconventional anchor involving the phenylalanine side chain at position p-11 of PTEN (cyan).

domain of microtubule-associated serine and threonine kinase 2 (MAST2) (Fig. 2B) (Prehaud et al., 2010). The amino acid change Q(-3)E in the PBM (-ETRL$_{COOH}$) allows G protein to recruit additional PDZ partners, notably the tyrosine-phosphatase PTPN4 (Fig. 2C). PTPN4 has been shown to function as an inhibitor of cell death while MAST2 acts as an inhibitor of neurite outgrowth in human neuroblastoma (Loh et al., 2008; Prehaud et al., 2010). Our hypothesis is that RABV G proteins antagonize PTPN4 and MAST2 functions by mimicking the PBM of endogenous protein partners and disrupting the interactions of their PDZ domains with their respective cellular ligands. Fig. 3 illustrates that virulent and attenuated viral peptides have similar affinities than that of endogenous partners of MAST2 and PTPN4 respectively. However, the high level of expression of RABV G proteins in infected cells allows the efficient competition of viral PBM with PBM from endogenous partners for the binding to PDZ domains.

2. Neuronal survival and the kinase MAST2

2.1. RABV G protein competes with PTEN for binding to MAST2

MAST2 is one of regulators of the phosphatase PTEN (Valiente et al., 2005). PTEN is encoded by one of the most frequently mutated tumour suppressor gene in human cancer (Bunney and Katan, 2010; Li et al., 1997; Steck et al., 1997). Both MAST2 and PTEN are widely expressed in the brain where they control neuronal survival (Garland et al., 2007; Ventruti et al., 2011). They form a complex through the interaction of the PDZ domain of MAST2 (MAST2-PDZ) with the PBM located at the C-terminus of PTEN. The phosphorylation of PTEN by MAST2 is drastically increased when the MAST2/PTEN complex is stabilized by the PDZ/PBM interaction (Fig. 2B) (Valiente et al., 2005). Phosphorylation at several sites of the C-terminal domain of PTEN modulates its cellular trafficking, stability and conformation and consequently its activity (Leslie et al., 2009). RABV specifically targets the MAST2/PTEN interaction to alter the intracellular localization of PTEN and to promote survival of the infected neurons (Terrien et al., 2012).

The dynamical and structural properties of MAST2-PDZ in complex with the C-terminal sequences of PTEN and of the RABV G protein have been studied to understand the respective binding specificities of PTEN and viral G protein for MAST2-PDZ (Terrien et al., 2012). Both viral and PTEN PBM interact with MAST2-PDZ through a similar network of connectivities, defining a shared and previously uncharacterized interaction surface with a PDZ domain (PDB id. 2KQF and 2KYL). The unusually large surface of interaction for a PDZ/PBM complex involves two anchoring points (Fig. 1). This results in a large enthalpic contribution and a high affinity of ligand for a PDZ domain (below 1 μM). The binding occurs between the PBM binding pocket and the canonical C-terminal PBM and also between a hydrophobic cluster exposed at the surface

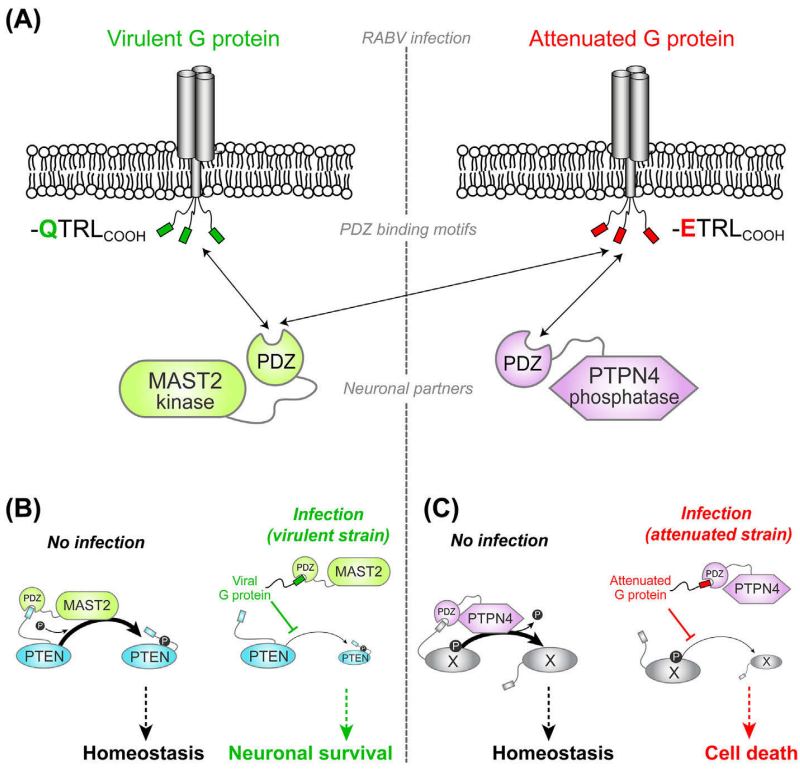


Fig. 2. Model of neuron infection by a virulent or an attenuated strain of RABV and associated phenotypes. (A) Upon infection by the virulent strain (left), the viral G protein with the -QTRL$_{COOH}$ PBM interacts with the PDZ domain of MAST2 kinase whereas upon infection by the attenuated strain (right), the viral G protein with the -ETRL$_{COOH}$ PBM interacts not only with MAST2 but also with other cellular partners, particularly PTPN4. (B) In the absence of infection, PTEN is phosphorylated notably by MAST2. Upon infection by the virulent strain, the viral G protein interacts *via* its PBM with the PDZ domain of the MAST2. This interaction could prevent MAST2-controlled phosphorylation of PTEN, which alters the homeostasis of the infected cell and leads to its survival. (C) In the absence of infection, the cellular partner X is dephosphorylated by PTPN4. Upon infection by the attenuated strain, the viral PBM interacts with the PDZ domain of PTPN4. This interaction prevents efficient dephosphorylation of X, which alters the homeostasis of the infected cell and triggers apoptosis.

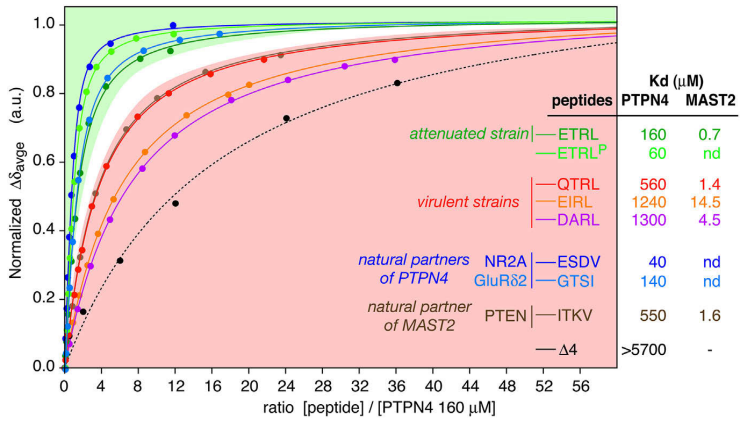


Fig. 3. Titrations of PTPN4-PDZ with viral and endogenous PBM peptides. NMR titration curves of PTPN4-PDZ with peptides derived from various viral glycoproteins and endogenous partners. The curves represent averaged chemical shifts changes normalized over the 10 most affected residues upon variation of the ratio peptide/PDZ for each complex. The dissociation constants Kd of complexes are indicated. The dissociation constants values obtained by ITC experiments with MAST2-PDZ are also reported.

of the MAST2-PDZ β2/β3 sheet and an aromatic residue at the ligand p(-11) (Fig. 1). Long-range effects induced by the PBM binding prevent MAST2-PDZ auto-association by modifying the structure and the dynamic of the PDZ dimerization surface (submitted). These results illustrate a potential allosteric regulation in PDZ domain through communication between binding sites and distal surfaces that could provide an interesting clue for enzyme overall regulation.

2.2. The phosphorylation of PTEN

Expression of the viral G protein in neuroblastoma cells promotes the nuclear exclusion of PTEN in a PBM−dependent manner (Terrien et al., 2012). The cellular trafficking of PTEN has been clearly related to the phosphorylation modifications of its C-terminal tail (PTEN-Ctail). The phosphorylation state of PTEN-Ctail has been proposed to have consequences on PTEN stability and controls a conformational switch of PTEN (open/closed model) (Rahdar et al., 2009; Vazquez et al., 2001, 2000). We recently deciphered by NMR the phosphorylation pattern of PTEN-Ctail by two major neuronal kinases, CKII and GSK3β (Cordier et al., 2012, 2014). Two cascades of phosphorylation events controlled by CK2 and GSK3β have been identified *in vitro* and in human neuroblastoma cell extracts. They involve two nearby clusters of sites and occur independently on different timescales. We proposed that these cascades, known to play distinct functional roles, could activate different time-delayed regulatory responses in cells. In this phosphorylation process, the contribution of MAST2 still remains elusive.

3. Neuronal apoptosis and the phosphatase PTPN4

3.1. RABV G protein acts as an apoptosis inducer by binding to PTPN4

The mutation Q(-3)E in the PBM of RABV G protein generates interactions with new cellular partners, especially with the phosphatase PTPN4, leading to attenuated rabies virulence (Fig. 2).

PTPN4 is a non-receptor tyrosine phosphatase functioning as an anti-apoptotic protein. The PBM peptide stemming from RABV attenuated G protein triggers apoptosis by interacting with the PDZ domain of PTPN4 (PTPN4-PDZ) that flanks the catalytic phosphatase domain. *In vitro*, the Q to E change significantly increases the affinity of the viral PBM for PTPN4-PDZ and favors the formation of the viral G protein/PTPN4 complex *in cellulo* as showed by colocalization monitored by confocal microscopy (Prehaud et al., 2010) (Fig. 3). Our analysis of the structure of viral peptides/PTPN4-PDZ complexes by NMR and X-ray diffraction reveals that this affinity modification is due to slight modifications of the interaction network at the PDZ binding site leading to the drastic functional effect reported for the conservative Q(-3)E change (Babault et al., 2011). We demonstrated that neuronal cell apoptosis levels reflect the affinities of PBM for PTPN4-PDZ (Babault et al., 2011). The mere binding of PBM to PTPN4-PDZ is sufficient to induce apoptosis despite a possible concommitent binding to MAST2-PDZ (Fig. 3).

3.2. Allosteric regulation of the catalytic activity of PTPN4 by its PDZ domain

In the present case, the complex formed by PTPN4-PDZ with its cellular ligand appears to be mandatory for PTPN4 anti-apoptotic function. The disruption of this interaction by the attenuated RABV PBM is the crucial initial step to hijack its anti-apoptotic function. We documented the regulation of the phosphatase activity of PTPN4 by its PDZ domain and its PBM ligands. The PDZ domain inhibits the flanking catalytic domain of PTPN4 through an intramolecular mixed-inhibition (Fig. 4) (Maisonneuve et al., 2014). The mere binding of a PBM, from endogenous or viral origin, to the PDZ domain abolishes the auto-inhibition and restores the catalytic activity of PTPN4 (Fig. 4). Both free inactive and liganded active forms of PTPN4 show a predominant compact conformation in solution indicating a close proximity between PDZ and PTP (protein tyrosine phosphatase) domains. However, the two forms adopt different hydrodynamic behavior as probed by NMR. A dynamical rearrangement between PDZ and PTP domains arises from the binding of the PDZ ligand that destabilizes the auto-inhibited state of the phosphatase through long−range interactions to stimulate its catalytic activity (Fig. 4). Thus, the PDZ domain of PTPN4 contributes both to the recruitment of substrates and to the catalytic regulation by directly modulating the dephosphorylation of endogenous partners.

As shown for MAST2, we postulated that the viral sequence acts as an inhibitor of PTPN4 function only by disrupting the interaction

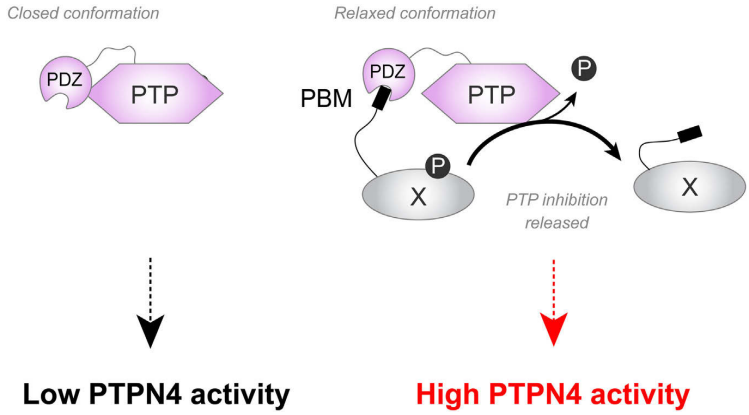


Fig. 4. Regulation of the PTPN4 phosphatase activity by its PDZ domain. In the absence of a PBM-containing partner, the PDZ and the phosphatase domains of PTPN4 adopt a compact conformation. In this condition, the PTP activity is inhibited by the PDZ domain that keeps the partner dephosphorylation at lower level (left). The PBM binding to the PDZ domain of PTPN4 induces a relaxed conformation of PDZ-PTP domains with a release of the inhibition of the PTPN4 phosphatase activity allowing an efficient dephosphorylation of a PBM-containing partner X (right).

of PTPN4-PDZ with its cellular ligand. However, the interactome of PTPN4 has been only partially described. The natural ligands of PTPN4-PDZ involved in the regulation of cell homeostasis remain unknown despite several studies exploring the proteome by yeast double hybrid system. Nevertheless, the p38γ-mitogen-activated protein kinase (MAPK) was recently identified as an endogenous partner of the PDZ domain of PTPN3, the closest homologous protein of PTPN4, in the RAS proliferative signaling (Hou et al., 2010). We presumed that p38γ could also be the ligand of PTPN4-PDZ and that this p38γ/PTPN4 interaction could promote anti-apoptotic signaling in neuronal cells. p38γ is the only member of MAPK family proteins with a PBM at its carboxy-terminus. We recently showed that p38γ and PTPN4 interact tightly both *in vitro* and in cell extracts (unpublished data). We are now interested in elucidating in details the molecular mechanism of the p38γ/PTPN4 interplay between association and phosphorylation/dephosphorylation reactions.

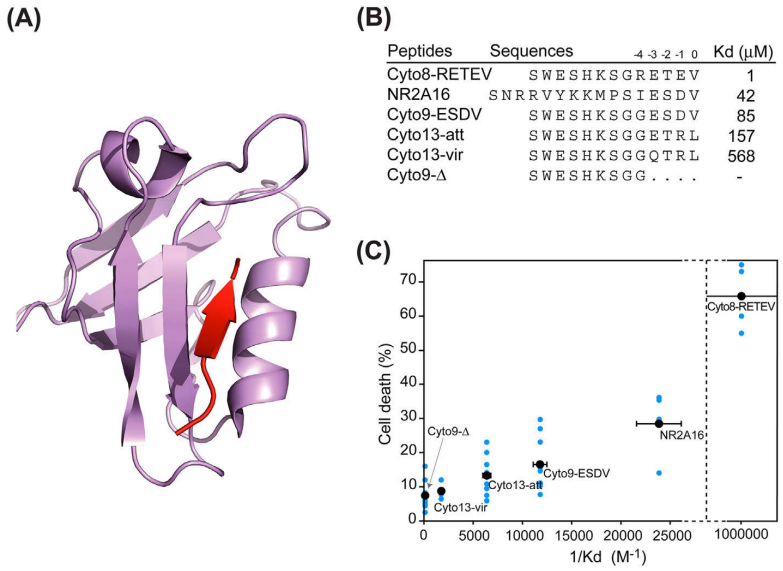


Fig. 5. Rational design of anti-proliferative peptides. (A) Structure of PTPN4-PDZ (purple) bound to the optimized peptide Cyto8-RETEV (red) (B) Peptide sequences and their affinities for PTPN4-PDZ. Cyto8-RETEV, an optimized sequence; GluN2A-16 including the C-terminal sequence of the endogenous partner of PTPN4-PDZ, GluN2A; Cyto13-att and Cyto13-vir from the G proteins of attenuated and virulent rabies virus strains, respectively; the chimeric peptide Cyto9-ESDV encoding the PBM of NR2A16 with the core of Cyto13-att and Cyto9-Delta corresponding to Cyto13-att truncated of its PBM (C) Induction of the death of human glioblastoma cells is correlated to the peptide affinity for PTPN4-PDZ. Data from Babault et al., 2011.

4. Applications

Understanding how viruses manipulate cellular homeostasis to their advantages opens up many perspectives to mimic viral sequences and to develop innovative therapies. Our general objective is to select optimized deliverable short peptides and small molecules that can compete with the natural ligands for binding to PTPN4 (Fig. 5A) (Babault et al., 2011) or MAST2 and that can trigger death or survival, respectively (Khan and Lafon, 2014). These peptides and small molecules will be able to recognize at high affinity and specificity the targeted PDZ domain.

We showed that short polypeptides of viral origin display spectacular death capacities or neuroprotecting properties upon intra-cellular delivery or expression (Babault et al., 2011). We used structural and biophysical data (thermodynamical and kinetics parameters) collected on numerous peptides/PDZ complexes to run rational screening of peptides. Starting from the viral sequences, we obtained peptides with affinities of 1 μM for PTPN4-PDZ (Fig. 5B) and of 60 nM for MAST2-PDZ, after two rounds of optimization. The best candidate peptides for PTPN4 and for MAST2 have the capacity to trigger cell death or to induce neurite outgrowth in several cell lines, respectively. In both cases, their capacity to alter the cellular homeostasis is correlated to their affinity for their targeted PDZ domains (Fig. 5C). Pro-survival peptides targeting MAST2 are currently in trial for their neuro-regeneration properties in animal models while the activity of the anti-oncogenic peptides will be evaluated in appropriate mouse model of sub-cutaneous xenografted tumors.

5. Conclusions

The structure/function analysis of MAST2 and PTPN4 and their ligands provides deep insight in the mechanisms acquired by viruses through evolution to hijack the cell machinery to their own benefit. Studying these mechanisms allow us to address unsolved questions about signaling and spatial cell biology upon infection. It also provides information about fundamental aspects of neuronal cell death and survival mechanisms and generates the development of more applied research. For instance, Rabies virus targets PDZ domains that are the starting point of an intramolecular signaling mechanism that ultimately modulates the activity of the attached catalytic domain and consequently alters cellular homeostasis.

More generally, viral genomes are highly promising sources of discovery for new survival or death peptide sequences and sustained robust pathways. These discoveries may lead to the development of new neuroprotective compounds or anticancer molecules. With the increasing knowledge of the structure and *in vivo* functions of PDZ proteins, PDZ interactions are promising targets for pharmaceutical treatment to cure nervous disorders as well as cancers.

Acknowledgments

Prof. Henri Buc is ackowledged for helpul discussion. This work was supported by grants from the Institut Pasteur, Agence Nationale pour la Recherche and Institut Carnot Pasteur Maladies Infectieuses. N.B., E.T., P.M. and F.D. are recipients of fellowships from the Ministère de l'Enseignement Supérieur et de la Recherche and the Fondation pour la Recherche Médicale.

References

Babault, N., Cordier, F., Lafage, M., Cockburn, J., Haouz, A., Prehaud, C., Rey, F.A., Delepierre, M., Buc, H., Lafon, M., Wolff, N., 2011. Peptides targeting the PDZ domain of PTPN4 are efficient inducers of glioblastoma cell death. Structure 19, 1518–1524.
Bard, L., Groc, L., 2011. Glutamate receptor dynamics and protein interaction: lessons from the NMDA receptor. Mol. Cell. Neurosci. 48, 298–307.
Boerkoel, C.F., Takashima, H., Bacino, C.A., Daentl, D., Lupski, J.R., 2001. EGR2 mutation R359W causes a spectrum of Dejerine-Sottas neuropathy. Neurogenetics 3, 153–157.
Bunney, T.D., Katan, M., 2010. Phosphoinositide signalling in cancer: beyond PI3K and PTEN. Nat. Rev. Cancer 10, 342–352.
Chang, N., El-Hayek, Y.H., Gomez, E., Wan, Q., 2007. Phosphatase PTEN in neuronal injury and brain disorders. Trends Neurosci. 30, 581–586.
Cho, K.O., Hunt, C.A., Kennedy, M.B., 1992. The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein. Neuron 9, 929–942.
Cordier, F., Chaffotte, A., Terrien, E., Prehaud, C., Theillet, F.X., Delepierre, M., Lafon, M., Buc, H., Wolff, N., 2012. Ordered phosphorylation events in two independent cascades of the PTEN C-tail revealed by NMR. J. Am. Chem. Soc. 134, 20533–20543.
Cordier, F., Chaffotte, A., Wolff, N., 2014. Quantitative and dynamic analysis of PTEN phosphorylation by NMR. Methods. http://dx.doi.org/10.1016/j.ymeth.2014.10.007.
Davey, N.E., Trave, G., Gibson, T.J., 2011. How viruses hijack cell regulation. Trends Biochem. Sci. 36, 159–169.
Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M., MacKinnon, R., 1996. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 85, 1067–1076.
Feng, W., Zhang, M., 2009. Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. Nat. Rev. Neurosci. 10, 87–99.
Garland, L.L., Rankin, C., Gandara, D.R., Rivkin, S.E., Scott, K.M., Nagle, R.B., Klein-Szanto, A.J., Testa, J.R., Altomare, D.A., Borden, E.C., 2007. Phase II study of erlotinib in patients with malignant pleural mesothelioma: a southwest oncology group study. J. Clin. Oncol. 25, 2406–2413.
Griffin, R.J., Moloney, A., Kelliher, M., Johnston, J.A., Ravid, R., Dockery, P., O'Connor, R., O'Neill, C., 2005. Activation of Akt/PKB, increased phosphorylation of Akt substrates and loss and altered distribution of Akt and PTEN are features of Alzheimer's disease pathology. J. Neurochem. 93, 105–117.
Hou, S.W., Zhi, H.Y., Pohl, N., Loesch, M., Qi, X.M., Li, R.S., Basir, Z., Chen, G., 2010. PTPH1 dephosphorylates and cooperates with p38gamma MAPK to increase ras oncogenesis through PDZ-mediated interaction. Cancer Res. 70, 2901–2910.
Javier, R.T., Rice, A.P., 2011. Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. J. Virol. 85, 11544–11556.
Khan, Z., Lafon, M., 2014. PDZ domain-mediated protein interactions: therapeutic targets in neurological disorders. Curr. Med. Chem. 21, 2632–2641.
Kim, E., Sheng, M., 2004. PDZ domain proteins of synapses. Nat. Rev. Neurosci. 5, 771–781.
Lafon, M., 2011. Evasive strategies in rabies virus infection. Adv. Virus Res. 79, 33–53.
Leslie, N.R., Maccario, H., Spinelli, L., Davidson, L., 2009. The significance of PTEN's protein phosphatase activity. Adv. Enzyme Regul. 49, 190–196.
Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S.I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R., Bigner, S.H., Giovanella, B.C., Ittmann, M., Tycko, B., Hibshoosh, H., Wigler, M.H., Parsons, R., 1997. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 275, 1943–1947.
Loh, S.H., Francescut, L., Lingor, P., Bahr, M., Nicotera, P., 2008. Identification of new kinase clusters required for neurite outgrowth and retraction by a loss-of-function RNA interference screen. Cell. Death Differ. 15, 283–298.
Maisonneuve, P., Caillet-Saguy, C., Raynal, B., Gilquin, B., Chaffotte, A., Perez, J., Zinn-Justin, S., Delepierre, M., Buc, H., Cordier, F., Wolff, N., 2014. Regulation of the catalytic activity of the human phosphatase PTPN4 by its PDZ domain. FEBS J. 281, 4852–4865.
Nourry, C., Grant, S.G., Borg, J.P., 2003. PDZ domain proteins: plug and play! Sci. STKE 2003, RE7.
Prehaud, C., Lay, S., Dietzschold, B., Lafon, M., 2003. Glycoprotein of nonpathogenic rabies viruses is a key determinant of human cell apoptosis. J. Virol. 77, 10537–10547.
Prehaud, C., Wolff, N., Terrien, E., Lafage, M., Megret, F., Babault, N., Cordier, F., Tan, G.S., Maitrepierre, E., Menager, P., Chopy, D., Hoos, S., England, P., Delepierre, M., Schnell, M.J., Buc, H., Lafon, M., 2010. Attenuation of rabies virulence: takeover by the cytoplasmic domain of its envelope protein. Sci. Signal. 3, ra5.
Rahdar, M., Inoue, T., Meyer, T., Zhang, J., Vazquez, F., Devreotes, P.N., 2009. A phosphorylation-dependent intramolecular interaction regulates the membrane association and activity of the tumor suppressor PTEN. Proc. Natl. Acad. Sci. U. S. A. 106, 480–485.
Steck, P.A., Pershouse, M.A., Jasser, S.A., Yung, W.K., Lin, H., Ligon, A.H., Langford, L.A., Baumgard, M.L., Hattier, T., Davis, T., Frye, C., Hu, R., Swedlund, B., Teng, D.H., Tavtigian, S.V., 1997. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat. Genet. 15, 356–362.
Terrien, E., Chaffotte, A., Lafage, M., Khan, Z., Prehaud, C., Cordier, F., Simenel, C., Delepierre, M., Buc, H., Lafon, M., Wolff, N., 2012. Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. Sci. Signal. 5, ra58.

Thoulouze, M.I., Lafage, M., Schachner, M., Hartmann, U., Cremer, H., Lafon, M., 1998. The neural cell adhesion molecule is a receptor for rabies virus. J. Virol. 72, 7181–7190.
Valiente, M., Andres-Pons, A., Gomar, B., Torres, J., Gil, A., Tapparel, C., Antonarakis, S.E., Pulido, R., 2005. Binding of PTEN to specific PDZ domains contributes to PTEN protein stability and phosphorylation by microtubule-associated serine/threonine kinases. J. Biol. Chem. 280, 28936–28943.
Vazquez, F., Grossman, S.R., Takahashi, Y., Rokas, M.V., Nakamura, N., Sellers, W.R., 2001. Phosphorylation of the PTEN tail acts as an inhibitory switch by preventing its recruitment into a protein complex. J. Biol. Chem. 276, 48627–48630.
Vazquez, F., Ramaswamy, S., Nakamura, N., Sellers, W.R., 2000. Phosphorylation of the PTEN tail regulates protein stability and function. Mol. Cell. Biol. 20, 5010–5018.
Ventruti, A., Kazdoba, T.M., Niu, S., D'Arcangelo, G., 2011. Reelin deficiency causes specific defects in the molecular composition of the synapses in the adult brain. Neuroscience 189, 32–42.
Verpy, E., Leibovici, M., Zwaenepoel, I., Liu, X.Z., Gal, A., Salem, N., Mansour, A., Blanchard, S., Kobayashi, I., Keats, B.J., Slim, R., Petit, C., 2000. A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C. Nat. Genet. 26, 51–55.
Wang, C.K., Pan, L., Chen, J., Zhang, M., 2010. Extensions of PDZ domains as important structural and functional elements. Protein Cell. 1, 737–751.
Weatheritt, R.J., Luck, K., Petsalaki, E., Davey, N.E., Gibson, T.J., 2012. The identification of short linear motif-mediated interfaces within the human interactome. Bioinformatics 28, 976–982.

Liste des Abréviations

Å	Angströms
Ac	Acétylation
Ank	Ankyrin repeats
ARN	Acide RiboNucléique
ATP	Adénosine Tri-Phosphatae
AUC	Analytical UltraCentrifugation
BEST	Band-selective Excitation Short-Transient
BI	Binding intensities
BLAST	Basic Local Alignment Search Tool
CCE	Cellules ciliées Externes
CCI	Cellules ciliées Internes
Cdh23	Cadherin 23
CSP	Perturbations de déplacement chimique
Da	Dalton
dB	Décibel
DD	Death Domain
DFNA	Deafness autosomal dominant
DFNB	Deafness autosomal recessive
DFNX	Deafness X-linked
DLS	Dynamic Light Scattering
Dmax	Maximum interatomic distance
E. *coli*	*Escherichia coli*
EAR	Epilepsy Associated Repeat
FERM	4.1 protein, Ezrin, Radixin, Moesin
FRET	Transfert d'énergie par résonance de type Förster
GST	Glutatione-S-transferase
GuK	Guanlylate Kinase
HHD	Harmonin Homology domain
Hrmn	Harmonine
HSQC	Heteronuclear Single Quantum Coherence
Ig	Immunoglobuline
IQ	Isoleucine, Glutamine enriched
ITC	Isothermal titration calorimetric
K.O	Knock Out
MAGUK	Membrane-Associated guanylate kinase
MIP	Motifs d'interaction aux domaines PDZ
Mw	Molecular Weight
MyTH4	Myosin Tail Homology 4
PAH	Paired Amphipatic Helix
PB1	Phox and Bem1
Pcdh15	Protocadherin 15
PDB	Protein Data Bank

PDZ	PSD-95, Dlg1, Zonula occludens 1
PEG	Poly-éthylène Glycol
pH	Potentiel hydropgène
pI	Potentiel Ionique
PIP	Phosphatidyl inositol Phosphate
PRE	Paramagnetic Relaxation Enhancement
PSD	Post Synaptic density
PSG	PDZ, SH3, GuK
PSI-BLAST	Position-specific Iteration Basic Local Alignment Search Tool
PSSM	Position-specific scoring matrix
PTM	Post Translationnal Modification
RDC	Residual Dipolar Coupling
Rg	Radius of gyration
RMN	Résonance Magnétique Nucléaire
RMSD	Root Mean Square Deviation
RP	Retinite Pigmentaire
RRM	RNA recognition motif
SAH	Stable Single Alpha Helix
SAM	Sterile alpha motif
SAP	SAF-A/B, Acinus, PIAS
SAXS	Small Angle X-Ray Scattering
SH3	SRC Homology 3
SLiM	Short linear motif
SNC	Système Nerveux central
SPR	Surface Plasmonic resonance
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus
TMC	Trans-Membrane, Channel-like
TROSY	Transverse relaxation optimized spectroscopy
Whrn	Whirline

Liste des Figures

1	01_01	Anatomie de l'oreille
2	01_02	Le système vestibulaire
3	01_03	Morphologie de la cochlée
4	01_04	L'organe de Corti
5	01_05	traitement de l'onde sonore par la cochlée
6	01_06	Cellules ciliées
7	01_07	Structure des touffes ciliaires
8	01_08	Mécanisme de la mécano-transduction
9	01_09	Développement de la touffe ciliaire
10	02_01	Causes des surdités congénitales
11	02_02	Morphologie des photorécepteurs humains
12	03_01	Les modules d'interaction et leurs motifs associés
13	03_02	Modes de régulation des voies de signalisations au niveau protéique
14	03_03	Similarité de séquences entre les domaines HHD identifiés et des domaines PAH
15	03_04	Interactions canoniques et non-canoniques des domaines HHD.
16	03_05	Occurrence des domaines PDZ et diversité des protéines les contenant
17	03_06	Structure consensus des domaines PDZ et leur interaction canonique
18	03_07	Complexes PDZ/PBM caractéristiques des trois classes
19	03_08	Modification de classe de PSD95-PDZ3 par mutagénèse dirigée
20	03_09	Dimérisation des domaines PDZ
21	03_10	Chemin énergétique conservé des domaines PDZ
22	03_11	Extensions des domaines PDZ
23	03_12	Structures de tandem PDZ résolues à ce jour
24	03_13	Supramodules connues des domaines PDZ
25	04_01	Organisation en domaines des protéines Usher et de leurs isoformes identifiées
26	04_02	Interactions entre protéines Usher et protéines associées
27	04_03	Structure du lien apical
28	04_04	Supramodule hétérotypique de l'Harmonine
29	04_05	Interactions entre protéines Usher1.
30	04_06	Organisation du complexe du point inférieur d'ancrage
31	04_07	Organisation du complexe des liens basaux
32	05_01	Isoformes et mutations symptomatiques de la Whirline
33	05_02	Structures des domaines PDZ de la Whirline humaine
34	05_03	Localisation des isoformes de la Whirline dans les photorécepteurs et les cellules ciliées
35	05_04	Changement de morphologie des touffes ciliaires chez les souris whirler
36	05_05	Effet de mutations de l'isoforme longue de Whirline sur les photorécepteurs et les cellules ciliées externes
37	05_06	Interactants de la Whirline, par domaine
38	05_07	Organisation moléculaire du complexe Usher2
39	05_08	Organisation moléculaire du complexe de l'apex des stéréocils
40	06_01	Information sur la Whirline au début de la thèse
41	07_01	Prédiction de tendance au désordre de la séquence de la Whirline

42	07_02	Organisation en domaines des paralogues de la Whirline
43	07_03	Conservation des séquences épingle-ß de l'audition
44	07_04	Constructions produites au cours de la thèse
45	07_05	Sédimentation de Nter-PDZ1
46	07_06	Substitution sur le PDZ1 de la Whirline
47	07_07	Meilleurs conditions de purification de Nter-PDZ1
48	08_01	HHD in proteins
49	08_02	Structure of the HHD2 in the five-helix conformation
50	08_03	HHD2 swapped-dimer conformation
51	08_04	HHD2 behaviour in solution and dynamic
52	08_05	HHD2 in the molecular context of whirlin
53	08_06	Diffusion aux petits angles de HHD2 en solution
54	08_07	K473 du HHD2
55	09_01	1H-15N HSQC spectrum of 0.6 mM 15N/13C/2H-labelled Whirlin PDZ1-PDZ2
56	09_02	Secondary structure prediction and relaxation of Whirlin PDZ1-PDZ2
57	09_03	Whirlin PDZ1-2 tandem dynamic in solution.
58	09_04	Effect of mutation in the C-terminal hairpin extensions on the overall conformation of the PDZ1-2
59	09_05	Correlation between the structural ensemble and the experimental PRE and RDC used as restraints in the docking calculation
60	09_06	Structural model of whirlin PDZ1-2 supramodule
61	09_07	Comparison of PDZ1-2 mutation effects with the PDZ1-2 supramodule model
62	09_08	Selection of conformations describing PDZ1-2 equilibrium
63	09_09	Interaction between whirlin and the peptide Sans
64	09_10	Superimposition of whirlin PDZ tandem with harmonin N-terminal supramodule
65	09_11	Effet de la mutation PDZ1-2$^{\Delta 15}$ sur la conformation du tandem PDZ1-2.
66	10_01	Comparaison des interactions décrites dans la littérature et de celles observées *in vitro* en SPR
67	10_02	BI des constructions de Whirline pour leur partenaires
68	10_03	Emission de fluorescence des constructions de la Whirline en presence de SANS
69	10_04	Affinités déterminées par titration Octet entre SANS et la Whirline.
70	10_05	Interaction de PDZ1 avec les lipides
71	11_01	Questions ouvertes sur les complexes de la Whirline
72	12_01	Séquence de la Whirline avec son organisation en domaine et régions

Liste des Tables

1 02_1 Caractéristiques cliniques des syndromes Usher
2 02_2 Protéines associées aux différents types de syndrome Usher
3 07_1 Taux d'identité entre HHD
4 10_1 Affinités déterminé par SPR des domaines de la Whirline pour les peptides Usher
5 10_2 Interactions déterminés par Hold-Up des domaines de la Whirline pour les peptides Usher.
6 10_3 Affinités déterminées par titrage RMN entre SANS et la Whirline
7 10_4 Affinités déterminées par titrage Octet entre SANS et la Whirline
8 12_1 Résidus du PDZ1-2 en équilibre conformationnel
9 12_3 Interactants décris de la Whirline

Bibliographie

1. Moore, B. C. J., Tyler, L. K. & Marslen-wilson, W. in *The perception of speech : from sound to meaning* 917–921 (2008). doi:10.1098/rstb.2007.2195
2. World Health Organization. Deafness and hearing loss - Fact sheet. *Who.int* (2017). Available at: http://www.who.int/mediacentre/factsheets/fs300/en/.
3. Bialek, W. & Wit, H. P. Quantum limits to oscillator stability : Theory and experiments on acoustic emissions from the human ear. *Phys. Lett. A* **104,** 173–178 (1984).
4. Avan, P., Büki, B. & Petit, C. Auditory distortions: origins and functions. *Physiol. Rev.* **93,** 1563–619 (2013).
5. Shaw, S. J. Directional Perception in the Human Auditory System. *J. Undergrad. Sci.* **140,** 135–140 (1996).
6. Kandel, E. R., Schwartz, J. H. & Jessell, T. M. *Principal of Neural Science*. *4th edition* (2000).
7. Widmaier, E. P., Raff, H. & Strang, K. *Physiologie Humaine - Les mécanismes du fonctionnement de l'organisme*. (2009).
8. Droual, R. & Stanfield, C. L. Notes from Principles of Human Physiology Chapter 10 - Sensory. (2010). Available at: http://droualb.faculty.mjc.edu/Course Materials/Physiology 101/Chapter Notes/Fall 2007/chapter_10 Fall 2007.htm.
9. Von Békésy, G. *Experiments in hearing*. (1960).
10. Alberti, P. W. in *Occupational exposure to noise: Evaluation, prevention, and control* (2001).
11. Helmhotz, H. *The Bonn lecture: popular lectures on scientific subjects*. (1857).
12. Fettiplace, R. & Kim, K. X. The physiology of mechanoelectrical transduction channels in hearing. *Physiol. Rev.* **94,** 951–86 (2014).
13. Verpy, E. *et al.* Stereocilin connects outer-hair-cell stereocilia to one another and to the tectorial membrane. *J Comp Neurol* **519,** 194–210 (2012).
14. Pujol, R., Lenoir, M., Ladrech, S., Tribillac, F. & Rebillard, G. in *Auditory, Physiology and Perception* (1992). doi:10.1016/B978-0-08-041847-6.50011-3
15. Lim, D. J. Functional structure of the organ of Corti: a review. *Hear. Res.* **22,** 117–146 (1986).
16. Mann, Z. F. & Kelley, M. W. Development of tonotopy in the auditory periphery. *Hear. Res.* **276,** 2–15 (2011).
17. Oghalai, J. S. The cochlear amplifier: augmentation of the traveling wave within the inner ear. *Curr. Opin. Otolaryngol. Head Neck Surg.* (2004).
18. 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. *J. Membr. Biol.* **209,** 119–134 (2006).
19. Ashmore, J. *et al.* The remarkable cochlear amplifier. *Hear. Res.* **266,** 1–17 (2010).
20. Appler, J. M. & Goodrich, L. V. Connecting the ear to the brain: Molecular mechanisms of auditory circuit assembly. *Prog. Neurobiol.* **93,** 488–508 (2011).
21. Dallos, P., Billone, M. C., Durrant, J. D., Wang, C.-Y. & Raynor, S. Cochlear Inner and Outer Hair Cells: Functional Differences. *Science (80-.).* (1972).
22. Safieddine, S., El-Amraoui, A. & Petit, C. The auditory hair cell ribbon synapse: from assembly to function. *Annu. Rev. Neurosci.* **35,** 509–28 (2012).
23. Saenz, M. & Langers, D. R. M. Tonotopic mapping of human auditory cortex. *Hear.*

Res. **307,** 42–52 (2014).
24. Hackney, C. M. & Furness, D. N. The composition and role of cross links in mechanoelectrical transduction in vertebrate sensory hair cells. *J. Cell Sci.* **126,** 1721–31 (2013).
25. Tilney, L. G., Derosier, D. J. & Mulroy, M. J. The organization of actin-filaments in the stereocilia of cochlear hair cells. *J. Cell Biol.* **86,** 244–259 (1980).
26. McGrath, J., Roy, P. & Perrin, B. J. Stereocilia morphogenesis and maintenance through regulation of actin stability. *Semin. Cell Dev. Biol.* 1–8 (2016). doi:10.1016/j.semcdb.2016.08.017
27. Assad, J. A., Shepherd, G. M. G. & Corey, D. P. Tip-link integrity and mechanical transduction in vertebrate hair cells. *Neuron* **7,** 985–994 (1991).
28. Zhao, B. & Müller, U. The elusive mechanotransduction machinery of hair cells. *Curr. Opin. Neurobiol.* **34,** 172–179 (2015).
29. Markin, V. S. & Hudspeth, A. J. GATING-SPRING MODELS OF MECHANOELECTRICAL TRANSDUCTION BY HAIR CELLS OF THE INTERNAL EAR. *Annu. Rev. Biophys. Biomol. Struct.* **24,** 59–83 (1995).
30. Schwander, M., Kachar, B. & Müller, U. The cell biology of hearing. *J. Cell. Biochem.* **190,** 9–20 (2010).
31. Lu, X. & Sipe, C. W. Developmental regulation of planar cell polarity and hair-bundle morphogenesis in auditory hair cells : lessons from human and mouse genetics. *WIREs develpmental Biol.* **5,** 85–101 (2016).
32. Michalski, N. & Petit, C. Genetics of auditory mechano-electrical transduction. *Pflugers Arch.* **467,** 49–72 (2015).
33. Zou, J. *et al.* Individual USH2 proteins make distinct contributions to the ankle link complex during development of the mouse cochlear stereociliary bundle. *Hum. Mol. Genet.* **24,** 6944–6957 (2015).
34. Richardson, G. P., de Monvel, J. B. & Petit, C. How the genetics of deafness illuminates auditory physiology. *Annu. Rev. Physiol.* **73,** 311–34 (2011).
35. WHO. WHO global estimates on prevalence of hearing loss : Mortality and Burden of Diseases. (2012).
36. Pujol, R. Journey into the world of hearing. (1999). Available at: http://www.cochlea.eu/.
37. Korver, A. M. H. *et al.* Congenital hearing loss. *Nat. Rev. Dis. Prim.* **3,** 1–17 (2017).
38. Van Camp, G., Smith, R. J. H., Sommen, M. & Shearer, A. Hereditary Hearing loss - Database. Available at: http://hereditaryhearingloss.org.
39. Morton, C. C. & Nance, W. E. Newborn Hearing Screening — A Silent Revolution. *New Engl. J. Med.* **354,** 2151–2164 (2006).
40. Lentz, J. & Keats, B. in *Gene Review* 1–29 (2017).
41. El-amraoui, A. & Petit, C. The retinal phenotype of Usher syndrome : Pathophysiological insights from animal models. *Comptes rendus - Biol.* **337,** 167–177 (2014).
42. Lentz, J. & Keats, B. in *Gene Review* 1–20 (2017).
43. Molday, R. S. & Moritz, O. L. Photoreceptors at a glance. *J. Cell Sci.* **128,** 4039–4045 (2015).

44. Kocaoglu, O. *et al.* Photoreceptor disc shedding in the living human eye. *Biomed. Opt. Express* **7,** 4554–4568 (2016).
45. Sahly, I. *et al.* Localization of Usher 1 proteins to the photoreceptor calyceal processes, which are absent from mice. *J. Cell Biol.* **199,** 381–99 (2012).
46. Mathur, P. & Yang, J. Usher syndrome: Hearing loss, retinal degeneration and associated abnormalities. *Biochim. Biophys. Acta* **1852,** 406–420 (2015).
47. Cosgrove, D. & Zallocchi, M. Usher protein functions in hair cells and photoreceptors. *Int. J. Biochem. Cell Biol.* **46,** 80–9 (2014).
48. Edelman, G. M. & Gall, W. E. The antibody problem. *Annu. Rev. Biochem.* **38,** (1969).
49. Wetlaufer, D. B. Nucleation, Rapid Folding, and Globular Intrachain Regions in Proteins. *Proc. Natl. Acad. Sci.* **70,** 697–701 (1973).
50. Liu, W. *et al.* The INAD scaffold is a dynamic, redox-regulated modulator of signaling in the Drosophila eye. *Cell* **145,** 1088–1101 (2011).
51. Yan, J. *et al.* Structure of the split PH domain and distinct lipid- binding properties of the PH – PDZ supramodule of a -syntrophin. *EMBO J.* **24,** 3985–3995 (2005).
52. Lin, M. M. & Zewail, A. H. Hydrophobic forces and the length limit of foldable protein domains. *Proc. Natl. Acad. Sci.* **109,** (2012).
53. Bork, P., Downing, A. K., Kieffer, B. & Campbell, I. D. Structure and distribution of modules in extracellular proteins. *Q. Rev. Biophys.* **29,** 119–167 (1996).
54. Patthy, L. Exons - original building blocks of Proteins. *BioEssays* **13,** (1991).
55. Pawson, T. & Nash, P. Assembly of Cell Regulatory Systems Through. *Science (80-.).* **300,** 445–452 (2003).
56. Van Roey, K. *et al.* Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* **114,** 6733–6778 (2014).
57. Brocchieri, L. & Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* **33,** 3390–3400 (2005).
58. Hegyi, H. & Bork, P. On the classification and evolution of Protein Modules. *J. Protein Chem.* **16,** 6–13 (1997).
59. Bhattacharyya, R. P., Reményi, A., Yeh, B. J. & Lim, W. a. Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75,** 655–680 (2006).
60. Maisonneuve, P. *et al.* Regulation of the Catalytic Activity of the Human Phosphatase PTPN4 by Its PDZ Domain. *FEBS J.* **281,** 4852–4865 (2014).
61. Van Itallie, C. M. & Anderson, J. M. Architecture of tight junctions and principles of molecular composition. *Semin. Cell Dev. Biol.* **36,** 157–165 (2014).
62. Feng, W. & Zhang, M. Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. *Nat. Rev. Neurosci.* **10,** 87–99 (2009).
63. 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* **82,** 897–903 (2014).
64. Pan, L., Yan, J., Wu, L. & Zhang, M. Assembling stable hair cell tip link complex via multidentate interactions between harmonin and cadherin 23. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 5575–80 (2009).
65. Fisher, O. S. *et al.* Structural studies of cerebral cavernous malformations 2 (CCM2)

reveal a folded helical domain at its C-terminus. *FEBS Lett.* **587,** 272–277 (2013).
66. Wang, X. *et al.* Structural Insights into the Molecular Recognition between Cerebral Cavernous Malformation 2 and Article Structural Insights into the Molecular Recognition between Cerebral Cavernous Malformation 2 and Mitogen-Activated Protein Kinase Kinase Kinase 3. *Structure* **23,** 1–10 (2015).
67. Fisher, O. S. *et al.* Structure and vascular function of MEKK3–cerebral cavernous malformations 2 complex. *Nat. Commun.* **6,** 7937 (2015).
68. 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,** 33460–71 (2012).
69. van Ingen, H., Baltussen, M. A. H., Aelen, J. & Vuister, G. W. Role of Structural and Dynamical Plasticity in Sin3: The Free PAH2 Domain is a Folded Module in mSin3B. *J. Mol. Biol.* **358,** 485–497 (2006).
70. Xie, T. *et al.* Structure of the 30-kDa Sin3-associated protein (SAP30) in complex with the mammalian Sin3A corepressor and its role in nucleic acid binding. *J. Biol. Chem.* **286,** 27814–27824 (2011).
71. 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,** 4040–5 (2010).
72. Kennedy, M. B. Origin of PDZ (DHR, GLGF) domains. *Trends Biochem. Sci.* **20,** 350 (1995).
73. 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,** 929–42 (1992).
74. Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* **114,** 3219–31 (2001).
75. Pallen, M. J. & Ponting, C. P. PDZ domains in bacterial proteins. *Mol. Microbiol.* **26,** 411–413 (1997).
76. Luck, K., Charbonnier, S. & Travé, G. The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett.* **586,** 2648–61 (2012).
77. Nourry, C., Grant, S. G. N. & Borg, J.-P. PDZ domain proteins: plug and play! *Sci. STKE* **2003,** RE7 (2003).
78. Kim, E. & Sheng, M. PDZ domain proteins of synapses. *Nat. Rev. Neurosci.* **5,** 771–81 (2004).
79. Hurd, T. W., Gao, L., Roh, M. H., Macara, I. G. & Margolis, B. Direct interaction of two polarity complexes implicated in epithelial tight junction assembly. *Nat. Cell Biol.* **5,** 137–142 (2003).
80. Suzuki, T., Ohsugi, Y., Uchida-Toita, M., Akiyama, T. & Yoshida, M. Tax oncoprotein of HTLV-1 binds to the human homologue of Drosophila discs large tumor suppressor protein, hDLG, and perturbs its function in cell growth control. *Oncogene* **18,** 5967–72 (1999).
81. Reiners, J. *et al.* Scaffold protein harmonin (USH1C) provides molecular links between Usher syndrome type 1 and type 2. *Hum. Mol. Genet.* **14,** 3933–43 (2005).

82. Préhaud, C. *et al.* Attenuation of rabies virulence: takeover by the cytoplasmic domain of its envelope protein. *Sci. Signal.* **3,** ra5 (2010).
83. Nair, D. *et al.* Super-Resolution Imaging Reveals That AMPA Receptors Inside Synapses Are Dynamically Organized in Nanodomains Regulated by PSD95. *J. Neurosci.* **33,** 13204–13224 (2013).
84. Terrien, E. *et al.* Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. *Sci. Signal.* **5,** ra58 (2012).
85. Cabral, J. H. M. *et al.* Crystal structure of a PDZ domain. *Nature* **382,** 649–652 (1996).
86. Doyle, D. A. *et al.* Crystal structures of a complexed and peptide-free membrane protein- binding domain: Molecular basis of peptide recognition by PDZ. *Cell* **85,** 1067–1076 (1996).
87. Saras, J. & Heldin, C.-H. PDZ domains bind carboxy.terminal sequences of target proteins. *Trends Biochem. Sci.* **4,** 455–458 (1996).
88. Songyang, Z. *et al.* Recognition of Unique Carboxyl-Terminal Motifs by Distinct PDZ Domains. *Science (80-.).* **275,** 73–77 (1997).
89. Gerek, Z. N. & Ozkan, S. B. Change in allosteric network affects binding affinities of PDZ domains: analysis through perturbation response scanning. *PLoS Comput. Biol.* **7,** e1002154 (2011).
90. Erlendsson, S. *et al.* PICK1 (protein interacting with C-kinase 1) binding promiscuity relies on unconventional PDZ (PSD-95/Discs-large/ZO-1 homology) binding modes for non-class II PDZ ligands. *J. Biol. Chem.* **Epub ahead,** 1–32 (2014).
91. Delhommel, F. *et al.* Deciphering the unconventional peptide binding to the PDZ domain of MAST2. *Biochem. J.* **469,** 159–168 (2015).
92. Chi, C. N. *et al.* Interactions outside the boundaries of the canonical binding groove of a PDZ domain influence ligand binding. *Biochemistry* **51,** 8971–9 (2012).
93. 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,** E3081-90 (2016).
94. Tonikian, R. *et al.* A specificity map for the PDZ domain family. *PLoS Biol.* **6,** e239 (2008).
95. Bezprozvanny, I. & Maximov, A. Classification of PDZ domains. *FEBS Lett.* **509,** (2001).
96. McLaughlin, R. N., Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491,** 138–42 (2012).
97. Raman, A. S., White, K. I. & Ranganathan, R. Origins of Allostery and Evolvability in Proteins: A Case Study. *Cell* **166,** 468–481 (2016).
98. Penkert, R., DiVittorio, H. M. & Prehoda, K. E. Internal Recognition Through PDZ Domain Plasticity in the Par-6 - Pals1 Complex. *Nat. Struct. Mol. Biol.* **11,** 1122–1127 (2004).
99. Lemaire, J. F. & McPherson, P. S. Binding of Vac14 to neuronal nitric oxide synthase: Characterisation of a new internal PDZ-recognition motif. *FEBS Lett.* **580,** 6948–6954 (2006).
100. 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).
101. Hillier, B. J., Bredt, D. S. & Lim, W. A. Unexpected Modes of PDZ Domain Scaffolding Revealed by Structure of nNOS-Syntrophin Complex. *Science (80-.).* **812,** (2008).
102. Tochio, H. *et al.* Formation of nNOS/PSD-95 PDZ dimer requires a preformed beta-finger structure from the nNOS PDZ domain. *J. Mol. Biol.* **303,** 359–370 (2000).
103. Harris, B. Z., Hillier, B. J. & Lim, W. A. Energetic determinants of internal motif recognition by PDZ domains. *Biochemistry* **40,** 5921–5930 (2001).
104. 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,** 671 (2010).
105. 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).
106. Ivarsson, Y. Plasticity of PDZ domains in ligand recognition and signaling. *FEBS Lett.* **586,** 2638–47 (2012).
107. Erlendsson, S. & Madsen, K. Membrane Binding and Modulation of the PDZ Domain of PICK1. *Membranes (Basel).* **5,** 597–615 (2015).
108. Gallardo, R., Ivarsson, Y., Schymkowitz, J., Rousseau, F. & Zimmermann, P. Structural diversity of PDZ-lipid interactions. *ChemBioChem* **11,** 456–467 (2010).
109. Zimmermann, P. *et al.* PIP2-PDZ domain binding controls the association of syntenin with the plasma membrane. *Mol. Cell* **9,** 1215–1225 (2002).
110. Pan, L. *et al.* Clustering and synaptic targeting of PICK1 requires direct interaction between the PDZ domain and lipid membranes. *Embo J* **26,** 4576–4587 (2007).
111. Chen, Y. *et al.* Genome-wide Functional Annotation of Dual-Specificity Protein- and Lipid-Binding Modules that Regulate Protein Interactions. *Mol. Cell* **46,** 226–237 (2012).
112. Wu, H. *et al.* PDZ Domains of Par-3 as Potential Phosphoinositide Signaling Integrators. *Mol. Cell* **28,** 886–898 (2007).
113. 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,** e54581 (2013).
114. Yan, J. *et al.* Structure of the split PH domain and distinct lipid-binding properties of the PH-PDZ supramodule of alpha-syntrophin. *Embo J* **24,** 3985–3995 (2005).
115. Chang, B. H. *et al.* A systematic family-wide investigation reveals that ~30% of mammalian PDZ domains engage in PDZ-PDZ interactions. *Chem. Biol.* **18,** 1143–52 (2011).
116. Im, Y. J. *et al.* Crystal Structure of GRIP1 PDZ6-Peptide Complex Reveals the Structural Basis for Class II PDZ Target Recognition and PDZ Domain-mediated Multimerization. *J. Biol. Chem.* **278,** 8501–8507 (2002).
117. Im, Y. J. *et al.* Crystal structure of the Shank PDZ-ligand complex reveals a class I PDZ interaction and a novel PDZ-PDZ dimerization. *J. Biol. Chem.* **278,** 48099–104 (2003).
118. Sugi, T. *et al.* Crystal structures of autoinhibitory PDZ domain of Tamalin: implications for metabotropic glutamate receptor trafficking regulation. *EMBO J.* **26,**

2192–205 (2007).
119. Elkins, J. M. *et al.* Unusual binding interactions in PDZ domain crystal structures help explain binding mechanisms. *Protein Sci.* **19,** 731–41 (2010).
120. Chen, J., Pan, L., Wei, Z., Zhao, Y. & Zhang, M. Domain-swapped dimerization of ZO-1 PDZ2 generates specific and regulatory connexin43-binding sites. *EMBO J.* **27,** 2113–23 (2008).
121. Ponting, C. P., Phillips, C., Davies, K. E. & Blake, D. J. PDZ domains: targeting signalling molecules to sub-membranous sites. *BioEssays* **19,** 469–477 (1997).
122. Fanning, A. S. & Anderson, J. M. Protein – protein interactions : PDZ domain networks. *Curr. Biol.* **6,** 1385–1388 (1996).
123. Das, S., Dawson, N. L. & Orengo, C. A. Diversity in protein domain superfamilies. *Curr. Opin. Genet. Dev.* **35,** 40–49 (2015).
124. Ye, F. & Zhang, M. Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. *Biochem. J.* **455,** 1–14 (2013).
125. Smock, R. G. & Gierasch, L. M. Sending signals dynamically. *Science (80-.).* **324,** 198–203 (2009).
126. Lockless, S. W. & Ranganathan, R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science (80-.).* **286,** 295–299 (1999).
127. Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138,** 774–86 (2009).
128. Fuentes, E. J., Der, C. J. & Lee, A. L. Ligand-dependent Dynamics and Intramolecular Signaling in a PDZ Domain. *J. Mol. Biol.* **335,** 1105–1115 (2004).
129. Dhulesia, A., Gsponer, J. & Vendruscolo, M. Mapping of two networks of residues that exhibit structural and dynamical changes upon binding in a PDZ domain protein. *J. Am. Chem. Soc.* **130,** 8931–9 (2008).
130. Reynolds, K. A., Mclaughlin, R. N. & Ranganathan, R. Hot Spots for Allosteric Regulation on Protein Surfaces. *Cell* **147,** 1564–1575 (2011).
131. Gianni, S. *et al.* Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. *Structure* **14,** 1801–9 (2006).
132. Wang, C. K., Pan, L., Chen, J. & Zhang, M. Extensions of PDZ domains as important structural and functional elements. *Protein Cell* **1,** 737–51 (2010).
133. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 3-18249–18254 (2009).
134. Zhang, J., Petit, C. M., King, D. S. & Lee, A. L. Phosphorylation of a PDZ Domain Extension Modulates Binding Affinity and Interdomain Interactions in Postsynaptic Density-95 (PSD-95) Protein , a Membrane-associated Guanylate Kinase (MAGUK) * □. *J. Biol. Chem.* **286,** 41776–41785 (2011).
135. Bhattacharya, S. *et al.* Ligand-Induced Dynamic Changes in Extended PDZ Domains from NHERF1. *J. Mol. Biol.* (2013). doi:10.1016/j.jmb.2013.04.001
136. Charbonnier, S. *et al.* The Structural and Dynamic Response of MAGI-1 PDZ1 with Noncanonical Domain Boundaries to the Binding of Human Papillomavirus E6. *J. Mol. Biol.* **406,** 745–763 (2011).
137. Ramirez, 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).
138. Feng, W., Wu, H., Chan, L.-N. & Zhang, M. Par-3-mediated junctional localization of the lipid phosphatase PTEN is required for cell polarity establishment. *J. Biol. Chem.* **283,** 23440–9 (2008).
139. Elkins, J. M. *et al.* Structure of PICK1 and other PDZ domains obtained with the help of self-binding C-terminal extensions. *Protein Sci.* **16,** 683–694 (2007).
140. Balana, B. *et al.* Mechanism underlying selective regulation of G protein-gated inwardly rectifying potassium channels by the psychostimulant-sensitive sorting nexin 27. **108,** 5831–5836 (2011).
141. Im, Y. J. *et al.* Structural Basis for Asymmetric Association of the β PIX Coiled Coil and Shank PDZ. *J. Mol. Biol.* **397,** 457–466 (2010).
142. Birrane, G., Chung, J. & Ladias, J. a a. Novel mode of ligand recognition by the Erbin PDZ domain. *J. Biol. Chem.* **278,** 1399–402 (2003).
143. Kang, B. S. *et al.* PDZ tandem of human syntenin: Crystal structure and functional properties. *Structure* **11,** 459–468 (2003).
144. 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).
145. 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 *. **275,** 19933–19941 (2000).
146. Grembecka, J. *et al.* The Binding of the PDZ Tandem of Syntenin to Target Proteins †,‡. *Biochemistry* **45,** 3674–3683 (2006).
147. Dong, H. *et al.* GRIP: a synaptic PDZ domain-containing protein that interact with AMPA receptors. *Nature* **386,** 279–284 (1997).
148. Long, J. *et al.* Supramodular Nature of GRIP1 Revealed by the Structure of Its PDZ12 Tandem in Complex with the Carboxyl Tail of Fras1. *J. Mol. Biol.* **375,** 1457–1468 (2008).
149. 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. *Nat. Struct. Biol.* **10,** 972–978 (2003).
150. Long, J. F. *et al.* Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95. *J. Mol. Biol.* **327,** 203–214 (2003).
151. Bach, A. *et al.* A high-affinity, dimeric inhibitor of PSD-95 bivalently interacts with PDZ1-2 and protects against ischemic brain damage. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 3317–22 (2012).
152. Mishra, P. *et al.* Dynamic scaffolding in a G protein-coupled signaling system. *Cell* **131,** 80–92 (2007).
153. Huang, J., Liu, C., Hughes, S. A., Postma, M. & Schwiening, C. J. Article Activation of TRP Channels by Protons and Phosphoinositide Depletion in Drosophila Photoreceptors. *Curr. Biol.* **20,** 189–197 (2010).
154. Ren, J. *et al.* Interdomain interface-mediated target recognition by the Scribble PDZ34 supramodule. *Biochem. J.* **468,** 133–144 (2015).
155. Jiang, J. *et al.* Activation of DegP chaperone-protease via formation of large cage-like

oligomers upon binding to substrate proteins. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 11939–11944 (2008).
156. Zhu, J., Shang, Y. & Zhang, M. Mechanistic basis of MAGUK-organized complexes in synaptic development and signalling. *Nat. Rev. Neurosci.* **17,** 209–223 (2016).
157. Li, Y. *et al.* Structure of Crumbs tail in complex with the PALS1 PDZ-SH3-GK tandem reveals a highly specific assembly mechanism for the apical Crumbs complex. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 2–7 (2014).
158. Nomme, J. *et al.* The Src homology 3 domain is required for junctional adhesion molecule binding to the third PDZ domain of the scaffolding protein ZO-1. *J. Biol. Chem.* **286,** 43352–43360 (2011).
159. Pan, L., Chen, J., Yu, J., Yu, H. & Zhang, M. The structure of the PDZ3-SH3-GuK tandem of ZO-1 protein suggests a supramodular organization of the membrane-associated guanylate kinase (MAGUK) family scaffold protein core. *J. Biol. Chem.* **286,** 40069–40074 (2011).
160. Chen, K.-E. *et al.* Reciprocal allosteric regulation of p38γ and PTPN3 involves a PDZ domain-modulated complex formation. *Sci. Signal.* **7,** 1–10 (2014).
161. 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. *J. Biol. Chem.* **291,** 16699–16708 (2016).
162. Krojer, T. *et al.* Interplay of PDZ and protease domain of DegP ensures efficient elimination of misfolded proteins. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7702–7 (2008).
163. Wilken, C., Kitzing, K., Kurzbauer, R., Ehrmann, M. & Clausen, T. Crystal Structure of the DegS Stress Sensor. *Cell* **117,** 483–494 (2004).
164. Sohn, J., Grant, R. A. & Sauer, R. T. Allosteric Activation of DegS, a Stress Sensor PDZ Protease. *Cell* **131,** 572–583 (2007).
165. Pan, L. & Zhang, M. Structures of usher syndrome 1 proteins and their complexes. *Physiology* **27,** 25–42 (2012).
166. Boëda, B. *et al.* Myosin VIIa, harmonin and cadherin 23, three Usher I gene products that cooperate to shape the sensory hair cell bundle. *EMBO J.* **21,** 6689–99 (2002).
167. Müller, U. Cadherins and Mechanotransduction by Hair Cells. *Curr. Opin. Cell Biol.* **20,** 557–566 (2008).
168. Kachar, B., Parakkal, M., Kurc, M., Zhao, Y. & Gillespie, P. G. High-resolution structure of hair-cell tip links. *Proc. Natl. Acad. Sci.* **97,** 13336–13341 (2000).
169. Araya-Secchi, R., Neel, B. L. & Sotomayor, M. An Elastic Element in the Protocadherin-15 Tip Link of the Inner Ear. *Nat. Commun.* **7,** 1–14 (2016).
170. Kazmierczak, P. *et al.* Cadherin 23 and protocadherin 15 interact to form tip-link filaments in sensory hair cells. *Nature* **449,** 87–92 (2007).
171. Sotomayor, M., Weihofen, W. A., Gaudet, R. & Corey, D. P. Structure of a force-conveying cadherin bond essential for inner-ear mechanotransduction. *Nature* **492,** 128–132 (2012).
172. Powers, R. E., Gaudet, R. & Sotomayor, M. A Partial Calcium-Free Linker Confers Flexibility to Inner-Ear Protocadherin-15. *Structure* **25,** 482–495 (2017).
173. Li, J. *et al.* Ca 2 + -Induced Rigidity Change of the Myosin VIIa IQ Motif-Single a Helix Lever Arm Extension Article Ca 2 + -Induced Rigidity Change of the Myosin

VIIa IQ Motif-Single a Helix Lever Arm Extension. *Structure* **25,** 579–591.e4 (2017).
174. Weck, M. L., Grega-larson, N. E. & Tyska, M. J. MyTH4-FERM myosins in the assembly and maintenance of actin-based protrusions. *Curr. Opin. Cell Biol.* 1–11 (2016). doi:10.1016/j.ceb.2016.10.002
175. Sakamoto, T., Zhang, F., Kiehart, D. P., Sellers, J. R. & Yang, Y. Dimerized Drosophila myosin VIIa : A processive motor. *Proc. Natl. Acad. Sci.* **103,** (2006).
176. Wei, Z., Liu, X., Yu, C. & Zhang, M. Structural basis of cargo recognitions for class V myosins. *Proc. Natl. Acad. Sci.* **110,** (2013).
177. Morgan, C. P. *et al.* PDZD7-MYO7A complex identified in enriched stereocilia membranes. *Elife* 1–34 (2016). doi:10.7554/eLife.18312
178. Delprat, B. *et al.* Myosin XVa and whirlin, two deafness gene products required for hair bundle growth, are located at the stereocilia tips and interact directly. *Hum. Mol. Genet.* **14,** 401–10 (2005).
179. Zou, J. *et al.* The roles of USH1 proteins and PDZ domain-containing USH proteins in USH2 complex integrity in cochlear hair cells. *Hum. Mol. Genet.* **26,** 624–636 (2016).
180. Wu, L., Pan, L., Wei, Z. & Zhang, M. Structure of MyTH4-FERM domains in myosin VIIa tail bound to cargo. -Supp. *Science (80-.).* **331,** 757–60 (2011).
181. Weil, D. Usher syndrome type I G (USH1G) is caused by mutations in the gene encoding SANS, a protein that associates with the USH1C protein, harmonin. *Hum. Mol. Genet.* **12,** 463–471 (2003).
182. Caberlotto, E. *et al.* Usher type 1G protein sans is a critical component of the tip-link complex, a structure controlling actin polymerization in stereocilia. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 5825–30 (2011).
183. Sorusch, N. *et al.* Characterization of the ternary Usher syndrome SANS/ush2a/whirlin protein complex. *Hum. Mol. Genet.* (2017). doi:doi: 10.1093/hmg/ddx027
184. 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. *Nat. Genet.* **26,** 51–5 (2000).
185. Grillet, N. *et al.* Harmonin mutations cause mechanotransduction defects in cochlear hair cells. *Neuron* **62,** 375–87 (2009).
186. Li, J., He, Y., Lu, Q. & Zhang, M. Mechanistic Basis of Organization of the Harmonin/USH1C-Mediated Brush Border Microvilli Tip-Link Complex. *Dev. Cell* **36,** 179–189 (2016).
187. Kazmierczak, P. & Müller, U. Sensing sound: molecules that orchestrate mechanotransduction by hair cells. *Trends Neurosci.* **35,** 220–9 (2012).
188. Wu, L., Pan, L., Wei, Z. & Zhang, M. Structure of MyTH4-FERM Domains in Myosin VIIa Tail Bound to Cargo. *Science (80-.).* **757,** (2011).
189. Stauffer, E. A. & Holt, J. R. Sensory Transduction and Adaptation in Inner and Outer Hair Cells of the Mouse Auditory System. *J. Neurophysiol.* **98,** 3360–3369 (2007).
190. Peng, A. W., Effertz, T. & Ricci, A. J. Adaptation of Mammalian Auditory Hair Cell Mechanotransduction Is Independent of Calcium Entry. *Neuron* **80,** 960–972 (2013).
191. Riazuddin, S. *et al.* Alterations of the CIB2 calcium- and integrin-binding protein cause Usher syndrome type 1J and nonsyndromic deafness DFNB48. *Nat. Genet.* **44,** 1265–71 (2012).

192. Wu, Z. & Müller, U. Molecular Identity of the Mechanotransduction Channel in Hair Cells : Not Quiet There Yet. *J. Neurosci.* **36,** 10927–10934 (2016).
193. Corey, X. D. P. & Holt, J. R. Are TMCs the Mechanotransduction Channels of Vertebrate Hair Cells ? *J. Neurosci.* **36,** 10921–10926 (2016).
194. Xiong, W. *et al.* TMHS is an integral component of the mechanotransduction machinery of cochlear hair cells. *Cell* **151,** 1283–95 (2012).
195. Zhao, B. *et al.* TMIE Is an Essential Component of the Mechanotransduction Machinery of Cochlear Hair Cells. *Neuron* **5,** (2014).
196. Kurima, K. *et al.* Dominant and recessive deafness caused by mutations of a novel gene , TMC1 , required for cochlear hair-cell function. *Nat. Genet.* **30,** 277–284 (2002).
197. Kurima, K. *et al.* TMC1 and TMC2 Localize at the Site of Mechanotransduction in Mammalian Inner Ear Hair Cell Stereocilia. *Cell Rep.* **12,** 1606–1617 (2015).
198. Maeda, R. *et al.* Tip-link protein protocadherin 15 interacts with transmembrane channel-like proteins TMC1 and TMC2. *Proc. Natl. Acad. Sci.* **111,** 12907–12912 (2014).
199. Wu, Z. *et al.* Mechanosensory hair cells express two molecularly distinct mechanotransduction channels. *Nat. Neurosci.* **20,** 24–35 (2017).
200. Fettiplace, R. Biophysical Perspective Is TMC1 the Hair Cell Mechanotransducer Channel ? *Biophys. J.* **111,** 3–9 (2016).
201. Besnard, T. *et al.* Non-USH2A mutations in USH2 patients. *Hum. Mutat.* **33,** 504–10 (2012).
202. Adato, A. *et al.* Usherin, the defective protein in Usher syndrome type IIA, is likely to be a component of interstereocilia ankle links in the inner ear sensory cells. *Hum. Mol. Genet.* **14,** 3921–32 (2005).
203. Michalski, N. *et al.* Molecular characterization of the ankle-link complex in cochlear hair cells and its role in the hair bundle functioning. *J. Neurosci.* **27,** 6478–88 (2007).
204. McMillan, D. R., Kayes-Wandover, K. M., Richardson, J. a & White, P. C. Very large G protein-coupled receptor-1, the largest known cell surface protein, is highly expressed in the developing central nervous system. *J. Biol. Chem.* **277,** 785–92 (2002).
205. Yagi, H., Noguchi, Y., Kitamura, K. & Sato, M. Deficiency of Vlgr1 resulted in deafness and susceptibility to audiogenic seizures while the degree of hearing impairment was not correlated with seizure severity in C57BL/6- and 129-backcrossed lines of Vlgr1 knockout mice. *Neurosci. Lett.* **461,** 190–5 (2009).
206. McGee, J. *et al.* The very large G-protein-coupled receptor VLGR1: a component of the ankle link complex required for the normal development of auditory hair bundles. *J. Neurosci.* **26,** 6543–53 (2006).
207. Sun, J.-P. *et al.* The very large G protein coupled receptor (Vlgr1) in hair cells. *J. Mol. Neurosci.* **50,** 204–14 (2013).
208. Ebermann, I. *et al.* PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. *J. Clin. Invest.* **120,** 1812–1823 (2010).
209. Vona, B. *et al.* Confirmation of PDZD7 as a Nonsyndromic Hearing Loss Gene. *Ear Hear.* **37,** 238–246 (2016).
210. 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 (2014). doi:10.1093/hmg/ddt629

211. Manor, U. *et al.* Regulation of stereocilia length by myosin XVa and whirlin depends on the actin-regulatory protein Eps8. *Curr. Biol.* **21,** 167–72 (2011).
212. Bird, J. E. *et al.* Chaperone-enhanced purification of unconventional myosin 15 , a molecular motor specialized for stereocilia protein trafficking. **111,** 1–6 (2014).
213. Alper, M. *et al.* Mutational Spectrum of MYO15A : The Large N-Terminal Extension of Myosin XVA Is Required for Hearing. *Hum. Mutat.* **28,** 1014–1019 (2007).
214. 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. *J. Biol. Chem.* **289,** 36070–88 (2014).
215. Mustapha, M. *et al.* DFNB31, a recessive form of sensorineural hearing loss, maps to chromosome 9q32-34. *Eur. J. Hum. Genet.* **10,** 210–212 (2002).
216. Mburu, P. *et al.* Defects in whirlin, a PDZ domain molecule involved in stereocilia elongation, cause deafness in the whirler mouse and families with DFNB31. *Nat. Genet.* **34,** 421–8 (2003).
217. 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* 1–17 (2013). doi:10.1002/prot.24438
218. Kersten, F. F. J. *et al.* Association of Whirlin with Ca v 1 . 3 (‗ 1D) Channels in Photoreceptors , Defining a Novel Member of the Usher Protein Network AND. *Invest. Ophthalmol. Vis. Sci.* **3,** 2338–2346 (2009).
219. Delprat, B. *et al.* Myosin XVa and whirlin , two deafness gene products required for hair bundle growth , are located at the stereocilia tips and interact directly. *Hum. Mol. Genet.* **14,** 401–410 (2005).
220. van Wijk, E. *et al.* The DFNB31 gene product whirlin connects to the Usher protein network in the cochlea and retina by direct association with USH2A and VLGR1. *Hum. Mol. Genet.* **15,** 751–65 (2006).
221. Mathur, P. D. *et al.* Distinct expression and function of whirlin isoforms in the inner ear and retina: An insight into pathogenesis of USH2D and DFNB31. *Hum. Mol. Genet.* **24,** 6213–6228 (2015).
222. Wright, R. N., Hong, D.-H. & Perkins, B. RpgrORF15 connects to the usher protein network through direct interactions with multiple whirlin isoforms. *Invest. Ophthalmol. Vis. Sci.* **53,** 1519–29 (2012).
223. 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).
224. Kikkawa, Y. *et al.* Mutant analysis reveals whirlin as a dynamic organizer in the growing hair cell stereocilium. *Hum. Mol. Genet.* **14,** 391–400 (2005).
225. Mburu, P. *et al.* Whirlin complexes with p55 at the stereocilia tip during hair cell development. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 10973–8 (2006).
226. Ebrahim, S. *et al.* Alternative Splice Forms Influence Functions of Whirlin in Mechanosensory Hair Cell Stereocilia. *Cell Rep.* **15,** 935–943 (2016).
227. Belyantseva, I. A. *et al.* Myosin-XVa is required for tip localization of whirlin and differential elongation of hair-cell stereocilia. *Nat. Cell Biol.* **7,** (2005).

228. Mburu, P. *et al.* Gelsolin plays a role in the actin polymerization complex of hair cell stereocilia. *PLoS One* **5,** e11627 (2010).
229. Maerker, T. *et al.* A novel Usher protein network at the periciliary reloading point between molecular transport machineries in vertebrate photoreceptor cells. *Hum. Mol. Genet.* **17,** 71–86 (2008).
230. Holme, R. H., Kiernan, B. W., Brown, S. D. M. & Steel, K. P. Elongation of hair cell stereocilia is defective in the mouse mutant whirler. *J. Comp. Neurol.* **450,** 94–102 (2002).
231. Audo, I. *et al.* A novel DFNB31 mutation associated with Usher type 2 syndrome showing variable degrees of auditory loss in a consanguineous Portuguese family. *Mol. Vis.* **17,** 1598–606 (2011).
232. Ebermann, I. *et al.* A novel gene for Usher syndrome type 2: Mutations in the long isoform of whirlin are associated with retinitis pigmentosa and sensorineural hearing loss. *Hum. Genet.* **121,** 203–211 (2007).
233. 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).
234. Grati, M. *et al.* Localization of PDZD7 to the stereocilia ankle-link associates this scaffolding protein with the Usher syndrome protein network. *J. Neurosci.* **32,** 14288–93 (2012).
235. Adato, A. *et al.* Interactions in the network of Usher syndrome type 1 proteins. *Hum. Mol. Genet.* **14,** 347–56 (2005).
236. van Wijk, E. *et al.* The DFNB31 gene product whirlin connects to the Usher protein network in the cochlea and retina by direct association with USH2A and VLGR1. *Hum. Mol. Genet.* **15,** 751–65 (2006).
237. Manor, U. *et al.* Regulation of stereocilia length by myosin XVa and whirlin depends on the actin-regulatory protein Eps8. *Curr. Biol.* **21,** 167–72 (2011).
238. Wang, L., Zou, J., Shen, Z., Song, E. & Yang, J. Whirlin interacts with espin and modulates its actin-regulatory function: an insight into the mechanism of Usher syndrome type II. *Hum. Mol. Genet.* **21,** 692–710 (2012).
239. Gosens, I. *et al.* MPP1 links the Usher protein network and the Crumbs protein complex in the retina. *Hum. Mol. Genet.* **16,** 1993–2003 (2007).
240. 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. *J. Neurochem.* **85,** 123–134 (2003).
241. Kersten, F. F. J. *et al.* Association of whirlin with Cav1.3 (alpha1D) channels in photoreceptors, defining a novel member of the usher protein network. *Invest. Ophthalmol. Vis. Sci.* **51,** 2338–46 (2010).
242. Zou, J., Lee, A. & Yang, J. The expression of whirlin and Cav1.3α_1 is mutually independent in photoreceptors. *Vision Res.* **75,** 53–9 (2012).
243. Gregory, F. D., Pangrˇ, T., Calin-jageman, I. E., Moser, T. & Lee, A. Harmonin enhances voltage-dependent facilitation of Ca v 1 . 3 channels and synchronous exocytosis in mouse inner hair cells. *J. Physiol.* **13,** 3253–3269 (2013).
244. Ciardo, M. G. *et al.* Whirlin increases TRPV1 channel expression and cellular stability. *Biochim. Biophys. Acta - Mol. Cell Res.* **1863,** 115–127 (2016).

245. Basei, F. L. *et al.* New interaction partners for Nek4.1 and Nek4.2 isoforms: from the DNA damage response to RNA splicing. *Proteome Sci.* **13,** 11 (2015).
246. de Nooij, J. C. *et al.* The PDZ-Domain Protein Whirlin Facilitates Mechanosensory Signaling in Mammalian Proprioceptors. *J. Neurosci.* **35,** 3073–84 (2015).
247. Green, J. a, Yang, J., Grati, M. H., Kachar, B. & Bhat, M. a. Whirlin, a cytoskeletal scaffolding protein, stabilizes the paranodal region and axonal cytoskeleton in myelinated axons. *BMC Neurosci.* **14,** 96 (2013).
248. Miyagi, Y. *et al.* Delphilin: a novel PDZ and formin homology domain-containing protein that synaptically colocalizes and interacts with glutamate receptor delta 2 subunit. *J. Neurosci.* **22,** 803–814 (2002).
249. Matsuda, K., Matsuda, S., Gladding, C. M. & Yuzaki, M. Characterization of the delta2 glutamate receptor-binding protein delphilin: Splicing variants with differential palmitoylation and an additional PDZ domain. *J. Biol. Chem.* **281,** 25577–25587 (2006).
250. Takeuchi, T. *et al.* Enhancement of both long-term depression induction and optokinetic response adaptation in mice lacking delphilin. *PLoS One* **3,** e2297 (2008).
251. 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,** 4040–5 (2010).
252. Cosgrove, D. & Zallocchi, M. Usher protein functions in hair cells and photoreceptors. *Int. J. Biochem. Cell Biol.* **46,** 80–9 (2014).
253. Hudspeth, A. J. & Gillespie, P. G. Pulling springs to tune transduction: Adaptation by hair cells. *Neuron* **12,** 1–9 (1994).
254. Wang, J. & Deretic, D. Molecular complexes that direct rhodopsin transport to primary cilia. *Prog. Retin. Eye Res.* **38,** 1–19 (2014).
255. Hirota-Nakaoka, N. & Goto, Y. Alcohol-induced denaturation of ß-lactoglobulin: A close correlation to the alcohol-induced α-helix formation of melittin. *Bioorganic Med. Chem.* **7,** 67–73 (1999).
256. Maerker, T. *et al.* A novel Usher protein network at the periciliary reloading point between molecular transport machineries in vertebrate photoreceptor cells. *Hum. Mol. Genet.* **17,** 71–86 (2008).
257. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36,** 1277–1282 (2003).
258. Guinier, A. Diffraction of X-rays of very small angles – application to the study of ultramicroscopic phenomenon. Ann. Phys. 12, 161–237. (1939). *Ann. Phys.* 161–237 (1939).
259. Svergun, D. I. Determination of the Regularization Parameter in Indirect- Transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr.* **25,** 495–503 (1992).
260. Delaglio, F. *et al.* Nmrpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
261. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinforma.* **59,** 687–696 (2005).
262. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from

NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56,** 227–241 (2013).
263. Ottiger, M., Delaglio, F. & Bax, A. Measurement ofJand Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J. Magn. Reson.* **131,** 373–378 (1998).
264. Dosset, P., Hus, J. C., Blackledge, M. & Marion, D. Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. *J. Biomol. NMR* **16,** 23–28 (2000).
265. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. *J. Am. Chem. Soc.* **104,** 4546–4559 (1982).
266. Gronenborn, A. M. Protein acrobatics-Dimerization via domain swapping. *Curr Opin Struc Biol* **19,** 39–49 (2010).
267. Assad, J. A. & Corey, D. P. An active motor model for adaptation by vertebrate hair cells. *J. Neurosci.* **12,** 3291–3309 (1992).
268. Gillespie, P. G. & Müller, U. Mechanotransduction by hair cells: models, molecules, and mechanisms. *Cell* **139,** 33–44 (2009).
269. Barbato, G., Mitsuhiko, I., Kay, L. E., Pastor, R. W. & Bax, A. Backbone Dynamics of Calmodulin Studied by 15N Relaxation Using Inverse Detected Two-Dimensional NMR Spectroscopy: The Central Helix Is Flexible? *Biochemistry* **31,** 5269–5278 (1992).
270. Zhu, G., Xia, Y., Nicholson, L. K. & Sze, K. H. Protein dynamics measurements by TROSY-based NMR experiments. *J. Magn. Reson.* **143,** 423–426 (2000).
271. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223 (2009).
272. 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. *Hum. Mol. Genet.* **18,** 655–66 (2009).
273. Ebrahim, S. *et al.* Alternative Splice Forms Influence Functions of Whirlin in Mechanosensory Hair Cell Stereocilia. *Cell Rep.* **15,** 935–943 (2016).
274. Ebermann, I. *et al.* A novel gene for Usher syndrome type 2: Mutations in the long isoform of whirlin are associated with retinitis pigmentosa and sensorineural hearing loss. *Hum. Genet.* **121,** 203–211 (2007).
275. Feng, W. & Zhang, M. Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. *Nat. Rev. Neurosci.* **10,** 87–99 (2009).
276. Hung, A. Y. & Sheng, M. PDZ domains: structural modules for protein complex assembly. *J. Biol. Chem.* **277,** 5699–702 (2002).
277. 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. *J. Biol. Chem.* **289,** 25327–40 (2014).
278. Delhommel, F., Wolff, N. & Cordier, F. 1H, 13C and 15N backbone resonance assignments and dynamic properties of the PDZ tandem of Whirlin. *Biomol. NMR Assign.* (2016). doi:10.1007/s12104-016-9701-z
279. Walsh, J. D., Meier, K., Ishima, R. & Gronenborn, A. M. NMR studies on domain diffusion and alignment in modular GB1 repeats. *Biophys. J.* **99,** 2636–2646 (2010).

280. Bernadó, P. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur. Biophys. J.* **39,** 769–780 (2010).
281. Durand, D. *et al.* NADPH oxidase activator p67phox behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.* **169,** 45–53 (2010).
282. Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39,** 5355-65 (2000).
283. Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125,** 1731–1737 (2003).
284. Clore, G. M. Practical Aspects of Paramagnetic Relaxation Enhancement in Biological Macromolecules. *Methods Enzymol.* **564,** 485–497 (2015).
285. Yang, S., Blachowicz, L., Makowski, L. & Roux, B. Multidomain assembled states of Hck tyrosine kinase in solution. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15757–62 (2010).
286. Fushman, D., Varadan, R., Assfalg, M. & Walker, O. Determining domain orientation in macromolecules by using spin-relaxation and residual dipolar coupling measurements. *Prog. Nucl. Magn. Reson. Spectrosc.* **44,** 189–214 (2004).
287. Huang, J. R. *et al.* Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: A combined NMR and SAXS study. *J. Am. Chem. Soc.* **136,** 7068–7076 (2014).
288. Bai, X. chen, McMullan, G. & Scheres, S. H. W. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40,** 49–57 (2015).
289. Göbl, C., Madl, T., Simon, B. & Sattler, M. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **80,** 26–63 (2014).
290. Hennig, J. & Sattler, M. The dynamic duo: Combining NMR and small angle scattering in structural biology. *Protein Sci.* **23,** 669–682 (2014).
291. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449,** 1078–82 (2007).
292. Pawson, T. Dynamic control of signaling by modular adaptor proteins. *Curr. Opin. Cell Biol.* **19,** 112–116 (2007).
293. Papal, S. *et al.* The giant spectrin βV couples the molecular motors to phototransduction and Usher syndrome type I proteins along their trafficking route. *Hum. Mol. Genet.* **22,** 3773–88 (2013).
294. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450,** 983–90 (2007).
295. Hura, G. L. *et al.* Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods* **10,** 453–4 (2013).
296. Rückert, M. & Otting, G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122,** 7793–7797 (2000).
297. Clore, G. M., Gronenborn, A. M. & Tjandra, N. Direct Structure Refinement against Residual Dipolar Couplings in the Presence of Rhombicity of Unknown Magnitude. *J.*

Magn. Reson. **131,** 159–162 (1998).
298. Zweckstetter, M. NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protoc.* **3,** 679–690 (2008).
299. Warren, J. J. & Moore, P. B. A maximum likelihood method for determining D(a)(PQ) and R for sets of dipolar coupling data. *J. Magn. Reson.* **149,** 271–275 (2001).
300. Linge, J. P., Williams, M. A., Spronk, C. A. E. M., Bonvin, A. M. J. J. & Nilges, M. Refinement of protein structures in explicit solvent. *Proteins Struct. Funct. Genet.* **50,** 496–506 (2003).
301. Simon, B., Madl, T., Mackereth, C. D., Nilges, M. & Sattler, M. An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. *Angew. Chemie - Int. Ed.* **49,** 1967–1970 (2010).
302. Nilges, M. *et al.* Accurate NMR Structures Through Minimization of an Extended Hybrid Energy. *Structure* **16,** 1305–1312 (2008).
303. Jacso, T. *et al.* The mechanism of denaturation and the unfolded state of the α-helical membrane-associated protein Mistic. *J. Am. Chem. Soc.* **135,** 18884–18891 (2013).
304. Eastman, P. *et al.* OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J Chem Theory Comput.* **9,** 461–469 (2013).
305. Ihms, E. C. & Foster, M. P. MESMER: Minimal ensemble solutions to multiple experimental restraints. *Bioinformatics* **31,** 1951–1958 (2015).
306. Hourdel, V. *et al.* MEMHDX: An interactive tool to expedite the statistical validation and visualization of large HDX-MS datasets. *Bioinformatics* **32,** 3413–3419 (2016).
307. Vincentelli, R. *et al.* Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat. Methods* **12,** 787–93 (2015).
308. 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).
309. Mathur, P. D. *et al.* A study of whirlin isoforms in the mouse vestibular system suggests potential vestibular dysfunction in *DFNB31* -deficient patients. *Hum. Mol. Genet.* ddv403 (2015). doi:10.1093/hmg/ddv403
310. Reddy Chichili, V. P., Kumar, V. & Sivaraman, J. Linkers in the structural biology of protein-protein interactions. *Protein Sci.* **22,** 153–167 (2013).
311. Li, J. *et al.* Structure of Myo7b/USH1C complex suggests a general PDZ domain binding mode by MyTH4-FERM myosins. *Proc. Natl. Acad. Sci.* 201702251 (2017). doi:10.1073/pnas.1702251114
312. Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C. & Teichmann, S. A. Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* **14,** 208–216 (2004).
313. Moore, A. D., Björklund, A. K., Ekman, D., Bornberg-bauer, E. & Elofsson, A. Arrangements in the modular evolution of proteins. *Trends Biochem. Sci.* **33,** 444–451 (2008).
314. Scott, J. D. & Pawson, T. Cell signaling in space and time: where proteins come together and when they're apart. *Science (80-.).* **326,** 1220–4 (2009).
315. Chothia, C., Gough, J., Vogel, C. & Teichmann, S. A. Evolution of the Protein

Repertoire. *Science (80-.).* **300,** 1701–1703 (2003).
316. Stiffler, M. A. *et al.* PDZ Domain Binding Selectivity Is Optimized Across the Mouse Proteome. *Science (80-.).* **317,** 364–369 (2007).
317. Bashton, M. & Chothia, C. The geometry of domain combination in proteins. *J. Mol. Biol.* **315,** 927–39 (2002).
318. Gordley, R. M., Bugaj, L. J. & Lim, W. A. in *Current Opinion in Structural Biology* **39,** 106–114 (Elsevier Ltd, 2016).
319. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450,** 983–90 (2007).
320. Popovych, N., Sun, S., Ebright, R. H. & Kalodimos, C. G. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.* **13,** 831–8 (2006).
321. Cooper, A. & Dryden, D. T. F. Allostery without conformational change: A plausible model. *Eur. Biophys. J.* **11,** 103–109 (1984).
322. Delaforge, E. *et al.* Large-Scale Conformational Dynamics Control H5N1 Influenza Polymerase PB2 Binding to Importin α. *J. Am. Chem. Soc.* **137,** 15122–15134 (2015).
323. Kassem, M. M., Wang, Y., Boomsma, W. & Lindorff-Larsen, K. Structure of the Bacterial Cytoskeleton Protein Bactofilin by NMR Chemical Shifts and Sequence Variation. *Biophys. J.* **110,** 2342–2348 (2016).
324. Prischi, F. & Pastore, A. in *Advanced Technologies for Protein Complex Production and Characterization* 351–367 (2016). doi:10.1007/978-3-319-27216-0_15
325. Delaforge, E. *et al.* Investigating the Role of Large-Scale Domain Dynamics in Protein-Protein Interactions. *Front. Mol. Biosci.* **3,** 1–8 (2016).
326. Aznauryan, M. *et al.* Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc. Natl. Acad. Sci. U. S. A.* 201607193 (2016). doi:10.1073/pnas.1607193113
327. Mackereth, C. D. *et al.* Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475,** 408–411 (2011).