UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)

THÈSE présentée par :

Pau JANÉ PALLÍ

soutenue le : **7 Octobre 2020**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Bio-informatique

Quantification de affinités PBM/PDZ et de leurs sites modulateurs par des approches expérimentales et informatiques à haut débit

THÈSE dirigée par :

M. TRAVÉ Gilles	Dr., Université de Strasbourg
M. NOMINÉ Yves	Dr., Université de Strasbourg

RAPPORTEURS :

Mme. IVARSSON Ylva	Prof., Uppsala University
Mme. LUCK Katja	Dr., IMB Mainz

AUTRES MEMBRES DU JURY :

Mme. THOMPSON Julie	Dr., Université de Strasbourg
M. GIBSON Toby	Dr., EMBL Heidelberg

"If you can't explain something in simple terms, you don't understand it."

-Richard Feynman

Acknowledgments

First of all, I would like to thank Gilles Travé for all the dedication on this project as a director of the thesis and all the time he had to share his knowledge with me. His passion for science converted any discussion with him to a master class. I also thank my co-director Yves Nominé.

I also want to thank the rest of the team, specially to Gergő Gógl with whom I performed grand part of my project. I learned a lot with his brightness mind and his dedication in science.

I would like to thank the PDZnet for the great training that offered to me. I also would like to thank all the amazing people I met during the program. I would like to particularly thank the other fellows of the PDZnet which did the meeting in the different cities of Europe more special.

I thank Carlos Fontes, Toby Gibson and Zsuzsanna Dosztányi who hosted me during my internships. I would also thank to all the people in their labs. I would like also to thank Nicolas Wolff and Renaud Vincentelli for the great collaboration in this project and the excellent scientific discussions we had.

I thank all the jury members who will read and evaluate this thesis. I hope to see them in the defense day although the difficulties of the situation.

I would like to thank to the Lunch Crew for all the special moments during lunch. Specially, I would like to thank to the "Lunch Crew the Return", which constituted the main day to day keep going's support. I would like to thank you for all the coffee hours in which we spent together talking about nonsense. These were the people who listened mostly to my daily concerns, such as... what's the difference between a pie and a cake?

Lastly, I would like to thank Dimitra for encouraging me in my moments of weakness. I can only wish to have more silly moments together and accomplish our dream trips, which have been cancelled during the pandemia time. Thanks for sharing your world with me.

Me gustaría agradecer a todos los embajadores de Estrasburgo por el tiempo que hemos pasado juntos. Estos cuatro años han estado plagado de aventuras, desde viajes al lago de Constanza hasta Bangkok; desde cenas "silenciosas" en que los vecinos golpeaban con escobas el suelo del piso hasta barbacoas en Baggersee. Nos hemos cuidado entre nostros en tal magintud que no somos conscientes, tanto haciendo terapia de cerveza como con deporte; tal como fútbol, pádel o running. Han sido unos años muy especiales que espero que perduren en nuestros recuerdos.

A la meva família qui sempre m'ha donat suport. Moltes gràcies per tot el que heu fet per mi, tant en l´esforç per la meva educació com en el per formar-me com a persona.

Una especial menció a la gent de Biologia Humana amb qui he viscut grans aventures durant aquest anys; des del Vinya Rock fins al singular concert d'Amics de les Arts a Luxemburg. Una especial menció també a Jan Huertas, qui ha dut la difícil tasca de visitar-me a cadascuna de les ciutats on he fet una estada durant el meu doctorat: Estrasburg, Lisboa, Heidelberg i Budapest.

Agrair a la Guàrdia, una gent entranyable amb qui passo uns estius inolvidables amb els que carrego les piles pel proper curs i que sempre ofereixen suport moral en tot moment.

A l'Armand i a la Laura, amb qui he passat grans moment, com la sortida de sol des del Sàhara. No us preocupeu, que encara que marxi d'Estrasburg podrem seguir menjant 'Tartes Flambées' fins a reventar.

En acabat, m'agradaria agrair a tota la gent del Balmes amb els qui he crescut i he passat grans moments. Espero que aquest temps de pandemia acabi ben aviat per seguir buscant aventures amb vosaltres.

Contents

List of Figures

List of Tables

List of Boxes

Part I

Prologue

Chapter 1

Resumé

1.1 Contexte

Au cours des dernières décennies, la fonction, la localisation et la séquence d'un grand nombre de protéines ont été systématiquement analysées et stockées dans des bases de données, comme UniProt [1]. Pourtant, les fonctions biologiques ne sont pas le produit de protéines isolées, mais émergent plutôt d'interactions complexes -et des réactions qui s'en suivent- impliquant de nombreuses protéines et d'autres espèces moléculaires. C'est pourquoi de nombreux efforts ont été déployés pour cartographier, dans divers organismes, dont l'humain, leur "interactome", c'est-à-dire l'ensemble du réseau de leurs interactions protéine-protéine (PPI) [2,3].

La plupart des approches interactomiques produisent des données binaires ("lie" ou "ne lie pas") [4,5]. Cependant, cette description binaire des interactomes reste inexacte et incomplète, car les interactions protéine-protéine peuvent présenter une très large palette d'affinités. Les constantes d'affinité à l'équilibre (K_d) peuvent varier de l'échelle picomolaire (10-12 M) à l'échelle milimolaire (10-3 M) [6–8]. En outre, au sein d'un interactome donné, la spécificité - une question clé en biologie moléculaire - émerge de la comparaison des affinités individuelles. Alors que l'affinité représente la force de liaison d'un complexe entre un ligand L et une protéine P, la spécificité de ce complexe ne peut être évaluée qu'en considérant le profil d'affinité global du ligand L ou une famille de protéines susceptibles de se lier au ligand L, dont la protéine P. Obtenir des informations sur la spécificité requiert de collecter puis de comparer un grand nombre de données d'affinité. Par

conséquent, traiter les questions de spécificité, et donc d'affinité, pour les réseaux PPI représente un défi méthodologique essentiel.

Dans cette thèse, un sous-ensemble de l'interactome PPI est étudié par l'approche "domainomique", qui se concentre sur des interactions entre fragments de protéines : domaines globulaires vs motifs linéaires courts (SLiM) [9]. Les SLiM ne dépassent généralement pas 10 résidus et sont situés dans des régions intrinsèquement dépliées des protéines. Ils peuvent évoluer rapidement en apparaissant *de novo* ou, à l'inverse, en disparaissant. Cela complique leur identification bioinformatique [10–12]. Malgré leur courte séquence, les SLiM, par leur reconnaissance spécifique de domaines globulaires, jouent un rôle essentiel dans réseaux PPI [10–13]. En règle générale, seuls 3 à 4 résidus d'un SLiM contribuent au motif consensus crucial pour la reconnaissance des domaines, les autres résidus jouant plutôt un rôle modulateur [14]. On estime que le protéome humain contient plus de 100 000 SLiM [13].

Ce travail a porté sur les domaines PDZ, une famille de domaines globulaires reconnaissant des motifs de liaison aux PDZ (appelés PBM, pour "PDZ-Binding Motifs") généralement situés à l'extrémité C-terminale de leurs protéines partenaires. Le protéome humain contient 266 domaines PDZ (le "PDZome") [15]. Les domaines PDZ comprennent 80 à 100 acides aminés. Leur repliement canonique, généralement décrit comme un tonneau β antiparallèle ou un β-sandwich, comprend trois feuillets β (β1 à β3), une courte hélice α (α1), deux feuillets β (β4 et β5), une longue hélice α (α2) et un dernier feuillet β (β6) [16, 17]. Ces éléments de structure secondaire peuvent varier en longueur [18].

Les PBM C-terminaux ont une nomenclature spécifique dans la littérature. L'ultime position C-terminale est appelée p0 (position 0), le résidu précédent p-1 et ainsi de suite. La position p0 est toujours hydrophobe, principalement Valine, Leucine ou Isoleucine [18–20]. Les PBM peuvent être regroupés en trois classes selon la nature de l'acide aminé présent en p-2 : les PBM de classe 1 contiennent une Sérine/Thréonine, les PBM de classe 2 une Valine/Tyrosine/Phénylalanine et les PBM de classe 3 un acide Aspartique/Glutamique [21]. La poche des domaines PDZ liant les peptides-cible est formée par les résidus de la boucle β1-β2 (séquence conservée [KR]...GϕGϕ), des brins β2 et β3, de la boucle β2-β3 et de l'hélice α2. ϕ est généralement une leucine ou une phénylalanine. "..." signifie tout acide aminé. Plusieurs résidus trouvés à l'interface entre le brin β2 et l'hélice α2 du domaine PDZ forment une poche hydrophobe accomodant le PBM, qui adopte une structure

secondaire β lors de l'interaction. À L'opposé de la poche de liaison, les deux extrémités N et C du domaine PDZ se rejoignent [15, 18–20]. L'interaction PDZ-PBM comprend plusieurs liaisons hydrogène établies entre les chaines principales du domaine PDZ et du peptide, typiques des feuillets β. Le groupement carboxyle de la position C-terminale p0 établit un contact avec la boucle [KR]...GϕGϕ β1-β2 du domaine PDZ. Le p-2 joue également un rôle crucial en se liant à plusieurs résidus de l'hélice α2- PDZ. Le résidu central de l'hélice α2 varie en fonction de la classe de PBM : il peut s'agir d'un résidu histidine, aliphatique ou tyrosine suivant qu'il est lié à un motif de classe 1, 2 ou 3, respectivement [18–20].

1.2 Contribution du travail de thèse

Les réseaux domaines-motifs sont souvent modulés par des modifications post-traductionnelles réversibles (PTM, pour "Post-Translational Modification"). La PTM la plus abondante est la phosphorylation, une réaction chimique réversible (catalysée par les kinases et en sens inverse par les phosphatases), qui transfère le groupe γ-phosphate d'une molécule d'ATP à un résidu récepteur, le plus souvent au groupe hydroxyle d'un résidu Ser/Thr, ou Tyr via la formation d'une liaison phosphoester [22]. Les acides aminés phosphorylés ont des propriétés uniques qui peuvent altérer de différentes manières les propriétés biochimiques des protéines modifiées. L'acétylation de lysines, une autre PTM fréquemment observée, altère une grande variété de substrats, tels que les histones, les acétylases, les facteurs de transcription liant l'ADN, les facteurs d'importation nucléaire et l'α-tubuline [23]. L'enzyme "lysine acétyl-coenzyme A acétyltransférase" transfère un groupe acétyle de l'enzyme au groupe ϵ d'un résidu lysine. Le processus est réversible par une autre enzyme, la lysine désacétylase. L'acétylation de la lysine neutralise la charge positive, tout en rendant la protéine "lisible" par les domaines reconnaissant l'acétyl-lysine. L'acétylation a un impact élevé sur la régulation cellulaire [24].

Les processus de phosphorylation et d'acétylation sont tous deux difficiles à reproduire in vitro. Si certaines stratégies visent à imiter ces processus [6, 25–29], d'autres cherchent plutôt à évaluer l'impact de ces modifications sur l'interactome [30]. Ici, nous avons utilisé des PBM synthétisés chimiquement pour reproduire différentes conditions: motifs non modifiés de diverses longueurs, acétylés, phosphorylés ou sujets à des mutations cherchant à "imiter" les phosphorylations. Ces

peptides ont été utilisés pour des études d'interaction à l'aide du test "Hold-Up", un test développé à l'origine dans notre laboratoire. Le Hold-Up est une approche de rétention chromatographique comparative qui quantifie l'affinité de liaison à l'équilibre des protéines envers les ligands (ici, les PBM biotinylés) attachés aux résines d'affinité. Les protéines (ici, des domaines PDZ) sont surexprimées dans E.coli, peuvent être purifiées ou non, et sont ensuite stockées sous forme d'extraits en attendant d'effectuer les tests. Après le test de liaison, les échantillons sont lus par un instrument d'électrophorèse capillaire microfluidique appelé "caliper". Nous avons ainsi observé comment diverses PTM, véritables ou imitées, ou l'ajout de résidus supplémentaires à un PBM donné, peuvent changer ses préférences de liaison au domaine PDZ. De plus, nous avons analysé le "potentiel d'impact de la liaison aux PDZ" de certaines positions de résidus des PBM lorsqu'ils sont mutés ou soumis à une PTM.

La réalisation du test holdup exige une approche rigoureuse. Plusieurs étapes critiques ont été identifiées: l'expression du PDZome, la vérification de sa qualité et l'étalonnage de ses concentrations. Depuis la publication du premier test holdup [7], d'autres améliorations ont été mises en œuvre pour atteindre une qualité élevée pour les étapes mentionnées [31]. Certaines d'entre elles sont développées dans les paragraphes suivants.

Un premier aspect concerne l'optimisation du rapport coût-efficacité des mesures. Nous avons effectué des mesures de holdup en "singlicate", dans le but de mesurer plus de PBM différents pour un même coût. Pour nous assurer de la qualité de nos données, nous avons vérifié l'interaction PBM/PDZ point par point, depuis ses niveaux d'expression jusqu'à l'étape de superposition de l'électrophérogramme. Un deuxième point concerne la qualité des extraits d'expression du PDZome. Une préparation optimale ne suffit pas toujours à garantir la qualité des mesures, car les protéines peuvent se dégrader lors du stockage des extraits. Nous avons mis en place une étape de traitement informatique pour mieux superposer les électrophérogrammes du caliper, ce qui a permis d'améliorer la fiabilité du test de rétention. Ces étapes de contrôle de la qualité nous permettent d'améliorer l'exactitude et la précision des données de retard, réduisant ainsi le nombre de faux positifs et de faux négatifs. Le protocole présenté dans la thèse a été appliqué à plus de 20 000 paires d'interactions, ce qui a permis de créer une base de données des interactions PBM/PDZ. Enfin, nous avons validé des sous-ensembles d'interactions positives résultant de l'essai de rétention par polarisation

de fluorescence (FP, pour Fluorescence Polarisation), une approche orthogonale de mesure quantitative d'affinité de liaison qui utilise des peptides marqués par fluorescence. Les données d'affinité de liaison obtenues par FP ont ensuite été utilisées pour extrapoler toutes les constantes d'affinité mesurées par holdup.

Il convient de noter que le test de holdup fournit des informations quantitatives à la fois sur l'"interactome" et sur le "négatome" des interactions PBM-PDZome [32]. En effet, au delà d'informations binaires nous avons quantifié toutes les affinités de chaque PBM mesuré pour tous les domaines constituant l'ensemble du PDZome. Ceci nous fournit une information précieuse pour l'analyse du réseau. L'interactome montre les préférences potentielles du PBM pour un sous-ensemble donné des domaines PDZ en supposant des conditions égales dans un système donné (sachant que, dans un système cellulaire, les protéines contenant chaque PBM et PDZ concernés varient dans leur expression et localisation). Malgré sa grande sensibilité, le test holdup n'a détecté que jusqu'à 50 domaines PDZ liant un PBM donné [7,31,33,34]. Un tel interactome, se bornant uniquement aux interactions détectables, révèle donc moins de 20% des informations du réseau PDZome. Or, la spécificité est déterminée non seulement par les interactions réelles, mais aussi par toutes les interactions n'ayant pas lieu. L'étude du "négatome" est donc hautement pertinente, d'autant plus qu'elle permet aussi, grâce à nos approches expérimentales, d'explorer la "zone grise" des interactions pouvant avoir lieu dans certaines circonstances (protéine à nombre de copies élevé, emplacement de la protéine, etc...). En ajoutant le négatome à nos études, nous pouvons donc couvrir près de 100% du réseau PBM/PDZome. Cela nous permet d'étudier davantage les variations de PBM, telles que la PTM et les mutations, dans le PDZome. Par conséquent, un point d'intérêt majeur du test holdup est sa capacité à fournir, grâce à la mesure systématique des affinités de liaison, des informations quantitatives sur les spécificités de liaison. Dans cette thèse, nous proposons également une approche simple mais robuste pour quantifier la spécificité de tout PBM mesurée par rapport à l'ensemble du PDZome, en utilisant les "profils de liaison au PDZome" obtenus à partir de l'essai de rétention. Nous avons utilisé cette approche pour comparer les spécificités des PBM soit sous forme sauvage, soit soumis à des mutations ou modifications par des PTM.

1.3 Perspectives

De nombreux chercheurs - en particulier dans le réseau de formation international "PDZnet" Marie Skłodowska-Curie dont je faisais partie - ont saisi le potentiel du test holdup et l'ont utilisé pour leurs recherches. Le nombre d'utilisateurs de ce test va certainement encore augmenter à l'avenir. La méthode a d'abord nécessité d'exprimer la banque complète de domaines PDZ. Ce travail a été réalisé par notre collaborateur Renaud Vincentelli [7], un autre membre du réseau. Les tests d'interactions ont été réalisés par trois laboratoires, dont le nôtre. Enfin, comme j'étais très impliqué dans le traitement et la conservation des données, j'ai eu accès à la plupart des données de holdup mesurées dans le réseau, et j'ai pu les traiter en utilisant mes approches. Afin de réduire les coûts, assez élevés, de l'approche holdup, notre laboratoire améliore actuellement la méthode en utilisant la fluorescence intrinsèque des protéines mesurées, pour quantifier les données d'affinité. Cela nous permettra à la fois d'éviter l'étape de Caliper et de réduire la complexité des processus bioinformatiques. Toutefois, les nouveaux développements de la méthode nécessiteront de pouvoir se référer à un ensemble de données solides et robustes, tel que celui que nous possédons actuellement (plus de 20 000 interactions). La nouvelle méthode sera plus rapide et moins coûteuse, ce qui permettra d'avancer encore plus vite dans le décryptage de l'interactome PBM/PDZome.

1.4 Conclusions

Cette thèse, a permis d'évaluer l'impact de diverses modifications des PBM (PTM, mutations, variations de longueurs) sur leurs interactomes avec l'ensemble du PDZome humain. Ces modifications conduisent à un changement global des profils de liaison -et donc de la capacité de liaison aux PDZ- fournissant des informations quantitatives sur l'effet biologique que de telles modifications peuvent avoir dans le contexte des protéines entières. Comme toute altération de la séquence de la PBM (comme les mutations, les modifications ou le changement de la longueur du peptide) peut entraîner des réarrangements globaux du profil de liaison au PDZome, ces altérations doivent être utilisées et interprétées avec une grande prudence en tenant compte des études supplémentaires sur la structure, les cellules ou même l'organisme entier. De tels résultats n'auraient pas pu être obtenus

sans le test holdup, et la stratégie d'analyse que nous avons développée par la suite, permettant d'obtenir des données d'affinité robustes. Nous avons également proposé un moyen de quantifier la spécificité. Dans l'ensemble, nous espérons que l'approche proposée pour la quantification de l'affinité et de la spécificité des réseaux d'interactions domaines-motifs, aura contribuée à nous rapprocher un peu plus d'une approche quantitative de l'interactome humain, basée sur les affinités d'interaction.

Chapter 2

Summary

2.1 Background

In the past decades, the function, localization and sequence of huge numbers of single proteins have been systematically analyzed and stored on databases, such as UniProt [1]. However, biological functions do not emerge from single proteins, but rather from complex interactions -and subsequent reactions- implicating many proteins and other molecular species. Therefore, many efforts have been invested to map, in various organisms including human, their "interactome", i.e. the entire network of their protein-protein interactions (PPI) [2, 3]. Most interactomic approaches deliver binary results (either "bind" or "not bind") [4, 5]. However, this binary information is an inaccurate and incomplete way to describe interactomes, since protein-protein interactions can display a very large palette of affinities. Equilibrium affinity constants (K_d) can range from the picomolar scale (10^{-12} M) to the milimolar scale (10^{-3} M) [6–8]. Furthermore, within a given interactome, specificity -a key issue in molecular biology- emerges from the comparison of individual affinities. While affinity represents the binding strength of a complex between a ligand L and a protein P, the specificity of L for P (or, conversely, of P for L) can only be assessed by considering the overall affinity profile of ligand L for a subset or family of proteins susceptible to bind ligand L (or the overall affinity of a family of ligands susceptible to bind to protein P). Gaining information about specificity implies to collect then compare a large number of affinity data. Therefore, addressing the specificity, and thus the affinity, issues for PPI networks of biological interest represents a key methodological challenge.

The human proteome comprises ∼20 000 proteins (not accounting for splicing isoforms), themselves estimated to contain 35 000 instances of folded domains [13, 35, 36] and up to a few milions instances of interaction motifs, which may be "pure" peptidic sequences or modified sequences bearing one or several of the ∼300 possible PTMs [13].

In this thesis, a subset of the PPI interactome is studied through the "domain-omics" approach, which focuses on minimal interacting blocks accounting for a large subset of PPI [9]: globular domains vs Short Linear Motifs (SLiM). SLiMs are usually not longer than 10 residues and located in intrinsically unfolded regions of proteins. They can evolve rapidly by appearing *de novo* or, conversely, disappearing. This complicates their search through the proteome [10–12]. Although being short in sequence, SLiMs have a high functionality as they are recognized by globular domains, which confers them a huge relevance in PPI [10–13]. Typically, only 3 to 4 residues of a SLiM contribute to the core consensus motif for domain recognition whereas the other residues may play a modulatory role [14]. The human proteome is estimated to contain over 100 000 SLiMs, which can be raised up to 1 000 000 peptide sequences instances due PTMs [13].

This work focuses on PDZ domains, a family of globular domains that bind to conserved PDZ-Binding Motifs (called henceforth PBMs) generally situated at the extreme C-terminus of their partner proteins. The human proteome contains 266 PDZ domains, mentioned henceforth as the PDZome [15]. PDZ domains comprise 80-100 amino-acids. Their canonical folding topology is as follows: three β-strands (β1 to β3), a short α-helix (α1), two β-strands (β4 and β5), a long α-helix (α2) and a last β-strand (β6), mostly described as an anti-parallel β-barrel or a β-sandwich [16, 17]. Although the fold is conserved, the secondary structure elements may vary in length [18].

PBMs have a specific nomenclature in the literature. The last position of a C-terminal PBM is called position 0 (p0), the neighbor residue is then numbered as position -1 (p-1) and so on backwards. All the PBMs have an hydrophobic residue in its p0, being mostly Valine, Leucine and Isoleucine [18–20]. PBMs can be grouped in three classes depending on the amino acid found in the p-2: class 1 motif contains a Serine/Threonine, class 2 motifs contains a Valine/Tyrosine/Phenylalanine and class 3 motifs contains an Aspartic/Glutamic acid [21].

The peptide-binding pocket of PDZ domains is formed by residues from the β1-

β2 loop known by its conserved [KR]. . . GϕGϕ sequence, the β2 and β3 strand, the β2-β3 loop and the α2-helix. ϕ is mainly Leucine and Phenylalanine in this order respectively, although other hydrophobic residues might also be found. "." means any amino acid. Among those, several residues found at the interface between the β2-strand and the α2-helix of the PDZ domain form a hydrophobic pocket that is recognized by the PBM, which adopts a β-strand upon binding. On the opposite site of the binding pocket, both of the N- and C-terminus of canonical PDZ domain are proximal to each other [15, 18–20]. The binding of these PBMs to their cognate PDZ domain is led by several hydrogen bonds established between backbone atoms of the peptide and backbone atoms of the PDZ domain observed typically between β-sheets. The p0 carboxylate group establishes contact with the conserved [KR]. . . GϕGϕ β1-β2 loop of the PDZ domain. The p-2 also plays a crucial role by binding to several residues of the PDZ α2-helix. The central residue of α2-helix varies depending on the PBM class: it is preferentially be a Histidine, an aliphatic or a Tyrosine residue when bound to a class 1, 2 or 3 motif, respectively [18–20].

2.2 Contribution of the thesis work

Domain-motif networks are often modulated by reversible post-translational modifications (PTMs). The most studied PTM is phosphorylation, a reversible chemical reaction (catalyzed by protein kinases and reverse-catalyzed by protein-phosphatases), that transfers the γ-phosphate group of an ATP molecule to a receiver residue, most often to the hydroxyl group of a Ser/Thr, or Tyr residues via forming a phosphoester bond [22]. Phosphorylated amino acids have unique properties that can alter biochemical properties of substrate proteins in different ways. Acetylation, an other common type of PTMs, occurs in a large variety of substrates, such as histones, acetylases, DNA-binding transcription factors, nuclear import factors and α-tubulin citekouzarides2000acetylation. In the case of Lysine acetylation, the process is promoted by an enzyme as lysine acetyl-coenzyme A acetyltransferase, which transfers an acetyl group from the enzyme to the ϵ-group of the Lysine residue in the protein. Lysine acetylation will neutralize the positive charge, but also will make the protein "readable" by specialized acetyl-lysine binding domains. The acetylation process is reversible by an other enzyme, as lysine deacetylase. Acetylation has a high impact in cell regulation [24].

Both phosphorylation and acetylation processes are difficult to reproduce in vitro. While some strategies aim at mimicking these processes [6, 25–29], others rather focus on the impact of PTMs on interactomes [30]. Here, we used chemically synthetized PBMs, hence we could better reproduce different conditions, such as a wild-type, acetylation or phosphorylation, or addition of extra exosites, and then to test residue mimication of the literature. These peptides were used for interaction studies using the Hold-Up assay, an assay originally developped in our laboratory. The holdup assay is a comparative chromatographic retention approach that quantifies the equilibrium binding affinity of proteins towards ligands (here, biotinylated PBMs) attached to affinity resins. The proteins (here, PDZ domains) are overexpressed in E.coli, may or may not be purified, and are then stored in the form of frozen extracts, waiting for the assay to be performed. After the binding assay, the samples are read through a microfluidic capillary electrophoresis instrument called "caliper". We observed how PTM as well as extra residues addition to a given PBM can switch its PDZ domain binding preferences. Moreover, we analyzed the "PDZ-binding impacting potential" of particular residue positions of PBMs when they are mutated or subject to a PTM.

The experimental holdup assay requires a rigorous approach for its performance. Several critical steps are identified in the following stages: the expression of the PDZome, the verification of its quality and the calibration of its concentrations. Since the first holdup assay was published [7], further improvements have been implemented to reach high quality for the mentioned stages [31]. Some of them are developped in the following paragraphs.

A first aspect concerns the optimization of the cost- and time-effectiveness of measurements. Here, we performed holdup measurements in singlicate, allowing us to increase the number of different PDZ-PBM pairs that we could address. To make sure about the accuracy of our data, we checked the PBM/PDZ interaction point by point from its expression levels to the electropherogram superimposition step. A second point concerns the quality of PDZome extracts. An optimal preparation does not always warrant the quality of the measurements, as protein degradation may occur during storage. We implemented a computational processing step to better superimpose the caliper's electropherograms leading to a better fiability of the holdup assay. These quality control steps allow us to improve accuracy and precision of the holdup data, reducing the amount of false positive and negative results. The protocol showed in the thesis has been applied to over 20 000

pairs of interactions resulting in a curated PBM/PDZ interaction database. Finally, we validated subsets of the positive interactions resulting from the holdup assay by Fluorescent Polarization (FP), an orthogonal powerfull approach for quantitative binding affinity measurements that uses fluorescently labelled peptides. The binding affinity data obtained by FP were then used to interpolate all the binding affinity constants (K_d) measured by the holdup assays.

Worthy of note, the holdup assay delivers powerful information about both, the PDZome "interactome" and the "negatome" [32]. Instead of providing "binary information" ("binds" or "does not bind"), we quantified PBM interactions against the whole PDZome. This provides us with an extra information to dig further in the PDZome network. The interactome shows the potential preferences of the PBM for a given subset of the PDZ domains assuming equal conditions in a giving system (whereas, in a cellular system, the expression and localization of the proteins comprising those PBMs and PDZs may greatly vary). Despite its high sensitivity, the holdup assay usually does not detect more than 50 PDZ domains binders for a given PBM [7, 31, 33, 34]. This means that a "binders-only" interactome delivers less than 20% of the PDZome network information. Yet, the specificity is determined, not only by the actual interactions, but also by all the interactions that do not take place. Therefore, the study of the "negatome" is highly relevant, specially when considering that it also allows us to explore the "grey zone" of interactions that only happen under certain circumstances (high copy number protein, protein localization, and so forth). By adding the negatome to our studies, we actually cover up to 100% of the PBM/PDZome network. This allows us to further study variations of PBM, such as PTM and mutations, in the PDZome. Therefore, a strong point of interest of the holdup assay is its ability to provide, thanks to the systematic measurement of binding affinities, quantitative information on binding specificities. In this thesis, we propose a simple yet robust approach to quantify the specificity of any measured PBM against the whole PDZome, using the "PDZome-binding profiles" obtained from the holdup assay. We used this approach to compare the specificities of PBMs either in wild-type form or subjected to mutations or modification by PTMs.

2.3 Future perspectives

Many researchers -particularly in the "PDZnet" Marie Skłodowska-Curie international training network I was part from- have realized the power of the holdup assay and used it for their research, and the demand for this assay will certainly increase in the future. The method first required to express all the PDZ library. This work was done by Vincentelli *et al.* [7], another member of the network. The holdup assay was performed by three laboratories including ours. Finally, since I was deeply involved in the data curation, I had access to -and could treat using my approaches- most of the holdup data measured in the team network. All the data curation performed in this thesis will allow the holdup to face one of its main problems: the costs. Our lab is currently improving the method by using fluorescence to quantify the affinity data. This will allow us to avoid the caliper step and to reduce the complexity of the bioinformatics processes. The use of FP will be still necessary to validate the data and calculate PBM concentrations. However, this cannot be done without a strong and robust data set like the one we possess now (over 20 000 interactions). The new method will be faster and cheaper, opening the way to the affinity-based quantification of the whole PBMome/PDZome interactome.

2.4 Conclusions

In this thesis the impact of PBM modifications (mimicking mutations in the core, PTM or addition of exosites) on their interactions with the PDZome is shown. These modifications lead to a global change in the binding profiles -and therefore in the PDZ-binding capability- providing quantitative information on the biological effect that such modifications may have in the context of full-length proteins. As any alteration in the sequence of the PBM (like mutations, modifications, or the change of the peptide length) may lead to global rearrangements of the PDZome-binding profile, such alterations are to be used and interpreted with great caution as concerns their impact at structural, cellular and whole-organism levels. Such findings could not have been possible without the holdup assay, and the subsequent curation strategy that we developed, allowing to obtain robust affinity data. We also proposed a way to quantify specificity. Altogether, the proposed approach for the quantification of affinity and specificity of motif-domain networks, might

bring the affinity-based unraveling of the quantitative human interactome a tiny bit closer.

Chapter 3

List of articles

- Rewiring of RSK-PDZ Interactome by Linear Motif Phosphorylation. Gergő Gógl, Beáta Biri-Kovács, Fabien Durbesson, Pau Jane, Yves Nomine, Camille Kostmann, Viktória Bilics, Márton Simon, Attila Reményi, Renaud Vincentelli, Gilles Trave, László Nyitray. Structure (2020). DOI: 10.1016/j.str.2020.03.010. **Published**

- Dual Specificity PDZ- And 14-3-3-Binding Motifs: A Structural and Interactomics Study. Gergő Gógl, Pau Jané, Célia Caillet-Saguy, Camille Kostmann, Goran Bich, Alexandra Cousido-Siah, Laszlo Nyitray, Renaud Vincentelli, Nicolas Wolff, Yves Nominé, Nikolai N Sluchanko, Gilles Travé. Journal of Molecular Biology (2019). DOI: 10.1016/j.jmb.2019.01.038. **Published**

- A computational protocol to analyze PDZ-PBM affinity data obtained by high-throughput holdup assay. Pau Jané, Lionel Chiron, Goran Bich, Gilles Travé, Yves Nominé. Methods in Molecular Biology. **In press**

- Interactomic affinity profiling by holdup assay: acetylation and distal residues impact the PDZome-binding specificity of PTEN phosphatase. Pau Jané, Gergő Gógl, Camille Kostmann, Goran Bich, Virginie Girault, Célia Caillet-Saguy, Pascal Eberling, Renaud Vincentelli, Nicolas Wolff, Gilles Travé, Yves Nominé. **Submitted to FEBS Letters**

Chapter 4

List of oral communications

- **Poster**: Quantitative evaluation of the PDZ-mediated human interactome and its hijacking by viral proteins. Pau Jané, Lionel Chiron, Yves Nominé, Gilles Travé. EMBO Workshop: Integrating Systems Biology: From Networks to Mechanisms to Models, 15 - 17 April 2018 ,EMBL Heidelberg (Germany).

- **Poster**: Quantitative evaluation of the PDZ-mediated human interactome and its hijacking by viral proteins. Pau Jané, Lionel Chiron, Yves Nominé, Gilles Travé. 4th symposium on non-globar proteins, 12-14 September 2018, Druskininkai (Lithuania).

- **Poster**: Quantitative evaluation of the PDZ-mediated human interactome and its hijacking by viral proteins. Pau Jané, Lionel Chiron, Yves Nominé, Gilles Travé. EMBO conference, Modularity of signaling proteins and networks, 16-21 September 2018, Seefeld (Austria).

- **Oral Presentation:** Quantitative evaluation of the PDZ-mediated human interactome and its hijacking by viral proteins. Pau Jané, Lionel Chiron, Yves Nominé, Gilles Travé. AFIPP (Association Francophone pour l'étude des Infections par les Papillomavirus et les Polyomavirus), 26 - 28 September 2018, Obernai (France)

- **Oral Presentation:** Quantitative evaluation of the PDZ-mediated human interactome and its hijacking by viral proteins. Pau Jané, Lionel Chiron, Yves Nominé, Gilles Travé. 1rst PDZnet Meeting, 2 - 4 October 2017, EMBL Heidelberg (Germany)

- **Oral Presentation:** Quantitative evaluation of the PDZ-mediated human interactome and its hijacking by viral proteins. Pau Jané, Lionel Chiron, Yves Nominé, Gilles Travé. 2nd PDZnet Meeting, 1 - 3 October 2019, Marseille (France)

- **Oral Presentation:** Quantitative evaluation of the PDZ-mediated human interactome and its hijacking by viral proteins. Pau Jané, Lionel Chiron, Yves Nominé, Gilles Travé. 3rd PDZnet Meeting, 30 September - 2 October 2019, Copenhagen (Denmark)

Part II

Introduction

Chapter 5

Protein-Protein interaction (PPI)

Cells are complex machines which develop tasks such as to grow and divide, to differentiate, to die or to keep maintaining survival functions. These tasks are the results of intricate regulations and cooperations between proteins as primary effectors. For multicellular organisms, these tasks are also performed cooperatively with other cells, either by proximity (cell-cell junctions) or at distance (cell signaling to change other cell tasks through hormones or short peptides, for instance) [37].

The "proteome" is the expressible proteins in a given organism. In the past decades, databases such Uniprot, have systematically analyzed and sort proteins depending on their localization, function and sequence [1]. However, biology do not emerge from single proteins functions. Instead, biology is the ensemble of complex interactions -and subsequent reactions- implicating many proteins and other molecules species. Therefore, the efforts are now driven to map the full "interactome" map of various organisms, which constitutes the entire network of their protein-protein interactions (PPI) [2, 3].

In this thesis, a subset of the PPI interactome is studied through the "domain-omics" approach, which focuses on minimal interacting blocks accounting for a large subset of PPI [9]: globular domains vs Short Linear Motifs (SLiM).

5.1 Globular domains

The first observation of globular domains existence in proteins was made by Wetlaufer after observing some crystal structures and their description by the authors. He defined it as nucleation (formation of three dimensional compact structures) along a native continuous polypeptide chain [38]. The definition was further improved adding fold, function and conservation as part of the globular domain concept:

- Fold: Globular domains are normally 40 to 150 residues length with a specific structure, which is normally a combination of unfolded regions (loops), α-helix and β-strands [39–42].
- Function: Globular domains are regions that provide physical contact in recognition events underlying functionality [39–43]. Globular domains may team up with other domains in the protein to develop its function or even carry their own function from the rest of the protein [44].
- Conservation: due to their functional importance, domains often spread across the proteome through genomic recombination events. Proteins may gain additional domains over evolution through several mechanisms [37, 45–47]:
  - gene fusion
  - exon extension
  - exon recombination
  - intron recombination
  - retransposition

A domain can be found in different proteins and a single protein may comprise several members of a given domain family [43]. The relevance of classifying all the globular domains by fold, conserved evolution and function raised the need of organizing them in families and store all the information in a database such as Pfam [36], SMART [48], CATH [49] CDD [50] or SCOP [51, 52]. According to Pfam 33.1 version on May 2020, there were 18259 domain families registered at the moment [36].

Globular domains can recognize other domains and sometimes cooperate with them to interact with a ligand. If globular domains interact with other globular domains, they form a domain-domain interaction (DDI). DDI can be classified in two types [47]:

- Heterotypic: The interaction is performed between two different domains which don't belong to the same domain family.
- Homotypic: The interaction involves two domains of the same family. This does not necessarily imply a homodimer interaction, since a homodimer is formed of two identical protein or domain sequences.

DOMINE is a DDI predictor database. They collect interactions inferred by the PDB [53] and 13 different computational approaches using the Pfam domain definitions [44, 54]. DOMINE database (manticore.niehs.nih.gov/cgi-bin/Domine) shows the following statistics at the date (August 2020): *"DOMINE contains a total of 26,219 domain-domain interactions (among 5,410 domains) out of which 6,634 are inferred from PDB entries, and 21,620 are predicted by at least one computational approach. Of the 21,620 computational predictions, 2,989 interactions are high-confidence predictions, 2,537 interactions are medium-confidence predictions (MCPs), and the remaining 16,094 are low-confidence predictions"*.

5.2 Short Linear Motifs (SLiM)

In 1990 Tim Hunt conceptualized the existence of linear motifs and described them as follows: *"These motifs are linear, in the sense that three-dimensional organization is not required to bring distant segments of the molecule together to make the recognizable unit. The conservation of these motifs varies: some are highly conserved while others allow substitutions that retain only a certain pattern of charge across the motif"* [14]. Over time, the concept of linear motif evolved to Eukaryotic Linear Motifs (ELM) or Short Linear Motif (SLiM), being the last one how we are going to refer to them, from now on, in this thesis.

SLiM are normally not longer than 10 residues and, typically, only 3 to 4 residues of a SLiM contribute to the core consensus motif for domain recognition,

whereas the other residues may play a modulatory role [14]. Although being short in sequence, SLiMs have a high functionality. Their low affinity binding range (1-150 μM) allows to engage the partner protein in a transient and reversible manner. As a consequence, dynamic networks may be affected by the rapid assembly or disassembly of protein complexes mediated by SLiMs [11, 13]. Every SLiM can be recognized by many similar globular domains inside a same family hugely increasing the interactome size [19]. SLiMs are also post-translational modifications (PTM) sites, which alters their physicochemical behavior as well as their affinity binding to their binding partners [13]. Moreover, the appearance of new motifs in a system, by convergent evolution or by viral hijacking (see section 6.4), have the capacity to rewire existing cellular networks. Altogether, these characteristics confer to SLiMs a huge relevance in PPI, which provides the biologist with a great opportunity to study relevant domain-motif sub-interactomes [10, 12, 55].

Detecting SLiM across the proteome can be computationally more challenging than identifying globular domains. As mentioned, SLiM are normally not longer than 10 residues and are located in intrinsically unfolded regions of proteins. Interestingly, some SLiM adopt secondary structures once bound to their protein partners. Moreover, their conservation is not as strong as in globular domains, as they are evolutionary plastic by rapidly appearing *de novo* or gradually disappearing. This complicates searches of SLiMs through the proteome since it makes it statistically difficult to discriminate between true and false positives. Only early in this century, computational tools started to make some progress in the SLiM identification. Therefore, most likely only a small fraction of the existing SLiM has been discovered at the moment [10–13, 56].

The human proteome is estimated to contain over 100 000 SLiMs, which can raise up to 1 000 000 by their capacity of accepting PTM [13]. The systematic computational identification and curation of domain-SLiM interactions has opened wide new perspectives for our understanding of PPI networks at a proteome wide scale. [10–13]. There are many resources nowadays to find SLiM in human proteome, such as SLiMSearch [57] or databases like ELM database, the most comprehensive repository of experimentally proved SLiMs, which stores consensus motifs as ligand sites, PTM sites, proteolytic cleavage and processing sites and subcellular targeting sites [58].

In this thesis, we will focus on the PDZ domains, which are a globular domain family that recognize SLiMs located at the C-terminus site (see chapter 6).

5.3 Post-translational modifications (PTM) as PPI modulators

Amino acids in proteins can be subject to reversible biochemical changes, which are called post-translational modifications (PTM). There are about 300 different modes of PTM in the human proteome, among which glycosylation, phosphorylation, ubiquitination, acetylation, and so forth [13]. The PTM is introduced due a chemical reaction catalyzed by another protein or ligand and may change the function of a protein or its cell localization. It might also change the motif consensus of a SLiM by shifting the binding preferences of their globular domain partners [25, 33]. According to Tompa *et al.* [13], there are around 1 000 000 PTM sites in total. This leads to a huge PPI interactome, from which we barely scratched the surface.

In the literature, PTM are usually referred to as "switches" for a given PPI [18]. In other words, PTM have the property of turning "on" and "off" an interaction. This is a simplification since interactions can sample different affinities; but this simplification arises from the fact that many interaction assays only deliver binary information ("binding" vs "not binding"). Here we will try to avoid this simplification and therefore we will not use the term "switch" onward. We believe that PTM play a regulatory function in PPI, but that affinities are a continuum, which are quantifiable within the limits of the sensitivity and quantitation threshold of the experimental methods that are used.

In this thesis, we studied phosphorylable and acetylable SLiM which binds to PDZ domains in both states, wild-type and in their PTM form (see chapter 9, chapter 10 and chapter 11). Our results led us to point out the inaccuracy of the term "PTM switches".

5.3.1 Phosphorylation

From the chemical point of view, phosphorus can form five covalent bounds due its 5 free electrons of its outer shell. When combined with four oxigen molecules, the phosphate group is created. In cells, phosphate forms esters with adenosine triphosphate (ATP), which constitutes the major energy storage compound. Moreover,

ATP is available to transfer the pospho group to other molecules, in particular proteins. The transfer of a phospho group into a molecule is called phosphorylation [22].

Phosphorylation is the most studied PTM in nature. Phosphorylated amino acids in proteins gain negative charges and can create extra hydrogen bonds or salt bridges, either intra- or intermolecularly, due to their higher density negative charge and the larger hydrated shell compared to the negative charged amino acids, acid Glutamic and acid Aspartic. This looks very suitable to form extra bounds to the guanidino group of the Arginines [59]. The unique size and charge opens new possibilities of recognition by phosphospecific-binding domains in proteins. Phosphorylated-dependant proteins interactions are crucial for the PPI in cells, which leads to changes in the interactome due alteration of the protein recognition, subcellular location changes or degradation of proteins [22].

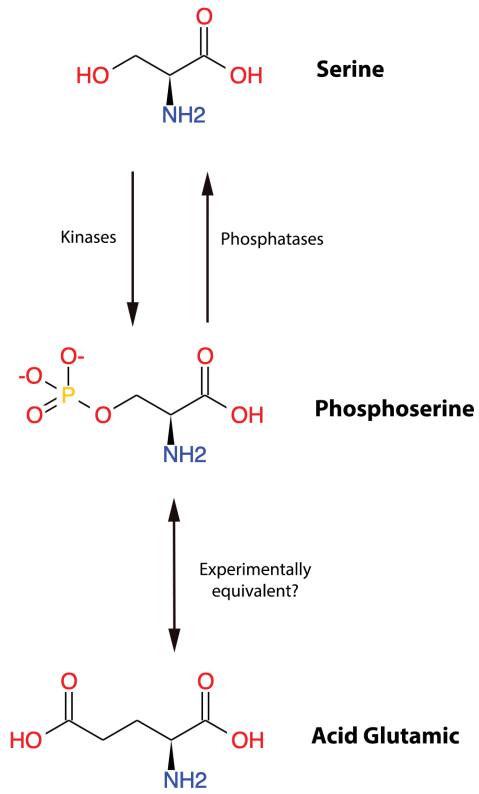


Figure 5.1: **Serine phosphorylation and its equivalently used Acid Glutamic.** A phosphogroup is transferred to the side group of the serine by a kinase. This process can be reverted with a phosphatase. Phosphoserine is sometimes phosphomimicked usind acid glutamic.

The phosphorylation of a residue in a protein is catalyzed by protein kinases. Kinases transfer a γ-phosphate group of an ATP molecule to a receiver residue, most often to the hydroxyl group of a Ser/Thr via forming a phosphoester bond. Depending on the kinase type, the γ-phosphate group is transferred to a Tyr residue. Despite its covalent nature, the phosphorylation event is reversible in the protein, as it can be removed by the reverse enzymes of the kinases, called protein phosphatases [60, 61].

Phosphorylation is not always easy to study since proteins have many phosphorylation sites and they are not so easy to control from the experimental point of view. Depending of the assay conditions, it might not be trivial to promote the addition of PTM selectively on a particular residue of interest, therefore many studies uses what is called "phosphomimication". The phosphomimication is the engineered mutation of a Serine or Threonine into an Glutamic/Aspartic acid with the aim to reproduce the biochemical effect of a phosphorylation, despite being chemically distinct [6, 19, 29, 62, 63] (fig. 5.1). In the practice, neither the charge nor the size of an acidic residue will be the same as in a residue modified by a phospho group. According to Pearlman *et al.*, a vicinal pair of Asp or Glu would serve better than a single Asp or Glu as a phosphomimic due to its double negative generated charge [64].

5.3.2 Acetylation

An acetyl group is a methyl group bounded to a carbonyl (C=O). The process of introducing an acetyl group to a molecule is called acetylation. Acetylation occurs in a large variety of substrates, such as histones, acetylases, DNA-binding transcription factors, nuclear import factors and α-tubulin [23]. Two kinds of acetylation may occur: the N-acetylation and the Lysine acetylation. While the first one is critical in synthesis, stability and localization of proteins, the second is more related to transcription factors, effector proteins, molecular chaperones, and cytoskeletal proteins [24]. In this thesis, we are more interested in the Lysine acetylation.

In Lysine acetylation, an acetyl group is transferred by an acetyl-coenzyme A to the ϵ-group of this amino acid in a protein. As other PTM, acetylation is promoted by an enzyme, which in this case is a lysine acetyltransferase (KAT),

and it can be reversed by a lysine deacetylase (KDAC) (fig 5.2). Lysine acetylation will neutralize the positive charge, but also will make the protein "readable" by specialized acetyl-lysine binding domains. Thus, acetylation has a high biological impact in cell regulation [24, 65, 66].

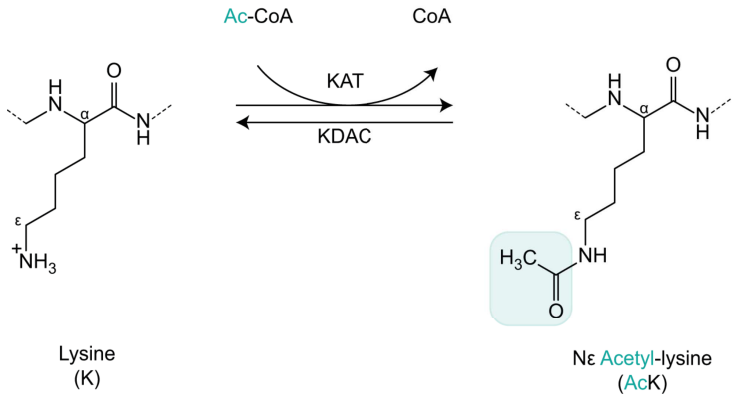


Figure 5.2: **Lysine acetylation mechanism.** An acetyl group is transferred by an acetyl-coenzyme A (KAT) to the ϵ-group of the Lysine. This is a reversible process due a lysine deacetylase (KDAC). Figure adapted from [66].

.

As with the phosphorylation, acetylation is being studied sometimes with some alternative mutations. In the PTEN protein case, the mutation of a lysine to an arginine creates a constitutively non-acetylable state of the protein. However, an arginine can be more reactive than lysine and therefore, it might change the binding properties of the mutated protein in the study [25, 67]. Other studies explored "more creative" ideas to mimic the acetylation, such as substituting the lysine residue by a glutamine or by an alkylated cysteine residue [26, 68].

Chapter 6

The PDZ domain as a PPI case study

PDZ domains were first discovered by Cho *et al.* when they identified highly conserved domains in postsynaptic density 95 protein (PSD-95) in the posynapse brain in rat was highly homologous to Drosophila lethal(1)discs-large-I (DLG) [16]. These domains were originally named GLGF domains (Gly-Leu-Gly-Phe) because of their repetitive and conserved features. They were also named DHR domains for disc large regions. The current PDZ domain name came later, after identifying these conserved sequence repeats for the **P**SD-95, **D**LG and zonula occludens-1 (**Z**O-1) proteins. We currently know that the PDZ domains are spread across the entire living kingdom as they are also found in metazoans, plants, bacteria and yeast [69].

We will here focus only on the human proteome and its 266 PDZ domains [15], mentioned henceforth as the PDZome, since it is a well characterized model for SLiM/Globular domain interaction.

6.1 Structure

PDZ domains are in average no longer than 80 to 100 residues. Some special cases of PDZ domains may be longer, such as GORASP2. Their structure were first solved in the apo state by Cabral *et al.* [70] . The first PDZ domain solved in a complex, a Cysteine-rich PDZ-binding protein (CRIPT), was solved by Doyle *et al.* [17]. The canonical folding topology of PDZ domains is as follows: three β-strands (β1 to β3), a short α-helix (α1), two β-strands (β4 and β5), a long α-

helix (α2) and a last β-strand (β6) (fig 6.1a), mostly described as an anti-parallel β-barrel or a β-sandwich [16, 17]. Although the fold is conserved, the secondary structure elements may vary in length [18].

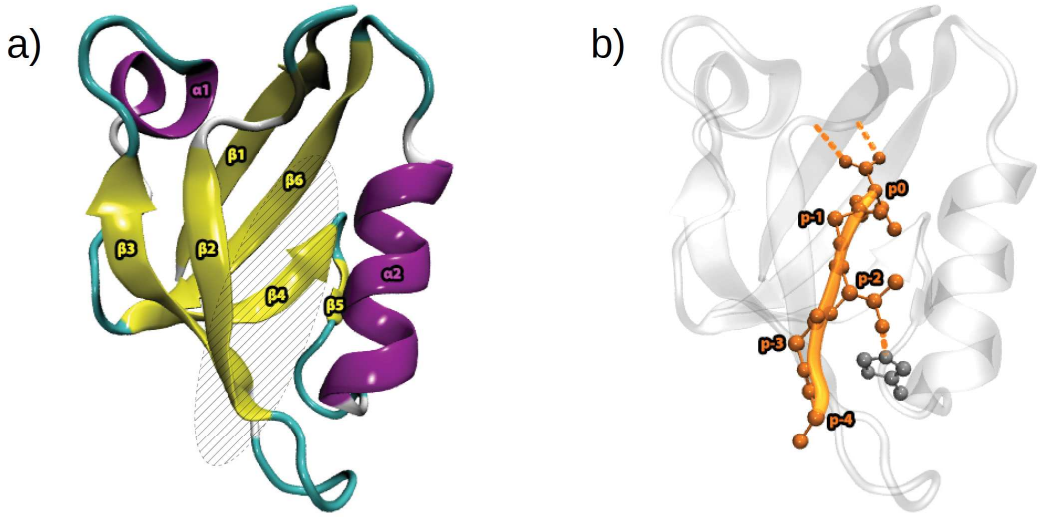


Figure 6.1: **Structure of a canonical PDZ domain and its recognition site.** (a) Canonical fold of a PDZ domain colored by secondary structure. The shadowed ellipse shows where the contact with the ligand happens. (b) PBM (in orange), fold as an antiparallel β-strand makes contacts with the PDZ domains in both, p0 to the β1-β2 carboxylated binding loop and p-2 establishing hydrogen bonds to a Histidine (in grey). This is an example of a class 1 motif.

The peptide-binding pocket of PDZ domains is mainly hydrophobic. It is formed by residues from the β1-β2 loop bearing the conserved [KR]...GϕGϕ sequence, the β2 and β3 strand, the β2-β3 loop and the α2-helix. ϕ is hydrophobic, mainly Leucine and to a lesser extent Phenylalanine. "." means any amino acid. Among those, several residues found at the interface between the β2-strand and the α2-helix of the PDZ domain is recognized by the PBM, which adopts a β-strand upon binding. On the opposite site of the binding pocket, the N- and C-terminus of a canonical PDZ domain are proximal to each other [15, 18–20].

However, not all 266 PDZ domains follow the same canonical fold. Some have undergone alterations or even circular permutations of the fold. For instance, while HTRA2 has five long β-strands instead of six, it comprises two extra small β-strands at the N- and C-terminus plus an extra α-helix found in the loop between β2 and β3. The Golgi-reassembly stacking protein 2 (GORASP2) has undergone a "circular permutation", since it has its N-terminal end at the beginning of the β3-strand (being therefore this its β1-strand), then it follows a regular PDZ domain

fold, but instead of finishing in the canonical β6-strand (β4 in its case) after the α2-helix, it forms the missing β1 and β2-strands of the canonical form. This PDZ domain therefore does not comprise the usual β2-β3 loop [18, 71, 72].

6.2 PDZ-Binding Motifs (PBM)

PDZ domains bind to conserved SLiM, called henceforth as PBMs (for PDZ-Binding Motifs). PBMs are generally situated at the extreme C-terminus of their partner proteins, yet they can also exist as "internal" motifs, i.e. the protein sequence is continued at both the N-terminal and C-terminal extremities of the motif. Besides PBM, some PDZ have also be found to interact with some lipids [19, 73–77]. However, in this thesis we will mainly focus on the C-terminus PBM.

6.2.1 C-terminal PBM

C-terminal PBMs have a specific nomenclature in the literature. The last position of a C-terminal PBM is called position 0 (p0), the preceding residue position -1 (p-1) and so forth. In all C-terminal PBMs, p0 is an hydrophobic residue, mostly Valine, Leucine and rarely, Isoleucine [18–20], according to a stringent definition. However ELM database [58] extens the p0 definition to include Alanine, Cysteine, Tyrosine and Phenilalanine. PBMs can be grouped in three classes depending on the amino acid found at p-2: class 1 motif contains a Serine/Threonine, class 2 motifs contains a Valine/Tyrosine/Phenylalanine/Leucine/Isoleucine and class 3 motifs contains an Aspartic/Glutamic acid (see table 6.1) [21].

Table 6.1: **Three defined PBM classes,** where X = any amino acid and Ψ = hydrophobic (VLIFY).

Class	p-2	p-1	p0
I	Ser/Thr	X	Ψ-COOH
II	Ψ	X	Ψ-COOH
III	Asp/Glu	X	Ψ-COOH

The binding of PBMs to their cognate PDZ domain involves several hydrogen bonds established between backbone atoms of the peptide and backbone atoms of the PDZ domain observed typically between β-sheets. The p0 carboxylate group

establishes contact with the conserved [KR]...GϕGϕ β1-β2 loop of the PDZ domain. The p-2 also plays a crucial role by binding to several residues of the PDZ α2-helix (fig 6.1b). The central residue of α2-helix varies depending on the PBM class: it may be a Histidine, an aliphatic residue or a Tyrosine when bound to a class 1, 2 or 3 motif, respectively [18–20]. Contrary to other cases of domain-motif interactions, PDZ/PBM interactions prioritize mostly size over biochemical characteristics of the amino acids due the limited size of the PDZ domain hydrophobic pocket. In this regard, the fact that phophoresidues and phosphomimetic amino acids display different sizes might explain some of the binding differences previously observed between phosphorylated and phosphomimetic motifs [6, 29, 62, 63].

According to the class consensus sequence definition of ELM [58], it is estimated to be approximately 3617 putative PBMs in the human proteome, when searched for possible candidates in SLiMSearch4 [57].

6.2.1.1 PTM in C-terminal PBM

Some important studies have addressed the relevance of phosphorylation in PBM/PDZ binding. Clairfeuille *et al.* pointed that phosphorylation enhances PBMs/SNX27 binding in *a priori* not relevant PBM positions [78]. Also acetylation in PTM are lately studied due its relevance. Ikenoue *et al.* pointed the PTEN/PDZ binding modulation via acetylation [25]. While these low scale studies were interesting proofs of concept, more work is need to explore the relevance of PTM in PBMs and its impact in the interactome. Fortunately, some efforts are moving to the large-scale PPI detection of PTM involved in PBM/PDZ interactions (see section 7.3) [33, 79, 80]. Noteworthy, most of the studies refers to PTM in PBMs as switch mechanisms for PDZ domain binding recognition. As happen with other studies, PTMs in PBMs are also studied with the mimication techiniques mentioned in subsection 5.3.1 and subsection 5.3.2.

In this thesis, RSK1 (box 1) and PTEN (box 2), two PBM containing proteins, which are affected by phosphorylation and acetylation respectively, are studied along the chapter 9, chapter 10 and chapter 11. For this, we needed the holdup assay (see subsubsection 7.3.2.4), a powerful high-throughput screening method with high sensitivity that can use fully synthetized PBMs.

Box 1 RSK1

RSK1 belongs to the ribosomal S6 kinases (RSK) related to the MAPK pathway. RSK is involved in cancer types such as glioblastoma and melanoma. The RSK1 C-terminus segment comprises a series of binding sites, sometimes overlapping, for its companion kinase ERK, S100B, a tyrosine kinase, and a phosphatase, and, at the extreme C-terminus, a class 1 PBM (RRVRKLPSTTL-COOH). This PBM also has three potential autophosphorylation sites (p-1, p-2 and p-3) [33, 81, 82]. From these, p-3 has been also previously studied in its phosphomimetic form [6]. We studied the RSK1 PBM in the wild type, the phosphorylated form (see chapter 9 and chapter 10) and the phosphomimic substitution (see chapter 10).

Box 2 PTEN

Phosphatase and tensin homolog deleted on chromosome 10 (PTEN) protein is well expressed in brain tissue. It is mostly found in neurons and synapses and has been related to neuronal survival (neurite outgrowth and axonal regeneration) [83–85]. PTEN harbors a class 1 C-terminus PBM (DEDQHTQITKV-COOH) which can be putatively acetylated at p-1 [25]. We studied the PTEN PBM in the wild-type and the acetylated form, as well as in the constituently unacetylated form by mutating K to R at p-1. We also tested the influence of extending the N-terminal sequence of the PTEN PBM (see chapter 11).

6.2.2 Internal PBM

Internal PBM are by far less studied than C-terminal PBM. They were first mentioned in the literature by Hillier *et al.* in 1999 [86]. Like C-terminal PBMs, internal PBMs also display a β-strand mechanism in which p-2 forms an hydrogen bond with the α-helix of the PDZ. An hydrophobic residue is also needed at p0, but in

this case it seems to require specifically a Leucine or Valine, due size matters, to establish bonds with the hydrophobic pocket in the PDZ. However, this p0 is not at the C-terminal end of the protein, hence an acidic residue is needed upstream of p0 to mimic the C-terminus carboxylated group. The carboxylated side chain of the acidic residue will form the same hydrogen bonds to the [KR]. . . GϕGϕ as a regular C-terminal carboxyle [86–88].

These "unconventional" PBM have been barely classified so far. The ELM database currently proposes the following class definition: "W.0,1[VIL].[ST]. KA0,1T...W" [58], although other instances have been described in the literature [76, 86–88]. Indeed, two other PhD students of the PDZnet european training network have dedicated their research to the experimental and bioinformatical identification of internal PDZ-binding sites.

6.3 PDZ participates in cell polarity

In cells, PDZ domains serve as scaffold domain regulating a wide range of biological processes such as protein traffic, signal transduction, cell-cell junctions and adhesion. Most of these processes are related to cell polarity, and therefore, to cell asymmetry. SCRIB, DLG, MPP5 and PAR proteins are PDZ containing proteins, which participate in the cell polarity control. The polarization also plays an important role in cell migration and asymmetric cell division, crucial during embryonic stages or tissue reparation. These are regulated by Rho GRPases and PAR3, SCRIB and Crumbs PDZ-containing proteins [89].

A variety of PDZ proteins, like SCRIB, MAGI, DLG, have been associated to cell polarity related diseases in particular viral-induced ones, such as neuronal disorders (rabies, Tick-borne encephalitis. . .) and cancers (cervical cancer, HPV-positive head and neck cancer, Merkel cell carcinoma. . .) [56, 90].

PDZ in Epithelial Cells Epithelia cells display an apical-basal polarity. The apical side serves as a barrier between the outside organism and the inner side, while the basal side is attached to extracellular matrix. From cell to cell, a series of specialized junctions are formed in the lateral part of the cells. These form a paracellular barrier and increase the complexity of the cell polarization, discrimi-

nating, besides the apical and the basal domains, the tight junction, the adherent junction and the basolateral domain [91].

Adherent junctions regulate the cell-cell adhesion linking the cell to the cell layers. Cadherins and catenins link the cytoskeleton to AFAD, a PDZ containing protein. Tight junctions are located above the adherent junctions constituting the border between the apical and the lateral border. Tight junctions create a channel in which diffusion of soluble molecules between cells are allowed. These are mainly organised by the TJP and MAGI proteins, both of them containing also PDZ domains [89, 92–94].

PDZ in neurons Like in epithelial cells, neurons also have polarization displayed as dendrites in the cell body and extended to the terminal axon. The axon specification is controled by PAR3 and TIAM1 PDZ containing proteins.

Synaptic transmission communication between pre- and post-synaptic neurons is essential in neuronal cell signalling. The pre-synaptic neuron comprises the axon terminal while the post-synaptic cell comprises the dendrites. Both cells are connected via cell-adhesion molecules moldulated by APBA1, LIN7 and DLG4 PDZ domain containing proteins. GRIP1, PICK1, DLG1, TIAM1, SHANK, nNOS, regulate mostly the excitatory receptors in the post-synaptic neurons by regulating the clustering and localization of the NMDA and AMPA receptor channels (fig. 6.2) [69, 95, 96].

Whirlin is a PDZ domain scaffold protein found in the auditory hair cells and photoreceptor cells, which are specialized neurons of the sound and vision senses. Whirlin organizes the complexes that maintain the membrane proteins to the cytoskeleton and are crucial to develop and maintain the cilia. Deafness and blindness diseases are related to the mutation of this protein in the N-terminal or its C-terminal regions, the latest containing a PBM as well [97–99].

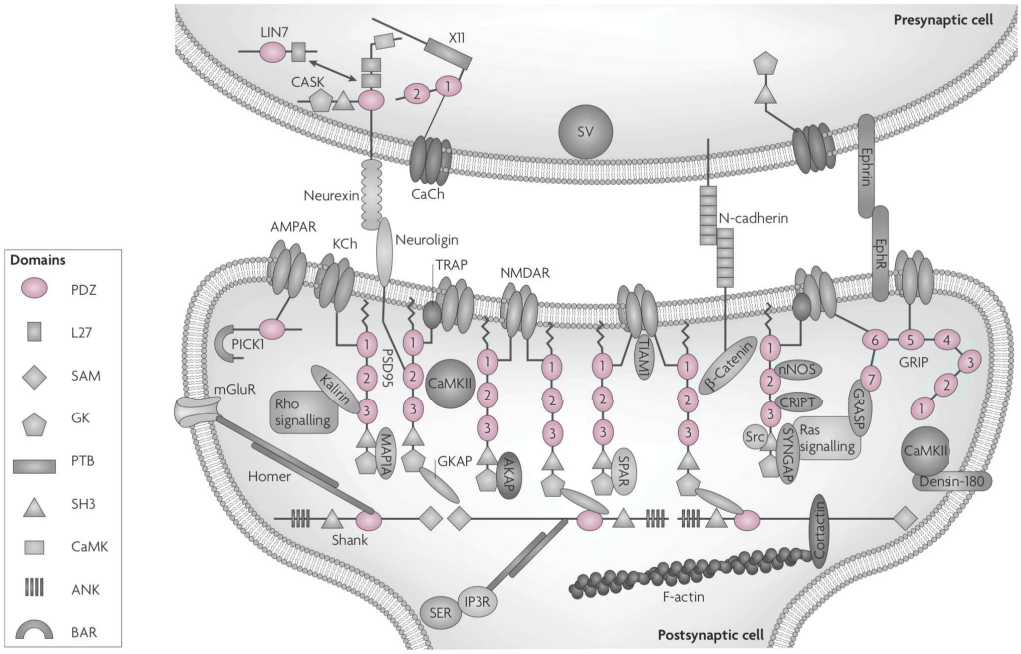


Figure 6.2: **Organization of the postsynaptic density proteins.** It is comprised by membrane receptors, ion channels, adaptor proteins and scaffolds, signaling proteins, cell-adhesion molecules and cytoskeleton. PDZ domains are highlighted in pink. This figures has been adapted from [95].

6.4 Viral hijacking

Differentiated cells are polarized and do not replicate. Many viruses perturb the polarization to initiate transcription and replication of the cell, and therefore, to replicate the viral genome. In the past 20 years many viruses have been found to harbor a PBM at the C-terminus in at least one of their proteins [90, 100].

It is estimated that 12% of all humans cancers is caused by virus [101]. The majority of these, are long-term persistent infections in the host that uses the cellular differentiation program for their replication. In such cases, the viral oncoproteins will interact with the cell environment initiating processes ending to malignant transformation of the cell into cancer. Examples of such oncoproteins which contains PBM at its C-terminus are Tax protein of Human T-lymphotropic virus type 1 (HTLV1), E4-ORF1 protein of Human Adenovirus and E6 protein of Human Papillomavirus (HPV) [90, 100, 101].

Other viruses, which are not long-term oncogenic persistent may also harbor

a C-terminus PBMs, like Influenza A, SARS coronavirus and Rabies virus [100]. These are also highly pathogenic. In the case of Influenza, NS1 protein from human influenza A virus harbors the PBM. The E protein of the integral membrane protein in SARS coronavirus harbors a PBM able to interact with PALS1 PDZ protein. This is the interaction suspected to affect the integrity of the lung epithelia, which furthers evolves to the acute respiratory syndrome constituting the main problem in the current pandemia [90, 100, 102]. G protein from Rabies virus targets neural enzymes during its infection. In this last case, it has been suggested that the virus strategy may inspire us to learn more about neurodegenerative processes, for instance, and how to use mimic viral sequences to develop innovative therapies to manipulate cellular homeostasis [34, 103].

6.4.1 Human Papillomavirus

Human Papillomavirus (HPV) are part of a large family of small DNA virus which targets skin or mucosal epithelial cells [104]. HPV can be grouped in "high-risk" and "low-risk" depending on their oncogenic risk for humans. High-risk groups are these that when infecting epithelial mucosa may cause warts, precancerous lesions or cancer in long-term, mainly cervical, mouth, anogenital or throat cancer. Moreover, the high-risk HPVs are linked by containing a PBM class 1 motif. [90, 100, 105]. The number of raising cancers associated to HPVs are over 500 000 cases worldwide [106, 107]. The most common examples of the high-risk HPV are HPV16, HPV18 and HPV35 [90, 100].

Tumorigenesis induced by HPVs is associated to their early proteins E6 and E7 [108–110]. Our group is experienced with the HPV and solved the 3D structure of the MAGI1-2 bound to HPV16 E6 protein via its PBM. [105]. Moreover, E6 degradates p53 protein due to its recruitment of ubiquitin ligase E6AP and subsequent degradation. This recruitment occurs due to an acidic leucine(L)-rich motif containing LxxxLL consensus sequence within E6AP that is captured by E6. Most of the PDZ domains found to interact with E6 PBM are related to cell polarity regulators, such as MAGI, DLG, SCRIB and PATJ proteins (fig. 6.3). All together, allows E6 to interact with many cellular proteins related to cancer pathway [56, 93, 111, 112].

We performed a search of all the HPV E6 oncogenic strains collected from the

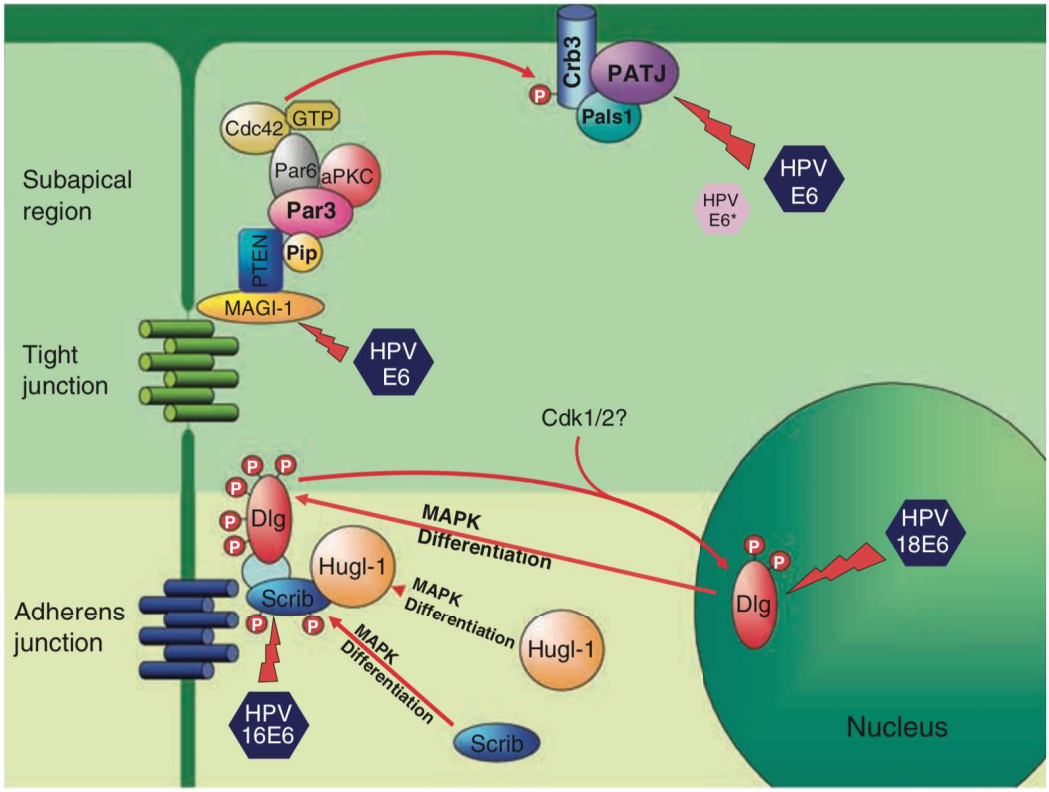


Figure 6.3: **Human papilloma's oncogenic viral protein E6 targeting.** E6 targets in a coordinated way to cell polarity regulators in epithelial cells. Figure adapted from [93]

.

literature [113–117], obtaining 'RRETQV' as a consensus sequence (fig. 6.4). It seems that E6 oncogenic proteins requires from a positive charge tailed located in the p-4 upstream, while an acid glutamic shows to be also necessary at p-3. A similar analysis was performed with 65 reviewed E6 proteins in Uniprot, where 19 of them showed a canonical class 1 PBM [118]. Their consensus sequence at its C-terminal was "RRETQV" as well.

Interestengly, this consensus sequence is, besides a PBM, a protein kinase A consensus motif (x-R-R-x-S/T-x), where "x" is any amino acid. The phosphorylation seems to occur at p-2, which potentially disrupts E6 binding to the PDZ domains [118–120]. This phosphorylation has been found to turn the E6 PBM into another motif binding for the 14-3-3 proteins [121, 122].

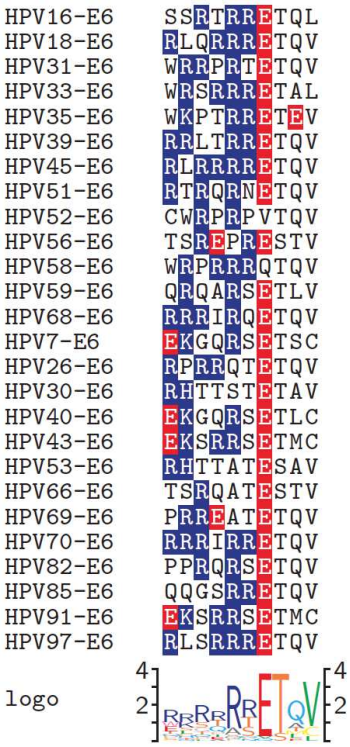


Figure 6.4: **E6 oncogenic C-terminal consensus sequence.** Logo analysis of PDZ motif-satisfying HPV E6 sequences which are, according to the literature, oncogenic strains [113–117]. This E6 oncoproteins from the extracted HPV strains reveals that the most common PDZ-targeting sequence motif in these proteins is RRETQV, with a mixture of more positive amino acids in the upstream site.
.

6.5 14-3-3 domain family

The lab started recently the study of the 14-3-3 domains. The 14-3-3 term was coined for the first time in 1968 by Carlson *et al.* who discovered an abundant protein class in mammalian brain. The authors eluted such proteins in the 14th fraction of bovine homogenate of a homemade DEAE cellulose column and the fractions 3.3 in the starch gel electrophoresis, hence the name "14-3-3" [123]. Many authors later discovered or rediscovered proteins of the same family and gave them other names, such as BAP-1, DER, FBP, MSF or NIP among many others. The 14-3-3 family now comprises 7 isoforms in mammalian cells, with at least two isoforms in every known organism [124].

A 14-3-3 protein was first crystalized in 1995 by Xiao *et al.* [125]. Each

monomer of the family comprises nine antiparallel α-helices, which together create a negatively charged channel [125]. 14-3-3 domains recognize 3 different motifs being these the "RSx**pS**xP", "Rx[Y/F]x**pS**xP" and the " Rxx[**pS**/**pT**]xx-COOH", where "x" means any amino acids and **pS** or **pT** are phosphoserine or phosphothreonine respectively. When searched in the ELM database, this last motif is further described as "R[^DE]{0,2}[^DEPG]([ST])[^P]{0,1}-COOH", where "^" means not allowed amino acids and between "{ }" how many amino acid extra by position are allowed [58]. Interestingly, this last motif is a kinase motif and a PBM at the same time [122]. 14-3-3 proteins are involved in signal transduction, apoptotic cell death, and cell cycle control [126, 127].

6.5.1 14-3-3 in the PDZ context

14-3-3 domains in the PDZ context are barely studied, but it is suggested that this protein family can indeed recognize phosphorylated class 1 PBM [128]. It has been proposed that some PBM alternates their binding to PDZ or 14-3-3 by the phosphorylation of one of their residues due its included kinases recognition motif, affecting E6 proteins from HPV as well [122]. This reveals a paradigm of phosphorylation regulation that affects biology in a dimension barely explored before. In this thesis, we will try to go a step further and study the interplay between PDZ and 14-3-3 for PBM binding domains (see chapter 10).

Chapter 7

Deliverables in PPI study

PPI needs a way to be measured, and if possible, to be quantified depending on the strength of the formed complex and on the chance of meeting of the involved proteins. In the literature these attributes are generally described as "affinity" and "specificity". While affinity determines the strength in a complex between a ligand and a protein, specificity describes the overall affinity strength and chances of binding of a ligand 'A' for a subset or protein family in which belongs protein 'B'.

7.1 Affinity

The affinity value is crucial in molecular biology as it determines the binding strength of complexes involving biological macromolecules, including proteins. This means that we can study and understand the biology of molecules in a quantitative way rather than in a qualitative way ("binds" or "does not bind"). The binding interaction of a protein A and a ligand B can be defined as the dissociation constant (K_d):

$$K_d = \frac{[A][B]}{[AB]} \tag{7.1}$$

Where [A] is the concentration of protein A, [B] the concentration of ligand B and [AB] is the concentration of the formed complex. The higher the relative

concentration of complex [AB] -and therefore the stronger the interaction- the lower the K_d. The inverse of the dissociation constant K_d is the association constant K_a.

Another way to depict affinity is through the Gibbs free equation:

$$\Delta G_{AB} = \Delta H - T\Delta S = -RTlnK_a \tag{7.2}$$

Where H is the enthalpy, T the temperature and S the entropy of the system. R is the gas constant.

Quantifying interactions in any assay will add rigor and therefore will minimize the amount of incorrect information produced in the obtained PPI dataset. This can result in a "high-quality" dataset which can be further used in computational studies aiming to predict PPI or to model certain biological process [129–133].

7.2 Specificity

Specificity is mostly tackled in the bibliography as a qualitative measure dependent of the affinity. In other cases, specificity is quantitatively dependent of the affinity for only few selected interactions, that represent only a small fraction of all the putative ones. However, these specificity descriptors are not enough to assess the complexity of a full interactome. For a PPI, such as SLiM/globular domain interaction, only a combination of interface complementarity, extra interactions outside the SLiM core consensus, avidity from a multidomain protein and spatial and temporal regulation of expression will solve the real specificity of the system [134]. However, all this factors do not act in the same level in a determined PPI:

- **Macro state level:** This is related to temperature, protein expression and localization. All these parameters have a strong influence in PPI and will play a role in specificity as will modify the encounter between a subset of the proteins for a determined ligand in cell. For instance, a high affinity ligand 'A' may never encounter its theoretical protein partner if a low affinity ligand 'B' is over expressed at enough level to cancel the first ligand 'A' out. Noteworthy, if a protein and ligand are never found to be co-expressed

simultaneously or in a same compartment, their *in vitro* measured affinity and specificity may have little relevance.

- **Micro state level:** This is related to the single state of each amino acid in the SLiM (in the core consensus or out of it), interface complementarity and avidity from multidomain or tandem domains. Although I introduced our PPI approach happening as "blocks" in protein (i.e. globular domain and SLiM), these should not be considered as static structures. Specially in the case of SLiMs, since their unfold nature in the protein allows its amino acids to adapt different conformations (i.e. rotation of the sidechain in the space). This confers to the disorder states a high variety of freedom degrees changing the entropy for the PPI interaction, and therefore, playing an important role in equation 7.2. A simplified example of it can be observed with the Ramachandran plots. On the one hand, Glycine has specially a lot of freedom degrees, due its flexibility because of the lack of sidechain, being found around the Ramachandran plot (fig. 7.1) [135, 136]. When glycine folds, the freedom degrees drops to zero decreasing the entropy and therefore decreasing the affinity and the specificity. On the other hand, proline only has one freedom degree, due its rigid full cyclic sidechain, and when folded in a protein, its degree of freedom remains zero making the entropy variation to stay 0.

For a SLiM/Globular domain interaction, specificity would be ideally approached with transcriptomics (macro state) and proteomics (micro state level). Noteworthy, the proteomic approach should deliver affinity data and be obtained for an entire families of proteins, needing therefore, high-throughput screening (HTS) methods [131, 137, 138]. Together, transcriptomics and proteomics, could help to determine the specificity of a given system or subset of proteins according to physiological conditions [139, 140]. In this thesis, we will focus on proteomic analysis using the holdup, which delivers affinity data from a given PBM against the whole PDZome (see subsubsection 7.3.2.4). We will also see a quantitative way to measure specificity from the proteomics point of view (see chapter 11).

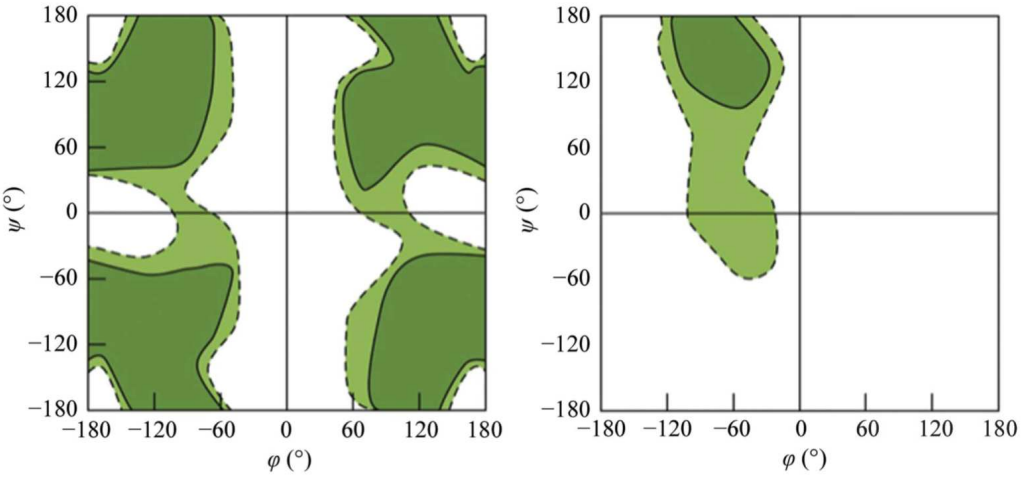


Figure 7.1: **Ramachandran plots for glycine (left) and proline (right).** Glycine is flexible and proline rigid due their absence of side chain and the presence of penta-atomic heterocyclic ring respectively. The continuous lines shows their allowed regions in the space, while the dotted lines shows the partially allowed. This figure has been taking from [136].

7.3 HTS techniques to study PPI

More and more, the literature is shifting from studying a single protein to take a big picture of PPI [141]. Biology does not emerge from single proteins, but rather from complex interactions. This made it mandatory for scientists to change the scope and focus on the real network of cells. The mapping of the human PPI is one of the key challenges in the post genome era [8]. Many strategies have been developed that are up to this task. In the past years several high-throughput experimental approaches have been used to map the interactome, and most likely, new PPI methods are going to be developed in the next years. The challenge of the interactome study relies not only in the quantity of measured data, but also in the quality of these.

7.3.1 Full length protein approaches

Full length protein measurements are performed either in small or in large scale. Most of the large scale direct physical PPI interactions are measured with a "binary" outcome (i.e. does it binds or does not binds). One of the most used

approach for that purpose is the yeast two-hybrid (Y2H) [2,142]. PPI interactions can be also studied by the interaction of several proteins in a complex (known as co-complex methods), one of the most used approach being the tandem affinity purification coupled to mass spectrometry [2,143].

On the one hand, a direct partner (the prey) of the target protein (the bait), can be found using direct assignment assays, like Y2H. Noteworthy to mention, efforts have been done to detect PTM interactions in Y2H assays [80]. On the other hand, co-complex methods catch both direct and indirect partners of the bait in the PPI. For this method, the bait is tagged with a molecular marker, which will "fish" a group of preys. This will be split from the other proteins either by pull down or by co-immunoprecipitation (CoIP). In such cases, the interpretation of which bait is targeted by which prey is usually more difficult to perform (fig. 7.2) [2].

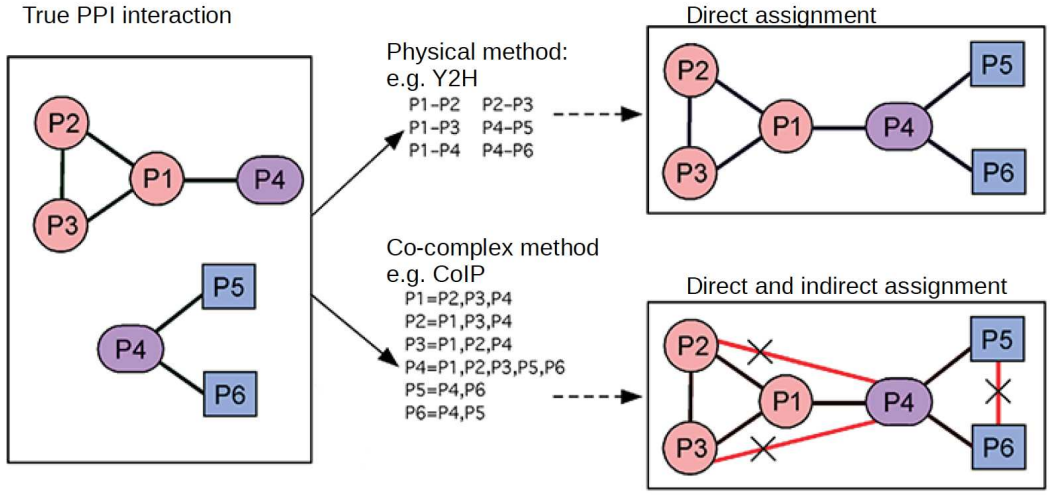


Figure 7.2: **Differences between direct and indirect assignment.** Two different PPI networks are derived from two different types of experimental data. Although in a binary way, Y2H is able to measure direct interactions between proteins. Co-complex measures both, direct and indirect interaction, leading to a wrong PPI map if it is not carefully interpreted. The "X" in the red edge indicate false positive interactions. This figure has been adapted from [2].

Full length protein approaches needs to be validated through many orthogonal assays. Recent efforts have been done to simplify these validations, like the NanoLuc two-hybrid (N2H) system [144]. Although, these methods lack a quantitative affinity values, and therefore may have a limited sensitivity, they are suitable to fully map the human protein interactome in a "binary and un-biased" way (es-

pecially in the direct assignment methods, like Y2H). An example of it is a recent database called HuRi, which contains 53 000 interaction PPI data constituting a good backbone to further study of the relevance of such interactions [140].

7.3.2 Domainomic approaches

As mentioned before, the "domainomics" approach is defined as: a small building block interaction, between a globular domain and a Short Linear Motifs (SLiM), which accounts for a large subset of PPI [9]. This kind of approaches requires a direct physical approach of measurements and needs from an extra calibration or validation method, such as Fluorescent Polarization (FP) (see Box 3).

Box 3 Fluorescent polarisation (FP)

FP is a PPI measure instrument, which uses labeled peptides. The ligand is labelled with a fluorophen, like fluorescent protein (GFP), and titrated with varying unlabeled analyte concentrations. The GFP will then emit light when stimulated with it. This light is polarized, making the emitted light also polarized in certain degrees depending on the rotational diffusion rates of the ligand. This rotational diffusion rate of a molecule depends on its molecular weight and shape. Therefore, the higher the rational diffusion rate, the lower the fluorescent polarization.
The principle is that if this ligand binds to a given protein, the complex of both, ligand and protein, will dramatically increase the molecular weight of our labelled ligand decreasing the rational diffusion rate of the formed complex and therefore, increase the polarization degree. The degree of light polarization is then plotted against the total ligand concentration determining therefore, a K_d value.

7.3.2.1 Phage Display

Phage display is a motif scanning method, in which a single PDZ domain is screened against many motifs [9]. The first phage display was first developed in 1985 by

George Smith [145]. Years after, Schdev Sidhu further developed and improved the protocol to screen C-terminal peptides [146], which further lead to perform large-scale studies in PBM/PDZ interactions [4, 6, 147, 148].

The method uses phages, more concretely the *Escherichia virus* M13, as a vector to infect bacteria. The phages are previously prepared by inserting DNA encoding SLiMs, which will be displayed in the phage surface of p8 or p3 in the majority of times. The fully randomized peptides displayed in the phages surfaces are going to be the prey on a surface full of immobilized baits (the protein domains). Unbound phages are washed and further repeated rounds of more phages are performed in order to enrich the binding clones. The resulting binding phages are then subjected to sequence analysis (fig. Figure 7.3) [149, 150]. The strength of this technique is the low cost manufacturing price in the creation of highly diverse library (10^{10}), being the bottle-neck the sequencing of the binding clones.

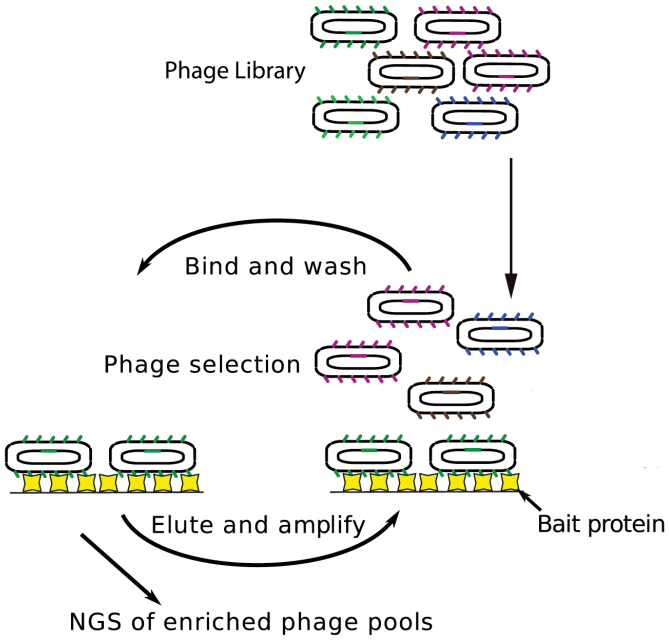


Figure 7.3: **Schematics of the Phage Display approach.** The library is incorporated to the phages. The library is screen through immobilized bait proteins for several rounds. The enriched binding phage clones are analyzed by next sequencing generation (NGS).This figure has been adapted from [6].

Tonikian *et al.* used the low cost method as an opportunity to "decodify" the amount of PBM class existing in the PDZome [4, 148]. However, according to Luck *et al.* this study got biased due the high rate of hydrophobic PBM (i.e. trypthophan rich) resulted from the Phage Display study with the premise that

most of them were not found in the human proteome [5]. Since then, efforts have been done to improve the phage display library, for instance, by enriching it in cDNA presence and open reading frames of the human proteome. The quality leap came with the Proteomic peptide-phage display (ProP-PD). ProP-Pd is a library carefully designed to display regions of the target proteome customized by oligonucleotide microarray. The library covers the intrinsically disordered regions of the human proteome using 16-mer peptides. The disordered regions in the human proteome were found using the IUPred algorithm [151, 152]. The peptides are overlapped by seven amino acids to optimize the coverage constituting 479 846 peptides derived from 18 682 proteins [153, 154].

Recently, phage display has started to move to the PTM studies (Phosphomimetic ProP-PD) [6]. However, it is not suited for the insertion of the PTM at a specific desired target site of the PBMs. Therefore, it needs to rely on in the mimication due close physic-chemical properties of other amino acids (Glutamic Acid in case of phosphorylation). Another drawback of phage display is that it cannot deliver quantitative affinities of the interactions, which can be compensated with its strength of binary mapping a given PDZ domain interactome. However, the enrichment process does not allow one to study to discriminate the non-binders from the weak binders phages clones as they will be removed during the washing step.

7.3.2.2 SPOT

SPOT array is also a technique used to study PBM/PDZ interactions with a scanning motif focus, in which a single PDZ domain is screened against many motifs [9]. It first approach is from the early ninteen's [155, 156]. The method uses, since their early stage, chemical synthesized peptides attached to a cellulose membrane. This means that PTM can be easily incorporated to the SPOT assay. Moreover, this technique can be fully automatized [157].

In a sheet of nitrocellulose membrane, many peptides are displayed in the form of ordered individual "spots", each corresponding to a particular sequence. Initially, peptides were synthetized from a C-terminal residue attached to the membrane, so that PBMs, which require a free C-terminus, could not be assayed with this technique. Boisguerin *et al.* had to improve the method by inverting the peptide synthesis to let the C-terminal part of the peptide free from the membrane.

Then, the domain of interest, fused to a detectable polypeptide tag (His-tag, MBP, GST...) is incubated together with the peptides and revealed by a chemiluminescence system using an anti-tag antibody [158].

The SPOT method is a semi-quantitative approach that provides some assessment of the binding strength [159], in particular when it is validated by another quantitative assay. Moreover, non-binding peptides (or putative weak binder peptides) are also revealed in the assay. An advantage of this method is that in the microarray, the peptides can be systematically mutated to explore the interaction map and the role of distinct amino acid at specific positions [157].

7.3.2.3 Protein Arrays

Protein arrays are a multiplex scanning method, in which multiple motifs are tested to bind multiple PDZ domains [9]. This method uses purified PDZ domains queried against fluorescent peptides. The protein of interest is immobilized on a surface and then the queried labelled peptide probed for binding.

FP is normally the method used to validate the method and discard false positives. It can detect medium to high affinity PPI due its semi-quantitative approach which has been used to produce interactomic map for ErbB receptors. [131, 137, 138]. Another advantage of this method is that it can also include PTM studies.

7.3.2.4 The holdup assay

We developed our own high-throughput screening assay, called the holdup assay [7, 31]. The holdup measures the PPI in a direct manner using a domain scanning approach, in which a motif is assayed against all the globular domains of the PDZ family [9].

The holdup assay was for the first time developed in our lab [160] and further improved together with our collaborators [7, 31]. In a summarized sentence, we could define the holdup assay as "a comparative chromatographic retention approach that quantifies the affinity-driven retention of protein on resins carrying their putative ligands" [7]. The assay has no washing steps, which both facilitates

and accelerates the procedure and provides access to equilibrium affinity information. The method uses chemical synthetized peptides fused with the mentioned carriers attached to the resin. Therefore, we can introduce PTM to our PBM allowing us to address, from the domain-motif perspective, some aspects of the "dynamic interactome" [33]. We can also measure affinities of very weak interactions, in submillimolar ranges that cannot be tackled by other methods mentioned here. This allows us to quantify the interactome in the continuum, which will display more information about the PPI network. Moreover, there is no need to use purified PDZ domain samples and we can express multi-PDZ constructs (e.g. tandems) if desired to study multivalencies.

In its automated robotic setup, the method uses 3 plates of 384 wells [7, 31]. Each plate is divided in 96 groups (four wells per group) resulting in a total of 288 groups. This amount of groups will allow us to cover the whole PDZome as each group will be used to test one PDZ interaction. The four wells are distributed in a reference plus three putative PBM (samples). The negative control will be further used as a reference data to obtain the affinities from the PBM/PDZ interactions (see chapter 8).

On the one hand, we overexpress MBP-fused PDZ domains in *E.coli* and prepare extracts where the domain concentration approximates 4 µM, which will be largely exceeded by the PBM concentration. In the theory, PBM concentration will be around 50-100μM [7]. In practice, we evaluate this concentration *a posteriori* by applying an orthogonal binding assay (FP) on a few selected interaction pairs (see chapter 8, chapter 10 and chapter 11). On the other hand, every well is filled with streptavidin resin and fully saturated with biotinylated putative peptide or reference (fig 7.4, step 1). We also add an internal control directly to the common PDZ storage tube before starting the experiment. This control is usually a soluble protein with an adequate molecular weight, avoiding in particular any overlap with the fusion MBP-PDZ protein. In our case, we use Lysozyme, which has a molecular weight of 18kDa, while the fusion MBP-PDZ protein oscillates between the 55 and 70kDa depending on the PDZ domain. The internal control will allow us to normalize the data in a further step of the method (see chapter 8).

The overexpressed lysate from *E.coli* is then incubated with the resin until binding equilibrium. At this stage, PDZ will be recruited in a variable proportion by the resin carrying the biotylinated PBM peptide as compared to the negative control. After the incubation, we perform a fast filtration and collect the flow-

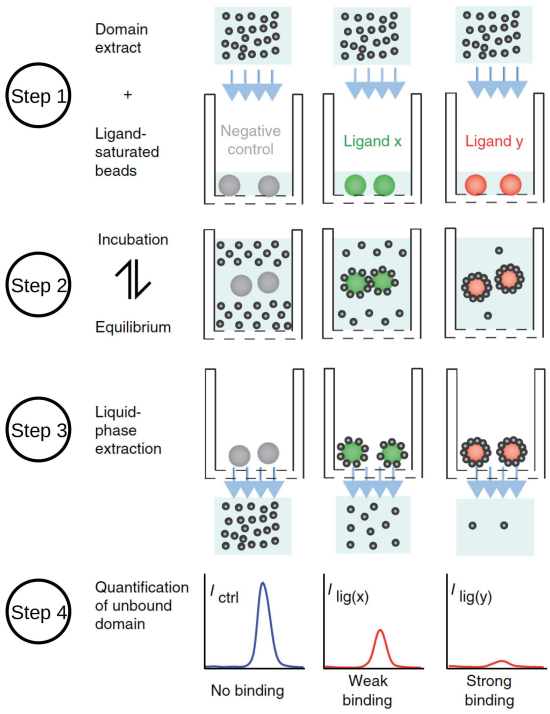


Figure 7.4: **Schematic approach of the holdup assay.** (Step 1) On one hand, we prepare overexpressed PDZ domains and on the other hand, we fully saturate resins with the negative control, ligand X and ligand Y (where ligand are meant as PBM). (Step 2) We then incubate both together until reaching equilibrium and (Step 3) we perform a fast filtration to collect the liquid phase. (Step 4) Lastly, we analyze the flow-through by microfluidic capillary electrophoresis. A decrease of the interested molecular weight (PDZ in our case), means an interaction with the resin during the incubation time, therefore an interaction. This figures has been adapted from [7].

through, since we are interested in the depletion of the MBP-PDZ fusion protein that may be captured by the biotinylated peptide attached to the resin. Therefore, the stronger the PDZ domain has bound to the PBM, the less protein we will find in the flow-through (fig 7.4, steps 2 to 4). To detect this depletion we use the *Perkin Elmer Caliper LabChip GXII* ®, a microfluidic capillary electrophoresis instrument. This instrument separates electrophoreticaly by size the proteins of the sample in an automatic way and with higher sensitivity and reproducibility. By superimposing both electropherograms, the sample and the negative control reference, we can extract the "Binding Intensities" (BI), which refers to our interaction strength quantification score, ranged from 0 to 1, allowing us to rank all the measured PBM/PDZ interaction according to their affinity. The BI is extracted using the following formula:

$$BI = \frac{I_{ctr} - I_{ref}}{I_{ctr}} \tag{7.3}$$

where I_{ctrl} refers to the intensity of the MBP-PDZ peak found in the 55 to 70 kDa range for the reference sample. I_{ref} refers to the intensity of the MBP-PDZ peak, found at the same position of the electropherogram for the PBM sample. As mentioned before, this value will range between 0 and 1, where BI $= 0$ stands for "absolutely no binding detect" (fig 7.5a) and BI $= 1$ stands for "maximal binding detect" (fig 7.5b). Vincentelli *et al.* measured the threshold value of the holdup assay by performing measurements with control resins, one bearing only biotin and the other bearing a neutral hydrophilic sequence (GSNSGNGNS) peptide. The results showed that 100% of the only biotin references and 98% of the neutral hydrophilic reference had a BI value lower than 0.2 and 0.1, respectively. Therefore, two thresholds were set: BI $= 0.2$ represents a very stringent threshold, which includes only high-confidence binding pairs, and BI $= 0.1$, which is a more relaxed threshold including additional PBM/PDZ pairs that produce a significant signal that may indicate weak binding [7]. A more detailed protocol of the holdup assay can be found in [31].

We developed in the frame of this thesis an accurate and sensitive computational protocol to perform the superimposition of the electropherograms. This leads to a better reproducibility of the BI values. Moreover, we propose an approach for converting the values of BI into K_d, which represent the most universal quantitative assessment of affinity for biological molecules. For a detailed protocol of this process, see chapter 8, chapter 10 and chapter 11.

The Binding Profiles: After having measured all individual interactions of each MBP-construct of the PDZome against a single PBM, we plot the BI values from left to right by decreasing BI intensities along the X-axis. These "PDZome-binding profiles" allow us to visualize and compare, at a glance, the specificity-promiscuity of different PBMs against the PDZome. When comparing different PBMs, we can also rank the different profiles by respecting the ranking order of one given PBM (which will be then the reference PBM). In this way, we can visualize, for each PBM profile, its degree of alignment (or similarity) with the PDZome-binding profile of the reference PBM. The more "disturbed" or "reshuffled" a PBM profile as compared to the reference one, the more divergent the prefences of that PBM

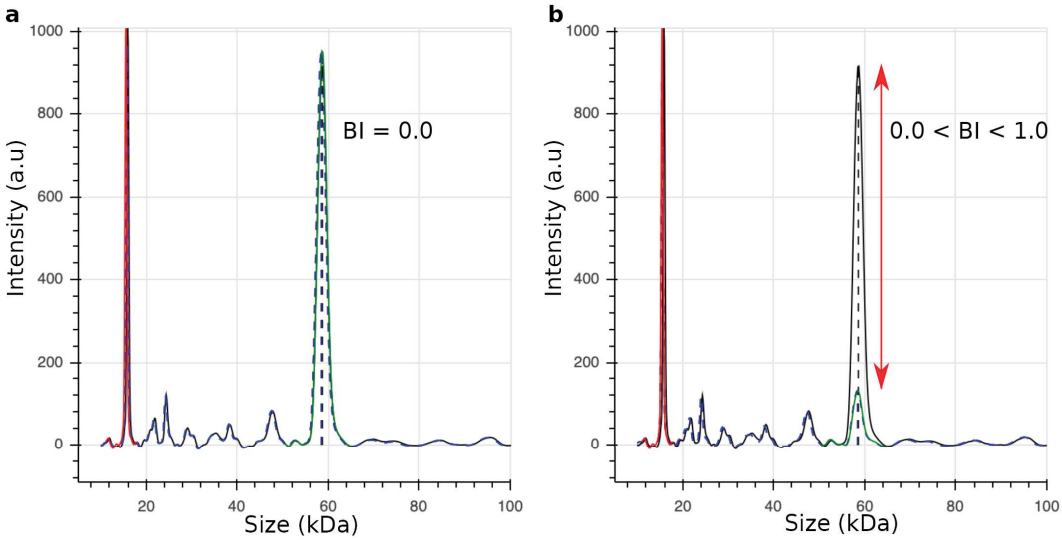


Figure 7.5: **Examples of PBM/PDZ binding cases in holdup, visualized after successful data treatment**. The peaks at 17 and 59 kDa correspond to the internal control protein (here, lysozyme), and the PDZ construct (protein of interest), respectively. Red: normalization peak. Green: peak of the PDZ construct as seen in the sample corresponding to the flow-through of the biotinylated PBM-loaded avidin resin. (a) Both electropherograms are perfectly aligned, and both PDZ peaks are undistinguishable: this PDZ construct did not detected bind to the PBM. (b) Both electropherograms are perfectly aligned, yet the PDZ peak intensity detected in the flow-through of the PBM-loaded resin has considerably decreased as compared to the PDZ peak of the flow-through of the biotin-loaded reference resin: this PDZ construct strongly bound to the PBM. Black: PDZ peak as detected in the reference, corresponding to the flow-through of the biotin-loaded avidin resin. This figure has been extracted from the presented protocol in this thesis, see chapter 8

.

from the preferences of the reference PBM (fig. 7.6).

Examples in which this representations has been applied are found in [7, 31, 33, 34] and in chapter 8, chapter 9, chapter 10 and chapter 11, which belong to the manuscripts (published or awaiting publication) I present in this thesis.

7.4 Relevance of large-scale PPI studies

Although initially the quality of the data produced in the screenings were questioned, quality standards and empirical validation ensured posterior high-quality

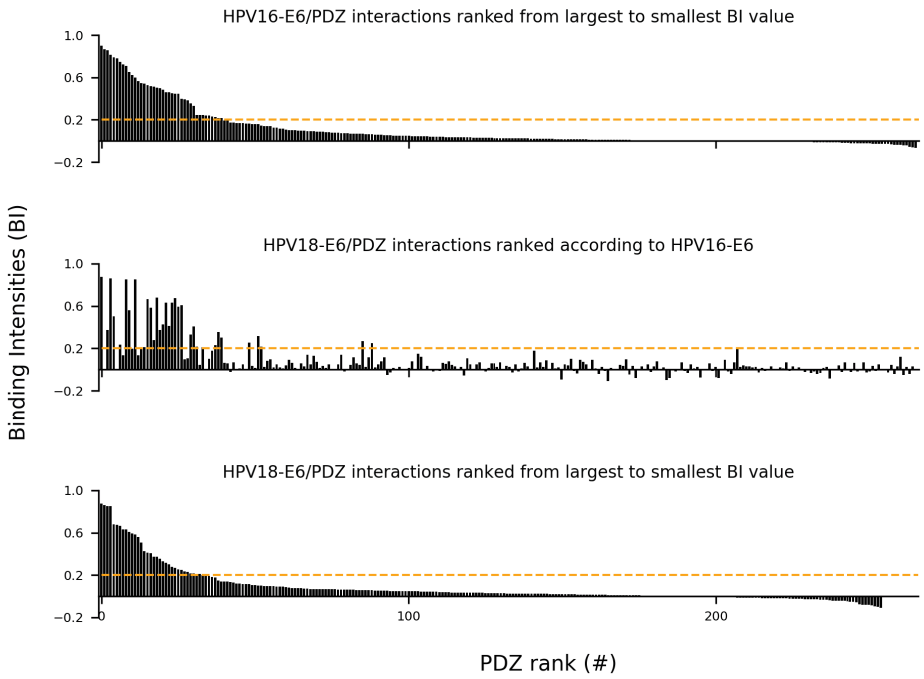


Figure 7.6: **Example of BI Profiles and its rearrangement.** The strongest PDZ binders are ranked from left to right of the plot in decreasing order along the X-axis for HPV16-E6 and HPV18-E6 (top and bottom panel). The curvature of the profile shows the specificity of the PBM/PDZ binding: the steeper the slope of the BI profile, the more specific is the PBM. Threshold for the confidence value of binding is set at 0.2 (orange dotted line). In the middle panel, PDZ domains are ranked on the basis of HPV16-E6 for HPV18-E6 to show the rearrangement of the binding.

data generation [161, 162]. The higher the quality of data, the more accurate the interactome description and predictions that can be developed. This helps not only in showing biologist which future networks to study but also to recover unmapped interactions at the moment [133]. Furthermore, through the integration with other "omics" data, interactomic datasets have proven highly valuable towards applications in different areas of clinical importance, such as understanding the multiple genotype-phenotype relationship in diseases [161–164].

Part III

Methodology

Chapter 8

A computational protocol to analyze PDZ-PBM affinity data obtained by high-throughput holdup assay.

Status: First author article in press in an issue of *Methods in Molecular Biology*.

8.1 Summary

We searched to optimize the transformation of capillary electrophoresis data that are produced as an output of the holdup assay, into accurate and reliable binding profiles. To this aim, we developed a series of steps, or algorithm, which were later wrote as a suite of computational routines, to perform an accurate reading, superimposition and BI extraction from the electropherograms:

- Read the Caliper data.
- Remove the noise by subtracting the baseline from the electropherogram raw data.
- Normalize the peak intensity using the introduced internal control.
- Adjust the X-axis of the sample electropherogram to the reference one by performing a linear transformation (translation and dilation).

- Perform an optional "secondary" correction of the signal intensities.
- Determine the affinity strength. In this assay, this is deduced from superimposing the electropherograms and calculating the depletion of the MBP-PDZ peak. The peak depletion can be quantified from 0 to 1 and we called it "Binding Intensity" (BI).

Moreover, we realized two main problems when delivering our data to the scientific community:

- Scientific community does not know the meaning of BI as it is not an international unit of affinity measurement.
- Since the PBM concentration ([PBM]) does not remain equal from one sample to the other due own peptide properties such as size, bulkiness, hydrophobicity etc.., the comparison of BI between different PBM might not be the most accurate.

To solve this problem, we decided to transform BI into steady-state affinity values using the following equation:

$$K_d = \frac{([PDZ] - BI \cdot [PDZ]) \cdot ([PBM] - BI \cdot [PDZ])}{BI \cdot [PDZ]} \tag{8.1}$$

where [PDZ] is the total PDZ concentration (set to 4 µM in usual cases in our assays) and [PBM] is the total peptide concentration. Since the [PBM] depends on each PBM as already mentioned, we cannot estimate its concentration in the resin. The conversion will be therefore impossible to be made directly.

In order to perform this conversion, we use affinity values obtained by an orthogonal experimental approach (competitive fluorescence polarization, or FP) or some of the PBM/PDZ interaction pairs measured in the profile. These affinity values are exploited to derive the [PBM] concentration in the holdup assay. In practice, we measure by competitive FP up to 20 PBM/PDZ interaction detected in the holdup assay (in treatcions with a BI>0.2 according to [7]). We then calculate the peptide concentration based on the $BI - K_d$ pairs obtained by both, holdup and FP. We lastly propose some new representations to show the obtained data by the holdup assay.

Contribution: I wrote the presented protocol (draft and final version) after successfully analyzing over 20 000 interaction with the software presented in chapter 15. Although the software was already in a pre-working phase when I joined the group, I participated to its improvement by proposing new functionalities and removing some not working features. I designed most of the representations. I create all the final figures related to the manuscript.

Besides the work in the holdup data analysis presented here, a lot of extra pipelines and infrastructure was performed. For instance, all the data were collected and transferred into a database allowing to retrieve and compare many different aspects. This showed us were the main problems to read the data came from or how long the PDZ domains were kept in good conditions, in the frozen bank of PDZ extracts, by comparing PDZ intensities between the 20 000 interactions. Additionally, we tried to relate sequences with the bindings. However, we still need more data to succeed in such purpose.

8.2 Chapter in Methods in Molecular Biology

A computational protocol to analyze PDZ/PBM affinity data obtained by high-throughput holdup assay.

Running title: A computational protocol to analyze PDZ/PBM data

Pau Jané[1], Lionel Chiron[2], Goran Bich[1], Gilles Travé[1] and Yves Nominé[1]

1. (Equipe labelisée Ligue, 2015) Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258 / CNRS UMR 7104 / Université de Strasbourg, 1 rue Laurent Fries, BP 10142, 67404 Illkirch, France.

2. CASC4DE, Le Lodge 20 av. du Neuhof, 67100, Strasbourg, France.

Abstract

The holdup assay is an automated high-throughput comparative chromatographic retention approach that allows to measure quantitative Binding Intensities (BI) for a large number of domain-motif pairs and deduce equilibrium binding affinity constants. We routinely apply this approach to obtain quantitative binding specificity profiles of particular PDZ-Binding Motifs (PBMs) towards the full library of known human PDZ domains (the PDZome). The quality of the electropherograms extracted from the capillary electrophoresis instrument at the final step of the holdup assay may vary, influencing the accuracy and reproducibility of the measurement. By using bioinformatic tools, we can solve these issues to extract more reliable BIs by means of a better superimposition of the electropherograms. The protocol presented in this chapter describes the main principles and strategies of our curated method to process holdup data and new ways to plot and compare the BIs for the PBM / PDZ interactions. For this particular protocol, all the necessary computing commands are freely available in open Python packages.

Keywords: holdup assay, PDZ/PBM interaction, computational approach, processing accuracy, electropherogram superimposition

1. Introduction

An important subset of protein-protein interactions are mediated by globular protein domains interacting with Short conserved Linear Motifs (SLiMs), mostly belonging to intrinsically unfolded regions of the proteome. Many high-throughput interatomic data allow to describe protein-SLiMs interaction only in a binary way (“binds” or “does not bind”) [1-4]. The quantification of the binding affinities would lead to a better understanding of the specificities and hierarchies of interactions.

In this regard, the holdup assay [5, 6] is well suited to extract quantitative affinity information for an entire network of domains-SLiMs interactions, such as the 266 known human PDZ domains and their target motifs, called PDZ-Binding Motifs (PBM). The holdup assay is a comparative chromatographic retention approach based on the reversible binding of constructs to affinity resins avoiding protein purification and washing steps. In this assay, soluble lysates of recombinant proteins overexpressed in bacteria (for instance: PDZ domain) are incubated with avidin beads, previously fully saturated by either a biotinylated peptide (for instance: PBM) or a reference (biotin). Once equilibrium is achieved, the resin-liquid mixture is subjected to a fast filtration. The resulting flow-through contains the remaining non-bound recombinant protein, which is further quantified using a capillary electrophoresis instrument. The electropherogram obtained for the PBM / PDZ pair is then superimposed and compared with the one for the reference with biotin / PDZ. A depletion of the recombinant protein of interest observed in the flow-through of the resin containing the peptide, as

compared to the reference flow-through, indicates that a protein / peptide binding event has occurred. The stronger the depletion of the recombinant protein of interest, as compared to the reference, the stronger the protein / peptide binding interaction. Ultimately, the binding intensity (BI), related to the interaction strength, is determined, from which in turn the equilibrium dissociation affinity constant (K_D) can be deduced.

A proper estimate of the BI relies on the precise and accurate pairwise comparison of electropherograms. When dealing with the same protein extract, the pattern of two electropherograms is highly conserved in most parts of the graphs, except for the PDZ peak which may have partly disappeared from the extract by being specifically retained on the PBM-coated resin. This makes possible to compare them in an automated way, using computational methods. However, capillary electrophoresis migration may be subject to variations linked to slight differences of input volume, migration in the capillary, buffer, well position or measurement temperature, possibly altering the migration of molecules and therefore their apparent sizes [7, 8]. As a consequence, slight fluctuations of the peaks on the X- (migration) and/or Y- (sensitivity) axis and baseline perturbation may alter the reproducibility of the electropherograms and subsequently the accuracy of BI values (Fig. 1a). For this reason, a strict comparison of two different electropherograms is not always easy to perform, even when data are recorded in the same experimental conditions. Therefore, the use of semi-automated methods to correct and compare the electropherograms, can dramatically improve their superimposition (Fig. 1b).

Here we present the main lines of a computational protocol to deal with these issues using commands available in the free Python Spike package. The accuracy of the results

dramatically increases when one uses this automated approach to improve superimposition of the data. This can be achieved by applying up to five consecutive processing steps to the electropherograms of both, the reference and the sample (Fig. 2):

1. Baseline subtraction in the raw data electropherograms.
2. Normalization of the peak intensities by the use of the added internal control.
3. X-axis adjustment of the sample electropherogram to the reference one by performing a linear transformation (translation and dilation).
4. "Secondary" correction of the signal intensities (optional).
5. BI determination deduced from superimposing the electropherograms and calculating the depletion of the peak for the protein of interest.

The standardization of electropherogram processing leads ultimately to a better comparison and reproducibility between different holdup assays, no matter if performed in the same laboratory or not. A nice example of the power of applying these strategies can be found in a recent publication, in which we achieved highly accurate and reproducible results [9].

Although the holdup assay is in principle adaptable to study diverse binding protein-ligand systems, here we will focus on the human PDZ domains (the PDZome) / PBMs interactions.

2. **Materials**

2.1 Capillary electrophoresis instrument's software

Every capillary electrophoresis instrument has its specific software to retrieve the data.

2.2 Computer

The use of a programming language is necessary to process all the data. For this project, Python has been chosen.

2.3 SPIKE package

A Python package available in https://spikedoc.bitbucket.io [10, 11].

3. Methods

All mentioned steps were performed with in-house Python scripts using different packages, including SPIKE. This protocol is meant to be applied to both electropherograms recorded for the reference and the sample. The protein extract loaded on the reference and the sample resins should come from the same expression source and in practice from the same storage tube. This protocol also requires an internal control added directly to the common storage tube, before starting the experiment. This control is usually a soluble protein with an adequate molecular weight, avoiding in particular any overlap with the molecular weight of the protein of interest. The peak intensity of the internal control protein (hereafter referred to as the "internal control peak") is used for normalization so that the two electropherograms obtained from the same source tube can be directly compared with the highest accuracy. Note that it is possible to include several internal control proteins with distinct molecular weights in the same sample. This provides the user with a "plan B", highly useful when the peak intensity of the regular internal control is too weak or too intense as compared to the peak of the protein of interest (hereafter referred to as the "PDZ peak"), or when the regular internal control peak, but not the rest of the electropherogram, has been altered due to a

migration defect. If any of the steps fails, the user can try to restart from point 2 in the "Transforming the input data" section by varying the default parameters.

3.1 Input data extraction from the quantitative capillary electrophoresis instrument

1. For each plate analyzed by holdup assay, open the software of the electrophoresis instrument and make sure that all the ladders and samples are well aligned (see **Note 1**).

2. Give the proper names to the samples and export the electropherogram's raw data including molecular weights (see **Note 2**).

3.2 Transforming the input data

1. Among other information, the extracted raw data files contain 3 columns with the migration time, the converted molecular weights and the fluorescence signal. Adjust their format to make them easy to read in order to manage the data for the following steps. A full access to the data allows to plot them. In particular, it is possible to replot the results of each individual in order to give a better understanding of the process and to detect any problem that would occur during the procedure.

2. Perform the baseline correction for both the reference and the sample electropherograms using the BC module of the SPIKE package [10, 11] (Fig. 3a). This will adjust the real intensity of each electropherogram by subtracting the background

signal. The user may need to adjust the parameters, for instance if the overexpressed domain migrates at a different size or with distinct expression level than the MBP-PDZ domain fusion used in our own assay.

3. Perform the normalization by selecting the internal control peak for both the reference and the sample (Fig. 3b). This allows to correct for potential variations of overall protein concentration between the two electropherograms, due to slight sensitivity changes or volume variations that may occur after the binding step, in particular when reagents are added to the samples prior to the caliper measurements.

4. Adjust the X-axis of the sample electropherogram to better match the reference electropherogram using a linear transformation (shift and dilation/contraction). For this purpose, the difference of signal between the two electropherograms is minimized using a least square approach by focusing on a molecular weight window that contains the peak of the protein of interest.

5. Superimpose the two electropherograms. When everything works properly, two different scenarios may occur: i) both electropherograms are strictly identical, so that no binding is observed (BI ~ 0.0; See below) (Fig. 4a); ii) only the peak corresponding to the molecular weight of the overexpressed protein has significantly decreased (Fig. 4b) (See **Note 3**), indicative of an interaction between the PDZ and the PBM. Noteworthy enough, the first scenario (no detectable binding) is the most likely to occur (in our experience, 80% to 90% of the cases).

6. No matter if the superimposition worked properly or not, the following part of the protocol details how to check the accuracy of your results.

3.3 Checking the data quality

In order to assay the overall quality of the electropherograms, we determined empirically several quality criteria that can be fulfilled:

1. Quality of the normalization peak: The intensity of the normalization peak should be in the same order of magnitude as for the PDZ peak observed in the reference.

2. Quality of the detection signal: Both normalization and PDZ peaks should be in the linear range of the capillary electrophoresis instrument.

3. Quality of the PDZ expression: The ratio of the PDZ intensity by the average of signal within the crude extract range (typically 20-40 kDa) should be as high as possible, and never below 1.5.

4. Quality of the PDZ construct dilution: We experimentally observed a threshold intensity of the PDZ peak (~200 units of Fluorescence) below which the processing might be critical.

5. Quality of the PDZ construct: The difference between experimental (from Caliper data after conversion) and theoretical (according to the sequence of the construct) molecular weights should be as low as possible, and not exceed 5 kDa (see **Note 4**).

6. Quality of the electropherogram superimposition on X-axis: The linear transformation used to match the X-axis of the two electropherograms should be as neutral as possible (i.e. dilation coefficient and shift as close as possible to 1.0 and 0.0, respectively).

7. Quality of the electropherogram comparison on Y-axis*:* It is also possible to improve the accuracy of the signal comparison when the overexpressed sample is not purified and contains peaks from the bacterial crude extract (see **Note 5**). If used, this optional processing step should lead to a correction factor as close to 1.0 as possible.

3.4 Extracting BIs and other data

1. Extract the BI of the peak of the protein of interest. After the superimposition, the Binding Intensity (BI) is calculated by subtracting the peak intensity of the sample (I_{lig}) from the PDZ peak intensity of the reference (I_{ref}), and then dividing by I_{ref} (Eq. 1). With this expression, 1.0 is the maximal value that can be obtained for BI. As previously shown [6], BI > 0.2 represents a high-confidence binding event. Below this threshold, 0.1 < BI < 0.2 may still represent a significant binding event, depending on data quality. For BI < 0.1, we consider that no binding has been detected.

$$BI = \frac{I_{ref} - I_{lig}}{I_{ref}}$$ Equation.1

2. The electropherograms contain a lot of information besides the visual part. In addition to the BI, it is worth to save additional values such as peak intensities and positions, correction factors, baseline levels, … (see **Note 6**). This might help for further analysis and can provide useful answers to several questions (for instance, is the overexpressed domain stable and expressed always at the same molecular weight?).

3.5 Storing and plotting the data

1. To combine all the data obtained with different plates, export them into tables and create a repository data base, for instance by using the SQLite3 Python engine. This type of interface allows to “ask queries” and retrieves all the needed data from this database at any moment for further comparisons and plots.

2. One way to plot the data from the database is to extract a Binding Profile (Fig. 5a). The Binding Profile for a given PBM is a bar plot displaying Binding Intensity (BI) values in which all the PDZ domains are ranked from the strongest to the weakest binder. This plotting mode captures the specificity of the recognition, imbedded in the curvature of the profile.

3. Another way to plot the data is to create a circle for each PDZ / PBM pair tested whose the diameter is proportional to the BI value, and to stack all the generated circles in the lower part of the largest one. The binding strengths, as well as the specificity, can then be easily appreciated by comparing the differences of the diameters (Fig. 5b).

4. The same type of plots can be generated by centering all the circles at the origin (Fig. 5c).

5. The usual approach to compare two BI profiles consists of plotting the PDZome binding specificity profiles of the two PBMs, both ranked from highest to smallest BI value, and adding in between the two profiles an extra plot showing the BI values of the second PBM sorted according to the ranking of the first PBM [6, 9] (Fig. 6a).

6. Alternatively, the BI values can also be transformed into -log(K_D), using the following equation (Eq. 2) (see **Note 7**):

$$-\log(K_D) = -\log\left(\frac{([PDZ_{tot}]-BI\cdot[PDZ_{tot}])\cdot([PBM_{tot}]-BI\cdot[PDZ_{tot}])}{BI\cdot[PDZ_{tot}]}\right) \qquad \text{Equation. 2}$$

For a given temperature, these values are proportional to ΔG, allowing to rank all the affinity values in a continuous heatmap (like in the binding profiles), and making it easier for experimentalists to compare the affinities using scaled values widely-known among all. The same kind of rearrangement as in figure 6a can be done with the colored heatmaps according to enthalpy changes (Fig. 6b).

7. Lastly, the circle plots shown in figure 5c can be divided into as many sectors as samples, allowing to compare several samples at one glance (Fig. 6c).

4 Notes

1. Electropherograms presented in this chapter are obtained using the Caliper GXII LabChip system (PerkinElmer). In some cases the operating system of the capillary instrument fails to detect the ladders at the proper molecular weights. This issue leads to an inconsistent X-axis, making very difficult to superimpose two electropherograms because of peak misalignment (Fig. 7). The problem is often linked to the appearance of some not-expected extra peaks (for instance contaminants or detection artifacts). To solve this problem, on the LabChip software, exclude the additional peaks and if needed, include the proper ones according to the molecular weights of the used ladders. A similar process may have to be applied to the sample peaks. Such a step often helps to optimize the data superimposition.

2. Exporting the electropherograms. In the Caliper GXII LabChip software, go to “Tools” and select “Sample Name Editor” to rename the samples. Select the data to be exported and click “Export” in the file tab menu. A window will pop up: click on “Raw Data” and “Include Size Data” to export all the necessary points of the electropherograms.

3. We consider an intensity decrease as significant when the amplitude change is much larger (at least 3 to 5 times more) than the variability observed in zones presenting constant signal (for instance around 100 kDa).

4. We experienced once a global shift of the molecular weights for all the ladders and samples. As far as it affects all the samples equally, this molecular weight shift will not affect the BI calculation while using this protocol.

5. The crude extract peaks from the bacterial expression can also help to perform a secondary correction in addition to the internal control peaks, since the intensities of those bacterial protein peaks should be the same in the two electropherograms to be compared. If one wishes to use the crude extract peaks as a secondary correction option, a range excluding the normalization peak and the peak of interest should be defined. We have noticed that the Y-axis superimposition can often be further improved by applying this secondary correction (Fig. 8).

6. We keep track of the binding strengths, the intensities and positions of the internal controls, the intensities and positions of the PDZ domains, the range of the crude extract and its intensity, the shift/dilation coefficients in the X-axis, and the correction factors when applying the secondary correction. We store all these data in CSV files and combine them in a unique Data Base using SQLite3. We then retrieve the data by asking "specific queries" and manage the large amount of data with the Pandas package provided by Python.

7. All the concentrations in Eq. 2 are known except that of the PBM. For this purpose, affinities of several PDZ domains for a given PBM are measured by fluorescence polarization assay and subsequently used to estimate the concentration of that PBM.

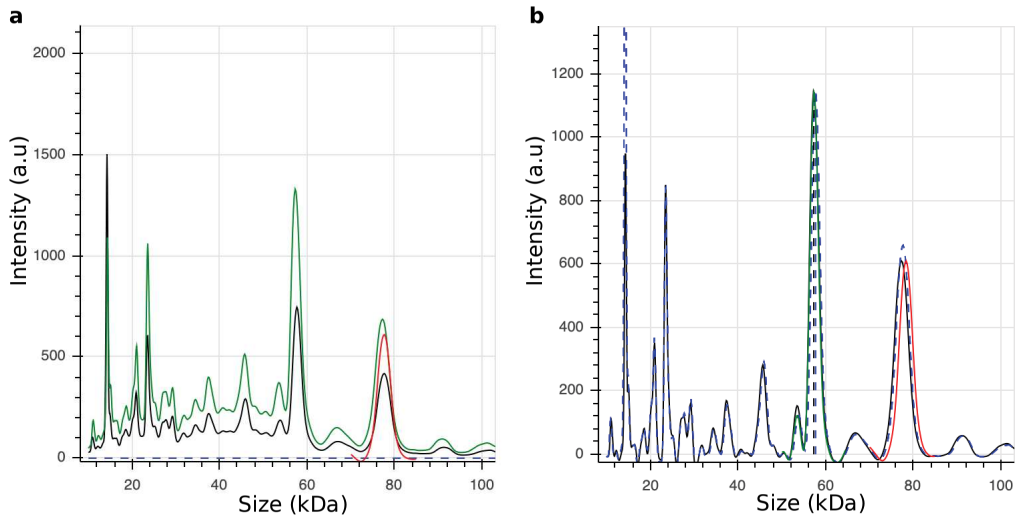
Acknowledgements

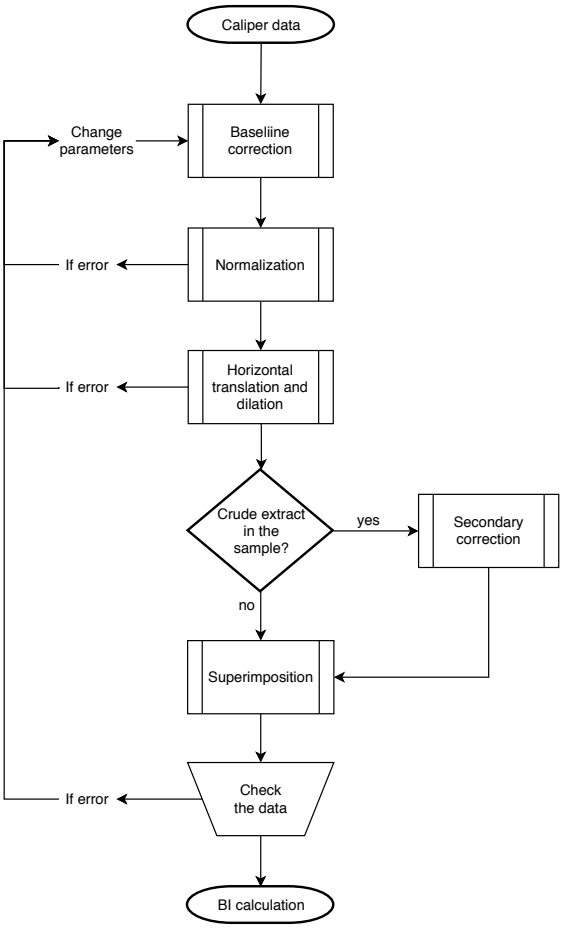
This work received institutional support from Centre National de la Recherche Scientifique (CNRS), Université de Strasbourg, Institut National de la Santé et de la Recherche Médicale (INSERM) and Région Alsace. The work was supported by funding from the European Union´s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 675341, by the Ligue contre le cancer (équipe labellisée 2015), by the National Institutes of Health (Grant R01CA134737), and by the French Infrastructure for Integrated Structural Biology (FRISBI).

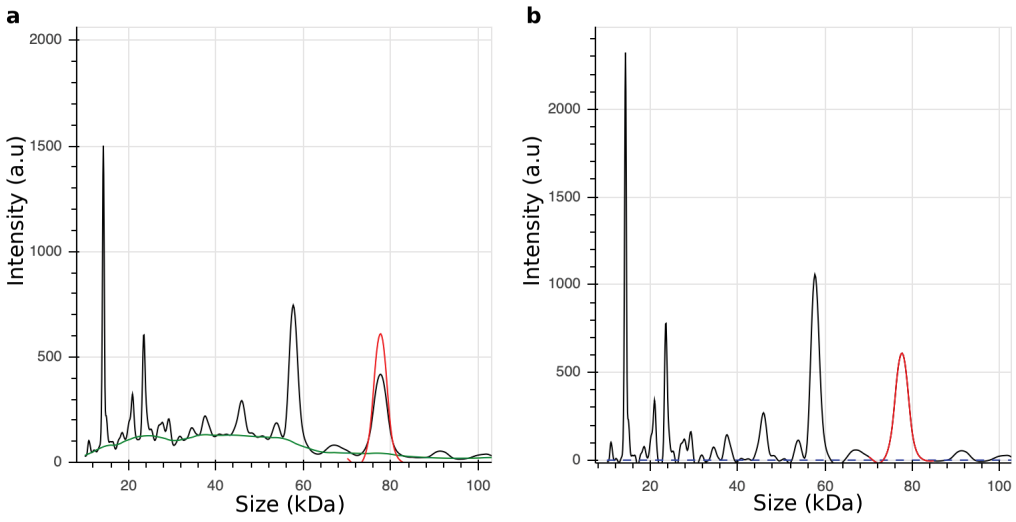
References

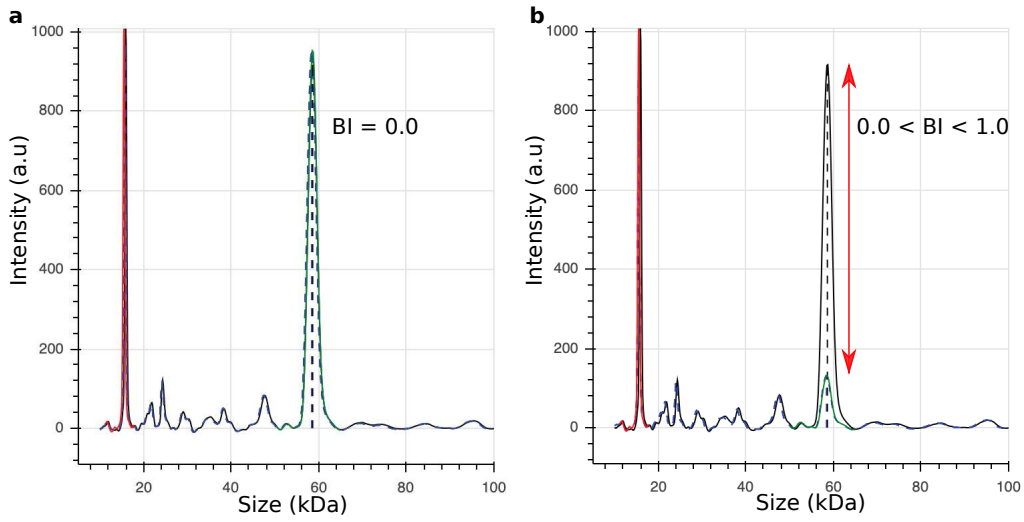
1. Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, Reva B, et al (2008) A specificity map for the PDZ domain family. PloS biology 6:e239
2. Luck K & Travé G (2011) Phage display can select over-hydrophobic sequences that may impair prediction of natural domain–peptide interactions. Bioinformatics 27:899–902
3. Smith CA, Kortemme T (2010) Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. J Mol Biol 402:460–474
4. Jadwin JA, Ogiue-Ikeda M, Machida K (2012) The application of modular protein domains in proteomics. FEBS Letters. 586:2586-2596

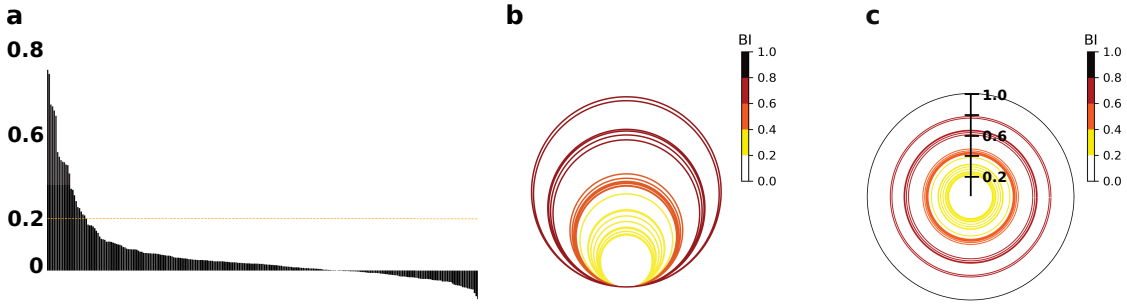
5. Charbonnier S, Zanier K, Masson M, Travé G (2006) Capturing protein-protein complexes at equilibrium: the holdup comparative chromatographic retention assay. Protein Expr Purif 50:89-101
6. Vincentelli R, Luck K, Poirson J, Polanowska J, Abdat J, Blémont M, et al (2015) Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. Nat Meth 12:787–793
7. Altria KD, Fabre H (1995) Approaches to optimisation of precision in capillary electrophoresis. Chromatographia 40:313–320
8. Ross GA (1997) Precision and quantitation in capillary electrophoresis. In: Shintani H, Polonský J Handbook of Capillary Electrophoresis Applications Springer, Dordrecht; p. 41-55
9. Gógl G, Biri-Kovács B, Durbesson F, Jane P, Nomine Y, et al (2019) Rewiring of RSK–PDZ interactome by linear motif phosphorylation. J Mol Biol 431:1234-1249
10. Tramesel D, Catherinot V, Delsuc MA (2007) Modeling of NMR processing, toward efficient unattended processing of NMR experiments. J Magn Resonance 188:56-67
11. van Agthoven MA, Chiron L, Coutouly MA, Delsuc MA, Rolando C (2012) Two-dimensional ECD FT-ICR mass spectrometry of peptides and glycopeptides. Anal Chem 84:5589-5595

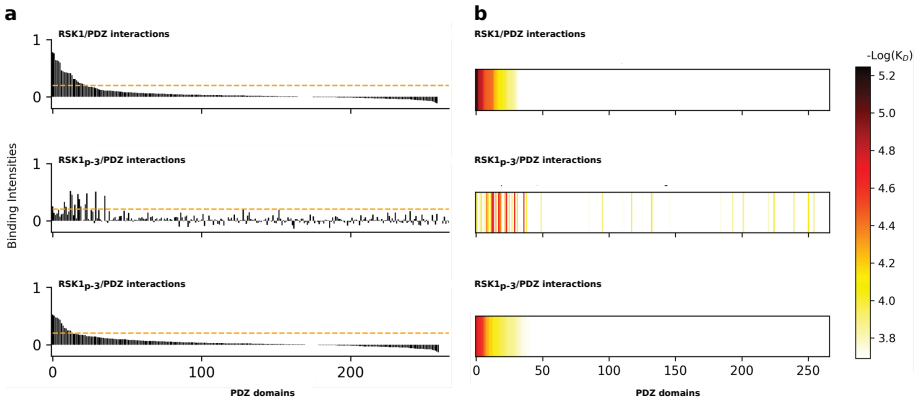

a
Intensity (a.u)
2000
1500
1000
500
0
20
40
60
80
100
Size (kDa)
b
Intensity (a.u)
1200
1000
800
600
400
200
0
20
40
60
80
100
Size (kDa)

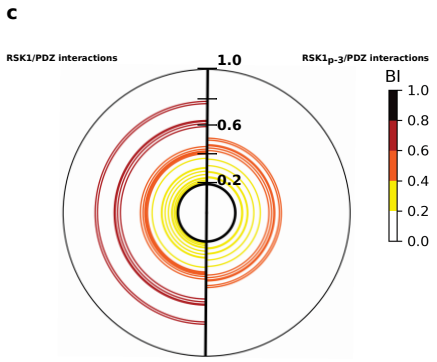

Caliper data
Change parameters
Baseliine correction
If error
Normalization
If error
Horizontal translation and dilation
Crude extract in the sample?
yes
Secondary correction
no
Superimposition
If error
Check the data
BI calculation

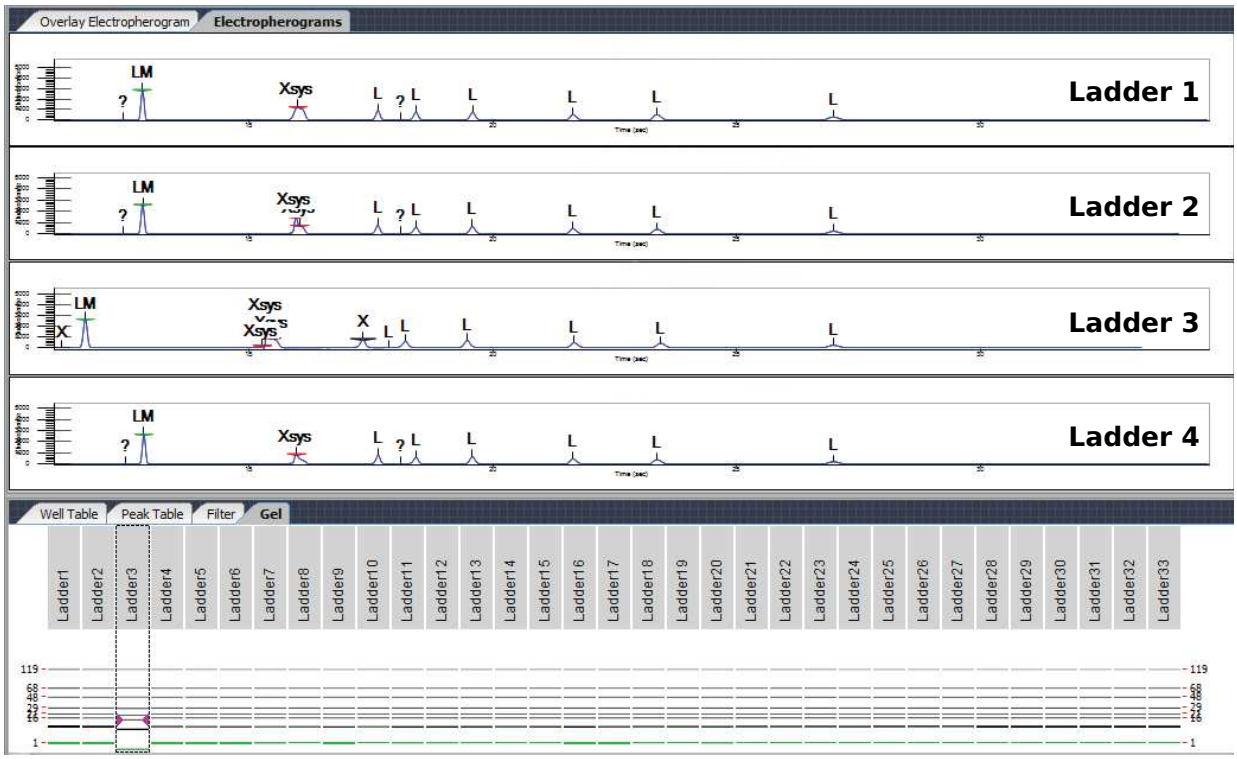

a
b
Intensity (a.u)
Size (kDa)
0
500
1000
1500
2000
20
40
60
80
100

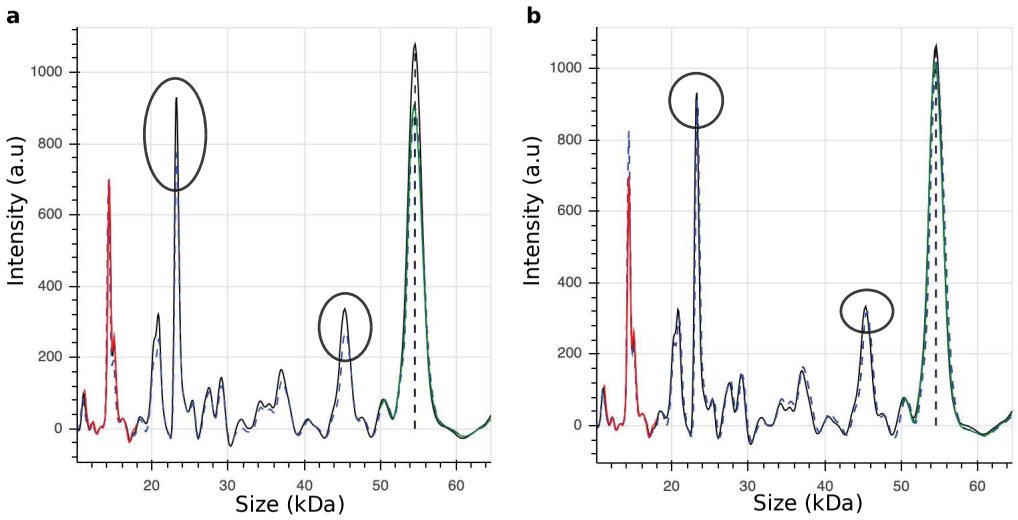

a
b
Intensity (a.u)
Size (kDa)
BI = 0.0
0.0 < BI < 1.0
0
200
400
600
800
1000
20
40
60
80
100


a
0.8
0.6
0.2
0
b
BI
1.0
0.8
0.6
0.4
0.2
0.0
c
BI
1.0
0.8
0.6
0.4
0.2
0.0
1.0
0.6
0.2


a
RSK1/PDZ interactions
RSK1p-3/PDZ interactions
RSK1p-3/PDZ interactions
Binding Intensities
1
0
0
100
200
PDZ domains
b
RSK1/PDZ interactions
RSK1p-3/PDZ interactions
RSK1p-3/PDZ interactions
-Log(KD)
5.2
5.0
4.8
4.6
4.4
4.2
4.0
3.8
0
50
100
150
200
250
PDZ domains


c
RSK1/PDZ interactions
RSK1p-3/PDZ interactions
1.0
0.6
0.2
BI
1.0
0.8
0.6
0.4
0.2
0.0


Overlay Electropherogram
Electropherograms
LM
Xsys
L
Ladder 1
Ladder 2
Ladder 3
Ladder 4
Well Table
Peak Table
Filter
Gel
Ladder1
Ladder2
Ladder3
Ladder4
Ladder5
Ladder6
Ladder7
Ladder8
Ladder9
Ladder10
Ladder11
Ladder12
Ladder13
Ladder14
Ladder15
Ladder16
Ladder17
Ladder18
Ladder19
Ladder20
Ladder21
Ladder22
Ladder23
Ladder24
Ladder25
Ladder26
Ladder27
Ladder28
Ladder29
Ladder30
Ladder31
Ladder32
Ladder33
119
1


a
b
Intensity (a.u)
Size (kDa)
0
200
400
600
800
1000
20
30
40
50
60

Part IV

Results

Chapter 9

Rewiring of RSK–PDZ Interactome by Linear Motif Phosphorylation

Status: Published in *Journal of Molecular Biology* as 4th author.

9.1 Summary

Background: Ribosomal S6 kinase 1 (RSK1) harbors a class C-terminus PBM with a role in cancer. It is related to the MAPK pathway being a strong partner of ERK. Its complex start with the phosphorylation of ERK which leads also to a phosphorylation in the PBM of RSK1.

Method: In order to see the impact of PTM in PBM, we needed to measure quantitatively RSK1 wild-type and RSK1 phosphorylated at p-3 against the PDZome. Therefore, we used the holdup assay. Furthermore, luciferase complementary assays in HEK293T cells were done to test full length RSK1/PDZ interactions.

Results: We obtained the binding profiles for both, RSK1 and RSK1 phosphorylated at p-3. We performed the rearrangement of the binding profiles for phosphorylated RSK1 towards the PDZome identifying some sequence preferences in the 2β-strand for the new PDZ domain partners.

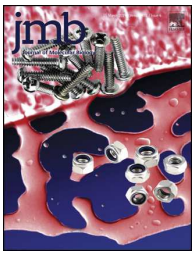
Conclusions: With the acquisition of all the RSK1/PDZ pair interactions, we could demonstrate the fact that biology is not governed by binary interactions

("binds" or "does not binds"), in which phosphorylation activates or deactivates interactions, but rather a continuum in the affinity scale.

Contribution: I performed the data curation of the elecropherograms in the holdup assay using the software presented in chapter 15. I processed the data up to three times and check individually them one by one. The curated electropherogramm in the supplementary figure has been provided by the same software used for the data curation.

Article

9.2 Publication in Journal of Molecular Biology



Rewiring of RSK–PDZ Interactome by Linear Motif Phosphorylation

Gergő Gógl [1], Beáta Biri-Kovács [1], Fabien Durbesson [2], Pau Jane [3], Yves Nomine [3], Camille Kostmann [3], Viktória Bilics [1], Márton Simon [1], Attila Reményi [4], Renaud Vincentelli [2], Gilles Trave [3] and László Nyitray [1]

1 - Department of Biochemistry, *ELTE Eötvös Loránd University, Budapest, Hungary*
2 - Unite Mixte de Recherche (UMR) 7257, *Centre National de la Recherche Scientifique (CNRS) Aix-Marseille Universite, Architecture et Fonction des Macromolécules Biologiques (AFMB), Marseille, France*
3 - Equipe Labellisee Ligue 2015, *Department of Integrated Structural Biology, Institut de Genetique et de Biologie Moleculaire et Cellulaire (IGBMC), INSERM U1258/CNRS UMR 7104/Universite de Strasbourg, 1 rue Laurent Fries, BP 10142, F-67404 Illkirch, France*
4 - Institute of Enzymology, *Research Center for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary*

Correspondence to Gilles Trave and László Nyitray: *trave@esbs.u-strasbg.fr, nyitray@elte.hu*
https://doi.org/10.1016/j.jmb.2019.01.038
Edited by Mingjie Zhang

Abstract

Phosphorylation of short linear peptide motifs is a widespread process for the dynamic regulation of protein–protein interactions. However, the global impact of phosphorylation events on the protein–protein interactome is rarely addressed. The disordered C-terminal tail of ribosomal S6 kinase 1 (RSK1) binds to PDZ domain-containing scaffold proteins, and it harbors a phosphorylatable PDZ-binding motif (PBM) responsive to epidermal growth factor stimulation. Here, we examined binding of two versions of the RSK1 PBM, either phosphorylated or unphosphorylated at position −3, to almost all (95%) of the 266 PDZ domains of the human proteome. PBM phosphorylation dramatically altered the PDZ domain-binding landscape of RSK1, by strengthening or weakening numerous interactions to various degrees. The RSK–PDZome interactome analyzed in this study reveals how linear motif-based phospho-switches convey stimulus-dependent changes in the context of related network components.

Background

Protein–protein interactions form a functional network, the interactome, which can be dynamically regulated by the phosphorylation of network components on disordered protein regions [1]. These so-called linear motifs most often bind to structured domains, such as (PSD95/DLG1/ZO-1) PDZ domains. PDZ domains belong to one of the most common families of globular domains, with 266 members in the human proteome [2]. They recognize short linear motifs called PDZ-binding motifs (PBMs) at the extreme C-terminus of their target proteins (canonical PBMs) or within internal regions (non-canonical PBMs). Canonical PBMs systematically contain a hydrophobic residue (most frequently Val or Leu) at their C-terminus (numbered as position 0) and are classified in three main classes based on the residue at minus 2 position (Ser/Thr in the most common class 1, hydrophobic in class 2 and acidic in class 3) [3]. In principle, the general consensus sequence determining a PBM allows for the presence of potentially phosphorylatable residues at any positions except the hydrophobic C-terminal position [4].

PDZ–PBM interactions are involved in various cellular processes and are especially common in intracellular signaling pathways. For example, all isoforms of the ribosomal S6 kinase (RSK) of the MAPK pathway contain a functional class 1 PBM [5]. RSK has an emerging role in multiple cancer types such as glioblastoma or melanoma [6,7]. Upon mitogenic stimulation, a series of phosphorylation events leads to the activation of the MAP kinase ERK1/2 [8]. RSK is one of the strongest interaction partners of ERK (compared to other docking motif facilitated MAPK interactions), and its complex activation mechanism is

also initiated by ERK phosphorylation (Fig. 1A) [9–11]. The C-terminal tail of RSK is a multifunctional linear motif as it contains partially overlapping binding sites for ERK, S100B, a tyrosine kinase, phosphatase(s), and PDZ domains [12–14]. In addition, activated RSK will autophosphorylate its own PBM within its intrinsically disordered tail, which will probably affect all of these interactions [15]. The RSK1 PBM contains three potential autophosphorylation sites, while other isoforms contain only two (Fig. 1B). Among these, Ser732 is found at the minus 3 position [16,17]. Thomas *et al.* [5] observed no change with RSK1/2 phosphomimics (at −3) in their interaction with MAGI1, SHANK1 or GRIP1, and they suggested that both inactive and active RSKs likely bind to PDZ domain proteins. Similarly, our recent work showed that phosphorylation of RSK1 only mildly changed the interaction with MAGI1 [15]. In contrast, a recent publication revealed that the phosphorylation (or a phosphomimetic mutation) at the analogous site triggered the association between RSK1/3 and the PDZ domain of SCRIBBLE and abolished the interaction between RSK3 and the PDZ domain of SHANK1 [18]. These results indicated that RSK activation might induce a complex reshuffling of its PDZ domain mediated interactome.

In order to elucidate the impact of phosphorylation of a given PBM, binding to all of its putative partners in the human proteome needs to be measured, quantitatively. In addition, *in vitro* observed changes need to be validated in cell-based assays. To address these challenges, we applied here a recently developed high-throughput approach [19] to measure the individual binding affinities of the 266 known human PDZ domains for both the unphosphorylated and phosphorylated RSK1 PBMs. Furthermore, we used luciferase complementation assay to measure the effect of epidermal growth factor (EGF) stimulation on full-length RSK–PDZ interactions in HEK293T cells. Our work reveals new kinase-scaffold complexes, the mechanism of PDZ domain-based RSK substrate targeting, and identifies new functions of RSK1.

Results

PDZome-binding profiles of native and phosphorylated RSK1 PBMs

To investigate how phosphorylation can modulate the binding of the RSK1 PBM to PDZ domains, we used the automated high-throughput holdup assay, which allows to measure binding intensities (BIs) for a large number of domain-motif pairs. As compared to the original work describing this approach [19], we used an updated version of our PDZ library, including all the 266 known human PDZ domains [20]. We were able to quantify the interaction of 255 PDZ for the unphosphorylated RSK1 peptide and 252 for the phosphorylated form (96% and 95% of the human PDZome, respectively). Both data sets were plotted in the form of "PDZome-binding profiles" (Fig. 2A) representing all the individual BIs measured for each PDZ domain for the unphosphorylated and phosphorylated RSK peptides, respectively. Using BI = 0.2 as the minimal threshold for a significant PDZ–peptide interaction, the holdup assay identified 34 potential RSK1 binders, including 26 PDZ binders for the unphosphorylated peptide and 25 binders for the phospho-peptide (Figs. 2A and S1,

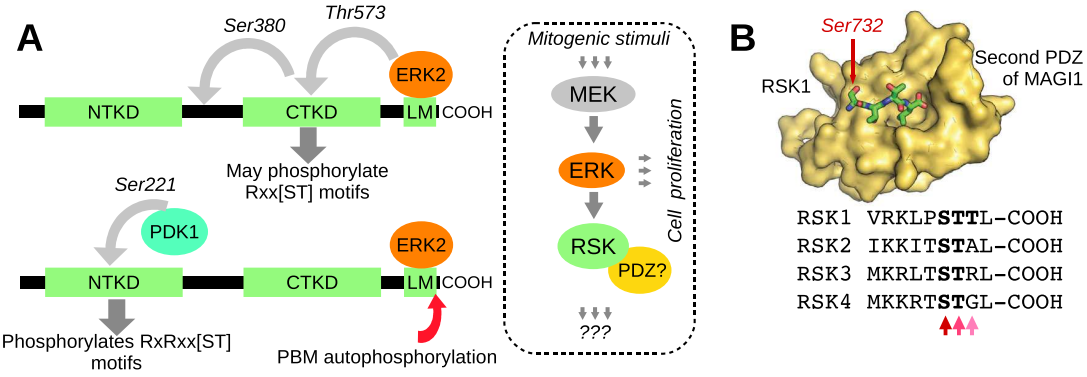


Fig. 1. The activation of RSK includes a feedback phosphorylation site that can affect PDZ binding. (A) Activation of the tandem kinase RSK is a multi-step process. Activation of RSK is initiated by ERK docking, which is followed by the phosphorylation of the C-terminal kinase domain (CTKD) [9]. The active CTKD phosphorylates a linker site between the kinase domains that creates a docking site for PDK1 [11]. In the end, PDK1 activates the N-terminal kinase domain (NTKD) [10]. Usually, only the NTKD is considered as an effector kinase, and the CTKD is only associated with a self-regulatory role, but one of these activated kinases phosphorylates its C-terminal PBM. While RSK is an effector of the mitogenic ERK pathway, its downstream effects are not well explored. (B) Each RSK isoforms contain a functional class 1 PBM. RSK1 contains 3 mutually exclusive autophosphorylation sites (at the minus 1, 2, and 3 positions) and the other isoforms contain only two (at the minus 2, 3 positions), but only the minus 3 site (Ser732 in RSK1) is considered as a major feedback site [16]. The structural panel shows RSK1 binding to the second PDZ domain of MAGI1.

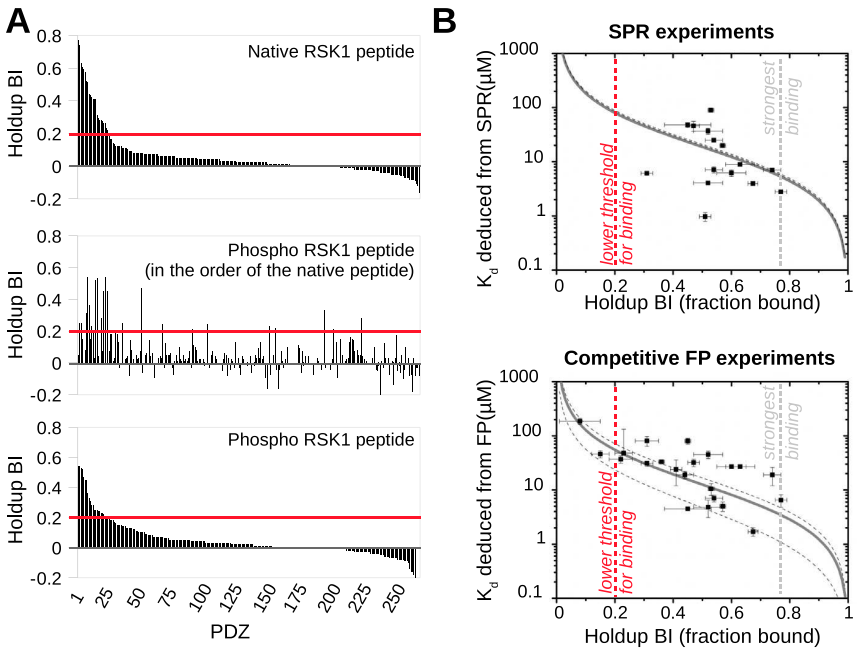


Fig. 2. PDZome binding of RSK1 explored by *in vitro* protein–peptide binding assays. (A) PDZome-binding profiles of unphosphorylated and phosphorylated RSK1 PBMs. A red line indicates the cutoff for a significant PDZ–PBM interaction (BI > 0.2). PDZ domains in the upper and lower plots are ranked on the basis of their BIs for the indicated peptide. In the middle plot, PDZ domains are ranked on the basis of their BIs for the unphosphorylated peptide, while the plotted BI values are those obtained for the phosphopeptide. Note the considerable reshuffling of binding ligands induced by phosphorylation. (B) Comparison of orthogonal binding data obtained by the holdup assay, SPR, and an FP-based assay. The correlation of BIs obtained by holdup assays to the affinities deduced from SPR or competitive FP measurements was fitted using a Monte Carlo approach. Despite the independent affinity measurement procedures, a similar correlation was observed in both cases. The fitting procedure delivers a value for the peptide concentration in the holdup assay, combining this with the free and peptide-bound PDZ domain concentrations (both delivered by the holdup assay); the dissociation constant of all human PDZ domains could be determined.

Table S1). The general distribution of the PDZome-binding profiles was similar in both cases. However, phosphorylation decreased the maximal and average BIs from 0.77 to 0.54, and from 0.42 to 0.33, respectively. Furthermore, the order of the PDZ domains that bind best to the unphosphorylated and phosphorylated RSK1 PBM was markedly different, as visually illustrated by the global reshuffling of their respective profiles (Fig. 2A). Using the same threshold for significant binding, the phosphorylated RSK1 PBM lost 12 of the detectable binders and gained 10 new binders as compared to the unphosphorylated peptide. This implies that at least 35% of the potential binders interact (often with variable affinities) to both phosphorylated and native RSK1 peptides, while the rest of them bind detectably to only one state of the RSK1 PBM.

In vitro validation of PDZ–RSK interactions by biophysical approaches

To validate the results of the holdup assay, we used orthogonal *in vitro* approaches: isothermal titration calorimetry (ITC), surface plasmon resonance (SPR-Biacore), direct and competitive fluorescence polarization (FP) (Table 1, Figs. S2 and S3). To benchmark the BIs of the holdup assay against steady-state dissociation constants, we decided to systematically test by high-throughput techniques (SPR-Biacore and competitive FP) those interactions that showed a BI value larger than 0.4 in any of the two holdup assays (Table 1, and Figs. S2 and S3). With these methods, we were able to accurately measure binding constants of 15 and 28 PDZ–PBM pairs, respectively. These experiments quantitatively confirmed the phosphorylation-induced changes in binding affinities, which were originally detected by the holdup assay. For example, a 3- to 6-µM dissociation constant was apparent between the PDZ of ARHGEF12 and the native RSK1 peptide, while no interaction could be detected with the phosphorylated state. Vica versa, the PDZ domain of SYNJ2BP interacted with the phosphorylated peptide with a 10- to 20-µM dissociation constant, while no interaction was detectable with the native state.

We used these data sets to estimate the quantitative correlation between measured BIs and the dissociation constants using Monte Carlo modeling and a general

Table 1. Summary of the *in vitro* experiments

	BI	K_D, direct FP (μM)	K_D, competitive FP (μM)	K_D, SPR (μM)	K_D, SPR, HPV16 E6 (μM)	K_D, estimated (μM)	Fold change
ARHGEF12	0.77 ± 0.02; 0.05 ± 0.05	7.5 ± 0.8; 29 ± 8	6.6 ± 1.7; >100	2.79 ± 0.11; no binding	10.9 ± 1.4	4.2; >100	0.04
GRID2IP-2	0.67 ± 0.02; 0.00 ± 0.01	5.1 ± 0.4; 47 ± 15	1.7 ± 0.3; 85 ± 11	3.96 ± 0.12; no binding	No binding	7.1; >100	0.07
MAST2	0.74 ± 0.03; 0.23 ± 0.03	7.9 ± 0.6; 13 ± 2	19 ± 7; 48 ± 84	7.02 ± 0.27; no binding	2.5 ± 0.2	4.9; 53.8	0.09
PDZD7–3	0.60 ± 0.03; 0.15 ± 0.03	0.80 ± 0.05; 1.8 ± 0.1	4 ± 1; 46 ± 7	6.2 ± 0.9; no binding	No binding	9.7; 92.9	0.10
MAST1	0.57 ± 0.01; 0.08 ± 0.03	26 ± 4; 34 ± 8	5 ± 1; 92 ± 12	20 ± 1; no binding	No binding	11.1; >100	0.11
GOPC	0.63 ± 0.05; 0.25 ± 0.10	20 ± 1; >100	27 ± 2; >100	8.92 ± 0.44; no binding	No binding	8.5; 48.0	0.18
MAGI1–2	0.43 ± 0.02; 0.15 ± 0.02	ND; ND	ND; ND	No binding, no binding	3.4 ± 0.8	20.3; 92.9	0.22
NHERF3–1	0.41 ± 0.01; 0.03 ± 0.01	80 ± 20; 220 ± 30	ND; ND	No binding, no binding	23 ± 3	22.1; >100	0.22
GORASP2	0.41 ± 0.02; 0.19 ± 0.01	67 ± 33; 114 ± 35	ND; ND	No binding, no binding	No binding	22.1; 69.2	0.32
GRASP	0.29 ± 0.01; 0.04 ± 0.01	ND; ND	ND; ND	ND; ND	ND	38.8; >100	0.38
PARD3B-1	0.52 ± 0.05; 0.31 ± 0.02	27 ± 3; 6.8 ± 0.5	45 ± 7; 31 ± 3.5	4.0 ± 0.3; 6.1 ± 0.4	No binding	13.8; 35.1	0.39
MAGI2–2	0.42 ± 0.01; 0.23 ± 0.03	420 ± 30; 430 ± 45	ND; ND	No binding, no binding	2.9 ± 0.14	21.2; 53.8	0.39
ARHGEF11	0.28 ± 0.06; 0.01 ± 0.02	ND; ND	ND; ND	ND; ND	ND	40.8; >100	0.40
SHANK3	0.27 ± 0.03; 0.07 ± 0.01	ND; ND	ND; ND	ND; ND	ND	43.0; >100	0.43
DFNB31–3	0.23 ± 0.04; −0.01 ± 0.02	ND; ND	ND; ND	ND; ND	ND	53.8; >100	0.53
NHERF2–2	0.20 ± 0.04; 0.07 ± 0.05	ND; ND	ND; ND	ND; ND	ND	64.8; >100	0.64
HTRA1	0.44 ± 0.03; 0.36 ± 0.01	30 ± 2; 11.3 ± 0.4	19 ± 3; 33 ± 2	No binding, no binding	No binding	19.4; 27.7	0.70
MAGI3–2	0.28 ± 0.03; 0.28 ± 0.06	ND; ND	ND; ND	ND; ND	ND	40.8; 40.8	1.00
PDZRN4–1	0.51 ± 0.02; 0.54 ± 0.03	33 ± 5; 14 ± 2	ND; ND	0.97 ± 0.18; 7.1 ± 0.9	6.0 ± 1.5	14.4; 12.6	1.14
SNTG2	0.41 ± 0.02; 0.52 ± 0.05	65 ± 2; 24 ± 2	24 ± 12; 4.8 ± 1.7	No binding; 37 ± 5	No binding	22.1; 13.8	1.60
PTPN3	0.05 ± 0.02; 0.21 ± 0.02	ND; ND	ND; ND	ND; ND	ND	>100; 60.8	1.66
SHROOM2	0.00 ± 0.01; 0.21 ± 0.01	ND; ND	ND; ND	ND; ND	ND	>100; 60.8	1.66
LIMK2	0.01 ± 0.06; 0.22 ± 0.07	ND; ND	ND; ND	ND; ND	ND	>100; 57.2	1.77
GORASP1	0.01 ± 0.02; 0.23 ± 0.03	ND; ND	ND; ND	ND; ND	ND	>100; 53.8	1.88
GRID2IP-1	0.06 ± 0.02; 0.24 ± 0.01	ND; ND	ND; ND	ND; ND	ND	>100; 50.8	1.99
LNX1–3	0.04 ± 0.02; 0.24 ± 0.07	ND; ND	ND; ND	ND; ND	ND	>100; 50.8	1.99
DLG4–2	0.11 ± 0.03; 0.25 ± 0.02	ND; ND	ND; ND	ND; ND	ND	>100; 48.0	2.10
PDZRN3–1	0.26 ± 0.01; 0.45 ± 0.01	90 ± 25; 17.5 ± 1.4	>100; 80 ± 10	No binding, no binding	8.6 ± 1.6	45.4; 18.6	2.45
LAP2	−0.02 ± 0.05; 0.28 ± 0.01	ND; ND	ND; ND	ND; ND	ND	>100; 40.8	2.47
SNTA1	0.31 ± 0.04; 0.53 ± 0.01	41 ± 11; 4.9 ± 0.4	81 ± 17; 10.5 ± 2.6	No binding; 90 ± 4	101 ± 40	35.1; 13.2	2.66
SNTB1	0.22 ± 0.04; 0.45 ± 0.08	18 ± 2; 1.5 ± 0.1	37 ± 6; 4.5 ± 0.3	No binding; 48 ± 5	27 ± 4	57.2; 18.6	3.08
PPP1R9A	0.00 ± 0.02; 0.33 ± 0.02	ND; ND	ND; ND	ND; ND	ND	>100; 31.8	3.17
SYNJ2BP	0.26 ± 0.07; 0.54 ± 0.03	39 ± 2; 16 ± 1	>100; 7 ± 1	No binding; 25 ± 1	33 ± 4	45.4; 12.6	3.59
SNX27	0.08 ± 0.07; 0.47 ± 0.02	25 ± 6; 4.4 ± 0.4	185 ± 25; 32 ± 5	No binding; 46 ± 9	No binding	>100; 17.1	5.92

Values after the semicolon correspond to the phosphorylated RSK1 peptides. HPV16 E6 was used as an internal standard during the SPR measurements. K_d estimation was calculated from BI values as described in Materials and Methods and using an estimated 17 μM peptide concentration. Fold changes were calculated by dividing the estimated unphosphorylated and the phosphorylated dissociation constants. For undetectable interactions, a very weak K_d was assumed (100 μM, which corresponds to a BI of 0.14). ND means not determined, while no binding means that it was impossible to quantitatively measure their affinities in our experimental conditions.

equation of the dissociation constant. While different experimental methods resulted slightly different affinities (where only two K_d pairs showed larger than a magnitude difference), their independent fits resulted similar conclusions. We have found that the peptide concentration in the holdup assays was between 14 and 23 μM (Fig. 2B, Table 1). Using this fitted parameter, it can be calculated that the holdup assay was capable of detecting any interaction with K_d < 65 μM (at the 0.2 BI cutoff).

Dynamic rearrangement of the RSK1–PDZ interactome *in vitro*

The holdup assay identified ARHGEF12 as the strongest interaction partner of the unphosphorylated peptide (BI = 0.77; K_D ≈ 4 μM) (Fig. 3 and Tables 1, S1). This protein is a RhoA GEF. It has recently been reported that its interaction with RSK2 is essential in RhoA activation and that this interaction leads to increased cell motility in the U87MG glioblastoma cell line [21]. We also identified strong interaction with MAST2, which is an AGC kinase similarly to RSK (BI = 0.74; K_D ≈ 5 μM) [22]. The previously characterized interaction between MAGI1 and RSK1 was found among the top binders of the unphosphorylated data set (BI = 0.43; K_D ≈ 20 μM). Interestingly, our approach shows that phosphorylation down-regulate this interaction by a factor of five in contrast to earlier works [5,15]. This is very likely due to the limited dynamic change of other methods (such as ITC) in cases of very weak interactions (e.g., compare Fig. S3B with S3B or S2). The strongest interaction partners of the phosphorylated PBM were three signal transducing adaptor proteins SYNJ2BP, SNTA1, and the E3 ubiquitin ligase PDZRN4 (in all cases BI ≈ 0.54; K_D ≈ 13 μM) (Fig. 3, Tables 1 and S1).

Approximately one third of the identified PDZ interaction partners of RSK1 were capable to interact with both states of the PBM (Fig. 3B). By using the holdup assay, we had the unique opportunity to gain quantitative insight into the dynamic changes that occur after a single phosphorylation event (Fig. 3C). At the two extremes, RSK1 was engaged in OFF and ON "phospho-switches" (according to our detection threshold) with ARHGEF12 and SNX27, respectively. All other interactions showed a gradual modulation by phosphorylation. In conclusion, we provided *in vitro* experimental evidence that phosphorylation reshuffles the whole RSK1–PDZ interactome.

The dynamics of RSK1–PDZ interactions in cells

The observed changes in steady-state binding affinities suggested large-scale rewiring of the RSK–PDZ interactions. To test this concept, we validated selected interactions in a cellular context using a split-luciferase fragment complementation system, called NanoBiT. This method is appropriate for detecting dynamic changes in PPIs [23]. Instead of using isolated, purified PDZ domains and RSK peptides, we used full-length proteins in HEK293T cells. Wild type

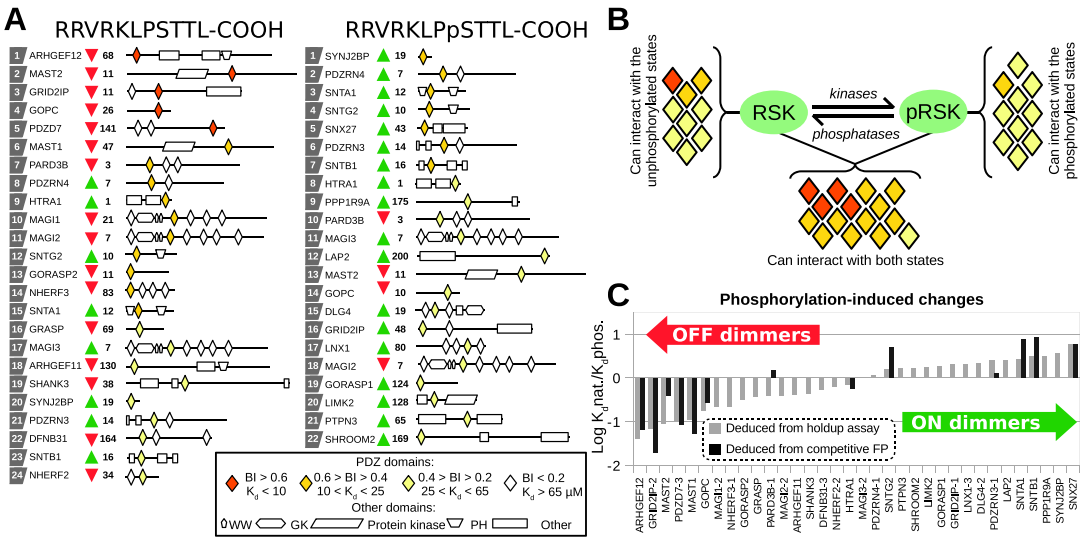


Fig. 3. Binding affinity changes elicited by PBM phosphorylation. (A) Domain architecture of the identified interaction partners. The PDZ domains are colored according to the measured BI values. (B) Using lowered cutoffs in the holdup assay (K_d < 100 μM), almost half of the identified RSK1 interaction partners showed detectable affinity to both states of the RSK1 C-terminal peptide. (C) Phosphorylation promotes a complex rearrangement in the RSK1 PDZ interactome. Instead of two definite classes (ON or OFF switching), a continuum (ON or OFF dimming) was measured in the phosphorylation-induced K_d differences of the holdup assay. Dark gray columns show the experimentally determined K_d differences from the competitive FP measurements.

(WT) and two mutant versions of RSK1 were used. The L714E mutation eliminates the interaction between ERK and RSK; therefore, RSK cannot be activated [9]. The ΔC1 truncation eliminates the last residue of RSK1 and thus suppresses the functional PBM of the protein [15]. We obtained high luminescence signals with the ARHGEF12, GOPC, PARD3B, MAGI1, and SYNJ2BP sensors in serum-starved cells (Fig. 4A). The C-terminal truncation significantly reduced the luminescent signal in all cases, while the L714E mutation decreased the luminescence outputs for PARD3B and SYNJ2BP.

EGF stimulation can be used to turn on the ERK signaling cascade, RSK activation and its PBM autophosphorylation [24]. Extracellular stimulation induced changes in NanoBiT sensor brightness within the same timescale as ERK–RSK dissociation (Fig. 4B). In all cases, the maximum change was

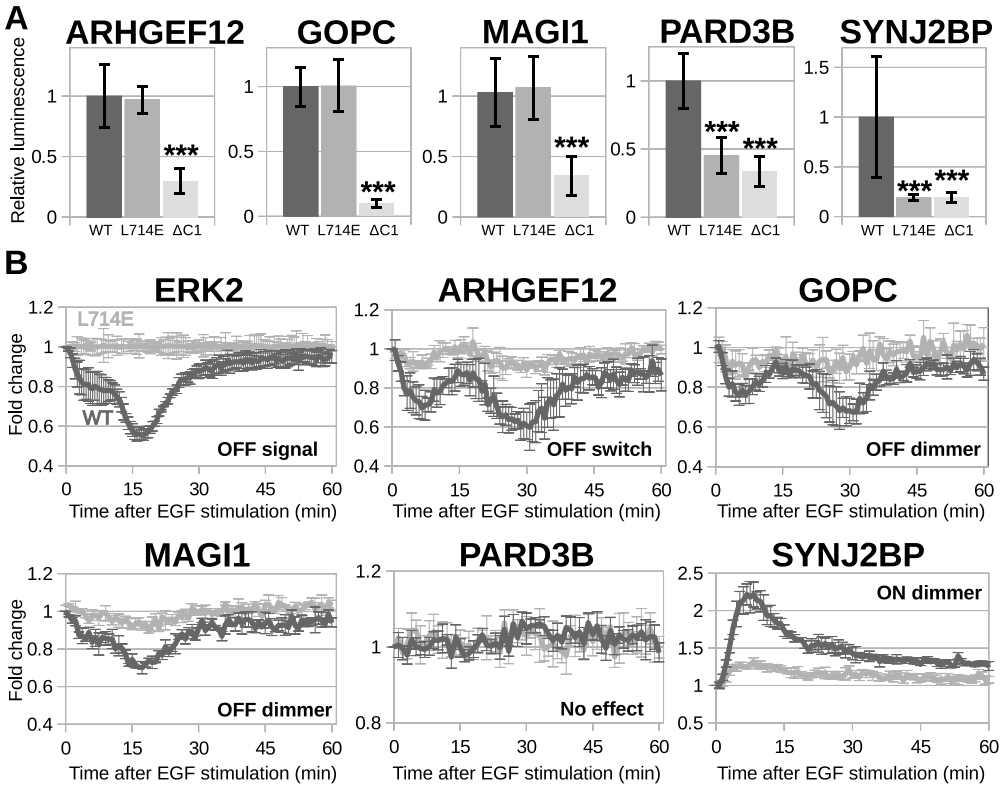


Fig. 4. Live-cell monitoring of RSK1 binding to PDZ-containing partners. (A) Monitoring steady-state luminescence with the interaction sensors between RSK1 and full-length PDZ proteins. Full-length proteins fused to two complementary fragments of nanoluc luciferase were co-expressed in serum starved HEK293T cells. The resulting luminescence signal was measured as indicated in the Materials and Methods. The luminescence signal obtained for the pair of WT constructs is used as reference (relative luminescence). The L714E RSK1 mutant is known to eliminate the interaction between RSK1 and ERK [9]. The ΔC1 RSK1 mutant does not contain the last C-terminal residue of RSK1 and therefore does not contain a functional PBM. The luminescence signal is consistently disrupted by the ΔC1 mutation, indicating that this signal reports on the PBM-mediated binding of RSK1 to its PDZ-containing targets. The L714E mutation disrupts the signal in cases where the interaction partner can significantly interact with the phosphorylated form of RSK1 (n = 6). Asterisks indicate statistical significance (***$P < 0.001$) calculated by two-tailed Student's *t* test between the luminescence signals of mutant and WT RSK1 constructs. (B) RSK1-based luminescence interaction sensors (with ERK2 and several proteins containing RSK1-binding PDZ domains) were co-expressed in serum-starved HEK293T cells. The luminescence signal in absence and in presence of EGF (20 ng/ml) was monitored for 60 min following EGF addition. The measured luminescence signal was normalized to the initial luminescence and to the spontaneous substrate (furimazine) decay based on the unstimulated cells. The dark and gray curves show the luminescence signals of the WT and the L714E mutant, respectively. EGF stimulation provokes a time-modulated decrease of the luminescence signal for co-expressed constructs of RSK1 and ERK2 as observed in our previous work [15]. Note that EGF stimulation diversely modulates the luminescence signal (increase, decrease, or no significant change) for each PDZ-containing protein in a comparable timescale to that of RSK-ERK dissociation. Remarkably, in this cell-based assay, using full-length proteins, EGF-induced luminescence signal modulation shows a good agreement to the results of *in vitro* measurements where only an RSK1 PBM peptide and PDZ domains constructs were used.

detectable between 10 and 20 min and the signal started to disappear after 30–45 min. As in our previous study [15], we observed periodic signals, which seem to be a characteristic feature of RSK-based interactions. ARHGEF12, GOPC, and MAGI1 showed a decrease in luminescence after stimulation. In contrast to these OFF signals, PARD3B did not show any change after activation of the pathway, while SYNJ2BP showed an increased luminescence after EGF stimulus. Results of this cell-based PPI tests showed a good agreement with *in vitro* measurements.

A compendium of potential RSK targets

Our high-throughput study identified a wealth of novel RSK-binding PDZ domains. The proteins that contain these RSK-binding PDZ domains represent in principle, candidate substrates of RSK kinases. In previous studies, only a few PDZ-containing partners of RSKs were assumed to be substrates [5,15,25]. To investigate this issue, we collected RSK-focused phosphoproteomic data sets for a meta-analysis. To our knowledge, there are three such data sets. (i) Galan *et al.* [26] searched for RSK substrates using specific inhibitors. (ii) Moritz *et al.* [27] tried to find tyrosine kinase activated AGC kinase substrates. (iii) Avey *et al.* [28] used the viral ORF45 protein to activate the ERK–RSK axis in cells, and they searched for up- or down-regulated phosphoproteins. (iv) In addition, a compendium of ERK targets has recently been published [29]. It is a systematic collection of ERK related phosphoproteomic studies containing both direct and indirect ERK substrates. The compendium is also a valuable resource for potential RSK phosphosites (Rxx[ST] and RxRxx[ST] motifs) [26]. The compendium contains 1477 [ST]P sites (from 892 proteins), 544 Rxx[ST] sites (from 430 proteins), and 458 other phosphorylation sites (from 330 proteins). We used this Rxx[ST] subset of the ERK compendium as an additional resource to our meta-analysis. The collection of four potential RSK substrates, termed here as RSK compendium, included 997 potential substrates, where 349 substrates were identified in more than one study (Fig. 5A, Table S2). Only 35 substrates were identified in all four phosphoproteomic data sets, including some well-characterized RSK substrates, such as ARHGEF12, EIF4B, EphA2, GSK3B, PFKFB2, PPP1R12A (MYPT1), RPS6, or SLC9A1 (NHE1) [21,30–35].

Direct and indirect phosphorylation by ERK and RSK

Of the potential RSK substrates, discussed above, only 28 were PDZ-containing proteins, about half of which were identified only in a single data set (Fig. 5B). Only four direct RSK1 binders were identified in both the RSK compendium and in our holdup assay: ARHGEF11 and 12, MAST2, and SHROOM2. Notably, ARHGEF12 was identified in all phosphoproteomic data sets as a PDZ-containing RSK substrate and was also the strongest binder of the unphosphorylated RSK1 peptide in the holdup assay. In addition, we have also found three additional partners (GRIP, SCRIB, NHERF1) binding to other RSK isoforms [18,5,25]. Conversely, it is worth noting that many of the strong RSK1 PBM binders (like GRID2IP, GOPC, PDZD7, or PDZRN4) do not contain any phosphorylation site matching the RSK1 consensus motif [16].

The RSK and the ERK compendiums show an overlap, indicating that some substrates can be phosphorylated by both RSK and ERK (Fig. 5AB). Although the MAPK- and the PBMs are found in the same C-terminal tail region of RSK where they are only separated by a few residues, it is stereochemically possible to form a ternary complex between the three domains [15]. Therefore, ERK can also phosphorylate RSK-bound PDZ proteins. We have found 8 RSK1 interaction partners that can be phosphorylated by ERK. One of them is ARHGEF12, which contains three RSK phosphorylation sites and a single MAPK phospho-site (Fig. 5C). In these cases, the C-terminal tail of RSK appears to serve a scaffolding role, bringing ERK and PDZ substrates close to each other.

To identify additional indirect, PDZ scaffold-mediated substrates, all potential interaction partners of our RSK1-binding PDZ scaffolds were collected from the IntAct PPI database [36]. This analysis revealed the significant enrichment of RSK and ERK substrates in many cases. For example, an interesting scaffold was MAGI1, which was not identified previously as a direct substrate of RSK (or ERK). MAGI1 has 74 potential interaction partners in that database; among which, more than 40% turn out to be potential RSK substrates. Similarly, 30% of MAGI1 potential interaction partners are potential ERK substrates, and 18% of them are potential substrates of both RSK and ERK (Fig. 5D, Table S3). We have found similarly significant enrichment of RSK/ERK substrates among various interaction partners, such as ARHGEF11. In conclusion, while only a small portion of RSK1-binding PDZ proteins may be direct substrates of RSK1, it appears that many of them may act as scaffolds, since many relevant potential RSK and ERK substrates can be found among their interaction partners.

Kinetic control of substrate phosphorylation

Next, we measured the kinetic parameters of PDZ–PBM interactions (Fig. S4). PDZ-bound fluorescent peptides were rapidly mixed with high molar excess of unlabeled peptides and changes in FP were monitored. Although the fluorescein label may alter the steady-state affinity of some interactions (Table 1), it probably affects only the dissociation rates, as usually observed for large hydrophobic groups. Under this assumption, unbiased off-rates for unlabeled peptides can be estimated (Fig. 6A). Our results revealed that

OFF dimmers have a generally slow binding kinetics (average $k_{off} \approx 210\ s^{-1}$), while ON dimmers showed faster dissociation rates (average $k_{off} \approx 1100\ s^{-1}$) (Fig. 6B). We used an *in silico* network-based modeling software to estimate substrate phosphorylation efficiency using these obtained kinetic parameters (Fig. S5) [37]. By using this simulation, the phosphorylated substrate levels, induced by the same amount of external stimulation, could be calculated and compared for ON and OFF switches (Fig. 6C). The analysis demonstrated that the presumed weaker interaction between OFF-dimmer PDZ domains and the active kinase should be compensated by a slower dissociation rate, thereby allowing for higher substrate phosphorylation.

Role of the RSK1 PBM in RhoA activation

RSK proteins have been proposed to play an important role in regulating cell motility, particularly through affecting the activity of the small GTPase protein, RhoA [38,21]. To this end, we have examined

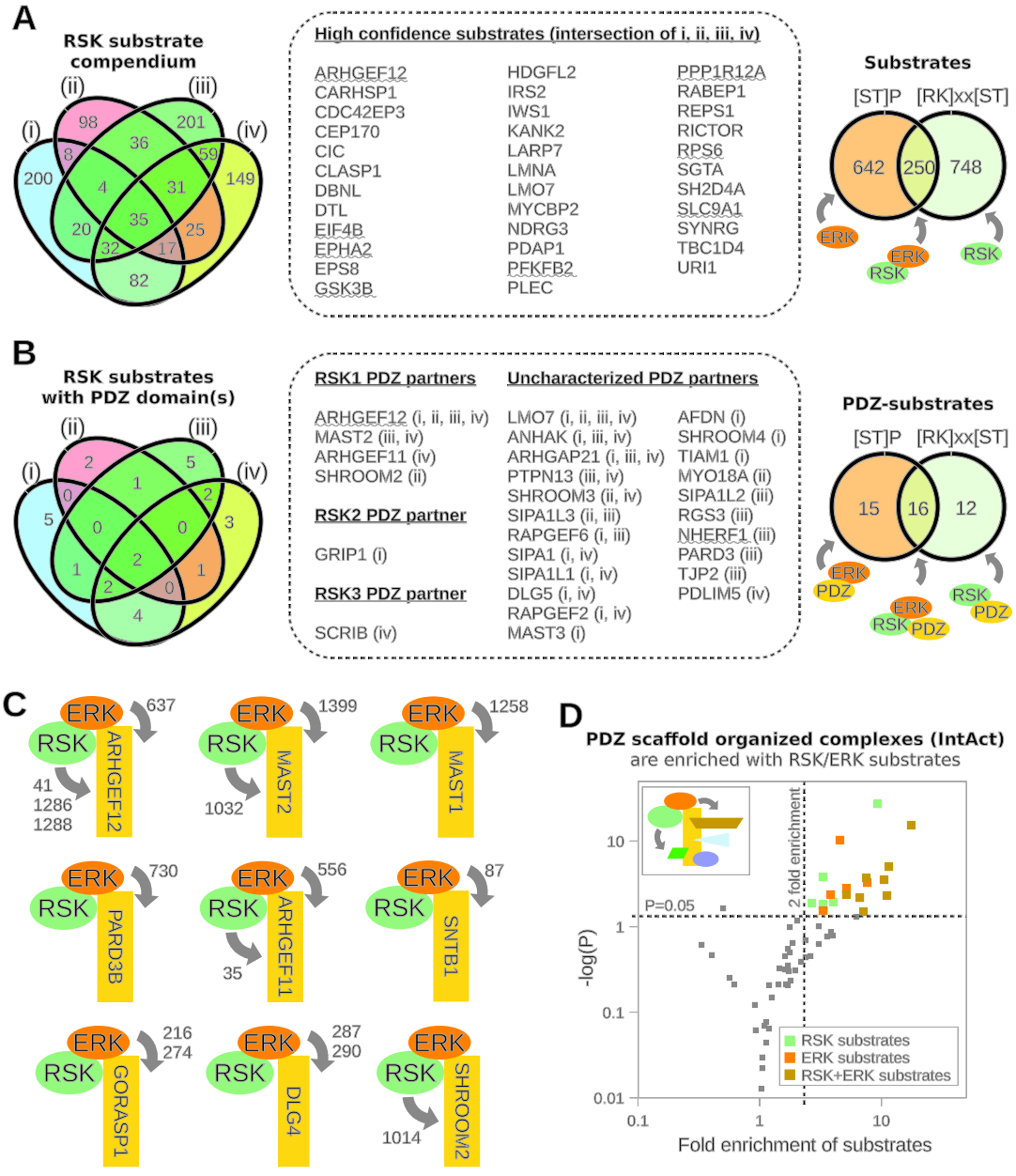


Fig. 5 (*legend on next page*)

A

	Measured k_{off} (s^{-1})	Calculated k_{on} ($s^{-1}\mu M^{-1}$)	$f_{bias}\left(\frac{K_{d,\ holdup}}{K_{d,\ direct\ FP}}\right)$	Estimated k_{off}* (s^{-1})
ARHGEF12	46 ± 2; ND	6; ND	0.56; 3.45	26; ND
MAST1	129 ± 12; ND	5; ND	0.43; 2.94	55; ND
MAST2	101 ± 5; ND	13; ND	0.62; 4.14	63; ND
GOPC	195 ± 18; ND	10;ND	0.43; 0.48	83; ND
GRID2IP-2	109 ± 5; ND	21; ND	1.39; 2.13	152; ND
PDZD7-3	72 ± 4; ND	90; ND	12.13; 51.61	873; ND
SNTG2	ND; 364 ± 87	ND; 15	0.34; 0.58	ND; 209
SYNJ2BP	ND; 472 ± 67	ND; 29	1.16; 0.79	ND; 372
SNTA1	ND; 610 ± 138	ND; 124	0.86; 2.69	ND; 1643
SNX27	ND; 562 ± 85	ND; 128	4.00; 3.89	ND; 2184
PARD3B-1	676 ± 219; 431 ± 48	25; 63	0.51; 5.16	346; 2225

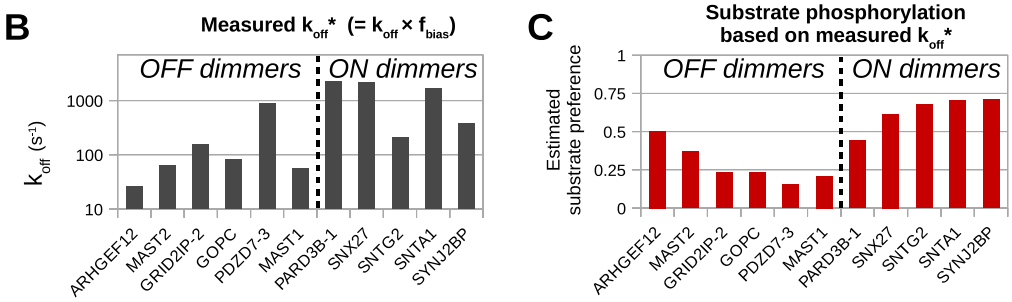


Fig. 6. Binding kinetics differ between ON and OFF dimmers. (A) Dissociation rates were deduced from stopped flow FP experiments. On-rates are calculated based on the steady-state affinities of the fluorescent peptides (deduced from direct FP measurements). The bias factor (the ratio of the binding affinities of the direct FP and the holdup assay) was applied as a correction factor to the fitted dissociation rates to estimate unbiased off-rates. Values after the semicolon correspond to the phosphorylated RSK1 peptides. ND means not determined. (B) The corrected dissociation rates (k_{off}*) of a set of RSK1–PDZ interactions. Partners with OFF dimmer behavior showed a slower binding kinetics while ON dimmers preferred faster binding rates. (C) Substrate phosphorylation was calculated using an *in silico* model with measured dissociation rates.

the role of the RSK1 C-terminal region in RhoA activation. We transiently transfected a RSK1/2 knock-out HEK293T cell line with either full-length RSK1 (WT, 1–735) or a RSK1 construct with its C-terminal residue truncated (ΔC1, 1–734). We have found that over-expressed and phosphorylated RSK1 localized in the cytoplasm, similarly to the endogenous phospho-RSK in WT HEK293T cells (Fig. 7A) [15]. Interestingly, an increased level of basal RhoA activity was only apparent in the presence of the WT RSK1 construct (Fig. 7B). This slight increase was more pronounced in cells that were stimulated by the addition of serum.

Fig. 5. Meta-analysis of phosphoproteomic studies and bioinformatic search to find potential direct and indirect PDZ-dependent substrates of RSK and of ERK. (A) Left panel: a graphical representation of the intersections of RSK substrate lists from four different HTP phosphoproteomic studies: (i) Galan *et al.* [26], (ii) Moritz *et al.* [27], (iii) Avey *et al.* [28], and (iv) [RK]xx[ST] subset of the ERK compendium [29]. Middle panel: the intersection of the four lists contains several previously characterized RSK substrates (underlined), suggesting that other proteins found in this group may also represent high-confidence RSK substrates. Right panel: the RSK compendium and the direct ERK compendium greatly overlap, suggesting that a set of substrates can be phosphorylated on both ERK ([ST]P) and RSK ([RK]xx[ST]) consensus sites. (B) Same representation as in panel A but focusing on RSK substrates with PDZ domains. Only a few PDZ domain-containing substrates are present in the whole data set, and only a handful of them were found in more than one HTP study. Moreover, only four RSK1 binding partners were identified, from which only ARHGEF12 was found in the common group of the HTP studies. Uncharacterized PDZ partners may be direct partners of other RSK isoforms, or may be PDZ-independent substrates or false positives. (C) Many RSK1 PDZ interaction partners contain an ERK phosphorylation site. In addition, a few substrates, such as ARHGEF12, can be phosphorylated by both kinases. (D) The IntAct database was used to estimate the enrichment of ERK and RSK substrates among the interaction partners of the RSK1 PDZ-dependent interaction partners. On the vulcano plot, each dot represents the enrichment of kinase substrates among the interaction partners of a PDZ scaffold. We have identified a high number of potential indirect RSK and ERK substrates among these interaction partners, which are indicated with colors in the upper right corner. *P* values indicate statistical significance compared to a random pool of intracellular proteins, calculated by chi-square test. Fold enrichment indicates the increased proportions of substrates compared to the same random pool.

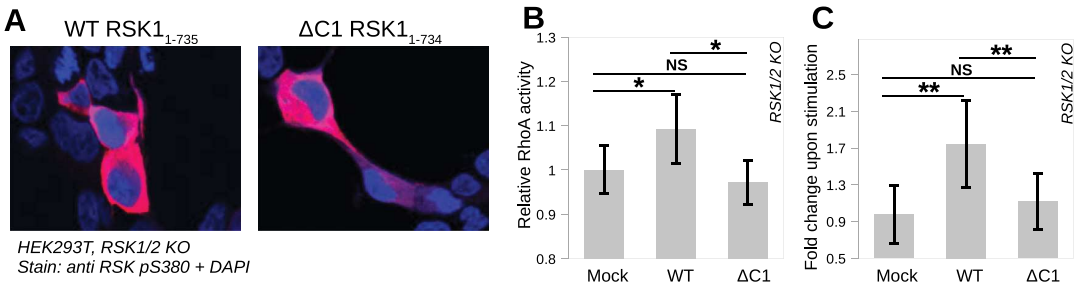


Fig. 7. The PBM of RSK1 links ERK activation to RhoA. (A) A RSK1/2 knockout HEK293T cell line was used to measure the role of the PBM of RSK1 in a native environment. Deletion of the RSK1 PBM does not affect the localization of active RSK1. (B) The presence of intact RSK1 increases the basal RhoA activity, but this effect is decreased without a functional PBM (n = 4). (C) Transfected and serum-starved cells were stimulated with serum (20% for 5 min). Without intact RSK1 (in the mock transfected knockout cell or in the presence of the PBM-lacking RSK1 construct), only minor increment was observed in the RhoA activity. The presence of intact, WT RSK1 enabled a proper response in RhoA activation upon stimulation (n = 4). Asterisks indicate statistical significance (**$P < 0.005$, *$P < 0.01$, NS $P > 0.01$) calculated by two-tailed Student's *t* test.

Stimulation increased RhoA activity in the presence of the WT RSK1 construct, but not in the presence of RSK ΔC1 (Fig. 7C). These results are consistent with a model postulating that the PBM of RSK1 serves as a docking motif for RSK1 to phosphorylate an important regulatory site in ARHGEF12, which then affects RhoA activation (Fig. S6) [39,40].

Discussion

Regulation of RSK1–PDZ interactions by PBM autophosphorylation

Previously, only a handful of PDZ interaction partners of RSK1 had been identified and their response to RSK1 autophosphorylation was largely unknown. Here, we characterized the PDZ interactome of RSK1 and examined how this changes upon PBM autophosphorylation at Ser732. Altogether, 34 interaction partners were identified with the holdup assay, most of them being novel, with the notable exception of MAGI1. In contrast to previous reports, we did not detect any interaction of RSK1 PBM with the first PDZ domain of NHERF (EBP50) and only detected a very weak affinity toward the first PDZ domain of Scribble (BI ≈ 0.17–0.18, corresponding to a dissociation constant of 60–70 μM) [25,18]. We do not think that this may be due to a lack of activity of these two domains in our assay, since both of them have already been positive with other PBMs in other holdup experiments. In particular, Scribble is positive with HPV16 E6 (BI ≈ 0.70, corresponding to a dissociation constant of 5–10 μM) [19]. Although most of the identified interactions were altered by PBM phosphorylation to some degree, we have found only a few cases that can be considered a genuine "phospho-switch." For example, detectable binding of RSK1 to ARHGEF12 and GRID2IP was mostly eliminated, while binding to the adapter protein SNX27 was promoted by phosphorylation. In contrast, most substrates showed a "phospho-dimmer" effect, where phosphorylation *modulated* binding rather than *switching* it ON or OFF. Approximately as much ON as OFF dimmers were identified. These partners are able to interact with both states of the RSK1 PBM, albeit with different affinities. The rest of the interaction partners (such as PARD3B) displayed comparable affinities to both states of the RSK1 PBM, and therefore, these interaction partners are likely unable to sense the presence or absence of the phosphoryl group. Similar dimming mechanism was described on phosphorylation of PDZ domains themselves [41].

Mitogenic stimulation, such as that mediated by EGF, activates the MAPK pathway. Eventually, the downstream signals will activate ERK, leading to RSK1 phosphorylation and subsequent autophosphorylation in Ser732 of its PBM. Therefore, upon stimulation, we can expect dynamic changes in the RSK PBM-PDZ interactome based on quantitative *in vitro* measurements. To test this assumption, we created five intracellular PPI sensors for selected PDZ-dependent RSK1 interactions. In our assays, ARHGEF12, GOPC, and MAGI1 showed a preference for the native PBM, while the PDZ domain of SYNJ2BP preferred the phosphorylated PBM. In contrast, PARD3B could interact with both versions of RSK1. This cell-based protein–protein interaction study showed that EGF stimulation induces a phosphorylation-mediated rewiring of the RSK1–PDZ interactome inside cells, following the trends of the *in vitro* observations.

Further analysis on RSK and its PDZ-containing binding partners indicated that some of the latter are phosphorylated by RSK. Among the unambiguously identified RSK substrates, ARHGEF12 has a prominent place. It is a strong partner of the RSK1 peptide and their interaction is responsive to EGF stimulation. Moreover, Shi *et al.* [21] have recently showed that the

association between RSK2 and ARHGEF12 (also known as leukemia-associated RhoGEF or LARG) is essential in RhoA activation in glioblastoma cells. They discovered that RSK can interact with ARHGEF12 and phosphorylate it at Ser1288. They demonstrated that the presence of RSK is essential for the association between RhoA and ARHGEF12, and for subsequent RhoA activation. Inactivation or inhibition of RSK eliminated RhoA activation in response to extracellular stimulation. Our experiments gave similar results with RSK1, highlighting the central role of the RSK PBM in this process (Figs. 7 and S6).

Kinetic compensation in dynamic networks

Many direct substrates or substrate-tethering scaffolds of RSK (e.g., ARHGEF12 or MAGI1) contain an OFF dimmer PDZ domain. This creates a paradoxical situation, because the active kinase will down-regulate complex formation, and thus, only a smaller fraction of the kinase should be capable of mediating the phosphoryl transfer. Despite this, phosphorylation of such OFF dimmer type substrates can be detected with high confidence (e.g., Ser1288 of ARHGEF12). Here, we propose that the lifetime of these OFF dimmer interactions can substantially increase their phosphorylation. In our experiments, OFF dimmer PDZ domains showed 5 times slower dissociation rates than ON dimmer interactions (Figs. 6 and S4). This kinetic compensation can largely contribute to substrate phosphorylation (Fig. S5). We should emphasize here that these are general principles and they should be true for many other feedback-coupled enzymatic processes [42].

Phosphorylation-sensitive PDZ domains

Phosphorylation of PBMs is a very common regulatory mechanism in human cells [18]. Based on our experiments, we identified a set of PDZ domains that are responsible for mediating the OFF or ON dimmer effects of the phosphorylatable −3 position of the RSK PBM. Comparison of PDZ sequences reveals that there is no obvious driving force behind OFF dimmer behavior, but there are at least three positions within the peptide binding groove that can be important for ON dimmers (Fig. 8). The first of them is the outward facing residue of the second strand (βB) of the PDZ domain. This side chain is positioned in close proximity of the phosphate group, and while it is usually a Ser/Thr residue in PDZ domains, an Asn residue is preferred within ON dimmers. The other two altered side chains are within the third strand (βC) of the PDZ domain. Here, both external side chains are altered in ON dimmers. Interestingly, the closest residue to the phosphosite is most frequently a Ser residue and the other one is a basic amino acid. The role of two of these residues in the coordination of the phosphate group was captured in a crystal structure of SNX27 [43].

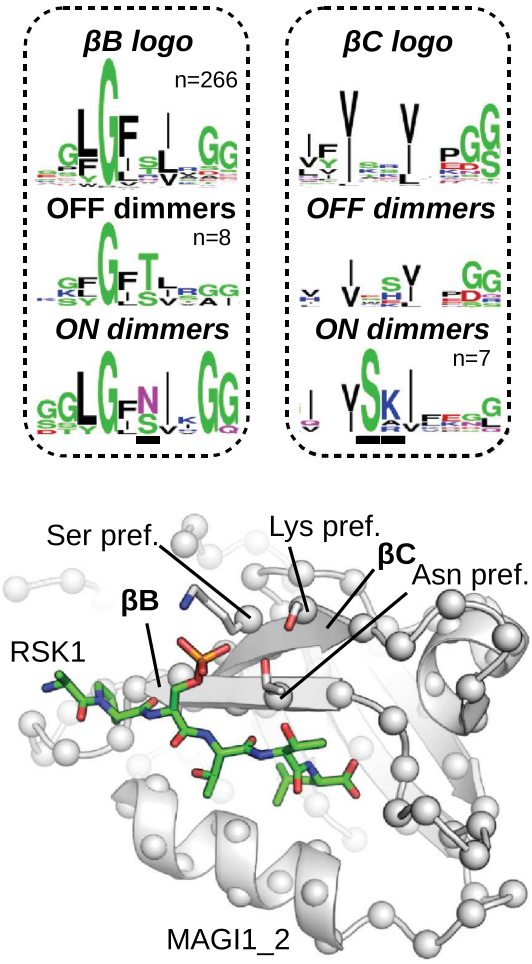


Fig. 8. Determinants of −3 phospho-PBM specificity. The PBM-binding groove of ON dimmer PDZ domains displays some notable sequence preferences. Sequence logos were generated from every human PDZ sequences or from identified dimmer subsets of the RSK1 peptide partners. Important differences are underlined in the ON dimmer sequence logo, and their side-chains are shown with sticks in the structure of the OFF dimmer MAGI1. In contrast to ON dimmers, no preferences was identified for OFF dimmers.

Asn56 from βB and Ser82 from βC mediate a hydrogen bond with the phosphate group of PBMs. Moreover, replacing the basic residue in the βC (Arg762) to Ala in Scribble can swap the RSK3 binding properties from ON- to OFF-dimmer [18]. These observations led us to the conclusion that ON dimmer propensity is determined by the presence of phosphate acceptor sites while OFF dimmer propensity is currently not understood. Further studies are needed to collect more evidence about such effects and classify PDZ domains on the basis of their response to phosphorylation events at different positions of their target PBM sequences.

Response to phosphorylation: switches and dimmers

Phosphorylation can alter linear motif binding by multiple ways. In the literature, most examples of phosphorylation-induced PPI changes are considered as switches (usually called "phospho- switches"), which can turn PPIs on or off. However, signaling processes are not solely based on binary events and may also involve fine-tuning mechanisms. A "switch" refers to binary transitions between two distinct states (the interaction occurs or does not occur), while a "dimmer" allows a fine tuning mechanism (smaller or larger changes in the affinity of an interaction). The dimming mechanism makes sense for describing events based on non-covalent interactions, however in some cases (e.g., in the context of additional binding events) synergism can enhance this effect resulting in switch-like changes. Our results demonstrate a continuum between ON and OFF switches, including many gradually altered dimmers, suggesting that, among phosphorylation-induced changes, ON/OFF dimmer effects may predominate, while ON/OFF switches represent only extreme cases.

Materials and Methods

Holdup assay

The automated holdup assay was carried out against peptides ($RSK1_{725–735}$) in triplicates as previously described [19] with minor modifications. In brief, we measured the fraction of PDZ depletion (BI) in the fluid phase during a pull-down experiment. For the detailed protocol, please look at Ref. [20]. The sequences of the clones of the PDZome v2 were designed according to [2]. All genes were codon optimized for *Escherichia coli* expression and cloned into a pETG41A plasmid. All protein constructs were expressed in *E. coli* following the previous protocol [19] with minimal modifications. All constructs were checked for solubility, and cell lysate soluble fractions were adjusted to approximately 4 μM concentration and frozen in 96-well plates. In addition, mass spectrometry was used to confirm the identity of each PDZ clones. For the detailed protocols of production and quality control, please look at Ref. [20]. We measured interactions against 255 proteins with the unphosphorylated peptide and against 252 proteins with the phosphorylated peptide. The missing proteins from the human PDZome (consisting of 266 proteins) showed problems with the expressions or stability of these constructs, or we had technical issues during the assay. In this work, BI = 0.2 was used as the minimal BI threshold value to define high-confidence PDZ–PBM pairs, as proposed previously [19]. Figure S1 contains the BI values of the RSK1 and phospho-RSK1 data sets. Data were analyzed as formerly described [19]. All plots and calculations in this work were done using these conventional data sets. In addition, we already provide the values calculated with an updated protocol in the supplemental file, because the new calculation approach will set the standard for future holdup papers. These were generated using an automated computational protocol awaiting for publication. This updated analysis revealed three new interaction partners of the native RSK1 peptide (SCRIB-3, MPDZ-10, and RHPN1) and four new partners of the phosphorylated peptide (SCRIB-3, LIN7A, PDZRN3–2, and DLG3). Apart from these weak interaction partners, most values are coherent between calculations.

Protein expression and purification and peptide synthesis

Tandem affinity (Ni- and MBP-) purified MBP-PDZ proteins were used in biochemical assays. Unphosphorylated $RSK1_{683–735}$ peptides were recombinantly expressed with an N-terminal cleavable GST tag. After affinity purification, the GST tag was removed and the peptide was isolated by reverse phase HPLC. A fraction of the isolated peptide was phosphorylated with a constitutively active (T573E mutant) RSK1 C-terminal kinase domain as formerly described [19]. Unphosphorylated, phosphorylated, and fluorescein-labeled or -unlabeled $RSK1_{729–735}$ peptides were all chemically synthesized on an automated PSE Peptide Synthesizer (Protein Technologies, Tucson, AZ, USA) with Fmoc strategy. Biotinylated $RSK1_{725–735}$ peptides were purchased from JPT Innovative Peptide Solutions with 70%–80% purity. The biotin group was attached to the N-terminal *via* a TTDS linker. Protein (and Tyr containing peptide) concentrations were determined by UV spectroscopy. For peptides that lacked an aromatic residue, their dry mass was directly measured. Predicted peptide masses were confirmed by mass spectrometry.

Isothermal titration calorimetry

ITC measurements were carried out in 20 mM Hepes (pH 7.5), 150 mM NaCl, and 500 μM TCEP using a VP-ITC apparatus (MicroCal). MBP-PDZ domain (50 μM) was titrated with concentrated peptides at 37 °C. The Origin for ITC 5.0 (Originlab) software package was used for data processing.

Surface plasmon resonance

SPR measurements were performed on a Biacore T200 instrument equipped with CM5 sensor chip. Streptavidine was immobilized on the sensor chip with EDC-MS using a standard protocol. Biotinylated peptides (RSK1, pRSK1, HPV16E6) were immobilized on streptavidine, and after an extensive washing step, MBP-PDZ domains were injected onto the chip at eight

different concentrations and with two additional replicates. Unfortunately, our SPR analysis did not reveal the kinetic parameters of the studied PDZ–peptide interactions due to biphasic and very fast behavior. The saturated phase of the reference channel subtracted data was fitted with a hyperbolic function.

Steady-state FP

FP was measured in 384-well plates (Corning) using Synergy H4 multi-mode reader (BioTek). For direct titration experiments, 50 nM reporter peptide ($RSK1_{729–735}$) was mixed with increasing amount of MBP-PDZ domains. In competitive measurements, the 50 nM reporter peptide was mixed with the PDZ domain in a concentration to achieve high degree of complex formation. Subsequently, increasing amount of unlabeled peptide ($RSK1_{683–735}$) was added to the reaction mixture. Titration experiments were carried out in triplicate, and the average FP signal was used for fitting the data to a quadratic or competitive binding equation.

Monte Carlo modeling

To estimate the dissociation constant of weak interactions, we used the measured BI values from the HU assay. This parameter equals the bound fraction of the PDZ domain; therefore, it can be inserted directly into the general binding equation:

$$K_d = \frac{[PDZ_{free}] * [RSK_{free}]}{[PDZ–RSKcomplex]} = \frac{([PDZ_{tot}] - BI * [PDZ_{tot}]) * ([RSK_{tot}] - BI * [PDZ_{tot}])}{BI * [PDZ_{tot}]}$$

Assuming that the total PDZ domain concentration is ~4 μM, the only unknown parameter is the total peptide concentration. Instead of a simple nonlinear fit, we have used an in-house Python script for Monte Carlo modeling and least squares fitting to utilize the standard deviations of the HU assay and the K_d measurements. Each fitting was repeated 10,000 times, and the average peptide concentration along with the lower and upper quartiles was plotted in Fig. 2B. Based on our SPR measurements, the RSK peptide concentration should be around 20 μM (most probably between 18 and 21 μM). Direct FP indicates that this concentration should be around 14 μM (most probably between 6 and 20 μM). In the case of the competitive FP, we have found that the peptide concentration should be around 14 μM (most probably between 9 and 18 μM). For K_d extrapolation, we have used a peptide concentration of 17 μM.

Protein–protein interaction assay

The NanoBiT PPI MCS starter system was purchased from Promega. The N-terminus of RSK1 was tagged with the short NanoBiT tag (SmBiT) and either the N- or the C-terminus of the interaction partner with the large NanoBiT tag (LgBiT). Full-length RSK1 was cloned into pBit2.1-N[TK/SmBiT] vector. Full-length MAGI1 and ERK2 constructs were previously cloned into the LgBiT vector. Full-length ARHGEF12 (isoform 2), GOPC (isoform 2), PARD3B (1–913), and SYNJ2BP were cloned into the pBit1.1-N[TK/LgBiT] vectors. All constructs were cloned from HEK293T or HeLa cDNA pools and were confirmed by sequencing. HEK293T cells were cultured in Dulbecco's modified Eagle medium (Lonza) containing 10% fetal bovine serum and 1% penicillin/streptomycin/amphotericin B. 2×10^4 cells/well were seeded onto a white, TC-treated 96-well plate (Greiner) 24 h prior to transfection. Transient transfections were carried out with FuGene HD reagent (Promega) according to the NanoBiT system's instructions. Four hours after transfection, cells were starved for 20 h in CO_2-independent medium (Thermo). Cells were assayed 24 h after transfection using Nano-Glo reagent (Promega) and a Synergy H4 plate reader (BioTek). Experiments were carried out according to the manufacturer's instructions. To validate the interaction sensors, we compared the steady-state luminescence signals of different mutants in serum-starved cells. Stimulation was performed using 20 ng/ml EGF (Sigma-Aldrich). Each experiment was performed with at least six biological replicates. We must note that we observed that the observed periodicity in the luminescence appeared environment dependent, as even under slightly modified conditions (i.e., different media, cell density or protein expression level) no periodic features appeared.

Signaling pathway modeling

Rule-based network modeling was carried out with the software package BioNetGen with the ordinary differential equation solver running on a desktop PC. The simulated pathway was described in Fig. S5A. Pathway activation was initiated from a pre-equilibrated state. The simulation was initiated by introducing the "Stim" to the system. This simplified, artificial signal generator was adjusted to mimic the natural activation profile of the ERK pathway upon EGF stimulation.

Stopped-flow FP

Fast kinetic measurements were performed with the stopped-flow instrument SFM-300 (Bio-Logic) with polarized excitation at 488 nm. Parallel and perpendicular fluorescent emissions were measured through a 550 ± 20-nm band-pass filter (Comar Optics). All reactions were measured at 25 °C in a buffer containing 20 nM Hepes (pH 7.5), 150 mM NaCl, and 150 μM TCEP. Post-mixing fluorescent peptide concentration was 0.5 μM. The fluorescent peptide ($RSK1_{729–735}$) was pre-complexed with high amount of MBP-PDZ domain (5–40 μM, post-mix). To

measure the dissociation of the labeled peptide, we rapidly mixed the PDZ bound complex with high molar excess of unlabeled peptide ($RSK1_{729–735}$ 100 μM, post-mix). Each experiment was carried out multiple times ($n > 9$), and the averaged transients were fitted using a single exponential function. Corrections were applied to estimate the unbiased binding of an unlabeled peptide based on the dissociation constant differences between the direct FP measurements and the unbiased HU assay.

Immunofluorescence

For detection of the intracellular localization of transfected proteins, 1×10^5 cells/well were seeded onto a cover slip-containing (Assistent) 24-well plate. Cells were fixed with 4% PFA solution and blocked for 1 h in 5% BSA and 0.3% Triton-X 100, dissolved in PBS at room temperature. The RSK1/2 knockout (CRISPR) HEK293 cell line was a kind gift from Fanxiu Zhu. To introduce exogeneous WT or mutant RSK1 into these cell lines, we created pIRES2-EGFP based vectors, which expressed untagged RSK1s along with a GFP transfection reporter gene. Phosphorylated RSK was detected with the help of anti-pRSK pSer380 (1:800, CST) primary and Alexa Fluor 647 (anti-rabbit, 1:800, Thermo) conjugated secondary antibodies. ARHGEF12 (isoform 2) was cloned into a pmCherry-C1 vector. Mutations were introduced by QuickChange site-directed mutagenesis. Nuclear staining was performed using DAPI (0.1 μg/ml). After washing, cover glasses were mounted to microscopy slides by Mowiol 4–88 mounting medium (Sigma-Aldrich). Confocal microscopy was carried out using a Zeiss LSM 710 system (Carl Zeiss Microscopy GmbH, Jena, Germany) with a 40× oil objective. Images were processed by the ImageJ software.

RhoA activation assay

The commercially available luminescence-based G-LISA RhoA activation assay (Cytoskeleton) was used to measure the GTP bound RhoA levels in cell cultures. 2×10^5 cells/well were seeded onto a 24-well plate. G-LISA assay was performed according to the manufacturer's recommendations, 24 h after transfection with the exception of the concentration and the antibody dilutions. Sample concentrations were equalized to 1 mg/ml. Primary and secondary antibodies were diluted to 1:500 and 1:1000, respectively. Luminescence signal was detected on a Synergy H4 plate reader (BioTek). The RSK inhibitor BI-D1870 treatment was performed at 100 nM for 1 h. The MEK inhibitor CI1040 was incubated ON at 100 nM. Inhibitor treatments were performed in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum. Serum stimulation (20%) was performed with serum-starved cells for 5 min.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.01.038.

Acknowledgments

This work was supported by the National Research Development and Innovation Office, Hungary: K 119359 (to L.N.), NN 114309 and KKP 126963 (to A.R.). G.G. and M.S. were supported through the New National Excellence Program of the Hungarian Ministry of Human Capacities. This work was completed in the ELTE Institutional Excellence Program (783-3/2018/FEKUTSRAT) supported by the Hungarian Ministry of Human Capacities. The work was also supported in part by the Ligue contre le cancer (équipe labellisée 2015 to G.T.), by the National Institutes of Health (Grant R01CA134737 to G.T.), by the European Union (PDZnet network, Marie Sklodowska-Curie Grant No. 675341 to G.T. and P.J.), and by the French Infrastructure for Integrated Structural Biology ANR-10-INBS-05.

Author Contributions: G.G. conceived the project, carried out the experiments, analyzed data, and wrote the paper. L.N., G.T., and A.R. supervised the research, analyzed data, and wrote the paper. F.B., P.J., Y.N., R.V., and G.T. performed and analyzed the holdup experiments. B.B.-K., C.K., V.B., and M.S. contributed by carrying out cell-based and *in vitro* experiments.

Conflict of Interest: There is no conflict of interest.

Received 10 December 2018;
Received in revised form 16 January 2019;
Accepted 28 January 2019
Available online 3 February 2019

Keywords:
phosphorylation;
PDZ;
RSK;
MAPK;
protein–protein interaction

Abbreviations used:
PBMs, PDZ-binding motifs; RSK, ribosomal S6 kinase; BIs, binding intensities; ITC, isothermal titration calorimetry; SPR, surface plasmon resonance; FP, fluorescence polarization; WT, wild type.

References

[1] V. Neduva, R.B. Russell, Linear motifs: evolutionary interaction switches, FEBS Lett. 579 (2005) 3342–3345, https://doi.org/10.1016/j.febslet.2005.04.005.

[2] K. Luck, S. Charbonnier, G. Travé, The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains, FEBS Lett. 586 (2012) 2648–2661, https://doi.org/10.1016/j.febslet.2012.03.056.
[3] M. Sheng, C. Sala, PDZ domains and the organization of supramolecular complexes, Annu. Rev. Neurosci. 24 (2001) 1–29, https://doi.org/10.1146/annurev.neuro.24.1.1.
[4] J. Saras, C.H. Heldin, PDZ domains bind carboxy-terminal sequences of target proteins, Trends Biochem. Sci. 21 (1996) 455–458, https://doi.org/10.1016/S0968-0004(96)30044-3.
[5] G.M. Thomas, G.R. Rumbaugh, D.B. Harrar, R.L. Huganir, Ribosomal S6 kinase 2 interacts with and phosphorylates PDZ domain-containing proteins and regulates AMPA receptor transmission, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15006–15011, https://doi.org/10.1073/pnas.0507476102.
[6] F.J. Sulzmaier, S. Young-Robbins, P. Jiang, D. Geerts, A.M. Prechtl, M.L.M.L. Matter, S. Kesari, J.W. Ramos, RSK2 activity mediates glioblastoma invasiveness and is a potential target for new therapeutics, Oncotarget 7 (2016) 79869–79884, https://doi.org/10.18632/oncotarget.13084.
[7] K.G. Hartman, M.I. Vitolo, A.D. Pierce, J.M. Fox, P. Shapiro, S.S. Martin, P.T. Wilder, D.J. Weber, Complex formation between s100b protein and the p90 ribosomal S6 kinase (RSK) in malignant melanoma is calcium-dependent and inhibits extracellular signalregulated kinase (ERK)-mediated phosphorylation of RSK, J. Biol. Chem. 289 (2014) 12886–12895, https://doi.org/10.1074/jbc.M114.561613.
[8] J. Pouysségur, V. Volmat, P. Lenormand, Fidelity and spatio-temporal control in MAP kinase (ERKs) signalling, Biochem. Pharmacol. 64 (2002) 755–763, https://doi.org/10.1016/S0006-2952(02)01135-8.
[9] A. Alexa, G. Gógl, G. Glatz, Á. Garai, A. Zeke, J. Varga, E. Dudás, N. Jeszenői, A. Bodor, C. Hetényi, A. Reményi, Structural assembly of the signaling competent ERK2–RSK1 heterodimeric protein kinase complex, Proc. Natl. Acad. Sci. 112 (2015) 2711–2716, https://doi.org/10.1073/pnas.1417571112.
[10] M. Frödin, C.J. Jensen, K. Merienne, S. Gammeltoft, A phosphoserine-regulated docking site in the protein kinase RSK2 that recruits and activates PDK1, EMBO J. 19 (2000) 2924–2934, https://doi.org/10.1093/emboj/19.12.2924.
[11] M. Frödin, T.L. Antal, B.a. Dümmler, C.J. Jensen, M. Deak, S. Gammeltoft, R.M. Biondi, A phosphoserine/threonine-binding pocket in AGC kinases and PDK1 mediates activation by hydrophobic motif phosphorylation, EMBO J. 21 (2002) 5396–5407, https://doi.org/10.1093/emboj/cdf551.
[12] Á. Garai, A. Zeke, G. Gógl, I. Törő, F. Ferenc, H. Blankenburg, T. Bárkai, J. Varga, A. Alexa, D. Emig, M. Albrecht, A. Reményi, Specificity of linear motifs that bind to a common mitogen-activated protein kinase docking groove, Sci. Signal. 5 (2012), ra74. https://doi.org/10.1126/scisignal.2003004.Specificity.
[13] G. Gógl, A. Alexa, B. Kiss, G. Katona, M. Kovács, A. Bodor, A. Reményi, L. Nyitray, Structural basis of ribosomal S6 kinase 1 (RSK1) inhibition by S100B protein: modulation of the extracellular signal-regulated kinase (ERK) signaling cascade in a calcium-dependent way, J. Biol. Chem. 291 (2015) 11–27, https://doi.org/10.1074/jbc.M115.684928.
[14] S. Kang, S. Dong, A. Guo, H. Ruan, S. Lonial, H.J. Khoury, T. Gu, J. Chen, Epidermal growth factor stimulates RSK2 activation through activation of the MEK/ERK pathway and Src-dependent tyrosine phosphorylation of RSK2 at Tyr-529, 283 (2008) 4652–4657, https://doi.org/10.1074/jbc.M709673200.
[15] G. Gógl, B. Biri-Kovács, Á.L. Póti, H. Vadászi, B. Szeder, A. Bodor, G. Schlosser, A. Ács, L. Turiák, L. Buday, A. Alexa, L. Nyitray, A. Reményi, Dynamic control of RSK complexes by phosphoswitch-based regulation, FEBS J. 285 (2018) 46–71, https://doi.org/10.1111/febs.14311.
[16] P.V. Hornbeck, B. Zhang, B. Murray, J.M. Kornhauser, V. Latham, E. Skrzypek, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations, Nucleic Acids Res. 43 (2015) D512–D520, https://doi.org/10.1093/nar/gku1267.
[17] K.N. Dalby, N. Morrice, F.B. Caudwell, J. Avruch, P. Cohen, Identification of regulatory phosphorylation sites in mitogen-activated protein kinase (MAPK)-activated protein kinase-1a/p90rsk that are inducible by MAPK, J. Biol. Chem. 273 (1998) 1496–1505, https://doi.org/10.1074/jbc.273.3.1496.
[18] G.N. Sundell, R. Arnold, M. Ali, P. Naksukpaiboon, J. Orts, P. Güntert, C.N. Chi, Y. Ivarsson, Proteome-wide analysis of phospho-regulated PDZ domain interactions, Mol. Syst. Biol. 14 (2018) 1–22, https://doi.org/10.15252/msb.20178129.
[19] R. Vincentelli, K. Luck, J. Poirson, J. Polanowska, J. Abdat, M. Blémont, J. Turchetto, F. Iv, K. Ricquier, M.-L. Straub, A. Forster, P. Cassonnet, J.-P. Borg, Y. Jacob, M. Masson, Y. Nominé, J. Reboul, N. Wolff, S. Charbonnier, G. Travé, Quantifying domain-ligand affinities and specificities by high-throughput holdup assay, Nat. Methods 12 (2015) 787–793, https://doi.org/10.1038/nmeth.3438.
[20] Y. Duhoo, V. Girault, J. Turchetto, L. Ramond, F. Durbesson, P. Fourquet, Y. Nominé, V. Cardoso, A.F. Sequeira, J.L.A. Brás, C.M.G.A. Fontes, G. Travé, N. Wolff, R. Vincentelli, High throughput production of a newly designed library of soluble human single and tandem PDZ domains allows semi-quantitative PDZ–peptide interaction screening through high throughput holdup assay., Methods Mol. Biol. in press (n.d.).
[21] G.-X. Shi, W.S. Yang, L. Jin, M.L. Matter, J.W. Ramos, RSK2 drives cell motility by serine phosphorylation of LARG and activation of Rho GTPases, Proc. Natl. Acad. Sci. 115 (2018) E190–E199, https://doi.org/10.1073/pnas.1708584115.
[22] L.R. Pearce, D. Komander, D.R. Alessi, The nuts and bolts of AGC protein kinases, Nat. Rev. Mol. Cell Biol. 11 (2010) 9–22, https://doi.org/10.1038/nrm2822.
[23] A.S. Dixon, M.K. Schwinn, M.P. Hall, K. Zimmerman, P. Otto, T.H. Lubben, B.L. Butler, B.F. Binkowski, T. MacHleidt, T.A. Kirkland, M.G. Wood, C.T. Eggers, L.P. Encell, K.V. Wood, NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells, ACS Chem. Biol. 11 (2016) 400–408, https://doi.org/10.1021/acschembio.5b00753.
[24] M. Cargnello, P.P. Roux, Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases, Microbiol. Mol. Biol. Rev. 75 (2011) 50–83, https://doi.org/10.1128/MMBR.00031-10.
[25] H.C. Lim, T.-S. Jou, Ras-activated RSK1 phosphorylates EBP50 to regulate its nuclear localization and promote cell proliferation, Oncotarget 7 (2016) 10283–10296, https://doi.org/10.18632/oncotarget.7184.
[26] J.A. Galan, K.M. Geraghty, G. Lavoie, E. Kanshin, J. Tcherkezian, V. Calabrese, G.R. Jeschke, B.E. Turk, B.A. Ballif, J. Blenis, P. Thibault, P.P. Roux, Phosphoproteomic analysis identifies the tumor suppressor PDCD4 as a RSK substrate negatively regulated by 14–3–3, Proc. Natl. Acad. Sci. 111 (2014) E2918–E2927, https://doi.org/10.1073/pnas.1405601111.
[27] A. Moritz, Y. Li, A. Guo, J. Villén, Y. Wang, J. MacNeill, J. Kornhauser, K. Sprott, J. Zhou, A. Possemato, J.M. Ren, P. Hornbeck, L.C. Cantley, S.P. Gygi, J. Rush, M.J. Comb, Akt–RSK–S6 kinase signaling networks activated by oncogenic receptor tyrosine kinases, Sci. Signal. 3 (2010)https://doi.org/10.1126/scisignal.2000998.

[28] D. Avey, S. Tepper, W. Li, Z. Turpin, F. Zhu, Phosphoproteomic analysis of KSHV-infected cells reveals roles of ORF45-activated RSK during lytic replication, PLoS Pathog. 11 (2015) 1–30, https://doi.org/10.1371/journal.ppat.1004993.
[29] E.B. Ünal, F. Uhlitz, N. Blüthgen, A compendium of ERK targets, FEBS Lett. 591 (2017) 2607–2615, https://doi.org/10.1002/1873-3468.12740.
[30] D. Shahbazian, P.P. Roux, V. Mieulet, M.S. Cohen, B. Raught, J. Taunton, J.W.B. Hershey, J. Blenis, M. Pende, N. Sonenberg, The mTOR/PI3K and MAPK pathways converge on eIF4B to control its phosphorylation and activity, EMBO J. 25 (2006) 2781–2791, https://doi.org/10.1038/sj.emboj.7601166.
[31] Y. Zhou, N. Yamada, T. Tanaka, T. Hori, S. Yokoyama, Y. Hayakawa, S. Yano, J. Fukuoka, K. Koizumi, I. Saiki, H. Sakurai, Crucial roles of RSK in cell motility by catalysing serine phosphorylation of EphA2, Nat. Commun. 6 (2015) 1–12, https://doi.org/10.1038/ncomms8679.
[32] R. Lara, M.J. Seckl, O.E. Pardo, The p90 RSK family members: common functions and isoform specificity, Cancer Res. 73 (2013) 5301–5308, https://doi.org/10.1158/0008-5472.CAN-12-4448.
[33] T. Houles, S.P. Gravel, G. Lavoie, S. Shin, M. Savall, A. Meant, B. Grondin, L. Gaboury, S.O. Yoon, J. St-Pierre, P.P. Roux, RSK regulates PFK-2 activity to promote metabolic rewiring in melanoma, Cancer Res. 78 (2018) 2191–2204, https://doi.org/10.1158/0008-5472.CAN-17-2215.
[34] M. Artamonov, K. Momotani, D. Utepbergenov, A. Franke, A. Khromov, Z.S. Derewenda, A.V. Somlyo, The p90 ribosomal S6 kinase (RSK) is a mediator of smooth muscle contractility, PLoS One 8 (2013), e58703. https://doi.org/10.1371/journal.pone.0058703.
[35] F. Cuello, A.K. Snabaitis, M.S. Cohen, J. Taunton, M. Avkiran, Evidence for direct regulation of myocardial Na+/H+ exchanger isoform 1 phosphorylation and activity by 90-kDa ribosomal S6 kinase (RSK): effects of the novel and specific RSK inhibitor fmk on responses to 1-adrenergic stimulation, Mol. Pharmacol. 71 (2006) 799–806, https://doi.org/10.1124/mol.106.029900.
[36] S. Kerrien, B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, M. Duesbury, M. Dumousseau, M. Feuermann, U. Hinz, C. Jandrasits, R.C. Jimenez, J. Khadake, U. Mahadevan, P. Masson, I. Pedruzzi, E. Pfeiffenberger, P. Porras, A. Raghunath, B. Roechert, S. Orchard, H. Hermjakob, The IntAct molecular interaction database in 2012, Nucleic Acids Res. 40 (2012) 841–846, https://doi.org/10.1093/nar/gkr1088.
[37] L.A. Harris, J.S. Hogg, J.J. Tapia, J.A.P. Sekar, S. Gupta, I. Korsunsky, A. Arora, D. Barua, R.P. Sheehan, J.R. Faeder, BioNetGen 2.2: advances in rule-based modeling, Bioinformatics 32 (2016) 3366–3368, https://doi.org/10.1093/bioinformatics/btw469.
[38] M.D. Larrea, F. Hong, S.A. Wander, T.G. da Silva, D. Helfman, D. Lannigan, J.A. Smith, J.M. Slingerland, RSK1 drives p27Kip1 phosphorylation at T198 to promote RhoA inhibition and increase cell motility, Proc. Natl. Acad. Sci. 106 (2009) 9268–9273, https://doi.org/10.1073/pnas.0805057106.
[39] E. Grabocka, P.B. Wedegaertner, Disruption of oligomerization induces nucleocytoplasmic shuttling of leukemia-associated rho guanine-nucleotide exchange factor, Mol. Pharmacol. 72 (2007) 993–1002, https://doi.org/10.1124/mol.107.035162.
[40] R. Kristelly, G. Gao, J.J.G. Tesmer, Structural determinants of RhoA binding and nucleotide exchange in leukemia-associated rho guanine-nucleotide exchange factor, J. Biol. Chem. 279 (2004) 47352–47362, https://doi.org/10.1074/jbc.M406056200.
[41] S.W. Pedersen, L. Albertsen, G.E. Moran, B. Levesque, S.B. Pedersen, L. Bartels, H. Wapenaar, F. Ye, M. Zhang, M.E. Bowen, K. Strømgaard, Site-specific phosphorylation of PSD-95 PDZ domains reveals fine-tuned regulation of protein–protein interactions, ACS 12 (2017) 2313–2323, https://doi.org/10.1021/acschembio.7b00361.
[42] J. Corzo, M. Santamaria, Time, the forgotten dimension of ligand binding teaching, Biochem. Mol. Biol. Educ. 34 (2006) 413–416, https://doi.org/10.1002/bmb.2006.494034062678.
[43] T. Clairfeuille, C. Mas, A.S.M. Chan, Z. Yang, M. Tello-Lafoz, M. Chandra, J. Widagdo, M.C. Kerr, B. Paul, R.D. Teasdale, N.J. Pavlos, V. Anggono, B.M. Collins, A molecular code for endosomal recycling of phosphorylated cargos by the SNX27–retromer complex, Nat. Struct. Mol. Biol. 23 (2016) 921–932, https://doi.org/10.1038/nsmb.3290.

Chapter 10

Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study

Status: Published in *Structure* as co-author.

10.1 Summary

Background: PTM in PBM/PDZ interactions are not always easy to measure. In the literature, it is usually try to solved the issue by using mimics [6,62,63]. In the case of RSK1, instead of the phosphorylation at p-3, a glutamic acid is used to compensate size and negative charge [6]. On the other hand, E6 protein of "high-risk" HPV has been proposed to switch their domain preferences to 14-3-3 domains when phosphorylated in a pivotal PBM recognition position like p-2 [118–120].

Results: We used the holdup assay and complementary FP measurements to screen the RSK1 wild-type RSK1 phosphorylation at p-3, RSK1 phosphomimic at p-3, E6 protein HPV16 in the wild-type, E6 phosphorylation at p-2 and E6 phosphomimic at p-2 against the whole PDZome. We obtain all the binding profiles and transformed them into K_d to make a better comparison of the PDZ domain preferences. This showed completely different BI profile between the phosphorylated RSK1 and the phosphomimic version. Lastly, we obtained the crystal structures of all the mentioned PBM in complex with MAGI1-2 and the E6 phosphorylated at

p-2 in complex with the 14-3-3σ domain. We obtained crystal structures for even those PBM/PDZ interaction that were unable to be measured experimentaly.

Conclusions: We showed the difference between altering pivotal PBM and regulatory sites. This showed the probable dynamic interplay between PDZ and 14-3-3 domains. The crystal structures solved for these PBM/PDZ interactions which were supposed to not bind may strength up the point of view of PPI interactions happening in the continuum. We lastly showed that the phosphomimic strategy might be suitable to reproduce clashes in pivotal positions but not for the regulatory sites.

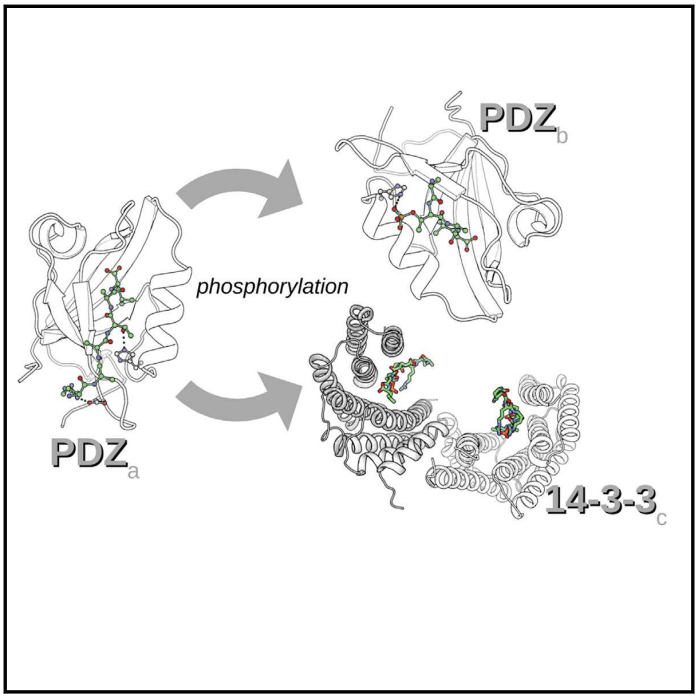
Contribution: I proposed the PBM to synthesize and conceived the first line of the topic to discuss. I wrote a preliminary stage draft of the paper before the introduction of the 14-3-3 topic. I participated in the final discussion of the results. I also performed the holdup analysis: I performed the data curation of the elecropherograms using the software presented in chapter 15, I processed the data and crosschecked them with previous results (chapter 9) and created the supplementary figures of the electropherogram data. I also calculated the conversion of Binding Intensities into the affinity scale.

Article

Structure

Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study

Graphical Abstract



Highlights

- A large proportion of PDZ-binding motifs are phosphorylatable
- Phosphorylated and phosphomimetic PBMs bind differently to PDZs and 14-3-3 proteins
- These differences are demonstrated by X-ray analysis and affinity profiling

Authors

Gergo Gogl, Pau Jane, Célia Caillet-Saguy, ..., Yves Nomine, Nikolai N. Sluchanko, Gilles Trave

Correspondence

goglg@igbmc.fr (G.G.), nikolai.sluchanko@mail.ru (N.N.S.), traveg@igbmc.fr (G.T.)

In Brief

Gogl et al. studied the effects of phosphorylation of PDZ domain-binding motifs. They demonstrated that there are many phosphorylatable and phosphorylated motifs, some of which are also putative binding targets of 14-3-3 proteins. Using quantitative interactomic assays and crystallography they showed how phosphorylation and phosphomimetic substitution alters their binding properties.

Gogl et al., 2020, Structure *28*, 1–13
June 2, 2020
https://doi.org/10.1016/j.str.2020.03.010

CellPress

Please cite this article in press as: Gogl et al., Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study, Structure (2020), https://doi.org/10.1016/j.str.2020.03.010

Structure
Article

Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study

Gergo Gogl,[1,6,7,*] Pau Jane,[1,6] Célia Caillet-Saguy,[2] Camille Kostmann,[1] Goran Bich,[1] Alexandra Cousido-Siah,[1] Laszlo Nyitray,[3] Renaud Vincentelli,[4] Nicolas Wolff,[2] Yves Nomine,[1] Nikolai N. Sluchanko,[5,*] and Gilles Trave[1,*]

[1]Equipe Labellisee Ligue 2015, Department of Integrated Structural Biology, Institut de Genetique et de Biologie Moleculaire et Cellulaire (IGBMC), INSERM U1258/CNRS UMR 7104/Universite de Strasbourg, 1 rue Laurent Fries, BP 10142, 67404 Illkirch, France
[2]Récepteurs-Canaux, Institut Pasteur, UMR 3571, CNRS, 75724 Paris, France
[3]Department of Biochemistry, ELTE Eotvos Lorand University, Budapest, Hungary
[4]Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS UMR 7257, Aix-Marseille Université, Marseille, France
[5]A.N. Bach Institute of Biochemistry, Federal Research Center of Biotechnology of the Russian Academy of Sciences, 119071 Moscow, Russia
[6]These authors contributed equally
[7]Lead Contact
*Correspondence: goglg@igbmc.fr (G.G.), nikolai.sluchanko@mail.ru (N.N.S.), traveg@igbmc.fr (G.T.)
https://doi.org/10.1016/j.str.2020.03.010

SUMMARY

Protein-protein interaction motifs are often alterable by post-translational modifications. For example, 19% of predicted human PDZ domain-binding motifs (PBMs) have been experimentally proven to be phosphorylated, and up to 82% are theoretically phosphorylatable. Phosphorylation of PBMs may drastically rewire their interactomes, by altering their affinities for PDZ domains and 14-3-3 proteins. The effect of phosphorylation is often analyzed by performing "phosphomimetic" mutations. Here, we focused on the PBMs of HPV16-E6 viral oncoprotein and human RSK1 kinase. We measured the binding affinities of native, phosphorylated, and phosphomimetic variants of both PBMs toward the 266 human PDZ domains. We co-crystallized all the motif variants with a selected PDZ domain to characterize the structural consequence of the different modifications. Finally, we elucidated the structural basis of PBM capture by 14-3-3 proteins. This study provides novel atomic and interactomic insights into phosphorylatable dual specificity motifs and the differential effects of phosphorylation and phosphomimetic approaches.

INTRODUCTION

Short linear motifs are peptide segments that are disordered in isolation yet fold upon complex formation with globular domains, thereby participating in protein-protein interaction (PPI) networks (Davey et al., 2012). Consensus sequence features help to define families of motifs, which generally correspond to families of domains that recognize these particular motifs (Kumar et al., 2019). Most domain-motif PPI networks are rather promiscuous, i.e., each individual domain can interact with numerous distinct motifs, and vice versa (Ivarsson and Jemth, 2019). Furthermore, domain-motif networks are often modulated by post-translational modifications (PTMs). The most abundant PTM is phosphorylation, a reversible biochemical reaction, catalyzed by protein kinases and reverse catalyzed by protein phosphatases, that transfers the γ-phosphoryl group of an ATP molecule to a receiver residue, most often to the hydroxyl group of a Ser/Thr, or Tyr residue via forming a phosphoester bond (Hunter, 2012). Phosphorylated amino acids have unique properties that can alter biochemical properties of substrate proteins in different ways. Many *in vitro* and *in cellulo* experiments involve "phosphomimetic" acidic (Glu/Asp) mutations that are easy to introduce by recombinant approaches and are meant to reproduce the biochemical effect of site-specific phosphorylation events, despite being chemically distinct (Sieracki and Komarova, 2013). Across evolution of orthologous proteins, acidic amino acids are often seen to replace phosphorylated sites, and conversely (Pearlman et al., 2011).

PDZs are globular protein domains displaying a conserved antiparallel β barrel fold composed of five to six β strands and one to two α helices. PDZ domains recognize short conserved PDZ-binding motifs (PBMs) mostly located at the extreme C-terminus of their target proteins (Songyang et al., 1997). The sequences of C-terminal PBMs fall into three main classes (Luck et al., 2012). The last C-terminal residue (position 0) is almost always hydrophobic (mainly, Leu/Val/Ile). The third last residue (position -2) can be Ser/Thr (class 1), Val/Tyr/Phe (class 2), or Asp/Glu (class 3). The human proteome contains ~266 PDZ domains (the PDZome) dispersed over ~150 proteins, and a few thousand putative PBMs (Luck et al., 2012). This creates an extensive PDZ/PBM interactome, which is often hijacked by viral intruder proteins bearing their own PBMs (Javier and Rice, 2011; Banks et al., 2012; James and Roberts, 2016). Many PBMs are potentially phosphorylatable (Sundell et al., 2018). The phosphorylation of a PBM may cause a general change in its "PDZome-binding profile," namely the list of binding strengths exhibited by the PBM toward each individual human PDZ domain. This was recently demonstrated for ribosomal protein S6 kinase 1 (RSK1), a kinase from the Ras/ERK-MAPK pathway

Please cite this article in press as: Gogl et al., Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study, Structure (2020), https://doi.org/10.1016/j.str.2020.03.010

Table 1. Ser/Thr Phosphorylatable PBMs: Summary of Our Bioinformatic Analysis of Human PBMs

	Class 1 PBM		Class 2 PBM		Class 3 PBM		All 3 Classes	
	Consensus	No. of Motifs	Consensus	No. of Motifs	Consensus	No. of Motifs	No. of Motifs	
	[ST]x[LVI]$	956	[VYF]x[LVI]$	458	[ED]x[LVI]$	512	1926	
	Phosphorylatable Class 1 PBM		Phosphorylatable Class 2 PBM		Phosphorylatable Class 3 PBM		All 3 Classes	
Position of Modification	Consensus	No. of Motifs	Consensus	No. of Motifs	Consensus	No. of Motifs	No. of Motifs	% of Total
-1	[ST][**ST**][LVI]$	172 (43)	[VYF][**ST**][LVI]$	73 (13)	[ED][**ST**][LVI]$	59 (15)	304	15.8
-2	[**ST**]x[LVI]$	956 (113)	not phosphorylatable by Ser/Thr kinases				956	49.6
-3	[**ST**][ST]x[LVI]$	138 (12)	[**ST**][VYF]x[LVI]$	80 (13)	[**ST**][ED]x[LVI]$	63 (9)	281	14.6
-4	[**ST**]x[ST]x[LVI]$	134 (18)	[**ST**]x[VYF]x[LVI]$	56 (5)	[**ST**]x[ED]x[LVI]$	64 (13)	254	13.2
-5	[**ST**]x(2)[ST]x[LVI]$	144 (26)	[**ST**]x(2)[VYF]x[LVI]$	59 (8)	[**ST**]x(2)[ED]x[LVI]$	57 (13)	260	13.5
-6	[**ST**]x(3)[ST]x[LVI]$	165 (29)	[**ST**]x(3)[VYF]x[LVI]$	65 (10)	[**ST**]x(3)[ED]x[LVI]$	67 (14)	297	15.4
-7	[**ST**]x(4)[ST]x[LVI]$	138 (24)	[**ST**]x(4)[VYF]x[LVI]$	57 (14)	[**ST**]x(4)[ED]x[LVI]$	70 (14)	265	13.8
-8	[**ST**]x(5)[ST]x[LVI]$	120 (24)	[**ST**]x(5)[VYF]x[LVI]$	65 (12)	[**ST**]x(5)[ED]x[LVI]$	76 (15)	261	13.6
-9	[**ST**]x(6)[ST]x[LVI]$	137 (33)	[**ST**]x(6)[VYF]x[LVI]$	37 (6)	[**ST**]x(6)[ED]x[LVI]$	56 (9)	230	11.9
-10	[**ST**]x(7)[ST]x[LVI]$	123 (26)	[**ST**]x(7)[VYF]x[LVI]$	50 (9)	[**ST**]x(7)[ED]x[LVI]$	71 (16)	244	12.7
No. of phosphorylatable motifs		956		291		331	1,578	81.9
No. of phosphorylated motifs		218		65		92	375	19.5
No. of phosphorylatable sites		2,227		542		583	3,352	
No. of phosphorylated sites		348		90		118	556	

Putative phosphorylatable PBMs were searched using SLiMSearch (Krystkowiak and Davey, 2017), using a disorder score cutoff of 0.3. Note that we used the most restrictive definition of PBM consensus motifs according to the ELM database (Kumar et al., 2019) and only focused on C-terminal PBMs. Numbers in parentheses indicate the number of motifs that were found to be phosphorylated in low- or high-throughput mass spectrometry datasets in the PhosphoSite database (Hornbeck et al., 2015). Phosphorylatable motifs contain at least a single Ser/Thr residue at their last 11 C-terminal sequences. Phosphorylated motifs are those found at least once in the PhosphoSite database on at least one site. Phosphorylatable Ser/Thr residues within consensus motifs are highlighted in bold. $ denotes the C terminus (-COOH). For further details, see Table S1.

which is regulated by auto-phosphorylation of its own C-terminal PBM (Gógl et al., 2019). In addition, phosphorylation of a PBM can also alter its interactions with other protein families. For example, the 14-3-3 family, which contains seven members in humans, can also recognize C-terminal motifs in a phosphorylation-dependent manner (Coblitz et al., 2006). Worthy of note, viral E6 oncoproteins of high-risk mucosal human papillomavirus (HPV) types responsible for cervical cancers (Suarez and Trave, 2018) comprise C-terminal PBMs, which are subjected to phosphorylation events prone to modulate their interaction with PDZ domains and 14-3-3 proteins (Boon and Banks, 2013).

PDZ/PBM interactions display weak affinities, with low-micromolar dissociation constants for the best binders (Luck et al., 2011). Thus, accurately measuring the steady-state dissociation constant of a PDZ/PBM interaction can be challenging. We developed a high-throughput experimental approach, called the holdup assay, that can accurately measure such weak interactions (with a limit of quantitation of up to 100–150 μM dissociation constant) (Vincentelli et al., 2015). We have specially adapted the holdup assay to the PDZ domain family, by cloning and expressing the 266 known human domains (Duhoo et al., 2019). The approach allows us to measure complete PDZome-binding affinity profiles of any peptide sequence with high sensitivity.

Here, we combined quantitative interactomics and crystallography to investigate the differential effects of phosphorylation and phosphomimetic substitution of two phosphorylatable PBMs found in the HPV16 E6 viral oncoprotein and the human RSK1 kinase. To this aim, we measured all the affinities of wild-type, phosphorylated, and phosphomimetic versions of RSK1 and E6 PBMs toward the 266 human PDZ domains. We crystallized wild-type and modified RSK1 and E6 PBMs with the second PDZ domain of MAGI1 (MAGI1_2). Finally, we solved the crystal structure of phosphorylated E6 PBM bound to 14-3-3σ.

RESULTS

Proteome-wide Identification of Ser/Thr Phosphorylatable PBMs

To evaluate the extent of human PBMs potentially modulatable by phosphorylation, we searched the human proteome with the "SLiMSearch" program (Krystkowiak and Davey, 2017) using the most restrictive definition of a PBM, defined by the "ELM" linear motif database (Kumar et al., 2019). This way, we identified 1,926 human PBMs (956 class 1, 458 class 2, and 512 class 3). In Table 1, we counted for each PBM class the number of potentially phosphorylatable Ser/Thr sites at different positions of these PBMs, as well as the numbers of such sites that have experimentally been proven to be phosphorylated *in cellulo* according to the PhosphoSite database (Hornbeck et al., 2015). We found that up to 82% (1,578) of the 1,926 predicted human PBMs bear one or more

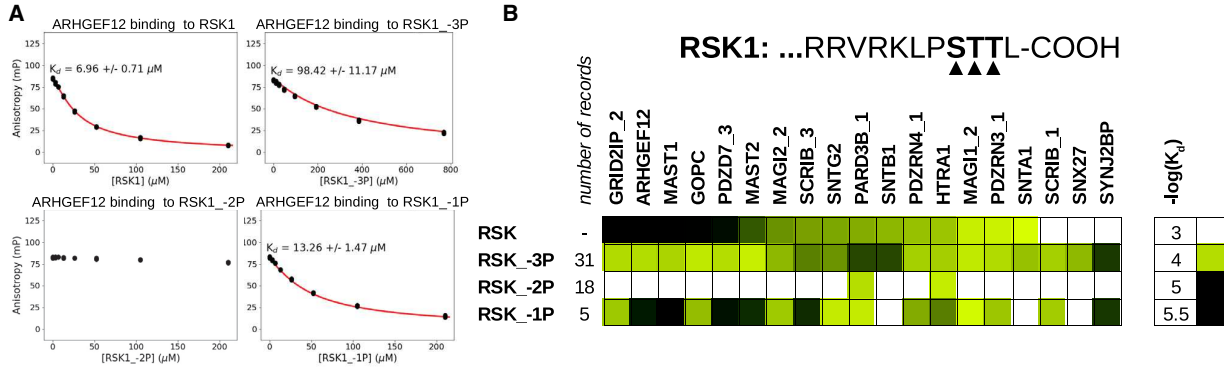


Figure 1. Affinity Measurements of Phosphorylated RSK1 Peptides with PDZ Domains

(A) Competitive fluorescence polarization assay was used to monitor 19 RSK interaction partners against four versions of the RSK1 PBM, native, or phosphorylated at position -1, -2, or -3. The interaction between the RSK1 and ARHGEF12 PDZ domains is shown as an example; all other data are shown in Figure S1.

(B) Heatmap of the measured interactions against the different, position-specific phospho-peptides. Based on the records of reported instances on the PhosphoSitePlus database, and also supported experimentally (Hornbeck et al., 2015; Gógl et al., 2018), RSK is most often modified at the -3 position and least often at the -1 position. Detectable interactions of RSK1 were modulated by phosphorylation at position -3 and -1 and were mostly eliminated by phosphorylation at position -2. The heatmap is colored by affinities, according to the scale bar on the right side.

potential phosphorylatable Ser/Thr sites between positions -1 and -10, and that 19% (375) have been found to be phosphorylated at least once on at least one site. In total, the predicted human PBMs contained, between positions -1 and -10, 3,352 potential and 556 experimentally proven potential phosphorylatable Ser/Thr sites (see also Table S1 for further details concerning a chosen subset of phosphorylatable class 1 motifs). Thus, phosphorylation of PBMs turns out to be very common, pointing to a dynamic regulation of their PPI interactome.

Class 1 PBMs, which systematically comprise a phosphorylatable Ser/Thr residue at position -2, are twice as frequent as class 2 or class 3 motifs and they harbor four times as many phosphorylatable sites. Worthy of note, phosphorylation of position -2 introduces an acidic charge, thereby creating a "pseudo-class 3 PBM," and its phosphomimetic substitution creates a class 3 PBM. The most common phosphorylated sites, proportionally to their occurrence, can be found at the position -1 of class 1 and class 2 PBMs. More than 25% of these sites (56 out of 245) are found to be phosphorylated *in cellulo*. In contrast, the least phosphorylated sites can be found at position -3 of class 1 PBMs. Less than 9% of these sites (12 out of 138) are found to be phosphorylated *in cellulo*. The most abundant PBM phosphorylation site affects position -2 of class 1 motifs with 113 experimentally proven instances.

The above list of phosphorylatable and/or phosphorylated PBMs is probably not exhaustive. As mentioned before, we used restrictive consensus motifs that may exclude several known PBMs (Vaccaro and Dente, 2002). For example, some functional PBMs may have a Cys, Met, and other residues at position 0 (Thomas et al., 2016), and others are not even positioned at the C terminus of proteins. Also, a number of phosphorylatable PBMs may only be modified under special conditions that have never been experimentally addressed. Finally, some phosphorylated PBMs may be difficult to detect by mass spectrometry (Lucrèce et al., 2011). For example, the HPV-E6 oncoproteins, whose expression is essential for HPV-transformed cells, such as HeLa, comprise a phosphorylatable PBM that is addressed in the present work. However, E6 oncoproteins are only expressed at a very low amount in cells and have a basic residue at position -4. Thus, a standard proteolytic digestion is expected to yield small amounts of a very small fragment, making it difficult to identify by mass spectrometry. Furthermore, the sequence of E6, being a viral protein, is not always present in standard lists of human proteins, so that a database search on the human proteome might skip E6-derived fragments.

Phosphorylation at Distinct Sites of the RSK1 PBM Differentially Impacts Its Binding Affinity for a Panel of PDZ Domains

The RSK1 kinase harbors a C-terminal class 1 PBM (…RRVRKLPSTTL-COOH) (Thomas et al., 2005). RSK1 can autophosphorylate its C-terminal tail at positions -3, -2, and -1, leading to a rearrangement in its PDZ specificity (Gógl et al., 2018). In addition, in the PhosphoSite database RSK1 PBM is the class 1 PBM most frequently phosphorylated at the -3 position (Table S1). Here, we used RSK1 to assess the potential impact of phosphorylation of different positions of a class 1 PBM on its binding affinity for PDZ domains. Based on previous results (Gógl et al., 2019), we selected 19 PDZ domains and tested their interactions with competitive fluorescence polarization assay against all possible (mono) phosphorylated RSK1 peptides (Figures 1 and S1). Within the detection range of fluorescence polarization, phosphorylation at position -2 apparently abolished most interactions, whereas phosphorylation at positions -3 and -1 rather modulated their binding affinities. Therefore, in terms of phosphorylation, the -2 position has a higher impact on PDZ interactions than the -3 and -1 positions. This corroborates a previous study that analyzed the effect of phosphorylation on the binding properties of 100 putative phosphorylatable PBMs, assayed against three distinct PDZ domains (Boisguerin et al., 2007).

PDZome-Binding Profiles of Native, Phosphorylated, and Phosphomimetic PBMs

For further studies, we set out to investigate at proteome-wide level how phosphorylation and phosphomimetic mutations

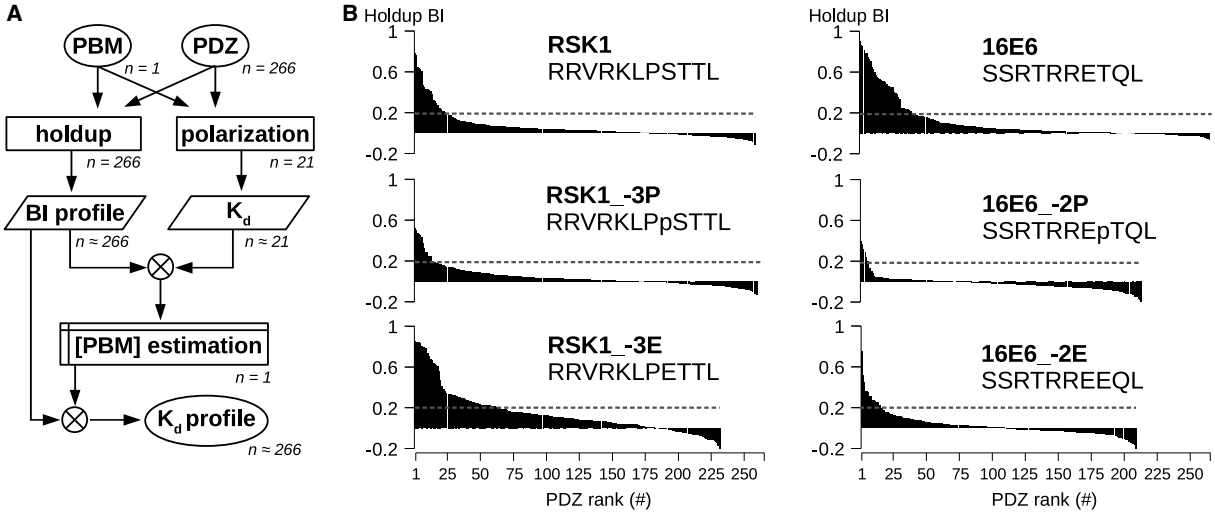


Figure 2. PDZome-Binding (BI) Profiles of the Studied PBMs, Measured by the Holdup Assay

(A) The general flowchart of the holdup assay and the conversion of binding intensities into dissociation constants with the aid of competitive fluorescence polarization assays. The holdup assay generates a BI profile that can be converted into steady-state dissociation constants, if we know the approximate peptide concentration during the holdup experiment. This concentration can be estimated using dissociation constants of a small set of interactions, determined by an orthogonal approach.

(B) BI profiles of the studied PBMs. Because each profile is ordered by decreasing BI values, the order of PDZ domains differs between panels. Gray dashed lines represent the conservative threshold of accurate binding quantitation by holdup assay (BI = 0.2).

directed at a higher- or a lower-impact position of class 1 PBMs would alter their full PDZome interactome. This question was addressed using the phosphorylatable PBMs of RSK1 kinase and of HPV16 E6 oncoprotein (herein defined as 16E6). The PBM of 16E6 (… SSRTRRETQL-COOH) harbors phosphorylatable sites at positions -2, -6, -8, and -9, and it was proven to be phosphorylated by various kinases at position -2 (e.g., by basophilic kinases due to a basic patch at position -5, such as PKA) (Kühne et al., 2000). To generate PDZome-binding profiles of RSK1 and 16E6 PBM variants, we performed holdup assays (Vincentelli et al., 2015) using an updated version of our PDZ library, which includes individual clones of all possible 266 PDZ domains as MBP-fused proteins (Duhoo et al., 2019). We measured the binding profiles of the 16E6 peptides (native, SSRTRRETQL; phosphorylated, SSRTRREpTQL; phosphomimetic, SSRTRREEQL) against all human PDZ domains. In the case of RSK1, we had previously determined the binding profiles of the native (RRVRKLPSTTL) and the phosphorylated (RRVRKLPpSTTL) peptides (Gógl et al., 2019). Here, we measured the binding profile of the phosphomimetic peptide (RRVRKLPETTL) along with an additional reference profile of the phosphorylated peptide (Figures 2 and S2; Table S2).

In a holdup experiment, a cell lysate containing an overexpressed PDZ domain of known concentration is incubated with a peptide-saturated resin, the mixture is rapidly filtered, and the remaining PDZ concentration in the filtrate is measured. The experiment provides, for each PDZ domain, a steady-state depletion factor (binding intensity, or BI), that can in principle be converted into a steady-state dissociation constant. This conversion is necessary if we need to compare multiple binding profiles, as each peptide might reach a different concentration during resin saturation. Estimating the dissociation constant requires access to three concentrations: free PDZ, free peptide, and PDZ-peptide complex. As stated above, the holdup assay delivers, for each PDZ-PBM pair, the concentrations of free PDZ and complex, while the concentration of free peptide remains unknown. However, using an orthogonal approach, we can obtain steady-state dissociation constants for a subset of PDZ-PBM pairs. We can then use these dissociation constants to back-calculate the peptide concentration in the holdup assay, which is expected to be the same for all PDZ-PBM pairs in that assay. We used a competitive fluorescence polarization assay and measured the binding affinities of the 6 studied peptides against 21 purified PDZ domains (Figure S1) (Roehrl et al., 2004). In the holdup assay, we had already determined the corresponding BI values for most of these 126 interactions. We complemented the binding profiles with 9 interactions that we only measured with the fluorescence polarization assay in Table S2. For each given interaction, where both a quantifiable (>0.2) BI value and a dissociation constant were available, we calculated the apparent peptide concentration present in the holdup assay. Then, we used the average peptide concentrations obtained in that way to convert the original BI profiles into profiles displaying actual dissociation constants (Figure 3A). In the case of the modified 16E6 peptides, we only detected very weak interactions with the holdup assay that we failed to detect with fluorescence polarization. For these peptides, we used the average of the other peptide concentrations (of RSK1s and 16E6) for the conversion.

Holdup experiments for 16E6_-2P, 16E6_-2E, and RSK1_-3E were performed in singlicate. These singlicate holdup runs provided highly reliable data, as shown by the strong agreement of dissociation constants obtained from holdup assays and fluorescence polarization assays (Figure 3B). Based on our previous experience, the holdup assay is highly sensitive with a limit of

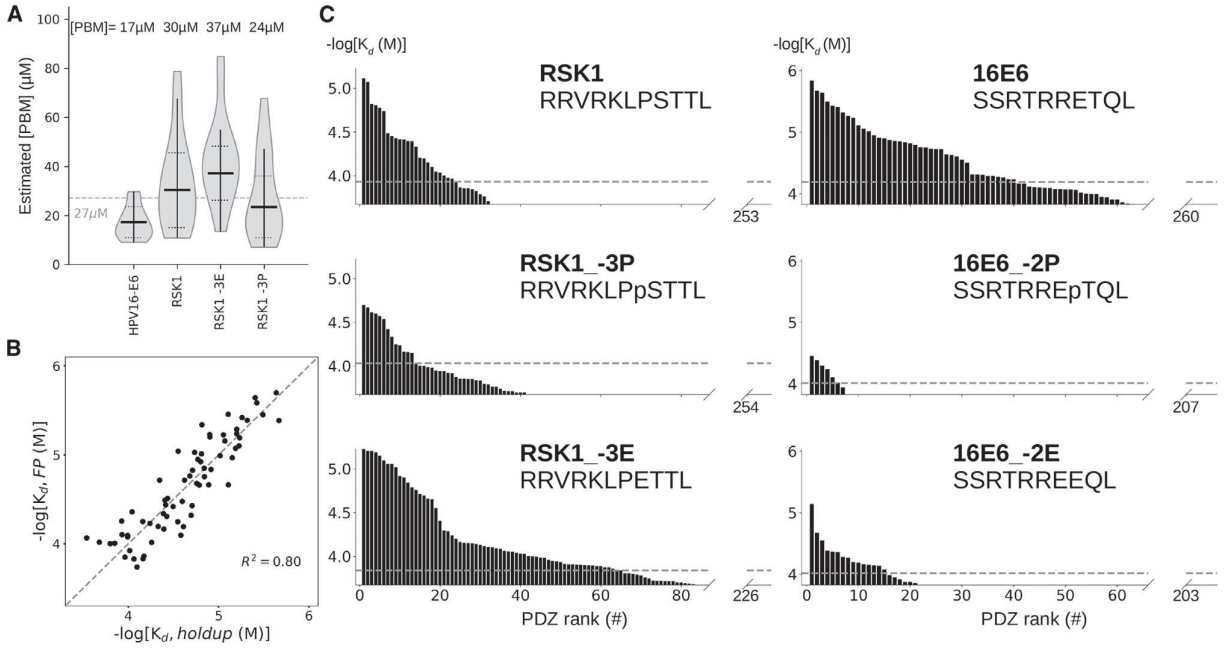


Figure 3. PDZome-Binding (K_d) Profiles of the Studied PBMs, Measured by the Holdup Assay

(A) Determined PBM peptide concentrations in the holdup assay, based on competitive fluorescence polarization experiments. Vertical lines in the violin plot show the minimal and maximal observed peptide concentration after outlier rejection. Horizontal lines show the mean and the standard deviation of the peptide concentration after outlier rejection. The gray dashed line shows the average peptide concentration that was used to convert the 16E6_-2P, 16E6_-2E profiles.

(B) Correlation of dissociation constants determined by the two orthogonal biochemical methods.

(C) K_d profiles of the studied PBMs. Because each profile is ordered by K_d values, the order of PBZ domains differs between panels. For clarity, we omitted most of the undetected binders. Gray lines represent the limit of accurate binding quantitation by the holdup assay (BI = 0.2).

detection of BI = 0.1 (10% PDZ depletion in solution) when experiments are run in triplicate (Vincentelli et al., 2015). Here, based on our comparison with fluorescence polarization, and on the fact that the holdup assays were run in singlicate, we set a conservative lower limit of quantitation at BI = 0.2 (20% PDZ depletion in solution), roughly corresponding to a 100 μM dissociation constant (Figure 3C).

Interactomic Consequence of Phosphorylation or Phosphomimetics of RSK1

Using the holdup assay, we were able to measure the interaction of RSK1 against 259, 260, and 232 PDZ domains for the native, phosphorylated, and phosphomimetic peptides, accounting for 97.4%, 97.7%, and 87.2% of coverage of the complete human PDZome, respectively (Figure 4). Using the threshold of BI = 0.2 for a quantifiable interaction, we detected 23, 13, and 67 significant interactions for the native, phosphorylated, and phosphomimetic peptides, respectively (8.9%, 5%, and 28.9% of the tested PDZome). Phosphorylation at position -3 generates a decrease in overall affinity and reorganizes the preferences of the RSK1 PBM for individual PDZ domains. Compared with the native RSK1 PBM, phosphorylated RSK1 retains, loses, and gains 10, 13, and 3 detectable partners, respectively. ARHGEF12 (an important substrate of RSK) (Shi et al., 2018) is among the lost binders, while SYNJ2BP is among the gained ones.

The phosphomimetic mutation also reshuffles the individual PDZ binding preferences of the native RSK1 PBM, but in contrast to phosphorylation it globally enhances the affinities in the binding profile, resulting in a larger number of detected interaction partners. Compared with the native RSK1 PBM, the phosphomimetic PBM retains, loses, and gains 22, 1, and 45 detectable partners, respectively. Overall these data point to strong differences between the phosphorylated and phosphomimetic RSK1 PBMs. Although phosphorylation strongly reorganizes the RSK1 PBM interactome, with numerous losses and gains of binders detected within a given affinity range, the phosphomimetic mutation expands the RSK1 interactome by increasing the number of detected binders without almost any loss. Far from accurately mimicking the phosphorylated state, the phosphomimetic peptide is essentially a very promiscuous peptide that binds all partners of both native and phosphorylated RSK1 PBM, plus numerous additional ones.

Structural Consequence of Phosphorylation or Phosphomimetics of RSK1

Previously, we determined crystal structures of the MAGI1_2 PDZ domain bound to the native RSK1 and RSK1_-3P peptides (Gógl et al., 2018). The PDZ domain was fused to Annexin A2 to enhance crystal formation, which enabled us to gain molecular insight into these PDZ/PBM complexes (Ecsédi et al., 2020). We used the same construct to obtain the complex with the

Please cite this article in press as: Gogl et al., Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study, Structure (2020), https://doi.org/10.1016/j.str.2020.03.010

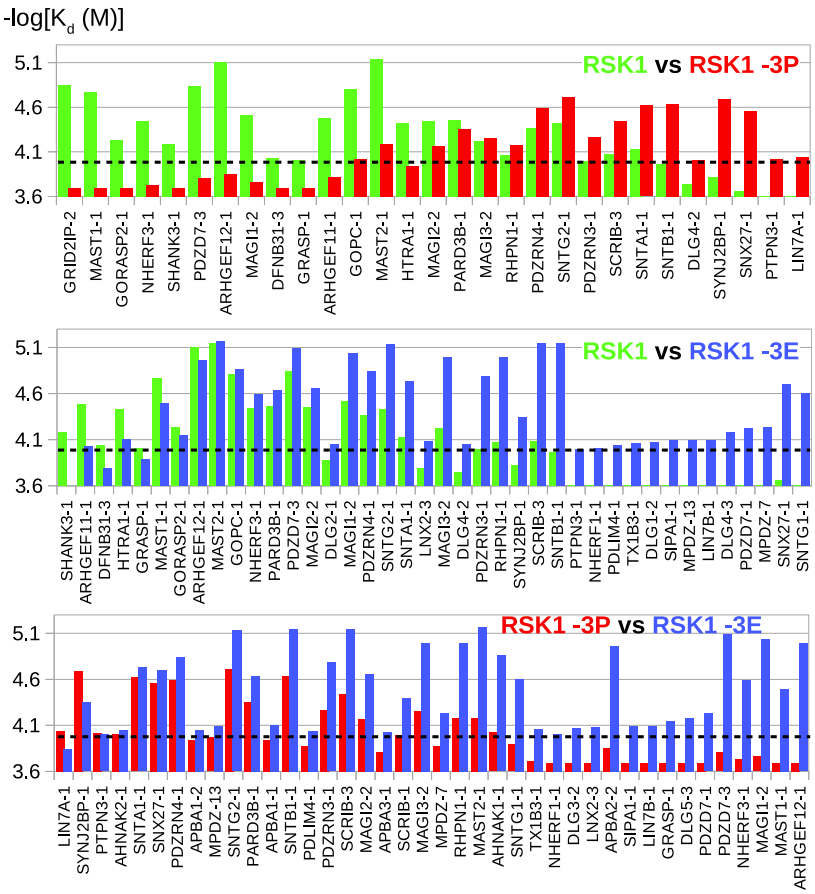


Figure 4. Pairwise Rearrangements in PDZome-Binding K_d Profiles of the Studied RSK1 PBMs

The PDZs are ordered by the determined fold-change (ΔΔG) between the two compared motifs. Dashed line represents the averaged limit of accurate binding quantitation by the holdup assay (BI = 0.2; [PBM] = 27 μM; K_d = 105 μM).

phosphomimetic RSK1 peptide (Figure 5; Table S3). In the crystal, MAGI1_2 exhibits the common PDZ fold consisting of five β strands and two α helices. The RSK1 peptides bind to the conventional binding groove of the PDZ domain (outlined by β-2, α-2, and the carboxylate-binding "GLGF loop").

The RSK1 peptides establish all canonical interactions of a class 1 PBM (Hung and Sheng, 2002). As a C-terminal anchor, the C terminus of the peptide interacts with the carboxylate-binding loop and the peptide adopts an antiparallel β strand conformation, which complements the exposed β-2 strand of the PDZ fold. A Thr residue at position -2 of the PBM mediates a hydrogen bond with His530 from the α-2 of the PDZ domain. Upon phosphorylation, this conformation is maintained, but some minor local conformation changes were observed around the phosphorylation site. Lys499 is located on the β-3 strand of the PDZ and its side chain is solvent-exposed, facing the -3 position of a PBM. As already observed in previous works (Zhang et al., 2007; Charbonnier et al., 2011), Lys499 is at an ideal position to mediate specific interactions with PBMs. Indeed, an interaction was observed between this residue and the Ser at position -3 of the RSK1 PBM (with a distance of 2.9 Å). Upon phosphorylation, this phospho-Ser residue was captured in two different rotamer conformations (in two closely related crystals), but both of them formed an interaction with the side chain of Lys499 (with a distance of 2.7–3.5 Å) (Gógl et al., 2018). The phosphomimetic peptide forms a similar interaction with the same amine group (with a distance of 2.2 Å).

Interactomic Consequence of Phosphorylation or Phosphomimetics of HPV16 E6

Using the holdup assay, we were able to measure the interaction of 16E6 PBM against 266 PDZ domains for the native peptide, 213 for the phospho-peptide, and 209 for the phosphomimetic peptide (Figure 6). These account to 100.0%, 80.1%, and 78.6% of coverage of the complete human PDZome, respectively. Using the limit of quantitation of our singlicate holdup assay (BI = 0.2, corresponding to a dissociation constant of 60–100 μM), we detected 41 significant interactions for the native-, 6 for the phospho-, and 15 for the phosphomimetic peptide (15.4%, 2.8%, and 7.2% of the tested PDZome, respectively). Moreover, the detected interactions of the modified peptides were markedly weaker compared with the average binding affinity of the native motif. In many cases, we failed to confirm the binding of these weak partners of the phosphorylated and the phosphomimetic peptides with the fluorescence polarization assay (Figures S1 and S3). Thus, we hardly detected any significant interaction partners upon modification of the HPV16-E6 motif at the -2 position.

Upon phosphorylation, 16E6 lost all of the experimentally significant binders of the native peptide with the exception of the weak binder FRMPD4 (Figure S3). Upon phosphomimetic substitution, 16E6 lost 31 significant binders of the native peptide. In addition, only a small overlap (consisting of 3 detected binders) was detected between the identified weak interaction partners of the phospho- and the phosphomimetic 16E6 PBM. An analogous phosphomimetic substitution on the 18E6 PBM was previously found to dramatically reduce its ability to interact with SNX27 (Ganti et al., 2016).

Structural Consequences of Phosphorylation or Phosphomimetics of HPV16 E6

MAGI1_2 is one of the strongest PDZ partners of 16E6 but, upon 16E6 phosphorylation at position -2, its affinity fell below the detection limits of both holdup and standard fluorescence polarization assays (Figure S1; Table S2). To estimate the very weak affinity of this complex, we repeated the competitive

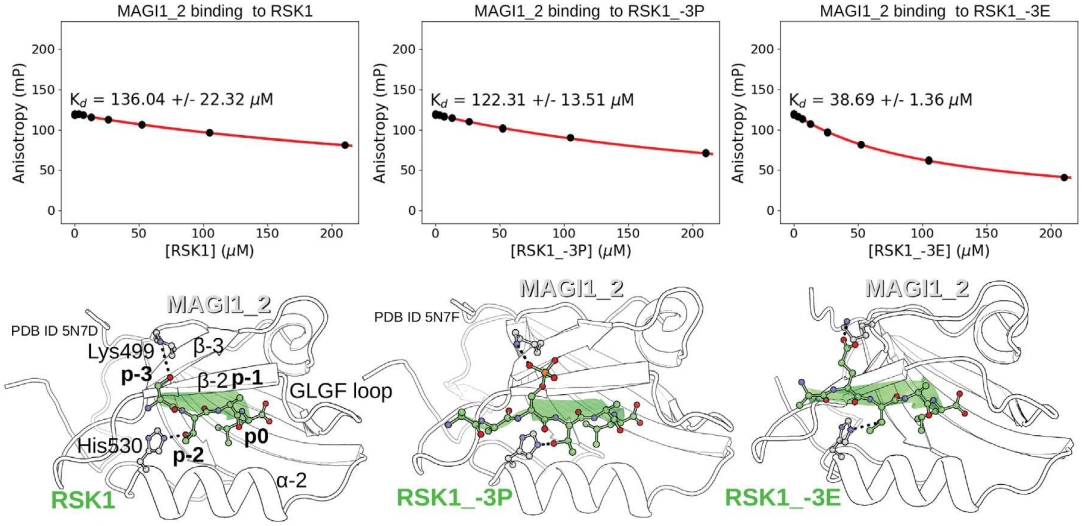


Figure 5. Crystal Structures Show the Molecular Consequence of RSK1 Phosphorylation on MAGI1_2 PDZ Binding

Isomorphous MAGI1_2-bound RSK1 peptides are shown in parallel with their affinities. Competitive fluorescence polarization assay was used to monitor the binding of RSK1 peptides to this PDZ domain (see Figure S2 for further details). RSK1 is modified at its position -3, facing Lys499 of MAGI1. The interaction between these residues remains possible, independently of the presence of the phosphorylation or the mutation. The crystal structures of MAGI1_2, bound to RSK1 (left panel) and RSK1_-3P (middle panel) were solved in our previous work (PDB: 5N7D and 5N7F) (Gógl et al., 2018). See Table S3 for statistical details about crystallographic data collection and refinement.

fluorescence polarization experiment by extending the range of titration from 200 μM to 2.5 mM. Under this modified condition, we managed to estimate the dissociation constants both for 16E6_-2P ($K_d \sim 2.5$ mM) and for 16E6_-2E ($K_d \sim 1$ mM) (Figure 7). Despite these fairly weak associations, we managed to crystallize both complexes, as well as a third complex, including wild-type 16E6 peptide, using the Annexin A2-fused MAGI1_2 construct. To push the reaction toward complex formation, we used an overall PDZ concentration of 135 μM (just as before), supplemented with 2 mM of modified peptides. In a crystal of this fused PDZ domain, the concentration of the PDZ site is approximately 11 mM (based on the typical unit cell dimensions and space group of the obtained crystals). In this situation, approximately 44% complex formation should be expected with the phosphorylated peptide. However, this calculation does not take into account synergistic effects within the crystal, which might significantly increase the fraction of complex formed. Indeed, co-crystallization resulted in isomorphous crystal with high peptide occupancy.

The 16E6 peptide adopts a classical PBM fold, satisfying all the required interactions with the PDZ domain of MAGI1 (Figure 7; Table S3). As described previously in detail (Charbonnier et al., 2011; Zhang et al., 2007), the 16E6 PBM mediates several bonds with the PDZ domain, including a side-chain-mediated contact between Glu494 of MAGI1 (of the β-2, β-3 loop) and an Arg from position -5 of the PBM. In contrast, in the phosphorylated complex the classical β strand structure of the PBM is disrupted. Here, the only interaction reminiscent of a class 1 motif is that involving the C terminus of the PBM and the GLGF loop of the PDZ. The Thr residue at position -2 is unable to mediate a bond with its hydroxy group with His530 from the α-2, because it is part of the phosphoester bond. This is somewhat compensated by a weak interaction between one of the oxygen atoms of the phosphate group and the imidazole ring of His530. The main chain of the PBM has moved away, impeding the β strand conformation of the bound peptide. The structure of the phosphomimetic peptide is similar, although not identical. We do not observe any direct contact between the acidic residue and His530, but the interactions of the peptide are similarly limited to the extreme C terminus. Thus, both the phosphorylation and the phosphomimetic mutation at position -2 resulted in a significant alteration in the bound β conformation of the PBM.

Phosphorylated PBMs Are Also Potential 14-3-3 Binding Motifs

C-Terminal motif binding is not an exclusive property of PDZ domains. For example, PBM phosphorylation can also create an alternative binding site for 14-3-3 proteins (Espejo et al., 2017). To test the possibility of these alternative interactions, we measured the interactions of 14-3-3σ with the PBM of RSK1, a kinase that was previously shown to interact with 14-3-3 proteins (Cavet et al., 2003). Although RSK1 did not detectably interact, RSK1_-3P showed indications of a weak phosphorylation-dependent interaction, and both RSK1_-2P and RSK1_-1P bound strongly to the tested 14-3-3σ (Figures 8A and S4). RSK1_-2P and RSK1_-1P are, according to experimental data, the minor PBM autophosphorylation sites of RSK1 (Figure 1) (Hornbeck et al., 2015; Gógl et al., 2018). All three phosphopeptides fit to the documented consensus of mode III 14-3-3-binding motifs (Figure 8B). 14-3-3 interactions are centered around the phosphate moiety, and mode III 14-3-3-binding motifs are C-terminal motifs phosphorylated at position -1, -2, or sometimes -3 relative to the C terminus, with Pro being excluded immediately after the phospho-residue (Kumar et al., 2019; Panni et al., 2011; Sluchanko, 2018). The latter

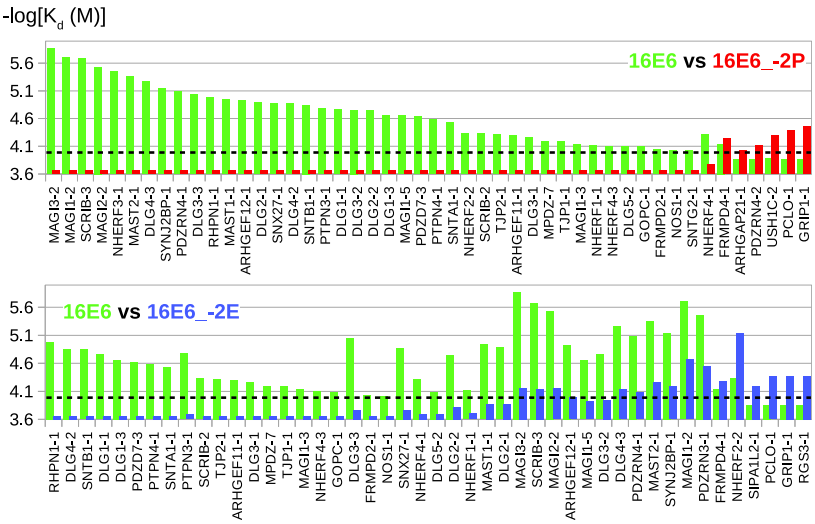


Figure 6. Pairwise Rearrangements in PDZome-Binding K_d Profiles of the Studied 16E6 PBMs

The PDZs are ordered by the determined K_d ratio ($\Delta\Delta G$) between the two compared motifs. Dashed line represents the average limit of accurate binding quantitation by the holdup assay (BI = 0.2; [PBM] = 27 μM; K_d = 105 μM).

restriction excluding proline only affects class 1 PBMs that are phosphorylated at position -2 (Figure 8B; Table S1). Among a large number of PBMs phosphorylatable at such positions, only 29 putative and 5 detectably phosphorylated PBMs do not satisfy the requirements of the mode III 14-3-3-binding consensus (34 in total out of all 956 class 1 motifs). Thus, all the remaining phosphorylatable PBMs (1,507 putative motifs) are also potential 14-3-3 binders (Figure 8C).

Crystal Structure of 14-3-3σ Protein Bound to Phosphorylated HPV16 E6 PBM

The HPV E6 oncoproteins were already described as interaction partners of several 14-3-3 proteins (Boon and Banks, 2013; Boon et al., 2015). We found that the C-terminal PBM of 16E6 can mediate phosphorylation-dependent interaction with 14-3-3σ, a well-characterized member of the 14-3-3 family (Figure 9A). We went on to solve their co-crystal structure using a crystallization-optimized protein (Sluchanko et al., 2017). As observed for most 14-3-3 binding motifs, the phosphate group is very well coordinated by the 14-3-3σ protein (Figure 9B) (Obsil and Obsilova, 2011; Sluchanko et al., 2017). It interacts with several residues of the 14-3-3 protein (e.g., Arg56, Arg129, and Tyr130) and it is also stabilized intra-molecularly by an Arg residue from position -4 of the same PBM peptide. In addition, the peptide forms several main-chain-mediated bonds, including an interaction between the C terminus of the PBM and Lys122 of 14-3-3σ. The Arg residue at position -5 of the PBM is involved in a stacking interaction with Arg60 of 14-3-3σ.

In contrast to the phosphorylated 16E6 PBM, the phosphomimetic 16E6 PBM did not detectably interact with 14-3-3σ (Figures 9A and S4), confirming previous data showing that a phosphomimetic 18E6 PBM failed to interact with 14-3-3ζ (Boon and Banks, 2013). In line with these observations, our phospho-PBM-bound 14-3-3 structure shows a strict coordination of the phosphate group. An acidic mutation would be inadequate to mediate the same mode of binding, as already observed for other binding partners of 14-3-3 proteins (Zheng et al., 2003).

DISCUSSION

New Perspectives in Quantitative Interactomics

In this work, we have quantitatively assessed a very large number of affinities, for ~1,500 distinct PDZ-peptide pairs. Although we obtained binding constants for ~170 pairs, we also showed that the remaining ~1,330 pairs displayed affinities below our quantitation threshold. Both types of data are important for building our understanding of interaction networks. Although the former inform us on the "interactome," the latter inform us on its complementary, often neglected side: the "negatome," comprising all non-favored interactions in the network.

The Consequences of PBM Phosphorylation

Although the human proteome contains thousands of putative phosphorylation sites within PBMs (Table 1), the only case when a PTM may target a high-impact, key motif position for PDZ binding are class 1 motifs phosphorylatable at position -2. All other modification sites affect lower-impact, modulatory positions.

Our results demonstrate that phosphorylation at position -3 of the RSK1 PBM reorganizes its PDZome-binding profile by increasing its affinity for some PDZ domains and decreasing it for others. This rearrangement in the RSK1 interactome can be measured both *in vitro* and *in cellulo* (Gógl et al., 2019). Phosphorylation at key position -2 of 16E6 outlines a more drastic effect as it suppresses most of the detectable interactions of the native motif, as shown by our structural data, by disrupting the β conformation of the bound peptide. This finding supports the previously described phospho-dependent disruption of different E6 PBMs with several PDZ domains and the phospho-regulation of other, non-viral PBMs (Boon et al., 2015; Cohen et al., 1996).

Although phosphorylation at a key motif position can disrupt molecular interactions to a degree that makes them undetectable with standard methods, phosphorylation at lower-impact sites tends to preserve the overall bound conformation, while modulating affinity and specificity profiles. As an example, the PBM of β2AR (… CSTNDSLL-COOH) harbors three phosphorylatable residues (Clairfeuille et al., 2016). Although the native β2AR PBM binds moderately to the PDZ domain of SNX27, its variants phosphorylated at modulatory positions -5 and -6 bind more strongly, and the variant phosphorylated at key position -2 binds significantly more weakly. Yet, solution NMR experiments showed that the chemical environment of the carboxylate-binding GLGF loop was altered in presence of the -2 phosphorylated

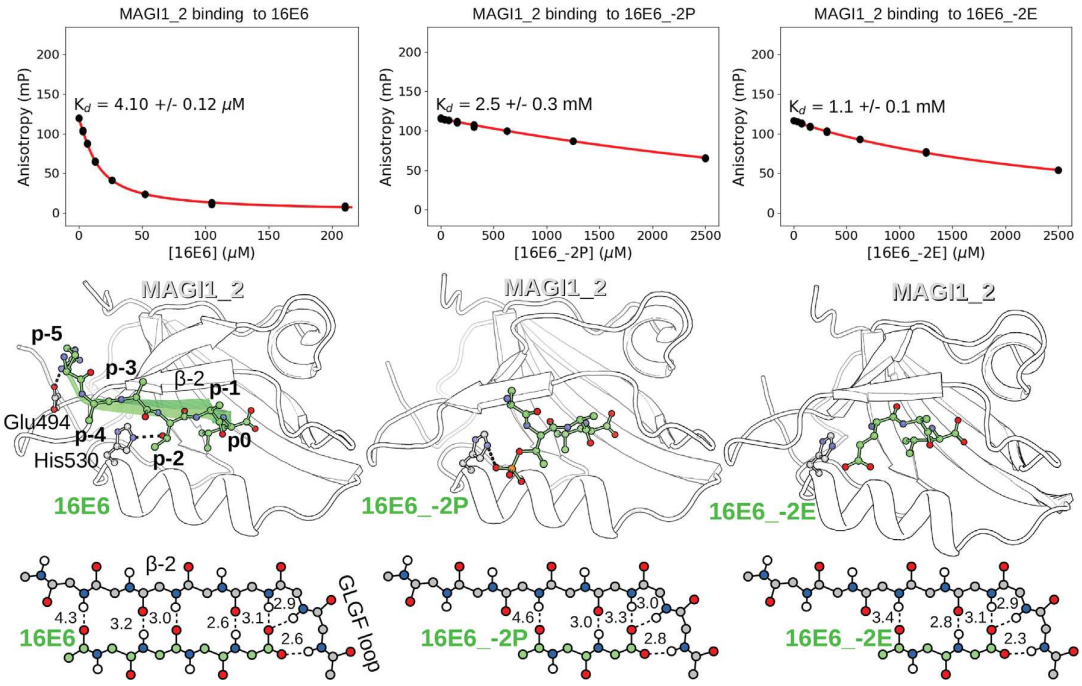


Figure 7. Crystal Structures Show the Molecular Consequence of 16E6 Phosphorylation on MAGI1_2 PDZ Binding

Isomorphous MAGI1_2-bound 16E6 peptides are shown in parallel with their affinities. Competitive fluorescence polarization assay was used to monitor the binding of 16E6 peptides to this PDZ domain (see Figure S1 for further details). 16E6 is modified at its position -2, which is involved in a bond with His530 in most class 1 PBZ/PBM complexes. Upon phosphorylation, this interaction is eliminated and only a minor contact remains possible between the His residue and the phosphate group. In the weak complexes of these modified peptides, the secondary structure of the peptide is also disturbed and the peptide seems to not adopt an optimized β strand conformation. The lower panels show the main-chain-mediated contacts (distances are given between non-hydrogen atoms). See Table S3 for statistical details of crystallographic data collection and refinement.

PBM, indicating that phosphorylation at key position -2 still allowed a residual binding mode compatible with our crystal structure of the MAGI1_2 bound 16E6_-2P peptide.

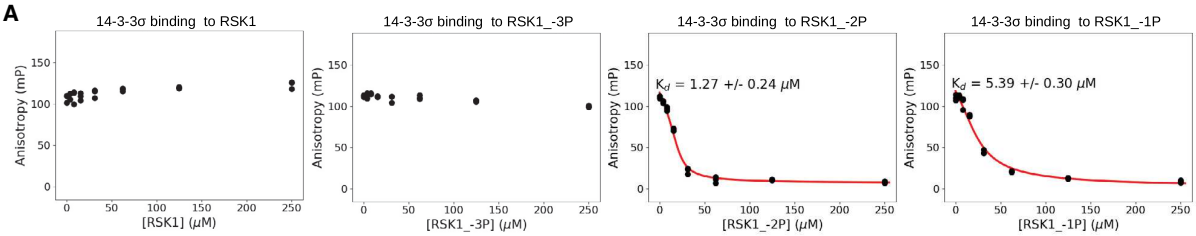
Out of all the putative PBM phosphorylation sites analyzed in this study, we have found that 1,507 PBMs are also putative binders of the 14-3-3 family, thereby presenting characteristics of dual specificity motifs (Van Roey et al., 2012). In these interactomic rearrangements, phosphorylation has a strong impact. Although some 14-3-3 binding affinities may seem moderate (like that of the 16E6/14-3-3σ interaction documented here), they may be relevant *in cellulo* since 14-3-3 isoforms are present at a high concentration in many tissues (Boston et al., 1982). Thus, an extensive dynamic interplay is likely to occur between the PDZ and 14-3-3 interaction networks. Moreover, phosphorylatable PBMs that are capable of binding PDZ and 14-3-3 proteins are also expected to bind transiently to kinases and phosphatases. That a motif bears information for the recognition by at least four different protein families has particular evolutionary consequences, since each interaction mode should impose its own constraints. This is remarkably illustrated by the HPV16 E6 PBM, which we captured here in crystal structures of two different types of complexes, one with a PDZ family member and one with a 14-3-3 family member. In both types of complexes, the same conserved residue of E6, namely Arg at position -5 of the PBM, was found to establish critical interactions, yet of a different nature. Furthermore, the same E6 PBM was, in our hands, very efficiently phosphorylated by protein kinase A (Figure S3). This kinase, as well as other members of the kinome, preferably acts on target consensus sequence presenting an Arg residue three positions upstream of the target threonine residue, i.e., precisely at the position of Arg-5 (Sarabia-Vega and Banks, 2019). The PBM of HPV16 E6 viral oncoprotein thus appears to have evolved sequence features that place it at a crossroad of four important protein families participating in the intricate, dynamic PPI network of the host organism.

PBM phosphorylation may affect not only the bound conformation, but also the free form of the motif. In the case of RSK1, we already showed that the free phosphorylated peptide adopts a transient structure in solution, where the phosphate group is involved in interactions with its preceding basic residues (Gógl et al., 2018). This intra-molecular interaction not only masks the strong negative charge of the phospho-residue, but also masks the site against inter-molecular interaction partners as it introduces an extra conformational selection step in the binding process. Such effects modulate both the kinetics and thermodynamics of the interaction. In the case of a phosphomimetic substitution, such transient charge-clamp interaction might not be as stable, resulting in a more accessible pseudo-phosphorylated PBM.

When Are Phosphomimetics a Lesser Evil Strategy?

Here, we presented a comprehensive quantitative interactomic approach to study the global effect of a PTM and its mimetic mutant

A



B

position	*-3*	*-2*	*-1*	*0*	*$*
class 1	x	[ST]	x	[LVI]	$
class 2	x	[VYF]	x	[LVI]	$
class 3	x	[ED]	x	[LVI]	$
mode III^{0}	x	x	[ST]	[^P]	$
mode III^{1}	x	[ST]	[^P]	x	$
mode III^{2}	[ST]	[^P]	x	x	$

C

	Phosphorylatable class 1 PBM		Phosphorylatable class 2 PBM	Phosphorylatable class 3 PBM
Position	**No. motifs (PhosphoSite)**	**Putative mode III 14-3-3 binder**	**Putative mode III 14-3-3 binder**	**Putative mode III 14-3-3 binder**
-1	172 (43)	172 (43)	73 (13)	59 (15)
-2	956 (113)	922 (108)	*not possible*	*not possible*
-3	138 (12)	138 (12)	80(13)	63 (9)
-4	134 (18)	*not possible*	*not possible*	*not possible*

Figure 8. Phosphorylated PBMs Are Potential 14-3-3 Binding Partners

(A) The phosphorylation of RSK1 PBM (… RKLPSTTL-COOH), affecting positions -1, -2, and -3, creates a putative mode III 14-3-3 binding site. Competitive fluorescence polarization assay was used to monitor the binding of RSK1 peptides to this 14-3-3 isoform (see Figure S4 for further details). Phosphorylation of these sites indeed increases RSK1 binding affinity to the 14-3-3σ isoform.

(B) The consensus of mode III 14-3-3 binding motifs and PBMs share some similarities. Mode III motifs require the phosphate group at positions -1, -2, or -3 and does not allow a Pro residue immediately after the phospho-residue. Positions are numbered based on the PBM numbering scheme. Numbers in the superscript indicates the linker between the phospho-residue and the C-terminal residue.

(C) According to the limitation presented in (B), most of the phosphorylatable PBMs (modified at positions -1, -2, or -3) fit into the consensus of mode III 14-3-3 binding motifs and only some class 1 motifs, phosphorylated at their -2 position, are excluded as a potential 14-3-3 binder. Numbers in parentheses indicate the number of motifs that were found to be phosphorylated in low- or high-throughput mass spectrometry datasets. $ denotes the C terminus (-COOH).

alternative on the PDZome-binding profiles of two PBMs. We further refined our previously described approach, which consists in experimentally measuring the affinities of all the possible interactions within the studied PPI network. This approach allows not only to demonstrate that phosphorylation can alter some particular interactions but also to exhaustively analyze, in quantitative terms, phosphorylation-dependent dynamics of the PDZ-PBM interactome. We hope that such approaches, which address domain-motif affinities and specificities and their potential PTM-induced changes for all possible interactions, rather than for a few selected ones, will be progressively adopted by the research community.

In an *in vitro* experiment, the phosphomimetic approach can be avoided by using a synthetic phospho-peptide, by phosphorylating a purified protein with a kinase, or by directly incorporating phospho residues during translation (Rogerson et al., 2015; Zhu et al., 2019). This is more difficult to achieve in a cellular assay. One can either activate specific pathways with external stimuli or treat the cells with phosphatase inhibitors to exceed basal phosphorylation levels of the target. However neither of these strategies will result in a pure, homogeneous phospho-state. Moreover the effect will not be specific for a single PTM site, since a huge number of phosphorylation or dephosphorylation events may simultaneously hit other sites within the same protein and in other ones. Because of its ease of use, the introduction of acidic residues to replace phosphorylated Ser/Thr residues will probably remain common in cellular (and even in *in vitro*) assays (Caria et al., 2019; Baliova and Jursky, 2019; Sundell et al., 2018).

Our study should raise further awareness about how the chemical discrepancy between phosphorylation and phosphomimetic substitution is prone to lead to quantitative binding discrepancies on an interactomic scale. Although a phosphomimetic substitution might be sufficient to reproduce the effect of a steric, structural clash induced by phosphorylation, it is likely to fail to imitate phosphorylation events that introduce novel interactions or even just modulate them, as exemplified here with 16E6 and RSK1, respectively. This limitation of the phosphomimetic strategy has already been stressed in other studies (Toto et al., 2017; Sundell et al., 2018). In addition, one should keep in mind that each motif can have multiple interaction partners (such as various PDZ domains and 14-3-3 proteins in the case of a PBM), due to their promiscuous binding properties. Although a phosphomimetic mutation may sometimes reproduce the effect of a phosphorylation for one particular interaction, it will fail to do so for many others. After all, a carboxyl group is not chemically identical to a phosphate ester.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - 1. MBP-PDZ Library Preparation, Peptide Synthesis
  - 2. Holdup Assay

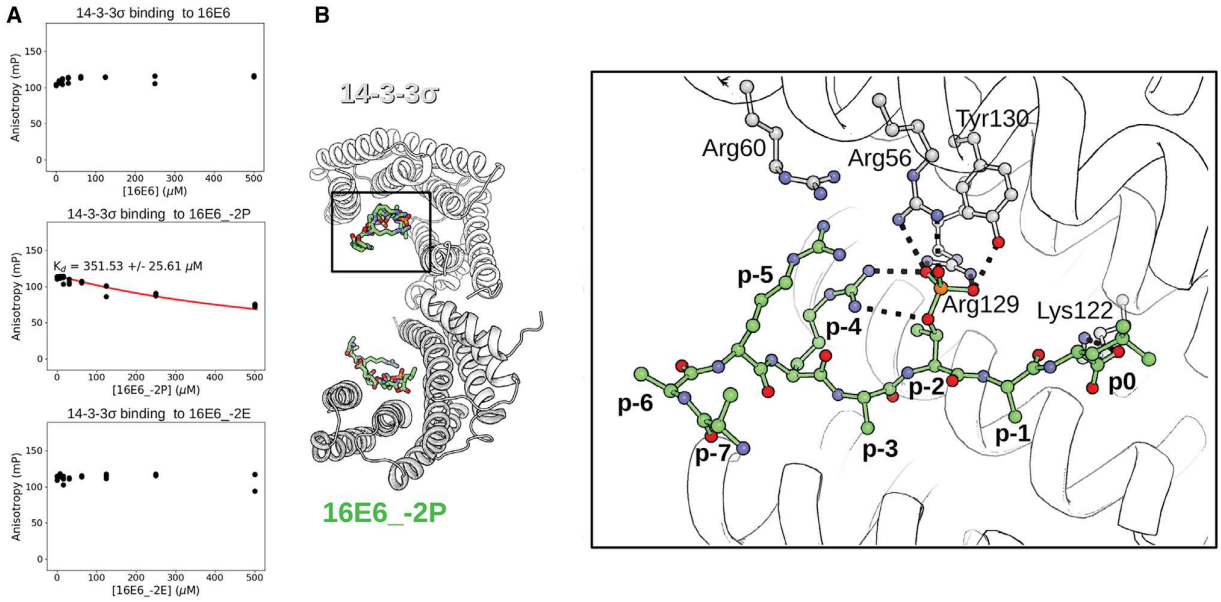


Figure 9. The Crystal Structure between 14-3-3σ and 16E6_-2P Shows the Molecular Details of a Phosphorylated PBM Bound to a 14-3-3 Protein

(A) Competitive fluorescence polarization assay was used to monitor the binding of 16E6 peptides to this 14-3-3 isoform (see Figure S4 for further details). 16E6 PBM can interact with 14-3-3σ in a phosphorylation-dependent manner.

(B) Crystal structure of 14-3-3σ bound to phosphorylated 16E6. 14-3-3 proteins form homodimers that capture two identical phosphopeptides. The inset shows the molecular determinants of the interaction. The phosphorylated residue is tightly coordinated and the C terminus of the peptide interacts with Lys122 of 14-3-3σ.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.str.2020.03.010.

ACKNOWLEDGMENTS

We thank Lawrence Banks for the shared plasmids and the useful discussions. We give thanks for the support of the Swiss Light Source synchrotron (P. Scherrer Institute, Villigen, Switzerland) and the help of the beam-scientist at the PXIII beamline. The work was also supported by the Ligue Contre le Cancer (Équipe Labellisée 2015 to G.T.), the French Infrastructure for Integrated Structural Biology (FRISBI), Instruct-ERIC, and by the European Union (PDZnet network, Marie Sklodowska-Curie grant no. 675341 to G.T. and P.J.). G.G. was supported by the Post-doctorants en France program of the Fondation ARC. N.N.S. is grateful to Kristina V. Tugaeva for help with sample preparation, purification and characterization and for general discussions, as well as to the Russian Science Foundation for the grant no. 19-74-10031.

AUTHOR CONTRIBUTIONS

G.G. carried out the fluorescence polarization experiments, performed the crystallographic studies, analyzed the data, and wrote the paper. P.J. analyzed the data and performed bioinformatic analysis. P.J., C.C.-S., C.K., G.B., and N.W. performed and analyzed the holdup experiments. A.C.-S. and N.N.S. contributed to the crystallographic experiments. Y.N., L.N., R.V., and N.N.S. contributed to data analysis and data interpretation. G.T. supervised the research, analyzed the data and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

Received: February 4, 2020
Revised: March 12, 2020
Accepted: March 20, 2020
Published: April 14, 2020

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

Baliova, M., and Jursky, F. (2019). Phosphomimetic mutation of glycine transporter GlyT1 C-terminal PDZ binding motif inhibits its interactions with PSD95. J. Mol. Neurosci. *70*, 488–493.

Banks, L., Pim, D., and Thomas, M. (2012). Human tumour viruses and the deregulation of cell polarity in cancer. Nat. Rev. Cancer *12*, 877–886.

Boisguerin, P., Ay, B., Radziwill, G., Fritz, R.D., Moelling, K., and Volkmer, R. (2007). Characterization of a putative phosphorylation switch: adaptation of SPOT synthesis to analyze PDZ domain regulation mechanisms. ChemBioChem *8*, 2302–2307.

Boon, S.S., and Banks, L. (2013). High-risk human papillomavirus E6 oncoproteins interact with 14-3-3ζ in a PDZ binding motif-dependent manner. J. Virol. *87*, 1586–1595.

Boon, S.S., Tomaić, V., Thomas, M., Roberts, S., and Banks, L. (2015). Cancer-causing human papillomavirus E6 proteins display major differences in the phospho-regulation of their PDZ interactions. J. Virol. *89*, 1579–1586.

Boston, P.F., Jackson, P., and Thompson, R.J. (1982). Human 14-3-3 protein: radioimmunoassay, tissue distribution, and cerebrospinal fluid levels in patients with neurological disorders. J. Neurochem. *38*, 1475–1482.

Caria, S., Stewart, B.Z., Jin, R., Smith, B.J., Humbert, P.O., and Kvansakul, M. (2019). Structural analysis of phosphorylation-associated interactions of human MCC with Scribble PDZ domains. FEBS J. *286*, 1–16.

Cavet, M.E., Lehoux, S., and Berk, B.C. (2003). 14-3-3β is a p90 ribosomal S6 kinase (RSK) isoform 1-binding protein that negatively regulates RSK kinase activity. J. Biol. Chem. *278*, 18376–18383.

Charbonnier, S., Nominé, Y., Ramírez, J., Luck, K., Chapelle, A., Stote, R.H., Travé, G., Kieffer, B., and Atkinson, R.A. (2011). 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.

Clairfeuille, T., Mas, C., M Chan, A.S., Yang, Z., Tello-Lafoz, M., Chandra, M., Widagdo, J., Kerr, M.C., Paul, B., Teasdale, R.D., et al. (2016). A molecular code for endosomal recycling of phosphorylated cargos by the SNX27-retromer complex. Nat. Struct. Mol. Biol. *23*, 921–932.

Coblitz, B., Wu, M., Shikano, S., and Li, M. (2006). C-terminal binding: an expanded repertoire and function of 14-3-3 proteins. FEBS Lett. *580*, 1531–1535.

Cohen, N.A., Brenman, J.E., Snyder, S.H., and Bredt, D.S. (1996). Binding of the inward rectifier K+ channel Kir 2.3 to PSD-95 is regulated by protein kinase A phosphorylation. Neuron *17*, 759–767.

Davey, N.E., Van Roey, K., Weatheritt, R.J., Toedt, G., Uyar, B., Altenberg, B., Budd, A., Diella, F., Dinkel, H., and Gibson, T.J. (2012). Attributes of short linear motifs. Mol. Biosyst. *8*, 268–281.

Duhoo, Y., Girault, V., Turchetto, J., Ramond, L., Durbesson, F., Fourquet, P., Nominé, Y., Cardoso, V., Sequeira, A.F., Brás, J.L.A., et al. (2019). High throughput production of a newly designed library of soluble human single and tandem PDZ domains allows semi-quantitative PDZ-peptide interaction screening through high throughput holdup assay. Methods Mol. Biol. *2025*, 439–476.

Ecsédi, P., Gógl, G., Hóf, H., Kiss, B., Harmat, V., and Nyitray, L. (2020). Structure determination of the transactivation domain of p53 in complex with S100A4 using annexin A2 as a crystallization chaperone. Sneak Peek. https://doi.org/10.2139/ssrn.3509911.

Espejo, A.B., Gao, G., Black, K., Gayatri, S., Veland, N., Kim, J., Chen, T., Sudol, M., Walker, C., and Bedford, M.T. (2017). PRMT5 C-terminal phosphorylation modulates a 14-3-3/PDZ interaction switch. J. Biol. Chem. *292*, 2255–2265.

Ganti, K., Massimi, P., Manzo-merino, J., Tomai, V., Pim, D., and Banks, L. (2016). Interaction of the human papillomavirus E6 oncoprotein with sorting nexin 27 modulates endocytic cargo transport pathways. PLoS Pathog *12*, 1–22.

Gógl, G., Biri-Kovács, B., Póti, Á.L., Vadászi, H., Szeder, B., Bodor, A., Schlosser, G., Ács, A., Turiák, L., Buday, L., et al. (2018). Dynamic control of RSK complexes by phosphoswitch-based regulation. FEBS J. *285*, 46–71.

Gógl, G., Biri-Kovács, B., Durbesson, F., Jane, P., Nomine, Y., Kostmann, C., Bilics, V., Simon, M., Reményi, A., Vincentelli, R., et al. (2019). Rewiring of RSK-PDZ interactome by linear motif phosphorylation. J. Mol. Biol. *431*, 1234–1249.

Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. *43*, D512–D520.

Hung, A.Y., and Sheng, M. (2002). PDZ domains: structural modules for protein complex assembly. J. Biol. Chem. *277*, 5699–5702.

Hunter, T. (2012). Why nature chose phosphate to modify proteins. Philos. Trans. R. Soc. B Biol. Sci. *367*, 2513–2516.

Ivarsson, Y., and Jemth, P. (2019). Affinity and specificity of motif-based protein-protein interactions. Curr. Opin. Struct. Biol. *54*, 26–33.

James, C.D., and Roberts, S. (2016). Viral interactions with PDZ domain-containing proteins—an oncogenic trait? Pathogens *5*, 1–22.

Javier, R.T., and Rice, A.P. (2011). Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. J. Virol. *85*, 11544–11556.

Kabsch, W. (2010). Xds. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Krystkowiak, I., and Davey, N.E. (2017). SLiMSearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. Nucleic Acids Res. *45*, W464–W469.

Kühne, C., Gardiol, D., Guarnaccia, C., Amenitsch, H., and Banks, L. (2000). Differential regulation of human papillomavirus E6 by protein kinase A: conditional degradation of human discs large protein by oncogenic E6. Oncogene *19*, 5884–5891.

Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J.A., Bukirova, D., Čalyševa, J., et al. (2019). ELM—the eukaryotic linear motif resource in 2020. Nucleic Acids Res. *48*, 1–11.

Luck, K., Fournane, S., Kieffer, B., Masson, M., Nominé, Y., and Travé, G. (2011). Putting into practice domain-linear motif interaction predictions for exploration of protein networks. PLoS One *6*, e25376.

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.

Lucrèce, M., Emmanuelle, S., Fabienne, B., Sandrine, S., Olivier, L., and Gérard, B. (2011). Sequence-dependent enrichment of a model phosphopeptide: a combined MALDI-TOF and NMR study. Anal. Chem. *83*, 3003–3010.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. *40*, 658–674.

Narayana, N., Cox, S., Shaltiel, S., Taylor, S.S., and Xuong, N.H. (1997). Crystal structure of a polyhistidine-tagged recombinant catalytic subunit of cAMP-dependent protein kinase complexed with the peptide inhibitor PKI(5-24) and adenosine. Biochemistry *36*, 4438–4448.

Obsil, T., and Obsilova, V. (2011). Structural basis of 14-3-3 protein functions. Semin. Cell Dev. Biol. *22*, 663–672.

Panni, S., Montecchi-Palazzi, L., Kiemer, L., Cabibbo, A., Paoluzi, S., Santonico, E., Landgraf, C., Volkmer-Engert, R., Bachi, A., Castagnoli, L., et al. (2011). Combining peptide recognition specificity and context information for the prediction of the 14-3-3-mediated interactome in *S. cerevisiae* and *H. sapiens*. Proteomics *11*, 128–143.

Pearlman, S.M., Serber, Z., and Jr, J.E.F. (2011). A mechanism for the evolution of phosphorylation sites. Cell *147*, 934–946.

Roehrl, M.H.A., Wang, J.Y., and Wagner, G. (2004). A general framework for development and data analysis of competitive high-throughput screens for small-molecule inhibitors of protein-protein interactions by fluorescence polarization. Biochemistry *43*, 16056–16066.

Van Roey, K., Gibson, T.J., and Davey, N.E. (2012). Motif switches: decision-making in cell regulation. Curr. Opin. Struct. Biol. *22*, 378–385.

Rogerson, D.T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S.M., Huguenin-Dezot, N., Muqit, M.M.K., Fry, A.M., Bayliss, R., et al. (2015). Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat. Chem. Biol. *11*, 496–503.

Sarabia-Vega, V., and Banks, L. (2019). Acquisition of a phospho-acceptor site enhances HPV E6 PDZ-binding motif functional promiscuity. J. Gen. Virol. https://doi.org/10.1099/jgv.0.001236.

Shi, G.-X., Yang, W.S., Jin, L., Matter, M.L., and Ramos, J.W. (2018). RSK2 drives cell motility by serine phosphorylation of LARG and activation of Rho GTPases. Proc. Natl. Acad. Sci. U S A *115*, E190–E199.

Sieracki, N.A., and Komarova, Y.A. (2013). Studying cell signal transduction with biomimetic point mutations. In Genetic Manipulation of DNA and Protein - Examples from Current Research, D. Figurski, ,, ed. (IntechOpen).

Simon, M., Gógl, G., Ecsédi, P., Póti, Á., Kardos, J., and Nyitray, L. (2020). High throughput competitive fluorescence polarization assay reveals functional redundancy in the S100 protein family. FEBS J. 1–13.

Sluchanko, N.N. (2018). Association of multiple phosphorylated proteins with the 14-3-3 regulatory hubs: problems and perspectives. J. Mol. Biol. *430*, 20–26.

Sluchanko, N.N., Beelen, S., Kulikova, A.A., Weeks, S.D., Antson, A.A., Gusev, N.B., and Strelkov, S.V. (2017). Structural basis for the interaction of a human small heat shock protein with the 14-3-3 universal signaling regulator. Structure *25*, 305–316.

Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., Chishti, A.H., Crompton, A., Chan, A.C., Andersen, J.M., and Cantley, L.C. (1997). Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science *275*, 73–77.

Suarez, I., and Trave, G. (2018). Structural insights in multifunctional papillomavirus oncoproteins. Viruses *10*, 1–22.

Sundell, G.N., Arnold, R., Ali, M., Naksukpaiboon, P., Orts, J., Güntert, P., Chi, C.N., and Ivarsson, Y. (2018). Proteome-wide analysis of phospho-regulated PDZ domain interactions. Mol. Syst. Biol. *14*, 1–22.

Thomas, G.M., Rumbaugh, G.R., Harrar, D.B., and Huganir, R.L. (2005). Ribosomal S6 kinase 2 interacts with and phosphorylates PDZ domain-containing proteins and regulates AMPA receptor transmission. Proc. Natl. Acad. Sci. U S A *102*, 15006–15011.

Thomas, M., Myers, M.P., Massimi, P., Guarnaccia, C., and Banks, L. (2016). Analysis of multiple HPV E6 PDZ interactions defines type-specific PDZ fingerprints that predict oncogenic potential. Plos Pathog. *12*, 1–21.

Toto, A., Mattei, A., Jemth, P., Gianni, S., Biochimiche, S., Fanelli, A.R., Pasteur, I., Cenci, F., Molecolari, P., Università, S., et al. (2017). Understanding the role of phosphorylation in the binding mechanism of a PDZ domain. Protein Eng. *30*, 1–5.

Vaccaro, P., and Dente, L. (2002). PDZ domains: troubles in classification. FEBS Lett. *512*, 345–346.

Vincentelli, R., Luck, K., Poirson, J., Polanowska, J., Abdat, J., Blémont, M., Turchetto, J., Iv, F., Ricquier, K., Straub, M.-L., et al. (2015). Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. Nat. Methods *12*, 787–793.

Zhang, Y., Dasgupta, J., Ma, R.Z., Banks, L., Thomas, M., and Chen, X.S. (2007). Structures of a human papillomavirus (HPV) E6 polypeptide bound to MAGUK proteins: mechanisms of targeting tumor suppressors by a high-risk HPV oncoprotein. J. Virol. *81*, 3618–3626.

Zheng, W., Zhang, Z., Ganguly, S., Weller, J.L., Klein, D.C., and Cole, P.A. (2003). Cellular stabilization of the melatonin rhythm enzyme induced by nonhydrolyzable phosphonate incorporation. Nat. Struct. Biol. *10*, 1054–1057.

Zhu, P., Gafken, P.R., Mehl, R.A., and Cooley, R.B. (2019). A highly versatile expression system for the production of multiply phosphorylated proteins. ACS Chem. Biol. *14*, 1564–1572.

Please cite this article in press as: Gogl et al., Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study, Structure (2020), https://doi.org/10.1016/j.str.2020.03.010

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Bacteria: E.coli BL21(DE3)	NEB	Cat#C2527I
Chemicals, Peptides, and Recombinant Proteins		
RSK1 peptide	(Gógl et al., 2019)	N/A
RSK1_-1P peptide	This paper	N/A
RSK1_-2P peptide	This paper	N/A
RSK1_-3P peptide	(Gógl et al., 2019)	N/A
RSK1_-3E peptide	This paper	N/A
16E6 peptide	(Vincentelli et al., 2015)	N/A
16E6_-2P peptide	This paper	N/A
16E6_-2E peptide	This paper	N/A
pB6 peptide	Severn Biotech	N/A
Deposited Data		
Crystal structure of RSK1 + MAGI1_2	(Gógl et al., 2018)	PDB ID: 5N7D
Crystal structure of RSK1_-3P + MAGI1_2	(Gógl et al., 2018)	PDB ID: 5N7F
Crystal structure of RSK1_-3E + MAGI1_2	This paper	PDB ID: 6TWY
Crystal structure of 16E6 + MAGI1_2	This paper	PDB ID: 6TWQ
Crystal structure of 16E6_-2P + MAGI1_2	This paper	PDB ID: 6TWX
Crystal structure of 16E6_-2E + MAGI1_2	This paper	PDB ID: 6TWU
Crystal structure of 16E6_-2P + 14-3-3σ	This paper	PDB ID: 6TWZ
Recombinant DNA		
PDZome library (His6-MBP-PDZ constructs in modified pET15b)	(Duhoo et al., 2019)	N/A
Plasmid: ANXA2-fused MAGI1_2 (His6-tagged in modified pET15b)	(Gógl et al., 2018)	N/A
Plasmid: 14-3-3σ (His6-tagged in modified pET28)	(Sluchanko et al., 2017)	N/A
Plasmid: 14-3-3σ optimized for crystallization (His6-tagged in modified pET28)	(Sluchanko et al., 2017)	N/A
Plasmid: PKA (His6-tagged in pET15b)	Narayana et al., 1997	Addgene Plasmid #14921
Software and Algorithms		
XDS	(Kabsch, 2010)	xds.mpimf-heidelberg.mpg.de/
Phaser	(McCoy et al., 2007)	www.phenix-online.org
Phenix	(Adams et al., 2010)	www.phenix-online.org
ProFit	(Simon et al., 2020)	github.com/GoglG/ProFit

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gergo Gogl (goglg@igbmc.fr). The study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The study focused on human proteins and peptides, expressed in E. coli BL21(DE3) or synthesized chemically.

METHOD DETAILS

1. MBP-PDZ Library Preparation, Peptide Synthesis

We used an updated version of our original PDZome library, that contains all the human (266) PDZ domains as soluble, isolated His6-MBP-PDZ constructs. This PDZome v2 library was prepared as previously described in details (Duhoo et al., 2019). Briefly, His6-MBP-PDZ constructs were individually overexpressed in E. coli BL21(DE3) with an autoinduction media. The His6-MBP-PDZ concentrations of soluble cell lysate fractions were evaluated with a microfluidic capillary electrophoretic system (Caliper LabChip GXII, PerkinElmer, Waltham, Massachusetts) and were adjusted to 4 μM by dilutions.

Tandem affinity purified proteins were used for affinity measurements. His6-MBP-PDZ constructs were purified on a Ni-IDA column and they was further purified by amylose affinity chromatography. For crystallization, ANXA2-fused MAGI1_2 was captured on a Ni-IDA column, the His6 tag was removed with a TEV protease and the protein was purified by cation exchange on a HiTrap SP HP column (GE Healthcare, Chicago, Illinois).

For affinity measurements an N-terminally His6-tagged 14-3-3σ (1-231) protein was used that lacked its flexible C-terminal extension. For crystallization, an engineered version of this protein was used, carrying amino acid substitutions 159KKE161 -> 159AAA161 to reduce surface entropy (Sluchanko et al., 2017). Both 14-3-3 proteins were expressed in E. coli BL21(DE3) and purified by Ni-affinity chromatography and gel-filtration. In the case of the engineered isoform, the expression tag was removed with 3C protease followed by a reverse Ni-affinity purification and gel-filtration.

Peptides used for holdup experiments (16E6: biotin-ttds-SSRTRRETQL; 16E6_-2P: biotin-ttds-SSRTRREpTQL; 16E6_-2E: biotin-ttds-SSRTRREEQL; RSK1: biotin-ttds-RRVRKLPSTTL; RSK1_-3P: biotin-ttds-RRVRKLPpSTTL; RSK1_-3E: biotin-ttds-RRVRKLPETTL) and the fluorescent peptides (f16E6: fluorescein-RTRRETQL; fRSK1: fluorescein-KLPSTTL and fpRSK1: fluorescein-KLPpSTTL) were chemically synthesized on an ABI 443A synthesizer with Fmoc strategy. In all cases, the biotin group was attached to the N-terminus via a TTDS linker and fluorescein was coupled directly to the N-terminus. The pB6 peptide (WLRRApSAPLPGLK) was commercially purchased (Severn Biotech, Kidderminster, UK). Predicted peptide masses were confirmed by mass-spectrometry.

His6-tagged PKA (Addgene #14921) (Narayana et al., 1997) was expressed in E. coli BL21(DE3) and was purified on a Ni-IDA column. Kinase reaction on f16E6 (800 μM) was performed in the presence of 5 mM MgCl2 and 2 mM ATP and 50 μM kinase for 3 h at room temperature. The kinase was removed from the reaction by boiling and centrifugation and the remaining peptide solution was buffer exchanged. pB6 peptide was labeled with sub-stoichiometric FITC (Sigma-Aldrich, St. Louis, Missouri) in a basic HEPES buffer (pH 8.2) and the reaction was stopped with 100 mM TRIS. The peptide was buffer exchanged in order to separate from the fluorescent contaminants.

The concentration of proteins, or peptides that contained aromatic residues were determined by their UV absorption at 280 nm. The concentration of peptide solutions which lacked aromatic residues were estimated based on the dry mass of the peptides and later confirmed by their far UV absorption (at 205 and 214 nm).

2. Holdup Assay

The automated holdup assay was carried out with the peptides 16E6, 16E6_-2P, 16E6_-2E, RSK1_-3P and RSK1_-3E in singlicates using a previously described protocol. (As 16E6 was used as a benchmark peptide for the new PDZome library, we measured its binding profile at least 5 times.) First, streptavidin resin was saturated with biotinylated peptides or with biotin (as a negative control). Then, the resins were depleted with an excess of biotin and were washed with a biotin-free buffer. Concentration-adjusted His6-MBP-PDZ containing cell lysates (PDZome v2 library) were incubated with the resins for a sufficient time for complex formation (30 min). The supernatant was separated from the resin by a fast filtration step, carried out by using filter plates (Millipore, Burlington, Massachusetts). PDZ concentrations were evaluated using a microcapillary electrophoretic system (Caliper; PerkinElmer, Waltham, Massachusetts) and BI values were calculated using Equation 1.

$$BI = \frac{I_{total} - I_{depleted}}{I_{total}} \quad \text{(Equation 1)}$$

where Itotal is the total Intensity of the PDZ peak (measured by the biotin control) and Idepleted is the intensity of PDZ peak in the peptide depleted reaction. In the holdup buffer at least a single internal standard was used (BSA and/or lysozyme) for peak intensity normalization (Figure S2).

3. Fluorescence Polarization (FP) Assay

Fluorescence polarization was measured with a PHERAstar (BMG Labtech, Offenburg, Germany) microplate reader by using 485 ± 20 nm and 528 ± 20 nm band-pass filters (for excitation and emission, respectively). In direct FP measurements, a dilution series of the MBP-PDZ or 14-3-3 protein was prepared in 96 well plates (96 well skirted pcr plate, 4ti-0740, 4titude, Wotton, UK) in a 20 mM HEPES pH 7.5 buffer that contained 150 mM NaCl, 0.5 mM TCEP, 0,01% Tween 20 and 50 nM fluorescently-labeled peptide. The volume of the dilution series was 40 μl, which was later divided into three technical replicates of 10 μl during transferring to 384 well micro-plates (low binding microplate, 384 well, E18063G5, Greiner Bio-One, Kremsmünster, Austria). In total, the polarization of the probe was measured at 8 different protein concentrations (whereas one contained no protein and corresponded to the free peptide). In competitive FP measurements, the same buffer was supplemented with the protein to achieve a complex formation of 60-80%,

based on the titration. Then, this mixture was used for creating a dilution series of the competitor (i.e. the studied peptides) and the measurement was carried out identically as in the direct experiment. Analysis of FP experiments were carried out using ProFit, an in-house developed, Python-based fitting program (Simon et al., 2020). The dissociation constant of the direct and competitive FP experiment was obtained by fitting the measured data with quadratic and competitive equation, respectively (Roehrl et al., 2004). Competitive titrations in the main figures are reproduced from Figures S1 or S4 (without their direct experimental pairs) unless it is defined differently in the text.

4. Crystallization

The MAGI1_2 complexes were reconstituted by mixing purified the ANXA2-fused PDZ domain (135 μM) and synthetic (biotinylated) peptides in a 1:3-1:15 stoichiometric ratio, depending on the dissociation constant of the crystallized complex. 14-3-3σ (300 μM) was supplemented with 16E6_-2P peptide in a 1:3 molar ratio. Crystallization conditions were screened using commercially available and in-house developed kits (Qiagen, Hampton Research, Emerald Biosystems) by the sitting-drop vapor-diffusion method in 96-well MRC 2-drop plates (SWISSCI, Neuheim, Switzerland), using a Mosquito robot (TTP Labtech, Cambridge, UK). After optimization, PDZ crystals grew rapidly in a drop made from 5 μl of protein solution and 5 μl of reservoir solution containing 20-25% polyethylene glycol 3000, 100 mM sodium citrate buffered at pH 5.5 and 100 mM trisodium-citrate at 20°C. The optimized condition of the 14-3-3σ crystals consisted of 20% polyethylene glycol 3350, 0.1M Bis-tris propane buffered at pH 6.5 and 0.2 M K/Na-tartarate at 4°C. All crystals were flash-cooled in a cryoprotectant solution containing 20% glycerol and stored in liquid nitrogen.

X-ray diffraction data were collected at the Synchrotron Swiss Light Source (SLS) (Switzerland) on the X06DA (PXIII) beamline and processed with the program XDS (Kabsch, 2010). The crystal structure was solved by molecular replacement with a previously determined crystal structure of the same chimera (PDB ID 5N7D), or with an apo 14-3-3σ structure (PDB ID 5LU2) using Phaser (McCoy et al., 2007) and structure refinement was carried out with PHENIX (Adams et al., 2010). TLS refinement was applied during the refinement. The crystallographic parameters and the statistics of data collection and refinement are shown in Table S3. The refined model and the structure factor amplitudes have been deposited in the PDB with the accession codes 6TWQ, 6TWU, 6TWX, 6TWY and 6TWZ.

A final electron density map, along with a simulated annealing difference omit map, is shown for all the determined crystal structures in Figure S5. A crystallographic dataset was also collected at a resolution of 2.9 Å of an APO ANXA2-fused MAGI1_2 crystal in order to calculate isomorphous difference maps on all the PDZ-bound complexes determined in this study.

QUANTIFICATION AND STATISTICAL ANALYSIS

1. Fluorescence Polarization Assay

The reported dissociation constants and errors are the average and the standard deviations of 500 independent Monte-Carlo simulations, calculated using ProFit as described in Simon et al., 2020.

2. Conversion of Holdup Binding Intensities to Dissociation Constants

Steady-state binding intensities (deduced from holdup assays) can be converted to steady-state dissociation constants using Equation 2:

$$K_d = \frac{([PDZ] - BI * [PDZ]) * ([PBM] - BI * [PDZ])}{BI * [PDZ]} \quad \text{(Equation 2)}$$

where [PDZ] is the total PDZ concentration (set to 4 μM in usual cases in our assays) and [PBM] is the total peptide concentration. This parameter is unknown that makes a direct and accurate conversion impossible.

To reveal this missing parameter for accurate conversion, we used the orthogonal affinity values from the fluorescence polarization assay. For each BI-Kd pairs, where both affinity measurements showed a detectable binding, we calculated an apparent peptide concentration based on Equation 2.

The calculated peptide concentrations showed a tailed distribution, with a few clear outliers in each cases. To estimate a global peptide concentration that is most compatible between every BI-Kd datasets, we performed an outlier rejection based on the absolute distances from the median (Figure 3A). Based on these criteria, we could calculate the average peptide concentration for each peptides using approximately 10 BI-Kd pairs. For the conversion of the modified 16E6 peptides, for which we lacked any reliable BI-Kd pairs, we used a peptide concentration of the average of the other peptides. We performed the conversion until the limit of detection of the holdup assay (BI = 0.1) in ‘ S2.

DATA AND CODE AVAILABILITY

The accession number for the crystal structure of RSK1_-3E + MAGI1_2 reported in this paper is PDB ID: 6TWY. The accession number for the crystal structure of 16E6 + MAGI1_2 reported in this paper is PDB ID: 6TWQ. The accession number for the crystal structure of 16E6_-2P + MAGI1_2 reported in this paper is PDB ID: 6TWX. The accession number for the crystal structure of 16E6_-2E + MAGI1_2 reported in this paper is PDB ID: 6TWU. The accession number for the crystal structure of 16E6_-2P + 14-3-3σ reported in this paper is PDB ID: 6TWZ. Any additional data, that is not directly available in the supplement, can be requested from the authors.

Chapter 11

Interactomic affinity profiling by holdup assay: acetylation and distal residues impact the PDZome-binding specificity of PTEN phosphatase

Status: Submitted to PLoS One as first author.

11.1 Summary

Background: Related to the previous chapter, we assayed the difference in acetylation as PTM and a synonym mutation in PTEN, as well as the impact of extending the PBM length. We prepared 11 residues long PTEN in the wild-type form, PTEN acetylated at p-1, PTEN K to R mutation at p-1 and 13 residue long PTEN in the wild-type form.

Results: We used the holdup assay and complementary FP measurements to analyze the impact of the acetylation, a synonym mutation and the length of PTEN in its interaction with the PDZome. We also introduced for the first time a specificity quantification for our BI Profiles.

Conclusions: We showed that synonym mutations may affect the PBM/PDZ interactions as well as inclusion of PTM and extension in length. We moreover conducted this paper in a methodological way to validate holdup assay complemented

by FP.

Contribution: I curated the holdup data using the software presented in chapter 15. I realized the differences among our PTEN and the one of our collaborator (turning out to be PTEN_11 and PTEN_13 respectively), hence I started to the story of the PBM length impact. I performed the bibliographical search of previous results in PTEN/PDZ interactions. I also proposed the PDZ index value which will be used for the specificity assessment of the PBMs. I wrote the first draft of the paper. I performed the draft of the figures. I participated in the final discussion of the results.

11.2 BiorXiv version of the manuscript

Interactomic affinity profiling by holdup assay: acetylation and distal residues impact the PDZome-binding specificity of PTEN phosphatase

Pau Jané[1], Gergő Gógl[1], Camille Kostmann[1], Goran Bich[1], Virginie Girault[2,3], Célia Caillet-Saguy[2], Pascal Eberling[1], Renaud Vincentelli[4], Nicolas Wolff[2], Gilles Travé[1] and Yves Nominé[1]

Affiliations:

1. (Equipe labelisée Ligue, 2015) Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258 / CNRS UMR 7104 / Université de Strasbourg, 1 rue Laurent Fries, BP 10142, 67404 Illkirch, France.

2. Unité Récepteurs-canaux, Institut Pasteur, UMR 3571, CNRS, 75724 Paris, France.

3. Current address: Technical University of Munich, School of Medecine, Institute of Virology, 81675 Munich, Germany.

4. Unité Mixte de Recherche (UMR) 7257, CNRS – Aix-Marseille Université, Architecture et Fonction des Macromolécules Biologiques (AFMB), Marseille, France.

Corresponding author:

Yves Nominé (mail: nominey@igbmc.fr; phone: +33.369.48.52.30) Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258 / CNRS UMR 7104 / Université de Strasbourg, 1 rue Laurent Fries, BP 10142, 67404 Illkirch, France. (Website: http://www.igbmc.fr/)

Abstract

Protein domains often recognize short linear protein motifs composed of a core conserved consensus sequence surrounded by less critical, modulatory positions. Here we used an accurate experimental approach combining high-throughput holdup chromatographic assay and fluorescence polarization to measure quantitative binding affinity profiles of the PDZ domain-binding motif (PBM) of PTEN phosphatase towards the 266 known human PDZ domains. Inclusion of N-terminal flanking residues, acetylation or mutation of a lysine at a modulatory position significantly altered the PDZome-binding profile of the PTEN PBM. A specificity index is also introduced to quantify the specificity of a given PBM over the complete PDZome. Our results highlight the impact of modulatory residues and post-translational modifications on PBM interactomes and their specificity.

Categories

Biophysics, bioenergetics

Keywords

PDZ / PBM interaction, PTEN, holdup assay, Acetylation, Binding specificity

Exosite, high-throughput interaction data, fluorescence polarization

Abbreviations

PDZ: PSD95/DLG/ZO-1

PBM: PDZ-binding motif

HPV: Human Papilloma Virus

FP: fluorescence polarization

MBP: Maltose-Binding Protein

TRX: Thioredoxin

Introduction

PDZs, named from the three proteins PSD-95, DlgA and ZO1, are globular protein domains that adopt a conserved antiparallel β-barrel fold comprising 5 to 6 β-strands and 1 to 2 α-helices. PDZ domains are involved in diverse cellular activities, such as cell junction regulation, cell polarity maintenance or cell survival [1]. PDZs recognize short linear motifs (called PDZ Binding Motif or PBMs) that follow particular sequence requirements and are mostly located at the extreme carboxy terminus of target proteins [2]. The human proteome contains 266 PDZ domains dispersed over 152 proteins [3] and thousands of presumably disordered C-termini matching a PBM consensus [4].

The core of a C-terminal PBM is formed by four residues, which are disordered in the unbound state but form, upon binding, an anti-parallel β-strand that inserts between a β-strand and a α-helix of the PDZ domain. A C-terminal PBM contains two conserved residues (positions are thereafter numbered backwards from the C-terminus, starting at p-0): a hydrophobic residue at p-0 and a characteristic residue at p-2, which actually determines the PBM classification: Ser / Thr for class I, a hydrophobic residue for class II and Asp / Glu for class III. Other positions located within or upstream of the core motif may also modulate the binding affinity ([5]–[8] and reviewed in [3]). In particular, systematic mutagenesis experiments have shown that amino acid replacements at positions -1, -3, -4 and -5, and sometimes even at upstream positions, can strongly alter the binding properties depending on the PDZ domain [9]–[11]. We and others have also shown that the length of the peptides or upstream or downstream sequences of the PDZ constructs used may influence the binding affinity in the assays [12]–[16].

Additionally, post translational modifications (PTM) at residues within or upstream of the PBM core are susceptible to alter the binding affinity for PDZ [17], and therefore the PDZ / PBM network. Protein acetylation is an example of PTM that mainly targets lysine residues. Acetyltransferases catalyze the transfer of an acetyl group from acetyl-coenzyme A to the ε-amino group of a lysine residue, inducing the neutralization of the positive charge of the lysine side chain. The reaction can be reversed by lysine deacetylases. By modifying the chemical nature of the protein, the acetylation process may alter its binding properties. In particular, an acetylated protein may become "readable" by specialized acetyl-lysine binding domains such as bromodomains [18]. Acetylation occurs in a large variety of protein substrates and plays important roles in protein regulation, DNA recognition, protein / protein interaction and protein stability [19]. Originally widely described for histone proteins, it has also been observed for a growing number of non-histone proteins [20], such as PTEN [21].

PTEN is a lipid phosphatase protein located in the cell nucleus with a prominent tumor suppressor activity. When brought to the plasma membrane, PTEN is able to antagonize the phosphatidylinositol 3-kinase (PI3K), inhibiting the PI3K-dependent cell growth, survival and proliferation signaling pathways [22]. Interestingly, PTEN harbors a class I PBM –ITKV$_{COOH}$– that appears to be critical for regulating its functions [23],[24]. The PDZ binding to the PTEN PBM leads to a stabilization of PTEN and an increase of its catalytic activity [25]. The PBM of PTEN presents several original characteristics. On the one hand, a structural study revealed an unconventional mode of binding of PTEN to the PDZ domain of the human kinase MAST2 [26]: while the core of the PTEN PBM displays a canonical interaction with the PDZ domain, a Phe residue at p-11 (F392) distal from the core PBM establishes additional contacts with MAST2 through a hydrophobic exosite outlined by β2- and β3-strands of the PDZ domain. On the other hand, lysine K402, located at the p-1 position of the PBM core in PTEN, has been suggested to represent a putative target of an acetylation reaction that might modulate PTEN binding to PDZ domains and thereby affects other PTEN activities [21]. Remarkably, those original characteristics of the PBM of PTEN (unconventional PDZ binding mode of PTEN and potential modulation by acetylation) have been examined only in context of interaction with

a few PDZ domains. It is thereafter interesting to cover their impact on the interactome with the full PDZome, thus requiring the use of a high-throughput screening method, as the holdup.

The holdup method is a chromatographic approach in solution developed in our group that allows to measure the binding strength of a peptide, attached to a resin, against a library of domains of a same family. We initially proposed this method to explore the interaction between PBM peptides and the human PDZ domains [27]. Briefly, a soluble cell lysate containing individually overexpressed PDZ domain is incubated until equilibrium with a calibrated amount of streptavidin-resin saturated either with the target biotinylated PBM peptide or with biotin as a reference. The flow-throughs containing the unbound protein fraction are recovered by filtration and loaded on a capillary electrophoresis instrument to quantify the amount of remaining free PDZ. The stronger the steady-state depletion of the PDZ domain in the flow-through as compared to the reference, the stronger the PDZ / PBM binding interaction. The assay is particularly suited to quantitatively evaluate and compare large numbers of interactions. This method delivers, for each PBM / PDZ pair, a "binding intensity" (BI), whose value can in principle range from 0.00 (no binding event detected) to 1.00 (maximal binding event). The approach has been automated [28] and the human PDZ library was recently extended to the complete 266 PDZ domains known in human proteome [29]. The high-throughput assay is implemented on 384 well-plates, and can probe a single peptide in triplicate or up to 3 different peptides in singlicate against the 266 PDZ domains. The full processing leads to a binding profile, i.e. a list of binding strengths in decreasing order exhibited by a given PBM towards the entire PDZome. The high accuracy and efficiency of the holdup assay has been validated previously [4],[15],[28],[30]. Very recently, a manual version of the holdup assay with purified samples and using widespread benchtop equipment has been implemented and has proven to be reliable [31].

In the present work, we investigated how the acetylation at position K402 in PTEN (–IT^{Ac}KV$_{COOH}$– thereafter corresponding to p-1 position in the PBM), would alter the binding affinity profile of the PTEN C-terminus to the full complement of known human PDZ domains (the PDZome). We also assessed the contribution of the K402R mutation, expected to preserve the positive charge and the overall bulkiness of the lysine residue, as well as the effect of the presence of the hydrophobic residue at p-11 (F392). For these purposes, we combined the updated high-throughput holdup assay with fluorescence polarization (FP) measurements allowing to convert each BI value into affinity. We obtained all the affinities of the complete human PDZ library for wild-type, acetylated and mutated versions of the PBM of an 11-mer PTEN C-terminal peptide as well as an extended 13-mer peptide. We also introduced a tentative "promiscuity index" to quantify the PDZome-binding specificity of each peptide. The results show that acetylation affects the affinities for the PDZome and highlight the importance of the exosite in modulating the PDZome specificity for the PDZ-binding motif of PTEN.

Material and Methods

Protein expression and purification

The 266 PDZ domains that constitute the used PDZ library (“PDZome V.2”) were produced using constructs with optimized boundaries as described previously [32]. All the genes were cloned into pETG41A or pETG20A plasmid. The expressions in *E.coli* resulted in a recombinant protein fused to an N-terminal solubility tag (His-MBP or TRX). The expressed tag-PDZ concentrations were quantified using capillary gel electrophoresis and cell lysates were diluted to reach approximately 4

µM tag-PDZ before freezing in 96-well plates. A detailed protocol of the PDZ library production, expression and benchmarking can be found in [29]. PDZ domains are named according to their originating protein name followed by the PDZ number (e.g. NHERF1-1 as the first PDZ domain of the NHERF1 protein).

For FP assay, tandem affinity purified His_6-MBP-PDZ proteins were used. Cell lysates were purified on Ni-IDA columns, followed by an MBP-affinity purification step. Protein concentrations were determined by far-UV absorption spectroscopy. A detailed protocol has been published previsouly [4].

Peptide synthesis

All 11-mer biotinylated peptides (PTEN_11: DEDQHTQITKV, PTEN_Ac: DEDQHTQIT***ac*****K**V and PTEN_KR: DEDQHTQIT**R**V) were chemically synthesized on an ABI 443A synthesizer with Fmoc strategy by the Chemical Peptide Synthesis Service of the IGBMC, while PTEN_13 (**PF**DEDQHTQITKV) was purchased from JPT Innovative Peptide Solutions with 70%–80% purity. A biotin group was systematically attached to the N-terminal extremity of the peptide via a TTDS linker while fluorescent peptides were prepared by directly coupling fluorescein to the N-terminus. Predicted peptide masses were confirmed by mass spectrometry. Due to the lack of aromatic residue, peptide concentrations were first estimated based on the dry mass of the peptide powders and subsequently confirmed by far-UV absorption (at 205 and 214 nm).

Holdup assay

The Holdup assay was performed in singlicate for the three 11-mer PTEN variants and the 13-mer PTEN variant as described in [28],[29]. Prior to interaction assay, the streptavidin resin was saturated with biotinylated PBM peptides and then washed with an excess of free biotin, while the reference resin was incubated only with biotin. Right before the holdup experiment, the PDZ library was spiked with an internal standard of lysozyme. Then, the biotin- or PBM-saturated resins were incubated with diluted cell lysates, each in a distinct well of a 384-well plate, allowing to adjust the concentration of tag-PDZ at around 4 µM. After a sufficient time for the complex to form (15 min.), a fast and mild filtration step is performed and the tag-PDZ concentrations were measured by capillary electrophoresis instrument (LabChip GXII, PerkinElmer, Massachusets, USA). A detailed protocol of how to run the holdup assay in an automatic way using liquid handling robots can be found in [29]. Standard markers were used to convert migration time into molecular weight on the LabChip software and inappropriate molecular weight markers were corrected or excluded.

Holdup data quality check and processing

Holdup data can be missing for some tested pairs mainly for three reasons: i/ biochemical issues, specially when the over-expressed domain is not concentrated enough in the sample, ii/ acquisition problems mainly because of a misreading of the Caliper data, iii/ technical difficulties related to data processing. For points i/ and ii/, many efforts have been made to optimize the expression and to run the LabChip GXII instrument in the best conditions. For point iii/, we developed bioinformatics processing tools in order to improve the accuracy and reproducibility of the intensity measurement of the tag-PDZ peak in the chromatogram [33]. Briefly a baseline correction of the electropherograms is first performed in order to remove the background noise and extract the real intensities using Python package available in https://spikedoc.bitbucket.io under the name of SPIKE.py [34],[35]. Then intensities are normalized using the internal standard (lysozyme as previously mentioned) to correct potential variations over all the protein concentrations. Lastly, both the sample and the reference

electropherograms were superimposed by adjusting the molecular weight on the X-axis according to a linear transformation (translation and dilation) of the sample electropherogram as compared to the reference one.

Beyond the purpose of this article, we have accumulated several tens of thousands of PDZ / PBM interaction data with the holdup protocol used here. Experienced holdup data curators combined four quantitatively evaluable quality criteria to retain or discard data during visual inspection. Individual electropherograms must display a sufficiently high intensity of the normalization peak (criterium 1) and of the tag-PDZ peak (criterium 2) while the signal of crude extract should be kept as low as possible compared to the tag-PDZ peak (criterium 3) (**Fig. 1A**). When comparing two electropherograms, the elution profiles must be sufficiently aligned (criterium 4) (**Fig. 1B**). In order to rationalize and accelerate data curation, we assigned to each criterium an individual quality score ranging from 0 to 1 from the lowest to the highest quality data (**Fig. 1C**). To avoid a cut-off effect, a linear or quadratic transition was introduced depending on the quality criteria type. The product of the resulting individual scores led to a global quality score in the 0-to-1 range. We calculated such scores for holdup data sets that had been treated by expert curators, then compared the scores of the data that had been either rejected or retained by the curators. This allowed us to semi-empirically set a threshold value of 0.6 which maximizes the true positive rate and minimizes the false negative rate. This threshold was automatically used to distinguish data to be rejected from those to be retained in a way that generally agrees with the expert curator's decision. For the datasets used in the present study, the percentage of rejection never exceeded 10%.

For filtered data, the BI was extracted with the following equation (**Eq. 1**) that estimates the depleted fraction after superimposition of the sample and reference electropherograms:

$$BI = \frac{I_{ref} - I_{lig}}{I_{ref}} \qquad \textbf{Eq. 1}$$

where I_{ref} and I_{lig} are the intensities of the tag-PDZ peaks measured in the reference and the sample electropherograms, respectively, for a given PDZ domain / PBM peptide interaction pair.

Data reproducibility has been previously explored for several PDZ / PBM pairs resulting in a standard error of the mean of about 0.07 BI unit (data not shown + [28]). This suggests that the maximal BI values differ significantly from PTEN_Ac or PTEN_KR constructs as compared to PTEN_11 and in a less extend to PTEN_13. In some cases, negative BI values as low as -0.20 can be observed and seem to be reproducible (data not shown). This could result from a lower intensity of the reference PDZ / PBM peak as compared to the sample PDZ / PBM peak, potentially due a preference of the PDZ domain for beads fully saturated with biotin as compared to beads with biotinylated peptide. As reported previously, we have also investigated the limit of detection by repeating the holdup experiments for an irrelevant "neutral" peptide owing no specific PBM consensus sequence. Almost all BI values were below 0.10 (98% of all measured PDZ / PBM pairs) and showed a standard deviation of less than 0.10 (considering 95% of the data) [28]. According to this, we applied a conservative safety factor of 2 that leads to a limit for BI of 0.20. This cut-off represents a very stringent threshold retaining only high-confidence PDZ / peptide interactions, and eliminating most of the false positives.

Steady-state fluorescence polarization

FP data were measured in 384-well plates (Greiner, Frickenhausen, Germany) using a PHERAstarPlus multi-mode reader (BMG labtech, Offenburg, Germany) with 485 ± 20 nm and 528 ± 20 nm band-pass filters for excitation and emission, respectively. N-terminal fluorescein-labeled HPV16E6 (fluorescein-RRETQL), RSK1 (fluorescein-KLPSTTL) and phospho-RSK1 (fluorescein-KLPpSTTL) were used as

tracers. In competitive measurements, the 50 nM fluorescent reporter peptide was first mixed in 20 mM HEPES pH 7.5 buffer (containing 150 mM NaCl, 0.5 mM TCEP, 0,01% Tween 20) with the PDZ domain at a sufficient concentration to achieve high degree of complex formation. Subsequently, increasing amount of unlabeled peptide was added to the reaction mixture with a total of 8 different peptide concentrations (including the 0 nM peptide concentration i.e. the absence of peptide). Titration experiments were carried out in triplicate. The average FP signal was used for fitting the data to a competitive binding equation with ProFit, an in-house Python-based program [36], allowing to extract the apparent affinity values. In our competitive assays, every tested PDZ domain detectably bound to at least one PBM peptide, in agreement with well folded PDZ domains.

Conversion from BI values to dissociation equilibrium constants

BIs were transformed into dissociation constants (K_D) using the following formula:

$$K_D = \frac{([PDZ_{tot}]-BI\cdot[PDZ_{tot}])\cdot([PBM_{tot}]-BI\cdot[PDZ_{tot}])}{BI\cdot[PDZ_{tot}]} \qquad \textbf{Eq. 2}$$

where $[PDZ_{tot}]$ and $[PBM_{tot}]$ correspond to the total concentrations of the PDZ domain (usually around 4 µM) and the PBM peptide used during the assay. Since the PBM_{tot} concentration in the resin during the holdup assay parameter may differ from one peptide to another and remains unknown, it is impossible to directly convert BI values into K_D constants. To extract the PBM concentration, we systematically determined by FP the K_D constants for a subset of PDZ / PBM pairs that were used to back-calculate the peptide concentrations in the holdup assays when quantifiable and significant (>0.20) BI values were available for the same pairs (**Eq. 2**). For each PBM, the average peptide concentration was calculated after outlier rejection based on the absolute distances from the median as compared to three times the standard deviation (3σ rule), with never more than 2 values rejected.

Results

An experimental strategy to measure large numbers of reliable affinity data.

For this study, we wished to generate accurate and complete PDZome-binding affinity profiles for four peptide variants of the C-terminal PBM of PTEN. In practise, this requires measuring the individual affinities of 4x266=1064 distinct PBM-PDZ pairs. A singlicate holdup experiment is well suited for such a task. Taking into account the additional ~360 biotin-PDZ negative control measurements required for data treatment, the assay delivers ~1400 filtrates of protein extracts, which must each be individually subjected to capillary electrophoresis. Next, individual electropherograms must be visually curated and analyzed by an expert user to extract the binding intensities (BI) values that will compose the final profiles. As described in the material and methods section, we rationalized the data curation step by introducing a numerical global quality score. Since the assay requires expensive materials and labor-intensive data treatment, one should favor an approach based on singlicate holdup runs. We therefore used a strategy that combines one holdup assay run in singlicate with a medium-throughput competitive FP protocol run on a large proportion of the PDZ / PBM interacting pairs detected in the holdup assay (see material and methods). This strategy warrants the obtention of highly reliable affinity data for all PDZ / PBM interacting pairs that pass the quality score filtering step after the holdup assay. Representative holdup data recorded for one PBM (PTEN_11) are shown in **Fig. 2A**. After normalization of the two capillary electropherograms recorded for both the PBM of interest and the biotin reference, the comparison of the intensities of the two resulting PDZ peaks informs about the strength of the interaction: the stronger the depletion, the stronger the binding. Representative FP data are shown in **Fig. 2B**. The apparent affinities were obtained by fitting the anisotropy data considering a

competitive binding model [37]. The holdup BI values and the binding strength derived from competitive FP measurements are consistent: higher the BI, stronger the affinity.

Generating PDZome-binding BI profiles of the four PTEN variant PBMs by holdup assay.

We applied the holdup assay to generate PDZome-binding profiles of three 11-mer peptides (PTEN_11 for the native sequence, PTEN_Ac and PTEN_KR for the acetylated and K402R mutated version of PTEN_11, respectively), as well as an extended 13-mer peptide (PTEN_13). Considering the quality score filtering step, we managed to quantify the interactions of 213, 233, 215 and 257 PDZ for the PTEN_11, PTEN_Ac, PTEN_KR and PTEN_13 peptides, respectively, which corresponds to 80%, 81%, 88% and 97% of the human PDZome. All holdup plots that detected a binding event with a binding intensity BI>0.20 are shown in **Supp. Fig. S1**. The four resulting holdup datasets were then plotted independently in the form of "PDZome-binding profiles" representing the individual BI values versus the PDZ domains ranked from higher to lower BI values (**Fig. 3**). PTEN_11 showed a maximal BI value of 0.71, i.e. a lower binding strength as compared to the ones of PTEN_KR, PTEN_Ac or PTEN_13 (BI = 0.86, 0.90 and 0.81, respectively). Using BI>0.2 as a minimal threshold value for retaining high-confidence PDZ / peptide interactions, the holdup assay identified 19, 43, 37 and 24 PDZ domains as potential binders for the PTEN_11, PTEN_Ac, PTEN_KR and PTEN_13 peptides, respectively. Altogether, they represent a total of 123 potential binders, of which 60 are non-redundant PDZ domains distributed over 46 distinct proteins.

Orthogonal validation by competitive FP and conversion of holdup BI data into dissociation constants of the four PTEN PBMs versus the human PDZome.

Calculation of an equilibrium constant for a PDZ-PBM interaction requires three concentrations: free PBM, free PDZ and PDZ-PBM complex. The holdup assay delivers for each PDZ / PBM pair the concentrations of free PDZ and PDZ-PBM complex, but not that of free PBM. To circumvent this problem we systematically measured by competitive FP, an orthogonal approach to holdup, the K_D constants for the 4 PTEN peptides against a subset of 20 PDZ domains (**Supp. Fig. S2**), resulting in approx. 8 to 10 significant K_D for each PBM. These accurate dissociation constants were used to back-calculate the peptide concentrations in the holdup assays (**Fig. 4A**). We found the concentrations of the different PBM peptides to vary between 10 and 90 µM, with averages between 17 and 34 µM depending on the PBM after outlier rejection. A global mean of 26 µM considering all the peptides was determined. A plot of experimental K_D obtained by FP *versus* BI superimposed well with the theoretical affinity values calculated using the global average peptide concentration of 26 µM (**Supp. Fig. S3**). This shows a very good agreement between the holdup BI values and the binding strength derived from competitive FP measurements.

Using the mean concentration obtained above for every PTEN peptide, the experimental BI values recorded by holdup for all tested domain / peptide pairs were subsequently transformed into equilibrium dissociation constants. A strong agreement is observed between the affinity constants obtained from holdup and FP assays with a coefficient of determination R^2 = 0.74 (**Fig. 4B**), confirming that singlicate holdup runs provided highly reliable data. At this stage, affinity data measured by FP assay were also included for the few PDZ domains (MAST1-1, MAST2-1, SNX27-1, MAGI1-2 and GRID2IP-2) for which holdup data were missing according to the quality score filtering, representing 1 to 3 additional PDZ binders per PTEN construct. A total of 215, 234, 218 and 259 interaction data were obtained for PTEN_11, PTEN_Ac, PTEN_KR and PTEN_13, respectively. The transformation into affinity values makes then possible to compare binding affinity profiles obtained for different peptides and different batches.

From binding profiles to specificity quantification

The above described holdup-FP strategy delivers binding affinity constants, a universal chemical property. The affinity values obtained for each PTEN peptide were plotted in a logarithmic scale, hence proportional to free energies of binding ΔG at a fixed temperature (**Fig. 5**). The resulting profiles contains information about specificity or promiscuity since a promiscuous peptide as seen by hodlup would bind to a large number of PDZ. We then looked for a numerical parameter that would express, in a quantitative way, this specificity or promiscuity information. For this purpose, we calculated for each profile the difference between the maximal and minimal affinity values detected by the assay, $\Delta G_{max} - \Delta G_{min}$. Next, we introduced a threshold affinity, called "half-maximal binding affinity" defined as follows: $\Delta G_{half} = \Delta G_{min} + (\Delta G_{max} - \Delta G_{min})/2$. We then defined the half-maximal binding promiscuity index I_P as the percentage of PDZ domains bound to the PBM with an affinity superior to the half-maximal affinity relative to the total number of PDZ domains that were successfully measured in the assay (**Fig 5**). Alternatively, the specificity index I_S could be defined as $1 - I_P$. Therefore, the lower the promiscuity index, the higher the specificity index, the higher the specificity of the PBM for a few selected domains across the PDZome. For instance, if 250 PDZ domains were fully assayed, and only 5 PDZ domains bound to the PBM with an affinity superior to the half-maximal affinity, the specificity index will be 98%. If 25 domains bound with an affinity superior to the half-maximal affinity, the specificity index will be 90%.

We probed the specificity index on the PDZome-binding profiles of the four PTEN peptides. In both the BI-based and the affinity-based representations (**Fig. 2 and 5**), the shapes of the profiles of PTEN_11, PTEN_Ac, PTEN_KR look similar, while the PTEN_13 presents a sharper, faster-decreasing profile. This indicates, in qualitative terms, that the PTEN_13 peptide selects PDZ domains in a more specific -less promiscuous- way than that of the three of other peptides. This is fully confirmed by the computed specificity indexes, which yield close values for PTEN_11, PTEN_Ac, and PTEN_KR (95.8%, 94.9%, 95.9%, respectively), while the extended wild-type peptide PTEN_13 displays a higher specificity index (98.5%) indicative of a higher specificity towards a few selected PDZ domains.

Rearrangements of the binding profiles due to minor changes in PTEN

The PTEN-bound PDZ domains are distributed over a diversity of PDZ-containing proteins (**Fig. 6**). Several PDZ domains such as MAST2-1, PDZD7-3, SNX27-1, MAGI1-3 and GRASP-1 were systematically among the strongest interaction partners of all four PTEN PBM variants. We compared our data to previously published studies, bearing in mind that sequences and boundaries of PTEN and PDZ constructs may differ (**Table 1**). Our results agree with isothermal titration calorimetry data obtained for SNX27-1 / PTEN [38] and MAST2-1 / PTEN complexes [26] and, in part, with FP data obtained for PARD3-1 / PTEN complex [39]. Interestingly, some of our newly identified PTEN-binding PDZ domains, such as MAGI1-3, MAGI2-3 and DLG4-1 bound wild-type PTEN peptides with a stronger affinity than the domains of the same proteins that were previously published to bind PTEN, such as MAGI1-2 [40], MAGI2-2 [24] and DLG4-3 [41], respectively. This result illustrates the strength of the complementary holdup / FP approach which can provide an affinity ranking of PDZ domains even within multi PDZ-containing proteins.

Although the shapes of the dissociation constant profiles for the three 11-mer PTEN variants were globally similar, the PDZ domains are reshuffled between the various profiles (**Fig 7**). We detected at least 20 additional new partners for PTEN_Ac, and 11 for PTEN_KR (**Fig. 7A & Supp. Info. S1**). The acetylated peptide is highly promiscuous and binds to all the partners of the native PTEN_11 PBM,

plus numerous additional ones. Furthermore, the arginine mutation does not seem able to efficiently reproduce the acetylated state as seen by the number of partners (8 over a total of 37) detected for PTEN_KR and not for PTEN_Ac. The opposite effect with 15 over a total of 43 detected for PTEN_Ac and not for PTEN_KR is even more pronounced, suggesting that the acetylation effect on binding is mainly due to the acetyl group rather than the size of the side chain carried by the acetylated lysine residue.

The impact of the PTEN peptide length was noticeable by comparing the dissociation constant profiles of PTEN_11 and PTEN_13 (**Fig. 7B**). The detected interactions of PTEN_13 were markedly stronger compared to the affinities observed for the same PDZ partners in PTEN_11. The strongest effect is observed for MAST2, the top binder for both PTEN_11 and PTEN_13, for which the $-\log(K_D)$ value increases from 4.9 to 5.9 in log(M) unit (i.e. a jump from 13 µM to 1 µM), corresponding to about a 10 fold stronger affinity. Only a few interactions, in the low range affinities, were potentially slightly strengthened although most likely not significantly. Moreover, 24 new binders appear due apparently to the presence of the two extra residues in the N-terminus of the peptide. These rearrangements are particularly noteworthy since the mutations or the Pro-Phe inclusion introduced for this work are located at positions described as non-critical for PBM binding.

Discussion

Insight into the holdup: a powerful semi-automated tool for medium-to-low affinity measurements

In this work, we quantitatively assessed more than 1,000 distinct PDZ-peptide affinities by using a "crude holdup assay" protocol, which quantifies the disappearance of a single protein peak (the tag-PDZ peak) out of a complex crude overexpression extract. This protocol requires a rigorous approach. Some critical biochemical steps have been previously identified [28] [29] including the standardized expression of the complete PDZome, the verification of its quality, the calibration of its concentrations in the crude extract, and a careful quality control of capillary electrophoresis runs. For data treatment, we developed a computational processing step for accurate superimposition of the electropherograms to improve the precision of binding intensities [33]. Here, a four-criteria quality score was introduced to further rationalize data curation. These improvements allow us to minimize the amount of false positive and false negative results. In addition, to spare costs and manpower for data treatment, holdup experiments were run in singlicate and combined with an orthogonal approach, the competitive FP. This generated high-confidence affinity data and allowed us to convert holdup binding intensities (BI) values into affinities (ΔG or K_D). The use of such an intrinsic universal parameter of molecular complexes also presented the advantage to facilitate the comparison with data available in the literature. In future developments of the automated holdup assay, we envision to replace crude overexpression extracts by purified proteins, which greatly facilitate both readout and data treatment [31].

Impact of PTEN PBM acetylation on its PDZ interactome

Lysine acetylation is a PTM difficult to study and reproduce *in vitro*. Some studies have explored lysine acetylation by proteomic approaches [42], while others have mutated lysine residues to glutamine or arginine to mimic acetylation or suppress acetylatability, respectively [21],[43]–[45]. In the present study, we investigated with chemically synthetized peptides that allow to fully control

PTM the differential effects of acetylation or mutation of a lysine residue on the PDZ interactome of PTEN. PTEN is a tumor suppressor that is frequently inactivated in human cancers [46],[47]. Some *in vivo* activities of PTEN such as PI3K signaling regulation seem to be abolished when PTEN is acetylated [48]. In addition, the Lys-to-Arg mutation at PTEN position 402 (corresponding to non-essential p-1 position of its C-terminal PBM) abolished PTEN acetylatability [21]. However, this may either mean that K402 is a direct acetylation target or indicates that the integrity of the PTEN PBM sequence is required for PBM-dependent acetylation of PTEN at other sites distinct from K402.

We found that K402 acetylation (inducing a loss of a positive charge and a slight increase of bulkiness) altered both the strength and the number of detected PDZ binders of PTEN. In contrast, the K402R mutation (preserving the positive charge but further increasing the bulkiness) did not alter the overall binding strength nor the number of binders. Furthermore, the K402R mutant retains binding to most partners of the native motif and also binds to a subset of the acetylated peptide partners. Therefore, at the p-1 position of the PTEN PBM, the presence or absence of a positive charge appears more critical for PDZ recognition than the bulkiness of the side chain.

Although a few PDZ domains including several from MAGI and NHERF detectably bound to all three peptides PTEN_11, PTEN_Ac and PTEN_KR, several PDZ domains bound only one or two of those peptides. For instance, both PTEN_Ac and PTEN_KR bound stronger than wild-type PTEN_11 to MAGI2_2 or DLG1_2 domains, in agreement with Ikenoue *et al*. Since our study is performed over the full PDZome, this implies that acetylation generally increases the affinity of PTEN for PDZ domains. Overall, the rather large number of PDZ partners associating with the PTEN PBM confirms that domain / motif networks are rather promiscuous [49].

Lessons from distal residues on the PTEN interactome

There is no consensus for the precise residue length of a given PBM needed to complete the interaction with a PDZ domain. Although the four C-terminal residues are usually thought to constitute the core of a PBM, it was shown that peptides comprising the last 10 positions of a PBM undergo a significant change in their PDZ-binding affinities as compared to peptides comprising only the last 5 positions [13]. Such affinity variations may result from differences of entropy of the free peptides, from altered interface contacts in the resulting PDZ-PBM complexes, or a combination of both. Accordingly, synthetic or recombinant PBMs employed for PDZ interactions generally include at least 9 to 11 residues [4],[5],[15],[17],[28],[50]. Indeed, the presence of distal sites altering PDZ-PBM binding has already been described [51], even at positions as far as at p-36 [52]. In the particular case of PTEN, Terrien *et al*. previously demonstrated the existence of a distal "exosite" at F392 (p-11), that triggers novel contacts within a secondary exposed hydrophobic surface of MAST2 [26]. Here, we showed that the inclusion of two extra residues, including F392, (PTEN_13 versus PTEN_11) affected both the PDZ interactome identified for PTEN and the specificity of its PBM. Indeed, several PDZ domains detectably bound only to the longer construct, in line with the idea of a global affinity increase because of the larger number of atomic contacts. Furthermore, while the three 11-mer peptides displayed equivalent PDZ-binding specificity, PTEN_13 showed an increased specificity. The addition of the distal exosite was therefore more influential for specificity than the chemical variations (Lys acetylation or Lys to Arg mutation) at p-1 position.

In principle, one may argue that domain-motif binding events may be altered by any distal region, so that only studies full-length protein / protein interactions are relevant. Notwithstanding the methodological issues (large full-length proteins can be very difficult to handle), one must keep in mind that most full-length multi-domain proteins are prone to many conformational changes (inducible by partner binding, ligand binding, PTM, molecular crowding, and so forth), which in turn influence

the availability of their globular domains or linear motifs for binding events. This justifies the 'domainomics' approach [53] undertaken in this work, that focuses on the binding properties of minimal interacting fragments of proteins, such as a globular domains (e.g., PDZs) and short linear motifs (e.g., PBMs). Even if our binder list might be incomplete as compared to studies involving full-length proteins, it provides a list of the PDZ domains capable to interact with the motif of the PTEN PBM, constituting the minimal block at the binding interface of protein / protein interaction.

To bind or not to bind

In this work, by covering almost the entire PDZ family, we quantified both the number of interacting and non-interacting partners for a given PBM. The knowledge of the two numbers is important since the count of 3 binding partners over a dataset of 10 domains, or 3 partners over a dataset of 100, is not reporting the same specificity. Over the years, we have accumulated holdup data for many peptides and noticed that more than 90% (244/266) of the PDZs in our expressed PDZome are functionally active since they interacted significantly with at least one PBM [29]. This indicates that most of the non-binders detected in our profiles are trustable. The holdup assay is therefore a reliable approach to address not only the specificity but also the 'negatome' in the sense of the negative interaction dataset as originally proposed [54].

In this work, we derived from the PDZome-binding profiles a single numerical index to evaluate the degree of specificity of a given PBM towards particular PDZ domains. One can assume the probability of binder occurrence to be all the more similar in the validated and untested PDZ datasets as the validated dataset is covering a large part (>~80%) of the entire human PDZome. The calculation of the specificity index will thus be roughly the same for both the validated and the complete PDZ datasets. One must notice that this index is not fully satisfying and cannot be considered as a universal parameter beyond our particular PBM-PDZome affinity profiling studies. In particular this index is only operative to compare profiles with a roughly continuous decreasing shape, e.g. in absence of discontinuous "breaks" or "stairs". But the concept of specificity index affords the advantage of introducing a numerical value attached to each PBM profile, that will ease their comparison.

Conclusion

In this study, we showed that the hydrophobic exosite at position p-11, not only impacts the interaction of the PTEN C-terminal tail with MAST2 as previously reported [26], but also affects its binding to a large set of other PDZ interaction partners, suggesting to well control the length of the polypeptide used for *in vitro* interaction studies. More importantly, we also showed that both, the K402 acetylation and even the K402R point mutation at p-1, a non-critical position of the canonical PBM motif for PDZ / PBM interaction, significantly increased the number of targeted PDZ domains. This could be of primary relevance, knowing that the activities of the tumor suppressor PTEN protein is regulated by acetylation. Finally, we also introduced a way to quantify specificity that could be extended to other interaction studies covering a whole domain family.

Supplementary information

- **Supp. Fig. S1**: contains the entire data set obtained by holdup for BI>0.20. For each panel, after superimposition of the two electropherograms recorded for the PBM of interest (blue dotted line) and for the biotin reference (black solid line), the normalization of the electropherogram of the PBM compared to the one of the reference is done using the signal of the lysozyme added in every sample at a constant concentration (red peak). The region between 20 and 60 kDa which contains peaks of the crude extract supposedly to be constant, is used to verify the proper intensity normalization of the two electropherograms. The intensities of the peak of interest after proper alignment along the molecular weight scale (region covered by the green dotted line) are subsequently used to quantify the depletion of an individual PDZ domain and then the BI value. All those normalization and alignment steps are performed automatically.
- **Supp. Fig. S2**: contains the entire data set obtained by FP. Average of FP data recorded in triplicate are represented with black dots. The reported dissociation constants and errors are the average and the standard deviations of 500 independent Monte-Carlo simulations, calculated using ProFit as described in Simon et al., 2020.
- **Supp. Fig. S3**: contains the experimental (BI, K_D) plot superimposed with K_D obtained with **Eq. 2**. Error bars are representative of peptide concentration uncertainty after propagation to the $-\log(K_D)$ values.
- **Supp. Info S1** file: contains the data set with all the BI values together with the transformed dissociation equilibrium constants for each PDZ-PBM interaction. All the plots in this study are performed according to this data set.

Acknowledgements

This work received institutional support from Centre National de la Recherche Scientifique (CNRS), Université de Strasbourg, Institut National de la Santé et de la Recherche Médicale (INSERM) and Région Alsace. G.G. was supported by the "Post-doctorants en France" program of the ARC foundation. The work was supported by funding from the European Union´s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 675341, by the Ligue contre le cancer (équipe labellisée 2015), by the National Institutes of Health (Grant R01CA134737), by the Canceropôle Grand-Est (projet Emergent), and by the French Infrastructure for Integrated Structural Biology (FRISBI).

Author contributions

PJ: conceptualization; data analysis and interpretation; drafting and revising the article. GG: conception and design; data acquisition; data analysis and interpretation; drafting and revising the article. CK: data acquisition; revising the article. GB: data analysis; revising the article. VG: data acquisition; revising the article. CCS: data acquisition; revising the article. PE: peptide synthesis; revising the article. RV: data acquisition, revising the article. NW: funding acquisition; conception and design; data analysis and interpretation; revising the article. GT: funding acquisition; conception and design; data analysis and interpretation; drafting and revising the article. YN: funding acquisition; conception and design; data analysis and interpretation; drafting and revising the article.

References

(1) Sheng M., & Sala C. (2001) PDZ domains and the organization of supramolecular complexes. *Annu. Rev. Neurosci.* 24(*1*); 1–29.

(2) Nourry C., Grant S.G.N., April J.B., Nourry C., Grant S.G.N., & Borg J. (2003) PDZ Domain Proteins : Plug and Play ! *Sci. STKE* 1791–13.

(3) Luck K., Charbonnier S., & Travé G. (2012) The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett.* 586(*17*); 2648–2661.

(4) Gogl G., Jané P., Caillet-Saguy C., Kostmann C., Bich G., Cousido-Siah A., Nyitray L., Vincentelli R., Wolff N., Nominé Y., Sluchanko N.N., & Travé G. (2020) Dual Specificity PDZ- and 14-3-3-Binding Motifs: A Structural and Interactomics Study. *Structure* 1–13.

(5) Fournane S., Charbonnier S., Chapelle A., Kieffer B., Orfanoudakis G., Travé G., Masson M., & Nominé Y. (2011) Surface plasmon resonance analysis of the binding of high-risk mucosal HPV E6 oncoproteins to the PDZ1 domain of the tight junction protein MAGI-1. *J. Mol. Recognit.* 24(*4*); 511–523.

(6) Ye F., & Zhang M. (2013) Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. *Biochem. J.* 455(*1*); 1–14.

(7) Mamonova T., Zhang Q., Khajeh J.A., Bu Z., Bisello A., & Friedman P.A. (2015) Canonical and noncanonical sites determine NPT2A binding selectivity to NHERF1 PDZ1. *PLoS One* 10(*6*); 1–17.

(8) Clairfeuille T., Mas C., Chan A.S.M., Yang Z., Tello-Lafoz M., Chandra M., Widagdo J., Kerr M.C., Paul B., Mérida I., Teasdale R.D., Pavlos N.J., Anggono V., & Collins B.M. (2016) A molecular code for endosomal recycling of phosphorylated cargos by the SNX27-retromer complex. *Nat. Struct. Mol. Biol.* 23(*10*); 921–932.

(9) Wiedemann U., Boisguerin P., Leben R., Leitner D., Krause G., Moelling K., Volkmer-Engert R., & Oschkinat H. (2004) Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *J. Mol. Biol.* 343(*3*); 703–718.

(10) Vouilleme L., Cushing P.R., Volkmer R., Madden D.R., & Boisguerin P. (2010) Engineering peptide inhibitors to overcome PDZ binding promiscuity. *Angew. Chemie - Int. Ed.* 49(*51*); 9912–9916.

(11) Amacher J.F., Cushing P.R., Brooks L., Boisguerin P., & Madden D.R. (2014) Stereochemical preferences modulate affinity and selectivity among five PDZ domains that bind CFTR: Comparative structural and sequence analyses. *Structure* 22(*1*); 82–93.

(12) Charbonnier S., Stier G., Orfanoudakis G., Kieffer B., Atkinson R.A., & Travé G. (2008) Defining the minimal interacting regions of the tight junction protein MAGI-1 and HPV16 E6 oncoprotein for solution structure studies. *Protein Expr. Purif.* 60(*1*); 64–73.

(13) Luck K., Fournane S., Kieffer B., Masson M., Nominé Y., & Travé G. (2011) Putting into practice domain-linear motif interaction predictions for exploration of protein networks. *PLoS One* 6(*11*); .

(14) Kozlov G., Gehring K., & Ekiel I. (2000) Solution structure of the PDZ2 domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptor. *Biochemistry* 39(*10*); 2572–2580.

(15) Gógl G., Biri-Kovács B., Durbesson F., Jané P., Nominé Y., Kostmann C., Bilics V., Simon M., Reményi A., Vincentelli R., Travé G., & Nyitray L. (2019) Rewiring of RSK–PDZ Interactome by Linear Motif Phosphorylation. *J. Mol. Biol.* 431(*6*); 1234–1249.

(16) Zhang Y., Yeh S., Appleton B.A., Held H.A., Kausalya P.J., Phua D.C.Y., Wong W.L., Lasky L.A., Wiesmann C., Hunziker W., & Sidhu S.S. (2006) Convergent and divergent ligand specificity among PDZ domains of the LAP and zonula occludens (ZO) families. *J. Biol. Chem.* 281(*31*); 22299–22311.

(17) Sundell G.N., Arnold R., Ali M., Naksukpaiboon P., Orts J., Güntert P., Chi C.N., & Ivarsson Y. (2018) Proteome‐wide analysis of phospho‐regulated PDZ domain interactions. *Mol. Syst. Biol.* 14(*8*); 1–22.

(18) Zeng L., & Zhou M. (2002) Bromodomain : an acetyl-lysine binding domain. *FEBS Lett.* 513124–128.

(19) Kouzarides T. (2000) Acetylation : a regulatory modification to rival phosphorylation ? *EMBO J.* 19(*6*); 1176–1179.

(20) Kim S.C., Sprung R., Chen Y., Xu Y., Ball H., Pei J., Cheng T., Kho Y., Xiao H., Xiao L., Grishin N. V., White M., Yang X.J., & Zhao Y. (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell* 23(*4*); 607–618.

(21) Ikenoue T., Inoki K., Zhao B., & Guan K.L. (2008) PTEN acetylation modulates its interaction with PDZ domain. *Cancer Res.* 68(*17*); 6908–6912.

(22) Chow L.M.L., & Baker S.J. (2006) PTEN function in normal and neoplastic growth. *Cancer Lett.* 241(*2*); 184–196.

(23) Takahashi Y., Morales F.C., Kreimann E.L., & Georgescu M.M. (2006) PTEN tumor suppressor associates with NHERF proteins to attenuate PDGF receptor signaling. *EMBO J.* 25(*4*); 910–920.

(24) Valiente M., Andrés-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(*32*); 28936–28943.

(25) Sotelo N.S., Schepens J.T.G., Valiente M., Hendriks W.J.A.J., & Pulido R. (2015) PTEN-PDZ domain interactions: Binding of PTEN to PDZ domains of PTPN13. *Methods* 77147–156.

(26) Terrien E., Chaffotte A., Lafage M., Khan Z., Préhaud 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(*237*); 1–12.

(27) Charbonnier S., Zanier K., Masson M., & Travé G. (2006) Capturing protein-protein complexes at equilibrium: The holdup comparative chromatographic retention assay. *Protein Expr. Purif.* 50(*1*); 89–101.

(28) Vincentelli R., Luck K., Poirson J., Polanowska J., Abdat J., Blémont M., Turchetto J., Iv F., Ricquier K., Straub M.-L., Forster A., Cassonnet P., Borg J.-P., Jacob Y., Masson M., Nominé Y., Reboul J., Wolff N., Charbonnier S., & Travé G. (2015) Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nat. Methods* 12(*8*); 787–93.

(29) Duhoo Y., Girault V., Turchetto J., Ramond L., Durbesson F., Fourquet P., Nominé Y., Cardoso V., Sequeira A.F., Bra J.L.A., Fontes C.M.G.A., Travé G., Wolff N., & Vincentelli R. (2019) High-throughput production of a new library of human single and tandem PDZ domains allows quantitative PDZ-Peptide interaction screening through high-throughput Holdup assay, in *Methods In Molecular Biology Clifton Nj*, pp 439–476.

(30) Khan Z., Terrien E., Delhommel F., Lefebvre-omar C., Vitry S., Bernard C., Ramirez J.,

Chaffotte A., Ricquier K., Vincentelli R., Buc H., Prehaud C., & Wolff N. (2019) Structure-based optimization of a PDZ-binding motif within a viral peptide stimulates neurite outgrowth. *J. Biol. Chem.* 294(*37*); 13755–13768.

(31) Bonhoure A., Forster A., Babah K.O., Gógl G., Eberling P., Kostmann C., Volkmer R., Mancilla V.T., Travé G., & Nominé Y. (2020) Benchtop Holdup Assay for Quantitative Affinity-Based Analysis of Sequence Determinants of Protein-Motif Interactions [published online ahead of print, 2020 May 16]. *Anal. Biochem.*

(32) Luck K., Charbonnier S., & Travé G. (2012) The emerging contribution of sequence context to the specificity of protein interactions mediated by PDZ domains. *FEBS Lett.* Federation of European Biochemical Societies.

(33) Jané P., Chiron L., Bich G., Travé G., & Nominé Y. (2020) A computational protocol to analyze PDZ/PBM affinity data obtained by high-throughput holdup assay. *Methods Mol. Biol.* In press.

(34) Tramesel D., Catherinot V., & Delsuc M.A. (2007) Modeling of NMR processing, toward efficient unattended processing of NMR experiments. *J. Magn. Reson.* 188(*1*); 56–67.

(35) Van Agthoven M.A., Chiron L., Coutouly M.A., Delsuc M.A., & Rolando C. (2012) Two-dimensional ECD FT-ICR mass spectrometry of peptides and glycopeptides. *Anal. Chem.* 84(*13*); 5589–5595.

(36) Simon M.A., Ecsédi P., Kovács G.M., Póti Á.L., Reményi A., Kardos J., Gógl G., & Nyitray L. (2019) High-throughput competitive fluorescence polarization assay reveals functional redundancy in the S100 protein family. *FEBS J.* 1–13.

(37) Roehrl M.H.A., Wang J.Y., & Wagner G. (2004) A general framework for development and data analysis of competitive high-throughput screens for small-molecule inhibitors of protein-protein interactions by fluorescence polarization. *Biochemistry* 43(*51*); 16056–16066.

(38) Shinde S.R., & Maddika S. (2017) PTEN Regulates Glucose Transporter Recycling by Impairing SNX27 Retromer Assembly. *Cell Rep.* 21(*6*); 1655–1666.

(39) Wu H., Feng W., Chen J., Chan L.N., Huang S., & Zhang M. (2007) PDZ Domains of Par-3 as Potential Phosphoinositide Signaling Integrators. *Mol. Cell* 28(*5*); 886–898.

(40) Kotelevets L., Hengel J., Bruyneel E., Mareel M., Roy F., & Chastre E. (2005) Implication of the MAGI‐1b/PTEN signalosome in stabilization of adherens junctions and suppression of invasiveness. *FASEB J.* 19(*1*); 115–117.

(41) Jurado S., Benoist M., Lario A., Knafo S., Petrok C.N., & Esteban J.A. (2010) PTEN is recruited to the postsynaptic terminal for NMDA receptor-dependent long-term depression. *EMBO J.* 29(*16*); 2827–2840.

(42) Dormeyer W., Ott M., & Schnölzer M. (2005) Probing lysine acetylation in proteins: Strategies, limitations, and pitfalls of in vitro acetyltransferase assays. *Mol. Cell. Proteomics* 4(*9*); 1226–1239.

(43) Hecht A., Laroche T., Strahl-Bolsinger S., Gasser S.M., & Grunstein M. (1995) Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: A molecular model for the formation of heterochromatin in yeast. *Cell* 80(*4*); 583–592.

(44) Huang R., Holbert M.A., Tarrant M.K., Curtet S., Colquhoun D.R., Dancy B.M., Dancy B.C., Hwang Y., Tang Y., Meeth K., Marmorstein R., Cole R.N., Khochbin S., & Cole P.A. (2010) Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation. *JACS* 132(*29*); 9986–9987.

(45) Wang X., & Hayes J.J. (2008) Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. *Mol. Cell. Biol.* 28(*1*); 227.

(46) Myers M.P., & Tonks N.K. (1997) PTEN: Sometimes taking it off can be better than putting it on. *Am. J. Hum. Genet.* 61(*6*); 1234–1238.

(47) Stambolic V., Suzuki A., De la Pompa J.L., Brothers G.M., Mirtsos C., Sasaki T., Ruland J., Penninger J.M., Siderovski D.P., & Mak T.W. (1998) Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell* 95(*1*); 29–39.

(48) Okumura K., Mendoza M., Bachoo R.M., DePinho R.A., Cavenee W.K., & Furnari F.B. (2006) PCAF modulates PTEN activity. *J. Biol. Chem.* 281(*36*); 26562–26568.

(49) Ivarsson Y., & Jemth P. (2019) Affinity and specificity of motif-based protein–protein interactions. *Curr. Opin. Struct. Biol.* Elsevier Ltd.

(50) Tonikian R., Zhang Y., Sazinsky S.L., Currell B., Yeh J.H., Reva B., Held H.A., Appleton B.A., Evangelista M., Wu Y., Xin X., Chan A.C., Seshagiri S., Lasky L.A., Sander C., Boone C., Bader G.D., & Sidhu S.S. (2008) A specificity map for the PDZ domain family. *PLoS Biol.* 6(*9*); 2043–2059.

(51) Appleton B.A., Zhang Y., Wu P., Yin J.P., Hunziker W., Skelton N.J., Sidhu S.S., & Wiesmann C. (2006) Comparative structural analysis of the Erbin PDZ domain and the first PDZ domain of ZO-1 Insights into determinants of PDZ domain specificity. *J. Biol. Chem.* 281(*31*); 22312–22320.

(52) Pascoe H.G., Gutowski S., Chen H., Brautigam C.A., Chen Z., Sternweis P.C., & Zhang X. (2015) Secondary PDZ domain-binding site on class B plexins enhances the affinity for PDZ-RhoGEF. *PNAS* 112(*48*); 14852–14857.

(53) Jadwin J.A., Ogiue-Ikeda M., & MacHida K. (2012) The application of modular protein domains in proteomics. *FEBS Lett.* 586(*17*); 2586–2596.

(54) Smialowski P., Pagel P., Wong P., Brauner B., Dunger I., Fobo G., Frishman G., Montrone C., Rattei T., Frishman D., & Ruepp A. (2009) The Negatome database: A reference set of non-interacting protein pairs. *Nucleic Acids Res.* 38(*SUPPL.1*); 540–544.

(55) Yang L., Wang Y., Chen P., Hu J., Xiong Y., Feng D., Liu H., Zhang H., Yang H., & He J. (2011) Na(+)/H(+) Exchanger Regulatory Factor 1 (NHERF1) is required for the estradiol-dependent increase of phosphatase and tensin homolog (PTEN) protein expression. *Endocrinology* 152(*12*); 4537–4549.

(56) Verrastro I., Tveen-Jensen K., Woscholski R., Spickett C.M., & Pitt A.R. (2016) Reversible oxidation of phosphatase and tensin homolog (PTEN) alters its interactions with signaling and regulatory proteins. *Free Radic. Biol. Med.* 9024–34.

Protein [a]	PDZ dom [b]	Method [c]	Ref	K_D [d]	PTEN_11 [e]	PTEN_Ac [e]	PTEN_KR [e]	PTEN_13 [e]
MAGI1	2	Co-IP	[40]		nd	nd	nd	nd
	3				30	10	15	10
MAGI2	2	Pull-down, IP, Co-IP	[24]		152	56	29	149
	3				47	15	21	30
MAGI3	2	Co-IP	[39]		nd	39	29	21
NHERF1	1	Pull-down, Co-IP	[55]		33	14	4	nm
NHERF2	1	Co-IP, Pull-down, Overlay assay	[23]		nd	nd	120	134
	2				35	nd	21	125
SNTB2	1	LC-MS	[56]		nd	nm	nd	64
SDCBP	1	LC-MS	[56]		nd	nd	nd	nd
PTPN13	2	Pull-down	[25]		148	nd	153	16
	4				nd	10	nd	36
DLG1	2	Pull-down	[24]		nd	36	71	81
DLG4	1				293	155	81	nd
	3	Co-IP	[41]		nd	nd	nd	nd
SNX27	1	ITC	[38]	38	14 (*)	4	8 (*)	6
MAST1	1	Pull-down	[24]		39	26	32	8 (*)
MAST2	1	ITC	[26]	2	13	4	7	1 (*)
MAST3	1	Pull-down	[24]		241	85	82	nd
MAST4	1	Pull-down	[24]		nd	nd	nd	nd
PARD3	1	FP	[39]	19	160	nd	56	96

Table 1. PDZ domains interactors for PTEN according to literature and the present study
Each row corresponds to a protein for which a binding to PTEN has been been described in literature. The main methods and the PDZ domain number are indicated. The four last columns contain information obtained by combining the holdup and FP methods in the present study.
[a] Protein name
[b] Domain interaction site for PTEN
[c] Detection methods described in literature
[d] Affinity provided in the literature when available (in µM)
[e] Affinity measured by holdup in this study (in µM)
[*] Affinity measured by FP in this study (in µM)
IP: Immunoprecipitation
Co-IP: Co-immunoprecipitation
nd: not detected in the holdup assay
nm: not measured in the holdup assay

Legends

Fig 1. Quality criteria and their conversion to the individual quality scores used to filter the holdup data. (**A**) A schematized electropherogram showing intensities of the normalization peak (*Norm_Int*) and of the MBP-PDZ peak (*Ref_Int*) visible in the red and blue regions, respectively. The region in green corresponds to the proteins of the crude extract, which is supposed to be kept low as compared to *Norm_Int* and *Ref_Int* in order to ensure that the MBP-PDZ is not underexpressed **(B)** The linear transformation used to superimpose the sample and reference electropherograms should be as neutral as possible: the *TranslX* translation factor and the *ScalX* scaling coefficient (>1 for dilation or <1 for a contraction) should be as close as possible to 0.0 and 1.0, respectively. **(C)** Profiles of the individual quality scores used to filter the data. In order to ensure that the analyzed samples were not too diluted, the scores vary linearly between 0 (low quality) and 1 (high quality) for the intensity of the normalization peak (Q_{norm}) or the MBP-PDZ peak (Q_{int}). Q_{2nd} is a quality score allowing to reject samples with low MBP-PDZ expression while Q_{delta} combines the *TranslX* and *ScalX* parameters and varies exponentially.
Double column fitting image.

Fig 2. Complementarity of holdup and fluorescent polarization data. The interaction data of PTEN_11 with MAST2-1, HTRA1-1 and SCRIB-3 are shown as examples of strong affinity, weak affinity or non-binding, respectively, all measured by holdup (**A**) and FP (**B**) methods. (**A**) After superimposition of the two electropherograms recorded for the PBM of interest (blue dotted line) and for the biotin reference (black solid line), the normalization of the electropherogram of the PBM compared to the one of the reference is done using the signal of the lysozyme added in every sample at a constant concentration (red peak). The region between 20 and 60 kDa which contains peaks of the crude extract supposedly to be constant, is used to verify the proper intensity normalization of the two electropherograms. The intensities of the peak of interest after proper alignment along the molecular weight scale (region covered by the green dotted line) are subsequently used to quantify the depletion of an individual PDZ domain. All those normalization and alignment steps are performed automatically and are important as the electropherogram overlap is never perfect. The holdup ultimately delivers "binding intensities" (BI) for each PBM/PDZ interaction pair, which in principle vary in a range from 0.00 (no binding) to 1.00 (strong binding). (**B**) In competitive FP measurements, polarization signal was recorded for increasing amounts of unlabeled peptide added to a solution of pre-formed PDZ / labeled peptide complex. The complexes consisted of MAST2-1, HTRA1-1 and SCRIB-3 mixed with 50 nM of labeled fpRSK1, fRSK1 and f16E6 peptides, respectively. The PDZ concentration depends on each sample and is adjusted to reach >50-80% complex formation to ensure a satisfactory signal-to-noise ratio. Each panel shows the average of three titration curves (black dots) and the fit results (red curves with the apparent K_D values) using competitive binding model.
Single column fitting image.

Fig 3. PDZ binding profiles of the four PTEN peptides. Holdup binding profiles obtained are shown for PTEN_11 (**A**), PTEN_Ac (**B**), PTEN_KR (**C**) and PTEN_13 (**D**). In each profile, the PDZ binders are ranked from left to right of the plot in BI decreasing order along the X-axis. Data for all the measured holdup data are shown. The grey dotted line shows the threshold for confidence value, set at BI = 0.20 (see main text). For each experiment, the number of PDZ

domains for which we obtained a measurement that passed the quality filtering step, and could therefore be included in the plot, is indicated (red case numbers). The holdup data for PDZ / PBM pairs with BI>0.20 are shown in Supp. Fig. S1.
Single column fitting image.

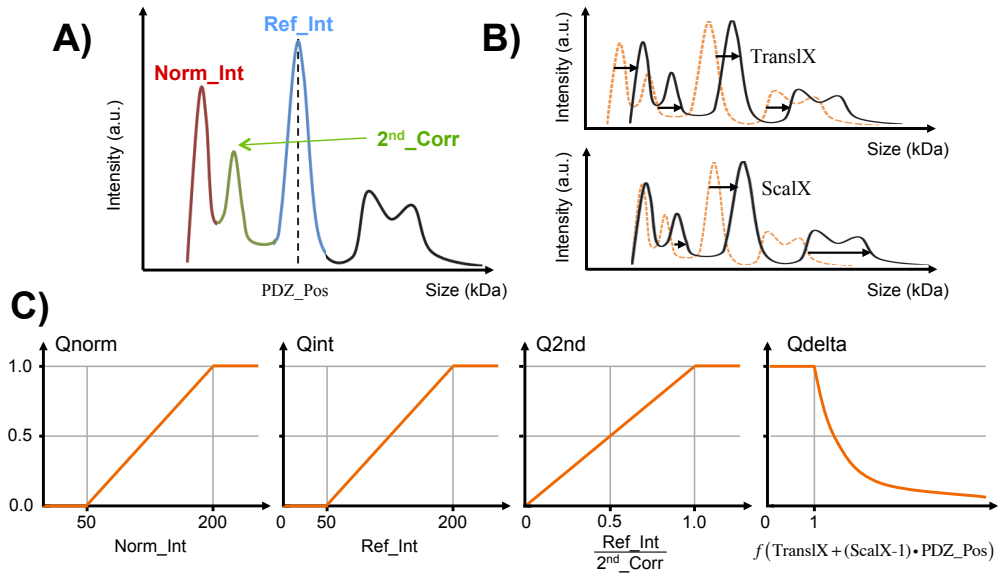
Fig 4. Conversion of the holdup binding intensities into affinities constants. (**A**) The violin plots shows the distribution of all the back-calculated apparent peptide concentrations obtained when both a quantifiable and significant (>0.20) BI value by holdup and a dissociation constant by FP were available for a given PDZ / PBM pair. The vertical line indicates the range of the distribution while the horizontal lines show the final mean peptide concentration and its final standard deviation after outlier exclusion (considering the 3σ rule). The final average peptide concentrations represented by the thick lines are used to convert the holdup BI values into K_D. (**B**) Comparison between the converted dissociation constants from the holdup assay and the dissociation constants directly measured by FP assay. The dotted line represents the perfect theoretical correlation. Since the data points seem to be randomly distributed on both sides of this dotted line, the R^2 is indicative of the goodness of fit.
Single column fitting image.

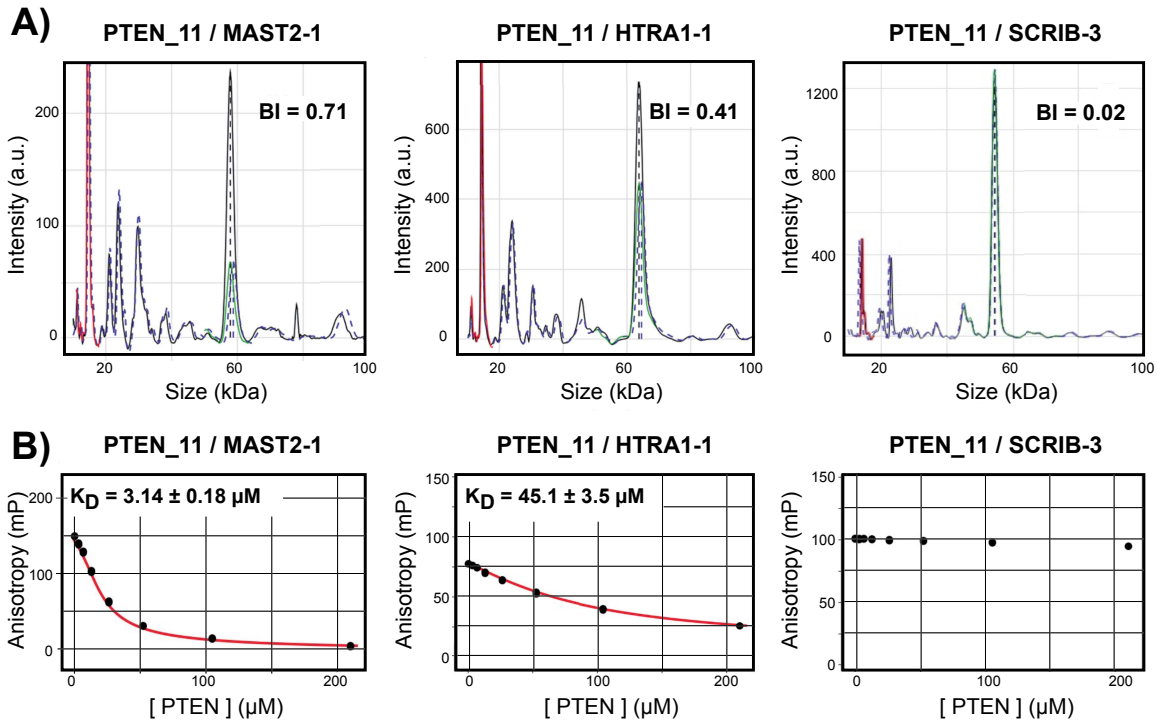
Fig 5. Determination of the specificity index for the PTEN binding profiles. For every profile, the significant PDZ binder affinity values are ranked from left to right along the X-axis in -log(K_D) decreasing order. The non-significant or undetected binders were omitted for clarity. The grey dotted line corresponds to the threshold BI value after converting it into -log(K_D) scale, while the blue and red dotted lines represent the highest affinity and the affinity at half the difference between the maximal and weakest significant affinity values, respectively. The reader can note that, for a constant threshold BI value (0.20), the weakest affinity values may vary moderately due to non-constant peptide concentrations. The numbers of PDZ domains above the half-maximal binding affinity" are indicated in red, while the numbers of tested and validated PDZ domains are in green. Values calculated for the promiscuity index (I_P) and the specificity index (I_S) are given (see main text). Full data sets for holdup and FP are visible in **Supp. Fig. S1 and S2,** respectively.
Double column fitting image.

Fig 6. Domain representations of the impacted PDZ domains by the different PTEN peptides. Proteins containing PDZ domains significantly bound to one PTEN peptide are colored and ranked from strongest to weakest binding strength depending on the best individual PDZ binder within each protein. The color code from white to black is indicative of the -log(K_D) values in the range of 4.0 – 6.0 after filtering step and BI conversion. The symbol (#) denotes PDZ domain for which the BI value could not be measured directly by holdup and has been inferred from FP measurements. Protein names appeared in bold when significant –log(K_D) values are observed for the four PTEN PBM.
Double column fitting image.

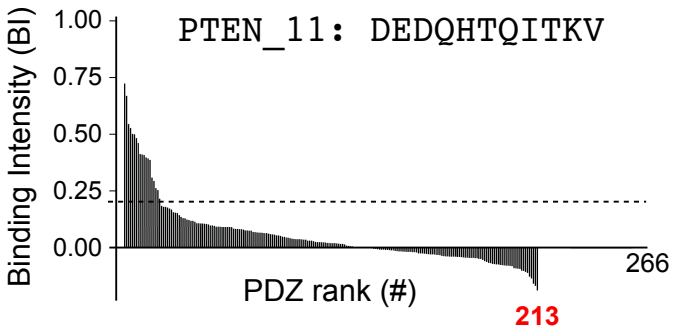
Fig 7. Changes in the PDZ binding profiles induced by changes in the PTEN peptides. (**A**) Comparison between PTEN_11 (grey), PTEN_KR (light purple) and PTEN_Ac (dark blue) using a shared PDZ axis. For the wild-type PTEN_11 peptide, the PDZ domains were ranked in descending affinity order along the X-axis, from left to right according to the significant affinities for PTEN_11, and from right to left according to the significant affinities solely detected for PTEN_13. The remaining PDZ domains that bind only to the PTEN_KR peptide were added in the middle region. (**B**) Comparison between PTEN_11 (grey) and PTEN_13

(orange) on a shared PDZ axis. The PDZ domains were ranked along the X-axis in descending order, from left to right according to the significant affinities for PTEN_11, and from right to left according to the significant affinities exclusively detected for PTEN_13. The left and right regions thus show PDZ domains that prefer the shorter or the longer PTEN PBM version, respectively. The overall uncertainty on log(K_D) values was estimated to be roughly ± 0.2 in log(M) unit by propagating BI uncertainty estimated in previous studies.
Double column fitting image.

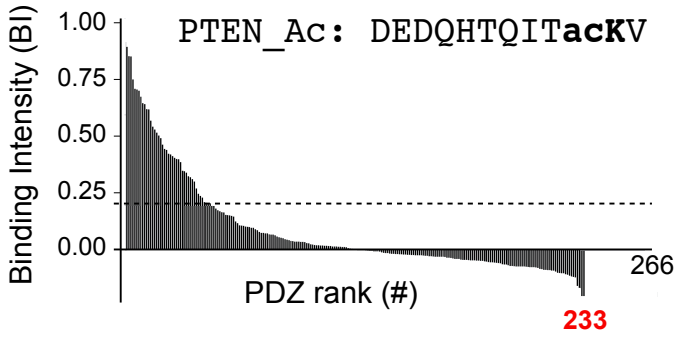

A)
Ref_Int
Norm_Int
2nd_Corr
Intensity (a.u.)
PDZ_Pos
Size (kDa)
B)
Intensity (a.u.)
TranslX
Size (kDa)
Intensity (a.u.)
ScalX
Size (kDa)
C)
Qnorm
1.0
0.5
0.0
50
200
Norm_Int
Qint
0
50
200
Ref_Int
Q2nd
0
0.5
1.0
Ref_Int
2nd_Corr
Qdelta
0
1
f(TranslX+(ScalX-1)•PDZ_Pos)


A)
PTEN_11 / MAST2-1
PTEN_11 / HTRA1-1
PTEN_11 / SCRIB-3
BI = 0.71
BI = 0.41
BI = 0.02
Intensity (a.u.)
Size (kDa)
B)
PTEN_11 / MAST2-1
PTEN_11 / HTRA1-1
PTEN_11 / SCRIB-3
K_D = 3.14 ± 0.18 µM
K_D = 45.1 ± 3.5 µM
Anisotropy (mP)
[PTEN] (µM)

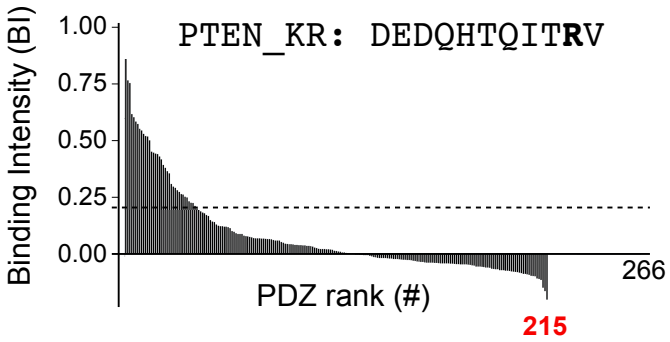
A)


PTEN_11: DEDQHTQITKV
Binding Intensity (BI)
1.00
0.75
0.50
0.25
0.00
PDZ rank (#)
266
213

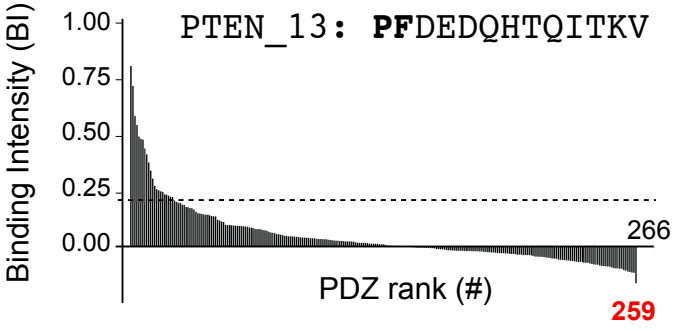
B)

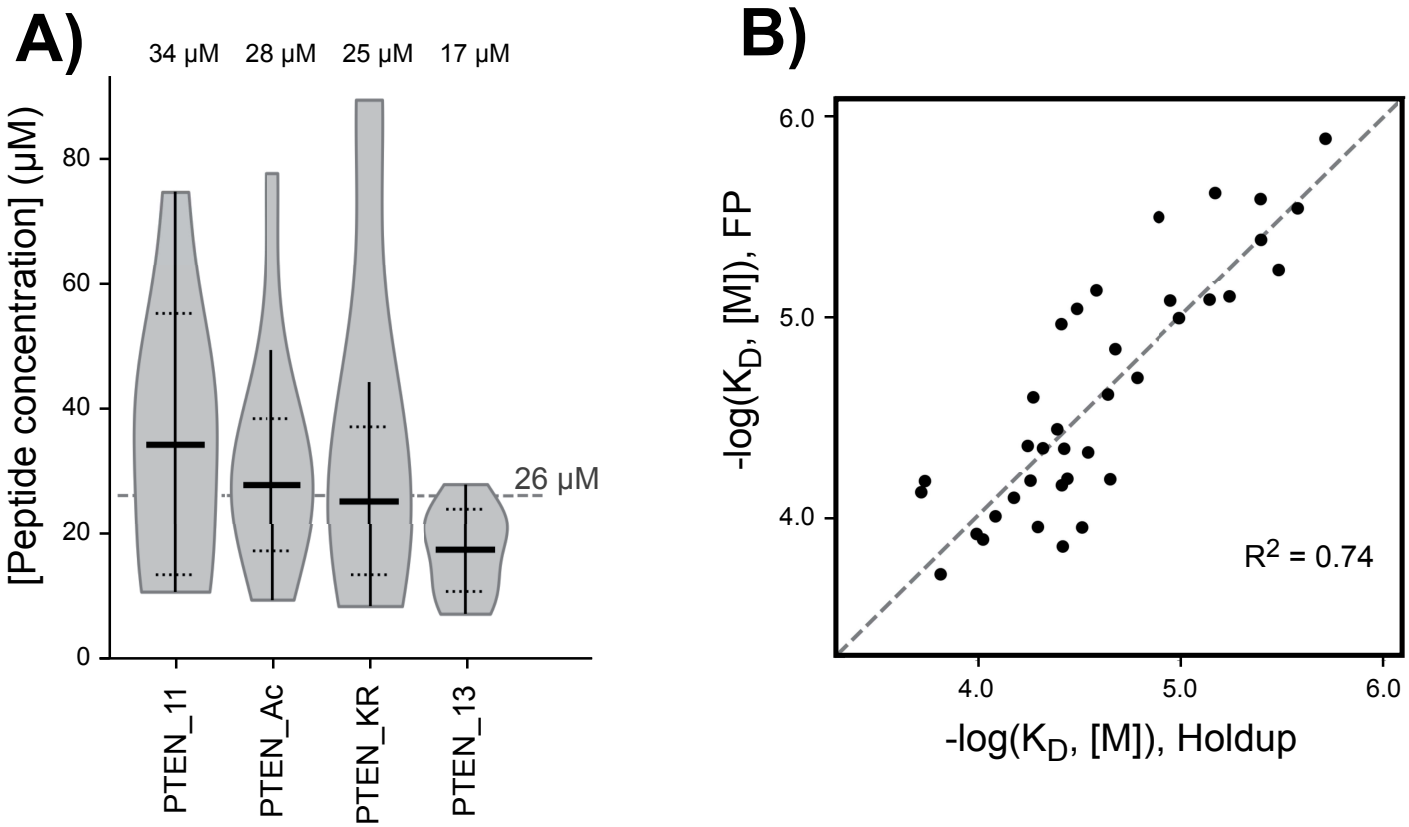

PTEN_Ac: DEDQHTQITacKV
Binding Intensity (BI)
1.00
0.75
0.50
0.25
0.00
PDZ rank (#)
266
233

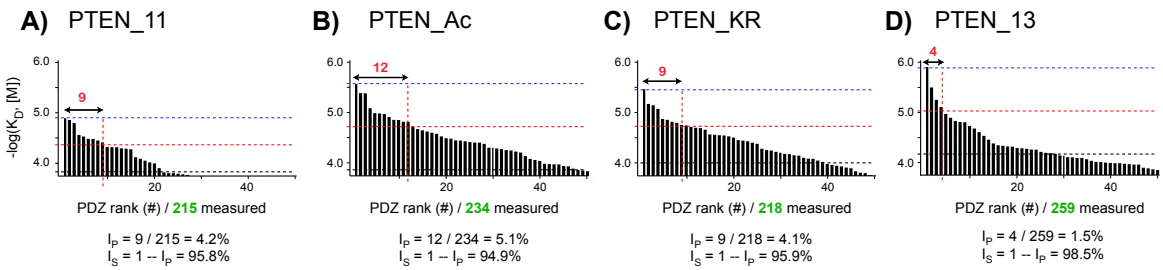
C)

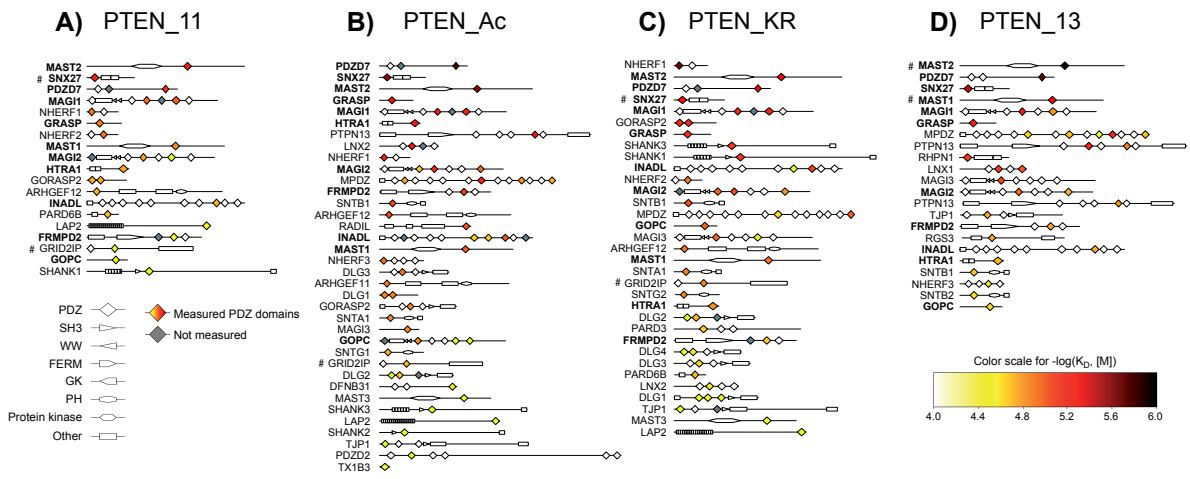

PTEN_KR: DEDQHTQITRV
Binding Intensity (BI)
1.00
0.75
0.50
0.25
0.00
PDZ rank (#)
266
215

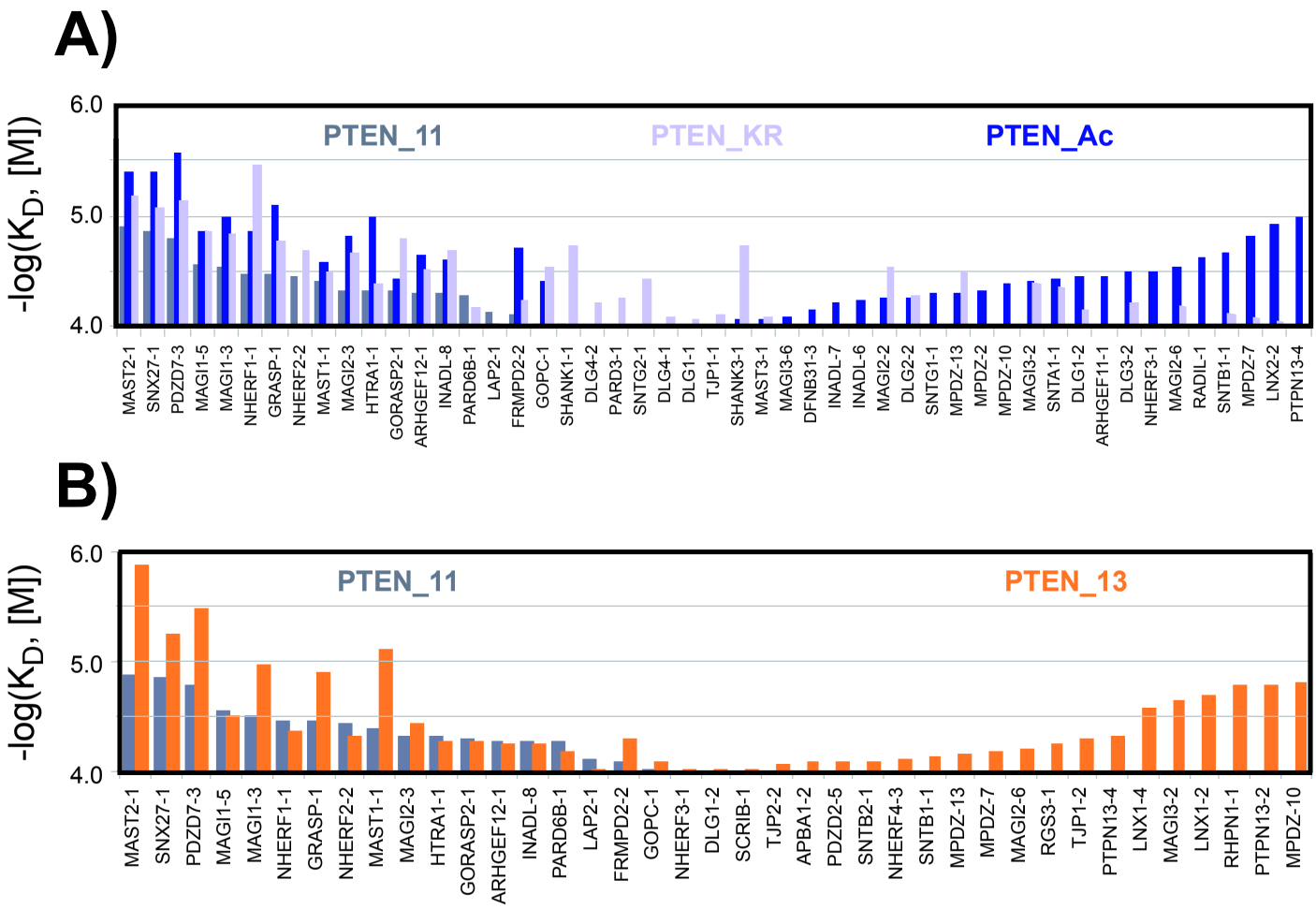
D)


PTEN_13: PFDEDQHTQITKV
Binding Intensity (BI)
1.00
0.75
0.50
0.25
0.00
PDZ rank (#)
266
259


A)
34 µM
28 µM
25 µM
17 µM
[Peptide concentration] (µM)
80
60
40
20
0
26 µM
PTEN_11
PTEN_Ac
PTEN_KR
PTEN_13
B)
6.0
5.0
4.0
-log(K_D, [M]), FP
R^2 = 0.74
4.0
5.0
6.0
-log(K_D, [M]), Holdup


A) PTEN_11
B) PTEN_Ac
C) PTEN_KR
D) PTEN_13
-log(K_D, [M])
6.0
5.0
4.0
20
40
9
12
9
4
PDZ rank (#) / 215 measured
PDZ rank (#) / 234 measured
PDZ rank (#) / 218 measured
PDZ rank (#) / 259 measured
I_P = 9 / 215 = 4.2%
I_S = 1 -- I_P = 95.8%
I_P = 12 / 234 = 5.1%
I_S = 1 -- I_P = 94.9%
I_P = 9 / 218 = 4.1%
I_S = 1 -- I_P = 95.9%
I_P = 4 / 259 = 1.5%
I_S = 1 -- I_P = 98.5%


A) PTEN_11
B) PTEN_Ac
C) PTEN_KR
D) PTEN_13
MAST2
SNX27
PDZD7
MAGI1
NHERF1
GRASP
NHERF2
MAST1
MAGI2
HTRA1
GORASP2
ARHGEF12
INADL
PARD6B
LAP2
FRMPD2
GRID2IP
GOPC
SHANK1
PDZ
SH3
WW
FERM
GK
PH
Protein kinase
Other
Measured PDZ domains
Not measured
PDZD7
SNX27
MAST2
GRASP
MAGI1
HTRA1
PTPN13
LNX2
NHERF1
MAGI2
MPDZ
FRMPD2
SNTB1
ARHGEF12
RADIL
INADL
MAST1
NHERF3
DLG3
ARHGEF11
DLG1
GORASP2
SNTA1
MAGI3
GOPC
SNTG1
GRID2IP
DLG2
DFNB31
MAST3
SHANK3
LAP2
SHANK2
TJP1
PDZD2
TX1B3
NHERF1
MAST2
PDZD7
SNX27
MAGI1
GORASP2
GRASP
SHANK3
SHANK1
INADL
NHERF2
MAGI2
SNTB1
MPDZ
GOPC
MAGI3
ARHGEF12
MAST1
SNTA1
GRID2IP
SNTG2
HTRA1
DLG2
PARD3
FRMPD2
DLG4
DLG3
PARD6B
LNX2
DLG1
TJP1
MAST3
LAP2
MAST2
PDZD7
SNX27
MAST1
MAGI1
GRASP
MPDZ
PTPN13
RHPN1
LNX1
MAGI3
MAGI2
PTPN13
TJP1
FRMPD2
RGS3
INADL
HTRA1
SNTB1
NHERF3
SNTB2
GOPC
Color scale for -log(K_D, [M])
4.0
4.4
4.8
5.2
5.6
6.0


A)
-log(K_D, [M])
6.0
5.0
4.0
PTEN_11
PTEN_KR
PTEN_Ac
MAST2-1
SNX27-1
PDZD7-3
MAGI1-5
MAGI1-3
NHERF1-1
GRASP-1
NHERF2-2
MAST1-1
MAGI2-3
HTRA1-1
GORASP2-1
ARHGEF12-1
INADL-8
PARD6B-1
LAP2-1
FRMPD2-2
GOPC-1
SHANK1-1
DLG4-2
PARD3-1
SNTG2-1
DLG4-1
DLG1-1
TJP1-1
SHANK3-1
MAST3-1
MAGI3-6
DFNB31-3
INADL-7
INADL-6
MAGI2-2
DLG2-2
SNTG1-1
MPDZ-13
MPDZ-2
MPDZ-10
MAGI3-2
SNTA1-1
DLG1-2
ARHGEF11-1
DLG3-2
NHERF3-1
MAGI2-6
RADIL-1
SNTB1-1
MPDZ-7
LNX2-2
PTPN13-4

B)
-log(K_D, [M])
6.0
5.0
4.0
PTEN_11
PTEN_13
MAST2-1
SNX27-1
PDZD7-3
MAGI1-5
MAGI1-3
NHERF1-1
GRASP-1
NHERF2-2
MAST1-1
MAGI2-3
HTRA1-1
GORASP2-1
ARHGEF12-1
INADL-8
PARD6B-1
LAP2-1
FRMPD2-2
GOPC-1
NHERF3-1
DLG1-2
SCRIB-1
TJP2-2
APBA1-2
PDZD2-5
SNTB2-1
NHERF4-3
SNTB1-1
MPDZ-13
MPDZ-7
MAGI2-6
RGS3-1
TJP1-2
PTPN13-4
LNX1-4
MAGI3-2
LNX1-2
RHPN1-1
PTPN13-2
MPDZ-10

11.3 Validating the Affinity Vs Specificity map through the showed BI Profiles in this thesis

In this chapter, we presented for the first time a manner to quantify specificity from our holdup results (see previous section). The aim in this section will be to assess specificity quantification with the other presented results in this thesis. As a refreshment from the papers, here are the residue composition of the used PBMs:

- **16E6:** SSRTRRETQL (see chapter 10)
- **16E6_-2P:** SSRTRRE**pT**QL (see chapter 10)
- **16E6_-2E:** SSRTRRE**E**QL (see chapter 10)

- **RSK1:** RRVRKLPSTTL (see chapter 10)
- **RSK1_-3P:** RRVRKLP**pS**TTL (see chapter 10)
- **RSK1_-3E:** RRVRKLP**E**TTL (see chapter 10)

- **PTEN_11:** DEDQHTQITKV (see chapter 11)
- **PTEN_KR:** DEDQHTQIT**R**V (see chapter 11)
- **PTEN_Ac:** DEDQHTQIT**acK**V (see chapter 11)
- **PTEN_13:** **PF**DEDQHTQITKV (see chapter 11)

As we have seen, the introduction of variations in the PBM affects up to 10 fold the affinity strength as well as the specificity. Once we calculate the half-maximal binding promiscuity index (see previous section), we plot it together with the max affinity of each BI profile on a scatter plot. This way, we can show the x- (specificity) and y- (affinity) axis descriptors that modulates our BI profile into a 2D map. When looking at figure 11.1, the conclusions mentioned in the presented papers are also visible here. We can see that not only affinities are different between phosphorylated or phosphomimicked residues for RSK1 and 16E6, but also the now quantified specificities are showed quantitatively different. This confirms what it was showed in the presented paper in chapter 10.

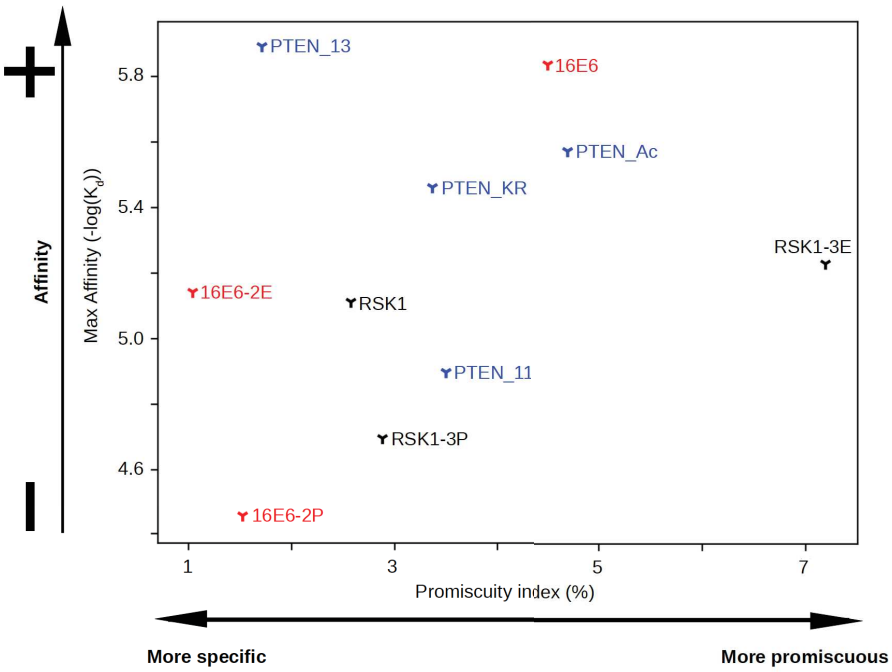


Figure 11.1: **Quantitative affinity-specificity map.** An Affinity Vs Specificity map is plotted to quantitatively asses the differences between the different profiles according to affinity and specificity (colored by PBM origin).

On the other hand the values obtained for PTEN_11 and PTEN_13 confirm clear differences in both affinity and specificity. PTEN_Ac and PTEN_KR are found to be closer to each other in terms of affinity and specificity. However, in the work of Ikenoue *et al.* [25], the original purpose of the PTEN_KR mutation was to make it more akin to a non acetylable PTEN (represented by PTEN_11 in our peptides) rather than to an acetylated PTEN (represented by PTEN_Ac in our peptides). The result is opposite to their expectation, as already showed in the presented paper in this chapter.

Overall we can see that mapping the PBMs assayed with the holdup into an Affinity-Specificity map helps to observe at glance, and quantitatively, the impact of PTM introduction, mutations or even differences between unrelated PBMs.

Part V

Discussion

Chapter 12

The interactome

12.1 The human interactome

The human interactome is a large subset of PPI which are related in a given biological system. In the case of the human proteome, it involves approximately 26 000 genes encoding 60 000 different proteins when accounting for splicing isoforms. Furthermore, these can be altered by multiple combinations of PTM, as already explained. Interactions between proteins are (as discussed previously) the core of any cellular process, which makes the rendering of these into a protein network relevant. It allows us to extract the overall functional organization of proteins in the cell [161]. Currently, the efforts are ongoing to "map" the human interactome assessing the 26 000 x 26 000 potentials interactions, excluding the spliced isoforms and multiple PTMs mentioned above [8, 140, 144].

In a given interactome the specificity emerges from the comparison of individual affinities within the same protein family. Most interactomic approaches deliver binary results (either "bind" or "not bind") [4, 5]. However, this binary information is an inaccurate and incomplete way to describe interactomes, since protein-protein interactions can display a very large palette of affinities. Equilibrium affinity constants (K_d) can range from the picomolar scale (10^{-12} M) to the millimolar scale (10^{-3} M) [6–8]. For this purpose, many large-scale technology tools have been developed to report large number of PPI to study a "subinteractome" given by a simplified system or a protein recognition family (like the PDZ domain family) [6–8].

12.2 The PDZome interactome

The full mapping of the affinities of the entire human PBM/PDZ network will require the measurement of 266 PDZ domains against 3617 putative PBM, if possible, in a quantitative affinity manner.

Measuring the affinity constants for an entire domain family (like the PDZ domains) is the best way to determine the specificity of a domain for a motif and to compare it with all the possible competing interactions involving either partner of the interaction. Computational prediction would be an appealing alternative, but to our knowledge there is no robust and accurate approach for predicting domain-motif affinities so far. Specially in the PDZ domain context, where PBM/PDZ binding has at least two critical disorders states: the PBM and the β2-β3 loop from the PDZ domains, increasing exponentially the variables that we should take into account. Therefore, the specificity of interactions is often tackled either by qualitative approaches (interaction assays performed in the presence of a large amount of potential competitors in background, washing steps/buffers supposed to retain only specific interactions, etc. . .) or by comparative assessment of small number of affinities for a given system [4, 5, 165].

Measuring domain-motif affinities at high-throughput thus represents a possible approach to describe quantitatively the specificity of domain-mediated interaction networks at a proteome wide scale [7]. Here, we presented a simple but quantitative approach to assess the specificity of a PBM/PDZome interactions allowing a comparison between wild-type PBMs and their mutations or modification by PTMs. Moreover, we introduce for the first time a specificity quantification measure for a whole interactome. This will allow in the future to further compare different PBM/PDZ profiles.

12.3 Future perspectives in the interactome study

Along all this thesis, I insisted in the relevance of measuring binding affinities in a quantitative way, but we already saw that not all the HTS methods are suitable to deliver such values. Instead, they compensate the lack of the affinity value with a huger interactomic data set measured in a semi-quantitative or qualitative way

("bind" or "does not bind").

Computers are perfect tools to deal with thousands of produced data by large-scale assays, and as consequence, some prediction tools started to be built taking profit of the produced data. In the past years, computers made a quality jump allowing to make tons of calculations in a shorter time. One may think that this could be the key of the success to produce predictors, but in reality, the sample and the quality data are the key of it. For instance, the prediction of most PPI relies on the amino acid representation, which are characterized by one-letter code, and therefore, in a qualitative manner, but biological networks in cells do not work in a qualitative way. Therefore, it implies a loss of resolution of the information in the computational study [166]. Some analysis requires to take into account extra information like the intrinsic properties of such letters, and more concretely, their physicochemical properties [166, 167]. To this, we need to stress that most predictors are not even taking into account more than a few amino acids of the sequence from both, ligand and binding domain, to perform their predictions.

In the PBM/PDZ predictors case, these are based on a total of 1000 to 2000 interactions measurements. Considering the putative PDZ interactome (266 PDZ domains x 3617 PBM), this does not even represents the 1% of the total of its interactome, and therefore, the trained dataset will be very limited and exclusively working for a very specific and reduced dataset.

This could explain why PBM/PDZ interaction predictors were poorly performant according to Katja Luck's thesis [168]. Since then, some of the public available PBM/PDZ predictors have actually disappeared, like iSPOT [169] or the one made by Wiedemann *et al.* [170], perhaps due lack of reliability or maintenance costs. Other predictors like POW!, from Bader's lab [171, 172], are poorly performant when comparing with our holdup results. One can barely find similitude between our holdup database and their predicted bindings.

Computational predictions should start to consider amino acids as a vectors, which could contain a range of up to 20 physicochemical properties (such as volume, charge, polarity etc) [167]. This way, we could treat the amino acids as Euclidian Vectors in the space and compute the metric distance between them [167]. The computation of amino acid differences will be the result of comparing metric matrices and the quantification of real distances. Many prediction tools might increase their reliability using such method instead of simply relying on

the amino acid letter code. I did something similar on a side-project with Dr. Bálint Mészáros (not shown in this thesis), which was correlating pretty well with the experimental results and could had work as a tool to find similar PBMs to asses with the holdup. However, the bottle-neck in this kind of projects always is going to be the same: the amount of available data, which are always too low, and its quality. Computational approaches can only be validated with a strong benchmarking using many affinity data i.e. from HTS methods. Robust HTS methods are the key to explore the landscape and improve benchmarking of the computational PPI approaches, and therefore, to serve as starting point to more reliable computational tools using strong algorithms, such as machine learning.

Chapter 13

The holdup

13.1 Insight into the holdup: a powerful measurement tool

Domain-motif networks are often modulated by reversible PTMs. Phosphorylation and acetylation processes are difficult to reproduce in vitro. While some strategies aim at mimicking these processes [6, 25–29], others rather focus on the impact of PTMs on interactomes [30]. Here, we used chemically synthetized PBMs to reproduce different conditions, such as a wild-type, acetylation or phosphorylation, or addition of extra exosites, and then to test residue mimication of the literature. These peptides were used for interaction studies using the holdup assay, an assay originally developed in our laboratory. The holdup assay is a comparative chromatographic retention approach that quantifies the equilibrium binding affinity of proteins towards ligands. We observed how PTM as well as extra residues addition to a given PBM can switch its PDZ domain binding preferences. Moreover, we analyzed the "PDZ-binding impacting potential" of particular residue positions of PBMs when they are mutated or subject to a PTM.

The experimental holdup assay requires a rigorous approach to warrant its best performances. Several critical steps are identified in the following stages: the expression of the PDZome, the verification of its quality and the calibration of its concentrations. Since the first automatized holdup assay was published [7], further improvements have been implemented to reach high quality for the mentioned stages [31].

A first aspect concerns the optimization of the cost- and time-effectiveness of measurements. Now, we performed holdup measurements in singlicate, also allowing to improve the number of different PDZ-PBM pairs that we can address. To make sure about the accuracy of our data, we checked the PBM/PDZ interaction point by point from its expression levels to the electropherogram superimposition step.

A second point concerns the quality of PDZome extracts. An optimal preparation does not always warrant the quality of the measurements, as protein degradation may occur during storage. We implemented a computational processing step to better superimpose the caliper's electropherograms leading to a better reliability of the holdup assay. These quality control steps allow us to improve accuracy and precision of the holdup data, reducing the amount of false positive and negative results. The protocol showed in the thesis has been applied to over 20 000 pairs of interactions resulting in a curated PBM/PDZ interaction database.

Finally, we validated subsets of the positive interactions resulting from the holdup assay by Fluorescent Polarization (FP), an orthogonal powerful approach for quantitative binding affinity measurements that uses fluorescently labelled peptides. The binding affinity data obtained by FP were then used to interpolate all the binding affinity constants (K_d) measured by the holdup assays.

Worthy of note, the holdup assay delivers powerful information about both, the PDZome "interactome" and the "negatome" [32]. Note that negatome in this context does not mean a strict not binding case but our limit of affinity quantification. Instead of providing "binary affinities" ("binds" or "does not bind"), we quantified PBM interactions against the whole PDZome. This provides us with an extra information to dig further in the PDZome network. The interactome shows the potential preferences of the PBM for a given subset of the PDZ domains assuming equal conditions in a giving system (whereas, in a cellular system, the expression and localization of the proteins comprising those PBMs and PDZs may greatly vary). Despite its high sensitivity, the holdup assay usually does not detect more than 50 PDZ domains "binders" for a given PBM [7, 31, 33, 34]. This means that a "binders-only" interactome would deliver less than 20% of the PDZome network information. Yet, the specificity is determined, not only by the actual interactions, but also by all the interactions that do not take place (or, to be more accurate, that are beyond the detection threshold of the binding assay utilized). Therefore, the study of the negatome is highly relevant, especially when considering

that it also allows us to explore the "grey zone" of interactions that only happen under certain circumstances (high copy number protein, protein localization, and so forth). By adding the negatome to our studies, we actually cover up to 100% of the PBM/PDZome network. This allows us to further study variations of PBM, such as PTM and mutations, in the PDZome. Therefore, a strong point of interest of the holdup assay is its ability to provide, thanks to the systematic measurement of binding affinities, quantitative information on binding specificities. In this thesis, we propose a simple yet robust approach to quantify the specificity of any measured PBM against the whole PDZome using the "PDZome-binding profiles" obtained from the holdup assay. We used this approach to compare the specificities of PBMs either in wild-type form or subjected to mutations or modification by PTMs.

13.2 Relevance of the holdup in protein interactomics and its future perspectives

An increasing number of PDZ biochemists -particularly in the "PDZnet" Marie Skłodowska-Curie international training network I was part from- have realized the power of the holdup assay and used it for their research. The demand for this assay will certainly increase in the future. The method first required to express all the PDZ library. This work was done by Vincentelli *et al.* [7], another member of the network. The holdup assay was performed by a consortium of three laboratories including ours. Finally, since I was deeply involved in the data curation, I had access to -and could treat using my approaches- most of the holdup data measured in the team network.

All the data curation performed in this thesis will allow the holdup to face one of its main problems: the costs. Our lab is currently improving the method by using fluorescence to quantify the affinity data. This will allow us to avoid the caliper step and to reduce the complexity of the bioinformatics processes. The use of FP will be still necessary to validate the data and calculate PBM concentrations. However, this cannot be done without a strong and robust dataset like the one we possess now (over 20 000 interactions). The new method will be faster and more cost-effective, paving the way to the future completion of the full quantitative PBM/PDZome interactome affinity map.

The holdup measures domain-motif affinities. In cells, proteins are full-length and therefore they may display affinities, which differ from the fragmental domain-motif affinities we measured in this thesis due to contextual effects, diverse conformational states, and so forth. Furthermore, the expression, the localization, the concentration, the turnover, the PTM of proteins may vary depending on the cell type and on the cell state. Nevertheless, it remains that the domain-motif affinities that we measure are intrinsic physical properties of elements of these cellular systems. These physical properties are then used by cells to produce phenotype. One measurable and quantifiable phenotype is the expressed transcriptome. Therefore, our lab is now starting to move into the transcriptomic field to investigate how intrinsic quantitative interactome properties such as those we could measure are utilized by cells to produce an expressed transcriptome.

Chapter 14

Conclusions

In this thesis the impact of PBM modifications (mimicking mutations in the core, PTM or addition of exosites) in its interaction with the PDZome is shown. These modifications lead to a global change in the binding profiles -and therefore in the PDZ-binding capability- providing quantitative information on the biological effect that such modifications may have in the context of full-length proteins. As any alteration in the sequence of the PBM (like mutations, modifications, or the change of the peptide length) may lead to global rearrangements of the PDZ binding profile, such alterations are to be used and interpreted with great caution considering additional structural, cellular or even whole-organism studies. The crystal structure solved for a "non-complex" between PDZ domain and a "switcher-off" PBM phosphorylated at p-2 (see chapter 10 of this thesis) shows that binding may happen below the detection limits of most usual methods. Such findings could not have been possible without the holdup assay, and the subsequent data curation strategy that we developed, allowing to obtain robust affinity data. We also proposed a way to quantify and compare the specificities of particular motifs towards a whole subset of the proteome (PDZome). Altogether, the proposed approach for the quantification of affinity and specificity of motif-domain networks, might bring the affinity-based unraveling of the quantitative human interactome a tiny bit closer.

Bibliography

[1] Rolf Apweiler, Amos Bairoch, Cathy H Wu, Winona C Barker, Brigitte Boeckmann, Serenella Ferro, Elisabeth Gasteiger, Hongzhan Huang, Rodrigo Lopez, Michele Magrane, Maria J Martin, Darren A Natale, Claire O'Donovan, Nicole Redaschi, and Lai-Su L Yeh. Uniprot: the universal protein knowledgebase. *Nucleic acids research*, 32(suppl_1):D115–D119, 2004.

[2] Javier De Las Rivas and Celia Fontanillo. Protein–protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS computational biology*, 6(6), 2010.

[3] Alexei Vazquez, Alessandro Flammini, Amos Maritan, and Alessandro Vespignani. Global protein function prediction from protein-protein interaction networks. *Nature biotechnology*, 21(6):697–700, 2003.

[4] Raffi Tonikian, Yingnan Zhang, Stephen L Sazinsky, Bridget Currell, Jung-Hua Yeh, Boris Reva, Heike A Held, Brent A Appleton, Marie Evangelista, Yan Wu, Xiaofeng Xin, Andrew C Chan, Somasekar Seshagiri, Laurence A Lasky, Chris Sander, Charles Boone, Gary D Bader, and Sidhu Sachdev S. A specificity map for the pdz domain family. *PLoS biology*, 6(9), 2008.

[5] Katja Luck and Gilles Travé. Phage display can select over-hydrophobic sequences that may impair prediction of natural domain–peptide interactions. *Bioinformatics*, 27(7):899–902, 2011.

[6] Gustav N Sundell, Roland Arnold, Muhammad Ali, Piangfan Naksukpaiboon, Julien Orts, Peter Güntert, Celestine N Chi, and Ylva Ivarsson. Proteome-wide analysis of phospho-regulated pdz domain interactions. *Molecular systems biology*, 14(8), 2018.

[7] Renaud Vincentelli, Katja Luck, Juline Poirson, Jolanta Polanowska, Julie Abdat, Marilyne Blémont, Jeremy Turchetto, François Iv, Kevin Ricquier,

Marie-Laure Straub, Anne Forster, Patricia Cassonnet, Jean-Paul Borg, Yves Jacob, Murielle Masson, Yves Nominé, Jérôme Reboul, Nicolas Wolff, Sebastian Charbonnier, and Travé Gilles. Quantifying domain-ligand affinities and specificities by high-throughput holdup assay. *Nature methods*, 12(8):787–793, 2015.

[8] Daniel Figeys. Mapping the human protein interactome. *Cell research*, 18(7):716, 2008.

[9] Joshua A Jadwin, Mari Ogiue-Ikeda, and Kazuya Machida. The application of modular protein domains in proteomics. *FEBS letters*, 586(17):2586–2596, 2012.

[10] Francesca Diella, Niall Haslam, Claudia Chica, Aidan Budd, Sushama Michael, Nigel P Brown, Gilles Trave, and Toby J Gibson. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci*, 13(6580):603, 2008.

[11] Norman E Davey, Kim Van Roey, Robert J Weatheritt, Grischa Toedt, Bora Uyar, Brigitte Altenberg, Aidan Budd, Francesca Diella, Holger Dinkel, and Toby J Gibson. Attributes of short linear motifs. *Molecular BioSystems*, 8(1):268–281, 2012.

[12] Roberto Mosca, Roland A Pache, and Patrick Aloy. The role of structural disorder in the rewiring of protein interactions through evolution. *Molecular & Cellular Proteomics*, 11(7), 2012.

[13] Peter Tompa, Norman E Davey, Toby J Gibson, and M Madan Babu. A million peptide motifs for the molecular biologist. *Molecular cell*, 55(2):161–169, 2014.

[14] T Hunt. Protein sequence motifs involved in recognition and targeting: a new series. *Trends Biochem. Sci*, 15(8):305–309, 1990.

[15] Katja Luck, Sebastian Charbonnier, and Gilles Travé. The emerging contribution of sequence context to the specificity of protein interactions mediated by pdz domains. *FEBS letters*, 586(17):2648–2661, 2012.

[16] Kyung-Ok Cho, Carol A Hunt, and Mary B Kennedy. The rat brain postsynaptic density fraction contains a homolog of the drosophila discs-large tumor suppressor protein. *Neuron*, 9(5):929–942, 1992.

[17] Declan A Doyle, Alice Lee, John Lewis, Eunjoon Kim, Morgan Sheng, and Roderick MacKinnon. Crystal structures of a complexed and peptide-free membrane protein–binding domain: molecular basis of peptide recognition by pdz. *Cell*, 85(7):1067–1076, 1996.

[18] Ho-Jin Lee and Jie J Zheng. Pdz domains and their binding partners: structure, specificity, and modification. *Cell Communication and Signaling*, 8(1):8, 2010.

[19] Ylva Ivarsson. Plasticity of pdz domains in ligand recognition and signaling. *FEBS letters*, 586(17):2638–2647, 2012.

[20] Baruch Z Harris and Wendell A Lim. Mechanism and role of pdz domains in signaling complex assembly. *Journal of cell science*, 114(18):3219–3231, 2001.

[21] Z Songyang, AS Fanning, C Fu, J Xu, SM Marfatia, AH Chishti, A Crompton, AC Chan, JM Anderson, and LC Cantley. Recognition of unique carboxyl-terminal motifs by distinct pdz domains. *Science*, 275(5296):73–77, 1997.

[22] Tony Hunter. Why nature chose phosphate to modify proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1602):2513–2516, 2012.

[23] Tony Kouzarides. Acetylation: a regulatory modification to rival phosphorylation? *The EMBO journal*, 19(6):1176–1179, 2000.

[24] Ibraheem Ali, Ryan J Conrad, Eric Verdin, and Melanie Ott. Lysine acetylation goes global: from epigenetics to metabolism and therapeutics. *Chemical reviews*, 118(3):1216–1252, 2018.

[25] Tsuneo Ikenoue, Ken Inoki, Bin Zhao, and Kun-Liang Guan. Pten acetylation modulates its interaction with pdz domain. *Cancer research*, 68(17):6908–6912, 2008.

[26] Rong Huang, Marc A Holbert, Mary Katherine Tarrant, Sandrine Curtet, David R Colquhoun, Beverley M Dancy, Blair C Dancy, Yousang Hwang, Yong Tang, Katrina Meeth, Ronen Marmorstein, Robert C Cole, Saadi Khochbin, and Philip A Cole. Site-specific introduction of an acetyl-lysine

mimic into peptides and proteins by cysteine alkylation. *Journal of the American Chemical Society*, 132(29):9986–9987, 2010.

[27] Andreas Hecht, Thierry Laroche, Sabine Strahl-Bolsinger, Susan M Gasser, and Michael Grunstein. Histone h3 and h4 n-termini interact with sir3 and sir4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell*, 80(4):583–592, 1995.

[28] Xiaodong Wang and Jeffrey J Hayes. Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. *Molecular and cellular biology*, 28(1):227–236, 2008.

[29] Nathan A Sieracki and Yulia A Komarova. Studying cell signal transduction with biomimetic point mutations. *Genetic Manipulation of DNA and Protein-Examples from Current Research*, pages 381–392, 2013.

[30] Wilma Dormeyer, Melanie Ott, and Martina Schnölzer. Probing lysine acetylation in proteins: strategies, limitations, and pitfalls of in vitro acetyltransferase assays. *Molecular & Cellular Proteomics*, 4(9):1226–1239, 2005.

[31] Yoan Duhoo, Virginie Girault, Jeremy Turchetto, Laurie Ramond, Fabien Durbesson, Patrick Fourquet, Yves Nominé, Vânia Cardoso, Ana Filipa Sequeira, Joana LA Brás, Carlos M G A Fontes, Gilles Travé, Nicolas Wolff, and Renaud Vincentelli. High-throughput production of a new library of human single and tandem pdz domains allows quantitative pdz-peptide interaction screening through high-throughput holdup assay. In *High-Throughput Protein Production and Purification*, pages 439–476. Springer, 2019.

[32] Pawel Smialowski, Philipp Pagel, Philip Wong, Barbara Brauner, Irmtraud Dunger, Gisela Fobo, Goar Frishman, Corinna Montrone, Thomas Rattei, Dmitrij Frishman, and Andreas Ruepp. The negatome database: a reference set of non-interacting protein pairs. *Nucleic acids research*, 38(suppl_1):D540–D544, 2010.

[33] Gergő Gógl, Beáta Biri-Kovács, Ádám L Póti, Henrietta Vadászi, Bálint Szeder, Andrea Bodor, Gitta Schlosser, András Ács, Lilla Turiák, László Buday, Anita Alexa, László Nytray, and Attila Reményi. Dynamic control of rsk complexes by phosphoswitch-based regulation. *The FEBS journal*, 285(1):46–71, 2018.

[34] Zakir Khan, Elouan Terrien, Florent Delhommel, Cynthia Lefebvre-Omar, Delphine Bohl, Sandrine Vitry, Clara Bernard, Juan Ramirez, Alain Chaffotte, Kevin Ricquier, Renaud Vincentelli, Henri Buc, Christophe Prehaud, Nicolas Wolff, and Monique Lafon. Structure-based optimization of a pdz-binding motif within a viral peptide stimulates neurite outgrowth. *Journal of Biological Chemistry*, 294(37):13755–13768, 2019.

[35] Stephen John Sammut, Robert D Finn, and Alex Bateman. Pfam 10 years on: 10 000 families and still growing. *Briefings in bioinformatics*, 9(3):210–219, 2008.

[36] Sara El-Gebali, Jaina Mistry, Alex Bateman, Sean R Eddy, Aurélien Luciani, Simon C Potter, Matloob Qureshi, Lorna J Richardson, Gustavo A Salazar, Alfredo Smart, Erik L L Sonhammer, Layla Hirsh, Lisanna Paladin, Damiano Piovesan, silvio C E Tosatto, and Robert D Finn. The pfam protein families database in 2019. *Nucleic acids research*, 47(D1):D427–D432, 2019.

[37] John Boyle. Molecular biology of the cell, by b. alberts, a. johnson, j. lewis, m. raff, k. roberts, and p. walter. *Biochemistry and Molecular Biology Education*, 36(4):317–318, 2008.

[38] Donald B Wetlaufer. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proceedings of the National Academy of Sciences*, 70(3):697–701, 1973.

[39] Joël Janin and Shoshanna J Wodak. Structural domains in proteins and their role in the dynamics of protein function. *Progress in biophysics and molecular biology*, 42:21–78, 1983.

[40] Peer Bork, A Kristina Downing, Bruno Kieffer, and Iain D Campbell. Structure and distribution of modules in extracellular proteins. *Quarterly reviews of biophysics*, 29(2):119–167, 1996.

[41] Chris P Ponting and Robert R Russell. The natural history of protein domains. *Annual review of biophysics and biomolecular structure*, 31(1):45–71, 2002.

[42] Christine Vogel, Matthew Bashton, Nicola D Kerrison, Cyrus Chothia, and Sarah A Teichmann. Structure, function and evolution of multidomain proteins. *Current opinion in structural biology*, 14(2):208–216, 2004.

[43] Tony Pawson and Piers Nash. Assembly of cell regulatory systems through protein interaction domains. *science*, 300(5618):445–452, 2003.

[44] Sailu Yellaboina, Asba Tasneem, Dmitri V Zaykin, Balaji Raghavachari, and Raja Jothi. Domine: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic acids research*, 39(suppl_1):D730–D735, 2011.

[45] Arshan Nasir, Kyung Mo Kim, and Gustavo Caetano-Anollés. Global patterns of protein domain gain and loss in superkingdoms. *PLoS computational biology*, 10(1):e1003452, 2014.

[46] Macarena Toll-Riera and M Mar Albà. Emergence of novel domains in proteins. *BMC evolutionary biology*, 13(1):47, 2013.

[47] Zohar Itzhaki, Eyal Akiva, Yael Altuvia, and Hanah Margalit. Evolutionary conservation of domain-domain interactions. *Genome biology*, 7(12):R125, 2006.

[48] Ivica Letunic, Tobias Doerks, and Peer Bork. Smart: recent updates, new developments and status in 2015. *Nucleic acids research*, 43(D1):D257–D260, 2015.

[49] Natalie L Dawson, Tony E Lewis, Sayoni Das, Jonathan G Lees, David Lee, Paul Ashford, Christine A Orengo, and Ian Sillitoe. Cath: an expanded resource to predict protein function through structure and sequence. *Nucleic acids research*, 45(D1):D289–D295, 2017.

[50] Shennan Lu, Jiyao Wang, Farideh Chitsaz, Myra K Derbyshire, Renata C Geer, Noreen R Gonzales, Marc Gwadz, David I Hurwitz, Gabriele H Marchler, James S Song, Narmada Thanki, Roxanne A Yamashita, Mingzhang Yang, Dachuan Zhang, Chanjuan Zheng, Christopher J Lanczycki, and Aron Marchler-Bauer. Cdd/sparcle: the conserved domain database in 2020. *Nucleic acids research*, 48(D1):D265–D268, 2020.

[51] Antonina Andreeva, Dave Howorth, Cyrus Chothia, Eugene Kulesha, and Alexey G Murzin. Scop2 prototype: a new approach to protein structure mining. *Nucleic acids research*, 42(D1):D310–D314, 2014.

[52] Antonina Andreeva, Eugene Kulesha, Julian Gough, and Alexey G Murzin. The scop database in 2020: expanded classification of representative family

and superfamily domains of known protein structures. *Nucleic acids research*, 48(D1):D376–D382, 2020.

[53] Protein Data Bank. Protein data bank. *Nature New Biol*, 233:223, 1971.

[54] Balaji Raghavachari, Asba Tasneem, Teresa M Przytycka, and Raja Jothi. Domine: a database of protein domain interactions. *Nucleic acids research*, 36(suppl_1):D656–D661, 2008.

[55] Monika Fuxreiter, Peter Tompa, and István Simon. Local structural disorder imparts plasticity on linear motifs. *Bioinformatics*, 23(8):950–956, 2007.

[56] Katia Zanier, Sebastian Charbonnier, Abdellahi Ould M'hamed Ould Sidi, Alastair G McEwen, Maria Giovanna Ferrario, Pierre Poussin-Courmontagne, Vincent Cura, Nicole Brimer, Khaled Ould Babah, Tina Ansari, Isabelle Muller, Roland H Stote, Jean Cavarelli, Scott Vande Pol, and Gilles Travé. Structural basis for hijacking of cellular lxxll motifs by papillomavirus e6 oncoproteins. *Science*, 339(6120):694–698, 2013.

[57] Izabella Krystkowiak and Norman E Davey. Slimsearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. *Nucleic acids research*, 45(W1):W464–W469, 2017.

[58] Manjeet Kumar, Marc Gouw, Sushama Michael, Hugo Sámano-Sánchez, Rita Pancsa, Juliana Glavina, Athina Diakogianni, Jesús Alvarado Valverde, Dayana Bukirova, Jelena Čalyševa, Nicolas Palopoli, Norman E Davey, Lucía Chemes, and Toby J Gibson. Elm—the eukaryotic linear motif resource in 2020. *Nucleic acids research*, 48(D1):D296–D306, 2020.

[59] Daniel J Mandell, Ilya Chorny, Eli S Groban, Sergio E Wong, Elisheva Levine, Chaya S Rapp, and Matthew P Jacobson. Strengths of hydrogen bonds involving phosphorylated amino acid side chains. *Journal of the American Chemical Society*, 129(4):820–827, 2007.

[60] Ana Paula Oliveira and Uwe Sauer. The importance of post-translational modifications in regulating saccharomyces cerevisiae metabolism. *FEMS yeast research*, 12(2):104–117, 2012.

[61] Panayotis Vlastaridis, Pelagia Kyriakidou, Anargyros Chaliotis, Yves Van de Peer, Stephen G Oliver, and Grigoris D Amoutzias. Estimating the total num-

ber of phosphoproteins and phosphorylation sites in eukaryotic proteomes. *Gigascience*, 6(2):giw015, 2017.

[62] Sofia Caria, Bryce Z Stewart, Ruitao Jin, Brian J Smith, Patrick O Humbert, and Marc Kvansakul. Structural analysis of phosphorylation-associated interactions of human mcc with scribble pdz domains. *The FEBS Journal*, 286(24):4910–4925, 2019.

[63] Martina Baliova and Frantisek Jursky. Phosphomimetic mutation of glycine transporter glyt1 c-terminal pdz binding motif inhibits its interactions with psd95. *Journal of Molecular Neuroscience*, 70(4):488–493, 2020.

[64] Samuel M Pearlman, Zach Serber, and James E Ferrell Jr. A mechanism for the evolution of phosphorylation sites. *Cell*, 147(4):934–946, 2011.

[65] Chunaram Choudhary, Chanchal Kumar, Florian Gnad, Michael L Nielsen, Michael Rehman, Tobias C Walther, Jesper V Olsen, and Matthias Mann. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, 325(5942):834–840, 2009.

[66] Rasmus Ree, Sylvia Varland, and Thomas Arnesen. Spotlight on protein n-terminal acetylation. *Experimental & molecular medicine*, 50(7):1–13, 2018.

[67] G Singh and A M Chan. Post-translational modifications of pten and their potential therapeutic implications. *Current cancer drug targets*, 11(5):536–547, 2011.

[68] Bing-Rui Zhou, Hanqiao Feng, Rodolfo Ghirlando, Hidenori Kato, James Gruschus, and Yawen Bai. Histone h4 k16q mutation, an acetylation mimic, causes structural disorder of its n-terminal basic patch in the nucleosome. *Journal of molecular biology*, 421(1):30–37, 2012.

[69] Claire Nourry, Seth GN Grant, and Jean-Paul Borg. Pdz domain proteins: plug and play! *Science Signaling*, 2003(179):re7–re7, 2003.

[70] João H Morais Cabral, Carlo Petosa, Michael J Sutcliffe, Sami Raza, Olwyn Byron, Florence Poy, Shirin M Marfatia, Athar H Chishti, and Robert C Liddington. Crystal structure of a pdz domain. *Nature*, 382(6592):649–652, 1996.

[71] Yingnan Zhang, Brent A Appleton, Ping Wu, Christian Wiesmann, and Sachdev S Sidhu. Structural and functional analysis of the ligand specificity of the htra2/omi pdz domain. *Protein science*, 16(8):1738–1750, 2007.

[72] Steven T Truschel, Debrup Sengupta, Adam Foote, Annie Heroux, Mark R Macbeth, and Adam D Linstedt. Structure of the membrane-tethering grasp domain reveals a unique pdz ligand interaction that mediates golgi biogenesis. *Journal of Biological Chemistry*, 286(23):20125–20129, 2011.

[73] Rodrigo Gallardo, Ylva Ivarsson, Joost Schymkowitz, Frédéric Rousseau, and Pascale Zimmermann. Structural diversity of pdz–lipid interactions. *Chembiochem*, 11(4):456–467, 2010.

[74] Yong Chen, Ren Sheng, Morten Källberg, Antonina Silkov, Moe P Tun, Nitin Bhardwaj, Svetlana Kurilova, Randy A Hall, Barry Honig, Hui Lu, and Wonhwa Cho. Genome-wide functional annotation of dual-specificity protein-and lipid-binding modules that regulate protein interactions. *Molecular cell*, 46(2):226–237, 2012.

[75] Ylva Ivarsson, Anna Maria Wawrzyniak, Rudra Kashyap, Jolanta Polanowska, Stéphane Betzi, Frédérique Lembo, Elke Vermeiren, Driss Chiheb, Nicolas Lenfant, Xavier Morelli, Jean-Paul Borg, Jérôme Reboul, and Pascale Zimmermann. Prevalence, specificity and determinants of lipid-interacting pdz domains from an in-cell screen and in vitro binding experiments. *PloS one*, 8(2), 2013.

[76] Yi Mu, Pengfei Cai, Siqi Hu, Sucan Ma, and Youhe Gao. Characterization of diverse internal binding specificities of pdz domains by yeast two-hybrid screening of a special peptide library. *PloS one*, 9(2), 2014.

[77] Boyoung Cha, Jianbo Yang, Varsha Singh, Nicholas C Zachos, Rafiquel I Sarker, Tian-e Chen, Molee Chakraborty, Chung-Ming Tse, and Mark Donowitz. Pdz domain-dependent regulation of nhe3 protein by both internal class ii and c-terminal class i pdz-binding motifs. *Journal of Biological Chemistry*, 292(20):8279–8290, 2017.

[78] Thomas Clairfeuille, Caroline Mas, Audrey SM Chan, Zhe Yang, Maria Tello-Lafoz, Mintu Chandra, Jocelyn Widagdo, Markus C Kerr, Blessy Paul, Isabel Mérida, Rohan D Teasdale, Nathan J Pavlos, Victor Anggono, and Brett M Collins. A molecular code for endosomal recycling of phosphorylated cargos

by the snx27–retromer complex. *Nature structural & molecular biology*, 23(10):921, 2016.

[79] Michael B Yaffe and Stephen J Smerdon. The use of in vitro peptide-library screens in the analysis of phosphoserine/threonine-binding domain structure and function. *Annu. Rev. Biophys. Biomol. Struct.*, 33:225–244, 2004.

[80] Arndt Grossmann, Nouhad Benlasfer, Petra Birth, Anna Hegele, Franziska Wachsmuth, Luise Apelt, and Ulrich Stelzl. Phospho-tyrosine dependent protein–protein interaction network. *Molecular systems biology*, 11(3):794, 2015.

[81] Kira G Hartman, Michele I Vitolo, Adam D Pierce, Jennifer M Fox, Paul Shapiro, Stuart S Martin, Paul T Wilder, and David J Weber. Complex formation between s100b protein and the p90 ribosomal s6 kinase (rsk) in malignant melanoma is calcium-dependent and inhibits extracellular signal-regulated kinase (erk)-mediated phosphorylation of rsk. *Journal of Biological Chemistry*, 289(18):12886–12895, 2014.

[82] Anita Alexa, Gergő Gógl, Gábor Glatz, Ágnes Garai, András Zeke, János Varga, Erika Dudás, Norbert Jeszenői, Andrea Bodor, Csaba Hetényi, and Attila Reményi. Structural assembly of the signaling competent erk2–rsk1 heterodimeric protein kinase complex. *Proceedings of the National Academy of Sciences*, 112(9):2711–2716, 2015.

[83] Elouan Terrien, Alain Chaffotte, Mireille Lafage, Zakir Khan, Christophe Prehaud, Florence Cordier, Catherine Simenel, Muriel Delepierre, Henri Buc, Monique Lafon, and Nicolas Wolff. Interference with the pten-mast2 interaction by a viral protein leads to cellular relocalization of pten. *Sci. Signal.*, 5(237):ra58–ra58, 2012.

[84] A Ventruti, TM Kazdoba, S Niu, and Gabriella D'Arcangelo. Reelin deficiency causes specific defects in the molecular composition of the synapses in the adult brain. *Neuroscience*, 189:32–42, 2011.

[85] Lionel ML Chow and Suzanne J Baker. Pten function in normal and neoplastic growth. *Cancer letters*, 241(2):184–196, 2006.

[86] Brian J Hillier, Karen S Christopherson, Kenneth E Prehoda, David S Bredt, and Wendell A Lim. Unexpected modes of pdz domain scaffolding revealed by structure of nnos-syntrophin complex. *Science*, 284(5415):812–815, 1999.

[87] JoAnn Trejo. Internal pdz ligands: novel endocytic recycling motifs for g protein-coupled receptors. *Molecular pharmacology*, 67(5):1388–1390, 2005.

[88] Monimoy Banerjee, David L Zoetewey, Mohiuddin Ovee, Suman Mazumder, Valery A Petrenko, Tatiana I Samoylova, and Smita Mohanty. Specificity and promiscuity in human glutaminase interacting protein recognition: insight from the binding of the internal and c-terminal motif. *Biochemistry*, 51(35):6950–6960, 2012.

[89] Sandra Iden and John G Collard. Crosstalk between small gtpases and polarity proteins in cell polarization. *Nature reviews Molecular cell biology*, 9(11):846–859, 2008.

[90] Ronald T Javier and Andrew P Rice. Emerging theme: cellular pdz proteins as common targets of pathogenic viruses. *Journal of virology*, 85(22):11544–11556, 2011.

[91] Daniel St Johnston and Julie Ahringer. Cell polarity in eggs and epithelia: parallels and diversity. *Cell*, 141(5):757–774, 2010.

[92] Patrick Humbert, Sarah Russell, and Helena Richardson. Dlg, scribble and lgl in cell polarity, cell proliferation and cancer. *Bioessays*, 25(6):542–553, 2003.

[93] Miranda Thomas, Nisha Narayan, David Pim, Vjekoslav Tomaić, Paola Massimi, Kazunori Nagasaka, Christian Kranjec, Noor Gammoh, and Lawrence Banks. Human papillomaviruses, cervical cancer and cell polarity. *Oncogene*, 27(55):7018–7030, 2008.

[94] Lorenza González-Mariscal, Abigail Betanzos, and Antonia Ávila-Flores. Maguk proteins: structure and role in the tight junction. In *Seminars in Cell and Developmental Biology*, volume 11, page 315. London: Academic Press, c1996-, 2000.

[95] Wei Feng and Mingjie Zhang. Organization and dynamics of pdz-domain-related supramodules in the postsynaptic density. *Nature Reviews Neuroscience*, 10(2):87–99, 2009.

[96] Xu Liu and Ernesto J Fuentes. Emerging themes in pdz domain signaling: structure, function, and inhibition. In *International review of cell and molecular biology*, volume 343, pages 129–218. Elsevier, 2019.

[97] Philomena Mburu, Mirna Mustapha, Anabel Varela, Dominique Weil, Aziz El-Amraoui, Ralph H Holme, Andreas Rump, Rachel E Hardisty, Stéphane Blanchard, Roney S Coimbra, Isabelle Perfettini, Nick Parkinson, Ann-Marie Mallon, Pete Glenister, Mike J Rogers, Adam J Paige, Lee Moir, Jo Clay, Andre Rosenthal, Xue Zhong Liu, Gonzalo Blanco, Karen P Steel, Christine Petit, and Steve D M Brown. Defects in whirlin, a pdz domain molecule involved in stereocilia elongation, cause deafness in the whirler mouse and families with dfnb31. *Nature genetics*, 34(4):421–428, 2003.

[98] Amel Bahloul, Elise Pepermans, Bertrand Raynal, Nicolas Wolff, Florence Cordier, Patrick England, Sylvie Nouaille, Bruno Baron, Aziz El-Amraoui, Jean-Pierre Hardelin, Dominique Durand, and Christine Petit. Conformational switch of harmonin, a submembrane scaffold protein of the hair cell mechanoelectrical transduction machinery. *FEBS letters*, 591(15):2299–2310, 2017.

[99] Florent Delhommel, Florence Cordier, Frederick Saul, Lucas Chataigner, Ahmed Haouz, and Nicolas Wolff. Structural plasticity of the hhd 2 domain of whirlin. *The FEBS Journal*, 285(20):3738–3752, 2018.

[100] Miranda Thomas and Lawrence Banks. Upsetting the balance: when viruses manipulate cell polarity control. *Journal of molecular biology*, 430(19):3481–3503, 2018.

[101] Claire D James and Sally Roberts. Viral interactions with pdz domain-containing proteins—an oncogenic trait? *Pathogens*, 5(1):8, 2016.

[102] Angelo Toto, Sana Ma, Francesca Malagrinò, Lorenzo Visconti, Livia Pagano, Kristian Stromgaard, and Stefano Gianni. Comparing the binding properties of peptides mimicking the envelope protein of sars-cov and sars-cov-2 to the pdz domain of the tight junction-associated pals1 protein. *Protein Science*, 2020.

[103] Célia Caillet-Saguy, Pierre Maisonneuve, Florent Delhommel, Elouan Terrien, Nicolas Babault, Monique Lafon, Florence Cordier, and Nicolas Wolff. Strategies to interfere with pdz-mediated interactions in neurons: what we can learn from the rabies virus. *Progress in biophysics and molecular biology*, 119(1):53–59, 2015.

[104] Harald zur Hausen. Human papillomaviruses in the pathogenesis of anogenital cancer. *Virology*, 184(1):9–13, 1991.

[105] Sebastian Charbonnier, Yves Nominé, Juan Ramírez, Katja Luck, Anne Chapelle, Roland H Stote, Gilles Travé, Bruno Kieffer, and R Andrew Atkinson. The structural and dynamic response of magi-1 pdz1 with noncanonical domain boundaries to the binding of human papillomavirus e6. *Journal of molecular biology*, 406(5):745–763, 2011.

[106] Harald Zur Hausen. Papillomaviruses and cancer: from basic studies to clinical application. *Nature reviews cancer*, 2(5):342–350, 2002.

[107] D Maxwell Parkin and Freddie Bray. The burden of hpv-related cancers. *Vaccine*, 24:S11–S25, 2006.

[108] Elisabeth Schwarz, Ulrich Karl Freese, Lutz Gissmann, Wolfgang Mayer, Birgit Roggenbuck, Armin Stremlau, and Harald Zur Hausen. Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature*, 314(6006):111–114, 1985.

[109] EJ Androphy, Nancy L Hubbert, John T Schiller, and Douglas R Lowy. Identification of the hpv-16 e6 protein from transformed mouse cells and human cervical carcinoma cell lines. *The EMBO journal*, 6(4):989–992, 1987.

[110] David Smotkin and Felix O Wettstein. Transcription of human papillomavirus type 16 early genes in a cervical cancer and a cancer-derived cell line and identification of the e7 protein. *Proceedings of the National Academy of Sciences*, 83(13):4680–4684, 1986.

[111] Sandy S Tungteakkhun and Penelope J Duerksen-Hughes. Cellular binding partners of the human papillomavirus e6 protein. *Archives of virology*, 153(3):397, 2008.

[112] Yves Nominé, Murielle Masson, Sebastian Charbonnier, Katia Zanier, Tutik Ristriani, François Deryckère, Annie-Paule Sibler, Dominique Desplancq, Robert Andrew Atkinson, Etienne Weiss, Georges Orfanoudakis, Bruno Kieffer, and Gilles Travé. Structural and functional analysis of e6 oncoprotein: insights in the molecular pathways of human papillomavirus-mediated pathogenesis. *Molecular cell*, 21(5):665–678, 2006.

[113] Daan Geraets, Laia Alemany, Nuria Guimera, Silvia de Sanjose, Maurits de Koning, Anco Molijn, David Jenkins, Xavier Bosch, Wim Quint, and RIS HPV TT Study Group. Detection of rare and possibly carcinogenic human papillomavirus genotypes as single infections in invasive cervical cancer. *The Journal of pathology*, 228(4):534–543, 2012.

[114] Tina Dalianis. Human papillomavirus (hpv) and oropharyngeal squamous cell carcinoma. *La Presse Médicale*, 43(12):e429–e434, 2014.

[115] Silvia de Sanjosé, Laia Bruni, and Laia Alemany. Hpv in genital cancers (at the exception of cervical cancer) and anal cancers. *La Presse Médicale*, 43(12):e423–e428, 2014.

[116] F Xavier Bosch, Ann N Burchell, Mark Schiffman, Anna R Giuliano, Silvia de Sanjose, Laia Bruni, Guillermo Tortolero-Luna, Susanne Kruger Kjaer, and Nubia Muñoz. Epidemiology and natural history of human papillomavirus infections and type-specific implications in cervical neoplasia. *Vaccine*, 26:K1–K16, 2008.

[117] Kevin A Ault. Epidemiology and natural history of human papillomavirus infections in the female genital tract. *Infectious diseases in obstetrics and gynecology*, 2006, 2006.

[118] Jeanine F Amacher, Lionel Brooks 3rd, Thomas H Hampton, and Dean R Madden. Specificity in pdz-peptide interaction networks: Computational analysis and review. *Journal of Structural Biology: X*, page 100022, 2020.

[119] Craig P Delury, Elizabeth K Marsh, Claire D James, Siaw Shi Boon, Lawrence Banks, Gillian L Knight, and Sally Roberts. The role of protein kinase a regulation of the e6 pdz-binding domain during the differentiation-dependent life cycle of human papillomavirus type 18. *Journal of virology*, 87(17):9463–9472, 2013.

[120] Christian KuÈhne, Daniela Gardiol, Corrado Guarnaccia, Heinz Amenitsch, and Lawrence Banks. Differential regulation of human papillomavirus e6 by protein kinase a: conditional degradation of human discs large protein by oncogenic e6. *Oncogene*, 19(51):5884–5891, 2000.

[121] Siaw Shi Boon and Lawrence Banks. High-risk human papillomavirus e6 oncoproteins interact with 14-3-3ζ in a pdz binding motif-dependent manner. *Journal of virology*, 87(3):1586–1595, 2013.

[122] Alexsandra B Espejo, Guozhen Gao, Karynne Black, Sitaram Gayatri, Nicolas Veland, Jeesun Kim, Taiping Chen, Marius Sudol, Cheryl Walker, and Mark T Bedford. Prmt5 c-terminal phosphorylation modulates a 14-3-3/pdz interaction switch. *Journal of Biological Chemistry*, 292(6):2255–2265, 2017.

[123] Francis D Carlson. Physiological and biochemical aspects of nervous integration. In *Symposium on Physiological and Biochemical Aspects of Nervous Integration (Aug. 30-Sept. 2, 1967: Marine Biological Laboratory, Woods Hole, Mass.)*. NJ, Prentice-Hall, 1968.

[124] Alastair Aitken. 14-3-3 proteins: a historic overview. In *Seminars in cancer biology*, volume 16, pages 162–172. Elsevier, 2006.

[125] Bing Xiao, Stephen J Smerdon, David H Jones, Guy G Dodson, Yasmina Soneji, Alastair Aitken, and Steven J Gamblin. Structure of a 14-3-3 protein and implications for coordination of multiple signalling pathways. *Nature*, 376(6536):188–191, 1995.

[126] Haian Fu, Romesh R Subramanian, and Shane C Masters. 14-3-3 proteins: structure, function, and regulation. *Annual review of pharmacology and toxicology*, 40(1):617–647, 2000.

[127] Heiko Hermeking. The 14-3-3 cancer connection. *Nature Reviews Cancer*, 3(12):931–943, 2003.

[128] Surajit Ganguly, Joan L Weller, Anthony Ho, Philippe Chemineau, Benoit Malpaux, and David C Klein. Melatonin synthesis: 14-3-3-dependent activation and inhibition of arylalkylamine n-acetyltransferase mediated by phosphoserine-205. *Proceedings of the National Academy of Sciences*, 102(4):1222–1227, 2005.

[129] Patrick Aloy and Robert B Russell. Potential artefacts in protein-interaction networks. *FEBS letters*, 530(1-3):253–254, 2002.

[130] Joel S Bader, Amitabha Chaudhuri, Jonathan M Rothberg, and John Chant. Gaining confidence in high-throughput protein interaction networks. *Nature biotechnology*, 22(1):78–85, 2004.

[131] Michael A Stiffler, Viara P Grantcharova, Mark Sevecka, and Gavin MacBeath. Uncovering quantitative protein interaction networks for mouse

pdz domains using protein microarrays. *Journal of the American Chemical Society*, 128(17):5913–5922, 2006.

[132] Gianni Cesareni, Arnaud Ceol, Caius Gavrila, Luisa Montecchi Palazzi, Maria Persico, and Maria Victoria Schneider. Comparative interactomics. *FEBS letters*, 579(8):1828–1833, 2005.

[133] István A Kovács, Katja Luck, Kerstin Spirohn, Yang Wang, Carl Pollis, Sadie Schlabach, Wenting Bian, Dae-Kyum Kim, Nishka Kishore, Tong Hao, Michael A Calderwood, Marc Vidal, and Albert-Lázló Barabási. Network-based prediction of protein interactions. *Nature communications*, 10(1):1–8, 2019.

[134] Ylva Ivarsson and Per Jemth. Affinity and specificity of motif-based protein–protein interactions. *Current opinion in structural biology*, 54:26–33, 2019.

[135] Daniel Ting, Guoli Wang, Maxim Shapovalov, Rajib Mitra, Michael I Jordan, and Roland L Dunbrack Jr. Neighbor-dependent ramachandran probability distributions of amino acids developed from a hierarchical dirichlet process model. *PLoS computational biology*, 6(4):e1000763, 2010.

[136] Oliviero Carugo and K Djinović-Carugo. Half a century of ramachandran plots. *Acta Crystallographica Section D: Biological Crystallography*, 69(8):1333–1341, 2013.

[137] Michael A Stiffler, Jiunn R Chen, Viara P Grantcharova, Ying Lei, Daniel Fuchs, John E Allen, Lioudmila A Zaslavskaia, and Gavin MacBeath. Pdz domain binding selectivity is optimized across the mouse proteome. *Science*, 317(5836):364–369, 2007.

[138] Richard B Jones, Andrew Gordus, Jordan A Krall, and Gavin MacBeath. A quantitative protein interaction network for the erbb receptors using protein microarrays. *Nature*, 439(7073):168–174, 2006.

[139] Jorrit J Hornberg, Bernd Binder, Frank J Bruggeman, Birgit Schoeberl, Reinhart Heinrich, and Hans V Westerhoff. Control of mapk signalling: from complexity to what really matters. *Oncogene*, 24(36):5533–5542, 2005.

[140] Katja Luck, Dae-Kyum Kim, Luke Lambourne, Kerstin Spirohn, Bridget E Begg, Wenting Bian, Ruth Brignall, Tiziana Cafarelli, Francisco J Campos-Laborie, Benoit Charloteaux, Dongsic Choi, Atina G Coteé, Meaghan Da-

ley, Steven Deimling, Alice Desbuleux, Amélie Dricot, Hardy Madeleine F Gebbia, Marinella, Nishka Kishore, Jennifer J Knapp, István A Kovács, Irma Lemmens, Miles W Mee, Joseph C Mellor, Carl Pollis, Carles Pons, Aaron D Richardson, Sadie Schlabach, Bridget Teeking, Anupama Yadav, Mariana Babor, Dawit Balcha, Omer Basha, Christian Bowman-Colin, Suet-Feung Chin, Soon Gang Choi, Claudia Colabella, Georges Coppin, Cassandra D'Amata, David De Ridder, Steffi De Rouck, Miquel Duran-Frigola, Hanane Ennajdaoui, Florian Goebels, Liana Goehring, Anjali Gopal, Ghazal Haddad, Elodie Hatchi, Mohamed Helmy, Yves Jacob, Yoseph Kassa, Serena Landini, Roujia Li, Natascha van Lieshout, Andrew MacWilliams, Dylan Markey, Joseph N Paulson, Sudharshan Rangarajan, John Rasla, Ashyad Rayhan, Thomas Rolland, Adriana San-Miguel, Yun Shen, Dayag Sheykhkarimli, Gloria M Sheynkman, Eyal Simonovsky, Murat Taşan, Alexander Tejeda, Vincent Tropepe, Jean-Claude Twizere, Yang Wang, Robert J Weatheritt, Jochen Weile, Yu Xia, Xinping Yang, Esti Yeger-Lotem, Quan Zhong, Patrick Aloy, Gary D Bader, Javier De Las Rivas, Suzanne Gaudet, Tong Hao, Janusz Rak, Jan Tavernier, David E Hill, Marc Vidal, Frederick P Roth, and Michael A Calderwood. A reference map of the human binary protein interactome. *Nature*, 580(7803):402–408, 2020.

[141] Mark R Spaller. Act globally, think locally: systems biology addresses the pdz domain, 2006.

[142] Bernhard Suter, Saranya Kittanakom, and Igor Stagljar. Two-hybrid technologies in proteomics research. *Current opinion in biotechnology*, 19(4):316–323, 2008.

[143] Tord Berggård, Sara Linse, and Peter James. Methods for the detection and analysis of protein–protein interactions. *Proteomics*, 7(16):2833–2842, 2007.

[144] Soon Gang Choi, Julien Olivet, Patricia Cassonnet, Pierre-Olivier Vidalain, Katja Luck, Luke Lambourne, Kerstin Spirohn, Irma Lemmens, Mélanie Dos Santos, Caroline Demeret, Louis Jones, Sudharshan Rangarajan, Wenting Bian, Eloi P Coutant, Yves L Janin, Sylvie van der Werf, Philipp Trepte, Erich E Wanker, Javier De Las Rivas, Jan Tavernier, Jean-Claude Twizere, Tong Hao, David E Hill, Marc Vidal, Michael A Calderwood, and Yves Jacob. Maximizing binary interactome mapping with a minimal number of assays. *Nature communications*, 10(1):1–13, 2019.

[145] George P Smith. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228(4705):1315–1317, 1985.

[146] Germaine Fuh, M Teresa Pisabarro, Ying Li, Clifford Quan, Laurence A Lasky, and Sachdev S Sidhu. Analysis of pdz domain-ligand interactions using carboxyl-terminal phage display. *Journal of Biological Chemistry*, 275(28):21486–21491, 2000.

[147] Andreas Ernst, Stephen L Sazinsky, Shirley Hui, Bridget Currell, Moyez Dharsee, Somasekar Seshagiri, Gary D Bader, and Sachdev S Sidhu. Rapid evolution of functional complexity in a domain family. *Science Signaling*, 2(87):ra50–ra50, 2009.

[148] Andreas Ernst, David Gfeller, Zhengyan Kan, Somasekar Seshagiri, Philip M Kim, Gary D Bader, and Sachdev S Sidhu. Coevolution of pdz domain–ligand interactions analyzed by high-throughput phage display and deep sequencing. *Molecular BioSystems*, 6(10):1782–1790, 2010.

[149] Gustav N Sundell and Ylva Ivarsson. Interaction analysis through proteomic phage display. *BioMed research international*, 2014, 2014.

[150] Cecilia Blikstad and Ylva Ivarsson. High-throughput methods for identification of protein-protein interactions involving short linear motifs. *Cell Communication and Signaling*, 13(1):38, 2015.

[151] Bálint Mészáros, Gábor Erdős, and Zsuzsanna Dosztányi. Iupred2a: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic acids research*, 46(W1):W329–W337, 2018.

[152] Gábor Erdős and Zsuzsanna Dosztányi. Analyzing protein disorder with iupred2a. *Current Protocols in Bioinformatics*, 70(1):e99, 2020.

[153] Ylva Ivarsson, Roland Arnold, Megan McLaughlin, Satra Nim, Rakesh Joshi, Debashish Ray, Bernard Liu, Joan Teyra, Tony Pawson, Jason Moffat, Shawn Shun-Cheng, Sachdev S Sidhu, and Philip M Kim. Large-scale interaction profiling of pdz domains through proteomic peptide-phage display using human and viral phage peptidomes. *Proceedings of the National Academy of Sciences*, 111(7):2542–2547, 2014.

[154] Norman E Davey, Moon-Hyeong Seo, Vikash Kumar Yadav, Jouhyun Jeon, Satra Nim, Izabella Krystkowiak, Cecilia Blikstad, Debbie Dong, Natalia Markova, Philip M Kim, and Ylva Ivarsson. Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. *The FEBS journal*, 284(3):485–498, 2017.

[155] Stephen P Fodor, J Leighton Read, Michael C Pirrung, Lubert Stryer, A Tsai Lu, and Dennis Solas. Light-directed, spatially addressable parallel chemical synthesis. *science*, 251(4995):767–773, 1991.

[156] Ronald Frank. Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron*, 48(42):9217–9232, 1992.

[157] Rudolf Volkmer, Victor Tapia, and Christiane Landgraf. Synthetic peptide arrays for investigating protein interaction domains. *FEBS letters*, 586(17):2780–2786, 2012.

[158] Prisca Boisguerin, Rainer Leben, Bernhard Ay, Gerald Radziwill, Karin Moelling, Liying Dong, and Rudolf Volkmer-Engert. An improved method for the synthesis of cellulose membrane-bound peptides with free c termini is useful for pdz domain binding studies. *Chemistry & biology*, 11(4):449–459, 2004.

[159] Armin A Weiser, Michal Or-Guil, Victor Tapia, Astrid Leichsenring, Johannes Schuchhardt, Cornelius Frömmel, and Rudolf Volkmer-Engert. Spot synthesis: reliability of array-based measurement of peptide binding affinity. *Analytical biochemistry*, 342(2):300–311, 2005.

[160] Sebastian Charbonnier, Katia Zanier, Murielle Masson, and Gilles Travé. Capturing protein–protein complexes at equilibrium: the holdup comparative chromatographic retention assay. *Protein expression and purification*, 50(1):89–101, 2006.

[161] Sam Lievens, Sven Eyckerman, Irma Lemmens, and Jan Tavernier. Large-scale protein interactome mapping: strategies and opportunities. *Expert review of proteomics*, 7(5):679–690, 2010.

[162] Hon Nian Chua and Limsoon Wong. Increasing the reliability of protein interactomes. *Drug discovery today*, 13(15-16):652–658, 2008.

[163] Anupama Yadav, Marc Vidal, and Katja Luck. Precision medicine—networks to the rescue. *Current Opinion in Biotechnology*, 63:177–189, 2020.

[164] Thomas Rolland, Murat Taşan, Benoit Charloteaux, Samuel J Pevzner, Quan Zhong, Nidhi Sahni, Song Yi, Irma Lemmens, Celia Fontanillo, Roberto Mosca, Atanas Kamburov, Susan D Ghiassian, Xinping Yang, Lila Ghamsari, Dawit Balcha, Bridget E Begg, Pascal Braun, Marc Brehme, Martin P Brolu, Anne-Ruxandra Carvunis, Dan Convery-Zupan, Roser Corominas, Coulumbe-Huntinton Jasmin, Elizabeth Dann, Matija Dreze, Amélie Dricot, Changyu Fan, Eric Franzosa, Fana Gebreab, Bryan J Guiterrez, Madeleine F Hardy, Mike Jin, Shuli Kang, Ruth Kiros, Guan Ning Li, Katja Luck, Andrew MacWilliams, Jörg Menche, Ryan R Murray, Alexandre Palagi, Matthew M Poulin, Xavier Rambout, John Rasla, Patrick Reichert, Viviano Romero, Elien Ryssinck, Julie M Sahalie, Annemarie Scholz, Akash A Shah, Amitabh Sharma, Yun Shen, Kerstin Spirohn, Stanley Tam, Alexander O Tejeda, Shelly A Trigg, Jean-Claude Twizere, Kerwin Vega, Jennifer Walsh, Michael E Cusick, yu Xia, Albert-Lázló Barabási, Lila M Iakoucheva, Patrick Aloy, Javier De Las Rivas, Michael A Tavernier Jan, Calderwood, David E Hill, Tong Hao, Frederick P Roth, and Vidal Marc. A proteome-scale map of the human interactome network.

[165] Colin A Smith and Tanja Kortemme. Structure-based prediction of the peptide sequence space recognized by natural and synthetic pdz domains. *Journal of molecular biology*, 402(2):460–474, 2010.

[166] William R Atchley, Jieping Zhao, Andrew D Fernandes, and Tanja Drüke. Solving the protein sequence metric problem. *Proceedings of the National Academy of Sciences*, 102(18):6395–6400, 2005.

[167] Karel Zimmermann and Jean-François Gibrat. Amino acid" little big bang": Representing amino acid substitution matrices as dot products of euclidian vectors. *BMC bioinformatics*, 11(1):4, 2010.

[168] Katja Luck. *Towards a better understanding of protein interaction specificities in cell signalling - PDZ domains in the spotlight of computational and experimental approaches*. Theses, Université de Strasbourg, October 2012.

[169] Barbara Brannetti, Andreas Zanzoni, Luisa Montecchi-Palazzi, Gianni Cesareni, and Manuela Helmer-Citterich. ispot: a web tool for the analysis

and recognition of protein domain specificity: A presentation for the esf workshop 'proteomics: Focus on protein interactions'. *Comparative and functional genomics*, 2(5):314–318, 2001.

[170] Urs Wiedemann, Prisca Boisguerin, Rainer Leben, Dietmar Leitner, Gerd Krause, Karin Moelling, Rudolf Volkmer-Engert, and Hartmut Oschkinat. Quantification of pdz domain specificity, prediction of ligand affinity and rational design of super-binding peptides. *Journal of molecular biology*, 343(3):703–718, 2004.

[171] Shirley Hui and Gary D Bader. Proteome scanning to predict pdz domain interactions using support vector machines. *BMC bioinformatics*, 11(1):507, 2010.

[172] Shirley Hui, Xiang Xing, and Gary D Bader. Predicting pdz domain mediated protein interactions from structure. *BMC bioinformatics*, 14(1):27, 2013.

Part VI

Appendix

Chapter 15

Electropherogram curation software

15.1 Installation process

- Install mercurial:
  https://www.mercurial-scm.org
- Cloning the "binding_extract" project via terminal:
  hg clone https://holdup_user@bitbucket.org/lc/bindings_extract
- Introduce password if needed: billythekid
- Install Anaconda or upgrade Python to latest version. We recommend Anaconda since this way you will be sure that almost all the needed python packages are installed
- No matter which python are you using, run the following commands in the terminal to install flask-socketio and eventlet packages:
  - *pip install flask-socketio*
  - *pip install eventlet*

15.2 Running the software

- Go through the terminal to the "binding_extract" folder and type:
  python -m Interf.run

- Once the graphical interface is open:
  - Click the Begin button
  - Click the Processing button
  - Select the desired parameters
  - Click the Validation button and wait for the data processing

15.3 Troubleshooting

- If you decided to install Anaconda and the software does not run:
  - type on the terminal in the user folder: *nano .bash_profile*
  - add the following line to the file:
    export PATH="/Applications/anaconda/bin:$PATH"
  - The path of the previous line may change depending on where your Anaconda set up was installed
- If you did not install Anaconda and the software does not run, then pay attention to the error messages in the terminal. They will show you which python packages are you missing. To install them use the *pip install* command
- If the terminal looks running the problem but no graphical interface is displayed, then "ctrl+click" on the printed server link in the terminal.

Chapter 16

Scripts

16.1 Converting BI into Affinity scale

```
#!/usr/bin/env python

# Libraries
import pandas as pd
import sys
from math import log10
import numpy as np


def conversion(BI,ConCPept,threshold,PDZ_tot):
# BI: Binding Intensity of the PBM/PDZ interaction
# ConCPept: Determined concentration of the PBM for the holdup
                                assay
# threshold: High confident value to determine a detected
                                binder
# PDZ_tot: PDZ concentration


threshold = threshold
PDZ_tot = PDZ_tot

BI = BI.where(BI > threshold, threshold)
Kd = (PDZ_tot - BI*PDZ_tot) * (ConCPept - BI*PDZ_tot) / (BI*
                                PDZ_tot)

LogKd = -np.log10(Kd/1000000)
```

```
print(Kd, LogKd)
return LogKd


# Variables here:
def BI_2_LogKd():
# threshold: High confident value to determine a detected
                                    binder
# PDZ_tot: PDZ concentration
threshold = 0.2
PDZ_tot = 4

# Retrieve concentracions of our PBMs and construct a
                                    dictionary
ConcPept = pd.read_csv('/Users/janepalp/Documents/PhD/bin/
                                    DataBases/Concentrations/
                                    dict_Conc_Pept.csv')
ColumGrouping = sorted(list(ConcPept.columns.values))
dictConc = dict(zip(ConcPept[ColumGrouping[1]], ConcPept[
                                    ColumGrouping[0]]))

# Extract the Binding Intensities from our PBM/PDZ interactions
BIs = pd.read_csv('/Users/janepalp/Documents/PhD/bin/DataBases/
                                    Concentrations/PDZList.csv')
ListPBMs = []

for i in range(1,len(sys.argv)):
df = BIs[['PDZ','BI_'+str(sys.argv[i])]]
ConcPept = float(dictConc[str(sys.argv[i])])

# Convert our Binding Intensities into Affinity using the
                                    conversion function made above
df['Log(Kd)_'+str(sys.argv[i])] = conversion(BIs['BI_'+str(sys.
                                    argv[i])], ConcPept, threshold,
                                     PDZ_tot)
df['Log(Kd)_'+str(sys.argv[i])] = np.where(np.isnan(df['BI_'+
                                    str(sys.argv[i])]) == True , np
                                    .nan, df['Log(Kd)_'+str(sys.
                                    argv[i])])


# Removing PDZ names from index
df = df.set_index('PDZ')
```

```
df.index.name = None

# Extracting min value of affinity
Kd_Min = (PDZ_tot - threshold*PDZ_tot) * (ConcPept - threshold*
                                PDZ_tot) / (threshold*PDZ_tot)

# Extract affinity data coming from orthogonal assay (FP) and
                                transform it into the -log(KD)
                                scale
NewPoints = pd.read_csv('/Users/janepalp/Documents/PhD/bin/
                                DataBases/Concentrations/
                                NewKDPoints.csv')

NewPoints = NewPoints[NewPoints.PBM == str(sys.argv[i])]

NewPoints['KD'] = np.where(NewPoints['KD'] > Kd_Min, Kd_Min,
                                NewPoints['KD'])

NewPoints['KD'] = -np.log10(NewPoints['KD']/1000000)

print(NewPoints)

# Search empty holdup PBM/PDZ interaction measurements that
                                were succesfully measured by FP
                                 and add it to our list
ColumGroupingNewPoints = sorted(list(NewPoints.columns.values))
print(ColumGroupingNewPoints)

NewPointsdict = dict(zip(NewPoints[ColumGroupingNewPoints[3]],
                                NewPoints[
                                ColumGroupingNewPoints[0]]))

print(NewPointsdict)

# Build the dataframe
df['Log(Kd)_'+str(sys.argv[i])].update(pd.Series(NewPointsdict)
                                )

if i == 1:
result = df
else:
result = pd.concat([result, df], axis=1, join_axes=[result.
                                index])
```

```
print(result)

return result


BI_2_LogKd()
```

16.2 Calculating and plotting the affinity vs specificity map

```
import numpy as np
from scipy.stats import entropy
from math import log, e
import pandas as pd
import matplotlib.pyplot as plt
from random import gauss
from mpl_toolkits.mplot3d import Axes3D  # noqa: F401 unused
                                    import

import sys
sys.path.append('/Users/janepalp/Documents/PhD/bin/pybin')
from BI2LogKd import BI_2_LogKd


# Extract the PDZ index
def PDZIndex_calculation(df):
y2 = df.values
slope = True
Ymiddle = (max(y)+min(y))/2
if Ymiddle in y2:
Yup = Ymiddle
Ydown = Ymiddle
slope = False
else:
Ydown = y2[y2 < Ymiddle].max()
Yup = y2[y2 > Ymiddle].min()


for n, i in enumerate(y):
if i == Ydown:
```

```
Xdown = n + 1
break


for n, i in enumerate(y):
if i == Yup:
Xup = n + 1


if slope != False:
m = (Yup-Ydown)/(Xup-Xdown)

PDZIndex = (Ymiddle - Ydown + (m*Xdown))/m

else:
PDZIndex = (Xup + Xdown)/2

PDZIndex = (PDZIndex/len(y2)) * 100

return PDZIndex,Ymiddle,Xup,Xdown,Yup,Ydown


# Plot the Affinity Vs Specificity map
def ScatterMap2D(x_scatter,y_scatter,PBMName):
fig, ax = plt.subplots()

Virus = ['HPV16-E6','HPV16-E6-Ph2','HPV16-E6-Pm2','HPV18-E6','
                                HPV35-E6','HTLV-TAX1','H5N1-NS1
                                ','NS5-WNV']
CloseVirus = ['NET1', 'VNGL2']
PTEN = ['PTEN-L11','PTEN-L13','PTEN-L11-KR','PTEN-L11-Ac']
RSK1 = ['RSK1', 'RSK1-Ph3', 'RSK1-Pm3']

FirstVirus = True
FirstCloseVirus = True
FirstPTEN = True
FirstRSK1 = True
Others = True

for i,m in enumerate(PBMName):

x = x_scatter[i]
y = y_scatter[i]
```

```
if m in Virus:
if FirstVirus == True:
ax.scatter(x, y, marker='1', color='red',label='HPV16')

FirstVirus = False
else:
ax.scatter(x, y, marker='1', color='red')

elif m in CloseVirus:
if FirstCloseVirus == True:
ax.scatter(x, y, marker='1', color='black', label='Close to
                                Virus')

FirstCloseVirus = False
else:
ax.scatter(x, y, marker='1', color='black')

elif m in PTEN:
if FirstPTEN == True:
ax.scatter(x, y, marker='1', color='blue', label='PTEN')

FirstPTEN = False
else:
ax.scatter(x, y, marker='1', color='blue')

elif m in RSK1:
if FirstRSK1 == True:
ax.scatter(x, y, marker='1', color='black', label='RSK1')

FirstRSK1 = False
else:
ax.scatter(x, y, marker='1', color='black')

else:
if Others == True:
ax.scatter(x, y, marker='1', color='black', label='Others')
Others = False
else:
ax.scatter(x, y, marker='1', color='black')



ax.set_ylabel('Max Affinity ($-log(K_{D})$)')
ax.set_xlabel('Half-PDZ')
```

```
plt.savefig('ScatterMap2D.png',dpi=300)



'''
Backbone starts here
'''

#Convert BI into -Log(KD) values from the previous script (
                                    section 16.1)
values = BI_2_LogKd()

df = values.filter(like='Log(Kd)_')
vmin = df.min().min()
vmax = df.max().max() + 0.1

# Extracting the max affinity and the PDZ Index for each PBM
All_MaxAffinities = []
All_HalfPDZ = []
PBMName = []


for j in df.columns:
PBM = j[8:]

df_data = df.sort_values(by=j, ascending=False)
y = df_data[j].values.tolist()

affinity = max(y)
PDZIndex,Ymiddle,Xup,Xdown,Yup,Ydown = PDZIndex_calculation(
                                    df_data[j])

plotted_Profile = Scatter_Profile(y,PDZIndex,Ymiddle,Xup,Xdown,
                                    Yup,Ydown,PBM,vmax)


All_PDZIndex.append(PDZIndex)
All_MaxAffinities.append(affinity)

PBMName.append(PBM)
```

```
result2D = ScatterMap2D(All_HalfPDZ,All_MaxAffinities,PBMName)


ScatterMappResults = pd.DataFrame(
{
  'Half-PDZ': All_HalfPDZ,
  'Entropy': All_entropies,
  'Max Affinities': All_MaxAffinities,
  'PBM': PBMName
})

ScatterMappResults.to_csv('ScatterMappResults.csv')

print(ScatterMappResults)
```

16.3 Affinity Profiles

```
  import matplotlib.pyplot as plt
  import numpy as np
  import random
  import sys
  sys.path.append('/Users/janepalp/Documents/PhD/bin/pybin')
  from BI2LogKd import BI_2_LogKd
  import matplotlib.gridspec as gridspec
  from mpl_toolkits.axes_grid1 import make_axes_locatable
  from brokenaxes import brokenaxes
  from matplotlib.gridspec import GridSpec
  import pandas as pd
  import math

  # Calculating the threshold for each Affinity Profile
  def Threshold(ConCPept):
  BI = 0.2
  PDZ_tot = 4

  Kd = (PDZ_tot - BI*PDZ_tot) * (ConCPept - BI*PDZ_tot) / (BI*
                                  PDZ_tot)
  thresholdPlot = -np.log10(Kd/1000000)

  return thresholdPlot
```

```
# Plot the Affinity Profile using breaks in the barplot
def plot_BreakBar(df,MinCon,MaxCon):
vmin = MinCon
vmax = MaxCon + 0.2

# Retrieve concentracions of our PBMs and construct a
                                dictionary
ConcPept = pd.read_csv('/Users/janepalp/Documents/PhD/bin/
                                DataBases/Concentrations/
                                dict_Conc_Pept.csv')
ColumGrouping = sorted(list(ConcPept.columns.values))
dictConc = dict(zip(ConcPept[ColumGrouping[1]], ConcPept[
                                ColumGrouping[0]]))


# Plotting conditions for the break bar
sps1 = GridSpec(len(list(df.columns)),1)
sps1.update(wspace=0, hspace=0.4)

fig = plt.figure(figsize=(12,15))

brokelimit = []
for col in df.columns:
searchlen = df[df[col] > MinCon]
brokelimit.append(len(searchlen))

limitXleft = max(brokelimit)

# Loop to plot all the break profiles in a vertical manner
                                one above the other. Comparison
                                 gets easier and axis will
                                uniformly adjusted
for col, j in zip(df.columns,range(0,len(df.columns))):

df = df.sort_values([col], ascending = False)
y = df[col].values
a = len(y) - 1
PDZ = np.arange(len(y))
PDZ = PDZ + 1


bax = brokenaxes(xlims=((-1,limitXleft+3),(260,len(y))),
                                subplot_spec=sps1[j,0])
```

```
bax.bar(PDZ, y, color='black')



thresholdHU = Threshold(float(dictConc[col[8:]]))
print(thresholdHU)

bax.plot([-1, limitXleft+3], [thresholdHU, thresholdHU],
                              linewidth=3, color='grey',
                              linestyle='--')
bax.plot([260,len(y)], [thresholdHU, thresholdHU], linewidth=
                              3, color='grey', linestyle='--'
                              )

bax.set_ylim(t2lim)
bax.set_title(col[8:], fontsize=24, loc='center')
bax.tick_params(labelsize=24)


if len(sys.argv) - 2 != j:
bax.set_xticklabels([])


# Text axis and saving the plot
fig.text(0.55, 0.05, 'PDZ rank', ha='center', size=24)
fig.text(0.11, 0.9, '-Log(K$_{{D}}$)', ha='center', size=24)
plt.savefig('AffinityProfile.png',dpi=300) #, orientation='
                              landscape')
plt.show()

#Convert BI into -Log(KD) values from the previous script (
                              section 16.1)
values = BI_2_LogKd()
values.to_csv('Data_Points.csv')
LogKDs = values.filter(like='Log(Kd)_')

# Extract the minimal and maximal values to adjust axis
                              uniformly
MinCon = LogKDs.min().max()
MaxCon = LogKDs.max().max()

LogKDs.to_csv('AffinityProfile.csv')

result = plot_BreakBar(LogKDs,MinCon,MaxCon)
```

Pau JANÉ PALLÍ

Quantification de affinités PBM/PDZ
et de leurs sites modulateurs
par des approches expérimentales et informatiques
à haut débit

Résumé

Ce travail a porté sur les domaines PDZ, une famille de domaines globulaires reconnaissant des motifs de liaison aux PDZ (appelés PBM, pour 'PDZ-Binding Motifs') généralement situés à l'extrémité C-terminale de leurs protéines partenaires. Les réseaux domaines-motifs sont souvent modulés par des modifications post-traductionnelles réversibles (PTM). Nous avons utilisé des PBM synthétiques simulant différentes conditions: motifs sauvages de diverses longueurs, acétylés, phosphorylés ou portant des mutations 'imitant' les PTM. Ces peptides ont été utilisés pour des études d'interaction à l'aide du test 'Hold-Up', un test développé à l'origine dans notre laboratoire. Nous avons évalué l'impact de diverses modifications des complexes PBM/PDZ, qui conduisent à un changement global de leur capacité de liaison du PDZ. Ces résultats fournissent des informations quantitatives sur l'effet biologique que de telles modifications pourraient avoir dans le contexte des protéines entières.

Mots-clés : PDZ, interaction domaine-motif, modifications post-traductionnelles, analyse quantitative d'interactomes.

Résumé en anglais

This thesis focuses on PDZ domains, a family of globular domains that bind to conserved PDZ-Binding Motifs (called henceforth PBMs) generally situated at the extreme C-terminus of their partner proteins. Domain-motif networks are often modulated by reversible post-translational modifications (PTMs). We used synthetized PBMs to reproduce different conditions, such as a wild-type, acetylation or phosphorylation, addition of extra exosites or residue mimication of PTM in the literature. These peptides were used for interaction studies using the holdup assay, an assay originally developed in our laboratory. We evaluated the impact of diverse modifications of the PBM/PDZ interactions, which led to a global change of the PDZ-binding capability. These results provided quantitative information on the biological effects that such modifications may have in the context of full-length proteins.

Keywords : PDZ, domain-motif interaction, post-translational modifications, quantitative analysis of interactomes.